

Leakless DNA Strand Displacement Systems

Chris Thachuk¹, Erik Winfree¹, and David Soloveichik^{2,3}(✉)

¹ California Institute of Technology, Pasadena, USA

² University of California, San Francisco, San Francisco, USA

³ The University of Texas at Austin, Austin, USA

david.soloveichik@utexas.edu

Abstract. While current experimental demonstrations have been limited to small computational tasks, DNA strand displacement systems (DSD systems) [25] hold promise for sophisticated information processing within chemical or biological environments. A DSD system encodes designed reactions that are facilitated by three-way or four-way toehold-mediated strand displacement. However, such systems are capable of spurious displacement events that lead to *leak*: incorrect signal production. We have identified sources of leak pathways in typical existing DSD schemes that rely on toehold sequestration and are susceptible to toeholdless strand displacement (i.e. displacement reactions that occur despite the absence of a toehold). Based on this understanding, we propose a simple, domain-level motif for the design of leak-resistant DSD systems. This motif forms the basis of a number of DSD schemes that do not rely on toehold sequestration alone to prevent spurious displacements. Spurious displacements are still possible in our systems, but require multiple, low probability events to occur. Our schemes can implement combinatorial Boolean logic formulas and can be extended to implement arbitrary chemical reaction networks.

1 Introduction

Although biological in origin, nucleic-acids have proven to be versatile materials for *de novo* engineered molecular systems. In particular, cascades of prescribed molecular events can be systematically constructed with so-called strand displacement reactions (DNA strand displacement, abbreviated as DSD). In a basic DSD cascade, one nucleic-acid molecule hybridizes with a partially double-stranded complement, releasing the original binding partner, which in turn triggers a downstream strand displacement event [17]. These interactions can be readily programmed by designing strands to have appropriate complementarity. Dynamic molecular systems like logic circuits [15], amplification schemes [23, 26], neural networks [16], as well as mechanical devices like motors [11] have all been experimentally realized [25]. DSD cascades can also emulate any chemical reaction network (CRN, system of chemical reaction equations obeying mass-action rate laws) [1, 4, 19]. By realizing an appropriate CRN, DSD systems can generate temporal patterns, perform signal processing, remember states, compute distributed algorithms, as well as other tasks that have been studied in the language of CRNs.

However, to move beyond relatively small proof-of-principle demonstrations, we must tackle an important issue — one that has up to now limited the scale of strand displacement systems. DSD systems have been observed to be susceptible to various levels of *leak*: the triggering of undesired strand displacement reactions.

A number of ideas have been proposed to combat leak. For example, designing sequences to have strong C-G bonds at the ends of helices decreases the rate of fraying, and thus impedes the toeless displacement responsible for leak (see Sect. 2). By further securing helix ends, “clamp” domains of 1 to 3 nucleotides can decrease certain kinds of unintended displacements (see Sect. 3) [15, 17, 23]. Another approach involves adding small quantities of “threshold gates” that preferentially consume leaked strands before they have a chance to interact downstream [15, 17]. The idea is that small leaks get neutralized, but when the desired displacement occurs, the threshold gates are saturated and the signal propagates. Although such thresholding can be effective in the context of digital on/off behavior, it is not fitting for analog or dynamical systems, where information is carried in the temporally varying amount of released signal strands. In particular, the existing leak mitigation options are insufficient for the implementations of CRNs. Not only are such systems strongly analog, but the large concentration differences between “fuel” (aka auxiliary) complexes and signal strands may result in the situation that the amount of leak is comparable to the amount of signal. Other leak reduction strategies include introducing Watson-Crick mismatches [10], and physically segregating different complexes [21].

We are interested in a systematic method capable of reducing leak to arbitrary desired limits. In the form of our argument, we are motivated by “proof-reading” in algorithmic self-assembly, where constructing the same pattern at larger scales (with increasing redundancy) in principle arbitrarily decreases the error rate [2, 22]. Similarly, we describe an ensemble of constructions with different levels of redundancy (parameter N).

The simplest non-trivial DSD operation is sequence translation: a cascade of strand displacement reactions that upon initiation by a strand with sequence X , results in the release of a strand with sequence Y , such that sequences X and Y are unrelated. Note that the output strand can contain some additional “left-over” domains from X as long as these are insufficient to trigger unintended displacement. A pair of translators $Y := W$ and $Y := X$ can be thought as a logical $Y = \text{OR}(W, X)$ operation, since either W or X is sufficient to produce Y . In the context of implementing CRNs, translators serve as unimolecular reactions. Thus we start by analyzing regular translators, as well as their “leakproof” counterparts. We later describe AND gates: DSD modules that produce output Y only when both inputs W and X are present. Together with translators, such AND gates are sufficient to implement “dual-rail” Boolean formulas (where abstract signal $X = 1$ is indicated by the presence of active strand $X1$ and absence of active strand $X0$, and vice versa) [17]. At the end of this paper, we discuss the generalization to leakless CRNs.

We argue that with the smallest non-trivial redundancy parameter ($N = 2$), the leakless translator (“double-long domain” (DLD) scheme) exhibits signifi-

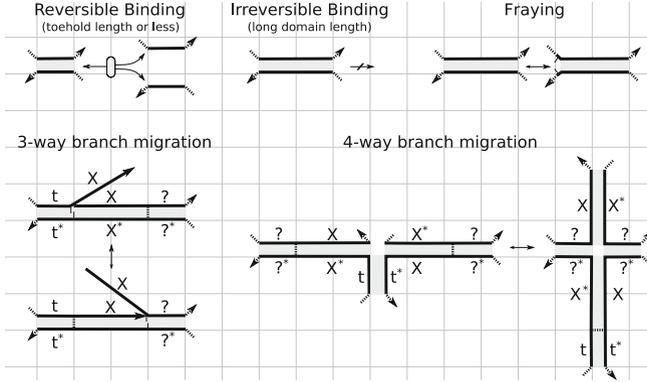


Fig. 1. DNA strand displacement events and conventions.

cantly less leak than a regular translator (“single-long domain” (SLD) scheme). Specifically we show a significant improvement at the usual experimental concentrations (e.g. 100 nM, where the leak is expected to be about five orders of magnitude less), as well as analyze the leak’s concentration scaling. If the concentrations decrease by a factor of α , the ratio of the leak of the DLD scheme compared to the SLD scheme is expected to decrease by a factor of α . The rates of the intended reactions are not significantly different for the SLD and DLD translator schemes. To estimate specific leak rates, and to verify that we have not missed any substantial leak pathways, we rely on an automated strand displacement reaction enumerator [9]. The scaling of leak with concentration is obtained analytically.

We then consider more generally translators and AND gates with arbitrary redundancy parameter N . Using a combinatorial argument, we prove that leak requires the joining of N initially separate complexes. We use this to argue that *even at thermodynamic equilibrium*, the amount of activated reporter (i.e. net leak) decreases exponentially with N . Importantly, the thermodynamic argument does not make strong assumptions on the types of reactions possible, or the types of admissible structures. For example, leaks that result in pseudoknots are not excluded. Our proof applies directly to a hypothetical experiment in which a translator or AND gate is composed with a reporter complex (which reports on the output via a fluorescence signal). However, more work is needed to extend this proof to arbitrary composition of leakless gates in networks.

2 Preliminary

We briefly summarize the DNA strand displacement conventions adopted in this paper, which are illustrated in Fig. 1. A detailed overview of these systems, including 3-way and 4-way DNA strand displacement can be found elsewhere [5, 25]. We study strand displacement systems at the domain level of abstraction. Domains have a defined length, and are an abstract representation

of sequences of that length. We define X^* to be the complement of domain X . Complementary domains can bind while non-complementary domains cannot. Complementary domains bound only by a single grid unit in length, called *toehold domains*, are reversibly bound (i.e., they can spontaneously disassociate). Complementary domains bound by at least two grid units in length, called *long domains*, are considered irreversibly bound (in reality long domains can be arbitrarily longer than the toehold domains – our graphical notation is just for convenience). Valid strand displacement events include ends of bound domains *fraying* (exposing base pairs that can bind to other domains), 3-way branch migration and 4-way branch migration. Binding of complementary domains can lead to two complexes combining into one, while branch migration (possibly followed by toehold unbinding) can lead to one complex separating into two. Unless otherwise qualified, the term *strand displacement* refers to the binding of a toehold domain followed by 3-way branch migration of the neighboring long domain (in other words, toehold mediated strand displacement). We use the term *toeless* strand displacement to refer to 3-way branch migration that is not preceded by binding of toehold-length domains. Rather it mechanistically occurs when double stranded long domains fray, followed by binding of a complementary invading strand to the momentarily opened bases and subsequent branch migration.

For the DNA strand displacement schemes proposed in this paper, we adopt the convention that each long domain X_i can be decomposed into a number of one grid unit parts labelled X_{ia} , X_{ib} , X_{ic} , etc. We use the subscript to denote when a domain is a proper suffix or prefix of X_i . For example, domain X_{iab} consists of the concatenation of X_{ia} and X_{ib} . Note that since X_{iab} is two grid units, it is itself a long domain, and thus irreversibly bound. By standard convention, also followed here, the molecules initially present in a strand displacement system can be classified as either *signal strands* or *fuel complexes*. Signal strands, or signals, propagate information through a system; these are typically at relatively low concentrations. Fuel complexes, typically held at a higher concentration, facilitate reactions by consuming input signals and producing output signals, via a sequence of strand displacement events and intermediate molecules. This process is initiated when input signals displace other strands on a fuel complex. As part of a multistranded complex, we say a signal strand is *sequestered* (or inactive) if it is unable to displace other strands. *External* signal strands carry information from component to component, and *intermediate* signal molecules (strands or complexes) carry information within a single component (e.g. between different fuel molecules). During strand displacement *waste* species can be created and are considered inert.

3 The Single Long Domain (SLD) Motif

Leak in typical DSD systems occurs when the following condition holds: *The unbound part of strand A shares a long domain with the bound part of strand B, but strand A is not currently supposed to displace B.* In the SLD motif, every sequestered strand is bound to a complex by at most one long domain

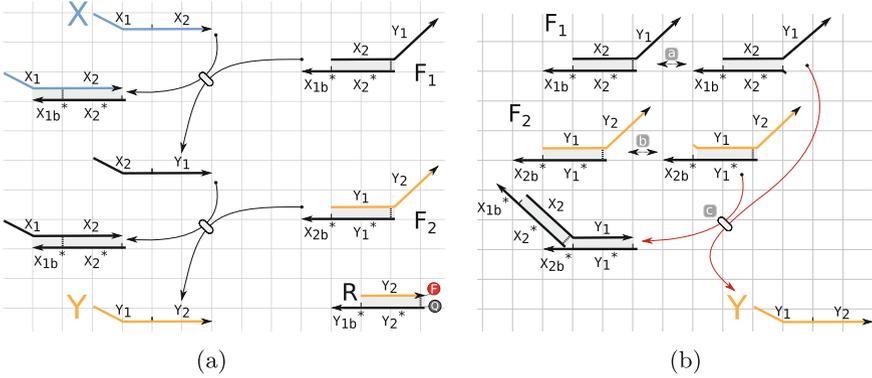


Fig. 2. A typical SLD implementation scheme of the translator $Y := X$ that uses *clamp* domains to help combat *leak*. F_1 and F_2 are fuel complexes. The output signal Y is sequestered on F_2 and should only be displaced in the presence of input X . R is a downstream reporter complex that is designed to interact with output signal Y . (a) The intended pathway when X is present consisting of two strand displacement reactions. (b) A leak pathway when X is not present: (reaction *a*) fraying of the clamp, (reaction *b*) fraying of the Y_1 long domain, (reaction *c*) toeless strand displacement via the fleetingly exposed 1-nt toehold at the left of Y_1 . When Y is spuriously produced it can successfully interact with downstream reporter complex, R .

and one toehold. Thus when the above condition holds, strand A can toelessly displace strand B . Although this rate is significantly slower than properly toehold mediated strand displacement, it nonetheless occurs at a non-negligible rate for relevant reaction regimes.

Consider a typical translator in a hypothetical experiment in which it is composed with a downstream reporter (Fig. 2a). The reporter emits an experimentally measurable fluorescence signal when the fluorophore (F) is dissociated from the quencher (Q). More generally, the reporter represents downstream complexes which receive input from this translator in a larger circuit. The intended pathway when input X is present is illustrated in Fig. 2a, consisting of two strand displacement reactions.

In the absence of input signal X it is still possible for output signal Y to be produced spuriously. Specifically, fuel F_1 can toelessly displace Y . Similarly, a toeless displacement interaction between F_2 and the reporter R can produce a fluorescence signal in the absence of X . Short *clamp* domains help mitigate these events, but they cannot eliminate them altogether due to fraying of the clamps. One of a number of leak pathways of the translator is illustrated in Fig. 2b.

Enumerator Analysis of the Leak. To systematically analyze the kinetics of the leak pathways we employed an automatic enumerator of strand displacement reactions which can be configured to capture 3-way and 4-way branch migration, toeless and remote toehold displacement, and cooperative hybridization [8,9].

(Since the enumerator ignores interactions forming pseudoknotted structures, and prunes reactions based on a number of assumptions, there could be plausible leak pathways that are not enumerated.) The enumerator input files for systems in this paper are included in the standard release of the enumerator [8]. Using size 5 toeholds, size 15 long domains, and size 2 clamps, the enumerator computes that the net leak rate between F_1 and F_2 resulting in free Y strand is $58 \text{ M}^{-1} \text{ s}^{-1} [F_1] \cdot [F_2]$. Additionally, there is also a leak between F_2 and reporter R that occurs with rate $13 \text{ M}^{-1} \text{ s}^{-1} [F_2] \cdot [R]$. For comparison, the intended path in the presence of X consists of three strand displacement reactions which the enumerator predicts occur with rate constants between $4 \cdot 10^5 \text{ M}^{-1} \text{ s}^{-1}$ and $8 \cdot 10^5 \text{ M}^{-1} \text{ s}^{-1}$ using the same parameters. The enumerator’s rate constant predictions are order-of-magnitude plausible compared to experimentally measured values of $1.4 \text{ M}^{-1} \text{ s}^{-1}$ and $9.6 \cdot 10^5 \text{ M}^{-1} \text{ s}^{-1}$ for toeless strand displacement and displacement via length-5 toeholds, respectively [27].

4 The Double Long Domain (DLD) Motif

The double long domain (DLD) motif dictates that sequestered signal strands are necessarily bound by at least two consecutive long domains. Schemes that use the DLD motif can be designed to satisfy the following *DLD scheme invariant*: *If a strand A is not intended to displace another strand B then any consecutive, unbound sequence of A differs from the bound sequence of B by at least one long domain (i.e. two grid units in our diagrams)*. In other words: there is never sufficient unbound sequence on a single strand to displace another that should not be displaced. This contrasts with typical SLD schemes that rely solely on the absence of open toehold domains to prevent certain reactions.

A DLD translator implementation is shown in Fig. 3a. In the absence of input signal X , it is not possible to fully displace output signal Y (without breaking bonds between long domains). The main leak pathway rather involves domains Y_1 and Y_2 becoming transiently unbound on the same strand due to the interaction of F_1 and F_2 . So in the absence of input signal X , reaching a state where the reporter is triggered requires multiple low probability and quickly reversible events to occur, and then not undo before the reporter has a chance to interact.

One possible leak pathway is illustrated in Fig. 3b. First, (reaction a) the clamp of F_2 must fray, then (reaction b) the bound domain Y_1 must fray, (reaction c) fuel F_1 can now form a first base pair, and then (reaction d) proceed to toelessly displace domain Y_1 with 3-way branch migration. At this point consecutive domains Y_1 and Y_2 are open and available to react with the reporter (i.e. signal Y is active). Reactions e and f show two other states in which both domains of the signal Y are active. In reaction e domain X_{2bc} on the signal strand can become bound to the invading complex via “open toehold” 4-way branch migration, which requires initiation by a slow loop-closing event [5]. In reaction f domain X_{2a} induces 3-way branch migration. Importantly, all these reactions are reversible and the states in which Y is active are transient. Then the triggering of the reporter requires catching the fuels in such an active Y state.

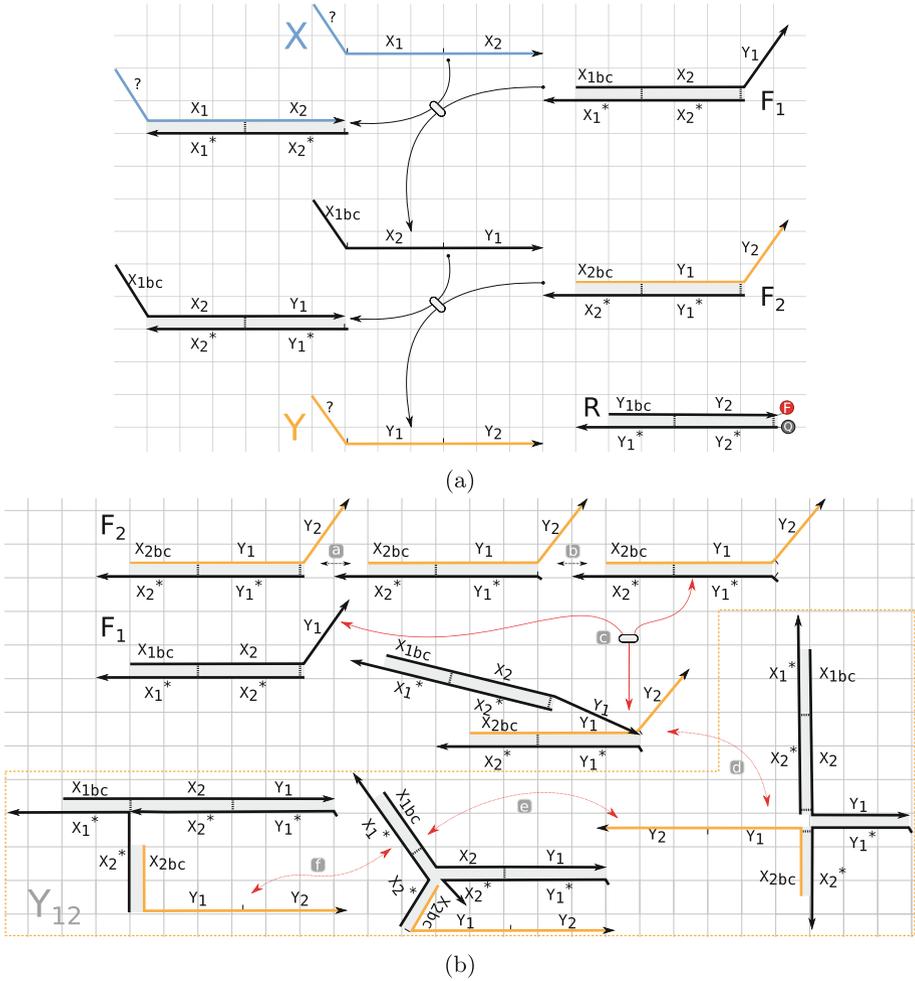


Fig. 3. An implementation of the translator $Y := X$ that uses the DLD motif. (a) The intended pathway when X is present consisting of two strand displacement reactions. R is a downstream reporter complex designed to interact with output signal Y . (b) Low probability leak pathway when input signal X is not present. See text for the description of reactions $a - f$. Transient complexes that have a functional Y signal and can successfully interact with downstream reporter complex, R , are shown outlined in yellow.

Enumerator Analysis of the Leak. The full DLD translator plus reporter system is too large for the current version of the enumerator to analyze at the necessary level of detail for evaluating leaks. As a result we decomposed the enumeration of the leak into two sub-problems: reversibly generating a complex with Y_1 and Y_2 domains open, and then the reaction of this complex with the reporter.

The enumerator reports that the net rate at which F_1 and F_2 react to produce a complex (call it ER) capable of reacting with the reporter is $3 \cdot 10^{-6} \text{ M}^{-1} \text{ s}^{-1} \cdot [F_1] \cdot [F_2]$. The reverse reaction occurs with rate constant $4.7 \cdot 10^{-4} \text{ s}^{-1}$. The reaction with the reporter is predicted to occur at $5.7 \cdot 10^5 \text{ M}^{-1} \text{ s}^{-1} \cdot [ER] \cdot [R]$. Since we can upper bound the concentration of complex ER by its equilibrium value, the overall rate at which leak is produced can be estimated to be:

$$3400 \text{ M}^{-2} \text{ s}^{-1} \cdot [F_1] \cdot [F_2] \cdot [R].$$

The enumerator also confirms that no interactions between the reporter and either fuel individually can result in the separation of the two reporter strands.

Both the existence (or non-existence) of reaction pathways and their rates will depend on the the assumptions used by the enumerator.¹ We adjusted enumerator parameters to strike a balance between ensuring that potential leak pathways were explored and yet combinatorial complexity remained tractable – which, on top of the uncertainty in the rate formulas, suggests that enumerator results should be regarded as provisional.

Comparing SLD and DLD Leak Rates. For typical 100 nM concentrations of fuels and reporter, SLD leak rate is roughly five orders of magnitude larger than DLD leak rate. Further, consider how the leak scales with the concentrations of the fuel complexes and the reporter. The SLD leak is a product of two concentrations, while the DLD leak scales as a product of three — the DLD leak effectively acts as a trimolecular reaction. The intended reaction pathway is bimolecular for both SLD and DLD schemes. Since the leak pathway is bimolecular for the SLD scheme, decreasing or increasing concentrations should not change the ratio of leak to intended rates. However, for the DLD scheme, decreasing the concentrations should linearly decrease the ratio of leak to intended rates.

DLD AND Gates. Going beyond translators, the basic computational DSD primitive has historically been the AND gate. We give two distinct AND gate constructions that maintain the DLD scheme invariant. Both are AND gates in the sense that signal X is produced (i.e. domains X_1 and X_2 are unbound) if and only if both input signals A and B are present. They also have the property that if one input is missing, the other input is not permanently consumed. The first scheme is simpler, while the second scheme has the potential advantage of better scaling to low concentrations.

¹ These assumptions include the approximate rate formulas for domain-level steps such as hybridization, fraying, 3-way and 4-way branch migration. There are also parameters set by the user that control the potential combinatorial explosion of the enumeration process, such as the granularity of domains (dividing a domain into subdomains allows the enumerator to explore more potential leak pathways, but makes the combinatorics worse), and the relevant time scales (opening a long double-stranded domain is “too slow to consider” and will not be enumerated, while a branch migration pathway may be “too fast” for considering bimolecular interactions prior to the end point).

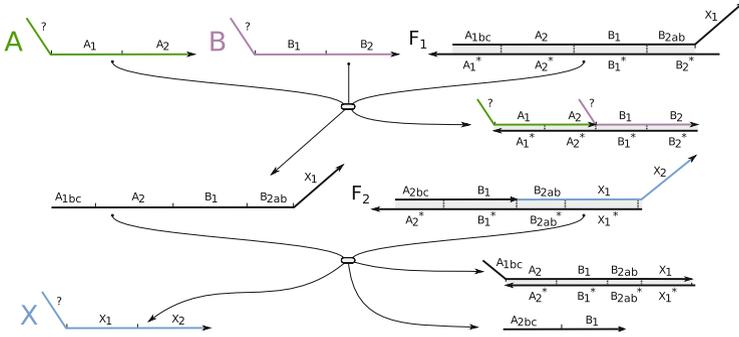


Fig. 4. An implementation of $X := \text{AND}(A, B)$ using the DLD motif and cooperative hybridization.

The scheme of Fig. 4 employs “cooperative” strand displacement [24]. Input signals A and B cooperate to displace the intermediate signal strand from fuel F_1 , which in turn displaces the output signal strand X bound to F_2 . The overall process is driven forward since the final state contains three additional bound toeholds and maintains the same number of complexes as the initial state. Note, however, that the cooperative step (input signals A and B displacing the intermediate signal from fuel F_1) is effectively trimolecular.² Thus, in contrast to the DLD translator, the intended displacement rate for the cooperative DLD AND gate (i.e. when both inputs are present) decreases just as quickly as the leak rate as concentrations are decreased.

The scheme of Fig. 5 employs “associative” strand displacement [3, 7, 13]. Two consecutive strand displacements (first with input A and then with input B) must occur on fuel complex F_1 to create a displacing complex with open domains A_1 and B_3 , capable of displacing the intermediate signal sequestered in F_2 . In other words these two strand displacement reactions “glue” (or “associate”) domains A_1 and B_3 together. The subsequent interaction with F_3 and F_4 is similar to a DLD translator, except the last step involves opening of a loop. The structure of F_4 is designed to ensure that another invariant is maintained: there is no signal strand that has open domains X_1 and X_2 but not X_3 . This invariant ensures that when the second reaction occurs in a downstream complex, it can only occur with the full X_1, X_2, X_3 signal strand, and does not become irreversibly blocked without properly gluing X_3 . The overall process is driven forward since the final state contains 4 additional bound toeholds and maintains the same number of complexes as the initial state. Although somewhat more complex, the associative scheme has the advantage that it lacks trimolecular

² Each partial displacement is reversible and quickly reaches a pseudo-equilibrium proportional to two concentrations (F_1 and an input). The second input then reacts, for an overall rate proportional to the product of $[F_1] \cdot [A] \cdot [B]$.

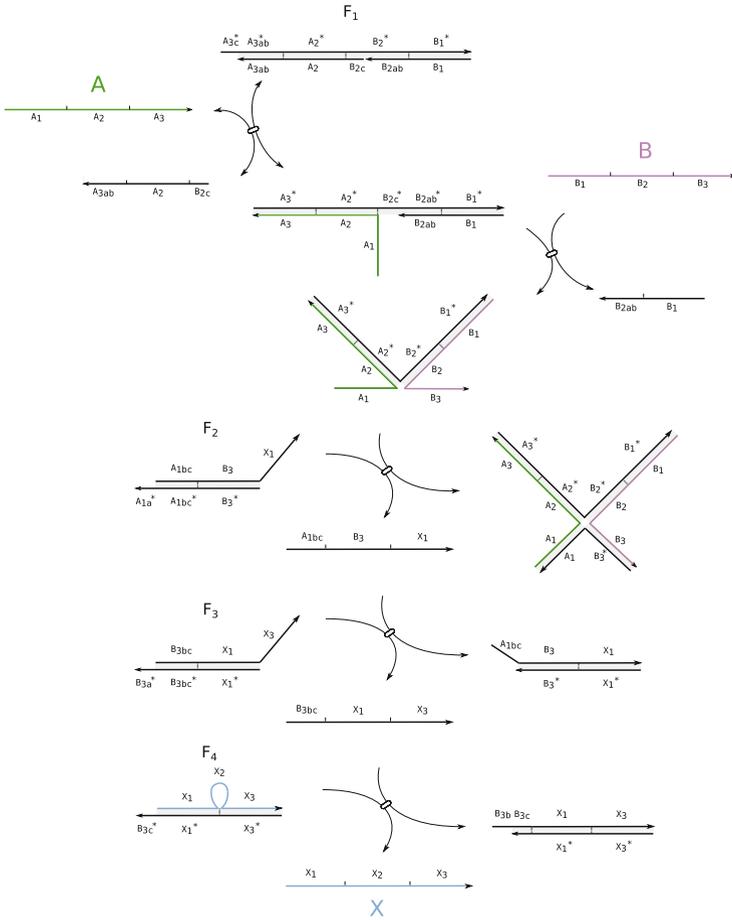


Fig. 5. An implementation of $X := \text{AND}(A, B)$ using the DLD motif and associative hybridization.

steps,³ and thus the associative AND gate should be faster than the cooperative AND gate in low concentration regimes.

The Triple Long Domain (TLD) Motif. The DLD motif can naturally be generalized to the triple long domain (TLD) motif, where each active signal is represented by three consecutive unbound long domains. The TLD translator is shown in Fig. 6b, and an example leak pathway is shown in Fig. 6c. In

³ Note that although the first reaction is reversible, the reverse reaction is bimolecular as opposed to unimolecular as is the case with the partial displacement by one input in the cooperative AND gate. Thus it is not as readily reversible, especially in low concentration regimes, and the associative gate avoids effectively trimolecular steps.

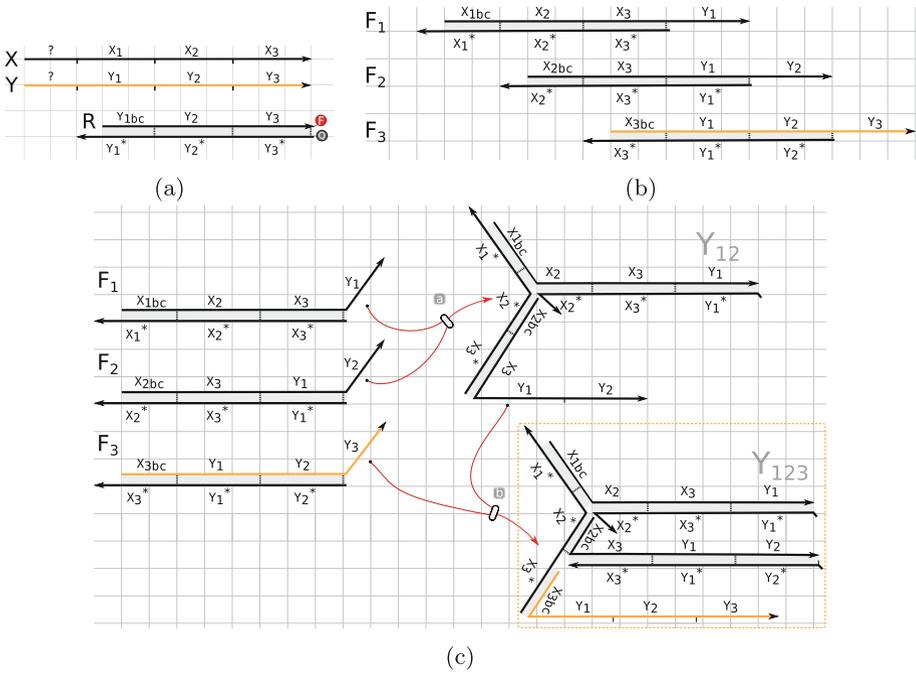


Fig. 6. An implementation of the translator $Y := X$ that uses the triple long domain (TLD) motif. (For simplicity, we show this scheme without clamps.) (a) Input signal and output signal format, and downstream reporter complex, R . (b) Fuel complexes. (c) A coarse grained leak pathway showing only bimolecular reactions.

this pathway, the reporter interacts with a transient structure (called Y_{123}) in which domains Y_1 , Y_2 and Y_3 are open. Intuitively, *all three* fuel complexes come together in a transient structure (Y_{123}) — a process which is more unlikely than two fuel complexes coming together in the DLD motif. Indeed, this idea is naturally generalized to signals represented by N long domains as described in the next section.

5 The N Long Domain (NLD) Motif

In this section, with the goal of decreasing the leak *arbitrarily*, we generalize the DLD motif to the NLD motif (“ N -long domains”, with arbitrary redundancy parameter N). We show constructions for translators and AND gates, and develop an asymptotic thermodynamic argument supporting the claim of leak reduction. In contrast to the main contributions of the previous section,

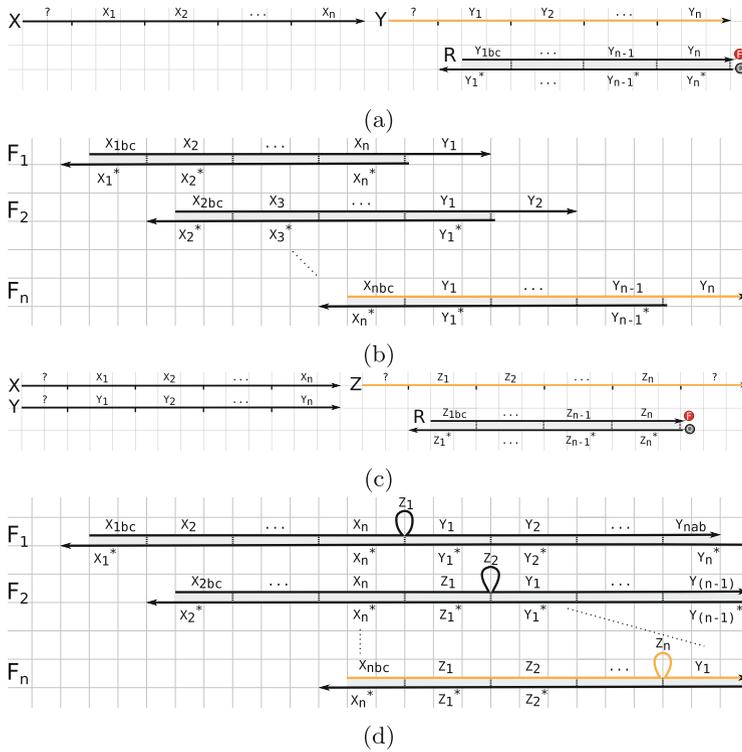


Fig. 7. Reaction implementations that use the N long domain (NLD) motif. (a) Input signal and output signal for translator $Y := X$, and downstream reporter complex, R . (b) Fuel complexes for translator $Y := X$. When input is present, at each step of the intended pathway, the intermediate signals that propagate through the cascade lose one X domain and gain one Y domain. (c) Input signal and output signal for $Z := \text{AND}(X, Y)$, and downstream reporter complex, R . (d) Fuel complexes for $Z := \text{AND}(X, Y)$. When both inputs are present, at each step of the intended pathway, the intermediate signals that propagate through the cascade lose one X domain and one Y domain and gain one Z domain (which is initially in a loop on the fuel complex).

the leak pathway does not necessarily require toeless displacement.⁴ Rather, the argument more generally reflects the decreasing likelihood of a sequence of thermodynamically unfavorable reactions. Since the argument is thermodynamic, it makes few assumptions on types of reactions and types of structures possible. The analysis even extends to leaks that involve pseudoknots.

⁴ For example, in the NLD AND gate described below, if input X is present, then there is a sequence of toehold-mediated reactions that can trigger the reporter. In particular, X displaces the top strand of F_1 from the left up to the hairpin, which in turn displaces the top strand of F_2 from the left up to the hairpin, and so forth. However, each of these reactions would quickly reverse because the partial displacement leaves each top strand attached. The associative hybridization AND gate of Fig. 5 also exhibits this behavior.

Figure 7b and d show the general NLD (i.e. redundancy N) translator and AND gates. The intended operation of the NLD translator when the input is present consists of a cascade of N strand displacement reactions. (The thermodynamic driving force is the formation of N new toehold bonds. The number of separate molecules before and after is the same, and thus there is no effective entropic driving force or penalty.) The intended operation of the AND gate when both inputs are present consists of cooperative strand displacement on F_1 , followed by a cascade of $N - 1$ toehold mediated strand displacement reactions. (The thermodynamic driving force is the formation of $N + 1$ new toehold bonds. While the number of separate molecules decreases by 1 as a result of cooperative strand displacement, this entropic penalty is overcome by the enthalpic gain of the toehold bonds.) With increasing N we expect the kinetics of the desired pathway to slow down quadratically with N .⁵

What about the leak pathway? Unlike for the DLD scheme, we do not examine the kinetics of the pathway. Rather we found we can bound the total amount of leak as we increase N , even if we allow the system to operate indefinitely (and reach thermodynamic equilibrium). In the rest of the section we develop an asymptotic thermodynamic argument examining the enthalpic and entropic driving and opposing forces for the leak in the general NLD scheme, leading to the conclusion that leak decreases exponentially with N .

5.1 Thermodynamics of Leak with Increasing N

The thermodynamic argument in this section relies on two assumptions about toehold and long domains. (1) *The binding of two separate molecules is thermodynamically unfavorable if only one new toehold bond forms ($\Delta G > 0$, bounded away from 0 independently of N).* (2) *The binding of long domains is effectively irreversible ($\Delta G \ll 0$).* The span of the regime where these assumptions are valid is determined by the concentrations involved, by the temperature, and by the length and sequence composition of the domains. Keeping these parameters fixed, we consider an asymptotic argument on increasing N .

In this section, we say *bottom strand* to refer to the strands that are illustrated on the bottom of their complex in Fig. 7b–d (and analogously *top strand*.) Similarly, we say a domain is a *bottom domain* if it occurs on a bottom strand and is therefore written as starred. Likewise *top domains* occur on top strands and are unstarred.

Imagine hypothetical experiments in which an NDL translator $Y := X$, or AND gate $Z := \text{AND}(X, Y)$, are paired with a reporter reading out their output. In the case of the translator, input X is absent; in the case of the AND gate, either input X or Y (or both) are absent. We do not assume a single molecule

⁵ The increasing length of the branch migration region is expected to lead to a linearly decreasing success probability per collision [20]. Thus each of the N strand displacement reactions slows down linearly with N . The time spent in the random walk of branch migration will increase quadratically, but will not be rate limiting for practical concentrations and values of N .

experiment; each present molecule can be present in many copies. Leak occurs when the bottom and top strands of a reporter molecule separate. We devote the following section (Sect. 5.2) to proving that producing a single activated reporter molecule necessarily involves the binding of N top fuel strands and N bottom fuel strands in a single complex together with the reporter bottom strand (assuming (2) above). Thus our intuition based on the TLD scheme is confirmed. To understand what this implies about the equilibrium amount of leak let us examine the enthalpic and entropic gains and costs.

Consider the thermodynamic equilibrium of the hypothetical translator and AND gate experiments described above. We argue that (assuming (1) above) the free energy difference between leaked and initial configurations increases linearly with N (with the leaked configurations unfavorable). What is the enthalpic driving force for leak? For the worst case, consider the AND gate when input X is present and input Y is absent: Note that there are N unbound bottom domains (X_1^*, \dots, X_N^*) that could potentially become bound in some leaked state if input X is present (all other single stranded domains have no complementary domains in the whole system). What is the entropic cost of leak? If any leaked configuration involves the binding of N fuel top strands and N fuel bottom strands in a single complex together with the reporter bottom strand, then the number of separate molecules decreases by N .

Assumption (1) implies that the formation of N new toehold bonds at the cost of decreasing the number of separate complexes by N is thermodynamically unfavorable, with the net increase in free energy proportional to N . This means that the amount of translator or AND gate in the leaked configuration decreases exponentially with N at thermodynamic equilibrium.

Does our thermodynamic argument about the individual translator and AND gate generalize to a circuit of these components? Unfortunately, the combinatorial argument in the next section fails to account for other types of strands with domains that overlap with the given translator or AND gate, as would occur in upstream and downstream components. Also note that in a circuit of these components, a leaked upstream signal may enable non-leak downstream reactions gaining N toehold bonds *per downstream component*. When we consider a fixed circuit, and increase N , this means that leak may result in αN new toehold bond formed, where $\alpha \geq 1$ depends on the circuit. To ensure that the amount of leak at the output of the whole circuit decreases exponentially with N , we need to be in a regime where forming α toehold bonds but losing entropy due to combining two separate molecules is still unfavorable — a strengthening of assumption (1).

5.2 Combinatorics: Leak Requires Binding of All Fuels

Translator. Recall the hypothetical experiment on the ND L translator paired with a downstream reporter (Fig. 7b). Suppose input X is absent, but the fluorophore and quencher are on separate molecules. We now show that assuming we do not decrease the number of bonded long domains (per assumption (2)

above), there are N fuel top strands and N fuel bottom strands bound in the same complex as the bottom strand of the leaked reporter.⁶

We ignore toeholds for the rest of the argument and assume for simplicity that the top strands are extended all the way to the left. Let C be the complex containing the bottom strand but not the top strand of the reporter. First we show that C must contain the same number of top fuel strands as bottom fuel strands. Consider top domain X_N and the complementary bottom domain X_N^* . Without the input, there is the same number of domains X_N as X_N^* in the system, and we say that X_N is *balanced*. Note that the initial state has the maximum number of long domain bonds. Thus, if we don't decrease the sum count of long domain bonds, every X_N must be bonded to a X_N^* and vice versa in C as well. Since each top fuel strand contains exactly one X_N domain, and every bottom fuel strand contains exactly one X_N^* domain, it must be that complex C contains the same number of top fuel strands as bottom fuel strands.

Let t and b be the total number of top domains and bottom domains in complex C , respectively. If C contains s top fuel strands (and therefore s bottom fuel strands), then the difference $t - b$ is: $s \cdot (N + 1) - s \cdot (N) - N$. The first term captures the contribution of $N + 1$ top domains in each top fuel strand. The second term captures the contribution of N bottom domains in each bottom fuel strand. The last term captures the contribution of the N bottom domains on the bottom reporter strand.

If we maximize the number of long domain bonds, it must be that complex C contains at least as many top domains as bottom domains (otherwise, we have an unbound bottom domain but we started with all bottom domains bound.) Thus, $s \cdot (N + 1) - s \cdot (N) - N \geq 0$. This implies that $s \geq N$. In other words, complex C contains N top fuel strands and N bottom fuel strands.

And Gate. Recall the hypothetical experiment on the NDL AND gate paired with a downstream reporter (Fig. 7d). Suppose one of the inputs is absent. We show that if the reporter top strand (fluorophore) is not in the same complex as the reporter bottom strand (quencher), then the complex containing the reporter bottom strand contains N fuel complexes (assuming we do not decrease the number of bonded long domains).

Again let C be the complex containing the bottom reporter strand but not the top reporter strand. If either input X or input Y is absent, then either X_N or Y_1 domains are balanced (i.e. have the same number of top as bottom types). Every fuel top strand contains exactly one X_N and Y_1 domain and every fuel

⁶ Note that it is not enough to notice that each fuel complex has one top Y domain in excess and thus assume that to replace the top reporter strand requires all N fuel complexes. As we saw before, there are possible cascades between fuel complexes that need to be taken into account. To drive home the point, consider removing the leftmost X_1 and X_1^* domains from F_1 . Then we could swap the top strands on F_1 and F_2 without decreasing the number of long domains bound, and then F_1 will contain two open Y domains: Y_1 and Y_2 . In this case, only $N - 1$ fuel complexes are sufficient to replace the top reporter strand.

bottom strand contains exactly one X_N^* and Y_1^* domain. Thus if one of the inputs is absent, the complex C contains the same number of fuel top strands as fuel bottom strands.

Suppose input X is absent. Suppose C contains s top fuel strands (and therefore s bottom fuel strands). Let t be the total count of top X and Z domains in complex C , and let b be the total count of bottom X and Z domains in complex C . Note that the contribution of any fuel top strand (and respectively fuel bottom strand) to the difference $t - b$ is identical, and thus we can avoid considering exactly *which* fuel strands are in complex C . Since each fuel top strand contains a total of $N + 1$ of top X and Z domains, and each fuel bottom strand contains a total of N of bottom X and Z domains, the difference $t - b$ is: $s \cdot (N + 1) - s \cdot (N) - N$. The last term again is the contribution of the N bottom Z domains on bottom reporter strand. Since globally there is an excess of top Z domains (and a balanced amount of X domains), it must be that on complex C , the above difference is non-negative. In other words, $s \cdot (N + 1) - s \cdot (N) - N \geq 0$, which implies that $s \geq N$.

If, on the other hand, input Y is absent (and X is possibly present), we proceed as before but let t be the total count of top Y and Z domains in complex C , and let b be the total count of bottom Y and Z domains in complex C . Note that similarly each fuel top strand contains a total of $N + 1$ of top Y and Z domains, and each fuel bottom strand contains a total of N of bottom Y and Z domains. Since without the Y input, there is a balanced amount of Y domains, and as before an excess of top Z domains, we again must satisfy the same inequality $s \cdot (N + 1) - s \cdot (N) - N \geq 0$, which implies $s \geq N$.

6 CRNs Using the DLD Motif

Can general chemical reaction networks be emulated using leakless DNA strand displacement systems? As a proof of principle, we give two implementations for the canonical reaction $A + B \rightarrow X + Y$ using the DLD motif. We note that this reaction can emulate any bimolecular reaction with at most two reactants and at most two products — A or B or both can be declared as *fuel* species (whose concentration is assumed to be constant) for reactions with less than two reactant input signals, whereas X or Y or both can be declared as *waste* species (which are considered inert) for reactions with less than two product signals. Furthermore, higher order reactions, or bimolecular reactions with more than two products, can be emulated by a cascade of these canonical reactions.

How can we determine if these schemes are *correct*, even in the absence of leak? Several general approaches to this question have been explored for DSD implementations [6, 12, 18]. The task is divided into two parts: enumerating all the (non-leak) reactions between the molecules (either using an enumerator [9, 14] or formal proofs) followed by applying some notion of program equivalence between the original CRN and the implemented CRN. Unfortunately, none of these approaches is fully satisfactory when applied to the systems described here. For example, Lakin, Phillips, and Stefanovic have previously shown that any

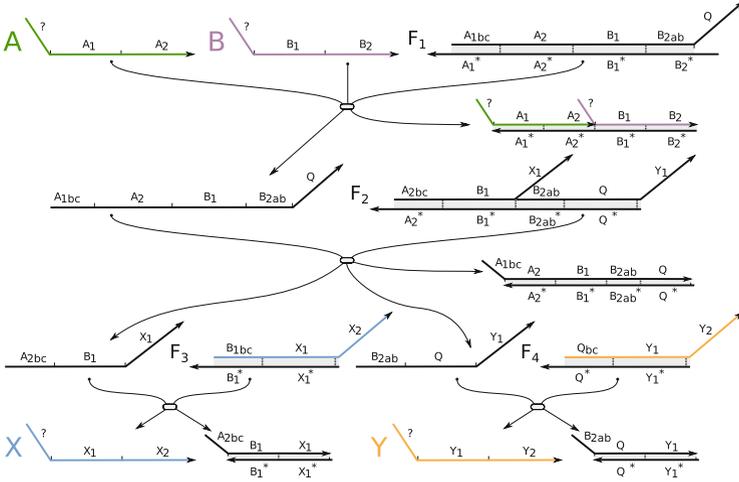


Fig. 8. An implementation of $A + B \rightarrow X + Y$ using the DLD motif and cooperative hybridization. Domain Q is unique to reactions that have $A + B$ as reactants.

DNA strand displacement encoding scheme satisfying certain modularity properties can be formally proven to correctly implement a chemical reaction network of interest, in terms of the notion of serializability [12]. Our proposed schemes do not satisfy the sufficient conditions to claim correctness with respect to serializability: both schemes have been optimized to share intermediates between different reaction encodings, and are therefore not strictly modular. However, both schemes do satisfy a key property identified in the serializability result: each DSD reaction cascade, called a *reaction encoding*, that emulates a formal reaction should be *transactional*. Informally, a reaction encoding $r_1, \dots, r_i, \dots, r_n$ that emulates the formal reaction $A + B \rightarrow X + Y$ is transactional if it can be partitioned into two parts, separated by its first irreversible reaction, r_i . The reactions r_1, \dots, r_{i-1} must be reversible and cannot produce signals representing X nor Y , and the signals representing formal species A and B must be consumed in r_1, \dots, r_i . The reactions r_i, \dots, r_n must produce the signals representing X and Y . The enumerator tool used in previous sections was applied to both scheme's encodings of $A + B \rightarrow X + Y$, and the resulting implementation CRNs were then easily verified to be transactional in this sense. At this point, we have not yet established that in implementations of larger CRNs, no unexpected crosstalk interactions between molecules would arise during the enumeration.

6.1 DLD Motif + Cooperative Hybridization

The DLD motif can be used in combination with cooperative hybridization [24] to implement arbitrary chemical reaction networks. The implementation of the

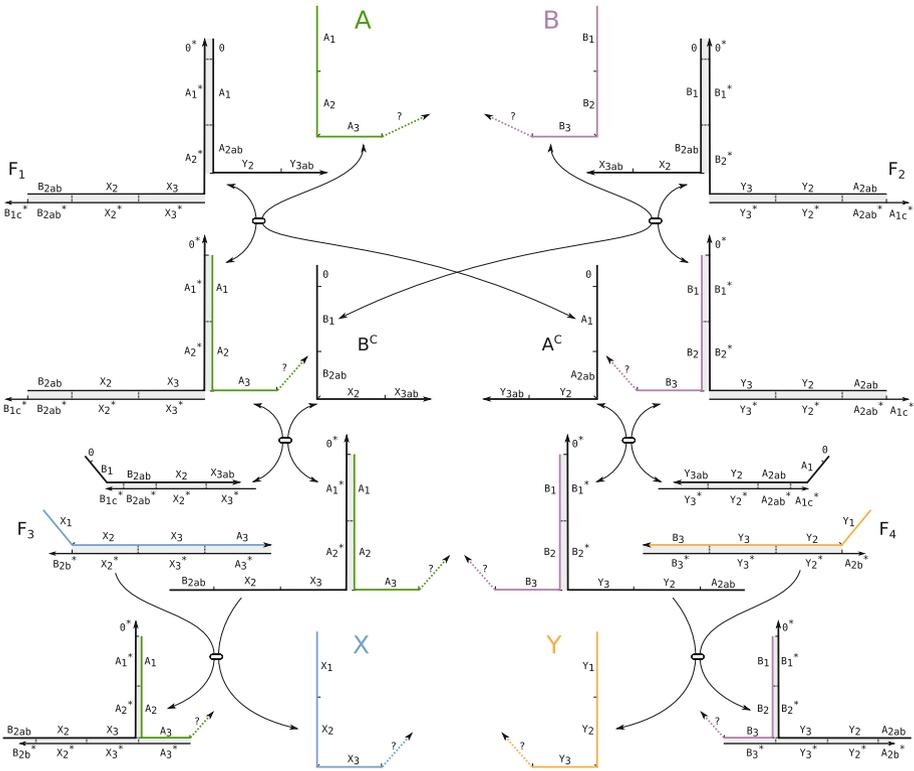


Fig. 9. An implementation of $A + B \rightarrow X + Y$ using the DLD motif and associative hybridization.

canonical reaction $A + B \rightarrow X + Y$ is illustrated in Fig. 8 and is a generalization of the cooperative AND gate implementation of Fig. 4. Input signals A and B cooperate to displace the long strand on fuel F_1 , which in turn displaces two intermediate signals on F_2 . One intermediate signal interacts with F_3 to produce output signal X , while the other interacts with F_4 to produce output signal Y . This process is driven forward since the final state contains four additional bound toeholds and maintains the same number of complexes as the initial state.

This reaction encoding is transactional. The process becomes irreversible only after the long strand on fuel F_1 is displaced by consuming both input signals. Thus, in the absence of one input the other cannot be irreversibly consumed. The remaining reactions produce the output species and render other species into waste.

6.2 DLD Motif + Associative Hybridization

Recall that cooperative strand displacement involves an initial effectively trimolecular step. Such trimolecular reactions can be prohibitively slow for low concentration systems. We next demonstrate that arbitrary bimolecular chemical reactions can in fact be implemented by utilizing the associative hybridization primitive. Although more complex, this scheme avoids the effectively trimolecular step and thus scales more favorably to low concentrations.

An implementation of the canonical reaction $A+B \rightarrow X+Y$ using associative strand displacement is illustrated in Fig. 9. Consider how the product X , initially bound to fuel complex F_3 , is produced. Two strand exchanges must occur on fuel complex F_1 to create a displacing complex with sufficient domains to displace the bound signal X . Firstly, input A exchanges with a “certificate” signal, A^C , which denotes that A has been consumed. The second exchange requires the certificate signal B^C , which denotes signal B has been consumed. After these strand exchanges, the domains X_2 , X_3 and A_3 have been “glued” together (associated) and displacement of X can occur. Production of signal Y is symmetric, by design. This process is driven forward since the final state contains two additional bound toeholds and maintains the same number of complexes as the initial state.

This reaction encoding is transactional. The process becomes irreversible only when the first output signal is produced (either X or Y). An output signal can only be produced after both input signals have been consumed. Once one output signal is produced, the other must eventually be produced since no series of backward reactions can occur in order to erroneously produce an input signal. (Suppose X has been produced. Then A is irreversibly bound and so is B^C . However, B^C is the only strand that could displace B).

7 Conclusion

The problem of leak has frustrated efforts to build complex DSD systems. In this work we begin a systematic effort to design DSD domain-level logic to reduce leak. In contrast to a number of previous approaches which relied on sequence-based leak reduction strategies, or subtle tweaks on existing designs (e.g. by introducing clamps), we rely on domain level redundancy. By utilizing more long domains in active signals, and more sequential strand displacement reactions to produce the output, we can increase the number of consecutive unfortunate events necessary for leak.

We focus on translator components, AND gates, and the implementation of CRNs. Our schemes rely on well-established types of strand displacement reactions, with the more complex components utilizing cooperative or associative displacement. It is, however, a natural open question whether it is possible to implement an NLD (or even DLD) AND gates and CRNs without using cooperative or associative strand displacement.

We advance two types of arguments to affirm leak reduction. First, we obtain leak rates for specific constructions (DLD motif) using a domain level reaction

enumerator. Second, we develop an analytical argument based on thermodynamics showing that increasing redundancy exponentially reduces the leak.

The principle remaining open questions concern the composition of the described components into circuits and networks. We have taken care to ensure that inputs and output signals have compatible form. However, we have not proven that the “leakless” properties of the individual components is preserved under composition. Further, properly ensuring the correctness of our CRN schemes when multiple reactions are implemented, even ignoring leak, requires more sophisticated arguments.

Acknowledgments. The authors are supported by a Banting Fellowship (CT), NSF CCF/HCC Grant No. 1213127, NSF CCF Grant No. 1317694, and NIGMS Systems Biology Center grant P50 GM081879 (DS). We thank Boya Wang and Robert Machinek for helpful discussions.

References

1. Cardelli, L.: Two-domain DNA strand displacement. *Math. Struct. Comput. Sci.* **23**(02), 247–271 (2013)
2. Chen, H.-L., Goel, A.: Error free self-assembly using error prone tiles. In: Ferretti, C., Mauri, G., Zandron, C. (eds.) *DNA 2004*. LNCS, vol. 3384, pp. 62–75. Springer, Heidelberg (2005)
3. Chen, X.: Expanding the rule set of DNA circuitry with associative toehold activation. *J. Am. Chem. Soc.* **134**(1), 263–271 (2011)
4. Chen, Y.-J., Dalchau, N., Srinivas, N., Phillips, A., Cardelli, L., Soloveichik, D., Seelig, G.: Programmable chemical controllers made from DNA. *Nat. Nanotechnol.* **8**(10), 755–762 (2013)
5. Dabby, N.L.: Synthetic molecular machines for active self-assembly: prototype algorithms, designs, and experimental study. Ph.D thesis, California Institute of Technology (2013)
6. Dong, Q.: A bisimulation approach to verification of molecular implementations of formal chemical reaction networks. Master’s thesis, Stony Brook University (2012)
7. Genot, A.J., Bath, J., Turberfield, A.J.: Combinatorial displacement of DNA strands: application to matrix multiplication and weighted sums. *Angew. Chem. Int. Ed.* **52**(4), 1189–1192 (2013)
8. Grun, C., Sarma, K., Wolfe, B., Shin, S.W., Winfree, E.: The peppercorn enumerator. <http://www.dna.caltech.edu/Peppercorn/>
9. Grun, C., Sarma, K., Wolfe, B., Shin, S.W., Winfree, E.: A domain-level DNA strand displacement reaction enumerator allowing arbitrary non-pseudoknotted secondary structures. In: *Verification of Engineered Molecular Devices and Programs (VEMDP)* (2014). <http://arxiv.org/abs/1505.03738>
10. Jiang, Y.S., Bhadra, S., Li, B., Ellington, A.D.: Mismatches improve the performance of strand-displacement nucleic acid circuits. *Angew. Chem.* **126**(7), 1876–1879 (2014)
11. Krishnan, Y., Simmel, F.C.: Nucleic acid based molecular devices. *Angew. Chem. Int. Ed.* **50**(14), 3124–3156 (2011)
12. Lakin, M.R., Phillips, A., Stefanovic, D.: Modular verification of DNA strand displacement networks via serializability analysis. In: Soloveichik, D., Yurke, B. (eds.) *DNA 2013*. LNCS, vol. 8141, pp. 133–146. Springer, Heidelberg (2013)

13. Machinek, R.R.F.: Control and observation of DNA nanodevices. Ph.D thesis, University of Oxford (2014)
14. Phillips, A., Cardelli, L.: A programming language for composable DNA circuits. *J. R. Soc. Interface* **6**(Suppl 4), S419–S436 (2009)
15. Qian, L., Winfree, E.: Scaling up digital circuit computation with DNA strand displacement cascades. *Science* **332**(6034), 1196–1201 (2011)
16. Qian, L., Winfree, E., Bruck, J.: Neural network computation with DNA strand displacement cascades. *Nature* **475**(7356), 368–372 (2011)
17. Seelig, G., Soloveichik, D., Zhang, D.Y., Winfree, E.: Enzyme-free nucleic acid logic circuits. *Science* **314**(5805), 1585–1588 (2006)
18. Shin, S.W.: Compiling and verifying DNA-based chemical reaction network implementations. Master’s thesis, California Institute of Technology (2011)
19. Soloveichik, D., Seelig, G., Winfree, E.: DNA as a universal substrate for chemical kinetics. *Proc. Nat. Acad. Sci.* **107**(12), 5393–5398 (2010)
20. Srinivas, N., Ouldridge, T.E., Šulc, P., Schaeffer, J.M., Yurke, B., Louis, A.A., Doye, J.P.K., Winfree, E.: On the biophysics and kinetics of toehold-mediated DNA strand displacement. *Nucleic Acids Res.* **41**(22), 10641–10658 (2013)
21. Teichmann, M., Kopperger, E., Simmel, F.C.: Robustness of localized DNA strand displacement cascades. *ACS Nano* **8**(8), 8487–8496 (2014)
22. Winfree, E., Bekbolatov, R.: Proofreading tile sets: error correction for algorithmic self-assembly. In: Chen, J., Reif, J. (eds.) *DNA Computing*. LNCS, vol. 2943, pp. 126–144. Springer, Heidelberg (2004)
23. Yin, P., Choi, H.M.T., Calvert, C.R., Pierce, N.A.: Programming biomolecular self-assembly pathways. *Nature* **451**(7176), 318–322 (2008)
24. Zhang, D.Y.: Cooperative hybridization of oligonucleotides. *J. Am. Chem. Soc.* **133**(4), 1077–1086 (2010)
25. Zhang, D.Y., Seelig, G.: Dynamic DNA nanotechnology using strand-displacement reactions. *Nat. Chem.* **3**(2), 103–113 (2011)
26. Zhang, D.Y., Turberfield, A.J., Yurke, B., Winfree, E.: Engineering entropy-driven reactions and networks catalyzed by DNA. *Science* **318**(5853), 1121–1125 (2007)
27. Zhang, D.Y., Winfree, E.: Control of DNA strand displacement kinetics using toehold exchange. *J. Am. Chem. Soc.* **131**(47), 17303–17314 (2009)