A simple DNA gate motif for synthesizing large-scale circuits

Lulu Qian

Bioengineering, California Institute of Technology, Pasadena, CA 91125, USA
`lulu@dna.caltech.edu`

Abstract. *The prospects of programming molecular systems to perform complex autonomous tasks has motivated research into the design of synthetic biochemical circuits. Of particular interest to us are cell-free nucleic acid systems that exploit non-covalent hybridization and strand displacement reactions to create cascades that implement digital and analog circuits. To date, circuits involving at most tens of gates have been demonstrated experimentally. Here, we propose a DNA catalytic gate architecture that we believe is suitable for practical synthesis of large-scale circuits involving possibly thousands of gates. We will review the current status of experimental implementations, suggest their future prospects, and discuss progress on developing a compiler that converts VHDL programs to sequences for DNA molecules.*

Biography

Lulu Qian is a Postdoctoral Researcher in Erik Winfree's lab at Caltech. She received a B.S. in Biomedical Engineering from Southeast University of China in 2002, and a Ph.D. in Biochemistry and Molecular Biology from Shanghai Jiao Tong University in 2007. Her current research in bionanotechnology concerns programming DNA circuits and DNA self-assembly. Significant past accomplishments include the design and synthesis of a DNA nanostructure in the shape of a 150 nm China map that folds from a single long strand and over 200 short strands. This was the first independent implementation of Rothemund's scaffolded DNA origami technique. Other work includes DNA computation as well as the genetics of schizophrenia.

A simple DNA gate motif for synthesizing large-scale circuits (extended abstract)

Lulu Qian and Erik Winfree

Computer Science, Computation & Neural Systems, and Bioengineering
California Institute of Technology, Pasadena, CA 91125, USA

Introduction. The prospects of programming molecular systems to perform complex autonomous tasks has motivated research into the design of synthetic biochemical circuits. DNA-based catalysts [9, 11, 13] and logic gates [8, 5, 6] have been proposed as general-purpose components for synthesizing chemical circuits [3, 2] with applications in medical therapeutics, nanotechnology, and embedded control of chemical reactions. Here we introduce a DNA catalytic gate motif that is suitable for scaling up to large circuits. We use an abstract circuit formalism that aids the design and understanding of circuit behavior. Thanks to the modular design of the gate motif, sequence design is straightforward. Furthermore, we argue that parallel synthesis and preparation of circuit components should also be scalable. Our estimates suggest that circuits involving thousands of distinct gates may be possible.

A simple catalytic gate with a threshold. Each catalytic gate may be represented abstractly as a node with one or more wires connected to each of the left and right sides (Fig. 1a). Each wire will correspond to a single-stranded DNA molecule with a specific sequence, which may be absent (an inactive wire) or present at some concentration (an active wire). As will be described below, an active wire on one side of a gate can catalyze the the exchange of activity on wires on the other side. Furthermore, there may be a threshold that must be exceeded before catalysis occurs. When connected into circuits involving many interacting catalytic gates, complex circuit behavior can be designed. The basic concept is explained below; greater detail will be provided in the full version of this paper [4].

The DNA implementation of elementary gate illustrated in Fig. 1a is shown in Fig. 1bc. Each line ("wire") indicates a free signal strand, e.g. S_2TS_3 ("fuel"), S_1TS_3 ("output"), or S_3TS_4 ("input"), while the circle ("gate") indicates the base strand $T'S_3'T'$. The state of the network is indicated by writing the amounts of each species in appropriate locations: the (relative) concentrations of signals on the lines, and the (relative) concentrations of gate:signal complexes within the circle at the end of the corresponding line. The concentration of a threshold complex is written as a negative number in the location for the corresponding gate:signal complex for the signal it is absorbing. Thus, Fig. 1a specifies the catalytic gate and signal strands of Fig. 1b with respective concentrations $10x$, $10x$, and $1x$; the threshold gate absorbing S_3TS_4, shown in Fig. 1c, is present at $0.5x$; where $1x$ is a standard concentration, perhaps 10 nM. With these initial concentrations, the S_3TS_4 strand will first overcome the threshold and then act catalytically to facilitate the equilibration of S_2TS_3 and S_1TS_3 to approximately $5.1x$ each. (This level can be estimated by noting that by symmetry the two wires will have similar activity, and that the total concentration on each wire and within the gate remains constant.)

The reaction mechanism is a simplified version of the entropy-driven catalytic gate introduced in [13, 14]. The basic gate design is shown in Fig. 1b, which shows the signal strand S_2TS_3, the gate complex in which signal strand S_1TS_3 is bound by its right end to the gate base strand $T'S_3'T'$, and the signal strand S_3TS_4. Example sequences are shown in Fig. 1d. The fundamental operation is toehold exchange, a toehold-mediate strand displacement reaction that results in a right-side signal strand replacing a left-side signal strand. For example, signal strand S_3TS_4 can bind via its toehold domain T, unbiased three-strand branch migration within the S_3 recognition domain will lead to a state where either S_1TS_3 or S_3TS_4 is bound by no more than the short toehold domain T, and that signal strand will quickly fall off. Now, the gate base strand has its left toehold revealed, and its right toehold hidden. (For simplicity, we ignore the abortive attempts, and only consider those reactions in which strand replacement occurs.) Note, however, that a left-side signal strand (such as S_2TS_3) cannot directly replace a bound left-side signal strand (such as S_1TS_3 bound to base strand $T'S_3'T'$). It can only do so in the presence of a right-side signal strand (such as S_3TS_4) by an indirect sequence of events: the right-side signal strand displaces the bound left-side signal strand, and in turn the other left-side signal strand displaces the now-bound right-side strand. Thus, the right-side signal

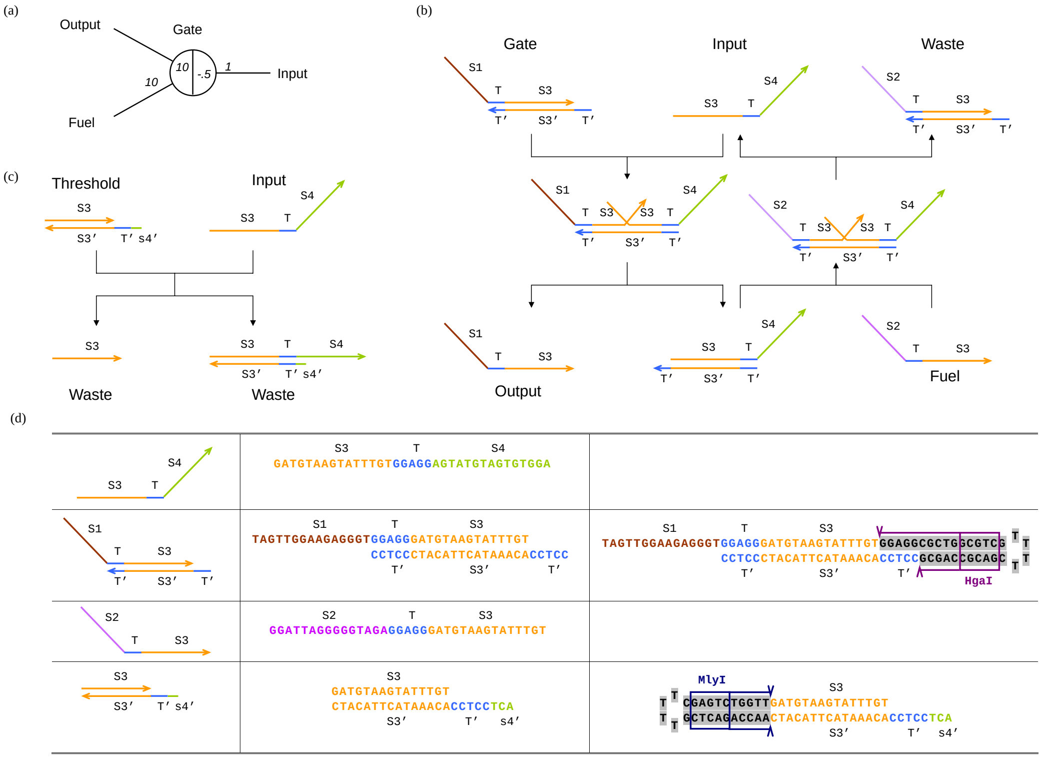

Fig. 1. The DNA motif for catalytic "seesaw" gates. **(a)** Abstract gate diagram. **(b)** The DNA gate motif and reaction mechanism. S_1, S_2, S_3, S_4 are the recognition domains; T is the toehold domain; T' is the Watson-Crick complement of T, etc. Arrowheads mark the $3'$ ends of strands. Black lines indicate elementary reactions via alternating back-and-forth toehold exchange replacements. **(c)** The threshold motif and reaction mechanism. Note that the toehold is extended by three nucleotides, providing an increased rate constant relative to the gate itself. **(d)** Example sequences for the gate and signal molecules, and hairpin precursor molecules that are cleaved by restriction enzymes to create functional gates.

strand has catalyzed the exchange of the two left-side signal strands, which cannot exchange in its absence. This back-and-forth motion is the inspiration for our name for this motif: "seesaw gates".

In the above discussion, recognition domains S_1, S_2, and S_4 played no active role in the mechanism. Rather, these domains allow the signal strands to interact with other gates and thus dictate connectivity within a circuit. For example, S_2TS_3 could also serve as a right-side signal strand for another gate with base strand $T'S_2'T'$, not shown here. However, while bound to base strand $T'S_3'T'$, signal strand S_1TS_3 cannot act similarly because its toehold domain is sequestered. Of course, if it is released into solution by a toehold exchange reaction, then it is free to act upon its target gate $T'S_1'T'$. In summary, each gate base strand consists of a single recognition domain identifying the gate, flanked by two toehold domains, only one of which is exposed at any given time; while each signal strand consists of two recognition domains identifying the two gates it connects, separated by a central toehold domain that is sequestered and thus inert when the signal strand is bound to a gate base strand.

In addition to catalysis, threshold behavior can be helpful for circuit function, for example for cleaning up after leaky reactions. Fig. 1c shows a threshold gate that serves as a competitive inhibitor of the signal strand S_3TS_4. Because of the slightly longer toehold on the threshold gate, the signal strand will react faster with the threshold gate than with the original gate. (Toehold-mediated strand displacement reactions rates depend exponentially on toehold length, for short toeholds [12].) Once the signal strand has reacted with threshold gate, it will never be released because all relevant toeholds are sequestered – effectively, inert waste

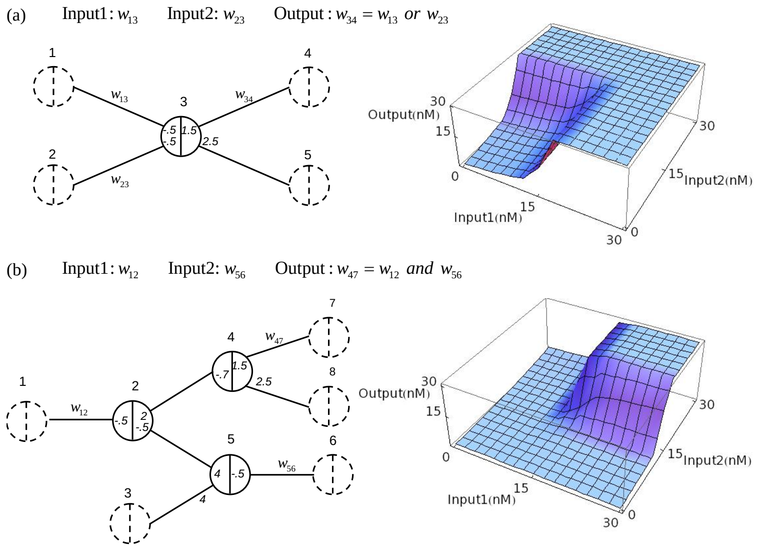

Fig. 2. The circuit diagram and input/output behavior of Boolean logic gates. **(a)** A two-input OR gate. **(b)** A two-input AND gate.

has been produced. Only after all the threshold gates have been used up, then the remaining signal strands can react (perhaps catalytically) with the original gate.

This consistent and modular motif makes systematic construction of circuits logically straightforward and further makes laboratory procedures for synthesizing gates and circuits plausible to carry out on a large scale.

Feedforward digital logic circuits. Digital logic has two compelling features for circuit construction: first, it has proven to be very expressive for the synthesis of a wide range of desired behaviors; and second, it is intrinsically robust to a variety of manufacturing and operational defects. It is well known that any boolean function can be computed (usually quite efficiently) by a well-designed feedforward circuit built from AND, OR, and NOT gates. Interestingly, with a small cost in circuit size and a minor change in representation, called "dual rail logic", AND and OR by themselves suffice. We will therefore provide constructions for these basic operations.

The OR gate is a simple extension of the basic catalyst shown in Fig. 1a in which there are two input wires, 13 and 23, that serve as catalysts for the release onto output wire 34 driven by exchange with the "power supply" wire 35. Under the assumption that the threshold gate reaction rate constant, k_1, is 10 times faster than the active gate reaction rate constant k_0, clean digital behavior is seen in the model (Fig. 2a). The AND gate is a little trickier, but follows from similar reasoning (Fig. 2b).

Discussion. Does our proposed seesaw gate motif live up to hopes and expectations as a DNA circuit component suitable for scaling up to large and complex circuits? Design of large feedforward digital circuits looks promising. At the highest level, abstract specifications for circuit function can be expressed concisely using existing hardware description languages such as Verilog [10, 1] and VHDL [7], then compiled down to a gate level netlist specifying elementary gates (AND, OR, NOT, NOR, NAND, XOR) and their connectivity. Thus, the sheer complexity of complex circuit design can be managed by off-the-shelf tools. The next step is compiling the digital logic netlist down to the seesaw gate circuit abstraction, using the constructions described above for dual rail logic. This is straightfoward if no circuit size optimizations are attempted. To achieve the final step of designing molecules, we must assign sequences to each gate base strand. For this purpose, a single large set of sufficiently distinct domain sequences would suffice for constructing any circuit containing up to the given number of seesaw gates.

However, the limitation to feedforward digital circuits – as opposed to sequential circuits containing buffers, flip-flops, resets, and clocks that orchestrate the re-use of circuit elements – is a concern. We do not know at this point whether this limitation is essential to the seesaw gate motif, or whether we have just not yet been clever enough to see how to do it. Similarly, we do not yet have a characterization of the class of analog dynamics that can be achieved, although it appears to be a rich space of behaviors. On the practical side, interfaces between DNA circuits and other chemical reactions will be necessary if DNA circuits are to serve as embedded controlers for molecular events.

Acknowledgements. The authors thank Dave Zhang, David Soloveichik, Georg Seelig, and Bernard Yurke for discussions and support. This work has been supported by NSF grant no. 0728703.

References

1. U. Golze. *VLSI Chip Design with the Hardware Description Language VERILOG*. Springer-Verlag, 1996.
2. J. Macdonald, Y. Li, M. Sutovic, H. Lederman, K. Pendri, W. Lu, B. L. Andrews, D. Stefanovic, and M. N. Stojanovic. Medium scale integration of molecular logic gates in an automaton. *Nano Letters*, 6:2598–2603, 2006.
3. R. Penchovsky and R. Breaker. Computational design and experimental validation of oligonucleotide -sensing allosteric ribozymes. *Nat. Biotechnol.*, 23(11):1424–1433, 2005.
4. L. Qian and E. Winfree. A simple DNA motif for synthesizing large-scale circuits. In *DNA Computing 14*, page to appear, Berlin Heidelberg, 2008. Springer-Verlag.
5. K. Sakamoto, H. Gouzu, K. Komiya, D. Kiga, , S. Yokoyama, T. Yokomori, and M. Hagiya. Molecular computation by DNA hairpin formation. *Science*, 288(5469):1223–1226, 2000.
6. G. Seelig, D. Soloveichik, D. Y. Zhang, and E. Winfree. Enzyme-free nucleic acid logic circuits. *Science*, 314:1585–1588, 2006.
7. M. Shahdad, R. Lipsett, E. Marschner, and K. Sheehan. Vhsic hardware description language. *IEEE Computer*, 18:94–103, 1985.
8. M. N. Stojanovic, T. E. Mitchell, and D. Stefanovic. Deoxyribozyme-based logic gates. *Journal of the American Chemical Society*, 124:3555–3561, 2002.
9. J. Tang and R. R. Breaker. Rational design of allosteric ribozymes. *Chem. Biol.*, 4:453–459, 1997.
10. D. E. Thomas and P. R. Moorby. *The Verilog Hardware Description Language*. Kluwer, 1991.
11. A. J. Turberfield, J. C. Mitchell, B. Yurke, A. P. Mills, Jr., M. I. Blakey, and F. C. Simmel. DNA fuel for free-running nanomachines. *Physical Review Letters*, 90(11):118102–1–4, 2003.
12. B. Yurke and A. P. Mills, Jr. Using DNA to power nanostructures. *Genetic Programming and Evolvable Machines*, 4:111–122, 2003.
13. D. Y. Zhang, A. J. Turberfield, B. Yurke, and E. Winfree. Engineering entropy-driven reactions and networks catalyzed by DNA. *Science*, 318:1121–1125, 2007.
14. D. Y. Zhang and E. Winfree. The toehold exchange primitive for controlling DNA hybridization reactions. Experimental work in progress.