Programmable Control of Nucleation for Algorithmic Self-Assembly
(Extended Abstract*)

Rebecca Schulman and Erik Winfree

California Institute of Technology, Pasadena, CA 91125, USA
{rebecka,winfree}@caltech.edu

Abstract. Algorithmic self-assembly has been proposed as a mechanism for autonomous DNA computation and for bottom-up fabrication of complex nanodevices. Whereas much previous work has investigated self-assembly programs using an abstract model of irreversible, errorless assembly, experimental studies as well as more sophisticated reversible kinetic models indicate that algorithmic self-assembly is subject to several kinds of errors. Previously, it was shown that proofreading tile sets can reduce the occurrence of mismatch and facet errors. Here, we introduce the zig-zag tile set, which can reduce the occurrence of spurious nucleation errors. The zig-zag tile set takes advantage of the fact that assemblies must reach a critical size before their growth becomes favorable. By using a zig-zag tile set of greater width, we can increase the critical size of spurious assemblies without increasing the critical size of correctly seeded assemblies, exponentially reducing the spurious nucleation rate. In combination with proofreading results, this result indicates that algorithmic self-assembly can be performed with low error rates without a significant reduction in assembly speed. Furthermore, our zig-zag boundaries suggest methods for exquisite detection of DNA strands and for the replication of inheritable information without the use of enzymes.

1 Introduction

Since Adleman first used DNA to perform a hard computation [1], researchers have explored the ability of biological molecules to carry out algorithms. Algorithmic self-assembly of DNA tiles [19] is Turing universal in theory, and tile sets for the construction of a variety of desired products have been suggested [12, 15, 3, 8]. An example of a structure that can be constructed using algorithmic self-assembly, a Sierpinski triangle, is shown in Figure 1. A simple generalization of this construction can be used to implement an arbitrary cellular automaton.

A tile program consists of labels for the sides of each of a set of square tiles, the strength with which each possible pair of labels binds, a designated seed tile, and a strength threshold τ. Polyomino tiles with labels on each unit-length of the perimeter can be used in addition to square tiles. The abstract tile-assembly

* A preprint of the full paper can be found at http://arxiv.org.

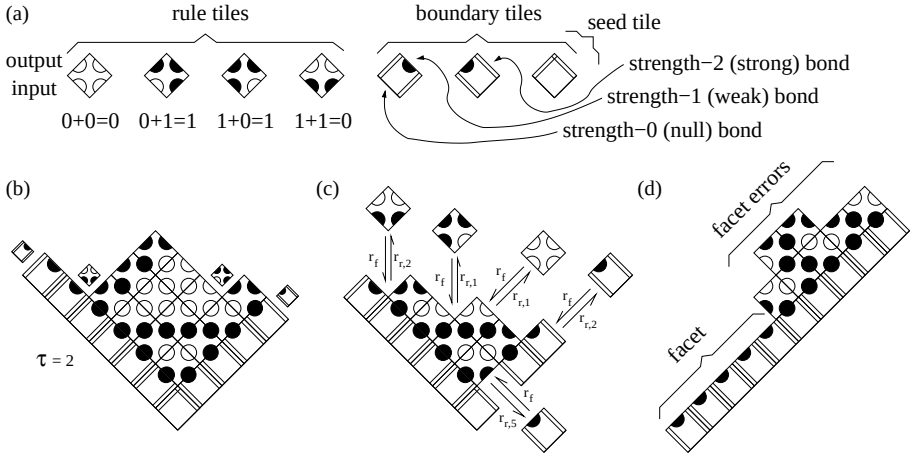

Fig. 1: **Tile assembly models (a)** The Sierpinski tile set. Tiles cannot be rotated. Whereas the rule tiles have sides that form strength-1 (weak, single-line) bonds, some sides on the boundary tiles form strength-2 (strong, double-line) bonds or strength-0 (null, thick-line) bonds. **(b)** Seeded growth of the Sierpinski tiles according to the aTAM at $\tau = 2$. The small tiles indicate the (only) four sites where growth can occur. At each location exactly one tile matches both exposed sides, so assembly results in a unique pattern. **(c)** For the growth of an isolated crystal under unchanging tile concentrations, the forward rate (association) is $r_f = k_f[tile] = k_f e^{-G_{mc}}$, while the reverse rate (dissociation) is $r_{r,b} = k_f e^{-bG_{se}}$ for a tile that makes bonds with total strength b. Parameters G_{mc} and G_{se} govern monomer tile concentration and sticky-end bond strength, respectively. A representative selection of possible events is shown here. The kTAM approximates the aTAM with threshold τ when $G_{mc} = \tau G_{se} - \epsilon$, in which case the same set of reactions are favorable or unfavorable in the two models. **(d)** An undesired assembly that can form due to unseeded growth of boundary tiles followed by facet errors.

model (aTAM) [20] describes the behavior of a tile program executed in the absence of assembly errors. Under the aTAM, assembly starts with designated tiles (usually just the seed tile) and proceeds by the addition of tiles at locations on the assembly's perimeter where the total strength of the connections between the tile and the assembly is greater than or equal to the threshold. Addition of tiles is irreversible but non-deterministic. Within the aTAM, it is possible to prove program correctness – that is, that growth from the seed tile always results in the unique desired structure. In this paper, unless stated otherwise mismatched labels will always bind with strength 0, bond strengths will be non-negative integers, and $\tau = 2$, as is the case for most prior work on algorithmic self-assembly.

In contrast to assembly in the aTAM, the assembly of DNA tiles is neither errorless nor irreversible. In practice a tile sometimes binds and sticks to a growing assembly even when the strength of the tile's attachments is smaller than

the threshold, an event called an *unfavorable attachment*. Unfavorable attachments can lead to three kinds of errors. First, an unfavorable attachment that only partially matches the adjacent tiles can occasionally become locked into place by succeeding attachments, forming a mismatch error. Second, a tile that attaches unfavorably to a facet and in turn allows the attachment of incorrect tiles nearby causes a facet error (Figure 1(d)). Lastly, a spurious nucleation error occurs when an assembly grows from a tile other than the designated seed tile.

Mismatch, facet and spurious nucleation errors have all been observed in algorithmic self-assembly experiments. In an experimental demonstration of the algorithmic self-assembly of a Sierpinski triangle [16], between 1% and 10% of tiles mismatched their neighbors, an effect that was attributed to both mismatch and facet errors. Furthermore, only a small fraction of the observed crystals were properly nucleated from seed molecules.

Why avoiding spurious nucleation can be difficult was clarified by experiments with just the boundary tiles shown in Figure 1 [17]. While the aTAM predicts that V-shaped assemblies should form, most observed assemblies were linear polymers without a seed tile. When all the tiles in the Sierpinski tile set were combined, most assemblies seen were spuriously nucleated rather than grown from a V-shaped boundary. The spuriously nucleated assemblies could have been produced either by linear boundaries growing wider due to multiple facet errors or by rule tiles assembling by themselves into crystals.

To theoretically study the rates at which these three kinds of errors occur, we need a model that includes energetically unfavorable events. The kinetic Tile Assembly Model (kTAM) [20] describes the dynamics of assembly according to an inclusive set of reversible chemical reactions: A tile can attach to an assembly anywhere that it makes even a weak bond, and any tile can dissociate from the assembly at a rate dependent on the total strength with which it adheres to the assembly (see Figure 1(c)). Several variants of the kTAM, reflecting different assumptions about how growth proceeds, have been developed [18]. In the kTAM as described in Figure 1(c), mismatch errors occur at least at a rate proportional to the square root of assembly speed [20]. Therefore, the mismatch error rate can be reduced by decreasing the temperature of the assembly reaction and/or decreasing the monomer concentration – but a 10-fold decrease in error rates requires a 100-fold decrease in assembly speed. A better solution to controlling mismatch and facet errors is to use "proofreading" tile sets that implement the same logic of an original tile set but assemble more robustly, reducing the error rates exponentially without significant slow-down [21, 6, 14].

In this paper we propose a method to control spurious nucleation errors without significant slow-down, exponentially reducing the rate at which assemblies without a seed tile grow large (unseeded growth), while maintaining the rate of growth that starts from a seed tile and proceeds roughly according to aTAM (seeded growth). To do so, a tile set must satisfy two conflicting constraints: When assembly begins from a seed tile, it must proceed quickly, whereas when assembly starts from a non-seed tile, it must go nowhere.

These two constraints are simultaneously satisfied by a phenomenon well-known to children who make rock candy: the nucleation of crystals in a super-

saturated solution. By cooling a solution slowly, it is possible to create a solution that has more solute dissolved than would be possible in standard conditions, called a supersaturated solution. Because of the interplay between surface and volume energy terms in a supersaturated solution, crystals smaller than a critical size will tend to shrink, whereas large crystals will grow. The seeded growth of crystals results from the mixing of a supersaturated solution with a small number of large crystals, called seed nuclei. When a seed nucleus is added to the solution, its growth is immediately favorable. Monomers attach to the seed, and a large crystal results.

To apply these principles for the control of nucleation in algorithmic crystals, it is enough to create a well-behaved boundary that plays the same role as the V-shaped boundary in Figure 1, but grows exclusively from a seed. Since rule tiles are not likely to spuriously nucleate on their own under optimal assembly conditions, once the well-behaved boundary has set up the correct initial information, algorithmic crystal growth will proceed correctly and without spurious side products. We use large seed tiles that serve the the same purpose as the large seed nuclei in the rock candy example. Tiles attach to the seed tile to produce a long boundary of predefined width. Because only full-width boundaries can grow by favorable attachments, without the seed tile there is a critical size barrier that prevents spurious nucleation – unlike the boundary tiles of Figure 1 for which the critical size is a single tile. The tile set that implements these ideas, called the zig-zag tiles, is described below.

2 The Zig-Zag Tile Set

The **zig-zag tile set** (see Figure 2(a)) of width k contains tiles that assemble to form a periodic ribbon of width k (see Figure 3(a)). Zig-zag tile sets can be constructed with any width $k \geq 2$. A zig-zag tile set includes a **top tile** and a **bottom tile**, each consisting of 2 horizontally connected square tiles. It also includes an L-shaped **seed tile** consisting of k vertically connected square tiles and a square tile horizontally connected to the bottom of the vertical connected tiles. In a zig-zag assembly, the top and bottom tiles stagger so that each column of tiles is connected to the columns on its right and left by either a top tile or a bottom tile. Each of the $k-2$ rows between the top and bottom tiles contains two unique **middle tiles** that alternate horizontally. Unique tiles in each row make assemblies of width less than k impossible to form without zero-strength attachments. Two tile types in each row enforce the staggering of the top and bottom tiles, which is essential for seeded growth to proceed quickly in a path that zig-zags up and down the width of the assembly. The seeded assembly path is shown in Figure 3(b).

The tile set is designed to operate in a physical regime where the attachment of a tile to another tile or assembly by two matching sides is energetically favorable, whereas an attachment by only one bond is energetically unfavorable. In this physical regime, algorithmic self-assembly is possible. In the aTAM, these conditions translate to growth with a threshold of 2. Growth from a seed tile occurs in a zig-zag shaped pattern; if assembly starts from a non-seed tile, no

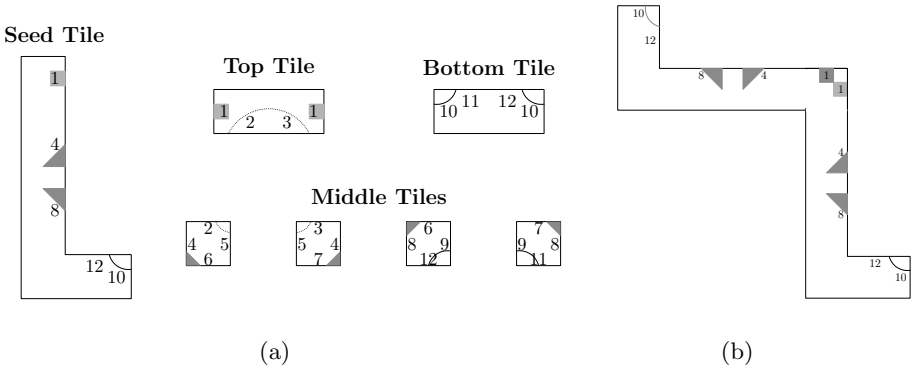

Fig. 2: **The zig-zag tile set. (a)** Each square, rectangle, and L shape represents a single tile. Excluding the seed tile, tiles are given unique bonds that determine where the fit in the assemble: each label has exactly one match on another tile. All correctly-matched bonds have strength 1. The geometric patterns shown on each tile identify them in subsequent figures. **(b)** The seed shown here, with appropriate tiles for vertical zig-zag growth, could be used instead of the L-shaped seed in (a) to form V-shaped assemblies.

growth occurs. In the kTAM, seeded growth occurs in the same pattern as in the aTAM, but there are also series of reactions that can cause spurious nucleation errors.

Spurious nucleation is a transition from assembly *melting*, where assemblies are more likely to fall apart than they are to get larger, to assembly *growth*, where each assembly step is energetically favorable. An assembly where melting and growth are both energetically favorable is a **critical nucleus**. Nucleation theory [9] predicts that the rate of nucleation is limited by the concentration of the critical nucleus, $[A_c]$. Since $[A_c] = e^{-\Delta G/kT}$, where ΔG is the free energy of a critical nucleus with respect to unbound tiles, linearly increasing the energy barrier, ΔG, exponentially decreases the rate of nucleation[1].

Since there is no energy barrier to seeded growth in the zig-zag tile set, growth from the seed tile is favorable. In contrast, there is an energy barrier for unseeded growth. The size of this barrier depends on the total concentration of critical nuclei. For a zig-zag tile set of width k, the critical nuclei are k tiles wide. Under the right conditions, the energy barrier depends linearly on the width of the critical nuclei, and thus the concentration of critical nuclei decreases exponentially with k. This argument is not rigorous, however, because unfortunately there are also many more kinds of critical nuclei for larger values of k. The rate of spurious nucleation is proportional to the sum of the concentrations of all these critical nuclei.

[1] In the kTAM, $\Delta G = (bG_{se} - nG_{mc})kT$ for an assembly involving n tiles and total bond strength b. k is Boltzmann's constant and T absolute temperature.

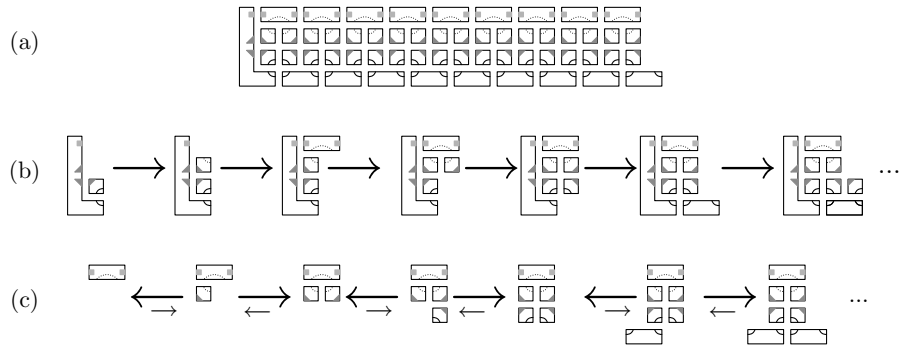

Fig. 3: **Zig-zag assembly. (a)** The structure formed by the zig-zag tile set according to aTAM with a threshold $\tau = 2$. **(b)** Seeded growth of a zig-zag tile set in the aTAM. The same growth pattern occurs reversibly in the kTAM with a threshold near $\tau = 2$. **(c)** A possible series of steps by which the tiles could spuriously nucleate in the kTAM. Under the conditions of interest, some steps are energetically favorable, but at least $k - 1$ must be unfavorable for a zig-zag tile set of width k. At this point, further growth is favorable.

To bound rather than explicitly calculate the rate of spurious nucleation, it is not necessary to calculate the rate of growth of each critical nucleus. Instead, we consider a set of subcritical assemblies, and we bound the total flux of assemblies leaving this set; it is assumed that (in the worst case) every assembly that leaves the set eventually becomes a long spuriously nucleated ribbon. This flux rate is a valid upper bound as long as single tiles are members of the set and spuriously nucleated assemblies are not.

For the zig-zag tile set of width k, we use the set of assemblies of width less than k. Because of the way the zig-zag tile set is designed, no assembly of width smaller than k can grow significantly longer without an unfavorable attachment. However, any assembly of width k can grow in a zig-zag fashion by exclusively favorable steps. Thus, we bound the rate of spurious nucleation by the rate at which assemblies of width $k - 1$ grow to a width of k.

To formally calculate such a bound, we make use of the kTAM as formulated for mass action dynamics [18], assuming constant tile concentrations. In mass-action dynamics the rate at which a reaction proceeds is proportional to a rate constant times the product of the concentrations of the reactants [10]. Given a tile set, we consider all possible accretion reactions: reactions either between two tiles or between a tile and an assembly in which tile concentrations remain constant. Since changes in the concentrations of unbound tiles are ignored[2], a reaction's rate is dependent on at most one changing concentration, so dynamics are linear and therefore easier to analyze. The concentration of tiles and the strength of formation are specified by the parameters G_{mc} and G_{se}. The concentration of each tile (except the seed tile) is $[tile] = e^{-G_{mc}}$ and the bond strength between

[2] In reality, tile concentrations will decrease as they are used, further decreasing the rate of spurious nucleation.

two matching tiles is G_{se}. The rate constant for each possible forward reaction is k_f, and the reverse rate constant for a reaction involving b bonds is $k_f e^{-bG_{se}}$. For a zig-zag tile set of width k, $J(k)$ is defined as the total rate of all addition reactions that exit the set of subcritical assemblies, i.e., reactions in which the reactant has width $k-1$ and the product has width k. We have proved the following theorem:

Theorem 1. *For a zig-zag tile set of width $k > 2$, if $G_{se} > 2(k \ln 2 + 1)$, $G_{mc} = 2G_{se} - \epsilon$, and $0 \leq \epsilon < \frac{1}{k}$, then, at all time points, $J(k) < 4k_f e^{\epsilon - kG_{se}}$.*

The proof appears in the full paper.

3 Discussion

3.1 Nucleation of Algorithmic Self-Assembly

Our original motivation for this work was to show that self-assembly programs that work in the aTAM, in which it is straightforward to design tile sets that algorithmically assemble any computationally defined structure, can also be made to work in the more realistic kTAM. Tiles sets that assemble correctly *via* unseeded growth in the aTAM with a threshold of $\tau = 1$ will assemble correctly in the kTAM under the right conditions. However, tile sets that are designed to assemble *via* seeded growth in the aTAM with a threshold $\tau = 2$ may fail in the kTAM because mismatch, facet and spurious nucleation errors occur. These problems are ameliorated in the limit of slow assembly speed [20]. Other work has described methods to control mismatch errors and facet errors without significant slowdown [21, 6, 14]. Here, we have developed a construction that corrects the last discrepancy, spurious nucleation errors, again without significant slowdown.

However, it remains to be formally proven that these constructions can be combined to control all types of errors simultaneously for any tile set of interest. No major difficulties are expected, in large part because mismatch and facet errors can both be controlled by a single mechanism [6] and the control of spurious nucleation errors works independently of this mechanism. Both methods work by transforming an original tile set which works in the aTAM at $\tau = 2$ into a new (typically larger) tile set that is more robust to particular kinds of errors in the kTAM. The transforms are simple : each tile in the original tile set is replaced by a $k \times k'$ block of tiles with a specified pattern of labels that implement the original tile's logic. The proofreading methods [21, 6, 14] transform rule tiles, while the zig-zag tile set can be considered a transform of the seed and boundary tiles. The cost of both these transformations is a moderate increase in spatial scale and the number of tile types.

As an alternative to these methods, one might wonder whether it is possible to also design tile sets capable of any desired computation that rely only on unseeded growth, which appears to be easier to implement experimentally. However, seeded growth, and therefore control of nucleation, appears to be necessary for practical, algorithmic construction by self-assembly: The seed sets up

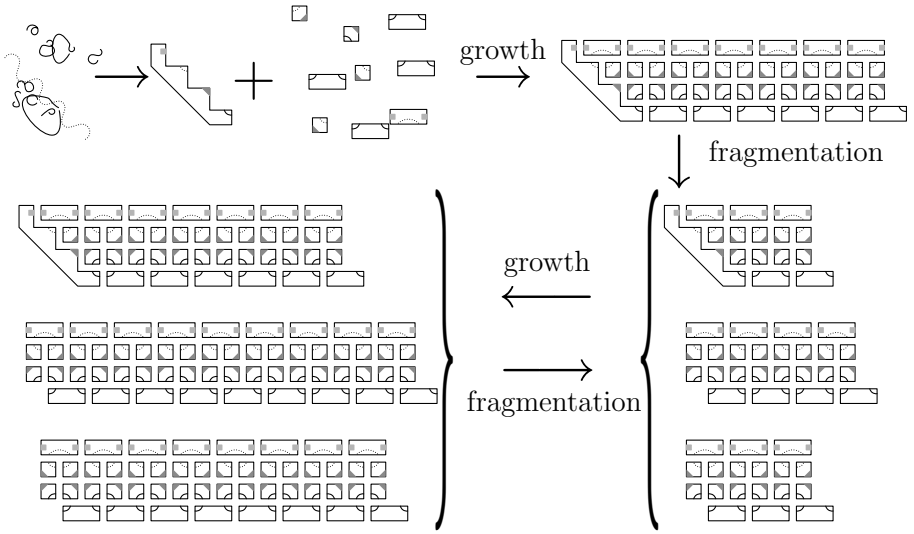

Fig. 4: **Exponential amplification of assemblies.** Probe strands assemble onto a target sequence to create a seed assembly, which nucleates zig-zag growth. Periodic fluid shear causes fragmentation of zig-zag assemblies, leading to exponential amplification. The diagonal structure of the seed assembly shown here is the natural shape for assembling DAE-E tiles on a scaffold strand [16].

the correct initial inputs and directs computation to proceed from beginning to end. (As a consequence, existing mismatch and facet error correction techniques have only been shown to reduce errors in properly seeded assemblies.) Unseeded growth is much more difficult to program and to analyze than is seeded self-assembly, because the "computation" can begin in the middle and proceed in either direction. Although it is possible to assemble computationally defined sets of structures using unseeded growth [1, 22, 4], we would not expect them to assemble efficiently a set of structures as rich as that generated by seeded self-assembly.

3.2 Exquisite Detection of DNA Sequences

Control over nucleation in algorithmic self-assembly can be seen as a special case of exquisite detection (the detection of a single molecule) [2]. For a tile set of sufficiently large width, essentially nothing happens when no seed tiles are present, whereas if even a *single* seed tile is added, growth by self-assembly will result in a macroscopic assembly. Theorem 1 shows that the *false-positive* rate for detection can be made arbitrarily small by design; the *false-negative* rate in the kTAM is 0. Although this idealized model does not consider many factors that could lead to poorer detection in a real system, we don't know of any insurmountable problems with implementing exquisite detection.

There are, however, two immediate drawbacks. First, detecting seed-tile assemblies is not as useful as detecting arbitrary DNA sequences. Second, the

linear growth of a single zig-zag assembly would require a long time lapse before a macroscopic change is perceptible. As sketched in Figure 4, we can surmount both obstacles. First, as in [13, 23], a set of strands can be designed to assemble double-crossover molecules on a (sufficiently long) target strand with nearly arbitrary sequence, thus creating the seed assembly if and only if the target strand exists. Second, since fluid shear forces can fragment large DNA assemblies, intermittent pipetting or vortexing will break large zig-zag assemblies, thus at least doubling the number of growing ends with each fragmentation episode. This fragmentation process can be expected to lead to exponential growth in the number of zig-zag assemblies without increasing the false-positive rate.[3]

3.3 Exponential Replication of Inheritable Information.

The zig-zag constructions detailed in this paper propagate a single bit of information: the presence or absence of the seed tile. Using a tile set that simply copies information, we could use the exponential amplification reaction to detect and identify one of several different target strands, by creating a tile set where the seed assemblies for each target strand contain a different pattern of 1s and 0s.

Furthermore, considering the amplification process as replication, the information encoded in the strip's width can be seen as a form of inheritable information. A zig-zag assembly replicates (in the appropriate culture medium consisting of tiles) by growth of new layers followed by random fission [11]. Errors during growth, bit flips as well as errors that increase or decrease the width of the assembly, are inherited. If one sequence of tiles has a greater reproductive fitness than other sequences – for example, by having a different growth or fission rate – then natural evolution can be expected to occur. Cairns-Smith considered related ideas about crystal growth as a possible scenario for the origin of life on Earth [5]. However, additional mechanisms would have to be present for this inheritable information to be useful in directing the reproduction of tile sequences. Such an enzyme-free system would be considerably less complex than that controlling the replication of chemical information in modern biological organisms or in processes such as polymerase chain reaction (PCR) that provide the basis for most *in vitro* evolution studies.

Acknowledgments. The authors are grateful to Ho-Lin Chen, Ashish Goel, Rizal Hariadi, Paul Rothemund, Bernie Yurke, and Dave Zhang for helpful advice and discussions. This work was supported by NSF CAREER Grant No. 0093486 to EW and an NSF Graduate Fellowship to RS.

References

1. L. M. Adleman. Molecular computation of solutions to combinatorial problems. *Science*, 266:1021–1024, Nov. 11, 1994.

[3] When a spuriously nucleated assembly does eventually form, of course, it will also be exponentially amplified.

2. L. M. Adleman. Personal communication, 2004.
3. L. M. Adleman, Q. Cheng, A. Goel, and M.-D. Huang. Running time and program size for self-assembled squares. *Symposium on the Theory of Computing (STOC)*, 2001.
4. L. M. Adleman, J. Kari, L. Kari, and D. Reishus. On the decidability of self-assembly of infinite ribbons. *Symposium on Foundations of Computer Science (FOCS)*, 43:530, 2002.
5. A. G. Cairns-Smith. *The life puzzle: on crystals and organisms and on the possibility of a crystal as an ancestor*. Oliver and Boyd, New York, 1971.
6. H.-L. Chen and A. Goel. Error free self-assembly using error prone tiles. Accepted to *DNA Computing 10*.
7. J. Chen and J. Reif, editors. *DNA Computing 9*, volume LNCS 2943, Berlin Heidelberg, 2004. Springer-Verlag.
8. M. Cook, P. W. K. Rothemund, and E. Winfree. Self-assembled circuit patterns. In Chen and Reif [7], pages 91–107.
9. R. Davey and J. Garside. *From Molecules to Crystallizers*. Oxford University Press, Oxford, UK, 2000.
10. K. A. Dill and S. Bromberg. *Molecular Driving Forces: Statistical Thermodynamics in Chemistry and Biology*. Garland Science, 2002.
11. G. Egan. Wang's carpets. In G. Bear, editor, *New Legends*. Legend, London, 1995.
12. M. G. Lagoudakis and T. H. LaBean. 2-D DNA self-assembly for satisfiability. In E. Winfree and D. K. Gifford, editors, *DNA Based Computers V*, volume 54 of *DIMACS*, pages 141–154, Providence, RI, 2000. American Mathematical Society.
13. C. Mao, T. H. LaBean, J. H. Reif, and N. C. Seeman. Logical computation using algorithmic self-assembly of DNA triple-crossover molecules. *Nature*, 407(6803):493–496, 2000.
14. J. H. Reif, S. Sahu, and P. Yin. Compact error-resilient computational DNA tiling assemblies. Accepted to *DNA Computing 10*.
15. P. W. K. Rothemund. Using lateral capillary forces to compute by self-assembly. *Proceedings of the National Academy of Sciences*, 97:984–989, 2000.
16. P. W. K. Rothemund, N. Papadakis, and E. Winfree. Algorithmic self-assembly of DNA Sierpinski triangles. 2004. Submitted.
17. R. Schulman, S. Lee, N. Papadakis, and E. Winfree. One dimensional boundaries for DNA tile self-assembly. In Chen and Reif [7], pages 108–125.
18. R. Schulman and E. Winfree. Kinetic models of DNA tile self-assembly. In preparation.
19. E. Winfree. On the computational power of DNA annealing and ligation. In R. J. Lipton and E. B. Baum, editors, *DNA Based Computers*, volume 27 of *DIMACS*, pages 199–221, Providence, RI, 1996. American Mathematical Society.
20. E. Winfree. Simulations of computing by self-assembly. Technical Report CS-TR:1998.22, Caltech, 1998.
21. E. Winfree and R. Bekbolatov. Proofreading tile sets: Error-correction for algorithmic self-assembly. In Chen and Reif [7], pages 126–144.
22. E. Winfree, X. Yang, and N. C. Seeman. Universal computation via self-assembly of DNA: Some theory and experiments. In L. F. Landweber and E. B. Baum, editors, *DNA Based Computers II*, volume 44 of *DIMACS*, pages 191–213, Providence, RI, 1998. American Mathematical Society.
23. H. Yan, T. H. LaBean, L. Feng, and J. H. Reif. Directed nucleation assembly of DNA tile complexes for barcode-patterned lattices. *Proc. Nat. Acad. Sci. USA*, 100(14):8103–8108, 2003.