Overview of Error Correction for of Tiling Self-Assemblies

(material from papers by Winfree, Goel, Chen, Reif
and their coauthors)
Algorithmic self-assembly (without errors):

- Crystal growth is programmed by designing a set of tiles with binding interactions enforce specific local assembly rules.

- Growth begins from a nucleating structure and consists of a series of attachments of single tiles.

- Under slightly supersaturated conditions, the attachment of a tile to a growing crystal is energetically favorable only if it attaches to a growing crystal by at least two binding sites.

- The tiles are designed so that during correct assembly, at every step a tile in the pattern attaches by a particular set of two or more binding sites.

- These binding sites are the tile’s inputs: the identities of the binding sites together determine which tile can attach at a given site.
Theoretical and Algorithmic Issues

• Efficiently assembling basic shapes with precisely controlled size and pattern
  – Constructing N X N squares with $\Omega(\log n/\log \log n)$ tiles
    [Adleman, Cheng, Goel, Huang, ’01]
  – Perform universal computation by simulating BCA
    [Winfree ’99]

• Library of primitives to use in designing nano-scale structures [Adleman, Cheng, Goel, Huang, ’01]

• Automate the design process
  [Adleman, Cheng, Goel, Huang, Kempe, Moisset de espanes, Rothemund ’01]

• Robustness

(Cheng, Goel, Cheng)
Robustness

• In practice, self-assembly is a thermodynamic process. When $T=2$, tiles with 0 or 1 matches also attach; tiles held by total strength 2 also fall off at a small rate.

• Currently, there are 1-10% errors observed in experimental self-assembly. [Winfree, Bekbolatov, ’03]

• Possible schemes for error correction
  – Biochemistry tricks
  – Coding theory and error correction

(Cheng, Goel, Cheng)
Errors of Algorithmic self-assembly:
- self-assembly is stochastic:
- unfavorable attachments of tiles with one or more incorrect or absent inputs also occur.

Errors of insufficient attachment:
- While tiles that attach unfavorably usually fall off quickly, occasionally such a tile is locked in by the subsequent favorable attachment of an adjacent tile, an event we will call an.

Tiling Error: when a tile that attaches unfavorably does not match some of its input binding sites, so may not be the correct tile in the desired pattern
- Subsequent algorithmic pattern formation can be severely disrupted, resulting in a grossly malformed product.

Figure illustrates correct and erroneous assembly steps.

Growth errors:
Insufficient attachments at sites where a correct tile could have attached; they involve both a correctly matching binding site and a mismatch.
Insufficient attachments on facets involve no mismatches; nonetheless, incorrect.
Modelling Errors

**Temperature:** A positive integer. (Usually 1 or 2)

**A set of tile types:** Each tile is an oriented rectangle with glues on its corners. Each glue has a non-negative strength (0, 1 or 2).

**An initial assembly (seed).**

**Rules:** A tile can attach to an assembly iff the combined strength of the “matched glues” is greater or equal than the temperature. Tiles with combined strength equal to temperature can fall off.

**Errors:** Once a while, there will be two tiles attach at the same time and both are held by strength at least two after the attachment. We call this an “**insufficient attachment**”.

**Our goal:** minimize the impact of insufficient attachments
Example

T=2

(Cheng, Goel, Cheng)
Example

(Cheng, Goel, Cheng)
Example

T=2

(Cheng, Goel, Cheng)
What can go wrong?

\[ T=2 \]

(Cheng, Goel, Cheng)
What can go wrong?

T=2

(Cheng, Goel, Cheng)
Why it may not matter:

$T=2$

(Cheng, Goel, Cheng)
Why it may not matter:

(Cheng, Goel, Cheng)
What can go *really* wrong?

$(Cheng, Goel, Cheng)$
What can go *really* wrong?

(Cheng, Goel, Cheng)
What can go *really* wrong?

\[ T=2 \]

(Cheng, Goel, Cheng)
Error-Reducing Designs

• Biochemistry tricks
  – Strand Invasion mechanism
    [Chen, Cheng, Goel, Huang, Moisset de espanes, ’04]

• Coding theory and error correction
  – Proofreading tiles
    [Winfree, Bekbolatov, ’03]
  – Snake tiles
    [Chen, Goel 04]
  – Compact Redundant Tiles
    [Reif,Sahu,Yin 04]
Example: Sierpinski Tile System

(Cheng, Goel, Cheng)
Example: Sierpinski Tile System

(Cheng, Goel, Cheng)
Example: Sierpinski Tile System

(Cheng, Goel, Cheng)
Example: Sierpinski Tile System

(Cheng, Goel, Cheng)
Crystallization Error

(Cheng, Goel, Cheng)
Crystallization Error

(Cheng, Goel, Cheng)
Crystallization Error

(Cheng, Goel, Cheng)
Crystallization Error

(Cheng, Goel, Cheng)
Crystallization Error

(Cheng, Goel, Cheng)
Proofreading Tile Sets: Error Correction for Algorithmic Self-Assembly

Erik Winfree and Renat Bekbolatov
2003
Proofreading Tiles

Each tile in the original system corresponds to four tiles in the new system.

The internal glues are unique to this block.

(Cheng, Goel, Cheng)
How does this help?

(Cheng, Goel, Cheng)
How does this help?

(Cheng, Goel, Cheng)
How does this help?

(Cheng, Goel, Cheng)
How does this help?

(Cheng, Goel, Cheng)
How does this help?

(Cheng, Goel, Cheng)
How does this help?

(Cheng, Goel, Cheng)
How does this help?

(Cheng, Goel, Cheng)
Fig. 2. (a) Assembly of two double-crossover tiles via hybridization of 5-nucleotide sticky ends. $k_f$ is the forward rate constant, in $/M/sec$, and $k_{r,1} = k_f e^{-G_{se}}$ is the reverse rate constant, in $/sec$. (b) Assembly of a double-crossover tile into a site on the growth front of a crystal via hybridization of two 5-nucleotide sticky-end pairs. The forward rate constant is assumed to be the same as for the single sticky-end reaction of (a), while the reverse rate constant is assumed to require twice as much energy to simultaneously break both sticky-end bonds — i.e., binding is cooperative — and thus $k_{r,2} = k_f e^{-2G_{se}}$. $G_{se}$ is the free energy of dissociation for a single sticky end, in units of $RT$. 
Kinetic Analysis of Original Assembly:
Assume a continuous-time Markov process: (satisfying detailed balance) for modeling the 3 D growth of a single crystal in a solution of free monomer tiles.

Assume a monomer tile whose interactions (in the strength units of the aTAM) with the crystal sum to

\[ b = \text{number of unit-strength sticky ends binding the tile to the crystal}. \]

Absolute Rates:

Rate Association of a new monomer tile at any given site:

\[ r_f = k_f [\text{monomer tile}] = k_f e^{-Gmc} \]

Rate Disassociation:

\[ r_{r,b} = k_{r,b} = k_f e^{-bGse} = k_f e^{-2Gse} \quad \text{for case } b = 2 \]

(Winfree,Bekbolatov)
Kinetic Analysis of Original Assembly, Cont:

\( G_{mc} = \) the entropic cost of putting a tile at a binding site (depends on the *monomer* tile concentration)
\( G_{se} = \) free energy cost of breaking a single strength-1 bond

Optimal Growth Rates are near Melting Temp of crystals:
When \( G_{mc} \approx b \ G_{se} \approx 2 \ G_{se} \) when \( b=2 \)

\( \Delta G = \text{dif free energy between an assembly with mismatched tile and assembly with a correct tile}. \)

Assume thermodynamic limit: Error Rate \( \varepsilon \approx e^{-\Delta G/RT} \approx e^{-\Delta G_{se}} \)

Growth Rate of Assembly:

\[ r \approx \text{monomer tile concentration} \]
\[ = \beta \ \text{[monomer tile]} = \beta \ e^{-G_{mc}} \approx \beta \ e^{-2G_{se}} = \beta \varepsilon^2 \]

(Winfree,Bekbolatov)
Fig. 3. (a) Phase diagram [28] for crystal growth of tiles implementing a BCA, under the kTAM. “Good crystals” (growth rate comparable to $k_f [DX]$ and error rate smaller than $\epsilon$) are obtained for large $G_{se}$ and $G_{mc}$, below the $\tau = 2$ boundary marking the melting transition where $G_{mc} = 2G_{se}$. (b) Model for kinetic trapping. The growth site may ($E$) be empty; ($C$) contain a correct tile; ($M$) contain a mismatched tile; ($FC$) be “frozen” with the correct tile in place; or ($FM$) be “frozen” with the mismatched tile. $r^*$ represents the rate at which tiles on the growth front are covered. The error rate is taken to be the probability that, starting in $E$, the system reaches $FM$. 

(Winfree, Bekbolatov)
Fig. 6. (a) The general $2 \times 2$ proofreading construction for rule tiles. (b) The original Sierpinski tiles. (c) The $2 \times 2$ proofreading Sierpinski tiles. (d) Growth of the proofreading Sierpinski tiles. Small tiles illustrate that when a mismatched tile is incorporated, further growth on one side must involve a second mismatch.
Fig. 4. (a) Growth of the original (1 × 1) Sierpinski tile set at $G_{mc} = 13.9$ and $G_{se} = 7.0$, to a size of $\sim$ 32 layers in $\sim$ 530 simulated seconds. Two errors can be seen; the first occurs in the third frame and is indicated by an arrow. Subsequent error-free growth correctly propagates the erroneous information. (b) Growth of the 2 × 2 proofreading tiles at $G_{mc} = 12.9$ and $G_{se} = 6.5$, to a size of $\sim$ 64 layers in $\sim$ 460 simulated seconds. (c) Growth of the 3 × 3 proofreading tiles at $G_{mc} = 11.9$ and $G_{se} = 6.0$, to a size of $\sim$ 96 layers in $\sim$ 310 simulated seconds.
**Fig. 9.** Proofreading tile sets are often able to heal a puncture in the crystal. Sometimes, as in this case, some of the tiles that fill in the puncture do not perfectly match their neighbors – a form of “scar tissue.”
**Kinetic Analysis of Proof-Reading Assembly:**
Again assume a continuous-time Markov process:
(satisfying detailed balance) for modeling the 3D growth of a single crystal in a solution of free monomer tiles.

*In Proof-Reading Assembly each monomer tile is replaced with a K x K subassembly.*
Again Optimal Growth Rates are near Melting Temp of crystals, where \( G_{mc} \approx 2 \ G_{se} \)

Assume thermodynamic limit: error rate for an entire block:
now determined by K mismatched tiles
*Error Rate* \( \varepsilon \approx e^{-\Delta G/RT} \approx e^{-\Delta G_{se}} \approx e^{-KG_{se}} \) \[ \text{so } e^{-G_{se}} \approx \varepsilon^{1/k} \]

*Growth Rate of Proof-Reading Assembly:*
\[ r \approx \text{monomer tile concentration} \]
\[ = \beta \ [\text{monomer tile}] \]
\[ = \beta \ e^{-G_{mc}} \approx \beta \ e^{-2G_{se}} = \beta \varepsilon^{2/K} \]

(Winfree,Bekbolatov)
**Fig. 7.** (a) A kinetic trapping model including all 29 states (up to symmetry) representing “well-associated” tiles within a $2 \times 2$ growth site (see text for details). Arrows inside tiles indicate that the tile belongs to a proofreading block that has a mismatch to the input in the indicated direction; there are $N$ such tiles for an $(N + 1)$-state BCA. Arrows between states indicate reversible reactions (association or dissociation of a tile); reverse reaction rates are given at the head of each arrow, and forward reaction rates are given near the tail. States within dotted circles each have an irreversible reaction to a frozen state (either FM or FC) with rate $r^*$. (b) A simplified kinetic trapping model with 9 states considers only the major reaction pathways in (a), which are indicated by the red and green reaction arrows for pathways leading to mismatched or correct blocks, respectively.

(Winfree, Bekbolatov)
Error Free Self-assembly Using Snaked Proof-Reading Tiles

Ho-Lin Chen & Ashish Goel

2005
Nucleation Error

(Cheng, Goel, Cheng)
Nucleation Error

• First tile attaches with a weak binding strength

(Cheng, Goel, Cheng)
Nucleation Error

• First tile attaches with a weak binding strength
• Second tile attaches and secures the first tile

(Cheng, Goel, Cheng)
Nucleation Error

• First tile attaches with a weak binding strength
• Second tile attaches and secures the first tile
• Other tiles can attach and forms a layer of (possibly incorrect) tiles

(Cheng, Goel, Cheng)
Snake Tiles

- Each tile in the original system corresponds to four tiles in the new system.
- The internal glues are unique to this block.

(Cheng, Goel, Cheng)
How does this help?

• First tile attaches with a weak binding strength

(Cheng, Goel, Cheng)
How does this help?

- First tile attaches with a weak binding strength
- Second tile attaches and secures the first tile

(Cheng, Goel, Cheng)
How does this help?

• First tile attaches with a weak binding strength
• Second tile attaches and secures the first tile
• No Other tiles can attach without another nucleation error

(Cheng, Goel, Cheng)
Four by Four Snake Tiles

(Cheng, Goel, Cheng)
Four by Four Snake Tiles

(Cheng, Goel, Cheng)
Four by Four Snake Tiles

(Cheng, Goel, Cheng)
Four by Four Snake Tiles

(Cheng, Goel, Cheng)
Four by Four Snake Tiles

(Cheng, Goel, Cheng)
Four by Four Snake Tiles

(Cheng, Goel, Cheng)
Four by Four Snake Tiles

(Cheng, Goel, Cheng)
Four by Four Snake Tiles

(Cheng, Goel, Cheng)
Four by Four Snake Tiles

(Cheng, Goel, Cheng)
Four by Four Snake Tiles

(Cheng, Goel, Cheng)
Four by Four Snake Tiles

(Cheng, Goel, Cheng)
Four by Four Snake Tiles
Nucleation Errors (T=2) for Diagonal Tile Assemblies

Starting from an initial assembly
Nucleation Error (T=2)

The first tile attaches with strength 1. (usually falls off fast)
Nucleation Error (T=2)

The second tile attaches and now both tiles are held by strength 2.
Nucleation Error (T=2)

Error propagates.

(Chen)
Snaked Tile System for Diagonal Tile Assemblies

- Replace a tile by a block of 4 tiles
- Internal glues are unique

[Chen, Goel, 2004]
Nucleation Error (T=2) for Diagonal Tile Assemblies

Starting from an initial assembly
Nucleation Error (T=2)

Two tiles attach and both tiles are held by strength 2.
Nucleation Error (T=2)

No other tiles can attach.

Inert edge
Generalization for Diagonal Tile Assemblies

(Chen)
Experimental Verification

Joint work with

Ashish Goel
Rebecca Schulman
Erik Winfree

(Chen)
Snaked Tile System

- Replace a tile by a block of 4 tiles
- Internal glues are unique

[Chen, Goel, 2004]
Tile sets used in experiments

Proofreading block

Snaked block

(Chen)
Width-4 Zig-Zag Ribbon

[Schulman, Winfree, DNA 10, 2004]

6 tile types
AFM of Zig-Zag Ribbons

(Chen)
ZZ + Snake Tiles

10 tile types

Side A

Side B (no glue)
Slow nucleation and growth!

ZZ + Snake Tiles

10 tile types

Side A

Side B (no glue)
Fast nucleation and growth!

**ZZf + Snake Tiles**

10 tile types

(Chen)
## Zig-Zag ribbons used in experiments

<table>
<thead>
<tr>
<th>name</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Normal Zig-Zag (ZZ)</strong></td>
<td>Side A: Glues 2, 4 slow</td>
</tr>
<tr>
<td></td>
<td>Side B: blunt</td>
</tr>
<tr>
<td><strong>Flipped (ZZf)</strong></td>
<td>Side A: Glues 1, 3 fast</td>
</tr>
<tr>
<td></td>
<td>Side B: blunt</td>
</tr>
<tr>
<td><strong>Double_sided (ZZ_DS)</strong></td>
<td>Side A: Glues 2, 4 slow</td>
</tr>
<tr>
<td></td>
<td>Side B: Glues 1, 3 fast</td>
</tr>
<tr>
<td><strong>Flipped + double_sided (ZZ_DSf)</strong></td>
<td>Side A: Glues 1, 3 fast</td>
</tr>
<tr>
<td></td>
<td>Side B: Glues 2, 4 slow</td>
</tr>
</tbody>
</table>

(Chen)
Experiment Results
ZZf + 100 nM Snaked block
Zig-Zag + 100 nM Snaked block
Zig-Zag + 100 nM Proofreading

(Chen)
ZZf + 100 nM Proofreading
ZZ_DS + 10 nM Proofreading

(Chen)
ZZ_DS + 10 nM Snaked block
## Statistics

| Ratio of chunks on each side | Zig-Zag  
Side A: glues 2, 4  
Side B: glues 1, 3 | Zig-Zag  
Side A: glues 1, 3  
Side B: glues 2, 4 |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Snaked block</td>
<td>4.7</td>
<td>4.2</td>
</tr>
<tr>
<td>Proofreading block</td>
<td>1.1</td>
<td>1.5</td>
</tr>
</tbody>
</table>
## Statistics

<table>
<thead>
<tr>
<th>Ratio of tiles on each side</th>
<th>Zig-Zag Side A: glues 2, 4 Side B: glues 1, 3</th>
<th>Zig-Zag Side A: glues 1, 3 Side B: glues 2, 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Snaked block</td>
<td>4.3</td>
<td>3.9</td>
</tr>
<tr>
<td>Proofreading block</td>
<td>1.0</td>
<td>1.2</td>
</tr>
</tbody>
</table>
Theoretical Analysis

• The snake tile design can be extended to 2k by 2k blocks.

• Prevents tile propagation even after k-1 insufficient attachments happen.
Why it works?

: insufficient attachments

: erroneous tiles falling off

(Chen)
Why it works?

: happens with rate $O(e^{-G}) \times r_f$

: erroneous tiles falling off

(Chen)
Theoretical Analysis

• The snake tile design can be extended to $2k$ by $2k$ blocks.

• Prevents tile propagation even after $(k-1)$ insufficient attachments happen.

• When $< k$ insufficient attachments happened locally, all the erroneous tiles are expected to fall off in time $\text{poly}(k)$.
Why it works?

- Happens with rate $O(e^{-G}) \times r_f$
- Happens with rate $1/poly(k)$

(Chen)
Why it works?

: happens with rate $O(e^{-kG})$

if backward rate $>>$ forward rate

(Chen)
Theoretical Analysis

If we want to assemble a structure with size $N$, we can use Snaked Tile System with block size $k = O(\log N)$.

The assembly process is expected to finish within time $\tilde{O}(N)$ and be error free with high probability.
Fig. 1. (a) The parity tile system. (b) Illustrating the action of the parity tile system on the ”input” string 1111. The arrow at the top represents the order in which tiles must attach in the absence of errors.

Fig. 2. (a) The original 10 tile. (b) The four proof-reading tiles for the 10 tile, using the construction of Winfree and Bekbolatov [11]. (c) The snaked proof-reading tiles for the parity tile system. The internal glues are all unique to the $2 \times 2$ block corresponding to the 10 tile. Notice that there is no glue on the right side of $10_A$ or the left side of $10_C$ and that the glue between the top two tiles is of strength 2. This means that the assembly process doubles or “snakes” back onto itself, as demonstrated by the arrow.
Theoretical Analysis

• The snake tile design can be extended to 2k by 2k blocks.
  – Prevents tile propagation even after k+1 nucleation error happen.

• With 2k by 2k snake tile system, we can assemble an N by N square of blocks with time $\tilde{O}(N^{1+4/k+1})$ and (with high probability) remain stable for time $\tilde{O}(N^{1+4/k+1})$.
  – Assuming tiles held by strength 3 does not fall off
  – The error probability at each location changes from $p$ to $p^k$

(Cheng, Goel, Cheng)
Errors of Insufficient Attachment

Lemma 1. The rate at which an insufficient attachment happens at any location in a growing assembly is $f^2 r e^{-G_{se}} = O(e^{-3G_{se}})$.

Proof. The rate of an insufficient attachment can be modeled as the Markov Chain shown in figure 3. For a nucleation error to happen, first a single tile must attach (at rate $f$). The fall-off rate of the first tile is $r e^{G_{se}}$ and the rate at which a second tile can come and attach to the first tile is $f$. After the second tile attaches, an insufficient attachment has happened. So the overall rate of an insufficient attachment is $f * \frac{f}{f + r e^{G_{se}}} \approx \frac{f^2}{r} e^{-G_{se}}$.
Nucleation Errors

\[ f_2 \cdot r \cdot e^{-G_{se}} = O\left(e^{-3G_{se}}\right). \]

**Proof.** The rate of an insufficient attachment can be modeled as the Markov Chain shown in figure 3. For a nucleation error to happen, first a single tile must attach (at rate \( f \)). The fall-off rate of the first tile is \( r \cdot e^{-G_{se}} \) and the rate at which a second tile can come and attach to the first tile is \( f \). After the second tile attaches, an insufficient attachment has happened. So the overall rate of an insufficient attachment is

\[ f \cdot f \cdot f + r \cdot e^{-G_{se}} \approx f_2 \cdot r \cdot e^{-G_{se}}. \]

Without proof-reading, or even using the proof-reading system of Winfree and Bekbolatov, a single insufficient attachment can cause a nucleation error, and hence the error propagates.

**Fig. 3.** The C tiles represent the existing assembly, and the E tiles are new erroneous tiles.
Nucleation Errors

**Lemma 2.** The rate at which a nucleation error takes place in our snaked proof-reading system is $O(e^{-4G_{se}})$.

*Proof.* In the snaked system, two insufficient attachments need to happen next to each other for a nucleation error to occur. According to lemma 1, the first insufficient attachment happens at rate $O(e^{-3G_{se}})$. After the first insufficient attachment, the error will eventually be corrected unless another insufficient attachment happens next to the first. The second insufficient attachment happens at rate $O(e^{-3G_{se}})$; but the earlier insufficient attachment gets “corrected” at rate $O(e^{-2G_{se}})$ (remember that $a \approx 1$ and hence a tile attached with strength 2 falls off at roughly the growth rate). Hence, the probability of another insufficient attachment taking place before the previous insufficient attachment gets reversed is $O(e^{-G_{se}})$, bringing the nucleation error rate down to $O(e^{-4G_{se}})$.
Snaked Assembly Growth

Fig. 4. (a) The structure of 4x4 block. (b) The order of the growth
Snaked Assembly Growth

**Theorem 2.** With a $2k \times 2k$ snaked tile system, $k = O(\log n)$, assuming that we can set $e^{G_{se}}$ to be $O(k^6)$, an $n \times n$ square of blocks can be assembled in time $\tilde{O}(n)$ and with high probability, no block errors happen for $\tilde{\Omega}(n)$ time after that.
Snaked Assembly Growth

Lemma 3. Consider any connected structure caused by $m$ insufficient attachments ($1 \leq m \leq k$). Then the width of the structure can be at most $2m$, and the height of the structure can be at most $2k$ (i.e., this connected structure can only span two blocks). This structure will fall off in expected time $O\left(\frac{k^5}{r}\right)$ unless there’s a block error somewhere in the assembly or an insufficient attachment happens within the (at most two) blocks spanned by the structure.

Proof Outline: The proof of this lemma involves a lot of technical details. Due to space constraints, we only present a sketch in this version. In the structure of $2k \times 2k$ snaked tiles, all the glues between the $(2i)$-th row and $(2i+1)$-th row have strength 1 for all $i$. So, to increase the width from $2i$ to $2i+1$, we must have at least one insufficient attachment. So, with $m$ insufficient attachments, the width of the structure can be at most $2m$. Using similar arguments, the height of the structure can be at most $2k$. Also, the attached tiles can be partitioned into $O(k)$ parts. Each of these parts can be viewed as a $2 \times O(k)$ rectangle with every internal glue having strength 1. The process of tiles attaching to and detaching from each rectangle can be modeled using two orthogonal random walks and hence, each rectangle will fall off in expected time $O\left(\frac{k^4}{r}\right)$. The different rectangles can fall off sequentially, and after one rectangle falls off completely, none of its tiles will attach again unless an insufficient attachment happens. Thus, the structure will fall off in expected time $O\left(\frac{k^5}{r}\right)$ unless there’s a block error (anywhere in the assembly) or an insufficient attachment happens (within the two blocks) before the structure has a chance to fall off. $\square$
Snaked Assembly Growth

**Theorem 3.** Assume that we use a $2k \times 2k$ snaked tile system and $G_{mc} = 2G_{se}$. Then for any $\epsilon$, there exists a constant $c$ such that, with probability $1 - \epsilon$, no $k$-bottleneck will happen at a specific location within time $c \frac{1}{f} e^{G_{se}} \left( \frac{e^{-G_{se}} + 1/k^6}{e^{-G_{se}}} \right)^{k-1}$.

**Proof.** By definition, $k$ insufficient attachments are required before a $k$-bottleneck happens. After $i$ insufficient attachments take place, one of the following is going to happen:

- One more insufficient attachment. Consider any structure $X$ caused by $i$ insufficient attachments. By Lemma 3, the size of $X$ cannot exceed two blocks, hence the number of insufficient attachment locations that can cause this structure to grow larger is at most $4k$. So, the rate of the $(i + 1)$-th insufficient attachment happening is at most $4kfe^{-G_{se}}$.

- All the attached tiles fall off. By Lemma 3, the expected time for all the attached tiles to fall off is $O\left(\frac{k^5}{r}\right)$

So, after $i$ insufficient attachments happen, the probability of the $(i + 1)$-th insufficient attachment happening before all tiles fall off is $O\left( \frac{ke^{-G_{se}}}{kfe^{-G_{se} + r/k^6}} \right) = O\left( \frac{e^{-G_{se}}}{e^{-G_{se} + 1/k^6}} \right)$. So, after the first insufficient attachment takes place, the probability of a $k$-bottleneck happening before all the attached tiles fall off is less than $O\left( \left( \frac{e^{-G_{se}}}{e^{-G_{se} + 1/k^6}} \right)^{k-1} \right)$. As shown in Lemma 1, the expected time for the first insufficient attachment is $O\left( \frac{1}{f} e^{G_{se}} \right)$. So, the expected time for a $k$-bottleneck to happen at a certain location is at most $O\left( \frac{1}{f} e^{G_{se}} \left( \frac{e^{-G_{se}} + 1/k^6}{e^{-G_{se}}} \right)^{k-1} \right)$. Hence, for any small $\epsilon$, we can find a constant $c$ such that, with probability $1 - \epsilon$, no $k$-bottleneck will happen at a specific location within time $c \frac{1}{f} e^{G_{se}} \left( \frac{e^{-G_{se}} + 1/k^6}{e^{-G_{se}}} \right)^{k-1}$.
Theorem 4. If we assume there are no $k$-bottlenecks, and the rate of insufficient attachments is at most $O\left(\frac{f}{k^6}\right)$, then an $n \times n$ square of $2k \times 2k$ snaked tile blocks can be assembled in expected time $O\left(\frac{k^5 n}{f}\right)$.

Proof. With the snaked tile system, after all the tiles in a block attach, all the tiles are held by strength at least 3 and will never fall off. Using the running time analysis technique of Adleman et al. [2], the system finishes in expected time $O(n \times T_B)$, where $n$ is the size of the terminal shape and $T_B$ is the expected time for a block to assemble. Without presence of $k$-bottlenecks, when we want to assemble a block, the erroneous tiles that currently occupy that block are formed by at most $k - 1$ insufficient attachments. By Lemma 3, without any further insufficient attachments happening, the erroneous tiles will fall off in time $O\left(\frac{k^5}{f}\right)$ and the correct block can attach within time $O\left(\frac{k^4}{f}\right)$. By assumption, the rate of insufficient attachment happening is at most $O\left(\frac{f}{k^6}\right)$, and there are at most $O(k)$ locations for insufficient attachments to happen and affect this process. So, there’s a constant probability that no insufficient attachments will happen during the whole process and thus the time required to assemble a block, $T_B$, is at most $O\left(\frac{k^5}{f}\right)$. 
Reducing Facet Nucleation during Algorithmic Self-Assembly

Ho-Lin Chen, Rebecca Schulman, Ashish Goel, and Erik Winfree

NANOLETTERS 2007
Errors of Algorithmic self-assembly:

(1) Insufficient attachments

(2)- Facet Nucleation Errors

- Insufficient attachments on assembly facets that involve no mismatches

In standard crystal growth:
- nucleation on facets is part of the desired growth process,

In a proper algorithmic self-assembly:
- every tile attaches by two or more binding sites, so nucleation on facets may cause errors of assembly

(Chen,Schulman,Goel,Winfree)
Figure 1. Algorithmic self-assembly and proofreading blocks. (a) During algorithmic self-assembly, a tile attaches to a growing crystal by binding domains on its edges. Here, the four labels on a tile's corners indicate specific binding domains; asterisk indicates complementary localities (X binds to X*). The attachment of a tile where both its input (bottom) edges match the available edges on the crystal is preferred over the attachment of a tile where a single (or no) match occurs. Growth errors occur when a tile attaches by one matching bond and one nonmatching bond. Facet errors occur when a tile attaches by only one matching bond. In both cases, for an error to occur, the incorrect tile must be "locked in" by a second tile before it detaches. (b) The logical structure of a 2 × 2 uniform proofreading block. Each tile in the original tile set is converted into four tiles that, as a logical block, redundantly encode the same input and output information on the perimeter of the block, while binding domains on the interior of the block encode the identity of the original tile. Correct assembly at a growth site proceeds one tile at a time, either in the order pT1→pT2→pT4→pT3 or in the order pT1→pT4→pT2→pT3. The unique labels inside a proofreading block reduce the rate of growth errors because for a block to be completed, one of the tiles that attaches on top of an incorrect tile must also be incorrect—it cannot match both the label inside the block and the label presented by the crystal. (c) The structure of a 2 × 2 snaked proofreading block. Binding labels on the perimeter of the block are the same as in a uniform proofreading block, but the interior has two modifications: there is an interaction between sT1 and sT2, and the other two tiles are fused to create the "double tile" sT34. This forces correct assembly to proceed in the order sT1→sT34→sT2. Snaked proofreading tile sets, like uniform proofreading tile sets, force a subsequent tile that attaches after an incorrect attachment to be incorrect also. (d) Zigzag ribbons. While only three repeat units are shown, ribbons can be arbitrarily long. The 6 tiles interact through 12 distinct pairwise binding domains, all shown as flat sides, as their logic is not essential here. In addition to the double tiles shown, we use variants of the double tiles that present binding localities to create a desired facet (e.g., Z78H presents H1* and H2* for the hard facet) or present inert "blunt ends" (e.g., Z56B) to which nothing may bind. (Chen, Schulman, Goel, Winfree)
The larger $3 \times 3$ snaked proofreading block (Figure 3) can protect against facet errors on both facet orientations, as is necessary for the full range of algorithmic growth. Simulation and theory predict that these and larger $k \times k$ blocks further reduce both growth errors and facet errors.

The larger blocks use the basic mechanism of the $2 \times 2$ block multiple times to reduce the rate of facet errors; for example, the $3 \times 3$ block uses the $2 \times 2$ snaked motif once for each facet orientation. Thus, experimental investigation of the $2 \times 2$ system assesses the essential principle used by the larger systems.

In this paper, we investigate experimentally whether $2 \times 2$ snaked proofreading tiles have a lower rate of facet nucleation than $2 \times 2$ uniform proofreading tiles. We use DNA tiles (Figure 4) to implement both uniform and snaked proofreading blocks and study their growth on long facets created using zigzag ribbons (Figure 1d).

See Supplementary Figures S1-S11 for sequences and diagrams of all molecules used in this work. We show that with snaked proofreading blocks, facet nucleation errors are reduced.

**Figure 3.** $3 \times 3$ snaked blocks reduce facet growth on both facet types. On either facet, an isolated insufficient attachment (initiated by the tiles marked with dots) can grow by favorable attachment to a maximal size of three tiles, at which point it is no longer possible to attach a tile by two binding sites. However, at a proper growth site, the series of exclusively favorable assembly steps following the snaked path shown can complete the block quickly.
Zigzag Ribbons of ProofReading Tiles have to deal with various Facet Types:

**Easy:** where snaked proofreading tiles can nucleate growth with just one insufficient attachment

**Hard:** where two adjacent insufficient attachments are required for snaked proofreading tiles to nucleate facet growth

**Blunt:** which contain no binding sites and therefore do not allow growth.

(Chen, Schulman, Goel, Winfree)
Figure 2. Facet nucleation and growth. (a) Facet nucleation of uniform proofreading blocks. Following a single insufficient attachment (tiles with dots indicate the unfavorable attachments that were locked in), subsequent growth by favorable attachments can grow an entire layer of tiles. Subsequent rows are each nucleated by a single insufficient attachment event. (b) Facet nucleation for snaked proofreading blocks along the hard facet. Here, a single insufficient attachment results in a pair of tiles on the facet, but further favorable growth steps are impossible because of the inert bonds interior to the snaked blocks. Two adjacent insufficient attachments are necessary to nucleate two layers of facet growth. Each additional two layers of growth requires another two adjacent insufficient attachments. (c) Facet nucleation for snaked proofreading blocks along the easy facet. Here, an insufficient attachment consists of a single tile and an adjacent double tile. Thus, two layers of snaked proofreading tiles can be nucleated by just one insufficient attachment.

(Chen, Schulman, Goel, Winfree)
Figure 4. DNA implementation of uniform and snaked proofreading blocks. (a) DNA tiles for a uniform proofreading block. Each tile shown is a double-crossover molecule known as the DAO-E molecule. A DAO-E molecule is composed of four strands of DNA. While the “core” of the molecule is double stranded, there are four five-nucleotide single-stranded regions (sticky ends) on each molecule that can bind to complementary sticky ends on other tiles. Two hairpins are present on each of the shaded tiles in Figure 1b to provide AFM contrast. (b) DNA tiles for a 2 × 2 snaked proofreading block, which in the rest of the paper will be referred to simply as a “snaked proofreading block”. To make an inert bond between sT1 and sT2, the sticky ends of the tiles are double stranded and truncated by two and three bases, respectively. The double tile sT34 is implemented by a larger molecule which has the structure of two DAO-E tiles fused together. Hairpins are used on tiles shaded in Figure 1c.
We used single-sided ribbons to measure the rate at which lattices grow on a facet. Since each row of growth must be nucleated by one or more insufficient attachments (Figure 2), the number of rows that grow on a facet in a fixed period of time increases with increasing nucleation rate. For each kind of proofreading lattice, we assembled ribbons that presented either an easy or hard facet as described above and then diluted them to 10 nM. We immediately added 50 nM of each preformed proofreading tile, waited for 10 min, then deposited the samples onto mica and imaged them using AFM.

Both kinds of proofreading lattices grew on both facets (parts d-g of Figure 5). The uniform proofreading lattices grew more than 10 layers on both facets, as did the snaked proofreading lattice on the easy facet. In contrast, the snaked proofreading lattice grew only two to six rows on the hard facet during the experiment, suggesting that the nucleation rate of the snaked proofreading lattice on the facet is less than that of the other cases. However, because most of the tiles were used up during the experiment (50 nM is enough to grow an average of 10 rows on each ribbon), it was not possible to quantify how much smaller the nucleation rate of snaked proofreading lattices on the hard facet is than the other rates. Reliability was also limited by trial-to-trial variations in experiment timing and in tile stoichiometry for both the lattices and the ribbons and by concerns about the difference in the melting temperatures of the two lattices.

Figure 5. AFM images. Missing tiles are due to damage during AFM scanning. Scale bars are 300 nm. (a) Uniform proofreading lattices (100 nM). (b) Zigzag ribbons, ZZhardeasy (10 nM). (c) Snaked proofreading lattices (100 nM). (d) ZZeasy (10 nM) with uniform proofreading (50 nM). (e) ZZeasy (10 nM) with snaked proofreading (50 nM). (f) ZZhard (10 nM) with uniform proofreading (50 nM). (g) ZZhard (10 nM) with snaked proofreading (50 nM). (h) ZZhardeasy (10 nM) with uniform proofreading (10 nM). (i) ZZhardeasy (10 nM) with snaked proofreading (10 nM). (j) A long ZZhardeasy ribbon with snaked proofreading.
Self-healing Tile System

• Goal: When a big portion of the lattice is removed, it should be able to grow back correctly.

• Method: For each tile in the original system, we create a unique block in the new system.

• Idea: Use the block to prevent tile from growing backwards.
Assumption

• Use abstract tile assembly model.

• Requires a fix set of incoming and outgoing edges for each tile in the original system.
Example: Sierpinski System

T=2

(Chen)
Example: Sierpinski System
Example: Sierpinski System

$$T=2$$

(Chen)
Example: Sierpinski System

(T=2) destroyed (Chen)
Example: Sierpinski System

\[
\begin{array}{cccccc}
0 & 0 & 1 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 \\
1 & 1 & 1 & 1 & 0 & 0 \\
T=2 & & & & & \\
\end{array}
\]
Example: Sierpinski System

T=2

(Chen)
Example: Sierpinski System

T=2

\[(\text{Chen})\]
Example (T=2)

- Replace a tile by a block of 4 tiles
- Internal glues are unique

(Chen)
Poly-amino safe

• The system is save even when several tiles can form a bigger block before attach to the assembly.

(Chen)
Example (T=2)

- Replace a tile by a block of 4 tiles
- Internal glues are unique

(Chen)
Poly-amino safe

- The system is save even when several tiles can form a bigger block before attach to the assembly.
Poly-amino safe

• The system is save even when several tiles can form a bigger block before attach to the assembly.
Poly-amino safe

- The system is save even when several tiles can form a bigger block before attach to the assembly.
Poly-amino safe

- The system is save even when several tiles can form a bigger block before attach to the assembly.
Poly-amino Safe

- The system is safe even when several tiles can form a bigger block before attach to the assembly.

(Chen)
Not Poly-amino Safe

- Replace a tile by a block of 4 tiles
- Internal glues are unique
Invadable Self-Assembly: Combining Robustness with Efficiency

*Ho-Lin Chen, Qi Cheng & Ashish Goel*

*2009*
Strand Invasion

Figure 1: Illustrating strand invasion
Tile Invasion

Invadable Self-Assembly: Consider a tile system $T$, a supertile $\Gamma$ of $T$ and a tile $t \in T$ that is attachable to $\Gamma$ at some position $p$. We say $t$ has a north-foothold in $\Gamma$ at $p$ iff $f_{N,\Gamma,t}(p) > 0$ and no tile in $T$ but $t$ has glue $\sigma_N(t)$ on its north edge. We define $(\text{south,west,east})$-foothold similarly.

Definition 2.1. We say tile $t$ has a foothold in $\Gamma$ at $p$ iff $t$ has a north, south, east or west-foothold in $\Gamma$ at $p$.

Definition 2.2. We say the attachment of $t$ to $\Gamma$ at $p$ is safe iff $t$ has a foothold in $\Gamma$ at $p$ and no tile in $T$ other than $t$ is attachable to $\Gamma$ at $p$.

Figure: The attachment of $A$ at $p$ is safe at $\tau=2$ since $A$ has a south-foothold, while the attachment of $B$ at $p'$ is non-safe.

Figure shows an example of a safe and of a non-because no tile other than $A$ has glue $y$ on its south safe attachment at temperature 2.

The attachment of side $A$ has a south-foothold, and neither $B$ nor is not $C$ can attach at $p$.

The attachment of $B$ at $p'$ is safe because $B$ has no foothold in the supertile at $p'$ even though $B$ fits perfectly there. In a lab experiment (Chen, Cheng & Goel 2009)
Compact Error Resilient Computational DNA Tiling Assemblies

John Reif, Sudheer Sahu, Peng Yin

Department of Computer Science, Duke University
Self-Assembly of DNA Tiles

- Perform universal computation.
- Manufacture patterned nanostructures from smaller unit nanostructures.
Assembly of Binary Counter (Winfree)

(Reif, Sahu, Yin)
Errors in Self-Assembly of DNA Tiles

- Binding rules are not strict.

- A tile might get assembled to a binding site where it was not supposed to go.
Example of a Computational Error

(Reif, Sahu, Yin)
How to Decrease Errors?

• Errors can be arbitrarily decreased by
  – Decreasing concentration of tiles.
  – Increasing binding strengths.
  – Drawback : Reduce speed.

• Another approach:
  – Change the logical design of the tiles.

(Reif, Sahu, Yin)
Error Resilient Tilings by Winfree

Original tiles:

Error resilient tiles:

- Error rate $\in \Rightarrow \in^2$
- Assembly size increased by 4

(Reif, Sahu, Yin)
Original tiles:

Error resilient tiles:

(Reif, Sahu, Yin)
Original tiles:  

Error resilient tiles:

(Reif,Sahu,Yin)
Original tiles:

Error resilient tiles:

(Reif, Sahu, Yin)
A Computational Tile

\[ V' = U \, \text{OP}_1 \, V \]
\[ U' = U \, \text{OP}_2 \, V \]
Compact Error Resilient Construction

- Wholeness of pad: Single pad per side.

(Reif, Sahu, Yin)
One Mismatch causes more Mismatch

Case 1

(Reif, Sahu, Yin)
One Mismatch causes more Mismatch

Case 2

(Reif, Sahu, Yin)
One Mismatch causes more Mismatch

Case 3

(Reif, Sahu, Yin)
One Mismatch causes more Mismatch

Case 4

(Reif, Sahu, Yin)
Result of Compact Error Resilient Scheme

• We saw:
  – Two way overlay scheme.
  – One mismatch caused at least one more mismatch.
  – Error is reduced from $\in$ to $\in^2$.

• Next we will see:
  – Three way overlay scheme.
  – One mismatch will cause at least two more mismatches.
  – Error will reduce from $\in$ to $\in^3$.

(Reif, Sahu, Yin)
Compact Error Resilient Tiles (3-way overlay)

Reduce Error from $\varepsilon$ to $\varepsilon^3$

(Reif, Sahu, Yin)
Examples of Error Resilient Assembly

(Reif, Sahu, Yin)
Examples of Error Resilient Assembly

(Pads) 00 01 10 11

(Tiles)

(Assembled Binary Counter)

(Reif, Sahu, Yin)
Computer Simulation (Xgrow, Winfree)

Size of errorfree aggregate vs Probability of single mismatch

- No error correction
- Our $T_1$ construction
- Winfree 2x2 construction
- Our $T_2$ construction
- Winfree 3x3 construction

(Reif, Sahu, Yin)
Conclusions

• Assembly size **not** increased.
• Two way overlay: error rate $\epsilon (5\%) \Rightarrow \epsilon^2 (0.25\%)$.
• Three way overlay: error rate $\epsilon (5\%) \Rightarrow \epsilon^3 (0.0125\%)$.
• Open question: error rate $\epsilon \Rightarrow \epsilon^k$?