# Complexities for Generalized Models of Self-Assembly

Gagan Aggarwal* Michael H. Goldwasser† Ming-Yang Kao‡ Robert T. Schweller§

**Abstract**

In this paper, we extend Rothemund and Winfree's examination of the tile complexity of tile self-assembly [6]. They provided a lower bound of $\Omega(\frac{\log N}{\log\log N})$ on the tile complexity of assembling an $N \times N$ square for almost all $N$. Adleman et al. [1] gave a construction which achieves this bound. We consider whether the tile complexity for self-assembly can be reduced through several natural generalizations of the model. One of our results is a tile set of size $O(\sqrt{\log N})$ which assembles an $N \times N$ square in a model which allows flexible glue strength between non-equal glues (This was independently discovered in [3]). This result is matched by a lower bound dictated by Kolmogorov complexity. For three other generalizations, we show that the $\Omega(\frac{\log N}{\log\log N})$ lower bound applies to $N \times N$ squares. At the same time, we demonstrate that there are some other shapes for which these generalizations allow reduced tile sets. Specifically, for thin rectangles with length $N$ and width $k$, we provide a tighter lower bound of $\Omega(\frac{N^{\frac{1}{k}}}{k})$ for the standard model, yet we also give a construction which achieves $O(\frac{\log N}{\log\log N})$ complexity in a model in which the temperature of the tile system is adjusted during assembly. We also investigate the problem of verifying whether a given tile system uniquely assembles into a given shape, and show that this problem is NP-hard.

## 1 Introduction

The tile assembly model extends the theory of Wang tilings of the plane [7] by adding a natural mechanism for growth. Under this model, we can consider a system of Wang tiles to be analogous to a computer program and the shape produced by the system analogous to the output of the program. Rothemund and Winfree [6] introduced the study of the *tile complexity* of self-assembled shapes, defined as the minimum number of distinct Wang tiles required to assemble the shape. They studied the tile complexity of self-assembled squares. We extend that study by considering the tile complexity of self-assembled squares and rectangles for natural generalizations of the tile assembly model.

In the *flexible glue* model, we remove the restriction imposed by Rothemund and Winfree [6] that different glue types have bonding strength of zero. In the *multiple temperature* model, we permit the temperature of the system to change during the assembly process. In the *multiple tile* or *q-tile* model, we allow tiles to cooperate by assembling into supertiles before being added to the growing result. In the *unique shape* model, we say that a system *uniquely assembles* a shape if the shape of its resultant supertiles is unique, though its produced supertile may not be unique.

In the standard model, Kolmogorov complexity dictates a lower bound of $\Omega(\frac{\log N}{\log\log N})$ for self-assembling $N \times N$ squares for almost all values of $N$ [6]. Adleman et al. [1] have shown how to reach this bound for all $N$. We show (in Thm. 6.1) that this lower bound also applies to each of our models except for the flexible glue model. For this model, Kolmogorov complexity only dictates an $\Omega(\sqrt{\log N})$ lower bound (as shown in Thm. 6.2). We show how to achieve this bound for all $N$ by encoding an arbitrary $(\log N)$-bit binary number into the glue function (Thm. 5.1).

Additionally, we provide a lower bound of $\Omega(\frac{N^{\frac{1}{k}}}{k})$ (see Thm. 3.1) for assembling *thin* $k \times N$ rectangles (rectangles whose width $k$ is less than $\frac{\log N}{\log\log N - \log\log\log N}$ in their length $N$). These lower bounds show that it can require significantly larger tile sets to assemble thinner rectangles than thicker rectangles. With this in mind, we utilize the multiple temperature model to construct a thin rectangle by first constructing a thicker rectangle using a small tile set. We then raise the temperature of the system so that portions of the larger rectangle fall apart, leaving the desired rectangle (Thm. 4.1).

Given a tile system, Adleman et al. [2] give an

*Department of Computer Science, Stanford University, Stanford, CA 94305. Email: gagan@cs.stanford.edu. Supported in part by Stanford Graduate Fellowship, NSF Grant IIS-0118173, NSF ITR Award Number 0331640, an SNRC grant, and grants from Microsoft and Veritas.

†Department of Mathematics and Mathematical Computer Science, Saint Louis University, 221 North Grand Blvd., St. Louis, MO 63103. Email: goldwamh@slu.edu.

‡Department of Computer Science, Northwestern University, Evanston, IL 60201. Email: kao@cs.northwestern.edu. Supported in part by NSF grant EIA-0112934.

§Department of Computer Science, Northwestern University, Evanston, IL 60201. Email: schwellerr@cs.northwestern.edu. Supported in part by NSF grant EIA-0112934.

algorithm to verify whether the tile system uniquely assembles into a given supertile. However, in the *unique shape* model, a tile system could *uniquely assemble* into a shape even if it does not produce a unique supertile. We investigate the problem of verifying whether a given tile system uniquely assembles into a given shape in the *unique shape* model, and show that this problem is co-NP-complete (Thm. 7.1). This result is in contrast to the polynomial-time verification algorithm for the standard model.

**Paper Layout:** In Section 2 we define the standard tile model as well as our generalized models. In Section 3 we introduce a construction for assembling $k \times N$ rectangles and provide a lower bound on the tile complexity of such rectangles. In Section 4 we use our construction to show how the multiple temperature model can reduce tile complexity when $k \ll N$. In Section 5 we use the flexible glue model to reduce the tile-complexity of assembling $N \times N$ squares from $\Theta(\frac{\log N}{\log \log N})$ to $\Theta(\sqrt{\log N})$. In Section 6 we discuss the lower bounds for assembling $N$-dimensional shapes for our models as dictated by Kolmogorov complexity. Finally, in Section 7, we show the co-NP-completeness result for the problem of verifying whether a given tile system uniquely assembles into a given shape.

## 2 Basics

**2.1 The Standard Model.** For alternate descriptions of the tile assembly model, see [1, 2, 6]. Briefly, the tiles in the model are four sided Wang tiles denoted by ordered quadruples $(\sigma_n, \sigma_e, \sigma_s, \sigma_w)$. The entries of the quadruples are glue types taken from an alphabet $\Sigma$ representing the north, east, south, and west edges of the Wang tile, respectively. For a given tile $t$, we use the function north($t$) to denote the glue on the north edge of $t$, and similarly define east($t$), south($t$) and west($t$). A *tile system* is an ordered quadruple $\langle T, s, G, \tau \rangle$ where $T$ is a set of tiles called the *tileset* of the system, $\tau$ is a positive integer called the *temperature* of the system, $s \in T$ is a single tile called the *seed* tile, and $G$ is a function from $\Sigma^2$ to $\{0, 1, \ldots, \tau\}$ called the *glue function* of the system. It is assumed that $G(x, y) = G(y, x)$, and $\exists$ $null \in \Sigma$ such that $G(null, x) = 0$ $\forall x \in \Sigma$. In the standard tile assembly model [1, 2, 6], the glue function is such that $G(x, y) = 0$ when $x \neq y$. When we are dealing with the standard glue model, we denote $G(x, x)$ by $G(x)$.

We define a *configuration* to be a mapping from $\mathbb{Z}^2$ to $T \bigcup \{\texttt{empty}\}$ where `empty` is a special tile whose edges have no bonding strength with any glue type. For a configuration $C$ we say a tile $t \in T$ is attachable at $(i, j)$ if $C(i, j) = \texttt{empty}$ and $G(\sigma_n, \text{south}(C(i, j + 1))) + G(\sigma_e, \text{west}(C(i + 1, j))) + G(\sigma_s, \text{north}(C(i, j - 1))) + G(\sigma_w, \text{east}(C(i - 1, j))) \geq \tau$. Informally, for a tile $t$ attachable to configuration $C$ at $(i, j)$, we define the act of attaching $t$ to $C$ at $(i, j)$ as a way to produce a new configuration from $C$ in which the empty tile at $(i, j)$ is replaced by $t$.

We define the adjacency graph of a configuration $C$ as follows. Let the set of vertices be the set of coordinates $(i, j)$ such that $C(i, j)$ is not empty. Let there be an edge between vertices $(x_1, y_1)$ and $(x_2, y_2)$ iff $|x_1 - x_2| + |y_1 - y_2| = 1$. We refer to a configuration whose adjacency graph is finite and connected as a *supertile*. For a supertile $S$, we denote the number of vertices (tiles) in the supertile by size($S$). For any supertile $S'$ whose adjacency graph is a subgraph of the graph of a supertile $S$, we say $S'$ is a *sub-supertile* of $S$.

A *cut* of a supertile is a cut of the adjacency graph of the supertile. In addition, for each edge $e_i$ in a cut define the *edge strength* $s_i$ of $e_i$ to be the glue strength (from the glue function) of the glues on the abutting edges of the adjacent tiles corresponding to $e_i$. Now define the *cut strength* of a cut $c$ to be $\sum s_i$ for each edge $e_i$ in the cut.

In the standard model, assembly takes place by *growing* a supertile starting with tile $s$ at position $(0, 0)$. We permit any $t \in T$ that is attachable at some position $(i, j)$ to attach and thus increase the size of the supertile. For a given tile system, any supertile that can be obtained by starting with the seed and attaching arbitrary attachable tiles is said to be *produced*. If this process comes to a point at which no tiles in $T$ can be added, the resultant supertile is said to be *terminally* produced. For a given shape $S$, we say a tile system $\Gamma$ uniquely produces shape $S$ if $\Gamma$ terminally produces a unique supertile with shape $S$. The *tile complexity* of a shape $S$ is the minimum tile set size required to uniquely assemble $S$ under a given assembly model.

### 2.2 Four Generalized Models

**The Multiple Temperature Model.** In the *multiple temperature* model, we replace the integer temperature $\tau$ in the tile system description with a sequence of integers $\{\tau_i\}_{i=1}^{k}$ called the *temperature sequence* of the system. We refer to a system with $k$ temperatures in its temperature sequence to be a *k-temperature* system.

In a $k$-temperature system, assembly takes place in $k$ phases. In the first phase, assembly takes place as in the standard model under temperature $\tau_1$. Phase 1 continues until no tiles can be added. In phase two, tiles can be added or removed under $\tau_2$. Specifically, at any point during phase 2 if there exists a cut of the

| Tile Complexities of Self-Assembling $k \times N$ Rectangles, $k \leq N$ | | | | |
|---|---|---|---|---|
| | Thin Rectangles LB | Thin Rectangles UB | Thick Rectangles LB | Thick Rectangles UB |
| Standard | $\frac{N^{\frac{1}{k}}}{k}$ (Thm. 3.1) | $N^{\frac{1}{k}}$ (Thm. 3.2) | $\frac{\log N}{\log\log N}$ (see [6]) | (see [1]) |
| Flexible Glue | $\frac{N^{\frac{1}{k}}}{k}$ (Thm. 3.1) | $N^{\frac{1}{k}}$ (Thm. 3.2) | $\sqrt{\log N}$ (Thm. 6.2) | (Thm. 5.1) |
| Multi-Temperature | $\frac{\log N}{\log\log N}$ (Thm. 6.1) | (Thm. 4.1) | $\frac{\log N}{\log\log N}$ (Thm. 6.1) | (see [1]) |
| $q$-Tile | $\frac{\log N}{\log\log N}$ (Thm. 6.1) | $N^{\frac{1}{k}}$ (Thm. 3.2) | $\frac{\log N}{\log\log N}$ (Thm. 6.1) | (see [1]) |
| Unique Shape | $\frac{N^{\frac{1}{k}}}{k}$ (Thm. 3.1) | $N^{\frac{1}{k}}$ (Thm. 3.2) | $\frac{\log N}{\log\log N}$ (Thm. 6.1) | (see [1]) |

Table 1: This table gives the general flavor of our results regarding upper and lower asymptotic bounds on tile complexity under our models for $k \times N$ rectangles, but the reader should reference precise details in the stated theorems. A rectangle is *thin* when $k < \frac{\log N}{\log\log N - \log\log\log N}$ and *thick* otherwise.

resultant supertile with cut strength less than $\tau_2$, we can remove the portion of the supertile occurring on the side of the cut not containing the seed tile. Also, at any point in the second phase we can add any tile in $T$ to the supertile if the tile is attachable at said position under temperature $\tau_2$. The second phase of this assembly continues until no tiles can be added or removed. We then go to phase 3 in which we can add and remove tiles under temperature $\tau_3$. We continue this process up through $\tau_k$. If during each phase of the assembly we must reach a point when no more tiles can be added or removed regardless of the choice of additions and removals, then we say the tile system *terminally* produces the shape assembled after phase $k$. If a tile system always ends with a unique terminally produced supertile $R$, we say the tile system *uniquely assembles* the shape of $R$ under the $k$-temperature model.

**The Flexible Glue Model.** In the *flexible glue* model, we remove the restriction that $G(x, y) = 0$ for $x \neq y$. This reduces the lower bound for assembling N dimensional shapes from $\Omega(\frac{\log N}{\log\log N})$ to $\Omega(\sqrt{\log N})$.

**The Unique Shape Model.** In the *unique shape* model, we redefine what we mean for a system to uniquely produce a shape $S$. In this model, a tile system uniquely produces a shape $S$ if the only terminal supertiles produced by the tile system are of shape $S$. Thus we allow the system to produce many different supertiles as long as they all have the desired shape.

**The $q$-Tile Model.** In the *$q$-tile* or *multiple tile* model we allow tiles in the system to combine into larger supertiles of size at most $q$ before being added to the growing seed supertile. Specifically, for a tile set $T$, we define a set of *addable* supertiles $W_T$ as follows: First, $T \subseteq W_T$. Second, if a supertile $r$ is produced by abutting two supertiles $s, t \in W_T$, then $r$ is also in $W_T$ if the sum of the edge strengths of the abutting edges of $s$ and $t$ reach or exceed the temperature of the system and $\text{size}(s) + \text{size}(r) \leq q$. In the $q$-tile assembly model, we allow the addition of supertiles from the set $W_T$ as well as $T$. Specifically, a supertile in $W_T$ can be attached to the growing seed supertile at any position such that the edge strengths of the abutting edges sum to at least the temperature of the system. For $q = 1$, we have $W_T = T$, which gives us the standard model of assembly.

## 3 The Assembly of $k \times N$ Rectangles

In this section, we present both an upper and lower bound for assembling $k \times N$ rectangles for arbitrary $k$ and $N$. The upper bound comes from a simple construction which constitutes a $k$-digit, base-$N^{\frac{1}{k}}$ counter and has tile complexity $\Theta(N^{\frac{1}{k}} + k)$. We also conjecture that it has optimal *time complexity* (as defined in [1]) of $\Theta(k + N)$. In later sections, we use modifications of this counter to show how both the 2-temperature and flexible glue model can reduce tile complexity of certain shapes. Additionally, for the case of $k = \log N$, this construction constitutes a $\log N$-bit, base-2 counter which assembles a $\log N \times N$ block. The tile set created by this special case was independently discovered by Cheng and Espanes [3]. They show that it assembles a $\log N \times N$ block in optimal time complexity $\Theta(N)$. This permits the assembly of $N \times N$ squares in optimal tile complexity and optimal time complexity as in [1], but does so in a much simpler fashion and uses only temperature $\tau = 2$.

THEOREM 3.1. *The tile complexity of self-assembling a $k \times N$ rectangle is $\Omega(\frac{N^{\frac{1}{k}}}{k})$ for the standard model, the unique shape model, and the flexible glue model.*

*Proof.* Suppose we had a tile system with fewer than $(\frac{N}{2(k!)})^{\frac{1}{k}}$ distinct tile types. Then there must be fewer than $\frac{N}{2(k!)}$ distinct $k$ tile columns consisting of these tiles. So in a $k \times N$ block, there must exist some $k$-tile column configuration which is repeated in more than $2(k!)$ columns. For each of these identical columns we can assign an ordering on the $k$ tiles that corresponds to a possible relative order in which the $k$ tiles of the given column could be placed. Since there are at most $k!$ orderings possible, we get that at least 3 of the identical columns must also have an identical ordering. From this we derive our contradiction as follows. Wherever the seed tile of the construction occurs, it lies to either the west or east of two of the identically placed columns. And if the seed tile occurs to the west (east) of a column, then all tiles east (west) of the column are determined by (1) the tiles and their positions in the column, and (2) the relative placement order of the tiles in the column. Thus we have a contradiction. This implies that the size of the tile set is at least $(\frac{N}{2(k!)})^{\frac{1}{k}} = \Omega(\frac{N^{\frac{1}{k}}}{k})$.

*Remark.* We also point out that the argument for Theorem 3.1 applies to any length $N$ shape whose height at each column is bounded by a value $k$.

THEOREM 3.2. *The tile complexity of self-assembling a $k \times N$ rectangle is $O(N^{\frac{1}{k}} + k)$ for the standard model, the unique shape model, and the $q$-tile model.*

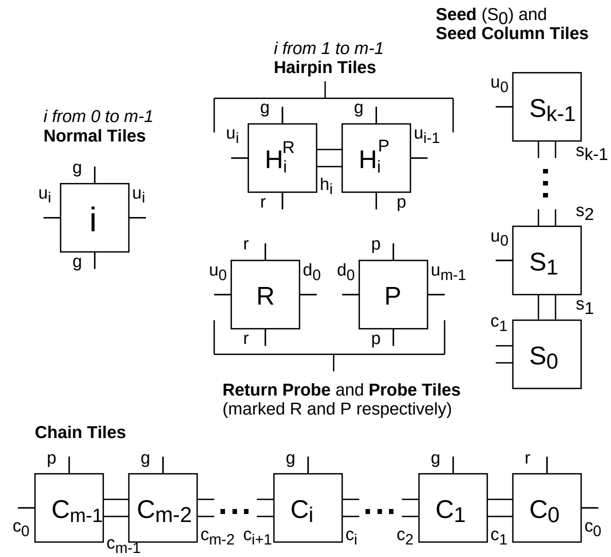



Figure 1: A tile set to assemble a $k \times m^k$ rectangle in $\Theta(k + m)$ tile complexity under the standard assembly model.

We show this by first providing a construction for the standard model and then arguing that the construction works for the unique shape model and the $q$-tile model as well.

*Proof for the Standard Model.* For a given $N$, let $m = \lceil N^{\frac{1}{k}} \rceil$. To show this bound, we give a general tile set to assemble a $k \times m^k$ rectangle in $O(N^{\frac{1}{k}} + k)$ tiles under the standard model. We then show how to adjust the tile set so it produces a $k \times N$ rectangle. The tile system we use constitutes a $k$-digit base-$m$ counter. The system has temperature $\tau = 2$ and the tile set and glue strength function are given in Figure 1.

The assembly takes place as follows. The north edge of the seed tile produces a length $k$ *seed column* from the seed column tiles. The west edge of the seed produces a length $m$ *chain* from the chain tiles. The 0 normal tile can then fill in all the rows and columns up until column $m-1$. In this column the $H_1^P$ hairpin tile must be placed. This causes a *hairpin* growth which causes another length $m$ chain of chain tiles to be placed in the first row. The next section of $m$ columns are then filled with the 0 normal tile in all rows but the second, which will contain the 1 normal tile or a hairpin tile. In general, when the $C_{m-1}$ chain tile is placed, probe tiles are added on top of each other until a row is found that does not consist of the $m-1$ normal tile. In such a row the appropriate pair of hairpin tiles are placed and a downward growing column of return probe tiles are placed until the bottom row is encountered, at which point the $C_0$ chain tile is placed to start the length $m$ chain growth over again in the first row. If no such row is encountered, the assembly finishes. The idea here is that the bottom *chain* row represents the least significant digit of the counter, thus incrementing every column. After each block of $m$ we need to find the most significant digit that requires incrementation, as well as rollover all the trailing digits displaying $m-1$ to 0. The hairpin tiles perform the incrementation, while the probe and return probe tiles perform the rollover.

It is easy to see that for a given $k$ and $m$, this construction assembles into a $k \times m^k$ block. In addition, we can adjust the glue types of the west edges of the seed column and seed tiles to represent any $k$-digit, base-$m$ number. By doing this we can start the counter at any number between 0 and $m^k - 1$. Thus we can designate the length of the constructed shape to be any number between 1 and $m^k$. Therefore, to assemble a $k \times N$ block, we use the above tile set with $m = \lceil N^{\frac{1}{k}} \rceil$ and the west edges of the seed and seed columns tiles set to represent the number $m^k - N$. Thus we can construct a $k \times N$ block in $4m + k = O(N^{\frac{1}{k}} + k)$ tiles.

*Proof for the Unique Shape Model.* This follows from the construction for the standard model.

*Proof for the q-Tile Model.* We omit this proof in this version.

## 4 Reducing Tile Complexity with the 2-Temperature Model

Now we show how a multiple temperature model can reduce tile complexity for assembling thin rectangles. For a given $k$ and $N$ with $k \ll N$, the idea is to use a modification of the construction from section 3 to build a $j \times N$ rectangle for optimal (bigger than $k$) $j$ such that the top $j-k$ rows are less stable than the bottom $k$. We then raise the temperature of the system to cause all but the bottom $k$ rows to fall apart. We then compare the complexity of this construction with the lower bound for the standard model given in section 3 to show that the 2-temperature model can beat a lower bound for the standard model.

**Minimizing the Complexity.** For a given $j$ and $N$, assembling a $j \times N$ rectangle using the construction of Theorem 3.2 yields an upper bound that is a function of $j$. If we are only interested in constructing a rectangle of length $N$, but do not care about the width, we can choose $j$ as a function of $N$ such that the tile complexity is minimized. To do this we choose a value for $j$ that balances the size of $N^{\frac{1}{j}}$ and $j$. For

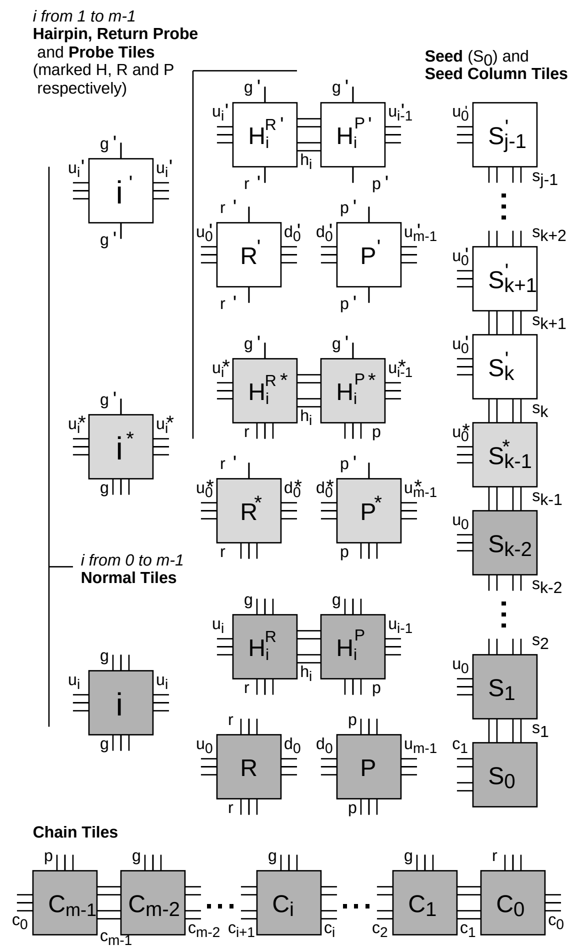



Figure 2: A tile set to assemble a $k \times N$ rectangle in $\Theta(j+m)$ tile complexity for $m = \lceil N^{\frac{1}{j}} \rceil$ under the 2-temperature model with $\tau_1 = 4$ and $\tau_2 = 6$.

$j = \frac{\log N}{\log\log N - \log\log\log N} = \Theta(\frac{\log N}{\log\log N})$, the term $N^{\frac{1}{j}}$ is $\frac{\log N}{\log\log N}$. Thus the number of tiles used in our construction for building a $j \times N$ block for such a $j$ is $j + N^{\frac{1}{j}} = \Theta(\frac{\log N}{\log\log N})$, which meets the lower bound dictated by Kolmogorov complexity [6]. We use this result in the following theorem.

Theorem 4.1. *Under the 2-temperature model, the tile complexity of self-assembling a $k \times N$ rectangle for an arbitrary integer $k \geq 2$ is $O(\frac{\log N}{\log\log N})$.*

*Proof.* In the case that $k$ exceeds $\lceil \frac{\log N}{\log\log N - \log\log\log N} \rceil$, we can simply use a single temperature model to build the two perpendicular axes of the rectangle in optimal $O(\frac{\log N}{\log\log N})$ complexity. The addition of a constant number of tiles can then fill in the rest of the

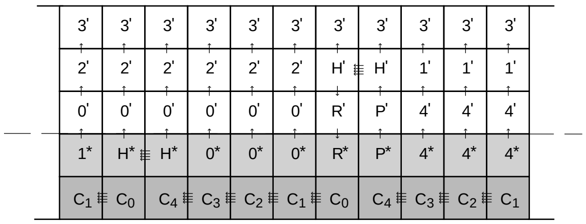



Figure 3: Here is a typical section of a self-assembled $5 \times N$ block from the tile set in Figure 2 that will break down to a $2 \times N$ block under temperature 6. Edges with strength 4 are marked by 4 adjacent arrows, while edges with strength 1 are marked with 1. All remaining edges have strength 3.

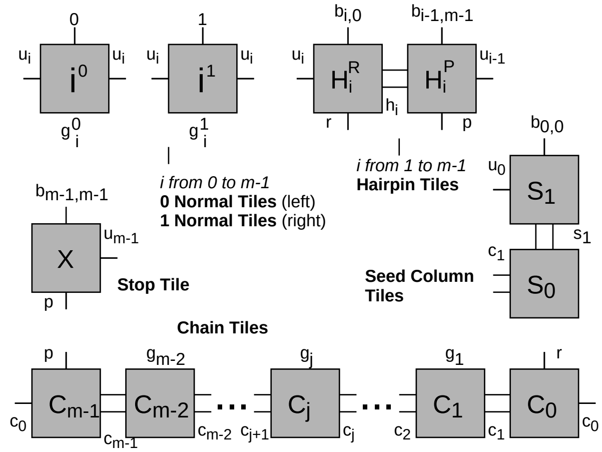



Figure 4: This tile set creates a $2 \times n$ block whose top row represents a given $n$-bit binary number $b$. Here $b_{ij}$ is the value of the bit in position $im + j$ in $b$. The glue function for glues $g_i^1$ and $g_j$ for $i$ from 0 to $m-1$ and $j$ from 1 to $m-2$ is $G(g_i^1, g_j) = b_{ij}$. All other pairs of non-equal glues have strength 0.

rectangle. Otherwise, to reach the bound we define a tile system that assembles a $j \times N$ block for optimal value $j = \lceil \frac{\log N}{\log\log N - \log\log\log N} \rceil$ under temperature $\tau_1$ and breaks down to a smaller $k \times N$ block in the second phase of the two-temperature assembly process. As in the construction from Theorem 3.2, let $m = \lceil N^{\frac{1}{j}} \rceil$. Consider the two-temperature tile system with $\tau_1 = 4$ and $\tau_2 = 6$ and the tile set and glue strength function given in Figure 2.

Under temperature $\tau_1$, this tile system assembles a $j \times N$ block in exactly the same fashion as our single temperature system from Theorem 3.2. However, for each tile in the top $j-k$ rows other than the seed column tiles, the north and south edges have glue strength 1. Therefore, since no glue in the system exceeds strength 4, we are assured that any cut consisting of one north-south edge and one east-west edge has cut strength less than $\tau_2 = 6$. Therefore, under the two-temperature model each tile in the top $j - k$ rows can be removed one at a time, starting at the northwest corner of the $j \times N$ block. We are then left with the bottom $k$ rows. Since our design ensures that each edge in the bottom $k$ rows has strength 3 or greater, no cut of the bottom $k \times N$ supertile can have cut strength less than 6. Thus no cut can break up the remaining $k \times N$ block. And since any alternate choice of cuts would only expedite the process to this end, we have uniquely constructed a $k \times N$ block in $10m + j = O(N^{\frac{1}{j}} + j) = O(\frac{\log N}{\log\log N})$ tile complexity under the two-temperature assembly model.

For small values of $k$, this beats the lower bound for the standard model. Consider $k = \frac{\log N}{2\log\log N}$. For such $k$ the value of $N^{\frac{1}{k}}$ is

$$N^{\frac{2\log\log N}{\log N}} = (2^{\log N})^{\frac{2\log\log N}{\log N}} = (\log N)^2.$$

Thus from Theorem 3.1 the lower bound for the standard model is $\Omega(\frac{N^{\frac{1}{k}}}{k}) = \Omega((\log N)(\log\log N))$. Since this bound only gets higher for smaller values of $k$, the 2-temperature model beats it for all $k$ between 2 and $\frac{\log N}{2\log\log N}$.

## 5 Assembling $N \times N$ Squares in $O(\sqrt{\log N})$ Tiles

Kolmogorov complexity dictates an $\Omega(\frac{\log N}{\log\log N})$ lower bound on the tile complexity of self-assembling $N \times N$ squares for the standard model in which we limit the glue function so that $G(x, y) = 0$ for $x \neq y$. However, if we lift this restriction, the bound no longer applies. In this section, we show that the tile complexity of self-assembling $N \times N$ squares is $\Theta(\sqrt{\log N})$ under the flexible glue model for almost all $N$.

Theorem 5.1. *The tile complexity of self-assembling $N \times N$ squares is $O(\sqrt{\log N})$ under the flexible glue model.*

*Proof.* The trick as introduced in [6] is to be able to initialize a fixed length binary counter to any arbitrary $(\log N)$-bit binary number. This can be done trivially in $O(\log N)$ tile complexity. By simulating base conversion, it can be done in $O(\frac{\log N}{\log\log N})$ [1]. Here we construct a $2 \times \log N$ block in $O(\sqrt{\log N})$ tiles such that the top row of the block encodes a given binary number $b$. We accomplish this by taking advantage of the flexible

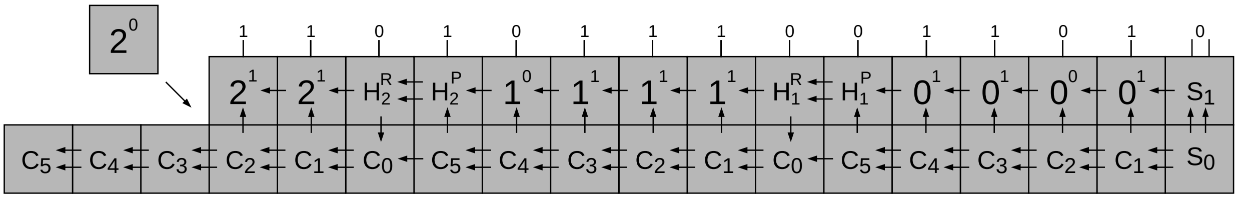



Figure 5: Assembling an arbitrary $n$-bit binary number in $O(\sqrt{n})$ tiles. Here we show the construction for $n = 36$ and binary number $b = \ldots 0110101110011010$.

glue function and encoding $b$ into the glue function. Let $n = \lfloor \log N \rfloor + 1$ and $m = \lceil \sqrt{n}\ \rceil$. Let $b$ be a given $n$-bit binary number. Let $b_{ij}$ be the bit in position $im + j$ in $b$. We use the tile set and glue function from Figure 4 to construct a $2 \times n$ block such that the top row of the block represents the number $b$. For convenience we denote some glues by the symbols 0 and 1.

To simplify the illustration, the tile set in Figure 4 assumes that $m^2 = n$. If this is not the case, we can do as before in Theorem 3.2 and initialize our 2-digit counter to an arbitrary value $c$ so that we assemble a block of length exactly $n$. At any rate, our construction uses $5\lceil \sqrt{n}\ \rceil + 1$ tiles and constructs a $2 \times n$ block in the same fashion as the general $k \times n$ assembly. Additionally, our glue function assures us that for the $(i, j)$ digit of $b$, the corresponding position in the top row of our construction can only be tiled with either a hairpin, seed column, or normal tile with north edge glue equal to $b_{ij}$.

To complete the $N \times N$ square we need to create a fixed length binary counter that is initialized to our given binary number $b$ and grows north, incrementing row by row, until $2^n$ is reached. The addition of a constant number of stairstep tiles can then finish off the square, as shown in [6]. The fixed length binary counter can also be implemented in a constant number of tiles as is done in [1, 6]. Alternatively, we can use our construction from Section 3 for building a $k \times N$ block for $k = \log N$. This yields a binary counter consisting of eight tiles plus $\log N$ seed column tiles. However, since we already have a "seed column" via our $2 \times n$ *seed block*, we can omit the seed column tiles. Thus the total number of tiles used for the assembly of an $N \times N$ square is only a constant greater than the $\Theta(\sqrt{\log N})$ tiles used to build the $2 \times n$ seed block.

## 6 Kolmogorov Lower Bounds

Rothemund and Winfree [6] have shown that Kolmogorov complexity dictates a $\Omega(\frac{\log N}{\log \log N})$ lower bound on the tile complexity for self-assembling $N \times N$ squares for almost all $N$. We show that their proof generalizes to the multiple temperature model, the unique shape model, and the $q$-tile model. For the flexible glue model, we modify their argument to get a $\Omega(\sqrt{\log N})$ lower bound from Kolmogorov complexity, making our construction in Theorem 5.1 tight.

Theorem 6.1. *The tile complexity of self-assembling $N \times N$ squares is $\Omega(\frac{\log N}{\log \log N})$ for the $q$-tile model, the multiple temperature model, and the unique shape model.*

*Proof.* We note that any tile system with tileset size of $n$ in the above models can be represented in $O(n \log n)$ bits assuming the maximal temperature of the system is bounded by a constant. Additionally, for each model there exists a constant size Turing machine that takes as input a tile system and outputs the maximum length of the shape produced by the given tile system under the corresponding model. As discussed in [6], Kolmogorov complexity thus dictates a $\Omega(\frac{\log N}{\log \log N})$ lower bound on the tile complexity of assembling an $N \times N$ square for almost all $N$. We omit the details of this proof in this version.

Theorem 6.2. *The tile complexity of self-assembling $N \times N$ squares is $\Omega(\sqrt{\log N})$ under the flexible glue model for almost all $N$.*

*Proof.* First we note that there exists a Turing machine of constant size that takes as input a tile system and outputs the maximal dimension of the produced terminal supertile, assuming there is one, under the flexible glue model. Thus, if a tile set that assembles an $N \times N$ square is given as input, the output is $N$. Therefore, by Kolmogorov complexity, the sum of the size of the Turing machine plus the length of the corresponding binary representation of the tile set must have the lower bound $\Omega(\log N)$ for almost all $N$. And since the machine's size is fixed, this is a bound on the number of binary digits required to represent the tile set. We can represent a flexible glue tile set of size $n$

tiles and temperature $\tau$ bounded by a constant with $f(n) = 4n \log(4n) + (4n)^2 \log \tau = O(n^2)$ bits as follows. For each tile we have four sides for which we must denote a glue. We need at most $4n$ distinct glues, and for each glue we need to store the bonding strength with every other glue in the model. This can be stored in a $4n \times 4n$ matrix. Now for a given $N$, let $\beta(N)$ be the cardinality of the minimum tile set for assembling an $N \times N$ square under our flexible model. Then for almost all $N$ there exist constants $C_1$, $C_2$, and $C_3$ such that

$$C_1 \log N \leq f(\beta(N)) \leq C_2 \beta(N)^2 \leq C_3 \beta(N) \sqrt{\log N}.$$

So $\beta(N) = \Omega(\sqrt{\log N})$ for almost all N.

## 7 Shape Verification in the *Unique Shape* Model

Given a shape, Adleman et al. [2] studied the problem of finding the minimum set of tiles which uniquely produces a supertile $A$, such that $A$ has the given shape. To show that the decision version of the problem is in NP, they gave an algorithm to verify whether a given set of tiles uniquely produces a supertile which has the given shape. Note that they insist that the tile system should assemble into a unique terminal supertile. Under the *unique shape* model, we say that a tile system $\mathbb{T}$ uniquely produces a shape if all terminal supertiles of $\mathbb{T}$ (henceforth called $Term(\mathbb{T})$) have the given shape, and no larger supertiles are produced. We note that this definition automatically implies that if a tile system uniquely produces a shape, then $Term(\mathbb{T})$ is non-empty.

Definition 7.1. Unq-Shape($\langle T, \mathcal{S}, G, \tau \rangle$, $W$) *is the problem of verifying whether the tile system* $\langle T, \mathcal{S}, G, \tau \rangle$ *uniquely assembles into the shape* $W$ *under the unique shape model.*

Theorem 7.1. Unq-Shape *is co-NP-complete. It remains NP-hard even when the glue function is restricted such that* $G(\alpha, \beta) = 0$ *for* $\alpha, \beta \in \Sigma$ *with* $\alpha \neq \beta$ *and the temperature* $\tau = 2$.

We begin by showing that Unq-Shape is in co-NP, and then show that it is NP-hard (and thus co-NP-hard).

Lemma 7.1. Unq-Shape *is in co-NP.*

*Proof.* For this, we need to show that if an instance of Unq-Shape is false, i.e. if the tile system in the instance does not assemble uniquely into the given shape, then there is a short proof of the fact. By definition, a given tile system $\mathbb{T} = \langle T, \mathcal{S}, G, \tau \rangle$ does not uniquely assemble into the given shape $W$ iff one of the following occurs:

1. A terminal supertile of a shape different from $W$ can be assembled. In this case, $Term(\mathbb{T})$ contains a supertile $A$ with a shape different from $W$. Then $A$, along with the order in which the tiles join to assemble $A$ would suffice as a proof. In order to check this proof, we first verify that $A$ can indeed be assembled by adding the tiles in the order specified. Then we check that $A$ is a terminal supertile by simply testing whether any tile is attachable at any of the empty sites adjacent to it.

2. A supertile $C$ of size one larger than the given shape can be assembled. Note that this supertile need not be terminal. We note that if any such $C$ exists, then there exists such a $C$ with size one larger than $W$. In this case, $C$, along with the order in which tiles are added to assemble $C$ would suffice as a proof. We verify this proof by checking that $C$ can indeed be assembled by adding the tiles in the order specified, and that the size of the supertile is larger than the shape.

In both cases, the size of the proof is linear in the instance size, and the verification algorithm runs in time quadratic in the instance size.

Lemma 7.2. Unq-Shape *is NP-hard.*

*Proof.* The proof uses a construction proposed by LaBean and Lagoudakis in [4] to solve SAT by exploiting the massive parallelism possible in DNA operations to emulate a non-deterministic device that solves SAT.

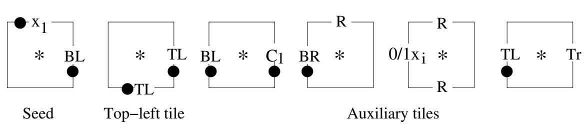
We reduce 3-SAT to Unq-Shape using their construction. Given any instance of 3-SAT with $m$ clauses and $n$ different variables, we construct an instance of Unq-Shape, ($\langle T, \mathcal{S}, G, 2 \rangle$, $W$) with $|T| = O(m + n)$, and the size of the shape $W$ of order $O(mn)$. There is a one-to-one correspondence between variable assignments and the supertiles in $Term(\mathbb{T})$, i.e every possible assignment to variables is associated with a distinct terminal supertile. If the assignment satisfies the formula, the terminal supertile associated with it has the shape of an $(n+2) \times (m+3)$ rectangle; else the shape associated with the assignment has the shape of an $(n+2) \times (m+3)$ rectangle with its top-right corner missing. Thus if the tile system uniquely assembles into a rectangle with its top-right corner missing, then no assignment satisfies the 3-SAT formula; on the other hand, if the tile system does not uniquely assemble into this corner-missing rectangle, then there is at least one assignment which satisfies the 3-SAT formula. Thus a polynomial time algorithm to decide Unq-Shape will result in a polynomial time algorithm for 3-SAT.

The idea of the reduction is to have the bottom row of the rectangle encode the clauses, the left column encode the variables, and let the second column correspond to a possible assignment of values to the variables. The rest of the assembly evaluates the formula at that assignment, and the row below the top row represents which of the clauses get satisfied by the assignment. A tile gets added to the top-right corner iff the assignment satisfies all the clauses.

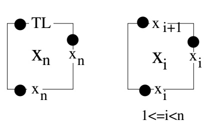
Figures 6(a) and (b) show two of the terminal supertiles of a tile system corresponding to the formula $(x_1 + x_2 + \overline{x_3})(x_1 + \overline{x_2} + x_3)(\overline{x_1} + x_2 + \overline{x_3})$. Each small square represents a tile. The label of an edge corresponds to the binding glue on the edge. All glues are of strength 1, except for the glues corresponding to those edge labels which are accompanied by a black circle – these are of strength 2. The glues on edges without any labels do not match. The temperature $\tau = 2$. Note that since the assignment $x_1 = 0$, $x_2 = 1$, $x_3 = 1$ satisfies the formula, a tile gets attached in the top-right corner in Figure 6(a). Also, since the assignment $x_1 = 0$, $x_2 = 1$, $x_3 = 0$ does not satisfy the formula, the rectangle in Figure 6(b) is missing a tile in the top-right corner.

Take any 3-SAT formula with $n$ variables, $x_1, x_2, \ldots, x_n$, appearing in $m$ clauses $C_1, C_2, \ldots, C_m$. We define a tile system, $\langle T, \mathcal{S}, G, 2\rangle$ (the temperature is fixed at 2), corresponding to it as follows:

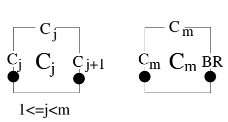
- **Seed and auxiliary tiles.**



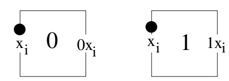

- **Variable tiles.** There are $n$ tiles encoding the variables for the first column. These tiles attach on top of the seed tile to give the first column. The east side of tiles $x_i$ provides a strength-2 glue to the assignment tiles.




- **Clause tiles.** There are $m$ tiles encoding the clauses for the bottom row. These tiles attach to the right of the seed tile to give the bottom row. The north side of $C_j$ has a strength-1 glue $C_j$ to attach the computation tiles.

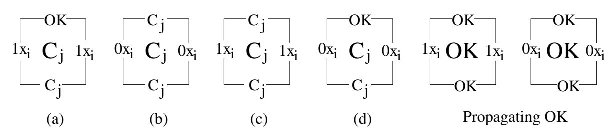



- **Assignment tiles.** Each variable can take one of two values, so there is a total of $2n$ tiles for assigning values. Tile $A0_i$ and $A1_i$ can attach with strength 2 to the east side of $x_i$, and have glue $0x_i$ and $1x_i$ respectively on their east side to attach the computation tiles. Any supertile with a full second column has either $A0_i$ or $A1_i$ attached to the east to $i^{th}$ variable tile, representing the variable assignment which will be evaluated by the terminal supertile resulting from this supertile.



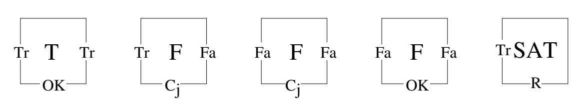
- **Computation tiles.** For any clause-variable pair, $C_j$ and $x_i$, three cases arise:

  1. $x_i$ is present as a positive literal: In this case, tiles (a) and (b) shown below are present.
  2. $x_i$ is present as a negative literal: In this case, tiles (c) and (d) shown below are present.
  3. $x_i$ is not present in clause $C_j$: For this case, tiles (b) and (c) shown below are present.




In all the three cases, there are $2n$ tiles to propagate $OK$ upwards. All these tiles have $OK$ on their north and south sides; $n$ of the tiles have $1x_i$ on their east and west sides, while the remaining $n$ tiles have $0x_i$ on their east and west sides.

- **Final check tiles.**

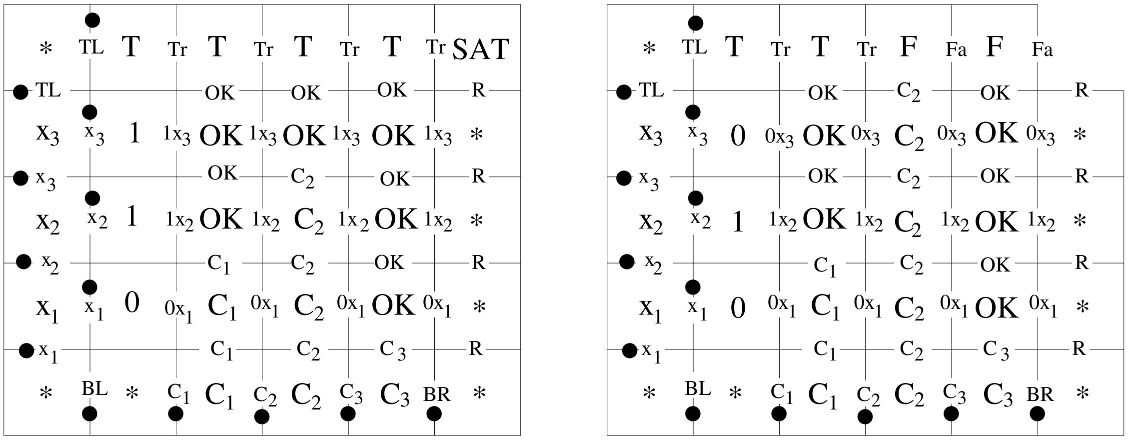


The tile labeled $SAT$ attaches to the top-right corner only if the formula is satisfied by the assignment in the second column of the supertile.

## 8 Further Research Directions

- The Kolmogorov lower bounds for the assembly of $N \times N$ squares do not apply if the temperature of the system is a large exponential function of $N$. If this is allowed, is it possible to reduce the tile complexity of assembling $N \times N$ squares?

- Are there any shapes in which the $q$-tile model or the unique shape model can be used to reduce the size of the minimum tileset, or can it be shown that




(a) (b)

Figure 6: Two terminal supertiles for the formula $(x_1 + x_2 + \overline{x_3})(x_1 + \overline{x_2} + x_3)(\overline{x_1} + x_2 + \overline{x_3})$.

these models are equivalent to the standard model with respect to tile complexity?

- In the $q$-tile model we put a bound of size $q$ on the size of addable supertiles. If we remove this bound, the set of supertiles that can be added is possibly infinite and our proof of the Kolmogorov lower bound therefore no longer holds. For this $\infty$-*tile*, or *two-handed* assembly model, what can we say about the lower bound for assembling $N \times N$ squares?

- Can it help to use more than two temperatures in the multiple temperature model? If so, does the temperature need only be monotonically increasing or can it help to raise and lower the temperature?

- Our construction for the flexible glue model shows how to create rectangles with width at least logarithmic in their length, but does not work for thinner rectangles. However, for thick rectangles with less than logarithmic width we can combine the flexible glue model with the 2-temperature model to attain the tight $\Theta(\sqrt{\log N})$ tile complexity. Is it possible to assemble such a rectangle using only the flexible glue model?

## References


[1] L. Adleman, Q. Cheng, A. Goel, and M. Huang. Running time and program size for self-assembled squares. In *Proceedings of the 33nd Annual ACM Symposium on Theory of Computing*, pages 740–748, 2001.

[2] L. Adleman, Q. Cheng, A. Goel, M. Huang, D. Kempe, P. Espanes, and P. Rothemund. Combinatorial optimization problems in self-assembly. In *Proceedings of the 34nd Annual ACM Symposium on Theory of Computing*, pages 23–32, 2002.

[3] Q. Cheng and P. M. de Espanes. Resolving two open problems in the self-assembly of squares. Technical Report 793, University of Southern California, June 2003.

[4] M. Lagoudakis and T. LaBean. 2D DNA Self-assembly for Satisfiability. In *Proceedings of the 5th DIMACS Workshop on DNA Based Computers held at MIT, Cambridge*, 1999.

[5] M. Li and P. Vitanyi. *An Introduction to Kolmogorov Complexity and Its Applications (Second Edition).* Springer Verlag, New York, 1997.

[6] P. Rothemund and E. Winfree. The program-size complexity of self-assembled squares. In *Proceedings of the 32nd Annual ACM Symposium on Theory of Computing*, pages 459–468, 2000.

[7] H. Wang. Proving theorems by pattern recognition. *Bell Systems Technical Journal*, 40:1–42, 1961.

[8] E. Winfree. *Algorithmic Self-Assembly of DNA.* PhD thesis, California Institute of Technology, Pasadena, 1998.