# Two lower bounds for self-assemblies at temperature 1

Ján Maňuch
School of Computing Science, Simon Fraser University
8888 University Drive, Burnaby BC, V5A 1S6, Canada
Email: jmanuch@sfu.ca

Ladislav Stacho
Department of Mathematics, Simon Fraser University
8888 University Drive, Burnaby BC, V5A 1S6, Canada
Email: lstacho@sfu.ca

Christine Stoll
Department of Mathematics, Simon Fraser University
8888 University Drive, Burnaby BC, V5A 1S6, Canada
Email: cstoll@sfu.ca

**Abstract**

Using the Tile Assembly Model proposed by Rothemund and Winfree, we give two lower bounds on the minimum number of tile types needed to uniquely assemble a shape at temperature 1 under a natural assumption that there are no binding domain mismatches (any two adjacent tiles either form a bond or else both touching sides of the tiles are without glues). Rothemund and Winfree showed that uniquely assembling a full $N \times N$ square (a square where there is a bond between any two adjacent tiles) at temperature 1 requires $N^2$ distinct tile types, and conjectured that the minimum number of tile types needed to self-assemble an $N \times N$ square (not a full square) is $2N - 1$. Our lower bounds imply that a tile system that uniquely assembles an $N \times N$ square without binding domains mismatches, requires at least $2N - 1$ tile types.

## 1 Introduction

Self-assembly is the process by which simple parts autonomously assemble into larger, more complex objects. Self-assembly occurs in nature, for example, when atoms combine to form molecules, and molecules combine to form crystals. It has been suggested that intricate self-assembly schemes will ultimately be useful for circuit fabrication, nano-robotics, DNA computing, and amorphous computing [1, 2, 6, 12]. To study the process of self-assembly we use the Tile Assembly Model proposed by Rothemund and Winfree [7] which uses Wang square tiles with glues of different types on the sides. Branched DNA molecules [9] provide a direct physical motivation for this model. DNA double-crossover molecules, each bearing four "sticky ends" analogous to the four sides of a Wang tile, have been designed to self-assemble into a periodic two dimensional lattice [3, 4, 8, 11]. The binding interactions between double-crossover molecules may be redesigned by changing the base sequence of their sticky ends, thus allowing arbitrary sets of Wang tiles to be investigated in the laboratory.

The Tile Assembly Model considers the assembly of square blocks called "tiles" and a set of glues called "binding domains". Each of the four sides of a tile can have a glue on it that determines interactions with neighbouring tiles. We assume that there is an infinite supply of tiles of each tile

type. The process of self-assembly is initiated by a single seed tile and proceeds by attaching tiles one by one. A tile can only bind to the growing complex if it binds strongly enough, as determined by the *temperature* $\tau$. The temperature measure the "cooperativity" of the binding interactions. In this paper we only consider self-assembly at temperature $\tau = 1$. At this temperature, there is no cooperativity, one matching glue is enough for a tile to attach to the growing structure. Self-assemblies at temperature $\tau > 1$ have higher computational power, but they might be more difficult to set-up in lab setting.

A measure of complexity of self-assembly is the minimum number of distinct tile types needed to uniquely assemble a certain shape. If we want to assemble any scaled version of a given shape, then the complexity depends on the expressibility (Kolmogorov complexity) of the shape [10]. If we want to assemble a given shape with a prescribed size, then in [7] it was observed that assembling an $N \times N$ full square (a square where there is a bond between any two adjacent tiles) at $\tau = 1$ requires $N^2$ distinct tile types. In fact, we will show that any uniquely produced 2-connected full assembly with $p$ tiles requires $p$ tile types (all tiles must be distinct) at $\tau = 1$. If we do not require a bond between every two adjacent tiles of a uniquely produced assembly the number of used tile types can be significantly smaller. In particular, in [7] a tile system that uniquely assembles an $N \times N$ square using only $2N - 1$ tile types was described. The construction is based on a comb-like backbone graph of the $N \times N$ square. Moreover, it was conjectured in [7] that this is best possible.

In this paper we give two lower bounds on the number of tile types in tile systems that uniquely produce a (non-full) assembly at temperature 1 and using the simplest energy model which uses glues with the same strengths and does not penalize mismatches. Our lower bounds are based on a natural assumption that there are no binding domain mismatches (any two adjacent tiles either form a bond of glues of the same type or both sides have no glue on them). Without this assumption the tile system would be prone to errors during a self-assembly process. For instance, consider the assembled shape on right in Figure 2. This assembly contains one domain mismatch: between starting tile $S$ and tile 5. Even though this assembly is terminal (cannot be extended), there is a chance that another copy of tile 6 would push out and replace the starting tile which would lead to assembly incorrect infinitely growing assemblies. We also note that all tile systems used in the literature are without binding domain mismatches.

One of the lower bounds is the radius of the subgraph of the lattice $\mathbb{Z} \times \mathbb{Z}$ induced by the assembled shape plus one, and the other is its Manhattan diameter plus one. As a consequence of these results, we also show that a tile system that uniquely assembles an $N \times N$ square where the binding domains of any two adjacent tiles match, requires at least $2N - 1$ tiles, thus confirming the conjecture in [7] in this case.

An extended abstract of this paper appeared in [5].

## 2   Self-assembly model

We will consider the square lattice, i.e., the graph with vertex set $\mathbb{Z} \times \mathbb{Z}$ and edge set $\{(u, v) : |u, v| = 1\}$. The directions $\mathcal{D} = \{N, E, S, W\}$ are used to indicate the natural directions in the lattice. Formally, they are functions from $\mathbb{Z} \times \mathbb{Z}$ to $\mathbb{Z} \times \mathbb{Z}$: $N(x, y) = (x, y + 1)$, $E(x, y) = (x + 1, y)$, $S(x, y) = (x, y - 1)$, and $W(x, y) = (x - 1, y)$. Note that $E^{-1} = W$ and $N^{-1} = S$.

A tile is a square with the north, east, south, and west edges labeled from some alphabet $\Sigma$ of binding domains (glues). Formally, a tile $t$ is a 4-tuple $(t_N, t_E, t_S, t_W) \in \Sigma^4$, indicating the binding domains on the north, east, south, and west side, respectively. We will use *null* to indicate the lack of a binding domain, and will assume $null \in \Sigma$. The special tile $empty = (null, null, null, null)$ represents an empty space when placed onto the grid. A *configuration* on a set of tiles $T$ is a map

$C : \mathbb{Z} \times \mathbb{Z} \to T$. We define the *vertex set* of configuration $C$ as $V(C) = \{(x,y) : C(x,y) \neq empty\}$. A configuration $C$ is *finite* if $V(C)$ is finite. We will refer to $C(x,y)$ as the tile at the vertex $(x,y)$ in $C$. Given a configuration $C$ and a set of vertices $V \subseteq \mathbb{Z} \times \mathbb{Z}$, a *sub-configuration* of $C$ induced by $V$ is the map $C[V] : \mathbb{Z} \times \mathbb{Z} \to T$ such that $C[V](x,y) = C(x,y)$ for all $(x,y) \in V$, and $C[V](x,y) = empty$, otherwise. If $G$ is any subgraph of the lattice graph, then we sometimes abuse the notation of $C[V(G)]$ to simply $C[G]$. Given two configurations $C$ and $D$, we define their union to be the following map from $\mathbb{Z} \times \mathbb{Z}$ to $T \cup \{\infty\}$:

$$(C \cup D)(x,y) = \begin{cases} C(x,y) & \text{if } D(x,y) = empty \\ & \text{or } C(x,y) = D(x,y), \\ D(x,y) & \text{if } C(x,y) = empty \\ & \text{or } C(x,y) = D(x,y), \\ \infty & \text{otherwise.} \end{cases}$$

Note that $C \cup D$ is a configuration whenever it is a map to $T$. Equivalently, $C \cup D$ is not a configuration if there exists $(x,y) \in V(C) \cap V(D)$ such that $C(x,y) \neq D(x,y)$.

A *strength function* $g : \Sigma \times \Sigma \to \mathbb{N} = \{0,1,2,...\}$ measures the interaction strength between binding domains. We denote by $s_\Sigma$ the strength function satisfying $s_\Sigma(\sigma, \sigma') = 1$, if $\sigma = \sigma' \neq null$, and $s_\Sigma(\sigma, \sigma') = 0$ otherwise. We call $s_\Sigma$ a *simple* strength function. In this paper we restrict ourselves to simple strength functions.

Given a tile $t$, a configuration $C$, and a direction $d$, we denote the interaction strength in configuration $C$ between tile $t$ at position $(x,y)$ and its respective neighbouring tile by

$$g_d^C(t,x,y) = g(t_d, C(d(x,y))_{d^{-1}}).$$

Note that we do not require that $C(x,y) = t$. In particular, if $C(x,y) \neq t$, then $g_d^C(t,x,y)$, $d \in \mathcal{D}$ tells us how $t$ would bind if it were in $C$. Given $(x,y) \in \mathbb{Z} \times \mathbb{Z}$ and $d \in \mathcal{D}$, we say that there is a bond between positions $(x,y)$ and $d(x,y)$ in $C$ if $g_d^C(C(x,y),x,y) = 1$, i.e. the binding domain on the abutting sides of the two tiles is the same.

Under the Tile Assembly Model a *tile system* is a 5-tuple $\mathbf{T} = (\Sigma, T, S, g, \tau)$, where $T$ is a finite set of tiles with binding domains from $\Sigma$ and contains the tile *empty*, $S$ is a configuration on $T$ called *seed configuration*, $g$ is a strength function, and $\tau$ is a threshold parameter called *temperature*. We will be working with seed configurations consisting of a single tile; formally a configuration $C_t$, where $t \in T$, satisfying $C_t(0,0) = t$, and $C_t(x,y) = empty$ for all $(x,y) \in \mathbb{Z} \times \mathbb{Z} \backslash \{(0,0)\}$.

Self-assembly is now defined as a relation between configurations on $T$. Let $C$ and $D$ be two configurations of $T$, such that $C = D$ except at position $(x,y)$, where $C(x,y) = empty$, and $D(x,y) = t$, for some $t \in T \backslash \{empty\}$. Then we write $C \to_{\mathbf{T}} D$, if

$$\sum_{d \in \mathcal{D}} g_d^C(t,x,y) \geq \tau.$$

This means that a tile can be added to a configuration at position $(x,y)$, if and only if the sum of the interaction strengths of $t$ with its neighbours reaches or exceeds $\tau$. For simple strength functions and $\tau = 1$, which are considered in this paper, $C \to_{\mathbf{T}} D$ if and only if there exists $d \in \mathcal{D}$ such that $g_d^C(t,x,y) = 1$. The relation $\to_{\mathbf{T}}^+$ is the transitive closure of $\to_{\mathbf{T}}$.

We are interested in a subclass of configurations that arise from the self-assembly process. A tile system $\mathbf{T}$ and the relation $\to_{\mathbf{T}}^+$ define the partially ordered set of configurations called *assemblies* of $\mathbf{T}$: $Asmb(\mathbf{T}) = \{A : S \to_{\mathbf{T}}^+ A\}$, and the set of *terminal assemblies* of $\mathbf{T}$: $Term(\mathbf{T}) = \{A \in Asmb(\mathbf{T}) : \nexists B \text{ such that } A \to_{\mathbf{T}}^+ B\}$. A tile system *uniquely produces* $A$ if $Term(\mathbf{T}) = \{A\}$. An

assembly $A$ is said to have *no binding domain mismatches* if for any two neighbouring positions $(x, y)$ and $d(x, y)$, $d \in \mathcal{D}$, we have $A(x, y)_d = A(d(x, y))_{d-1}$. The uniqueness assumption on the tile system has an interesting consequence:

**Observation 1.** *Let $\mathbf{T}$ be a tile system hat uniquely produces a terminal assembly $U$. Let $A$ and $B$ be two assemblies of $\mathbf{T}$, then $A \cup B$ is an assembly of $\mathbf{T}$ as well.*

## 3   Results

### 3.1   Lower bound based on graph radius

An assembly $A$ is an $N \times N$ *square* if there exists a position $(x_0, y_0)$ such that $(x, y) \in A$ if and only if $x_0 \leq x < x_0 + N$ and $y_0 \leq y < y_0 + N$.

The following lemma is based on the simple argument from [7] showing that a unique assembly of the full square requires all tiles to be distinct. In our setting, where we do not require the terminal assembly to be full, the argument is a little bit more involved. We need the following definitions.

**Definition 1** (Backbone graph)**.** *Given a tile system $\mathbf{T} = (\Sigma, T, C_t, s_\Sigma, 1)$, and a configuration $C$ of $T$, the backbone graph of $C$, $G(C) = (V, E)$ is the subgraph of the square lattice whose vertex set is $V(C)$, and two vertices $(x, y)$, $(x', y')$ form an edge if and only if there is a bond between $(x, y)$ and $(x', y')$ in $C$. Note that if the configuration $C$ is also an assembly of $\mathbf{T}$, then its backbone graph is connected. If $G$ is a subgraph of a backbone graph we say a tile $t \in T$ appears on $G$, if there is a vertex $(x, y) \in V(G)$ such that $C(x, y) = t$.*

A *translation $\phi$ mapping $p_0 = (x_0, y_0)$ to $p_1 = (x_1, y_1)$* is a mapping from $\mathbb{Z} \times \mathbb{Z}$ to $\mathbb{Z} \times \mathbb{Z}$ such that a point $p = (x, y)$ is mapped to $\phi(p) = (x + (x_1 - x_0), y + (y_1 - y_0))$. The composition of $n \geq 0$ copies of a translation $\phi$ will be denoted as $\phi^{(n)}$. Similarly, $\phi^{(-n)}$ denotes the composition of $n \geq 0$ copies of the inverse translation $\phi^{-1}$.

**Lemma 1.** *Let $\mathbf{T} = (\Sigma, T, C_t, s_\Sigma, 1)$ be a tile system that uniquely produces a finite assembly $A$. Suppose there are no binding domain mismatches in $A$. Let $P$ be a path in the backbone graph of $A$ with one endpoint at $(0, 0)$ (the position of the seed tile). Then all tiles of $A$ on $P$ are distinct.*

*Proof.* We will think of $P$ as being directed, with the position of the seed tile being the first vertex of $P$. Towards a contradiction, suppose there is a tile $r$ that is repeated on $P$. Say $r$ occurs at positions $p_0 = (x_0, y_0)$ and $p_1 = (x_1, y_1)$, where $p_0$ precedes $p_1$ on $P$. Let $\phi$ be the translation mapping $p_0$ to $p_1$. Let $Q$ be the sub-path of $P$ from $p_0$ to $p_1$. Recall that $A[Q]$ is the sub-configuration of $A$ induced by the vertices of $Q$. For every integer $n \geq 0$, let $Q_n = \phi^{(n)}(Q)$ be the translated version of the path $Q$, i.e., the vertex $(x, y)$ of $Q$ corresponds to the vertex $\phi^{(n)}(x, y)$ of $Q_n$. Similarly, for every $n \geq 0$, let $A_n$ be the translated version of the configuration $A$ such that for every $(x, y) \in \mathbb{Z} \times \mathbb{Z}$, $A_n(\phi^{(n)}(x, y)) = A(x, y)$, denoted as $A_n = \phi^{(n)}(A)$. Let $R$ be the sub-path of $P$ from $(0, 0)$ to $p_0$. Note that $A[R] \in Asmb(\mathbf{T})$, since $(0, 0) \in V(R)$.

Let $B_n = \bigcup_{i=0}^{n} A_i[Q_i]$. In what follows, we show that $A[R] \cup B_n \in Asmb(\mathbf{T})$, for all $n \geq 0$. Since the terminal assembly $A$ is finite and unique, this will inevitably lead to a contradiction. To prove this claim, it is enough to show that for all $n \geq 0$, $A[R] \cup B_n$ is a configuration, its vertex set contains $(0, 0)$, and its backbone graph is connected. We show this by induction on $n$.

For $n = 0$, the claim is trivially true, since $B_0 = A_0[Q_0] = A[Q]$. As induction hypothesis, suppose it is also true for some $n - 1 \geq 0$. We will show that it is true for $n$ as well. Towards
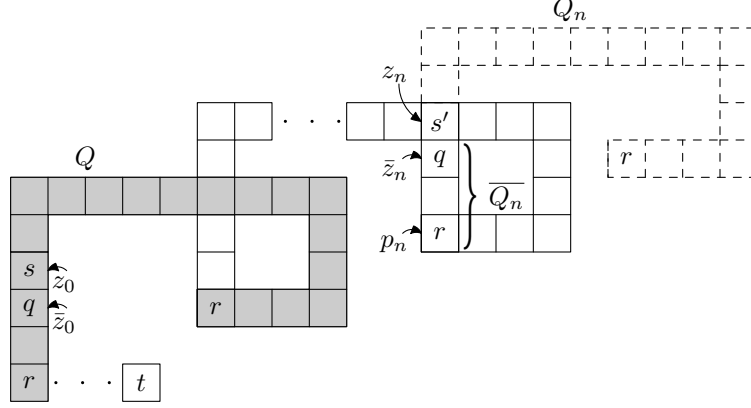
4

Figure 1: Forming $(A[R] \cup B_{n-1}) \cup A_n[Q_n]$ with a conflict at vertex $z_n$.

a contradiction suppose $A[R] \cup B_n = (A[R] \cup B_{n-1}) \cup A_n[Q_n]$ is not a configuration. Let $z_n$ be the first vertex on the path $Q_n$ starting from $p_n = \phi^{(n)}(p_0)$ such that $(A[R] \cup B_{n-1})(z_n) = s' \neq s = A_n[Q_n](z_n)$, and $s, s' \neq empty$ (see Figure 1). Note that $z_n$ is not the first vertex $(p_n)$ on $Q_n$, since $(A[R] \cup B_{n-1})(p_n) = A_n[Q_n](p_n) = r$. Let $\bar{z}_n$ be the neighbour of $z_n$ on $Q_n$ closer to $p_n$, say $\bar{z}_n = d(z_n)$ for some direction $d \in \mathcal{D}$ ($d = S$ in the example of Figure 1). Denote by $\bar{Q}_n$ the sub-path of $Q_n$ from $p_n$ to $\bar{z}_n$. By our choice of $\bar{z}_n$, $(A[R] \cup B_{n-1}) \cup A_n[\bar{Q}_n]$ is a configuration of $T$. Moreover, since the vertex sets of $A[R] \cup B_{n-1}$ and $A_n[\bar{Q}_n]$ intersect in $p_n$, the backbone graph of $(A[R] \cup B_{n-1}) \cup A_n[\bar{Q}_n]$ is connected and hence $(A[R] \cup B_{n-1}) \cup A_n[\bar{Q}_n] \in Asmb(\mathbf{T})$.

Let $q$ be the tile at $\bar{z}_n$ in $A_n[Q_n]$. The same tile $q$ also appears in $A[Q]$ at vertex $\bar{z}_0 = \phi^{(-n)}(\bar{z}_n)$. In $A[Q]$ there is a bond between the adjacent positions $\bar{z}_0$ and $z_0 = \phi^{(-n)}(z_n)$. This implies that the binding domain of $q$ on the side $d^{-1}$ cannot be $null$. Since $(A[R] \cup B_{n-1}) \cup A_n[\bar{Q}_n] \in Asmb(\mathbf{T})$ is a sub-configuration of the unique terminal assembly $A$ and $A$ has no binding domain mismatches, there must be a bond between positions $\bar{z}_n$ and $z_n$ in $(A[R] \cup B_{n-1}) \cup A_n[\bar{Q}_n]$, i.e., $s'_d = q_{d^{-1}} = s_d$. Let $\bar{Q} = \phi^{(-n)}(\bar{Q}_n)$ be the translated version of $\bar{Q}_n$ and let $B$ be a configuration such that $B_{z_0} = s'$ and $B(x,y) = empty$, for any $(x,y) \neq z_0$. Since vertex $\bar{z}_0$ precedes $z_0$ on the path $P$, and the both tiles, $s$ and $s'$, bind to $q$ on the side $d^{-1}$, $A[R] \cup A[Q] \cup B \in Asmb(\mathbf{T})$. This contradicts the uniqueness of assembly $A$. Thus $A[R] \cup B_n$ is indeed a configuration of $T$. Moreover, since the vertex sets of $A[R] \cup B_{n-1}$ and $A_n[Q_n]$ intersect in $p_n$, $A[R] \cup B_n \in Asmb(\mathbf{T})$. Hence, $A[R] \cup B_n \in Asmb(\mathbf{T})$ for all $n \geq 0$, contradicting that $\mathbf{T}$ uniquely assembles a finite assembly $A$. $\square$

Note that Lemma 1 is false if we drop either the condition that the seed tile is an endpoint of the path or the condition that there are no binding domain mismatches in $A$. Figure 2 shows that we can (uniquely) assemble a path of length four with only four non-empty tiles if the seed tile $S$ is not an endpoint of the path, and that we can (uniquely) assemble a path of length seven with only seven non-empty tiles if we allow binding domain mismatches (note the mismatch between the seed tile $S$ and tile 5 to the south of $S$).

We say that an assembly $A$ is *full* if for any two neighbouring positions $(x,y)$ and $d(x,y)$, $d \in \mathcal{D}$, such that $A(x,y), A(d(x,y)) \neq empty$, we have $A(x,y)_d = A(d(x,y))_{d^{-1}} \neq null$. As the first corollary we have that for 2-connected uniquely produced full assemblies, all tiles have to be distinct.

**Corollary 1.** *Let* $\mathbf{T} = (\Sigma, T, C_t, s_\Sigma, 1)$ *be a tile system that uniquely produces a finite full assembly $A$. If $A$ is 2-connected, then all tiles of $A$ are distinct, i.e., $\mathbf{T}$ contains at least $|V(A)|$ tile types.*
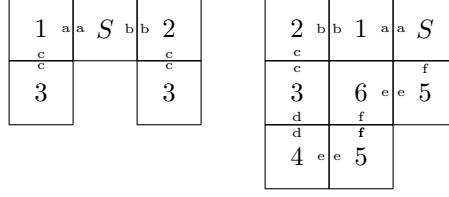
5

Figure 2: Fewer tiles suffice if we do not require the seed tile to be an endpoint of the path, or do not require that there are no binding domain mismatches.

*Proof.* The result follows by Lemma 1 and the fact that for any two points $p_0, p_1 \in V(A)$, there is a path starting in $(0,0)$ and containing both $p_0$ and $p_1$. □

As the second corollary we have an observation that on *any* path in the backbone graph of a uniquely produced assembly, each tile type can repeat at most twice. We will use this in the proof of our second lower bound in Subsection 3.2.

**Corollary 2.** *Let $\mathbf{T} = (\Sigma, T, C_t, s_\Sigma, 1)$ be a tile system that uniquely produces a finite assembly $A$. Suppose there are no binding domain mismatches in $A$. Let $P$ be a path in the backbone graph of $A$. If a tile is repeated on $P$, it is repeated exactly twice on $P$. Moreover, if a tile repeats at vertices $p_0$ and $p_1$ on $P$, then the first intersection of any path in the backbone graph starting at vertex $(0,0)$ with $P$ lies strictly between $p_0$ and $p_1$.*

*Proof.* Suppose a tile $r$ is repeated at least three times on $P$. Let $Q$ be any path starting at $(0,0)$ and intersecting $P$. Without loss of generality, we may assume $Q$ ends at the vertex of this intersection, say $w$. By our assumption, $Q$ has no other intersection with $P$. If $(0,0) \in V(P)$, then $Q = w = (0,0)$. Now tile $r$ appears twice on the path consisting of $Q$ together with one of the two sub-paths of $P$ from $w$ to an endpoint of $P$. This contradicts Lemma 1, and thus every repeated tile on $P$ occurs exactly twice on $P$. The second statement follows directly from Lemma 1. □

**Corollary 3.** *Let $\mathbf{T} = (\Sigma, T, C_t, s_\Sigma, 1)$ be a tile system that uniquely produces a finite assembly $A$. Suppose there are no binding domain mismatches in $A$. Let $C$ be a cycle in the backbone graph of $A$. Then all tiles of $A$ on $C$ are distinct.*

*Proof.* Suppose a tile $r$ is repeated twice on $C$, say $r$ occurs at position $p_0$ and $p_1$. As the backbone graph of $A$ is connected, there is a path starting at $(0,0)$ and containing both $p_0$ and $p_1$. This contradicts Lemma 1. □

The *eccentricity* of a vertex $v$ of a connected graph $G$ is defined as $\max_{u \in V(G)} d(u,v)$, where $d(u,v)$ is the length of a shortest path from $u$ to $v$. The *radius of $G$* is the minimum eccentricity of the vertices of $G$. The *radius of an assembly $A$*, denoted by $\mathrm{rad}(A)$, is the radius of the sub-graph of the lattice induced by V(A).

**Theorem 1.** *Let $\mathbf{T} = (\Sigma, T, C_t, s_\Sigma, 1)$ be a tile system that uniquely produces a finite assembly $A$. Suppose there are no binding domain mismatches in $A$. Then $T$ contains at least $\mathrm{rad}(A) + 1$ non-empty tiles.*

*Proof.* By the definition of $\mathrm{rad}(A)$, there is a path in the backbone graph of $A$ with one endpoint at $(0,0)$ of length at least $\mathrm{rad}(A)$. By Lemma 1 all tiles on this path are distinct. Thus $T$ contains at least $\mathrm{rad}(A) + 1$ non-empty tiles. □

Theorem 1 implies that for an $N \times N$ square we need at least $N$ or $N+1$ tile types (depending on the parity of $N$), while the upper bound in [7] is $2N-1$. In the following section we will provide another lower bound based on the Manhattan diameter of the assembly which will match the upper bound when applied for the square.

## 3.2 Lower bound based on Manhattan diameter

Let $P$ be a path in the backbone graph of an assembly $A$ and assign an orientation to $P$. For any vertex $v$ on $P$, we denote the side on which $P$ enters $v$ by $in^P(v)$ and the side on which $P$ leaves $v$ by $out^P(v)$. Let $r$ be a tile that is repeated (exactly twice) on $P$, say $r$ occurs at positions $p_0$ and $p_1$, where $p_0$ precedes $p_1$ on $P$. We say $r$ has a *good repetition* if $out^P(p_0) = in^P(p_1)$. Otherwise, $r$ has a *bad repetition*. Note that the definition of good repetition is independent of the orientation assigned to $P$. Also note that the repetition of tile 5 in Figure 2 is a bad repetition.

Given two vertices $(x, y)$ and $(x', y')$ in $\mathbb{Z} \times \mathbb{Z}$, their *Manhattan distance* is $|x - x'| + |y - y'|$. The *Manhattan diameter* of an assembly $A$, denoted by $\mathrm{Mdiam}(A)$, is the maximum Manhattan distance between any two vertices in $V(A)$.

We prove the result in two steps. In the first step we prove the bound for a path which does not contain any bad repetitions (Lemma 2). In the second step we will show that between any two points in $V(A)$ there is a path without bad repetitions (Lemma 3).

**Lemma 2.** *Let $\mathbf{T} = (\Sigma, T, C_t, s_\Sigma, 1)$ be a tile system that uniquely produces a finite assembly $A$. Let $P$ be a path in the backbone graph of $A$ such that the Manhattan distance between the endpoints of $P$ is $\mathrm{Mdiam}(A)$. If $P$ does not contain any bad repetitions, then $T$ contains at least $\mathrm{Mdiam}(A) + 1$ non-empty tiles.*

*Proof.* Assign an orientation to $P$ and let $a$ be the first and $b$ be the last vertex of $P$. Consider a direction $d \in \mathcal{D}$. If the Manhattan distance between $d(a)$ and $b$ is less than the Manhattan distance between $a$ and $b$, we say that $d$ is a *forward direction*. If neither $d$ nor $d^{-1}$ are forward, we pick any of them to be forward. We call an edge $uv$ of $P$ (with tail $u$ and head $v$) a *forward edge* if $d \in \mathcal{D}$ is forward and $v = d(u)$. Otherwise, the edge $uv$ is a *back edge*. Let $n_1$ be the number of tiles occurring exactly once on $P$, and let $n_2$ be the number of tiles that are repeated on $P$. Since $P$ has no bad repetitions, it is easy to see that each tile can repeat at most twice on $P$. Indeed, assume that a tile $r$ repeats 3 times at positions $p_0$, $p_1$ and $p_2$ (in this order). Then since $p_0$ and $p_1$ is a good repetition, $out^P(p_0) = in^P(p_1)$. Similarly, since $p_0$ and $p_2$ is a good repetition, $out^P(p_0) = in^P(p_2)$. Hence, we have $in^P(p_1) = in^P(p_2)$, which implies that $out^P(p_1) \neq in^P(p_2)$. This is a contradiction with the fact that $p_1$ and $p_2$ is a good repetition.

For any repeated tile $r$ on $P$ occurring at positions $p_0$ and $p_1$ such that $p_0$ precedes $p_1$, $out^P(p_0) = in^P(p_1)$, and hence there is at least one back edge incident with either $p_0$ or $p_1$. Summing over all repeated tiles, we count at least $n_2$ back edges. On the other hand, any back edge is incident with exactly two different vertices on $P$, i.e., it is counted at most twice. Thus, the number of back edges is at least $n_2/2$. By the definition of forward and back edges, $P$ contains $\mathrm{Mdiam}(A)$ more forward edges than back edges. Since the tiles can repeat at most twice, $n_1 + 2n_2$ is the number of vertices on $P$. Since $P$ contains at least $n_2/2$ back edges, this implies that there are at least $n_2/2 + \mathrm{Mdiam}(A)$ forward edges, and at least $\mathrm{Mdiam}(A) + n_2$ edges in total. Hence, the number of edges is $|P| - 1 = n_1 + 2n_2 - 1 \geq \mathrm{Mdiam}(A) + n_2$. This implies that $n_1 + n_2 \geq \mathrm{Mdiam}(A) + 1$. Since the set of tiles occurring exactly once on $P$ and the set of tiles being repeated on $P$ are disjoint subsets of $T$, $T$ contains at least $n_1 + n_2 \geq \mathrm{Mdiam}(A) + 1$ non-empty tile types. $\square$

Note that the above lemma holds without the assumption of no binding domain mismatches.

**Lemma 3.** *Let* $\mathbf{T} = (\Sigma, T, C_t, s_\Sigma, 1)$ *be a tile system that uniquely produces a finite assembly $A$. Suppose there are no binding domain mismatches in $A$. Let $u$ and $v$ be two vertices in $V(A)$. Then there is a path $P$ in the backbone graph of $A$ connecting $u$ and $v$ without any bad repetitions.*

*Proof.* Among all paths between $u$ and $v$ in the backbone graph of $A$, let $P$ be a path with the fewest number of bad repetitions. If $P$ does not contain any bad repetitions, we are done. So suppose that a tile $r$ has a bad repetition on $P$. Say $r$ occurs at positions $p_0 = (x_0, y_0)$ and $p_1 = (x_1, y_1)$, where $p_0$ precedes $p_1$ on $P$. Let $\phi$ be the translation mapping $p_0$ to $p_1$. We will show that there exists another path with the same endpoints as $P$ and with fewer bad repetitions than $P$, contradicting our assumption.

To this end, consider a shortest path $F$ from $(0,0)$ to $P$ in the backbone graph, and let $w$ be the intersection point of these paths. By Corollary 2, $w$ must be between $p_0$ and $p_1$ on $P$. Let $Q$ be the sub-path of $P$ from $p_0$ to $p_1$. Let $Q_1 = \phi(Q)$ be the translated version of $Q$, and let $A_1 = \phi(A)$ be the translated version of the configuration $A$. Recall that for every $p \in \mathbb{Z} \times \mathbb{Z}$, $A_1(\phi(p)) = A(p)$. Let $R$ be the subtree consisting of $F$ and $Q$. Note that $A[R]$ is an assembly of $\mathbf{T}$, since $(0,0)$ is on the tree $R$.

Suppose $A[R] \cup A_1[Q_1]$ is a configuration of $T$, then it is also an assembly of $\mathbf{T}$, since $R$ and $Q_1$ share a vertex. If the backbone graph of assembly $A[R] \cup A_1[Q_1]$ is a tree, then it contains a path with one endpoint at $(0,0)$ on which tile $r$ appears twice, namely at $p_1$ and $p_2 = \phi(p_1)$. This contradicts Lemma 1. Thus, either the backbone graph of the assembly $A[R] \cup A_1[Q_1]$ is not a tree, or $A[R] \cup A_1[Q_1]$ is not a configuration of $T$. In either case, there exists a vertex $z_1 \neq p_1$ such that $A[R](z_1) \neq empty$ and $A_1[Q_1](z_1) \neq empty$. Furthermore, suppose $z_1$ is the first such vertex on $Q_1$ starting from $p_1$. Let $\bar{z}_1$ be the neighbour of $z_1$ on $Q_1$ closer to $p_1$, say $\bar{z}_1 = d(z_1)$ for some direction $d \in \mathcal{D}$. Denote by $\bar{Q}_1$ the sub-path of $Q_1$ from $p_1$ to $\bar{z}_1$. By our choice of $\bar{z}_1$, $A[R] \cup A_1[\bar{Q}_1]$ is a configuration of $T$. Moreover, since $R$ and $\bar{Q}_1$ share the vertex $p_1$, $A[R] \cup A_1[\bar{Q}_1]$ is an assembly of $\mathbf{T}$.
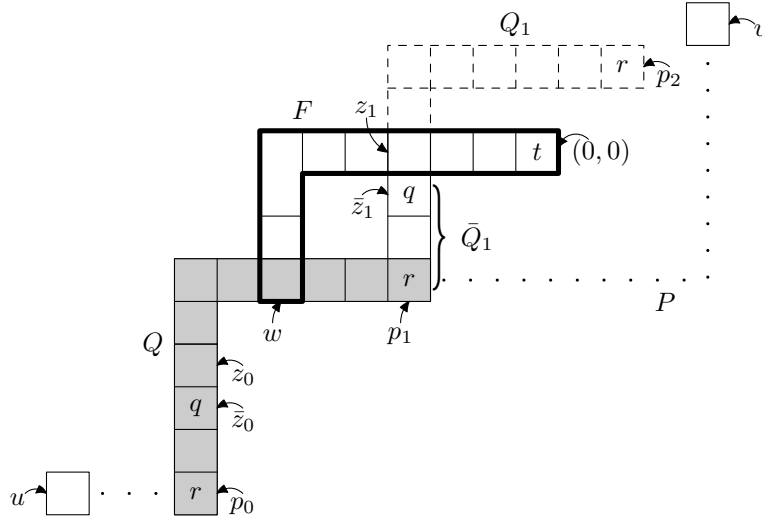


Figure 3: Intersection vertex $z_1$ is on the path $F$. Then there exists a path with one endpoint at $(0,0)$ on which tile $r$ appears twice.

Let $q = A_1[Q_1](\bar{z}_1)$. We also have $q = A[Q](\bar{z}_0)$, where $\bar{z}_0 = \phi^{-1}(\bar{z}_1)$. In $A[Q]$ there is a bond between the adjacent positions $\bar{z}_0$ and $z_0 = \phi^{-1}(z_1)$. This implies that the binding domain of $q$ on the side $d^{-1}$ cannot be *null*. Since $A[R] \cup A_1[\bar{Q}_1] \in Asmb(\mathbf{T})$ is a sub-configuration of the unique

terminal assembly $A$ and $A$ has no binding domain mismatches, there must be a bond between positions $\bar{z}_1$ and $z_1$ in $A[R] \cup A_1[\bar{Q}_1]$.

Now if $z_1$ is on the path $F$ and $z_1 \neq w$, then the path consisting of the sub-path of $F$ from $(0,0)$ to $z_1$, together with the sub-path of $Q_1$ from $z_1$ to $p_1$ and $Q$, is a path with one endpoint at $(0,0)$ on which tile $r$ appears twice (see Figure 3). This contradicts Lemma 1. Thus, $z_1$ must be on the path $Q$.

There are two cases to consider, depending on the relative position of $z_1$ and $z_0$ on $Q$. Note that $z_0 \neq z_1$. For if, $z_1 = z_0$, $\phi$ is the identity translation, and thus $p_1 = p_0$. This contradicts our assumption that tile $r$ is repeated twice.
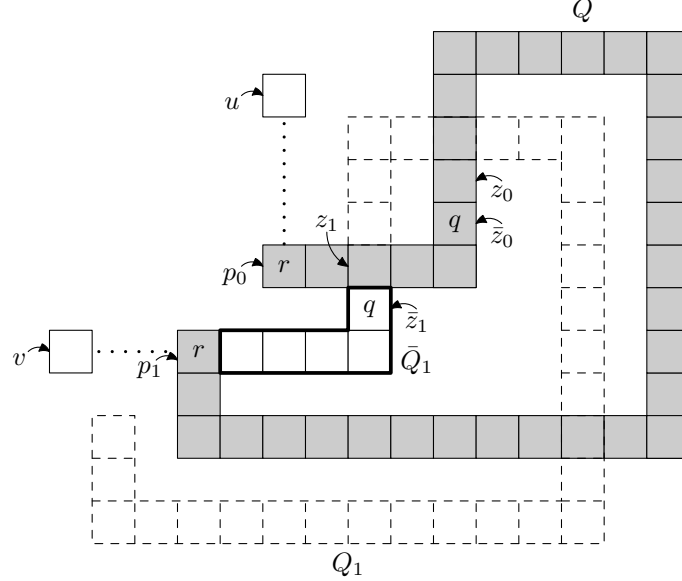


Figure 4: Intersection vertex $z_1$ precedes $z_0$ on $Q$. Then there is a path starting at $(0,0)$ on which tile $q$ appears twice.

First, suppose $p_0$ is closer to $z_1$ than $z_0$ on $Q$ (see Figure 4). Then there is a cycle (from $z_1$ to $p_1$ along $Q$ and back to $z_1$ along $\bar{Q}_1$) containing the tile $q$ twice, namely, at positions $\bar{z}_0$ and $\bar{z}_1$. This contradicts Corollary 3.

Next, suppose $p_0$ is closer to $z_0$ than $z_1$ on $Q$. We will exhibit a different path having the same endpoints as $P$ and fewer bad repetitions than $P$. Let the vertex $s_1$ be the first vertex on $\bar{Q}_1$ (starting from $\bar{z}_1$ towards $p_1$) that is also on $P$ and let $\bar{s}_1$ be the neighbour of $s_1$ on $Q_1$ closer to $z_1$.

First assume $s_1$ lies on the subpath of $P$ from $u$ to $p_0$ (see Figure 5). Let $P'$ be the path obtained from $P$ by replacing the subpath of $P$ from $s_1$ to $z_1$ with the subpath of $Q_1$ from $s_1$ to $z_1$. Assign to $P'$ the same orientation as $P$. Since $A[F] \cup A[P]$ and $A[R] \cup A_1[\bar{Q}_1]$ are assemblies of $\mathbf{T}$, by Observation 1, $B = A[F] \cup A[P] \cup A_1[\bar{Q}_1]$ is an assembly of $\mathbf{T}$ as well. It follows that $P'$ is a subgraph of the backbone graph of $B$, and hence also of $A$.

Removing vertices from $P$ cannot increase the number of bad repetitions. The only new tiles added to $P'$ are the tiles of the subpath of $\bar{Q}_1$ from $\bar{s}_1$ to $\bar{z}_1$ (it is possible that this subpath is empty if $s_1 = \bar{z}_1$). These tiles appear on the part of $P$ which is not included in $P'$. Moreover, each of these tiles is entered and left on $P'$ via the same directions as the corresponding tiles on $P$, i.e., for every $x$ on the subgraph of $P'$ from $\bar{s}_1$ to $\bar{z}_1$, $out^{P'}(x) = out^P(\phi^{-1}(x))$ and $in^{P'}(x) = in^P(\phi^{-1}(x))$. Hence, any bad repetition introduced by these new tiles only replaces a removed bad repetition involving corresponding tiles on the removed part of $P$ (the subpath of $P$ from $\phi^{-1}(\bar{s}_1)$ to $\phi^{-1}(\bar{z}_1)$).
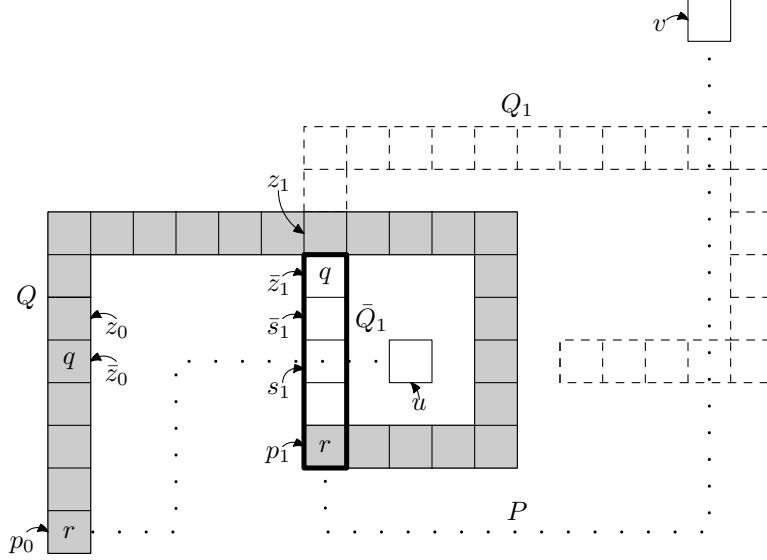
Figure 5: Intersection vertex $z_1$ succeeds $z_0$ on $Q$ and $s_1$ lies on the subpath of $P$ from $u$ to $p_0$. Then replacing the sub-path of $P$ from $s_1$ to $z_1$ (shaded) with the sub-path of $Q_1$ from $s_1$ to $z_1$ (unshaded) yields a path with fewer bad repetitions than $P$.

Since $z_0$ precedes $z_1$ on $Q$, $p_0 \neq z_1$. Furthermore, $s_1$ precedes $p_0$ on $P$. Thus, $p_0$ is in between $s_1$ and $z_1$ on $P$, and hence not on $P'$. Therefore, tile $r$ does not have a bad repetition on $P'$. Thus, $P'$ has fewer bad repetitions than $P$.

Next, suppose $s_1$ lies on the subpath of $P$ from $p_1$ to $v$ (it is possible that $s_1 = p_1$)(see Figure 6). Let $P'$ be the path obtained from $P$ by replacing the subpath of $P$ from $z_1$ to $s_1$ with the subpath of $Q_1$ from $z_1$ to $s_1$. The only new tiles added to $P'$ are the tiles of the subpath of $\bar{Q}_1$ from $\bar{s}_1$ to $\bar{z}_1$. These tiles already appear on $P$, in particular on its sub-path $Q$. Note that these tiles are repeated exactly twice on $P'$, otherwise we would have a contradiction with Corollary 2. Each of the tiles on the subpath of $P'$ from $\bar{z}_1$ to $\bar{s}_1$ is entered and left via opposite directions as the corresponding tiles of the subpath of $P$ from $\bar{s}_0 = \phi^{-1}(\bar{s}_1)$ to $\bar{z}_0$, i.e., for every $x$ on the subpath of $P'$ from $\bar{z}_1$ to $\bar{s}_1$, $out^{P'}(x) = in^P(\phi^{-1}(x))$ and $in^{P'}(x) = out^P(\phi^{-1}(x))$. Hence, these repeating tiles create good repetitions, and no new bad repetitions were introduced on $P'$. Moreover, if $s_1 = p_1$, then tile $r$ now has a good repetition on $P'$, since $s_1$ is on the sub-path of $Q_1$. If $s_1 \neq p_1$, then $r$ is not repeated on $P'$. Thus, $P'$ has fewer bad repetitions than $P$, contradicting the minimality of $P$. This proves the lemma.

$\square$

Note that Lemma 3 is again false if we drop the condition that there are no binding domain mismatches in $A$. This can be seen from the second example in Figure 2, where there is only one path between tile $S$ and the rightmost tile 5, and this path contains a bad repetition.

Combining Lemmas 2 and 3 together, we have the following claim.

**Theorem 2.** *Let* $\mathbf{T} = (\Sigma, T, C_t, s_\Sigma, 1)$ *be a tile system that uniquely produces a finite assembly* $A$. *Suppose there are no binding domain mismatches in* $A$. *Then* $T$ *contains at least* $\mathrm{Mdiam}(A) + 1$ *non-empty tiles.*

*Proof.* Consider a pair of vertices $u$ and $v$ in the backbone graph with the Manhattan distance equal to $\mathrm{Mdiam}(A)$. By Lemma 3, there is a path $P$ connecting $u$ and $v$ such that $P$ has no bad
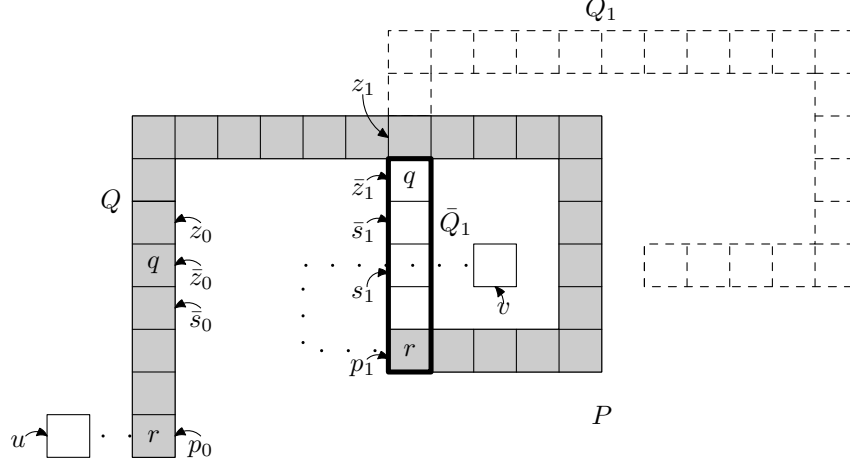
10

Figure 6: Intersection vertex $z_1$ succeeds $z_0$ on $Q$ and $s_1$ lies on the subpath of $P$ from $p_1$ to $v$. Then replacing the sub-path of $P$ from $z_1$ to $s_1$ (shaded) with the sub-path of $Q_1$ from $z_1$ to $s_1$ (unshaded) yields a path with fewer bad repetitions than $P$.

repetitions. Then by Lemma 2, $T$ contains at least $M diam(A) + 1$ non-empty tiles. □

**Corollary 4.** *Let* $\mathbf{T} = (\Sigma, T, C_t, s_\Sigma, 1)$ *be a tile system that uniquely produces an* $N \times N$ *square* $A$. *Suppose there are no binding domain mismatches in* $A$. *Then* $T$ *contains at least* $2N - 1$ *non-empty tiles.*
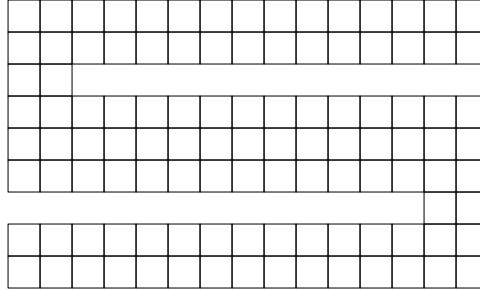


Figure 7: An example of a shape for which Theorem 1 gives a better lower bound (24) than Theorem 2 (23).

The second bound (Theorem 2) gives a better lower bound compared to the first bound (Theorem 1) for an $N \times N$ square. However, for many uniquely produced assemblies Theorem 1 gives a better bound. For instance, for an S-shaped assembly, see Figure 7, Theorem 1 gives a lower bound of 24 while Theorem 2 gives a lower bound of 23. Obviously, this difference can be made arbitrarily large by increasing the width of the shape.

## Conclusions

In this paper we gave two lower bounds on the number of tiles needed to uniquely assemble a shape using the energy model which does not penalize any mismatches. More specifically, if a tile system uniquely produces a terminal assembly $A$ at temperature 1 and $A$ does not have any

binding domain mismatches, then our first result states that at least $\mathrm{rad}(A) + 1$ non-empty tiles are needed. Our second result states that under the same assumptions, at least $\mathrm{Mdiam}(A) + 1$ non-empty tiles are required. A consequence of the second result is that $2N - 1$ non-empty tiles are needed to uniquely assemble an $N \times N$ square at temperature 1, provided the square assembles without any binding domain mismatches. This is a partial answer to an open problem from [7], where the conjecture on the lower bound was stated without the natural assumption of no binding domain mismatches. Therefore, a natural question is whether this assumption can be dropped, maybe by using a different energy function which penalizes mismatches.

An another interesting question is whether there is a shape for which would require a smaller number of tiles one assembled uniquely while allowing binding domain mismatches than when assembled uniquely without allowing them. When binding domain mismatches are allowed, one can use the following strategy to assemble any shape:

- assemble a surface of the shape using distinct tiles;

- for each connected component of internal tiles, use one tile with all sides containing the same glue $x$ to assemble the internal tiles; this assembly would be triggered by adding glue $x$ to internal side of the last surface tile which enclosed the connected component of the internal tiles.
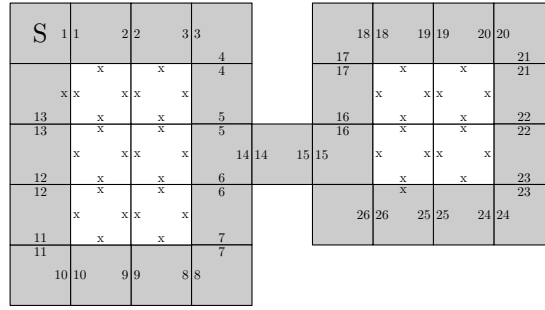


Figure 8: Example of the construction using domain mismatches. The border tiles are depicted gray. All inner tiles use the same tile type.

Figure 8 illustrates this construction one a particular shape. The inner tiles form two connected component and they are filled only after the whole component is surrounded with the surface tiles. Obviously, this construction requires the number of surface tiles of the shape plus one tile types. Although this construction is not always optimal, for instance, for a $N \times N$ square, it requires $4N - 3$ tiles, it cannot be use to construct the same shape without binding domain mismatches. Hence, there might be a shape which can be assembled using this construction with the smaller number of tile types than the number needed to uniquely assemble the same shape without using binding domain mismatches.

# References

[1] H. Abelson, D. Allen, D. Coore, C. Hanson, G. Homsy, T. F. Knight, R. Nagpal, E. Rauch, G. J. Sussman, and R. Weiss. Amorphous computing. *Communications of the ACM*, 43:74–82, 2000.

[2] M. Gomez-Lopez, J. Preece, and J. Stoddart. The art and science of self-assembling molecular machines. *Nanotechnology*, 7:183 – 192, 1996.

[3] T.H. LaBean, H. Yan, J. Kopatsch, F. Liu, E. Winfree, J. H. Reif, and N.C. Seeman. Construction, analysis, ligation, and self-assembly of DNA triple crossover complexes. *Journal of the American Chemical Society*, 122:1848–1860, 2000.

[4] C. Mao, T. H. LaBean, J.H. Reif, and N.C. Seeman. Logical computation using algorithmic self-assembly of DNA triple-crossover molecules. *Nature*, 407:493–496, 2000.

[5] J. Maňuch, L. Stacho, and C. Stoll. Two lower bounds for self-assemblies at temperature 1. In *Proc. of the 3rd international Conference on Bioinformatics and Biomedical Engineering (iCBBE, 2009)*. IEEE, 2009.

[6] J.H. Reif. Local parallel biomolecular computation. In *Proc. DNA-Based Computers*, pages 217–254, 1997.

[7] P. W. K. Rothemund and E. Winfree. The program-size complexity of self-assembled squares. In *Proceedings of STOC*, pages 459–468, 2000.

[8] P.W.K Rothemund, N. Papadakis, and E. Winfree. Algorithmic self-assembly of DNA sierpinski triangles. *PLoS Biology*, 2:2041–2053, 2004.

[9] N.C. Seeman. DNA nanotechnology: novel DNA constructions. *Annual Review of Biophysics and Biomolecular Structure*, 27:225–248, 1998.

[10] D. Soloveichik and E. Winfree. Complexity of self-assembled shapes. *SIAM Journal on Computing*, 36:1544 – 1569, 2007.

[11] E. Winfree, F. Liu, L. A. Wenzler, and N. C. Seeman. Design and self-assembly of two dimensional DNA crystals. *Nature*, 394:539–544, 1998.

[12] E. Winfree, X. Yang, and N. Seeman. Universal computation via self-assembly of DNA: Some theory and experiments. In *Proceedings of the Second Annual Meeting on DNA Based Computers*, pages 191–214, 1996.