Program Size and Temperature in Self-Assembly[*]

Ho-Lin Chen[†]   David Doty[†]   Shinnosuke Seki[‡]

Abstract

Winfree's abstract Tile Assembly Model (aTAM) is a model of molecular self-assembly of DNA complexes known as tiles, which float freely in solution and attach one at a time to a growing "seed" assembly based on specific binding sites on their four sides. We show that there is a polynomial-time algorithm that, given an $n \times n$ square, finds the minimal tile system (i.e., the system with the smallest number of distinct tile types) that uniquely self-assembles the square, answering an open question of Adleman, Cheng, Goel, Huang, Kempe, Moisset de Espanés, and Rothemund (*Combinatorial Optimization Problems in Self-Assembly*, STOC 2002). Our investigation leading to this algorithm reveals other positive and negative results about the relationship between the size of a tile system and its "temperature" (the binding strength threshold required for a tile to attach).

1 Introduction

Tile self-assembly is an algorithmically rich model of "programmable crystal growth". It is possible to design monomers (square-like "tiles") with specific binding sites so that, even subject to the chaotic nature of molecules floating randomly in a well-mixed chemical soup, they are guaranteed to bind so as to deterministically form a single target shape. This is despite the number of different types of tiles possibly being much smaller than the size of the shape and therefore having only "local information" to guide their attachment. The ability to control nanoscale structures and machines to atomic-level precision will rely crucially on sophisticated self-assembling systems that automatically control their own behavior where no top-down externally controlled device could fit.

A practical implementation of self-assembling molecular tiles was proved experimentally feasible in 1982 by Seeman [20] using DNA complexes formed from artificially synthesized strands. Experimental advances have delivered increasingly reliable assembly of algorithmic DNA tiles with error rates of 10% per tile in 2004 [19], 1.4% in 2007 [12], 0.13% in 2009 [5], and 0.05% in 2010 [11]. Erik Winfree [26] introduced the abstract Tile Assembly Model (aTAM) – based on a constructive version of Wang tiling [24, 25] – as a simplified mathematical model of self-assembling DNA tiles. Winfree demonstrated the computational universality of the aTAM by showing how to simulate an arbitrary cellular automaton with a tile assembly system. Building on these connections to

computability, Rothemund and Winfree [18] investigated a self-assembly resource bound known as *tile complexity*, the minimum number of tile types needed to assemble a shape. They showed that for most n, the problem of assembling an $n \times n$ square has tile complexity $\Omega(\frac{\log n}{\log \log n})$, and Adleman, Cheng, Goel, and Huang [2] exhibited a construction showing that this lower bound is asymptotically tight. Under natural generalizations of the model [1, 4, 6–10, 13, 14, 16, 22, 23], tile complexity can be reduced for the assembly of squares and more general shapes.

The results of this paper are motivated by the following problem posed in 2002. Adleman, Cheng, Goel, Huang, Kempe, Moisset de Espanés, and Rothemund [3] showed that there is a polynomial time algorithm for finding a minimum size tile system (i.e., system with the smallest number of distinct tile types) to uniquely self-assemble a given $n \times n$ square,[1] subject to the constraint that the tile system's "temperature" (binding strength threshold required for a tile to attach) is 2. They asked whether their algorithm could be modified to remove the temperature 2 constraint. Our main theorem answers this question affirmatively. Their algorithm works by brute-force search over the set of all temperature-2 tile systems with at most $O(\frac{\log n}{\log \log n})$ tile types, using the fact proven by Adleman, Cheng, Goel, and Huang [2] that such an upper bound on tile complexity suffices to assemble any $n \times n$ square. A simple counting argument shows that for any constant τ, the number of tile systems with glue strengths and temperature at most τ and $O(\frac{\log n}{\log \log n})$ tile types is bounded by a polynomial in n. One conceivable approach to extending the algorithm to arbitrary temperature is to prove that for any tile system with K tile types, the strengths and temperature can be re-assigned so that they are upper-bounded by a constant or slow-growing function of K, without affecting the behavior of the tile system.[2] However, we show that this approach cannot work, by demonstrating that for each K, there is a tile system with K tile types whose behavior cannot be preserved using any temperature less than $2^{K/4}$. The proof crucially uses *3-cooperative* binding, meaning attachment events that require three different glues of a tile to match the assembly. On the other hand, we show that any 2-cooperative tile system (which captures a wide and useful class of systems) with K tile types is behaviorally equivalent to a system with temperature at most $2K + 2$.

Of course, the choice of integer glue strengths is an artifact of the model. Nonetheless, our investigation does reflect fundamental questions about how finely divided molecular binding energies must be in a real molecular self-assembly system. The requirement of integer strengths is simply one way of "quantizing" the minimum distinction we are willing to make between energies and then re-scaling so that this quantity is normalized to 1.[3] Our 3-cooperative lower bound therefore shows that in general, certain self-assembling systems that have very large gaps between some of their binding energies nonetheless require other binding energies to be extremely close (exponentially small in terms of the larger gaps) and yet still unequal. This can be interpreted as an infeasibility result if one defines "exponentially fine control" of binding energies as "infeasible" to execute in any real laboratory, since no implementation of the specified tile behavior can use courser energies.

As a converse to the exponential temperature lower bound stated above, we show that there is a polynomial-time algorithm that, given any tile system $\mathcal{T}$ with K tile types specified by its desired binding behavior, finds a temperature and glue strengths at most $2^{O(K)}$ that implement

---

[1]The square is encoded by a list of its points, so the algorithm's running time is polynomial in n.

[2]We define "behavior" more formally in Section 3. Briefly, we consider a tile system's behavior unaltered by a reassignment of strengths and temperature if, for each tile type t, the reassignment has not altered the collection of subsets of sides of t that have sufficient strength to bind.

[3]Indeed, our proof does not require that strengths be integer, merely that the distance between the smallest energy strong enough to bind and the largest energy too weak to bind be at least 1.

this behavior or reports that no such strengths exist. This algorithm is used to show our main result, that there is a polynomial-time algorithm that, given an $n \times n$ square, determines the smallest tile assembly system (at any temperature) that uniquely self-assembles the square, answering the open question of [3].

2   Abstract Tile Assembly Model

This section gives a brief informal sketch of the abstract Tile Assembly Model (aTAM). See Section A for a formal definition of the aTAM.

A *tile type* is a unit square with four sides, each having a *glue label* (often represented as a finite string). We assume a finite set T of tile types, but an infinite number of copies of each tile type, each copy referred to as a *tile*. An *assembly* (a.k.a., *supertile*) is a positioning of tiles on (part of) the integer lattice $\mathbb{Z}^2$; i.e., a partial function $\mathbb{Z}^2 \dashrightarrow T$. For a set of tile types T, let $\Lambda(T)$ denote the set of all glue labels of tile types in T. We may think of a tile type as a function $t : \{\mathsf{N}, \mathsf{S}, \mathsf{E}, \mathsf{W}\} \to \Lambda(T)$ indicating, for each direction $d \in \{\mathsf{N}, \mathsf{S}, \mathsf{E}, \mathsf{W}\}$ ("north, south, east, west"), the glue label $t(d)$ appearing on side d. A *strength function* is a function $g : \Lambda(T) \to \mathbb{N}$ indicating, for each glue label ℓ, the strength $g(\ell)$ with which it binds. Two adjacent tiles in an assembly *interact* if the glue labels on their abutting sides are equal and have positive strength according to g. Each assembly induces a *binding graph*, a grid graph whose vertices are tiles, with an edge between two tiles if they interact. The assembly is τ-*stable* if every cut of its binding graph has strength at least τ, where the weight of an edge is the strength of the glue it represents. That is, the assembly is stable if at least energy τ is required to separate the assembly into two parts.

A *tile assembly system* (TAS) is a quadruple $\mathcal{T} = (T, \sigma, g, \tau)$, where T is a finite set of tile types, $\sigma : \mathbb{Z}^2 \dashrightarrow T$ is a finite, τ-stable *seed assembly*, $g : \Lambda(T) \to \mathbb{N}$ is a *strength function*, and τ is a *temperature*. In this paper, we assume that all seed assemblies σ consist of a single tile type (i.e., $|\mathrm{dom}\, \sigma| = 1$). Given a TAS $\mathcal{T} = (T, \sigma, g, \tau)$, an assembly α is *producible* if either $\alpha = \sigma$ or if β is a producible assembly and α can be obtained from β by placing a single tile type t on empty space (a position $p \in \mathbb{Z}^2$ such that $\beta(p)$ is undefined), such that the resulting assembly α is τ-stable. In this case write $\beta \to_1 \alpha$ (α is producible from β by the stable attachment of one tile), and write $\beta \to \alpha$ if $\beta \to_1^* \alpha$ (α is producible from β by the stable attachment of zero or more tiles). An assembly is *terminal* if no tile can be stably attached to it. Let $\mathcal{A}[\mathcal{T}]$ be the set of producible assemblies of $\mathcal{T}$, and let $\mathcal{A}_\square[\mathcal{T}] \subseteq \mathcal{A}[\mathcal{T}]$ be the set of producible, terminal assemblies of $\mathcal{T}$. Given a connected shape $S \subseteq \mathbb{Z}^2$, a TAS $\mathcal{T}$ *uniquely self-assembles* S if $\mathcal{A}_\square[\mathcal{T}] = \{\widehat{\alpha}\}$ and $\mathrm{dom}\, \widehat{\alpha} = S$.

3   Finding Strengths to Implement a Tile System

In this section we show that there is a polynomial-time algorithm that, given a desired behavior of a tile set with unspecified glue strengths or temperature, can find strengths and temperature to implement that behavior that are at most exponential in the number of tile types, or report that no such strengths and temperature exist. This algorithm is the primary technical tool used in the proof of our main result, Theorem 4.2.

First, we formalize what we mean by the "behavior" of a tile system. Let T be a set of tile types, and let $t \in T$. Given a strength function $g : \Lambda(T) \to \mathbb{N}$ and a temperature $\tau \in \mathbb{Z}^+$, define the *cooperation set of t with respect to g and τ* to be the collection $\mathcal{D}_{g,\tau}(t) = \big\{\, D \subseteq \{\mathsf{N}, \mathsf{S}, \mathsf{E}, \mathsf{W}\} \ \big| \ \sum_{d \in D} g(t(d)) \geq \tau \,\big\}$, i.e., the collection of subsets of sides of t whose glues have sufficient strength to bind cooperatively.

Let $\sigma : \mathbb{Z}^2 \dashrightarrow T$ be a seed assembly, let $\tau_1, \tau_2 \in \mathbb{Z}^+$ be temperatures, and let $g_1, g_2 : \Lambda(T) \to \mathbb{N}$ be strength functions. We say that the TAS's $\mathcal{T}_1 = (T, \sigma, g_1, \tau_1)$ and $\mathcal{T}_2 = (T, \sigma, g_2, \tau_2)$, differing only on their strength function and temperature, are *locally equivalent* if, for each tile type $t \in T$, $\mathcal{D}_{g_1,\tau_1}(t) = \mathcal{D}_{g_2,\tau_2}(t)$.[4] The behavior of an individual tile type during assembly is completely determined by its cooperation set, in the sense that if $\mathcal{T}_1$ and $\mathcal{T}_2$ are locally equivalent, then $\mathcal{A}[\mathcal{T}_1] = \mathcal{A}[\mathcal{T}_2]$ and $\mathcal{A}_\square[\mathcal{T}_1] = \mathcal{A}_\square[\mathcal{T}_2]$.[5]

Even without any strength function or temperature, by specifying a cooperation set for each tile type, one can describe a "behavior" of a TAS in the sense that its dynamic evolution can be simulated knowing only the cooperation set of each tile type. We call a system thus specified a strength-free TAS. More formally, a *strength-free TAS* is a triple $(T, \sigma, \mathcal{D})$, where T is a finite set of tile types, $\sigma : \mathbb{Z}^2 \dashrightarrow T$ is the size-1 seed assembly, and $\mathcal{D} : T \to \mathcal{P}(\mathcal{P}(\{\mathsf{N}, \mathsf{S}, \mathsf{E}, \mathsf{W}\}))$ is a function from a tile type $t \in T$ to a cooperation set $\mathcal{D}(t)$. For a standard TAS $\mathcal{T} = (T, \sigma, g, \tau)$ and a strength-free TAS $\mathcal{T}_{\mathrm{sf}} = (T, \sigma, \mathcal{D})$ sharing the same tile set and seed assembly, we say that $\mathcal{T}$ and $\mathcal{T}_{\mathrm{sf}}$ are *locally equivalent* if $\mathcal{D}_{g,\tau}(t) = \mathcal{D}(t)$ for each tile type $t \in T$.

Note that every TAS has a unique cooperation set for each tile type, and hence, has a locally equivalent strength-free TAS. Say that a strength-free TAS is *implementable* if there exists a TAS locally equivalent to it. Not every strength-free TAS is implementable. This is because cooperation sets could be contradictory; for instance, two tile types t_1 and t_2 could satisfy $t_1(\mathsf{N}) = t_2(\mathsf{N})$ and $\{\mathsf{N}\} \in \mathcal{D}(t_1)$ but $\{\mathsf{N}\} \notin \mathcal{D}(t_2)$.

Theorem 3.1. *There is a polynomial time algorithm that, given a strength-free TAS $\mathcal{T}_{\mathrm{sf}} = (T, \sigma, \mathcal{D})$, determines whether there exists a locally equivalent TAS $\mathcal{T} = (T, \sigma, g, \tau)$ and outputs such a $\mathcal{T}$ if it exists. Furthermore, it is guaranteed that $\tau \leq 2^{O(|T|)}$.*

Proof. Intuitively, we count the number of different total behaviors a tile can have, based on equating its behavior with its cooperation set. We then cast the problem of finding strengths to implement a given strength-free tile system as finding solutions to a certain system of linear inequalities, which is solved by Gaussian elimination. We then argue that the vertices of the polytope defined by this system have rational numbers with numerator and denominator that are at most exponential in the number of inequalities, which is itself linear in the number of tile types. This implies that by multiplying the rational numbers at one of these vertices by their least common multiple, which preserves the inequalities, we obtain an integer solution with values at most exponential in the number of tile types.

Formally, let $\mathcal{T}_{\mathrm{sf}}$ be a strength-free TAS with a tile set T of k tile types with $u \leq k+1$ different glues.[6] We would like to decide whether $\mathcal{T}_{\mathrm{sf}}$ is in fact realizable by a TAS. To have the tightest upper bound on the temperature, ideally we should solve the problem of finding the minimum temperature TAS that is locally equivalent to $\mathcal{T}_{\mathrm{sf}}$. This optimization problem can be cast as an integer linear program on a temperature variable τ and a set of glue-strength variables $s_1, s_2, \ldots, s_u$ as in the following example:

---

[4]Note that the definition of equivalence is independent of the seed assembly; we include it only to be able to talk about the equivalence of TAS's rather than the more cumbersome "equivalence of triples of the form (T, g, τ)."

[5]The converse does not hold, however. For instance, some tile types may have a subset of sides whose glues never appear together at a binding site during assembly, so it would be irrelevant to the definition of $\mathcal{A}[\mathcal{T}]$ and $\mathcal{A}_\square[\mathcal{T}]$ whether or not that combination of glues have enough strength to bind.

[6]Each tile type has 4 sides so it might seem that there could be $4k$ total glues if there are k tile types. However, in a nontrivial system (one that has no "effectively null" glues that appear on only one side of any tile type), for each side of a tile type, the choice of glue for that side is limited to those glues on the opposite side of the k tile types, or alternately we could choose the null strength-0 glue.

4

$$
\begin{array}{lrcl}
\text{Minimize} & \tau \\
\text{subject to} & \tau, s_1, s_2, \ldots, s_u & \in & \mathbb{N} \\
& s_1 + s_3 - \tau & \geq & 0 \\
& s_1 + s_4 + s_6 - \tau & \geq & 0 \\
& & \cdots & \\
& s_1 + s_2 + s_4 - \tau & \leq & -1 \\
& s_2 + 2s_3 - \tau & \leq & -1 \\
& & \cdots &
\end{array}
$$

The "≥ 0" inequalities correspond to the union of all cooperation sets of all tile types, and the "≤ -1" inequalities correspond to the union of all complements of the cooperation sets; i.e., each set $\mathcal{P}(\{\mathsf{N}, \mathsf{S}, \mathsf{E}, \mathsf{W}\}) \setminus \mathcal{D}(t)$, where $\mathcal{D}(t)$ is the cooperation set of t. Since we require each strength to be an integer, "≤ -1" is equivalent to "< 0". Since each tile type t has $|\mathcal{D}(t)|$ "≥ 0" inequalities (one for each subset of sides in its cooperation set) and $16 - |\mathcal{D}(t)|$ "≤ -1" inequalities, there are $16k$ inequalities in the integer linear program.

However, solving this optimization problem is not necessary to prove the theorem. Our goal will not be to find the smallest temperature TAS that satisfies the constraints above (which remains an open problem to do in polynomial time), but simply to find any feasible integer solution with temperature and strengths at most $2^{O(k)}$. Call the above system of constraints (including the integer constraint) S_1. Consider the real-valued system of linear inequalities S_2 defined as the above inequalities with the integer constraint $\tau, s_1, s_2, \ldots, s_u \in \mathbb{N}$ relaxed to simply $\tau, s_1, s_2, \ldots, s_u \geq 0$. Then we have the implication "S_1 has a solution" $\implies$ "S_2 has a solution". Conversely, any rational-valued solution to S_2 can be converted to an integer-valued solution to S_1 by multiplying each value by the least common multiple L of the denominators of the rational numbers.[7] Furthermore, since the input coefficients are integers, if the feasible polytope of S_2 is non-empty, then all of its vertices are rational. Therefore S_2 has a solution if and only if it has an integer solution. Since any integer solution to S_2 is a solution to S_1, we have the full bidirectional implication "S_1 has a solution" $\iff$ "S_2 has an integer solution". We can pick any n linearly independent inequalities of S_2, interpret them as equalities, and use Gaussian elimination (with exact rational arithmetic) to obtain some vertex of the feasible polytope described by the inequalities, and convert these to integer solutions to S_1 through multiplication as described above. If we cannot find n linearly independent inequalities (testable by computing the rank of the matrix defining the inequalities) then there is no TAS implementing the behavior of $\mathcal{T}_{\mathrm{sf}}$. It remains to show that in case there is a solution, the integers we obtain by this method obey the stated upper bound $2^{O(k)}$.

Each coefficient has absolute value at most 2 (since we may assume N/S glues are disjoint from E/W glues), and each equation has at most 5 nonzero left side terms since each tile type has only 4 sides (together with the -1 coefficient for τ). Applying Lemma 3.2 (stated and proven after the current proof) with $n = u + 1$, $c_1 = 2$, and $c_2 = 5$, we obtain that each vertex is a rational vector $\vec{x} = (\frac{p_1}{q_1}, \ldots, \frac{p_{u+1}}{q_{u+1}})$ such that, for each $1 \leq i \leq u + 1$, $|p_i|, |q_i| \leq 2^{u+1} 6^{(u+1)/2} = 2^{u+1+((u+1)/2) \cdot \log 6} < 2^{3u+3}$. Since we enforce nonnegativity, $p_i = |p_i|$ and $q_i = |q_i|$. We multiply the rational vector $\vec{x}$ by $L = \mathrm{LCM}(q_1, \ldots, q_{u+1})$ to obtain the integer vector $\vec{x}'$. By Lemma 3.2, $L \leq 2^{3u+3}$. Then each integer solution value x_i' obeys $x_i' = L \cdot x_i \leq (2^{3u+3})^2 \leq 2^{2(3(k+1)+3)} = 2^{O(k)}$. $\qquad\square$

---

[7]This actually enforces the stronger condition that each "≤ -1" inequality is actually "$\leq -L$". This is possible because we have no *upper* bound on the variables, which would prevent multiplication from necessarily preserving the inequalities.

Lemma 3.2. *Let $c_1, c_2 \in \mathbb{Z}^+$ be constants, and let $\vec{b}$ be an $n \times 1$ integer column vector and $A = (a_{ij})$ be a nonsingular $n \times n$ integer matrix such that for each i, j, $|a_{ij}| \leq c_1$ and $|b_j| \leq c_1$, and each row of A contains at most c_2 nonzero entries. Then the solution to the linear system $A\vec{x} = \vec{b}$ is a rational vector $\vec{x} \in \mathbb{Q}^n$ such that, if each component $x_i = \frac{p_i}{q_i}$ is written in lowest terms with $p_i, q_i \in \mathbb{Z}$, then $|p_i| \leq c_1^n (c_2 + 1)^{n/2}$ and $|q_i| \leq c_1^n c_2^{n/2}$. Furthermore, the least common multiple of all the q_i's is at most $c_1^n c_2^{n/2}$.*

Proof. Recall Hadamard's inequality $|\det A| \leq \prod_{i=1}^{n} \|v_i\|_2$, where v_i is the ith row of A.[8] Since v_i has at most c_2 nonzero entries that are each at most absolute value c_1, Hadamard's inequality tells us that $|\det A| \leq \prod_{i=1}^{n} \sqrt{c_2 \cdot c_1^2} = c_1^n c_2^{n/2}$. Similarly, letting A_i be A with column vector $\vec{b}$ replacing A's i^{th} column, A_i has at most $c_2 + 1$ nonzero entries per row, so a similar argument gives $|\det A_i| \leq c_1^n (c_2 + 1)^{n/2}$. The ith solution is $x_i = \frac{\det A_i}{\det A}$ by Cramer's rule. Since A and A_i are integer-valued, so are $\det A$ and $\det A_i$, whence the upper bounds on $|\det A|$ and $|\det A_i|$ also apply to $|q_i|$ and $|p_i|$, respectively, since they are x_i's lowest terms representation. Since the lowest terms representation of each q_i necessarily divides $\det A$, this implies that their least common multiple is also at most $c_1^n c_2^{n/2}$. $\qquad\square$

4 Finding the Minimum Tile Assembly System Assembling a Square at any Temperature

This section shows the main result of this paper, that there is a polynomial-time algorithm that, given an $n \times n$ square S_n, computes the smallest TAS that uniquely self-assembles S_n. Adleman, Cheng, Goel, Huang, Kempe, Moisset de Espanés, and Rothemund [3] showed that this problem is polynomial-time solvable when the temperature is restricted to be 2, and asked whether there is an algorithm that works when the temperature is unrestricted, which we answer affirmatively.

The next proposition is useful in the proof of Theorem 4.2.

Proposition 4.1. *For each $k \in \mathbb{Z}^+$, there are at most $168^k k^{4k+2}$ implementable strength-free TAS's with at most k tile types.*

Proof. If $\mathcal{T} = (T, \sigma, \mathcal{D})$ is an implementable strength-free TAS, then for all $t \in T$, the cooperation set $\mathcal{D}(t)$ of t is a collection of subsets of $\{\mathsf{N}, \mathsf{S}, \mathsf{E}, \mathsf{W}\}$ that is closed under the superset operation, i.e., if $D \subseteq D' \subseteq \{\mathsf{N}, \mathsf{S}, \mathsf{E}, \mathsf{W}\}$ and $D \in \mathcal{D}(t)$, then $D' \in \mathcal{D}(t)$. This closure property is due to the fact that having strictly more sides available to bind cannot inhibit binding that would otherwise occur, as long as strengths are assumed to be nonnegative. Each cooperation set $\mathcal{D}$ is defined by a unique *antichain*, which is a subcollection $\mathcal{D}' = \{D'_1, \ldots, D'_m\} \subseteq \mathcal{D}$ such that, for all $1 \leq i, j \leq m$, $D'_i \not\subseteq D'_j$, whose closure under the superset operation is equal to $\mathcal{D}$. The antichain consists of the minimal elements of $\mathcal{D}$ under the partial order $\subseteq$. The number of antichains of subsets of $\{\mathsf{N}, \mathsf{S}, \mathsf{E}, \mathsf{W}\}$ is given by the fourth Dedekind number $M(4) = 168$ [21]. Thus, a tile type has at most 168 different possible cooperation sets.

For each side of a tile type, there are at most k glue labels to choose, so there are at most k^4 ways to assign these labels to each side. Therefore, encoding each tile type as a list of 4 glue labels and cooperation set, and encoding a TAS as a list of tile types, there are at most $(168k^4)^k = 168^k k^{4k}$

---

[8]Hadamard's inequality is typically stated for v_i a column of A, but the determinant of a matrix and its transpose are equal so the bound holds when taking the product over rows as well.

different strength-free tile sets with k tile types. Since there are k choices for the seed tile, there are at most $168^k k^{4k+1}$ different strength-free TAS's with k tile types. Thus there are at most $\sum_{i=1}^{k} 168^i i^{4i+1} \leq 168^k k^{4k+2}$ such TAS's with *at most* k tile types. $\qquad \square$

The following theorem is the main result of this paper.

Theorem 4.2. *There is a polynomial-time algorithm that, given an $n \times n$ square S_n, outputs a minimal TAS $\mathcal{T}$ that uniquely self-assembles S_n.*

Proof. In [3], the authors study the variant of the minimum tile set problem restricted to squares where the temperature is fixed at $\tau = 2$. They use the following argument to show the problem is solvable in polynomial time. For all $n \in \mathbb{N}$, let $S_n = \{0, 1, \ldots, n-1\}^2$ denote the $n \times n$ square. Adleman, Cheng, Goel, and Huang [2] showed that for all $n \in \mathbb{N}$, there is a TAS with at most $O(\frac{\log n}{\log \log n})$ tile types that uniquely self-assembles S_n. The proof of [3] first shows by a simple counting argument that there are at most a polynomial in n number of temperature-2 TAS's with at most $O(\frac{\log n}{\log \log n})$ tile types, using the fact that all strengths may without loss of generality be assumed to be 0, 1, or 2. They then make use of a polynomial-time algorithm UNIQUE-SHAPE devised in the same paper [3] that, given any shape S and any TAS $\mathcal{T}$, determines whether $\mathcal{T}$ uniquely self-assembles S. Finding the minimum TAS for S_n then amounts to iterating over every "small" ($O(\frac{\log n}{\log \log n})$ tile types) TAS $\mathcal{T}$ and using the algorithm UNIQUE-SHAPE to check which of these systems assemble S_n. The upper bound of [2] guarantees that UNIQUE-SHAPE will report a positive answer for at least one of these systems.

Let $c \in \mathbb{Z}^+$ be the constant, shown to exist in [2], such that, for all $n \in \mathbb{N}$, some tile system with at most $c \log n / \log \log n$ tile types uniquely self-assembles S_n. Let $k = c \log n / \log \log n$. Enumerate each strength-free TAS with at most k tile types and test whether it has a locally equivalent standard TAS $\mathcal{T}$, using the algorithm described in Theorem 3.1. For each such $\mathcal{T}$, run the algorithm UNIQUE-SHAPE of [3] to test whether $\mathcal{T}$ is directed and strictly self-assembles S_n. Output the TAS $\mathcal{T}$ with the smallest number of tile types that passes this test.

Since UNIQUE-SHAPE and the algorithm of Theorem 3.1 run in polynomial time, it remains to show that there are a polynomial (in n) number of strength-free TAS's to search with at most $k = c \log n / \log \log n$ tile types. Proposition 4.1 tells us that the number of strength-free TAS's with at most k tile types is at most $168^k k^{4k+2}$. We have that $168^k \leq 168^{c \log n / \log \log n} \leq 168^{c \log n} \leq (2^{\log n})^{c \log 168} \leq n^{8c}$, and

$$
\begin{aligned}
k^{4k+2} &\leq k^2 (c \log n / \log \log n)^{4c \log n / \log \log n} \\
&= k^2 (2^{\log(c \log n / \log \log n)})^{4c \log n / \log \log n} \\
&= k^2 (2^{\log c + \log \log n - \log \log \log n})^{4c \log n / \log \log n} \\
&= k^2 2^{4c \log n (\log c + \log \log n - \log \log \log n) / \log \log n} \\
&= k^2 (2^{\log n})^{4c(\log c + \log \log n - \log \log \log n) / \log \log n} \\
&= k^2 n^{4c \log c / \log \log n + 4c - 4c \log \log \log n / \log \log n} \\
&\leq n^{4c \log c + 4c + 2},
\end{aligned}
$$

whence the number of strength-free TAS's to search is at most $n^{4c \log c + 12c + 2} = \operatorname{poly}(n)$. $\qquad \square$

We note that while we have stated the theorem for the family of square shapes, our method, as well as that of [3], works for any family of shapes $S_1, S_2, \ldots$ where $|S_n| = \operatorname{poly}(n)$ and the tile

complexity of S_n is at most $O(\frac{\log n}{\log\log n})$. This includes, for instance, the family $\{T_1, T_2, \ldots\}$, where T_n is a width-n right triangle, and for each $q \in \mathbb{Q}^+$ the family $\{R_{q,1}, R_{q,2}, \ldots\}$, where $R_{q,n}$ is the $n \times \lfloor qn \rfloor$ rectangle.

5 Bounds on Temperature Relative to Number of Tile Types

This section reports two bounds relating the number of tile types in a TAS to its temperature. The first bound, Theorem 5.1, shows that there are TAS's that require temperature *exponential* in the number of tile types (in the sense of local equivalence as defined in Section 3), if any combination of sides may be used for binding. This result can be interpreted to mean that the algorithm of [2] to find the minimum temperature-2 TAS for assembling an $n \times n$ square, which searches over all possible assignments of strengths to the glues, cannot be extended in a straightforward manner to handle larger temperatures, which is why it is necessary for the algorithm of Theorem 4.2 to "shortcut" through the behaviors of tile types rather than enumerating strengths. The second bound, Theorem 5.2, on the other hand, shows that if we restrict attention to those (quite prevalent) classes of tile systems that use only one or two sides of tiles to bind, then *linear* temperature always suffices.

5.1 Tile Assembly Systems Requiring Temperature Exponential in Number of Tile Types

In this section, we prove that a temperature that is exponential in the number of tile types given by Theorem 3.1 is optimal, although there is a gap between the exponents ($2^{|T|/4}$ for Theorem 5.1 below versus $O(2^{6|T|})$ for Theorem 3.1).

Theorem 5.1. *For every $n \in \mathbb{Z}^+$, there is a TAS $\mathcal{T} = (T, \sigma, g, \tau)$ such that $|T| = 4n$ and for every TAS $\mathcal{T}' = (T, \sigma, g', \tau')$ that is locally equivalent to $\mathcal{T}$, $\tau' \geq 2^n$.*

Proof. The tile set T is shown in Figure 1.[9] In each stage, each of the top two light tile types represents a triple (a pair in stage 1) of glues whose sum is at least τ. Each of the bottom two dark tiles represents a triple of glues whose sum is less than τ. For the dark tile types to be nontrivial, we could imagine that the (unlabeled) north glue is strong enough to cooperate with some of the other glues. The actual strengths of the glues are left as variables, but the caption of Figure 1 gives one example of strengths that would satisfy the inequalities that the tile types represent.

We prove by induction on n that $A''_n \geq A_n + 2^n$. For the base case, in the first stage, the top light tile type and top dark tile type enforce that $A'_1 + B'_1 \geq \tau > A_1 + B'_1$, so $A_1 < A'_1$. Similarly, the bottom light tile type and the bottom dark tile type enforce that $A'_1 < A''_1$. Therefore $A''_1 \geq A_1 + 2$.

For the inductive case, assume that $A''_{n-1} \geq A_{n-1} + 2^{n-1}$. The top dark tile type enforces that $\tau > A''_{n-1} + A_n + B'_n$, and by the induction hypothesis, $A''_{n-1} + A_n + B'_n \geq A_{n-1} + 2^{n-1} + A_n + B'_n$. The top light tile type enforces that $A'_n + A_{n-1} + B'_n \geq \tau$, which combined with the previous two inequalities shows that $A'_n \geq 2^{n-1} + A_n$. A similar analysis with the bottom light tile type and bottom dark tile type shows that $A''_n \geq 2^{n-1} + A'_n$, whence $A''_n \geq 2^{n-1} + 2^{n-1} + A_n$, establishing the inductive case.

---

[9] We do not specify the seed assembly since we are concerned only with the local behavior of the tiles. To make the local equivalence nontrivial, we would need to add a small number of tile types to the TAS to ensure that each tile shown is actually attachable at some point during assembly. However, but this would not affect the asymptotic size of the tile set as $n \to \infty$, so the exponential lower bound on the temperature would still hold.
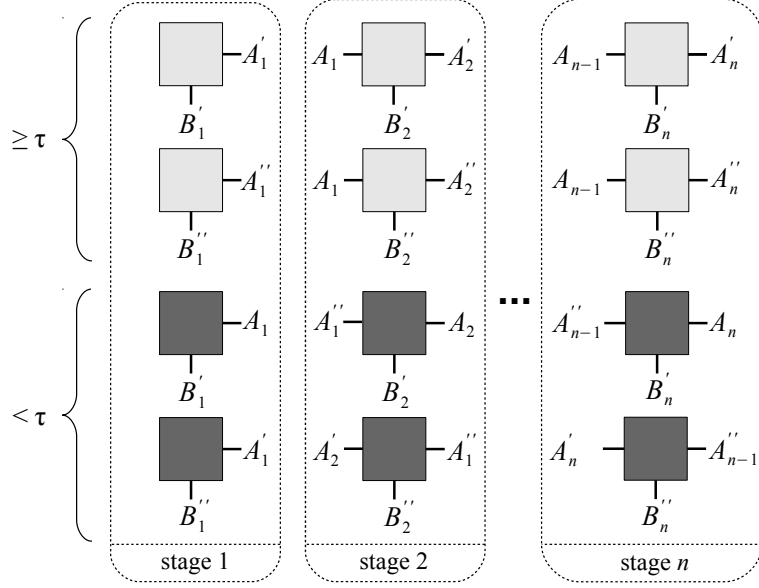
Figure 1: A set of tile types requiring temperature that is exponentially larger than the number of tile types. There are stages $1, 2, \ldots, n$, with stage i containing 4 tiles, and stage i ensuring that the gap between the largest and smallest strength in the stage is at least 2^i. In each stage, each of the top two light tiles represents a triple (a pair in stage 1) of glues whose sum is at least τ. Each of the bottom two dark tiles represents a triple of glues whose sum is less than τ. The inequalities are satisfiable, for instance, by setting $A_n = 3^{n-1}, A'_n = 2A_n, A''_n = 3A_n, B_n = \tau - A_n, B'_n = \tau - A'_n, B''_n = \tau - A''_n$.

Since it can be assumed without loss of generality that strengths are at most τ, this shows that the tile set consisting of n stages, having $|T| = 4n$ tile types, requires $\tau \geq 2^n$ to be realized. $\qquad\square$

5.2 Temperature Linear in the Number of Tile Types Suffices for 2-Cooperative Equivalence

Theorem 5.1 shows that temperature exponentially larger than the number of tile types is sometimes necessary for a TAS's behavior to be realized by integer strengths. However, the definition of local equivalence assumes that all possible combinations of sides of a tile type may be present in an assembly. Many TAS's are more constrained than this. There is a wide class of TAS's that we term *2-cooperative*, meaning that all binding events during all assembly sequences use only 1 or 2 sides that bind with positive strength. Nearly all theoretical TAS's found in the literature are 2-cooperative (indeed, temperature 2 systems by definition cannot *require* three sides to be present, although the model allows tile attachments with excess strength). In this section we show that the 3-cooperativity of Figure 1 is necessary, by showing that 2-cooperative systems can always be realized by strengths linear in the number of tile types.

Theorem 5.2. *Let $\mathcal{T} = (T, \sigma, g, \tau)$ be a TAS, and let $\mathcal{D}^{(2)}_{g,\tau}(t) \subseteq \mathcal{D}_{g,t}(t)$ be the cooperation set of t with respect to g and τ restricted to containing only subsets of $\{\mathsf{N}, \mathsf{S}, \mathsf{E}, \mathsf{W}\}$ of cardinality 1 or 2. Then there is a TAS $\mathcal{T}' = (T, \sigma, g', \tau')$ with $\tau' \leq 2|T| + 2$ such that, for each $t \in T$, $\mathcal{D}^{(2)}_{g,\tau}(t) = \mathcal{D}^{(2)}_{g',\tau'}(t)$.*

9

That is, $\mathcal{T}'$ is equivalent in behavior to $\mathcal{T}$, so long as all attachments involve only 1 or 2 sides.

Proof. Let $K = |T|$. Let $\Lambda(T)$ denote the set of all glue labels on tile types in T. Let $G = \{\, g(\sigma) \mid \sigma \in \Lambda(T) \,\} \setminus \{0, \tau\}$. That is, G is the set of all positive but insufficient glue strengths used in this system. $|G| \leq 2K$ since there are at most $|T|$ north-south glues and at most $|T|$ east-west glues.

We split G into two subsets $L = \{\, g \in G \mid 0 < g < \tau/2 \,\}$ and $H = \{\, g \in G \mid \tau/2 \leq g < \tau \,\}$. Let $L = \{\ell_1, \ell_2, \ldots, \ell_n\}$ such that $\ell_1 < \ell_2 < \ldots < \ell_n < \tau/2$, and let $H = \{h_1, h_2, \ldots, h_m\}$ with $\tau/2 \leq h_1 < h_2 \ldots h_m < \tau$. For descriptive purposes, define $\ell_0 = 0$ and $\ell_{n+1} = \tau/2$ (although these numbers may not be glue strengths).

We aim at designing an algorithm to find a glue function g' satisfying:

1. for any label $\sigma \in \{\sigma_1, \ldots, \sigma_{2K}\}, g(\sigma) \geq \tau \iff g'(\sigma) \geq 2n + 2$;

2. for any pair of labels $\{\sigma, \sigma'\} \subset \{\sigma_1, \ldots, \sigma_{2K}\}, g(\sigma) + g(\sigma') \geq \tau \iff g'(\sigma) + g'(\sigma') \geq 2n + 2$.

Then for $\tau' = 2n + 2$, the TAS $\mathcal{T}' = (T, \sigma, g', \tau')$ satisfies the 2-cooperative equivalence with $\mathcal{T}$ that we seek.

First, we define an equivalence relation $\equiv$ on H defined as: for $h, h' \in H$, $h \equiv h'$ if $(\forall\, 1 \leq i \leq n)(h + \ell_i \geq \tau \iff h' + \ell_i \geq \tau)$. This partitions H into subsets $H_{n+1}, \ldots, H_2, H_1$ such that $h \in H_j$ if and only if $\tau - \ell_j \leq h < \tau = \ell_{j-1}$. In other words, $h \in H_j$ if $h + \ell_j \geq \tau$ implies $i \geq j$.

A glue function $g' : \Lambda(T) \to \mathbb{N}$ is defined as follows: for a label $\sigma \in \Lambda(T)$,

$$g'(\sigma) = \begin{cases} 0, & \text{if } g(\sigma) = 0; \\ i, & \text{if } g(\sigma) = \ell_i \ (1 \leq i \leq n); \\ 2n + 2 - j, & \text{if } g(\sigma) \in H_j \ (1 \leq j \leq n + 1); \\ 2n + 2, & \text{if } \tau \leq g(\sigma). \end{cases}$$

It is trivial that this satisfies condition (1) above. Let $\sigma, \sigma' \in \Lambda(T)$ with $0 < g(\sigma) \leq g(\sigma') < \tau$. There are two cases.

1. Suppose that $\tau \leq g(\sigma) + g(\sigma')$. Note that $\tau/2 \leq g(\sigma')$. If $\tau/2 \leq g(\sigma)$, then by definition $n + 1 \leq g'(\sigma)$ and $n + 1 \leq g'(\sigma')$, whence their sum it at least $2n + 2$. If $g(\sigma) < \tau/2$, then let $g(\sigma) = \ell_i$ for some $1 \leq i \leq n$. Since $\tau \leq g(\sigma) + g(\sigma')$, this means that $\tau = \ell_i \leq g(\sigma')$, which in turn implies $g(\sigma') \in H_i \cup H_{i-1} \cup \ldots \cup H_1$. By definition, $g'(\sigma) = i$ and $g'(\sigma') \geq 2n + 2 - i$. Consequently, $2n + 2 \leq g'(\sigma) + g'(\sigma')$.

2. Suppose that $g(\sigma) + g(\sigma') < \tau$. In this case, $g(\sigma) < \tau/2$. If $g(\sigma') < \tau/2$, then by the definition of g', both $g(\sigma)$ and $g(\sigma')$ are at most n so that their sum cannot reach $2n + 2$. Otherwise, let $g(\sigma) = \ell_i$. An argument similar to the one above gives $g(\sigma') \in H_{n+1} \cup \ldots \cup H_{i+1}$, whence $g'(\sigma') < 2n + 2 - i$. Thus, $g'(\sigma) + g'(\sigma') < 2n + 2$.

This verifies that g' satisfies condition (2). $\qquad\square$

6 Open Questions

Our polynomial-time algorithm for finding the minimal tile system to self-assemble a square made crucial use of our polynomial-time algorithm that, given a strength-free tile system $\mathcal{T}_{\mathsf{sf}}$, finds

strengths and a temperature to implement a locally equivalent TAS $\mathcal{T}$, or reports that none exists. An open question is whether there is a polynomial-time algorithm that, given a strength-free TAS $\mathcal{T}_{\mathsf{sf}}$, outputs a TAS *of minimal temperature* that is locally equivalent to $\mathcal{T}_{\mathsf{sf}}$, or reports that none exists. In the proof of Theorem 3.1, we expressed the problem as an integer linear program, but needed only to find an integer feasible solution that is not necessarily optimal. If we instead wanted to find an *optimal* solution to the program, it is not clear whether the problem restricted to such instances is NP-hard.

The next question is less formal. Our results relating to 3-cooperative and 2-cooperative systems (Theorems 5.1 and 5.2, respectively), show that there is a difference in self-assembly "power" between these two classes of systems when we consider two tile systems to "behave the same" if and only if they are locally equivalent, which is a quite strict notion of behavioral equivalence. For example, two tile systems could uniquely self-assemble the same shape even though they have different tile types (hence could not possibly be locally equivalent). It would be interesting to know to what extent 3-cooperative systems – or high temperature tile systems in general – are strictly more powerful than temperature 2 systems under more relaxed notions of equivalence. For example, it is not difficult to design a shape that can be uniquely self-assembled with slightly fewer tile types at temperature 3 than at temperature 2. How far apart can this gap be pushed? Are there shapes with temperature-2 tile complexity that is "significantly" greater than their absolute (i.e., unrestricted temperature) tile complexity?

References

[1] Zachary Abel, Nadia Benbernou, Mirela Damian, Erik D. Demaine, Martin Demaine, Robin Flatland, Scott Kominers, and Robert Schweller. Shape replication through self-assembly and RNase enzymes. In *SODA 2010: Proceedings of the Twenty-first Annual ACM-SIAM Symposium on Discrete Algorithms*, Austin, Texas, 2010. Society for Industrial and Applied Mathematics.

[2] Leonard M. Adleman, Qi Cheng, Ashish Goel, and Ming-Deh Huang. Running time and program size for self-assembled squares. In *STOC 2001: Proceedings of the thirty-third annual ACM Symposium on Theory of Computing*, pages 740–748, Hersonissos, Greece, 2001. ACM.

[3] Leonard M. Adleman, Qi Cheng, Ashish Goel, Ming-Deh A. Huang, David Kempe, Pablo Moisset de Espanés, and Paul W. K. Rothemund. Combinatorial optimization problems in self-assembly. In *STOC 2002: Proceedings of the Thirty-Fourth Annual ACM Symposium on Theory of Computing*, pages 23–32, 2002.

[4] Gagan Aggarwal, Qi Cheng, Michael H. Goldwasser, Ming-Yang Kao, Pablo Moisset de Espanés, and Robert T. Schweller. Complexities for generalized models of self-assembly. *SIAM Journal on Computing*, 34:1493–1515, 2005. Preliminary version appeared in SODA 2004.

[5] Robert D. Barish, Rebecca Schulman, Paul W. Rothemund, and Erik Winfree. An information-bearing seed for nucleating algorithmic self-assembly. *Proceedings of the National Academy of Sciences*, 106(15):6054–6059, March 2009.

[6] Florent Becker, Ivan Rapaport, and Eric Rémila. Self-assembling classes of shapes with a minimum number of tiles, and in optimal time. In *FSTTCS 2006: Foundations of Software Technology and Theoretical Computer Science*, pages 45–56, 2006.

[7] Harish Chandran, Nikhil Gopalkrishnan, and John H. Reif. The tile complexity of linear assemblies. In *ICALP 2009: 36th International Colloquium on Automata, Languages and Programming*, volume 5555, pages 235–253. Springer, 2009.

[8] Erik D. Demaine, Martin L. Demaine, Sándor P. Fekete, Mashhood Ishaque, Eynat Rafalin, Robert T. Schweller, and Diane L. Souvaine. Staged self-assembly: Nanomanufacture of arbitrary shapes with $O(1)$ glues. *Natural Computing*, 7(3):347–370, 2008. Preliminary version appeared in DNA 13.

[9] Erik D. Demaine, Matthew J. Patitz, Robert T. Schweller, and Scott M. Summers. Self-assembly of arbitrary shapes using RNase enzymes: Meeting the Kolmogorov bound with small scale factor. In *STACS 2011: Proceedings of the 28th International Symposium on Theoretical Aspects of Computer Science*, 2011.

[10] David Doty. Randomized self-assembly for exact shapes. *SIAM Journal on Computing*, 39(8):3521–3552, 2010. Preliminary version appeared in FOCS 2009.

[11] Constantine Evans. Personal communication.

[12] Kenichi Fujibayashi, Rizal Hariadi, Sung Ha Park, Erik Winfree, and Satoshi Murata. Toward reliable algorithmic self-assembly of DNA tiles: A fixed-width cellular automaton pattern. *Nano Letters*, 8(7):1791–1797, 2007.

[13] Ming-Yang Kao and Robert T. Schweller. Reducing tile complexity for self-assembly through temperature programming. In *SODA 2006: Proceedings of the 17th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 571–580, 2006.

[14] Ming-Yang Kao and Robert T. Schweller. Randomized self-assembly for approximate shapes. In Luca Aceto, Ivan Damgård, Leslie Ann Goldberg, Magnús M. Halldrsson, Anna Ingólfsdóttir, and Igor Walukiewicz, editors, *ICALP 2008: International Colloqium on Automata, Languages, and Programming*, volume 5125 of *Lecture Notes in Computer Science*, pages 370–384. Springer, 2008.

[15] James I. Lathrop, Jack H. Lutz, and Scott M. Summers. Strict self-assembly of discrete Sierpinski triangles. *Theoretical Computer Science*, 410:384–405, 2009. Preliminary version appeared in CiE 2007.

[16] Ján Maňuch, Ladislav Stacho, and Christine Stoll. Step-assembly with a constant number of tile types. In *ISAAC 2009: Proceedings of the 20th International Symposium on Algorithms and Computation*, pages 954–963, Berlin, Heidelberg, 2009. Springer-Verlag.

[17] Paul W. K. Rothemund. *Theory and Experiments in Algorithmic Self-Assembly*. PhD thesis, University of Southern California, December 2001.

[18] Paul W. K. Rothemund and Erik Winfree. The program-size complexity of self-assembled squares (extended abstract). In *STOC 2000: Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing*, pages 459–468, 2000.

[19] Paul W.K. Rothemund, Nick Papadakis, and Erik Winfree. Algorithmic self-assembly of DNA Sierpinski triangles. *PLoS Biology*, 2(12):2041–2053, 2004.

[20] Nadrian C. Seeman. Nucleic-acid junctions and lattices. *Journal of Theoretical Biology*, 99:237–247, 1982.

[21] Neil J. Sloane. The On-Line Encyclopedia of Integer Sequences. http://oeis.org/wiki/A000372. Dedekind numbers: number of monotone Boolean functions of n variables or number of antichains of subsets of an n-set.

[22] David Soloveichik and Erik Winfree. Complexity of self-assembled shapes. *SIAM Journal on Computing*, 36(6):1544–1569, 2007. Preliminary version appeared in DNA 10.

[23] Scott M. Summers. Reducing tile complexity for the self-assembly of scaled shapes through temperature programming. *Algorithmica*. to appear.

[24] Hao Wang. Proving theorems by pattern recognition – II. *The Bell System Technical Journal*, XL(1):1–41, 1961.

[25] Hao Wang. Dominoes and the AEA case of the decision problem. In *Proceedings of the Symposium on Mathematical Theory of Automata (New York, 1962)*, pages 23–55. Polytechnic Press of Polytechnic Inst. of Brooklyn, Brooklyn, N.Y., 1963.

[26] Erik Winfree. *Algorithmic Self-Assembly of DNA*. PhD thesis, California Institute of Technology, June 1998.

A   Formal Definition of the Abstract Tile Assembly Model

This section gives a terse definition of the abstract Tile Assembly Model (aTAM, [26]). This is not a tutorial; for readers unfamiliar with the aTAM, [18] gives an excellent introduction to the model.

Fix an alphabet Σ. Σ^* is the set of finite strings over Σ. Given a discrete object O, $\langle O \rangle$ denotes a standard encoding of O as an element of Σ^*. $\mathbb{Z}$, $\mathbb{Z}^+$, $\mathbb{N}$, $\mathbb{R}^+$ denote the set of integers, positive integers, nonnegative integers, and nonnegative real numbers, respectively. For a set A, $\mathcal{P}(A)$ denotes the power set of A. Given $A \subseteq \mathbb{Z}^2$, the *full grid graph* of A is the undirected graph $G^{\mathrm{f}}_A = (V, E)$, where $V = A$, and for all $u, v \in V$, $\{u, v\} \in E \iff \|u - v\|_2 = 1$; i.e., if and only if u and v are adjacent on the integer Cartesian plane. A *shape* is a set $S \subseteq \mathbb{Z}^2$ such that G^{f}_S is connected.

A *tile type* is a tuple $t \in (\Sigma^*)^4$; i.e., a unit square with four sides listed in some standardized order, each side having a *glue label* (a.k.a. *glue*) $\ell \in \Sigma^*$. For a set of tile types T, let $\Lambda(T) \subset \Sigma^*$ denote the set of all glue labels of tile types in T. Let $\{N, S, E, W\}$ denote the *directions* consisting of unit vectors $\{(0, 1), (0, -1), (1, 0), (-1, 0)\}$. Given a tile type t and a direction $d \in \{N, S, E, W\}$, $t(d) \in \Lambda(T)$ denotes the glue label on t in direction d. We assume a finite set T of tile types, but an infinite number of copies of each tile type, each copy referred to as a *tile*. An *assembly* is a nonempty connected arrangement of tiles on the integer lattice $\mathbb{Z}^2$, i.e., a partial function $\alpha : \mathbb{Z}^2 \dashrightarrow T$ such that $G^{\mathrm{f}}_{\mathrm{dom}\,\alpha}$ is connected and $\mathrm{dom}\,\alpha \neq \varnothing$. The *shape of* α is $\mathrm{dom}\,\alpha$. Given two assemblies $\alpha, \beta : \mathbb{Z}^2 \dashrightarrow T$, we say α is a *subassembly of* β, and we write $\alpha \sqsubseteq \beta$, if $\mathrm{dom}\,\alpha \subseteq \mathrm{dom}\,\beta$ and, for all points $p \in \mathrm{dom}\,\alpha$, $\alpha(p) = \beta(p)$.

A *strength function* is a function $g : \Lambda(T) \to \mathbb{N}$ indicating, for each glue label ℓ, the strength $g(\ell)$ with which it binds. Let α be an assembly and let $p \in \mathrm{dom}\,\alpha$ and $d \in \{N, S, E, W\}$ such that $p + d \in \mathrm{dom}\,\alpha$. Let $t = \alpha(p)$ and $t' = \alpha(p + d)$. We say that the tiles t and t' at positions p and $p + d$ *interact* if $t(d) = t'(-d)$ and $g(t(d)) > 0$, i.e., if the glue labels on their abutting sides are equal and have positive strength. Each assembly α induces a *binding graph* G^{b}_α, a grid graph $G = (V_\alpha, E_\alpha)$, where $V_\alpha = \mathrm{dom}\,\alpha$, and $\{p_1, p_2\} \in E_\alpha \iff \alpha(p_1)$ interacts with $\alpha(p_2)$.[10] Given $\tau \in \mathbb{Z}^+$, α is τ-*stable* if every cut of G^{b}_α has weight at least τ, where the weight of an edge is the strength of the glue it represents. That is, α is τ-stable if at least energy τ is required to separate α into two parts. When τ is clear from context, we say α is *stable*.

A *tile assembly system* (TAS) is a triple $\mathcal{T} = (T, \sigma, g, \tau)$, where T is a finite set of tile types, $\sigma : \mathbb{Z}^2 \dashrightarrow T$ is the finite, τ-stable *seed assembly*, $g : \Lambda(T) \to \mathbb{N}$ is the *strength function*, and $\tau \in \mathbb{Z}^+$ is the *temperature*. Given two τ-stable assemblies $\alpha, \beta : \mathbb{Z}^2 \dashrightarrow T$, we write $\alpha \to^{\mathcal{T}}_1 \beta$ if $\alpha \sqsubseteq \beta$ and $|\mathrm{dom}\,\beta \setminus \mathrm{dom}\,\alpha| = 1$. In this case we say α $\mathcal{T}$-*produces* β *in one step*.[11] If $\alpha \to^{\mathcal{T}}_1 \beta$, $\mathrm{dom}\,\beta \setminus \mathrm{dom}\,\alpha = \{p\}$, and $t = \beta(p)$, we write $\beta = \alpha + (p \mapsto t)$. The $\mathcal{T}$-*frontier* of α is the set $\partial^{\mathcal{T}}\alpha = \bigcup_{\alpha \to^{\mathcal{T}}_1 \beta} \mathrm{dom}\,\beta \setminus \mathrm{dom}\,\alpha$, the set of empty locations at which a tile could stably attach to α.

A sequence of $k \in \mathbb{Z}^+ \cup \{\infty\}$ assemblies $\vec{\alpha} = (\alpha_0, \alpha_1, \ldots)$ is a $\mathcal{T}$-*assembly sequence* if, for all $1 \leq i < k$, $\alpha_{i-1} \to^{\mathcal{T}}_1 \alpha_i$. We write $\alpha \to^{\mathcal{T}} \beta$, and we say α $\mathcal{T}$-*produces* β (in 0 or more steps) if there is a $\mathcal{T}$-assembly sequence $\vec{\alpha} = (\alpha_0, \alpha_1, \ldots)$ of length $k = |\mathrm{dom}\,\beta \setminus \mathrm{dom}\,\alpha| + 1$ such that 1) $\alpha = \alpha_0$, 2) $\mathrm{dom}\,\beta = \bigcup_{0 \leq i < k} \mathrm{dom}\,\alpha_i$, and 3) for all $0 \leq i < k$, $\alpha_i \sqsubseteq \beta$. In this case, we say that β is the *result* of $\vec{\alpha}$, written $\beta = \mathrm{res}(\vec{\alpha})$. If k is finite then it is routine to verify that $\mathrm{res}(\vec{\alpha}) = \alpha_{k-1}$.[12] We say α is $\mathcal{T}$-*producible* if $\sigma \to^{\mathcal{T}} \alpha$, and we write $\mathcal{A}[\mathcal{T}]$ to denote the set of $\mathcal{T}$-producible canonical assemblies. The relation $\to^{\mathcal{T}}$ is a partial order on $\mathcal{A}[\mathcal{T}]$ [15, 17].[13] A $\mathcal{T}$-assembly sequence $\alpha_0, \alpha_1, \ldots$ is *fair* if, for all i and all $p \in \partial^{\mathcal{T}}\alpha_i$, there exists j such that $\alpha_j(p)$ is defined; i.e., no frontier location is "starved".

An assembly α is $\mathcal{T}$-*terminal* if α is τ-stable and $\partial^{\mathcal{T}}\alpha = \varnothing$. It is easy to check that an assembly sequence $\vec{\alpha}$ is fair if and only $\mathrm{res}(\vec{\alpha})$ is terminal. We write $\mathcal{A}_\square[\mathcal{T}] \subseteq \mathcal{A}[\mathcal{T}]$ to denote the set of

---

[10]For $G^{\mathrm{f}}_{\mathrm{dom}\,\alpha} = (V_{\mathrm{dom}\,\alpha}, E_{\mathrm{dom}\,\alpha})$ and $G^{\mathrm{b}}_\alpha = (V_\alpha, E_\alpha)$, G^{b}_α is a spanning subgraph of $G^{\mathrm{f}}_{\mathrm{dom}\,\alpha}$: $V_\alpha = V_{\mathrm{dom}\,\alpha}$ and $E_\alpha \subseteq E_{\mathrm{dom}\,\alpha}$.

[11]Intuitively $\alpha \to^{\mathcal{T}}_1 \beta$ means that α can grow into β by the addition of a single tile; the fact that we require both α and β to be τ-stable implies in particular that the new tile is able to bind to α with strength at least τ. It is easy to check that had we instead required only α to be τ-stable, and required that the cut of β separating α from the new tile has strength at least τ, then this implies that β is also τ-stable.

[12]If we had defined the relation $\to^{\mathcal{T}}$ based on only finite assembly sequences, then $\to^{\mathcal{T}}$ would be simply the reflexive, transitive closure $(\to^{\mathcal{T}}_1)^*$ of $\to^{\mathcal{T}}_1$. But this would mean that no infinite assembly could be produced from a finite assembly, even though there is a well-defined, unique "limit assembly" of every infinite assembly sequence.

[13]In fact it is a partial order on the set of τ-stable assemblies, including even those that are not $\mathcal{T}$-producible.

$\mathcal{T}$-producible, $\mathcal{T}$-terminal canonical assemblies.

A TAS $\mathcal{T}$ is *directed (a.k.a., deterministic, confluent)* if the poset $(\mathcal{A}[\mathcal{T}], \to^{\mathcal{T}})$ is directed; i.e., if for each $\alpha, \beta \in \mathcal{A}[\mathcal{T}]$, there exists $\gamma \in \mathcal{A}[\mathcal{T}]$ such that $\alpha \to^{\mathcal{T}} \gamma$ and $\beta \to^{\mathcal{T}} \gamma$.[14] We say that a TAS $\mathcal{T}$ *strictly self-assembles* a shape $S \subseteq \mathbb{Z}^2$ if, for all $\alpha \in \mathcal{A}_\square[\mathcal{T}]$, dom $\alpha = S$; i.e., if every terminal assembly produced by $\mathcal{T}$ has shape S. If $\mathcal{T}$ strictly self-assembles some shape S, we say that $\mathcal{T}$ is *strict*. Note that the implication "$\mathcal{T}$ is directed $\implies$ $\mathcal{T}$ is strict" holds, but the converse does not hold. We say that $\mathcal{T}$ *uniquely self-assembles* a shape S if $\mathcal{T}$ is directed and it strictly self-assembles S.

When $\mathcal{T}$ is clear from context, we may omit $\mathcal{T}$ from the notation above and instead write $\to_1$, $\to$, $\partial\alpha$, *frontier*, *assembly sequence*, *produces*, *producible*, and *terminal*. We also assume without loss of generality that every single glue or double glue occurring in some tile type in some direction also occurs in some tile type in the opposite direction, i.e., there are no "effectively null" single or double glues.

---

[14]The following two convenient characterizations of "directed" are routine to verify. $\mathcal{T}$ is directed if and only if $|\mathcal{A}_\square[\mathcal{T}]| = 1$. $\mathcal{T}$ is *not* directed if and only if there exist $\alpha, \beta \in \mathcal{A}[\mathcal{T}]$ and $p \in \text{dom } \alpha \cap \text{dom } \beta$ such that $\alpha(p) \neq \beta(p)$.