Natural Computing manuscript No.
(will be inserted by the editor)

Size-Separable Tile Self-Assembly

A Tight Bound for Temperature-1 Mismatch-Free Systems

Andrew Winslow

Received: date / Accepted: date

Abstract We introduce a new property of tile self-assembly systems that we call *size-separability*. A system is size-separable if every terminal assembly is a constant factor larger than any intermediate assembly. Size-separability is motivated by the practical problem of filtering completed assemblies from a variety of incomplete "garbage" assemblies using gel electrophoresis or other mass-based filtering techniques.

Here we prove that any system without cooperative bonding assembling a unique mismatch-free terminal assembly can be used to construct a size-separable system uniquely assembling the same shape. The proof achieves optimal scale factor, temperature, and tile types (within a factor of 2) for the size-separable system.

Keywords 2HAM · hierarchical · aTAM · glues · gel electrophoresis

1 Introduction

The study of theoretical tile self-assembly was initiated by the Ph.D. thesis of Erik Winfree [19]. He proved that systems of passive square particles (called *tiles*) that attach according to matching bonds (called *glues*) are capable of universal computation and efficient assembly of shapes such as squares. Soloveichik and Winfree [17] later proved that these systems are capable of efficient assembly of any finite shape, allowing for an arbitrary scaling of the shape.

An extended abstract of this work was previously published in Unconventional Computation and Natural Computation, LNCS, vol. 8553, pp. 367–378, Springer Berlin Heidelberg (2014)

A. Winslow
Université Libre de Bruxelles
Département d'informatique
ULB CP212, Belgium
E-mail: awinslow@ulb.ac.be

In this original *abstract Tile Assembly Model (aTAM)*, tiles attach singly to a growing seed assembly.

An alternate model, called the *two-handed assembly model (2HAM)* [1,2,4, 6], *hierarchical tile assembly model* [3,14], or *polyomino tile assembly model* [9, 10], allows “seedless” assembly, where tiles can attach spontaneously to form large assemblies that may attach to each other. Cannon et al. [2] proved that this seedless assembly is capable of simulating any seeded assembly process, while also achieving more efficient assembly of some classes of shapes.

A generalization of the 2HAM called the *staged tile assembly model* introduced by Demaine et al. [4] utilizes sequences of *mixings*, where each mixing combines a set of *input assemblies* using a 2HAM assembly process. The products of the mixing are the *terminal assemblies* that cannot combine with any other assembly produced during the assembly process (called a *producible assembly*). This set of terminal assemblies can then be used as input assemblies in another mixing, combined with the sets of terminal assemblies from other mixings.

After a presentation by the author of work [20] on the staged self-assembly model at DNA 19, Erik Winfree commented that the staged tile assembly model has an unrealistic assumption: at the end of each mixing process, all producible but non-terminal assemblies are removed from the mixing. A similar assumption is made in the 2HAM model, where only the terminal assemblies are considered to be “produced” by the system.

Ignoring large producible assemblies is done to simplify the model definition, but allows unrealistic scenarios where “nearly terminal” systems differing from some terminal assembly by a small number of tiles are presumed to be eliminated or otherwise removed at the end of the assembly process. While filtering techniques, including gel electrophoresis, may be employed to obtain filtering of particles at the nanoscale, these techniques generally lack the resolution to distinguish between macromolecules that differ in size by only a small amount. Moreover, these techniques separate particles according to not only size, but (to a lesser extent) shape, charge, rigidity, and other properties. One approach addressing both issues is to maximize the difference between product and intermediate assemblies with respect to an important filtering criteria, such as size, to an extent that low resolution and properties with lesser effects are overwhelmed.

Our results. In this work, we consider efficient assembly of shapes in the 2HAM model under the restriction that terminal assemblies are significantly larger than all non-terminal producible assemblies. We call a system *factor-c size-separable* if the ratio between the smallest terminal assembly and largest non-terminal producible assembly is at least c. Thus, high-factor size-separable systems lack large but non-terminal assemblies, allowing robust filtering of terminal from non-terminal assemblies in these systems.

Our main result is an algorithm for converting 2HAM systems of a special class into size-separable 2HAM systems. A 2HAM system $\mathcal{S} = (T, f, \tau)$ consists of a set of *tiles* T that attach by forming bonds according to their *glues* and a *glue-strength function* f, and two assemblies attach if they can be translated to

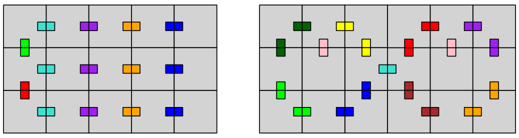

Fig. 1 Assemblies without partitions into equal-sized halves (left) or containing cycle-separated portions that assemble independently (right) cause difficulty in constructing high-factor size-separable systems.

a non-overlapping configuration such that total strength of the bonds formed meets or exceeds the temperature τ of the system. If a system is temperature-1 ($\tau = 1$), then any two assemblies can attach if they have a single matching glue. An assembly is said to be *mismatch-free* if no two coincident tile sides in the assembly or any assembly in the system have different glues. We prove the following:

Theorem 1. *Let $\mathcal{S} = (T, f, 1)$ be a 2HAM system with a mismatch-free unique terminal assembly A. Then there exists a factor-2 size-separable 2HAM system $\mathcal{S}' = (T', f', 2)$ with a unique mismatch-free finite terminal assembly A' such that $|T'| \leq 8|T|$ and A' has the shape of A scaled by a factor of 2.*

For some questions about temperature-1 systems, results have been far easier to obtain for mismatch-free systems than for general systems allowing mismatches. For instance, a lower bound of $2n - 1$ for the assembly of a $n \times n$ square by any temperature 1 aTAM system was conjectured by Rothemund and Winfree [16], and proved for mismatch-free systems by Maňuch, Stacho, and Stoll [11]. Meunier [12] was able to show the same lower bound for systems permitted to have mismatches under the assumption that the seed tile starts in the lower left of the assembly, and removing this restriction remains open. In a similar vein, Reif and Song [15] have shown that temperature-1 mismatch-free aTAM systems are not computationally universal, while the same problem for systems with mismatches permitted is a notoriously difficult problem that remains open, despite significant efforts [8, 7, 18, 13].

In spite of such results, constructing high-factor size-separable versions of temperature-1 mismatch-free systems remains challenging. One difficulty lies in partitioning the assembly into the two equal-sized halves that come together for the final assembly step, e.g. the left assembly in Figure 1. Another difficulty lies in coordinating assembly inside and outside cycles in the bond graph, e.g. during assembly of the right assembly in Figure 1.

Our approach is to first construct a version of A where the bond graph is a tree and there exists a vertex cut of $G(A)$ consisting of a path of *1-occurrence tiles*: tiles whose types appear once in the assembly. This modified version of A is then scaled in size and temperature by a factor of 2, using special 2×2 macrotiles that only assemble along the boundary of the scaled assembly via mixed-strength bonds. Finally, the 1-occurrence tiles forming a vertex cut are given weakened glues such that only completely formed subassemblies on both sides of the cut can attach across the cut.

2 Definitions

All of the definitions given here and used in this work are equivalent to those found in prior work on the two-handed tile assembly model, e.g. [1–3,14].

Assembly systems. In this work we study the *two-handed tile assembly model (2HAM)*, and instances of the model called *systems*. A 2HAM system $\mathcal{S} = (T, f, \tau)$ is specified by three parts: a *tile set* T, a *glue-strength function* f, and a *temperature* $\tau \in \mathbb{N}$.

The tile set T is a set of unit square *tiles*. Each tile $t \in T$ is defined by 4-tuple $t = (g_n, g_e, g_s, g_w)$ consisting of four *glues* from a set Σ of *glue types*, i.e. $g_n, g_e, g_s, g_w \in \Sigma$. The four glues g_n, g_e, g_s, g_w specify the glue types in Σ found on the north (N), east (E), south (S), and west (W) sides of t, respectively. Each glue also defines a *glue-side*, e.g. (g_n, N). Define $g_D(t)$ to be the glue on the side D of t, e.g. $g_{\mathrm{N}}(t) = g_n$.

The glue function $f : \Sigma^2 \to \mathbb{N}$ determines the strength of the *bond* formed by two coincident glue-sides. For any two glues $g, g' \in \Sigma$, $f(g, g') = f(g', g)$. A unique *null glue* $\varnothing \in \Sigma$ has the property that $f(\varnothing, g) = 0$ for all $g \in \Sigma$. For convenience, we sometimes refer to a glue-side with the null glue as a side *without a glue*. In this work we only consider glue functions such that for all $g, g' \in \Sigma$, $f(g, g') = 0$ and if $g \neq \varnothing, f(g, g) > 0$.

Configurations and assemblies. A *configuration* is a partial function $C : \mathbb{Z}^2 \dashrightarrow T$ mapping locations on the integer lattice to tiles. Define $L_D(x, y)$ to be the location in $\mathbb{Z}^2$ one unit in direction D from (x, y), e.g. $L_{\mathrm{N}}(0, 0) = (0, 1)$. For any pair of locations $(x, y), L_D(x, y) \in C$, the *bond strength* between the these tiles is $f(g_D(C(x, y)), g_{D^{-1}}(C(L_D(x, y))))$. If $g_D(C(x, y)) \neq g_{D^{-1}}(C(L_D(x, y)))$, then the pair of tiles is said to form a *mismatch*. A configuration with no mismatches is *mismatch-free*. If $g_D(C(x, y)) = g_{D^{-1}}(C(L_D(x, y)))$, then the glue and directions define a *glue-side pair* $(g_D(C(x, y)), \{D, D^{-1}\})$.

The *bond graph of* C, denoted $G(C)$, is defined as the graph with vertices $\mathrm{dom}(C)$ and edges $\{((x, y), L_D(x, y)) : f(g_D(C(x, y)), g_{D^{-1}}(C(L_D(x, y)))) > 0\}$. That is, the graph induced by the neighboring tiles of C forming positive-strength bonds.

A configuration C is a *τ-stable assembly* or an *assembly at temperature τ* if $\mathrm{dom}(C)$ is connected on the lattice and, for any partition of $\mathrm{dom}(C)$ into two subconfigurations C_1 and C_2, the sum of the bond strengths between tiles at pairs of locations $p_1 \in \mathrm{dom}(C_1)$, $p_2 \in \mathrm{dom}(C_2)$ is at least τ, the temperature of the system. Any pair of assemblies A_1, A_2 are equivalent if they are identical up to a translation by $\langle x, y\rangle$ with $x, y \in \mathbb{Z}$. The *size* of an assembly A is $|\mathrm{dom}(A)|$, and $t \in T$ is a *k-occurrence tile type* in A if $|\{(x, y) \in \mathrm{dom}(A) : A(x, y) = t\}| = k$. The *shape* of an assembly is the polyomino induced by $\mathrm{dom}(A)$, and a shape is *scaled by a factor k* by replacing each cell of the polyomino with a $k \times k$ block of cells.

Two τ-stable assemblies A_1, A_2 are said to *assemble* into a *superassembly* A_3 if A_2 is equivalent to an assembly A'_2 such that $\mathrm{dom}(A_1) \cap \mathrm{dom}(A'_2) = \emptyset$ and A_3 defined by the union of the partial functions A_1 and A'_2 is a τ-stable assembly. Similarly, an assembly A_1 is a *subassembly* of A_2, denoted $A_1 \subseteq A_2$, if A_2 is equivalent to an assembly A'_2 such that $\mathrm{dom}(A_1) \subseteq \mathrm{dom}(A'_2)$.

Producible and terminal assemblies. An assembly A is a *producible assembly* of a 2HAM system $\mathcal{S}$ if A can be assembled from two other producible assemblies or A is a single tile in T. A producible assembly A is a *terminal assembly* of $\mathcal{S}$ if A is producible and A does not assemble with any other producible assembly of $\mathcal{S}$.

We also consider *seeded* versions of some 2HAM systems, where an assembly is producible if it can be assembled from another producible assembly and a single tile of T or is a specified *seed* tile or assembly. Note that for any temperature-1 2HAM system $\mathcal{S}$, the seeded version of $\mathcal{S}$ has the same set of terminal assemblies as $\mathcal{S}$.

If $\mathcal{S}$ has a single terminal assembly A, we call A the *unique terminal assembly (UTA)* of $\mathcal{S}$. In the case that $|A|$ is finite and mismatch-free, we further call A the *unique mismatch-free finite terminal assembly (UMFTA)* of $\mathcal{S}$.

Size-separability. A 2HAM system $\mathcal{S} = (T, f, \tau)$ is a *factor-c size-separable* if for any pair of producible assemblies A, B of $\mathcal{S}$ with A terminal and B not terminal, $|A|/|B| \geq c$. Since this ratio is undefined when $\mathcal{S}$ has infinite producible assemblies, we define such a system to have undefined size-separability. Every system with defined size-separability has factor-c size-separability for some $1 \leq c \leq 2$.

3 Tree-ification

First, we prove that any $\tau = 1$ system producing a unique terminal assembly can be converted into a system with another unique terminal assembly with the same shape but whose bond graph is a tree. This is formalized in the Tree-ification Lemma (Lemma 3) at the end of this section.

Lemma 1 *Let $\mathcal{S} = (T, f, 1)$ be a 2HAM system with UMFTA A. Then no glue-side pair appears twice on a simple cycle of $G(A)$.*

Proof For the sake of contradiction, assume that a glue-side pair appears twice on a simple cycle of $G(A)$ between tile types t_1, t_2, and t_3, t_4. Without loss of generality, assume that the directions of the glue-side pair are E and W. If $t_1 = t_2$ or $t_3 = t_4$, then t_1 or t_3 has the same east and west glue and $\mathcal{S}$ produces an infinite assembly, a contradiction. Without loss of generality, three possible relationships between the tile types remain: (1) $t_1 = t_3$ and $t_2 = t_4$, (2) $t_1 \neq t_3$ and $t_2 = t_4$, (3) $t_1 \neq t_3$ and $t_2 \neq t_4$.

Case 1: $t_1 = t_3$ and $t_2 = t_4$. We carry out seeded assembly augmented with surgery, starting with a seed assembly C_1 consisting of tiles on the simple cycle containing the two occurrences of the adjacent pair t_1, t_2. Consider the types and locations of tiles as they appear along $G(C_1)$, starting just after the

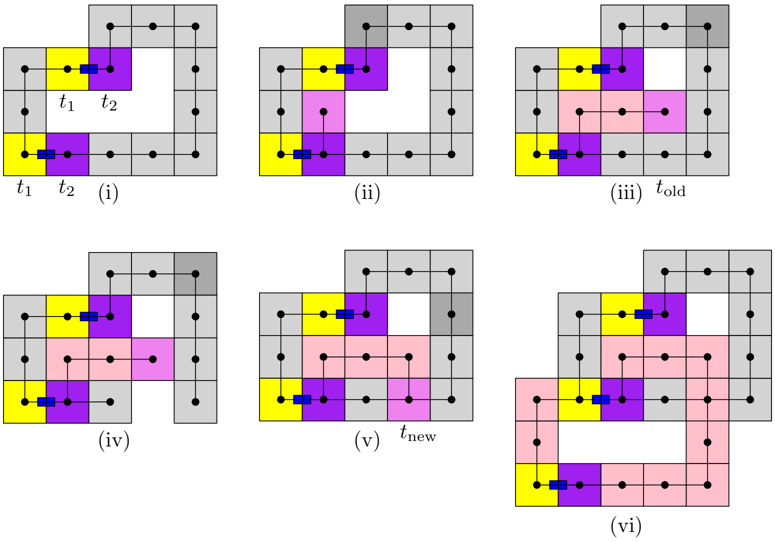

Fig. 2 Seeded assembly of C_2, starting with C_1. The graph $G(C_2)$ has a spanning subgraph of two cycles sharing two vertices.

first occurrence of adjacent pair t_1, t_2. These tiles are attached in order, but translated to start from the *second* occurrence of the adjacent pair t_1, t_2 (see Figure 2).

Because the cycle is simple and thus non-self-intersecting, no tile attachment is blocked by the presence of a tile appearing earlier along the cycle. However, tile attachments may be blocked by the presence of a tile of C_1. We replace the blocking tile, called t_{old}, with the next tile along the cycle, called t_{new} (steps (iii)-(v) of Figure 2).

Just before removing t_{old}, the assembly consists of C_1 and the path of tiles attached so far. The path and C_1 share two tiles: the second occurrence of t_1, t_2 in C_1. So removing t_{old} and placing t_{new} yields a 1-stable and thus producible assembly of $\mathcal{S}$. Since A has no mismatches, t_{new} attaches to both neighbors of t_{old} along the simple cycle of $G(C_1)$.

The sequence of tiles around C_1 form a cycle in $G(C_2)$, and the newly attached sequence of tiles form a second cycle in $G(C_2)$. The cycles also share a common pair of vertices: t_1 and t_2 in the second occurrence of the glue-side pair in C_1.

Additional cycles can be grown indefinitely to produce an infinite sequence of producible assemblies $\{C_n\}$ of $\mathcal{S}$, growing a path from the second occurrence of the glue-side pair on the cycle formed in the $(n-1)$st repetition. The assembly at the end of the nth repetition, called C_n, has a bond graph $G(C_n)$ spanned by a sequence of n cycles with each sharing two vertices with adjacent cycles in the sequence. Replacing a blocking tile with the next tile along the nth cycle is always possible, as removing the tile removes at most one vertex from any cycle and disconnecting the graph requires removing at least two vertices from a cycle.

Since each cycle is an identical translation of the previous cycle, each cycle contains at least one vertex not found in any previous cycle. So $\{C_n\}$ contains arbitrarily large assemblies producible by $\mathcal{S}$, and A cannot be the UMFTA of $\mathcal{S}$, a contradiction.

Case 2: $\boldsymbol{t_1 \neq t_3}$ and $\boldsymbol{t_2 = t_4}$. Consider the subassembly C of A consisting only of the tiles forming the cycle. Since $G(C)$ is a cycle, removing t_1 from C and replacing it with t_3 yields a 1-stable and thus producible assembly C' of $\mathcal{S}$. But C' has two occurrences of the same glue-side pair between tile types t_3 and t_2, forbidden by case 1.

Case 3: $\boldsymbol{t_1 \neq t_3}$ and $\boldsymbol{t_2 \neq t_4}$. Consider a seeded version of $\mathcal{S}$ with an arbitrary tile type as the seed. Additionally restrict the assembly process in two ways: any producible assembly with t_1 and no tile east of t_1 must immediately attach t_4 east of t_1, and any producible assembly with t_2 and no tile west of t_2 must immediately attach t_3 west of t_2. Since no producible assembly of this process contains occurrences of t_1 west of t_2, A is not a producible assembly.

However, any terminal assembly of this seeded version of $\mathcal{S}$ is also a terminal assembly of $\mathcal{S}$, since the restricted seeded assembly process is a special case that does not affect whether an assembly is terminal. Thus the seeded version also has a unique terminal assembly A, a contradiction. □

Lemma 2 *Let $\mathcal{S} = (T, f, 1)$ be a 2HAM system with UMFTA A. Let (g, p) be the glue-side pair of an edge e in $G(A)$. Then if e lies on a simple cycle in $G(A)$, all edges with glue-side pair (g, p) lie on simple cycles of $G(A)$.*

Proof Consider a seeded version of $\mathcal{S}$ with an arbitrary tile type as the seed. Additionally restrict the assembly process in the following way: each time an attaching tile t leaves g exposed on a side in p, carry out the sequence of tile attachments as they occur on the cycle in A containing e. Continue the attachments until attaching the next tile on the cycle is blocked by an existing tile. Each tile attached leaves an exposed glue matching a glue of the next tile in the cycle, so the blocking tile must form a bond with the last attached tile and a cycle is formed. Because this seeded version of $\mathcal{S}$ only modified the order in which tiles attach, it produces the unique terminal assembly A and all edges of A with glue-side pair (g, p) lie on simple cycles of $G(A)$. □

Lemma 3 (Tree-ification Lemma) *Let $\mathcal{S} = (T, f, 1)$ be a 2HAM system with UMFTA A. Then there exists a 2HAM system $\mathcal{S}' = (T', f', 1)$ with UMFTA A' and $|T'| \leq |T|$ such that A' has the shape of A and $G(A')$ is a tree.*

Proof The system $\mathcal{S}'$ and assembly A' are created by repeatedly breaking cycles in the bond graph of A. A cycle is broken by removing an edge of the cycle, i.e. an occurrence of a glue-side pair. The break is implemented in T by replacing all occurrences of glues of the glue-side pair with the null glue.

Such an operation could potentially disconnect $G(A')$, causing A' to no longer be a 1-stable (and producible) assembly of $\mathcal{S}'$. This could happen in two ways: either the glue-side pair occurs twice on a cycle, or the glue-side pair is a cut edge of the bond graph of A' elsewhere. By Lemma 1, no glue-side pair appears twice in any cycle. By Lemma 2, if a glue-side pair appears on a cycle, then it cannot be a single-edge cut anywhere in A. So removing all occurrences of glues in the glue-side pair from $\mathcal{S}'$ does not cause A' to not be 1-stable. Moreover, removing glues cannot add to the set of producible assemblies of $\mathcal{S}'$, so A' remains the unique terminal assembly of $\mathcal{S}$.

Since each break decreases the number of edges in $G(A')$, the breaking process must terminate. At termination, $G(A')$ has no cycles, and so is a tree. Since we only removed glues from tiles of T to create T', the tile set is not larger. However, the removal of glues may cause multiple tiles in T to become indistinguishable. So $|T'| \leq |T|$. □

4 1-Occurrence Tile Types

We also make use of the existence of *1-occurrence tile types*: tile types that appear only once in the terminal assembly of the system. These special tile types are utilized to create unique bonds at locations that partition the unique terminal assembly into equal-sized halves.

Lemma 4 *Let $\mathcal{S} = (T, f, 1)$ be a 2HAM system with UMFTA A with $G(A)$ a tree and $|A| \geq 2$. Then A has at least two 1-occurrence tile types.*

Proof We prove the result by induction on $|T|$, and assume that each tile in T has at least once occurrence in A.[1] For the base case of $|T| = 2$, it must be that $|A| = 2$, otherwise $\mathcal{S}$ produces an infinite assembly. So both tiles in A are 1-occurrence tile types.

For the inductive step, consider two tiles at the leaves of $G(A)$. If both leaf tiles are 1-occurrence tile types, then the claim holds. Otherwise let t be the type of a leaf tile that occurs multiple times in A.

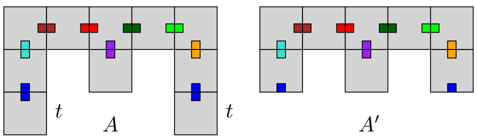

Fig. 3 Removing tile type t from $\mathcal{S} = (T, f, 1)$ yields a system $\mathcal{S}' = (T', f, 1)$ with a UMFTA A', $|T'| < |T|$ and the same number of 1-occurrence tile types.

Since A is mismatch-free, t must have only one non-null glue g and so all occurrences of t are leaves of $G(A)$ (see Figure 3). Without loss of generality, assume g is on the north side of t. No occurrence of t can be replaced with another tile type, since otherwise $\mathcal{S}$ has producible assembly with size $|A|$ not equal to A and A is not the UTA of $\mathcal{S}$. So removing all occurrences of t

[1] Otherwise $\mathcal{S}$ has a second terminal assembly containing a tile type not found in A.

from A yields a UTA A' for the system $\mathcal{S}' = (T - \{t\}, f, 1)$. By the inductive hypothesis, since $|T'| < |T|$, $\mathcal{S}'$ has at least two 1-occurrence tile types. Then since A is A' with the addition of tiles of type t, A also has at least two 1-occurrence tile types. □

Lemma 5 *Let $\mathcal{S} = (T, f, 1)$ be a 2HAM system with UMFTA A with $G(A)$ a tree. For any tile type $t \in T$, the simple path in $G(A)$ between any two occurrences of t uses the same glue-side of t on both occurrences.*

Proof For the sake of contradiction, assume there exists a simple path in $G(A)$ between two occurrences of t such that the path uses, without loss of generality, the north glue-side of the first occurrence and the east glue-side of the second occurrence. Start with the 1-stable assembly formed by the path, including both occurrences of t. Repeatedly carry out the sequence of tile attachments from the second occurrence of t as they appear along the path from the first to the second occurrence of t, including the final attachment of t. Since A is mismatch-free and $G(A)$ is a tree, the repeated sequence of attachments cannot be blocked by an existing tile in the assembly and thus an infinite assembly is formed, a contradiction with the UMFTA A. □

Lemma 6 *Let $\mathcal{S} = (T, f, 1)$ be a 2HAM system with UMFTA A with $G(A)$ a tree. Let v, v' be vertices of $G(A)$ with the tile t at v a 1-occurrence tile. Then there exists a 2HAM system $\mathcal{S}' = (T', f', 1)$ with $|T'| \leq 2|T|$ and UMFTA A' such that $G(A') = G(A)$ and the path from v to v' in $G(A')$ consisting entirely of 1-occurrence tiles.*

Proof The construction of $\mathcal{S}'$ is simple: replace all glues along the path from v to v' with unique glues. Let T' be the updated tile set. First, we prove that the resulting mismatch-free assembly A' is the unique terminal assembly of $\mathcal{S}'$. Certainly A' is a terminal assembly of $\mathcal{S}'$, as otherwise A was not a terminal assembly of $\mathcal{S}$. For any terminal assembly of $\mathcal{S}'$, all occurrences of newly created tiles in $T' - T$ can be swapped with the tiles in T they replaced to yield a terminal assembly of $\mathcal{S}$. So A' is the unique terminal assembly of $\mathcal{S}'$.

Second, we prove that no two tiles along the path from v to v' in A have the same type, implying $|T'| \leq 2|T|$. For the sake of contradiction, suppose that two tiles along the path have the same type t_{rep}. By Lemma 5, the path between two occurrences of t_{rep} must enter both occurrences using the same glue-side. So consider the seeded assembly process starting at t, growing along the path to the first occurrence of t_{rep}, then to the second occurrence, and then growing from the second occurrence using the sequence of tile placements encountered when traveling from the first occurrence *backwards* to t. Since A is the UMFTA of $\mathcal{S}$, this assembly is a subassembly of A and contains one occurrence of t. Since A is mismatch-free and $G(A)$ is a tree, this sequence of tile placements cannot be blocked or form a cycle. So the resulting assembly has two occurrences of t, a contradiction. □

5 A Size-Separable Macrotile Construction

A simple barrier to general high-factor size-separability is the fact that any system with a tree-shaped unique terminal assembly A cannot be factor-c size-separable for any $c > 1 + 1/|A|$, since assembly can always finish with the attachment of a single tile at a leaf of $G(A)$. A more subtle issue is how to partition assemblies into equal-sized 1-stable halves that will come together in the final assembly step. For instance, every partition of the left assembly in Figure 1 into a pair of 1-stable subassemblies has one assembly less than one-third the size of the total assembly. We resolve both of these issues by creating a temperature-2 2HAM system with a unique terminal assembly whose shape is the shape of A scaled by a factor of 2, and whose bond graph has an edge cut of two temperature-1 bonds that partitions $G(A)$ into two subgraphs of equal size.

Lemma 7 *Let $\mathcal{S} = (T, f, 1)$ be a 2HAM system with UMFTA A with $G(A)$ a tree. There exists a 2HAM system $\mathcal{S}' = (T', f', 2)$ with UMFTA A' and $|T'| \leq 4|T|$ such that A' has the shape of A scaled by a factor of 2. Moreover, the tiles of any producible assembly of $\mathcal{S}'$ form a path along the outer (infinite) face of $G(A')$.*

Proof We start by describing common properties of all occurrences of each tile type $t \in T$. Lemma 4 implies that there exists a 1-occurrence tile t_{solo} in A. Since $G(A)$ is a tree, Lemma 5 implies that any path between two occurrences t use the same glue-side pair on both occurrences. So a breadth-first search of $G(A)$ starting at t_{solo} visits all occurrences of t via the same incoming side. Then since A is mismatch-free, if the edges of $G(A)$ are oriented away from t_{solo}, all occurrences of t have the same set of incoming and outgoing edges. So all occurrences of t have their corners visited in the same order during a counterclockwise traversal of the boundary of A.

Macrotiles. We use these conditions to construct unique macrotile versions of each tile type according to their incoming and outgoing edges induced by the breadth-first search starting at t_{solo}. All possible macrotile constructions up to symmetry are shown in Figure 4. For each glue-side pair in the original system, we use two glue-side pairs in the scaled system. The glue-side pair visited first in the counterclockwise traversal of the boundary starting at t is strength-2, while the second is strength-1.

There are also three glues internal to each macrotile attaching each pair of adjacent tiles forming a macroedge whose corresponding edge of the tile either has an outgoing edge induced by the breadth-first search, or has no edge. These glues are unique to the macrotile type. The strengths of each of these glues is determined by whether the closest macroside has glues. If not, then the glue is strength-2, otherwise the glue is strength-1.

If the closest macroside does not contain a glue, then the strength-2 internal glue is necessary to allow assembly to continue along the boundary of the assembly in the counterclockwise direction, e.g. from the northwest to the

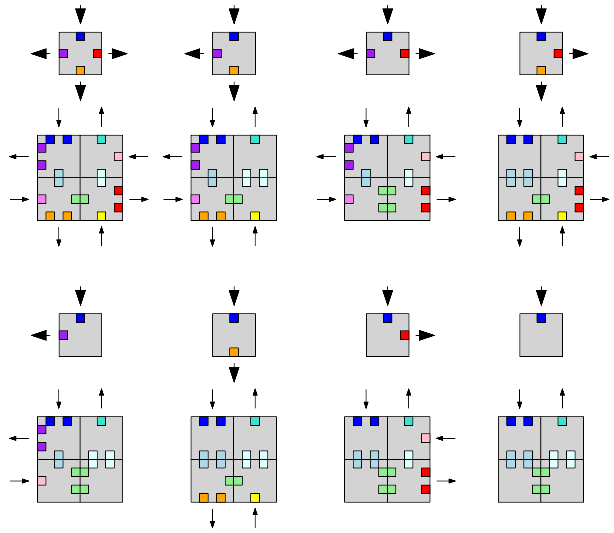

Fig. 4 All combinations, up to symmetry, of incoming and outgoing edges (large arrows) induced by a breath-first search of $G(A)$. The resulting macrotile system A' carries out A at scale 2 in the order that the tiles appear along the boundary (small arrows). All internal macrotile glues are unique to the tile type, while all external macrotile glues correspond to the glues found on the surface of the inducing tile.

southwest tile. If the closest macroside does contain a glue, then the strength-1 internal glue prevents the placement of the next tile in the macrotile until a second tile from an adjacent macrotile has been placed. So no pair of tiles in a macrotile can be present in an assembly unless all tiles between them along a counterclockwise traversal of the boundary of the macrotile assembly are also present.

Scaled assembly of A. By construction, A' is terminal because it corresponds to a mismatch-free terminal assembly in the original system with no exposed glues.

Now we prove by induction that that any subassembly of A' corresponding to a subtree of $G(A)$ is producible. A subtree of size 1 corresponds to a leaf node and is producible (the lower-rightmost case in Figure 4). For larger subtrees, the assembly forms by combining the 4 tiles of the root macrotile to the three (or less) subtree assemblies by growing the assembly in counterclockwise order around the boundary, attaching either a subtree assembly (if the macroside has a glue) or the next tile of the root macrotile. In either case, placing the second root tile along the macroedge is possible, since either the internal glue shared with the previous root tile is strength-2 or a second glue is provided by the subtree assembly.

A path in the outer face of $G(A')$. For the sake of contradiction, suppose a tile t is missing from a producible assembly when traversing the tiles of the assembly as they occur along the outer face of $\boldsymbol{G(A')}$. We claim that the subassembly A_{sub} consisting of the remainder of the tiles in the macrotile

m containing t and any subtrees visited by the traversal before reaching the counterclockwise-most tile of m (northeast in Fig. 4) are only attached to the rest of the assembly by a single strength-1 glue. Observe in Figure 4 that the northeast tile is attached to the next tile in the traversal (the southeast tile of the north macrotile) by a strength-1 glue. It can be verified exhaustively that removing any tile of m leaves only this single strength-1 glue connecting A_{sub} to the rest of the assembly. Thus this assembly is not 2-stable and thus not producible, a contradiction.

Terminal assembly uniqueness. We start by proving that every producible assembly can positioned on a 2×2 macrotile *grid*, where every tile in the southwest corner is a southwest tile of some macrotile, every tile in the northwest corner is the northwest tile of some macrotile, etc. Start by noticing that each glue type appears coincident to only one of 12 edges of the grid: the 4 internal edges of each macrotile, and the 8 external edges. Suppose there is some smallest producible assembly that does not lie on a grid. Then this assembly must be formed by the attachment of two smaller assemblies that do lie on grids, and whose glues utilized in the attachment are coincident to only one of 12 edges of the grid. So if these assemblies are translated to have coincident matching glue-sides, then their grids must also be aligned and the assembly resulting from their attachment also lies on the grid, a contradiction.

Let A'_p be a producible assembly of the macrotile system that is not A'. Construct an assembly A_p of the original input system $\mathcal{S}$ in the following way: replace each macrotile region with a single tile corresponding to one of the tiles in the macrotile region. If such a replacement is *unambiguous*, meaning that all tiles in each macrotile region belong to a common macrotile, then the resulting assembly is a 1-stable (and thus producible) assembly of $\mathcal{S}$.

For the sake of contradiction, assume that there is some A'_p such that replacement is ambiguous. The ambiguity must be due to two tiles in the same macrotile region bonded via external strength-2 glues on *different* macrosides to tiles in adjacent macrotiles, since no macroside has two strength-2 glues (see Figure 4). So there is some path in $G(A'_p)$ from the external glue of one of of these tiles to the external glue to the other consisting of length-2 and length-3 subpaths through other macrotile regions, each consisting of tiles of a common macrotile. So this path can be unambiguously replaced with a path from tiles in $\mathcal{S}$ from one side of a tile location to the other side, with some tile of $\mathcal{S}$ able to attach at this location. But this yields a producible assembly of $\mathcal{S}$ (and thus a subassembly of A) with a cycle, a contradiction. Since constructing A_p from A'_p is always unambiguous, and A_p is a subassembly of A, A'_p is a subassembly of A'.

So every producible assembly of the macrotile system that is not A' is a subassembly of A'. Doty [5] proved that this implies A' can be assembled by first assembling A'_p, then combining A'_p with other producible assemblies. So A'_p is not terminal and A' is the unique terminal assembly of $\mathcal{S}'$. □

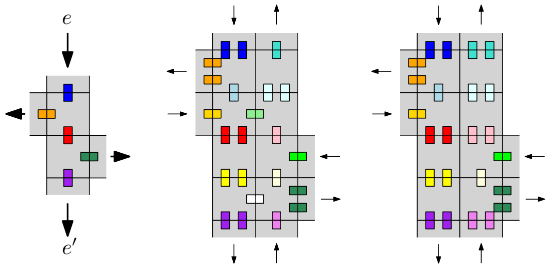

Fig. 5 Modifying 1-occurrence macrotiles along the path from e to e' to eliminate internal glues. The original assembly with breadth-first search directions (left) yields the macrotile assembly (center) by Lemma 7. The modification in the proof of Theorem 1 (right) removes internal glues by strengthening unique external glues.

Theorem 1 *Let $\mathcal{S} = (T, f, 1)$ be a 2HAM system with UMFTA A. Then there exists a factor-2 size-separable 2HAM system $\mathcal{S}' = (T', f', 2)$ with UMFTA A' and $|T'| \leq 8|T|$. Furthermore, A' has the shape of A scaled by a factor of 2.*

Proof Let t_{solo} be a 1-occurrence tile of A, guaranteed to exist by Lemma 4. Starting at t_{solo}, consider a complete counterclockwise traversal of the boundary of A and let e be the first edge encountered, and e' be the edge encountered halfway along the traversal. First, we apply Lemma 6 to create a path of 1-occurrence tiles containing two specific edges e and e'. Second, we apply Lemma 7 to obtain a macrotile system $\mathcal{S}'$ with unique terminal assembly A'.

Third, we modify the 1-occurrence macrotiles along the path from e to e' in A'. Recall that Lemma 7 uses two glues to attach each pair of adjacent macrotiles. The first glue visited during a counterclockwise traversal starting at t_{solo} has strength 2, while the second has strength 1. We increase the strength of the second glue to 2, and eliminate the internal glue closest to this macroside in the macrotile closer to t_{solo} (the macrotile further from t_{solo} does not have this glue). See Figure 5 for an example.

Clearly the same terminal macrotile assembly exists as before, as we have only decreased the constraints for tiles to attach. The addition of two strength-2 glues on the same macroside allows for the possibility that tiles within a macrotile region come from different macrotile tile sets, i.e. do not permit unambiguous replacement, because the multiple macrotiles share the same glue-side. However, since the macrotiles are 1-occurrence, the glue-side pair between them must appear only once in A', otherwise $\mathcal{S}$ produces an assembly containing multiple occurrences of some 1-occurrence tile in A. So it is still the case that producible assemblies of $\mathcal{S}'$ are unambiguous and the terminal assembly uniqueness argument in the proof of Lemma 7 still holds.

This modified A' has a path through the center of the macrotiles from e to e' that contains no glues. So $G(A')$ has a 2-edge cut consisting of half of each of the edges e and e': the half traversed first and halfway along a counterclockwise traversal of the boundary of A', starting at the macrotile corresponding to t_{solo}. The edge glues both have strength 2, since they are external glues between 1-

occurrence tiles just modified to no longer include strength-1 glues. We reduce the strength these two glues to 1, yielding a 2-edge cut with total strength 2.

Size-separability. This system is factor-2 size-separable if any 2-stable subassembly A'_{half} containing tiles adjacent to both edges of the 2-edge cut has size at least $|A'|/2$. By construction, the length of the path along the outer face of $G(A')$ between tiles adjacent to edges the 2-edge cut has length $|A'|/2$. By Lemma 7, all tiles along this path are also in A'_{half} and thus $|A'_{\text{half}}| \geq |A'|/2$.

The total size of T' is at most $8|T|$, since invoking Lemma 6 increases the size of T by at most a factor of 2 and so $|T'| \leq 4 \cdot 2|T|$. □

6 Open Problems

For temperature-1 systems with mismatch-free unique terminal assemblies, our result is nearly optimal as using scale factor and temperature at least 2 are both necessary. Nevertheless, our result is only a first step in understanding what is possible in size-separable systems. A natural next step is to try to extend this result to the same set of systems, except permitting mismatches. We conjecture that a similar result is possible there:

Conjecture 1 Let $\mathcal{S} = (T, f, 1)$ be a 2HAM system with unique finite terminal assembly A. Then there exists factor-2 size-separable system $\mathcal{S}' = (T', f', 2)$ with a unique finite terminal assembly A' and $|T'| = O(|T|)$. Furthermore, A' has the shape A scaled by a factor of $O(1)$.

Extending the result to mismatch-free systems at higher temperatures also is of interest because these systems are generally capable of much more efficient assembly. Soloveichik and Winfree [17] prove that temperature-2 systems using an asymptotically optimal number of tiles and unlimited scale factors can construct any finite shape, and their construction can be modified to be factor-2 size-separable. It remains open to achieve high-factor size-separable systems at temperature 2 using a small scale factor.

Conjecture 2 Let $\mathcal{S} = (T, f, 1)$ be a 2HAM system with UMFTA A. Then there exists factor-2 size-separable system $\mathcal{S}' = (T', f', 2)$ with a unique terminal assembly A' and $|T'| = O(T)$. Furthermore, A' has the shape A scaled by a factor of $O(1)$.

We close by conjecturing that not every system can be made size-separable by paying only a constant factor in scale and tile types, and ask for an example:

Conjecture 3 There exists a 2HAM system $\mathcal{S} = (T, f, \tau)$ with a unique finite terminal assembly A such that any factor-2 size-separable system $\mathcal{S}' = (T', f', \tau')$ with unique finite terminal assembly A' with the shape of A either has $|T'| \geq 100|T|$ or the scale of A' is at least 100.

Acknowledgements The author thanks the anonymous UCNC and Natural Computation reviewers for their comments that improved the presentation and correctness of the paper.

References

1. Abel, Z., Benbernou, N., Damian, M., Demaine, E.D., Demaine, M.L., Flatland, R., Kominers, S.D., Schweller, R.: Shape replication through self-assembly and RNase enzymes. In: Proceedings of the 21st ACM-SIAM Symposium on Discrete Algorithms (SODA), pp. 1045–1064 (2010)
2. Cannon, S., Demaine, E.D., Demaine, M.L., Eisenstat, S., Patitz, M.J., Schweller, R.T., Summers, S.M., Winslow, A.: Two hands are better than one (up to constant factors): Self-assembly in the 2HAM vs. aTAM. In: STACS 2013, *LIPIcs*, vol. 20, pp. 172–184. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik (2013)
3. Chen, H., Doty, D.: Parallelism and time in hierarchical self-assembly. In: Proceedings of the 23rd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), pp. 1163–1182 (2012)
4. Demaine, E.D., Demaine, M.L., Fekete, S.P., Ishaque, M., Rafalin, E., Schweller, R.T., Souvaine, D.L.: Staged self-assembly: nanomanufacture of arbitrary shapes with $O(1)$ glues. Natural Computing **7**(3), 347–370 (2008)
5. Doty, D.: Producibility in hierarchical self-assembly. In: O.H. Ibarra, L. Kari, S. Kopecki (eds.) Unconventional Computation and Natural Computation, *LNCS*, vol. 8553, pp. 142–154. Springer Berlin Heidelberg (2014)
6. Doty, D., Patitz, M.J., Reishus, D., Schweller, R.T., Summers, S.M.: Strong fault-tolerance for self-assembly with fuzzy temperature. In: Foundations of Computer Science (FOCS), pp. 417–426 (2010)
7. Doty, D., Patitz, M.J., Summers, S.M.: Limitations of self-assembly at temperature 1. In: R. Deaton, A. Suyama (eds.) DNA 15, *LNCS*, vol. 5877, pp. 35–44. Springer Berlin Heidelberg (2009)
8. Lathrop, J.I., Lutz, J.H., Patitz, M.J., Summers, S.M.: Computability and complexity in self-assembly. In: A. Beckmann, C. Dimitracopoulos, B. Löwe (eds.) Logic and Theory of Algorithms, *LNCS*, vol. 5028, pp. 349–358. Springer Berlin Heidelberg (2008)
9. Luhrs, C.: Polyomino-safe DNA self-assembly via block replacement. In: A. Goel, F.C. Simmel, P. Sosik (eds.) DNA 14, *LNCS*, vol. 5347, pp. 112–126. Springer Berlin Heidelberg (2009)
10. Luhrs, C.: Polyomino-safe DNA self-assembly via block replacement. Natural Computing **9**(1), 97–109 (2010)
11. Maňuch, J., Stacho, L., Stoll, C.: Journal of computational biology. Journal of Computational Biology **16**(6), 841–852 (2010)
12. Meunier, P.E.: The self-assembly of paths and squares at temperature 1. Tech. rep., arXiv (2014). Http://arxiv.org/abs/1312.1299
13. Meunier, P.E., Patitz, M.J., Summers, S.M., Theyssier, G., Winslow, A., Woods, D.: Intrinsic universality in tile self-assembly requires cooperation. In: Proceedings of the 25th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), pp. 752–771 (2014)
14. Padilla, J., Patitz, M.J., Pena, R., Schweller, R.T., Seeman, N.C., Sheline, R., Summers, S.M., Zhong, X.: Asynchronous signal passing for tile self-assembly:fuel efficient computation and efficient assembly of shapes. In: G. Mauri, A. Dennunzio, L. Manzoni, A.E. Porreca (eds.) Unconventional Computation and Natural Computation (UCNC), *LNCS*, vol. 7956, pp. 174–185. Springer Berlin Heidelberg (2013)
15. Reif, J., Song, T.: The computation complexity of temperature-1 tilings. Tech. rep., Duke University (2014)
16. Rothemund, P.W.K., Winfree, E.: The program-size complexity of self-assembled squares (extended abstract). In: Proceedings of ACM Symposium on Theory of Computing (STOC), pp. 459–468 (2000)
17. Soloveichik, D., Winfree, E.: Complexity of self-assembled shapes. SIAM Journal on Computing **36**(6), 1544–1569 (2007)
18. Summers, S.M.: Universality in algorithm self-assembly. Ph.D. thesis, Iowa State University (2010)
19. Winfree, E.: Algorithmic self-assembly of DNA. Ph.D. thesis, Caltech (1998)
20. Winslow, A.: Staged self-assembly and polyomino context-free grammars. In: D. Soloveichik, B. Yurke (eds.) DNA 19 (2013)