5-6-2019

Optimal staged self-assembly of linear assemblies

Cameron Chalk
University of Texas at Austin

Eric Martinez
The University of Texas Rio Grande Valley

Robert Schweller
The University of Texas Rio Grande Valley

Luis Vega
The University of Texas Rio Grande Valley

Andrew Winslow
The University of Texas Rio Grande Valley

See next page for additional authors

Recommended Citation

Chalk, C., Martinez, E., Schweller, R. et al. Optimal staged self-assembly of linear assemblies. Nat Comput 18, 527–548 (2019). https://doi.org/10.1007/s11047-019-09740-y

Authors
Cameron Chalk, Eric Martinez, Robert Schweller, Luis Vega, Andrew Winslow, and Tim Wylie

Optimal Staged Self-Assembly of Linear Assemblies*

Cameron Chalk[1], Eric Martinez[2], Robert Schweller[2], Luis Vega[2], Andrew Winslow[2], and Tim Wylie[2]

[1] Department of Electrical and Computer Engineering, University of Texas at Austin
ctchalk@utexas.edu
[2] Department of Computer Science, University of Texas - Rio Grande Valley
{eric.m.martinez02, robert.schweller, luis.a.vega01, andrew.winslow, timothy.wylie}@utrgv.edu

Abstract. We analyze the complexity of building linear assemblies, sets of linear assemblies, and $\mathcal{O}(1)$-scale general shapes in the staged tile assembly model. For systems with at most b bins and t tile types, we prove that the minimum number of stages to uniquely assemble a $1 \times n$ *line* is $\Theta(\log_t n + \log_b \frac{n}{t} + 1)$. Generalizing to $\mathcal{O}(1) \times n$ lines, we prove the minimum number of stages is $\mathcal{O}(\frac{\log n - tb - t\log t}{b^2} + \frac{\log\log b}{\log t})$ and $\Omega(\frac{\log n - tb - t\log t}{b^2})$.
Next, we consider assembling sets of lines and general shapes using $t = \mathcal{O}(1)$ tile types. We prove that the minimum number of stages needed to assemble a set of k lines of size at most $\mathcal{O}(1) \times n$ is $\mathcal{O}(\frac{k \log n}{b^2} + \frac{k\sqrt{\log n}}{b} + \log\log n)$ and $\Omega(\frac{k\log n}{b^2})$. In the case that $b = \mathcal{O}(\sqrt{k})$, the minimum number of stages is $\Theta(\log n)$. The upper bound in this special case is then used to assemble "hefty" shapes of at least logarithmic edge-length-to-edge-count ratio at $\mathcal{O}(1)$-scale using $\mathcal{O}(\sqrt{k})$ bins and optimal $\mathcal{O}(\log n)$ stages.

Keywords: Tile self-assembly, staged self-assembly, DNA computing, biocomputing

1 Introduction

Modern technology applications increasingly involve precise design and manufacture of materials and devices at the nanoscale. One approach to nanoscale design is to use *self-assembly*: local interaction rules that direct the aggregation of large numbers of simple units. Seeman [15] discovered that short strands of DNA whose interactions are controlled by attraction between their base sequences can be programmed to carry out such self-assembly. This approach was subsequently extended both experimentally and theoretically by Winfree [16], who introduced the *abstract Tile Assembly Model (aTAM)* to describe systems

* This research was supported in part by National Science Foundation Grants CCF-1117672 and CCF-1555626.

of four-sided planar *tiles* which randomly collide and attach if abutting sides have matching *glues* of sufficient bonding strength. This simple model is computationally universal [16] and experimentally capable of complex algorithmic behaviors [12].

Staged tile assembly. Here we study a tile assembly model introduced by Demaine et al. [9] that permits carrying out assembly in multiple *bins* whose products can be mixed together later, capturing the common experimental technique of decomposing a complex reaction into *stages* of simpler reactions. This model generalizes the *two-handed* [4] or *hierarchical* [7] *tile self-assembly model (2HAM)*. Unlike the aTAM, in which single tiles attach to a multi-tile seed *assembly*, the 2HAM permits arbitrary pairs of assemblies to attach provided they do so via glues of sufficient strength. Growth without a seed occurs naturally in experimental DNA tile systems [3,14], motivating the study of two-handed models.

Efficient assembly. One of the fundamental goals of self-assembly is the design of *efficient* systems that assemble given shapes or patterns. Staged systems have three combinatorial measures of efficiency: the number of tile types (*tile complexity*), the maximum number of bins used in any stage (*bin complexity*), and the number of stages of the system (*stage complexity*). Numerous constructions of efficient staged systems that assemble given shapes [9,11] and patterns [10,17] have been given. Here, we give new, more efficient constructions for assembling height-1 and height-$\mathcal{O}(1)$ rectangles called *lines*, sets of such lines, and *hefty* general shapes of sufficient edge-length-to-edge-count ratio. The results are summarized in Table 1 and described below.

Assembling $1 \times n$ lines. The construction of lines is often used as a subroutine in the assembly of more complex shapes [9,11] or as a simple benchmark [1,6]. In the 2HAM, assembling $1 \times n$ lines requires n tile types; as a corollary, staged systems with 1 bin, 1 stage, and n tile types assemble $1 \times n$ lines.

If $\mathcal{O}(1)$ bins and $\mathcal{O}(\log n)$ stages are permitted, then $\mathcal{O}(1)$ tile types suffice [9], demonstrating a trade-off between two measures of staged system complexity. However, no general trade-off relating all three complexity measures were known prior to this work for assembling $1 \times n$ lines. Here we obtain tight upper and lower bounds that completely characterize the trade-off: for systems of at most t tile types and b bins, the minimum number of stages needed to assemble any $1 \times n$ line is $\Theta(\log_t n + \log_b \frac{n}{t} + 1)$ (Theorems 1 and 2).

A precursor to the upper bound construction was used to generate a set of gadgets to achieve the primary results in [5]. The lower bound approach (Theorem 2) is novel and is not information-theoretic. As a result, it holds for all n rather than almost all n, a common limitation of information-theoretic lower bounds in tile self-assembly.

Assembling $\mathcal{O}(1) \times n$ lines. In the 2HAM, $\mathcal{O}(1) \times n$ lines can be assembled using $n^{\mathcal{O}(1)}$ tile types [8],[3] but a lower bound exceeding $\Omega(\frac{\log n}{\log \log n})$ remains open. The assembly of $\mathcal{O}(1) \times n$ lines has not been studied explicitly in the

[3] The result is given for the aTAM in [8] but the same tile set at temperature 2 in the 2HAM behaves identically.

staged model, however some constructions of Demaine et al. [11] utilize $\mathcal{O}(1) \times n$ line construction as a subroutine.

We give staged systems that use t tile types and b bins that assemble $\mathcal{O}(1) \times n$ lines in $\mathcal{O}(\frac{\log n - tb - t\log t}{b^2} + \frac{\log\log b}{\log t})$ stages (Theorem 4) and prove that for almost all n, $\Omega(\frac{\log n - tb - t\log t}{b^2})$ stages are required (Theorem 5). The upper bound implies a number of new results, including the assembly of $\mathcal{O}(1) \times n$ lines by systems with $\mathcal{O}(1)$ bins, $\mathcal{O}(1)$ stages, and $\mathcal{O}(\frac{\log n}{\log\log n})$ tile types, beating our lower bound of $\Omega(\log n)$ tile types for $1 \times n$ lines (Theorem 2).

This result utilizes the *bit-pad* gadget of [5], and the construction of this pad is the bottleneck for the complexity we achieve. Used naively, this bit-pad gadget can be used to assemble $\mathcal{O}(\log n) \times n$ rectangles within the stated complexity.

Here, we combine with bit-pad gadget with a a novel "sideways" counter to reduce the rectangle height from $\mathcal{O}(\log n)$ to $\mathcal{O}(1)$. This counter involves a non-deterministic guessing strategy for copying sets of $\log n$ bits though a $\mathcal{O}(1)$-height regions, "deactivating" incorrect copies. This technique solves a common difficulty in assembling shapes with narrow regions of low "geometric bandwidth" [2,8] and may have other applications in two-handed self-assembly.

Assembling $\mathcal{O}(1) \times n$ line sets and general shapes. Finally, we consider constructing a set of k $\mathcal{O}(1)$-height lines of differing lengths up to n, in service of general shape construction. The first result is a b-bin, $\mathcal{O}(\frac{k\log n}{b^2} + \frac{k\sqrt{\log n}}{b} + \log\log n)$-stage, $\mathcal{O}(1)$-tile system for assembling any such set of lines (Theorem 6). This is complemented by a lower bound of $\Omega(\frac{k\log n}{b^2})$ (Theorem 7), optimal within an additive $\mathcal{O}(\log\log n)$ factor for small b.

In the special case of systems with $\mathcal{O}(\sqrt{k})$ bins and $\mathcal{O}(1)$ tile types, we give a tight bound of $\Theta(\log n)$ stages (Theorem 8 and Corollary 1). We then use the upper bound to efficiently assemble *hefty* shapes whose edge lengths are at least logarithmic in the number of edges with a $\mathcal{O}(1)$ scale factor increase. This small scale factor contrasts with the results of [5], where more efficient assembly of shapes is obtained, but with unbounded scale factor.

We also prove that any such shape can be assembled by a system with $\mathcal{O}(1)$ tile types, $\mathcal{O}(\sqrt{k})$ bins, and $\mathcal{O}(\log n)$ stages (Theorem 9), optimal for nearly every choice of k and n (Theorem 10) and giving an affirmative answer to a question of [11].

2 The Staged Self-Assembly Model

Here, we give a technical introduction to the two-handed tile assembly model (2HAM) and the staged self-assembly model. The *two-handed tile assembly model* is a model of tile-based assembly processes in which large assemblies can combine freely, in contrast to the well-studied aTAM that limits assembly to single-tile addition to a growing seed assembly. An example system is shown in Figure 1a.

The *staged self-assembly model* is a generalization of the 2HAM in which the terminal assemblies of one 2HAM system can be used, in place of single tiles, as the input assemblies of another 2HAM system. Each system exists in a separate

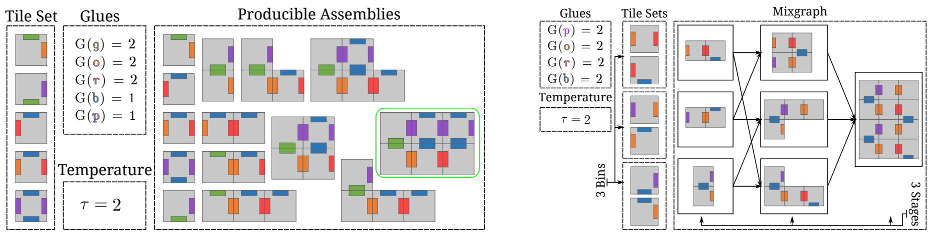
Bins	Tiles	Upper Bound	Lower Bound	Reference
		$1 \times n$ lines		
$\mathcal{O}(1)$	$\mathcal{O}(1)$	$\Theta(\log n)$		Cor. 1, Thm. 3 of [9]
b	t	$\Theta(\log_t n + \log_b \frac{n}{t} + 1)$		1, 2
		$\mathcal{O}(1) \times n$ lines (standard glues)		
1	$n^{\mathcal{O}(1)}$	1		Thm. 3.2 of [8]
b	t	$\mathcal{O}(\frac{\log n - t \log t - tb}{b^2} + \frac{\log \log b}{\log t})$	$\Omega(\frac{\log n - t \log t - tb}{b^2})$	4, 5
		Line sets		
b	$\mathcal{O}(1)$	$\mathcal{O}(\frac{k\sqrt{\log n}}{b} + \frac{k \log n}{b^2} + \log \log n)$	$\Omega(\frac{k \log n}{b^2})$	6, 7
$\mathcal{O}(\sqrt{k})$		$\Theta(\log n)$		8, 1
		Hefty hole-free shapes		
$\mathcal{O}(k)$	$\mathcal{O}(1)$	$\mathcal{O}(\log n)$	$\Omega(\frac{\log n}{k})$	Cor. 1 of [11], Thm. 3 of [9]
$\mathcal{O}(\sqrt{k})$		$\Theta(\log n)$		9, 10

Table 1: An overview of old and new results on problems considered in this paper. Variables t and b denote resource constraints on tile types and bins, respectively. For line sets, k denotes the number of lines in the set, while n denotes the length of the longest line. For general shapes, k denotes the number of edges in the shape, while n denotes the edge length of the minimum-diameter bounding square of the shape. A hefty shape is a shape whose edges are all length at least logarithmic in the number of edges.

bin, and the terminal assemblies of a set of bins can be combined as the input assemblies to another bin in the subsequent *stage*. A staged system then consists of a mixing "graph" that defines which bins' contents are mixed into each bin in the subsequent stage. Figure 1b shows a small example system.

Tiles. A *tile* is a non-rotatable unit square with each edge labeled with a *glue* from a set Σ. Each pair of glues $g_1, g_2 \in \Sigma$ has a non-negative integer *strength* $\text{str}(g_1, g_2)$, with $\text{str}(g_1, g_2) = 0$ unless $g_1 = g_2$. Every set Σ contains a special *null glue* whose strength with every other glue is 0.

Configurations, bond graphs, and stability. A *configuration* is a partial function $A : \mathbb{Z}^2 \to T$ for some set of tiles T, i.e. an arrangement of tiles on a square grid. For a given configuration A, define the *bond graph* G_A to be the weighted grid graph in which each element of $\text{dom}(A)$ is a vertex, and the weight of the edge between a pair of tiles is equal to the strength of the coincident glue



(a) 2HAM Example (b) Staged Self-Assembly Example

Fig. 1: (a) A 2HAM example that uniquely builds a 2×3 rectangle. The top 4 tiles in the tile set all combine with strength-2 glues building the 'L' shape. The tile with blue and purple glues needs two tiles to cooperatively bind to the assembly with strength 2. All possible producibles are shown with the terminal assembly highlighted. (b) A simple staged self-assembly example. The system has 3 bins and 3 stages, as shown in the mixgraph. There are six tiles in our system that we assign to bins as desired. From each stage only the terminal assemblies are added to the next stage. The result of this system is the assembly shown in the output bin in stage 3.

pair. A configuration is said to be τ*-stable* for positive integer τ if every edge cut of G_A has strength at least τ, and is τ*-unstable* otherwise.

Assemblies. For a configuration A and vector $\boldsymbol{u} = \langle u_x, u_y \rangle$ with $u_x, u_y \in \mathbb{Z}^2$, $A + \boldsymbol{u}$ denotes the configuration $A \circ f$, where $f(x, y) = (x + u_x, y + u_y)$. For two configurations A and B, B is a *translation* of A, written $B \simeq A$, provided that $B = A + \boldsymbol{u}$ for some vector $\boldsymbol{u}$. For a configuration A, the *assembly* of A is the set $\tilde{A} = \{B : B \simeq A\}$. An assembly $\tilde{A}$ is a *subassembly* of an assembly $\tilde{B}$, denoted $\tilde{A} \sqsubseteq \tilde{B}$, provided that there exists an $A \in \tilde{A}$ and $B \in \tilde{B}$ such that $A \subseteq B$. An assembly is τ*-stable* provided the configurations it contains are τ-stable. Assemblies $\tilde{A}$ and $\tilde{B}$ are τ*-combinable* into an assembly $\tilde{C}$ provided there exist $A \in \tilde{A}$, $B \in \tilde{B}$, and $C \in \tilde{C}$ such that $A \cup B = C$, $A \cap B = \varnothing$, and $\tilde{C}$ is τ-stable.

Two-handed assembly and bins. We define the assembly process in terms of bins. A *bin* is an ordered tuple (S, τ) where S is a set of *initial* assemblies and τ is a positive integer parameter called the *temperature*. For a bin (S, τ), the set of *produced* assemblies $P'_{(S,\tau)}$ is defined recursively as follows:

1. $S \subseteq P'_{(S,\tau)}$.
2. If $A, B \in P'_{(S,\tau)}$ are τ-combinable into C, then $C \in P'_{(S,\tau)}$.

A produced assembly is *terminal* provided it is not τ-combinable with any other producible assembly, and the set of all terminal assemblies of a bin (S, τ) is denoted $P_{(S,\tau)}$. Intuitively, $P'_{(S,\tau)}$ represents the set of all possible supertiles that can self-assemble from the initial set S, whereas $P_{(S,\tau)}$ represents only the set of supertiles that cannot grow any further.

The assemblies in $P_{(S,\tau)}$ are *uniquely produced* iff for each $x \in P'_{(S,\tau)}$ there exists a corresponding $y \in P_{(S,\tau)}$ such that $x \sqsubseteq y$. Thus unique production implies that every producible assembly can be repeatedly combined with others to form an assembly in $P_{(S,\tau)}$.

Staged assembly systems. An *r-stage b-bin mix graph M* is an acyclic r-partite digraph consisting of rb vertices $m_{i,j}$ for $1 \leq i \leq r$ and $1 \leq j \leq b$, and edges of the form $(m_{i,j}, m_{i+1,j'})$ for some i, j, j'. A *staged assembly system* is a 3-tuple $\langle M_{r,b}, \{T_1, T_2, \ldots, T_b\}, \tau \rangle$ where $M_{r,b}$ is an r-stage b-bin mix graph, T_i is a set of tile types, and τ is an integer temperature parameter.

Given a staged assembly system, for each $1 \leq i \leq r$, $1 \leq j \leq b$, we define a corresponding bin $(R_{i,j}, \tau)$ where $R_{i,j}$ is defined as follows:

1. $R_{1,j} = T_j$ (this is a bin in the first stage);
2. For $i \geq 2$, $R_{i,j} = \Big(\bigcup_{k:\ (m_{i-1,k}, m_{i,j}) \in M_{r,b}} P_{(R_{(i-1,k)}, \tau)} \Big)$.

Thus, the j^{th} bin in stage 1 is provided with the initial tile set T_j, and each bin in any subsequent stage receives an initial set of assemblies consisting of the terminally produced assemblies from a subset of the bins in the previous stage as dictated by the edges of the mix graph.[4] The *output* of the staged system is simply the union of all terminal assemblies from each of the bins in the final stage.[5] We say that this set of output assemblies is *uniquely produced* if each bin in the staged system uniquely produces its respective set of terminal assemblies.

Shapes. The *shape* of an assembly is the polyomino defined by the tile locations, i.e. dom(A), and is *scaled by a factor c* by replacing each cell of the polyomino with a $c \times c$ block of cells. A shape is *hole-free* provided it is simply connected.

Since every shape is a polyomino, its boundary consists of unit-length horizontal and vertical line segments. An *edge* of a shape is a maximal contiguous parallel sequence of such segments. A shape with k edges is *hefty* provided each edge has length at least $\frac{4 \log_2 k + 4}{26} = \Omega(\log k)$. A shape S is an $h \times w$ *line* provided $S = \{y+1, y+2, \ldots, y+h\} \times \{x+1, x+2, \ldots x+w\}$ for some $x, y \in \mathbb{Z}^2$.

3 Assembling $1 \times n$ Lines

We start by analyzing the parameterized staged complexity of assembling $1 \times n$ lines using systems with t tile types and b bins. The following upper bound follows immediately from combining the construction of Lemmas 1 and 2.[6]

Theorem 1. *There exists a constant c such that for any $b, t, n \in \mathbb{N}$ with $b, t > c$ there exists a staged assembly system with b bins and t tile types whose uniquely produced output is a $1 \times n$ line using $\mathcal{O}(\log_t n + \log_b \frac{n}{t} + 1)$ stages.*

[4] The original staged model [9] only considered $\mathcal{O}(1)$ distinct tile types, and thus for simplicity allowed tiles to be added at any stage. Because systems here may have super-constant tile complexity, we restrict tiles to only be added at the initial stage.

[5] This is a slight modification of the original staged model [9] in that the final stage may have multiple bins. However, all of our results apply to both variants of the model.

[6] The "+1" implies the trivial requirement of at least one stage.

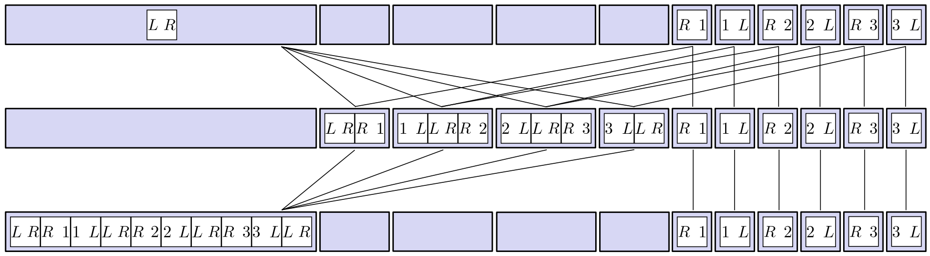


Fig. 2: A high level example using $t = 7$ tile types and 11 bins. Note that the growing assembly in the third stage's leftmost bin maintains the property that L and R glues are exposed on the left and right identical to the single tile in the first stage's leftmost bin. This two-stage mixing process repeats, each time increasing the length of the assembly in the leftmost bin by a factor of $\Theta(t)$.

Lemma 1. *For any $b, t, n \in \mathbb{N}$ with $t \geq 5$ and $b \geq \frac{3}{2}t + \frac{5}{2}$, there exists a staged assembly system with b bins and t tile types whose uniquely produced output is a $1 \times n$ line using $\mathcal{O}(\log_t n + 1)$ stages.*

Lemma 2. *For any $b, t, n \in \mathbb{N}$ with $b > 11$ and $\frac{3}{2}t + \frac{5}{2} > b$, there exists a staged assembly system with b bins and t tile types whose uniquely produced output is a $1 \times n$ line using $\mathcal{O}(\log_b \frac{n}{t-b} + 1)$ stages.*

Detailed proofs are omitted due to space constraints. We instead give a brief overview of the constructions here. Both constructions consider constant fractions t', b' of t, b, respectively.

In the case of Lemma 1 (when $b \geq \frac{3}{2}t + \frac{5}{2}$), t' copies of a $1 \times \ell$ assembly are assembled into a $1 \times \ell t'$ assembly in two stages (initially, $\ell = 1$). An example of this technique for a specific t and b can be seen in Figure 2. Growing by a factor of t' in $\mathcal{O}(1)$ stages implies $\mathcal{O}(\log_t n)$ stages suffice to assemble $1 \times n$ lines, where n is a power of t'. Since this system generates all powers of t' in intermediate stages, values of n that are not powers of two are handled by keeping a *partial growth bin* where k distinct $1 \times (t')^i$ assemblies are concatenated to a growing assembly each time the ith digit in the base t' expansion of n is k. If Lemma 1 does not apply but $b \geq \frac{t}{2}$, then shrinking t by a factor of 3 and applying Lemma 1 implies $\mathcal{O}(\log_{t/3} n + 1) = \mathcal{O}(\log_t n + 1)$ stages suffice.

Otherwise, $t/2 > b$ and Lemma 2 applies. In this case, the above technique fails because there are too few bins for the t' tiles used to connect t' copies of a $1 \times \ell$ assembly. Instead, the assembly is grown by factors of b' (rather than t') using b' tile types as connectors. The $t' - b'$ tiles not used as connectors create a $1 \times (t' - b')$ assembly that is assigned in the first stage to each of the connector tiles' bins, increasing the length of connectors in the first stage. Growing by a factor of b' in $\mathcal{O}(1)$ stages using assemblies which start at length $t' - b'$ implies $\mathcal{O}(\log_b \frac{n}{t-b} + 1)$ stage complexity. Lengths that are not powers of b' are handled identically as in Lemma 1, but utilizing the base b' (rather than base t') expansion of n. Since $\frac{t}{2} > b$, $\mathcal{O}(\log_b \frac{n}{t-b}) = \mathcal{O}(\log_b \frac{n}{t})$.

3.1 Lower bound

A lower bound can also be shown for assembling $1 \times n$ lines by proving an equivalent statement: that a system with s stages, b bins, and t tile types can uniquely assemble only lines of length $\mathcal{O}(\min(t^s, tb^s))$.

Theorem 2. *For any $b, t, n \in \mathbb{N}$, a staged system with b bins and t tile types whose uniquely produced output is a $1 \times n$ line must use $\Omega(\log_t n + \log_b \frac{n}{t} + 1)$ stages.*

4 Assembling $\mathcal{O}(1) \times n$ Lines

We now turn our attention to assembling $\mathcal{O}(1) \times n$ lines. Theorem 4 assembles a $\mathcal{O}(1) \times n$ line using a staged system with t tile types, b bins, and $\mathcal{O}(\frac{\log n - tb - t \log t}{b^2} + \frac{\log \log b}{\log t})$ stages, breaking the $\Omega(\log_t n + \log_b \frac{n}{t} + 1)$ lower bound for $1 \times n$ lines.[7] A complementary lower bound of $\Omega(\frac{\log n - tb - t \log t}{b^2})$ for any constant height is given by Theorem 5.

4.1 Special class of $\mathcal{O}(1) \times n$ lines

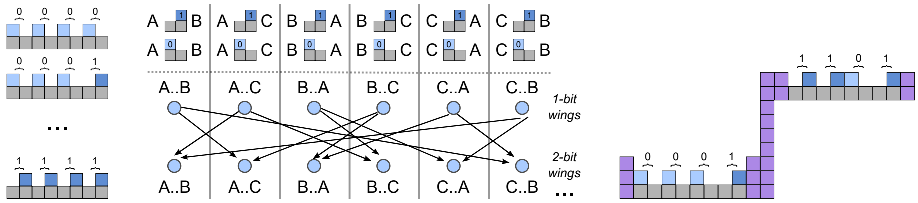
As a warmup, we describe a simpler construction restricted to an infinite set (but not all) of $\mathcal{O}(1) \times n$ lines. This simpler construction already beats the trivial lower bound of n for $1 \times n$ lines in the aTAM. Details of fine-tuning the termination of the counting, yielding the desired result for all n (Theorem 4), is omitted due to space constraints.

Theorem 3. *For any $t, b, n = \Omega(1)$ with $n \in \{i : i = 2^m(2m+3), m \in \mathbb{N}\}$, there exists a temperature-2 staged assembly system with b bins and t tile types whose uniquely produced output is a $\mathcal{O}(1) \times n$ line using $\mathcal{O}(\log \log n)$ stages.*

The construction has four phases:

1. *Counter gadgets* assemble a horizontal counter that counts from 0 to $2^m - 1$ for some $m \in \mathbb{N}$ with $n = 2^m(2m + 3)$. Nondeterminism enables efficiently building all such counter gadgets, but creates many unwanted counter gadgets.
2. *Deactivator gadgets* are assembled. They attach to and *deactivate* unwanted counter gadgets for later disposal.
3. The remaining desired counter gadgets assemble with each other with the help of *gum pads*. The horizontal counter of desired length is assembled.
4. Deactivated counter gadgets are "disposed" by attaching to the bottom of the resulting linear assembly, and the assembly is completed into a rectangle.

[7] Note that the first bound is missing the additive constant to ensure at least one stage. There is still a requirement of at least one stage, but '+1' may be insufficient as the term could be negative.



(a) (b) (c)

Fig. 3: (a) An example of how 4-bit wing gadgets geometrically encode binary strings. (b) Using $\mathcal{O}(1)$ bins and tile types, the number of bits represented on counter gadgets is doubled every stage. (c) Using vertical lines built from $\mathcal{O}(1)$ tile types, left and right wings are nondeterministically brought together to form a counter gadget.

Phase 1: assembling counter gadgets

- *Wing gadgets* are rectangular assemblies with geometric bumps on their north surface, where the bumps geometrically encode an *index* in binary using m bits (Figure 3a).
- A wing gadget *has index* i provided it geometrically encodes a binary string representing i, and all m*-bit* wing gadgets are nondeterministically built using $\mathcal{O}(1)$ tiles, $\mathcal{O}(1)$ bins, and $\mathcal{O}(\log m)$ stages using the mix-graph shown in Figure 3b.
- Two wing gadgets are nondeterministically brought together with $\mathcal{O}(1)$-size assemblies to form *counter gadgets*, as shown in Figure 3c.

Phase 2: deactivating bad counter gadgets

- A *deactivator gadget* detects counter gadgets whose left and right wings do not have the same index and *deactivates* them, preventing their assembly with other counter gadgets in a later stage (Figure 4c). A deactivator gadget is built by assembling an *error checker* and a *deactivator base*.
- Error checkers (Fig. 4a) are assemblies of $\mathcal{O}(1)$ width and $2m+3$ length that, given an m-bit left wing and right wing gadget, can bind to those gadgets if the binary strings represented by those gadgets differ at any of their m bit locations. These gadgets are built using $\mathcal{O}(1)$ tiles, $\mathcal{O}(1)$ bins, and $\mathcal{O}(\log m)$ stages.
- Alone, error checkers cannot completely guarantee that a counter gadget will not interact with the glues of other assemblies. To deactivate the counter gadgets, error checkers are combined with a *deactivator base* to create our deactivator gadgets (Figure 4b). The deactivator base is built $\mathcal{O}(1)$ tiles, $\mathcal{O}(1)$ bins, and $\mathcal{O}(\log m)$ stages.
- Deactivator gadgets are mixed with counter gadgets to deactivate *mismatched* counter gadgets encoding different values on east and west wings (Figure 4c). Deactivated counter gadgets are "disposed" later.

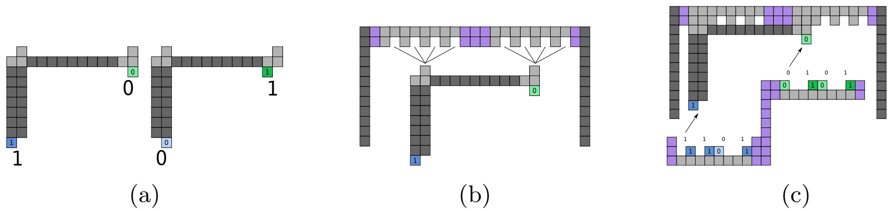


Fig. 4: (a) The two different kinds of error checkers. These attached nondeterministically to the deactivator base using their northern geometric teeth. (b) The error checker attaching to the base, nondeterministically choosing a location, completing our deactivator gadget. Through nondeterminism, deactivator gadgets can be created to detect mismatches at every possible bit location. (c) A deactivator gadget attaching to a mismatched counter gadget.

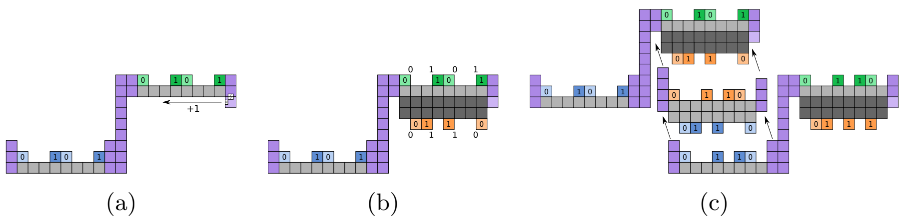


Fig. 5: (a) Increment tiles begin adding *geometric teeth* on the underside of the right wing. (b) The geometric teeth on the underside of the right wing. They represent the same number as the top of the right wing after being incremented by one. (c) A gum pad detects matching geometric teeth and adheres two counter gadgets together.

Phase 3: line formation

- Counter gadgets that have not been deactivated are mixed with $\mathcal{O}(1)$ *increment tiles* that bind to their right wings, exposing a geometric representation of each wing's binary string, incremented by 1 (Figs. 5a and 5b).
- *Gum pads* allow a pair of left and right wings on two counter gadgets to attach side-by-side if the indices of the two wings are identical (Figure 5c). Gum pads are built using $\mathcal{O}(1)$ tile types, $\mathcal{O}(1)$ bins, and $\mathcal{O}(\log m)$ stages.
- Gum pads are mixed with the counter gadgets, allowing them to self-assemble into a linear assembly of length n that counts horizontally from 0 to $2^m - 1$.

Phase 4: disposal and finishing

- Deactivated counter gadgets are disposed by attaching to the bottom of the linear assembly, increasing the assembly's width by $\mathcal{O}(1)$, as shown in Figure 6a.
- A final bin has $\mathcal{O}(1)$ tile types that finish the line by filling any gaps or jagged edges, so that the end result is a rectangle.

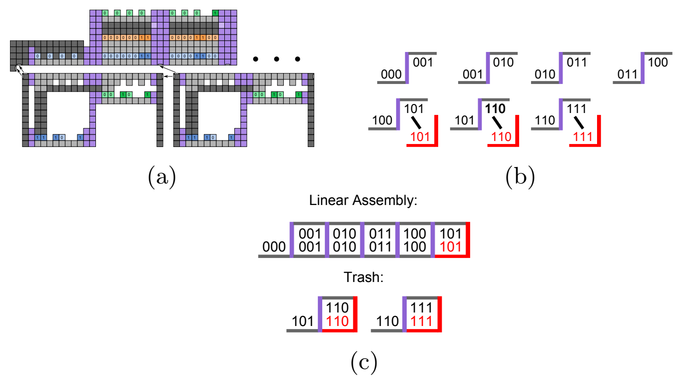


Fig. 6: (a) Disposing of trash assemblies. $\mathcal{O}(1)$ tiles are added to the westmost edge of the counter. Using these tiles, deactivators can attach to the bottom of the counter. The empty space is filled with $\mathcal{O}(1)$ filler tile types. (b) Stopper gadgets for every number at least 5 assembled and mixed with the counter gadgets. (c) Mixed with gum pads, the counter gadgets assemble, a horizontal counter counting from 0 to 5; with stopped counter gadgets as trash.

Complexity Counter gadgets, deactivator gadgets, and gum pads are all assembled using a common technique borrowed from [9] that uses $\mathcal{O}(1)$ tile types and $\mathcal{O}(\log m)$ stages to assemble $\Theta(m)$ assemblies (in $\mathcal{O}(1)$ bins). The same technique is also used to assemble the $\Theta(m)$ lines used in the deactivator gadgets and toothed gum and counter gadget "pads", starting with $\mathcal{O}(1)$ bit gadgets and also using $\mathcal{O}(1)$ bins and $\mathcal{O}(\log m)$ stages. Thus, all aforementioned gadgets can be assembled in parallel using $\mathcal{O}(1)$ tile types, $\mathcal{O}(1)$ bins, and $\mathcal{O}(\log m)$ stages. Since $n = 2^m(2m+3)$, $m = \Theta(\log n)$, and $\mathcal{O}(\log m) = \mathcal{O}(\log \log n)$.

4.2 Generalizing to all n

The construction of Theorem 3 builds counter gadgets using a horizontal counting method to count from 0 to $2^m - 1$ for any $m \in \mathbb{N}$, yielding assemblies of length $n = 2^m(2m+3)$ for all $m \in \mathbb{N}$. General values of n are achieved by fine-tuning length at two scales: "large scale" via terminating the counter early at a specific value before the desired n and "small scale" via attaching a smaller assembly to reach exactly n from where the counter terminated.

Terminating the counter early is achieved by deactivating "high-value" counter gadgets with values larger than a specified value using *stopper gadgets*, as shown in Figure 6. Encoding the counter termination value dominates the stage complexity, giving the following result:

Theorem 4. *For any $t, b, n \in \mathbb{N}$ with $t, b = \Omega(1)$, there exists a temperature-2 staged system with b bins and t tile types that assembles a $\mathcal{O}(1) \times n$ line using $\mathcal{O}(\frac{\log n - tb - t\log t}{b^2} + \frac{\log\log b}{\log t})$ stages.*

4.3 Lower bounds for $\mathcal{O}(1) \times n$ lines

Lower bounds for assembling $\mathcal{O}(1) \times n$ lines are obtained using information-theoretic arguments based on combining the bound on information content from [5] with the lower bound of $\lceil \log_2 n \rceil$ on the number of bits needed to specify n for almost all n:

Theorem 5. *For any $b, t \in \mathbb{N}$ and almost all $n \in \mathbb{N}$, any staged self-assembly system with b bins and t tile types and uniquely assembles a $\mathcal{O}(1) \times n$ line must use $\Omega(\frac{\log n - tb - t \log t}{b^2})$ stages.*

5 Assembling $\mathcal{O}(1) \times n$ Line Sets

Now we consider extending the construction of a $\mathcal{O}(1) \times n$ line to a set of k such lines, working towards the construction of hefty shapes in Section 6. The first upper bound construction uses parallel instances of the Theorem 4 construction to assemble multiple lines in parallel with a comparable number of stages.

Theorem 6. *Let $L = \{n_1, \ldots, n_k\} \subseteq \mathbb{N}$ with $n = \max(L)$. There exists a staged assembly system with $\mathcal{O}(1)$ tile types, b bins, and $\mathcal{O}(\frac{k\sqrt{\log n}}{b} + \frac{k \log n}{b^2} + \log \log n)$ stages whose uniquely produced output is a set of $\mathcal{O}(1) \times n_i$ lines for all $n_i \in L$.*

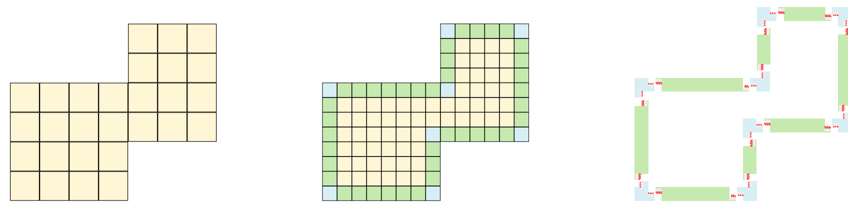
Theorem 7. *Let $L = \{n_1, \ldots, n_k\} \subseteq \mathbb{N}$ with $n = \max(L)$. For almost all L, any staged self-assembly system with $\mathcal{O}(1)$ tile types and b bins that assembles $\mathcal{O}(1) \times n_i$ lines for all $n_i \in L$ has $\Omega(\frac{k \log n}{b^2})$ stages.*

In the case that $b = \mathcal{O}(\sqrt{\log n})$, the prior two theorems are tight up to additive terms. However, as b increases, the "crazy mixing" approach [9] used in the modular construction of Theorem 6 fails to utilize the growing number of possible mix graphs. The next construction achieves optimal stage complexity for large bin counts, specifically bin counts scaling with k:

Theorem 8. *Let $L = \{n_1, \ldots, n_k\} \subseteq \mathbb{N}$ with $n = \max(L)$. There exists a staged self-assembly system with $\mathcal{O}(1)$ tile types, $\mathcal{O}(\sqrt{k})$ bins, and $\mathcal{O}(\log n)$ stages that assembles $\mathcal{O}(1) \times n_i$ lines for all $n_i \in L$.*

The following lower bound matches this construction and follows directly from Theorem 7.

Corollary 1. *Let $L = \{n_1, \ldots, n_k\} \subseteq \mathbb{N}$ with $n = \max(L)$. For almost all L, any staged self-assembly system with $\mathcal{O}(1)$ tile types and $\mathcal{O}(\sqrt{k})$ bins that assembles $\mathcal{O}(1) \times n_i$ lines for all $n_i \in L$ has $\Omega(\log n)$ stages.*



(a) (b) (c)

Fig. 7: (a) A hefty hole-free shape to be constructed. (b) The shape scaled by factor 2 with backbone (green) and vertices (blue). (c) The decomposition of the backbone into vertices and lines.

6 Assembling Hefty Shapes

The efficient line set assembly result of Theorem 8 can be combined with a technique of [11] to assemble general shapes optimally:

Theorem 9. *Let S be a hefty hole-free shape with k vertices and minimum-diameter bounding square of edge length n. There exists a $\tau = 2$ staged system with $\mathcal{O}(\sqrt{k})$ bins, $\mathcal{O}(1)$ tile types, and $\mathcal{O}(\log n)$ stages that uniquely produces S scaled by a factor $\mathcal{O}(1)$.*

Theorem 10. *Let S be a hefty shape with k edges and minimum-diameter bounding square of edge length n with $k = \mathcal{O}(n^{2-\varepsilon})$ for some $\varepsilon > 0$. For almost all S, any staged self-assembly system with $\mathcal{O}(1)$ tile types and $\mathcal{O}(\sqrt{k})$ bins that assembles S has $\Omega(\log n)$ stages.*

The technique of [11] is to first efficiently create the *backbone* of the given shape, then fill in the backbone of the shape using $\mathcal{O}(1)$ tile types and one stage (see Figure 7). For a shape with k vertices (and edges), this approach uses $\mathcal{O}(k)$ bins.

We reduce the bin complexity to $\mathcal{O}(\sqrt{k})$ by replacing k separate bins, each containing a different edge assembly, with $\mathcal{O}(\sqrt{k})$ bins, each containing many edge assemblies each labeled with geometric teeth, similar to the construction of Theorem 8. In exchange, $\mathcal{O}(\log n)$ additional stages must be used to assemble these edge assemblies.

References

1. Adleman, L., Cheng, Q., Goel, A., Huang, M.D., Wasserman, H.: Linear self-assemblies: equilibria, entropy and convergence rates. In: 6th International Conference on Difference Equations and Applications (2001)
2. Adleman, L.M., Cheng, Q., Goel, A., Huang, M.D.A., Kempe, D., de Espanés, P.M., Rothemund, P.W.K.: Combinatorial optimization problems in self-assembly. SIAM Journal on Computing 38(6), 2356–2381 (2009)

3. Barish, R.D., Schulman, R., Rothemund, P.W.K., Winfree, E.: An information-bearing seed for nucleating algorithmic self-assembly. Proceedings of the National Academy of Sciences 106(15), 6054–6059 (March 2009)
4. 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 (2013)
5. Chalk, C., Martinez, E., Schweller, R., Vega, L., Winslow, A., Wylie, T.: Optimal staged self-assembly of general shapes. Algorithmica (2017)
6. Chandran, H., Gopalkrishnan, N., Reif, J.: Tile complexity of linear assemblies. SIAM Journal on Computing 41(4), 1051–1073 (2012)
7. Chen, H.L., Doty, D.: Parallelism and time in hierarchical self-assembly. In: 23rd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA). pp. 1163–1182. SIAM (2012)
8. Cheng, Q., Aggarwal, G., Goldwasser, M.H., Kao, M.Y., Schweller, R.T., de Espanés, P.M.: Complexities for generalized models of self-assembly. SIAM Journal on Computing 34, 1493–1515 (2005)
9. 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)
10. Demaine, E.D., Eisenstat, S., Ishaque, M., Winslow, A.: One-dimensional staged self-assembly. Natural Computing 12(2), 247–258 (2013)
11. Demaine, E.D., Fekete, S.P., Scheffer, C., Schmidt, A.: New geometric algorithms for fully connected staged self-assembly. In: Proc. of the 21st Int. Conf. on DNA Comp. and Molecular Programming, LNCS, vol. 9211, pp. 104–116. Springer (2015)
12. Evans, C.: Crystals that Count! Physical Principles and Experimental Investigations of DNA Tile Self-Assembly. Ph.D. thesis, Caltech (2014)
13. Rothemund, P.W.K., Winfree, E.: The program-size complexity of self-assembled squares (extended abstract). In: Proc. of the 32nd ACM Sym. on Theory of Computing. pp. 459–468. STOC'00 (2000)
14. Schulman, R., Winfree, E.: Synthesis of crystals with a programmable kinetic barrier to nucleation. Proceedings of the National Academy of Sciences 104(39), 15236–15241 (2007)
15. Seeman, N.C.: Nucleic-acid junctions and lattices. Journal of Theoretical Biology 99, 237–247 (1982)
16. Winfree, E.: Algorithmic Self-Assembly of DNA. Ph.D. thesis, Caltech (1998)
17. Winslow, A.: Staged self-assembly and polyomino context-free grammars. Natural Computing 14(2), 293–302 (2015)