Staged Self-assembly: Nanomanufacture of Arbitrary Shapes with $O(1)$ Glues

Erik D. Demaine[1], Martin L. Demaine[1], Sándor P. Fekete[2], Mashhood Ishaque[3], Eynat Rafalin[4], Robert T. Schweller[5], and Diane L. Souvaine[3]

[1] MIT Computer Science and Artificial Intelligence Laboratory, 32 Vassar St., Cambridge, MA 02139, USA*
{edemaine,mdemaine}@mit.edu
[2] Institut für Mathematische Optimierung, Technische Universität Braunschweig, Pockelsstr. 14, 38106 Braunschweig, Germany
s.fekete@tu-bs.de
[3] Department of Computer Science, Tufts University, Medford, MA 02155, USA**
{mishaq01,dls}@cs.tufts.edu
[4] Google Inc.,***
erafalin@cs.tufts.edu
[5] Department of Computer Science, University of Texas Pan American, 1201 W. University Drive, Edinburg, Texas 78539, USA
schwellerr@cs.panam.edu

Abstract. We introduce *staged self-assembly* of Wang tiles, where tiles can be added dynamically in sequence and where intermediate constructions can be stored for later mixing. This model and its various constraints and performance measures are motivated by a practical nanofabrication scenario through protein-based bioengineering. Staging allows us to break through the traditional lower bounds in tile self-assembly by encoding the shape in the staging algorithm instead of the tiles. All of our results are based on the practical assumption that only a constant number of glues, and thus only a constant number of tiles, can be engineered, as each new glue type requires significant biochemical research and experiments. Under this assumption, traditional tile self-assembly cannot even manufacture an $n \times n$ square; in contrast, we show how staged assembly enables manufacture of arbitrary orthogonal shapes in a variety of precise formulations of the model.

1 Introduction

Self-assembly is the process by which an organized structure can form spontaneously from simple parts. It describes the assembly of diverse natural structures such as crystals, DNA helices, and microtubules. In nanofabrication, the idea is to co-opt natural self-assembly processes to build desired structures, such as a sieve for removing viruses from serum, a drug-delivery device for targeted chemotherapy or brachytherapy, a magnetic device for medical imaging, a catalyst for enzymatic reactions, or a

* Partially supported by NSF CAREER award CCF-0347776 and DOE grant DE-FG02-04ER25647.
** Partially supported by NSF grant CCF-0431027.
*** Work performed while at Tufts University. Partially supported by NSF grant CCF-0431027.

M.H. Garzon and H. Yan (Eds.): DNA 13, LNCS 4848, pp. 1–14, 2008.

biological computer. Self-assembly of artificial structures has promising applications to nanofabrication and biological computing. The general goal is to design and manufacture nanoscale pieces (e.g., strands of DNA) that self-assemble uniquely into a desired macroscale object (e.g., a computer).

Our work is motivated and guided by an ongoing collaboration with the Sackler School of Graduate Biomedical Sciences that aims to nanomanufacture sieves, catalysts, and drug-delivery and medical-imaging devices, using protein self-assembly. Specifically, the Goldberg Laboratory is currently developing technology to bioengineer (many copies of) rigid struts of varying lengths, made of several proteins, which can join collinearly to each other at compatible ends. These struts occur naturally as the "legs" of the *T4 bacteriophage*, a virus that infects bacteria by injecting DNA. In contrast to nanoscale self-assembly based on DNA [WLWS98, MLRS00, RPW04, BRW05, See98, SQJ04, Rot06], which is inherently floppy, these nanorod structures are extremely rigid and should therefore scale up to the manufacture of macroscale objects.

The traditional, leading theoretical model for self-assembly is the two-dimensional *tile assembly model* introduced by Winfree in his Ph.d. thesis [Win98] and first appearing at STOC 2000 [RW00]. The basic building blocks in this model are *Wang tiles* [Wan61], unrotatable square tiles with a specified glue on each side, where equal glues have affinity and may stick. Tiles then self-assemble into supertiles: two (super)tiles nondeterministically join if the sum of the glue affinities along the attachment is at least some threshold τ, called *temperature*. This basic model has been generalized and extended in many ways [Adl00, ACGH01, ACG$^+$02, SW04, ACG$^+$05, RW00, KS06]. The model should be practical because Wang tiles can easily simulate the practical scenario in which tiles are allowed to rotate, glues come in pairs, and glues have affinity only for their unique mates. In particular, we can implement such tiles using two unit-length nanorods joined at right angles at their midpoints to form a plus sign.

Most theoretical research in self-assembly considers the minimum number of distinct tiles—the *tile complexity* t—required to assemble a shape uniquely. In particular, if we allow the desired shape to be scaled by a possibly very large factor, then in most models the minimum possible tile complexity (the smallest "tile program") is $\Theta(K/\lg K)$ where K is the Kolmogorov complexity of the shape [SW04]. In practice, the limiting factor is the number of distinct glues—the *glue complexity* g—as each new glue type requires significant biochemical research and experiments. For example, a set of DNA-based glues requires experiments to test whether a collection of codewords have a "conflict" (a pair of noncomplementary base sequences that attach to each other), while a set of protein-based glues requires finding pairs of proteins with compatible geometries and amino-acid placements that bind (and no other pairs of which accidentally bind). Of course, tile and glue complexities are related: $g \leq t \leq g^4$.

We present the *staged tile assembly model*, a generalization of the tile assembly model that captures the temporal aspect of the laboratory experiment, and enables substantially more flexibility in the design and fabrication of complex shapes using a small tile and glue complexity. In its simplest form, staged assembly enables the gradual addition of specific tiles in a sequence of stages. In addition, any tiles that have not yet attached as part of a supertile can be washed away and removed (in practice, using a weight-based filter, for example). More generally, we can have any number of *bins* (in

reality, batches of liquid solution stored in separate containers), each containing tiles and/or supertiles that self-assemble as in the standard tile assembly model. During a stage, we can perform any collection of operations of two types: (1) add (arbitrarily many copies of) a new tile to an existing bin; and (2) pour one bin into another bin, mixing the contents of the former bin into the latter bin, and keeping the former bin intact. In both cases, any pieces that do not assemble into larger structures are washed away and removed. These operations let us build intermediate supertiles in isolation and then combine different supertiles as whole structures. Now we have two new complexity measures in addition to tile and glue complexity: the number of stages—or *stage complexity* s—measures the time required by the operator of the experiment, while the number of bins—or *bin complexity* b—measures the space required for the experiment.[1] (When both of these complexities are 1, we obtain the regular tile assembly model.)

Our results. We show that staged assembly enables substantially more efficient manufacture in terms of tile and glue complexity, without sacrificing much in stage and bin complexity. All of our results assume the practical constraint of having only a small constant number of glues and hence a constant number of tiles. In contrast, an information-theoretic argument shows that this assumption would limit the traditional tile assembly model to constructing shapes of constant Kolmogorov complexity.

For example, we develop a method for self-assembling an $n \times n$ square for arbitrary $n > 0$, using 16 glues and thus $O(1)$ tiles (independent of n), and using only $O(\log \log n)$ stages, $O(\sqrt{\log n})$ bins, and temperature $\tau = 2$ (Section 4.2). Alternatively, with the minimum possible temperature $\tau = 1$, we can self-assemble an $n \times n$ square using 9 glues, $O(1)$ tiles and bins, and $O(\log n)$ stages (Section 4.1). In contrast, the best possible self-assembly of an $n \times n$ square in the traditional tile assembly model has tile complexity $\Theta(\log n / \log \log n)$ [ACGH01, RW00], or $\Theta(\sqrt{\log n})$ in a rather extreme generalization of allowable pairwise glue affinities [ACG$^+$05].

More generally, we show how to self-assemble arbitrary shapes made up of n unit squares in a variety of precise formulations of the problem. Our simplest construction builds the shape using 2 glues, 16 tiles, $O(\text{diameter})$ stages, and $O(1)$ bins, but it only glues tiles together according to a spanning tree, which is what we call the *partial connectivity model* (Section 5.1). All other constructions have *full connectivity*: any two adjacent unit squares are built by tiles with matching glues along their shared edge. In particular, if we scale an arbitrary hole-free shape larger by a factor of 2, then we can self-assemble with full connectivity using 8 glues, $O(1)$ tiles, and $O(n)$ stages and bins (Section 5.2). We also show how to simulate a traditional tile assembly construction with t tiles by a staged assembly using 3 glues, $O(1)$ tiles, $O(\log \log t)$ stages, $O(t)$ bins, and a scale factor of $O(\log t)$ (Section 5.3). If the shape happens to be monotone in one direction, then we can avoid scaling and still obtain full connectivity, using 9 glues, $O(1)$ tiles, $O(\log n)$ stages, and $O(n)$ bins (details omitted in this version).

Table 1 summarizes our results in more detail, in particular elaborating on possible trade-offs between the complexities. The table captures one additional aspect of our

[1] Here we view the mixing time required in each stage (and the volume of each bin) as a constant, mainly because it is difficult to analyze precisely from a thermodynamic perspective, as pointed out in [Adl00]. In our constructions, we believe that a suitable design of the relative concentrations of tiles (a feature not captured by the model) leads to reasonable mixing times.

Table 1. Summary of the glue, tile, bin, and stage complexities, the temperature τ, the scale factor, the connectivity, and the planarity of our staged assemblies and the relevant previous work

$n \times n$ square	Glues	Tiles	Bins	Stages	τ	Scale	Conn.	Planar
Previous work [ACGH01, RW00]	$\Theta(\frac{\log n}{\log \log n})$		1	1	2	1	full	yes
Jigsaw technique (§4.1)	9	$O(1)$	$O(1)$	$O(\log n)$	1	1	full	yes
Crazy mixing (§4.2)	16	$O(1)$	B	$O(\lceil \frac{\log n}{B^2} \rceil \log B)$	2	1	full	yes
Crazy mixing, $B = \sqrt{\log n}$	16	$O(1)$	$\sqrt{\log n}$	$O(\log \log n)$	2	1	full	yes

General shape with n tiles	Glues	Tiles	Bins	Stages	τ	Scale	Conn.	Planar
Previous work [SW04]	$\Theta(K/\log K)$		1	1	2	unbounded	partial	no
Arbitrary shape with n tiles (§5.1)	2	16	$O(\log n)$	$O(\text{diameter})$	1	1	partial	no
Hole-free shape with n tiles (§5.2)	8	$O(1)$	$O(n)$	$O(n)$	1	2	full	no
Simulation of 1-stage tiles T (§5.3)	3	$O(1)$	$O(\lvert T\rvert)$	$O(\log \log \lvert T\rvert)$	1	$O(\log \lvert T\rvert)$	partial	no
Monotone shapes with n tiles (omitted)	9	$O(1)$	$O(n)$	$O(\log n)$	1	1	full	yes

constructions: Planarity. Consider two jigsaw puzzle pieces with complex borders lying on a flat surface. It may not be possible to slide the two pieces together while both remain on the table. Rather, one piece must be lifted off the table and dropped into position. Our current model of assembly intuitively permits supertiles to be placed into position from the third dimension, despite the fact that it may not be possible to assemble within the plane. A *planar* construction guarantees assembly of the final target shape even if we restrict assembly of supertiles to remain completely within the plane. This feature seems desirable, though it may not be essential in two dimensions because reality will always have some thickness in the third dimension (2.5D). However, the planarity constraint (or *spatiality* constraint in 3D) becomes more crucial in 3D assemblies, so this feature gives an indication of which methods should generalize to 3D; see Section 6.

Related Work. There are a handful of existing works in the field of DNA self-assembly that have proposed very basic multiple stage assembly procedures. John Reif introduced a step-wise assembly model for local parallel biomolecular computing [Rei99]. In more recent work Park et. al. have considered a simple hierarchical assembly technique for the assembly of DNA lattices [PPA$^+$06]. Somei et. al. have considered a microfluidic device for stepwise assembly of DNA tiles [SKFM05]. While all of these works use some form of stepwise or staged assembly, they do not study the complexity of staged assembly to the depth that we do here. Further, none consider the concept of bin complexity.

2 The Staged Assembly Model

In this section, we present basic definitions common to most assembly models, then we describe the staged assembly model, and finally we define various metrics to measure the efficiency of a staged assembly system.

Tiles and tile systems. A *(Wang) tile* t is a unit square defined by the ordered quadruple $\langle \text{north}(t), \text{east}(t), \text{south}(t), \text{west}(t) \rangle$ of glues on the four edges of the tile. Each *glue*

is taken from a finite alphabet Σ, which includes a special "null" glue denoted null. For simplicity of bounds, we do not count the null glue in the *glue complexity* $g = |\Sigma| - 1$.

A *tile system* is an ordered triple $\langle T, G, \tau \rangle$ consisting of the *tileset* T (a set of distinct tiles), the *glue function* $G : \Sigma^2 \to \{0, 1, \ldots, \tau\}$, and the *temperature* τ (a positive integer). It is assumed that $G(x, y) = G(y, x)$ for all $x, y \in \Sigma$ and that $G(\mathrm{null}, x) = 0$ for all $x \in \Sigma$. Indeed, in all of our constructions, as in the original model of Adleman [Adl00], $G(x, y) = 0$ for all $x \neq y$[2], and each $G(x, x) \in \{1, 2, \ldots, \tau\}$. The *tile complexity* of the system is $|T|$.

Configurations. Define a *configuration* to be a function $C : \mathbb{Z}^2 \to T \cup \{\mathrm{empty}\}$, where empty is a special tile that has the null glue on each of its four edges. The *shape* of a configuration C is the set of positions (i, j) that do not map to the empty tile. The shape of a configuration can be disconnected, corresponding to several distinct supertiles.

Adjacency graph and supertiles. Define the *adjacency graph* G_C of a configuration C as follows. The vertices are coordinates (i, j) such that $C(i, j) \neq \mathrm{empty}$. There is an edge between two vertices (x_1, y_1) and (x_2, y_2) if and only if $|x_1 - x_2| + |y_1 - y_2| = 1$. A *supertile* is a maximal connected subset G' of G_C, i.e., $G' \subseteq G_C$ such that, for every connected subset H, if $G' \subseteq H \subseteq G_C$, then $H = G'$. For a supertile S, let $|S|$ denote the number of nonempty positions (tiles) in the supertile. Throughout this paper, we will informally refer to (lone) tiles as a special case of supertiles.

If every two adjacent tiles in a supertile share a positive strength glue type on abutting edges, the supertile is *fully connected*.

Two-handed assembly and bins. Informally, in the two-handed assembly model, any two supertiles may come together (without rotation or flipping) and attach if their strength of attachment, from the glue function, meets or exceeds a given temperature parameter τ.

Formally, for any two supertiles X and Y, the *combination* set $C^{\tau}_{(X,Y)}$ of X and Y is defined to be the set of all supertiles obtainable by placing X and Y adjacent to each other (without overlapping) such that, if we list each newly coincident edge e_i with edge strength s_i, then $\sum s_i \geq \tau$.

We define the assembly process in terms of bins. Intuitively, a bin consists of an initial collection of supertiles that self-assemble at temperature τ to produce a new set of supertiles P. Formally, with respect to a given set of tile-types T, a *bin* is a pair (S, τ) where S is a set of initial supertiles whose tile-types are contained in T, and τ is a temperature parameter. For a bin (S, τ), the set of *produced* supertiles $P'_{(S,\tau)}$ is defined recursively as follows: (1) $S \subseteq P'_{(S,\tau)}$ and (2) for any $X, Y \in P'_{(S,\tau)}$, $C^{\tau}_{(X,Y)} \subseteq P'_{(S,\tau)}$. The set of *terminally* produced supertiles of a bin (S, τ) is $P_{(S,\tau)} = \{X \in P' \mid Y \in P', C^{\tau}_{(X,Y)} = \emptyset\}$. We say the set of supertiles P is *uniquely* produced by bin (S, τ) if each supertile in P' is of finite size. Put another way, unique production implies that every producible supertile can grow into a supertile in P.

Intuitively, P' represents the set of all possible supertiles that can self-assemble from the initial set S, whereas P represents only the set of supertiles that cannot grow any

[2] With a typical implementation in DNA, glues actually attach to unique complements rather than to themselves. However, this depiction of the glue function is standard in the literature and does not affect the power of the model.

further. In the case of unique assembly of P, the latter thus represents the eventual, final state of the self-assembly bin. Our goal is therefore to produce bins that yield desired supertiles in the uniquely produced set P.

Given a collection of bins, we model the process of mixing bins together in arbitrarily specified patterns in a sequence of distinct stages. In particular, we permit the following actions: We can *create* a bin of a single tile type $t \in T$, we can *merge* multiple bins together into a single bin, and we can *split* the contents of a given bin into multiple new bins. In particular, when splitting the contents of a bin, we assume the ability to extract only the unique terminally produced set of supertiles P, while filtering out additional partial assemblies in P'. Intuitively, given enough time for assembly and a large enough volume of tiles, a bin that uniquely produces P should consist of almost entirely the terminally produced set P. We formally model the concept of mixing bins in a sequence of stages with the *mix graph*.

Mix graphs. An *r-stage b-bin mix graph* M consists of $rb+1$ vertices, m_* and $m_{i,j}$ for $1 \leq i \leq r$ and $1 \leq j \leq b$, and an arbitrary collection of edges of the form $(m_{r,j}, m_*)$ or $(m_{i,j}, m_{i+1,k})$ for some i, j, k.

Staged assembly systems. A *staged assembly system* is a 3-tuple $\langle M_{r,b}, \{T_{i,j}\}, \{\tau_{i,j}\}\rangle$ where $M_{r,b}$ is an r-stage b-bin mix graph, each $T_{i,j}$ is a set of tile types, and each $\tau_{i,j}$ 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 $(R_{i,j}, \tau_{i,j})$ where $R_{i,j}$ is defined as follows:

1. $R_{1,j} = T_{1,j}$ (this is a bin in the first stage);
2. For $i \geq 2$, $R_{i,j} = \left(\bigcup_{k:\ (m_{i-1,k}, m_{i,j}) \in M_{r,b}} P_{(R_{(i-1,k)}, \tau_{i-1,k})} \right) \cup T_{i,j}$.
3. $R_* = \bigcup_{k:\ (m_{r,k}, m_*) \in M_{r,b}} P_{(R_{(r,k)}, \tau_{r,k})}$.

Thus, the jth bin in the ith stage takes its initial set of seed supertiles to be the terminally produced supertiles from a collection of bins from the previous stage, the exact collection specified by $M_{r,b}$, in addition to a set of added tile types $T_{i,j}$. Intuitively, the mix graph specifies how each collection of bins should be mixed together when transitioning from one stage to the next. We define the set of terminally produced supertiles for a staged assembly system to be $P_{(R_*, \tau_*)}$. In this paper, we are interested in staged assembly systems for which each bin yields unique assembly of terminal supertiles. In this case we say a staged assembly system uniquely produces the set of supertiles $P_{(R_*, \tau_*)}$.

Throughout this paper, we assume that, for all i, j, $\tau_{i,j} = \tau$ for some fixed global temperature τ, and we denote a staged assembly system as $\langle M_{r,b}, \{T_{i,j}\}, \tau\rangle$.

3 Assembly of $1 \times n$ Lines

As a warmup, we develop a staged assembly for the $1 \times n$ rectangle ("line") using only three glues and $O(\log n)$ stages. The assembly uses a divide-and-conquer approach to split the shape into a constant number of recursive pieces. Before we turn to the simple divide-and-conquer required here, we describe the general case, which will be useful later. This approach requires the pieces to be combinable in a unique way, forcing the creation of the desired shape. We consider the *decomposition tree* formed by

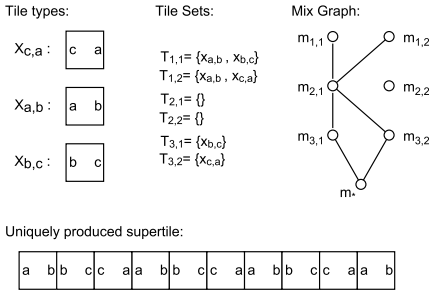

Fig. 1. A sample staged assembly system that uniquely assembles a 1×10 line. The temperature is $\tau = 1$, and each glue a, b, c has strength 1. The tile complexity is 3, the stage complexity is 3, and the bin complexity is 2.

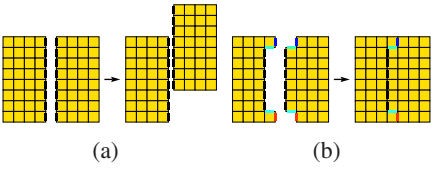

Fig. 2. (a) The shifting problem encountered when combining rectangle supertiles. (b) The jigsaw solution: two supertiles that combine uniquely into a fully connected square supertile.

the recursion, where sibling nodes should uniquely assemble to their parent. The staging proceeds bottom-up in this tree. The height of this tree corresponds to the stage complexity, and the maximum number of distinct nodes at any level corresponds to the bin complexity. The idea is to assign glues to the pieces in the decomposition tree to guarantee unique assemblage while using few glues.

Theorem 1. *There is a planar temperature-*1 *staged assembly system that uniquely produces a (fully connected)* 1×2^k *line using* 3 *glues,* 6 *tiles,* 6 *bins, and* $O(k)$ *stages.*

Proof. The decomposition tree simply splits a 1×2^k line into two $1 \times 2^{k-1}$ lines. All tiles have the null glue on their top and bottom edges. If the 1×2^k line has glue a on its left edge, and glue b on its right edge, then the left and right $1 \times 2^{k-1}$ inherit these glues on their left and right edges, respectively. We label the remaining two inner edges—the right edge of the left piece and the left edge of the right piece—with a third glue c, distinct from a and b. Because $a \neq b$, the left and right piece uniquely attach at the inner edges with common glue c. This recursion also maintains the invariant that $a \neq b$, so three glues suffice overall. Thus there are only $\binom{3}{2} = 6$ possible 1×2^k lines of interest, and we only need to store these six at any time, using six bins. At the base case of $k = 0$, we just create the nine possible single tiles. The number of stages beyond that creation is exactly k.

Corollary 1. *There is a planar temperature-*1 *staged assembly system that uniquely produces a (fully connected)* $1 \times n$ *line using* 3 *glues,* 6 *tiles,* 10 *bins, and* $O(\log n)$ *stages.*

Proof. We augment the construction of Theorem 1 applied to $k = \lfloor \log n \rfloor$. When we build the 1×2^i lines for some i, if the binary representation of n has a 1 bit in the ith position, then we add that line to a new output bin. Thus, in the output bin, we accumulate powers of 2 that sum to n. As in the proof of Theorem 1, three glues suffice to guarantee unique assemblage in the output bin. The number of stages remains $O(\log n)$.

4 Assembly of $n \times n$ Squares

Figure 2(a) illustrates the challenge with generalizing the decomposition-tree technique from $1 \times n$ lines to $n \times n$ squares. Namely, the naïve decomposition of a square into two $n \times n/2$ rectangles cannot lead to a unique assembly using $O(1)$ glues with temperature 1 and full connectivity: by the pigeon-hole principle, some glue must be used more than once along the shared side of length n, and the lower instance of the left piece may glue to the higher instance of the right piece. Even though this incorrect alignment may make two unequal glues adjacent, in the temperature-1 model, a single matching pair of glues is enough for a possible assembly.

4.1 Jigsaw Technique

To overcome this shifting problem, we introduce the *jigsaw technique*, a powerful tool used throughout this paper. This technique ensures that the two supertiles glue together uniquely based on geometry instead of glues. Figure 2(b) shows how to cut a square supertile into two supertiles with three different glues that force unique combination while preserving full connectivity.

Theorem 2. *There is a planar temperature-*1 *staged assembly of a fully connected* $n \times n$ *square using* 9 *glues,* $O(1)$ *tiles,* $O(1)$ *bins, and* $O(\log n)$ *stages.*

Proof. We build a decomposition tree by first decomposing the $n \times n$ square by vertical cuts, until we obtain tall, thin supertiles; then we similarly decompose these tall, thin supertiles by horizontal cuts, until we obtain constant-size supertiles. Table 2 describes the general algorithm. Figure 3 shows the decomposition tree for an 8×8 square. The height of the decomposition tree, and hence the stage complexity, is $O(\log n)$.

We assign glue types to the boundaries of the supertiles to guarantee unique assemblage based on the jigsaw technique. The assignment algorithm is similar to the $1 \times n$ line, but we use three glues on each edge instead of one, for a total of nine glues instead of three.

It remains to show that the bin complexity is $O(1)$. We start by considering the vertical decomposition. At each level of the decomposition tree, there are three types of intermediate products: leftmost supertile, rightmost supertile and middle supertiles. The leftmost and rightmost supertiles are always in different bins. The important thing

Table 2. Algorithm for vertical decomposition. (Horizontal decomposition is symmetric.).

Algorithm DecomposeVertically (supertile S):
— Here S is a supertile with n rows and m columns; S is not necessarily a rectangle.

1. **Stop vertical partitioning when width is small enough:**
   If $m \leq 3$, DecomposeHorizontally(S) and return.
2. **Find the column along which the supertile is to be partitioned:**
   Let $i := \lfloor (m+1)/2 \rfloor$.
   Divide supertile S along the ith column into a left supertile S_1 and right supertile S_2 such that tiles at position $(1, i)$ and (n, i) belong to S_1 and the rest of the ith column belongs to S_2.
3. **Now decompose recursively:**
   DecomposeVertically (S_1)
   DecomposeVertically (S_2)

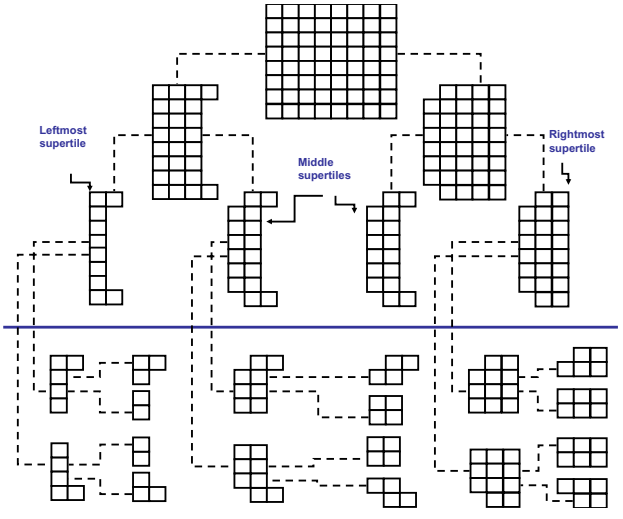

Fig. 3. Decomposition tree for 8×8 square in the jigsaw technique

to observe is that the middle supertiles always have the same shape, though it is possible to have two different sizes—the number of columns can differ by one. In one of these sizes, the number of columns is even and, in the other, the number is odd. Thus we need to separate bins for the even- and odd-columned middle supertiles. For each of the even- or odd-columned supertiles, each of left and right boundaries of the supertile can have three choices for the glue types. Therefore, there is a constant number of different types of middle supertiles at each level of the decomposition tree. Thus, for vertical decomposition, we need $O(1)$ bins. Each of the supertiles at the end of vertical decomposition undergoes horizontal decomposition. A similar argument applies to the horizontal decomposition as well. Therefore, the number of bins required is $O(1)$.

4.2 Crazy Mixing

For each stage of a mix graph on B bins, there are up to $\Theta(B^2)$ edges that can be included in the mix graph. By picking which of these edges are included in each stage, $\Theta(B^2)$ bits of information can be encoded into the mix graph per stage. The large amount of information that can be encoded in the mixing pattern of a stage permits a very efficient trade-off between bin complexity and stage complexity. In this section, we consider the complexity of this trade-off in the context of building $n \times n$ squares.

It is possible to view a tile system as a compressed encoding of the shape it assembles. Thus, information theoretic lower bounds for the descriptional or Kolmogorov complexity of the shape assembled can be applied to aspects of the tile system. From this we obtain the following lower bound:

Theorem 3. *Any staged assembly system with a fixed temperature and bin complexity B that uniquely assembles an $n \times n$ square with $O(1)$ tile complexity must have stage complexity $\Omega(\frac{\log n}{B^2})$ for almost all n.*

Our upper bound achieves a stage complexity that is within a $O(\log B)$ factor of this lower bound:

Theorem 4. *For any n and B, there is a temperature-2 fully connected staged assembly of an $n \times n$ square using 16 glues, $O(1)$ tiles, B bins, and $O(\frac{\log n}{B^2} \log B + \log B)$ stages.*

In the interest of space, the proofs of these two theorems are omitted in this version.

We conjecture that this stage complexity bound can be achieved by a temperature-1 assembly by judicious use of the jigsaw technique.

5 Assembly of General Shapes

In this section, we describe a variety of techniques for manufacturing arbitrary shapes using staged assembly with $O(1)$ glues and tiles.

5.1 Spanning-Tree Technique

The *spanning-tree technique* is a general tool for making an arbitrary shape with the connectivity of a tree. We start with a sequential version of the assembly:

Theorem 5. *Any shape S with n tiles has a partially connected temperature-1 staged assembly using 2 glues, at most 2^4 tiles, $O(\log n)$ bins, and $O(\mathrm{diameter}(S))$ stages.*

Proof. Take a breadth-first spanning tree of the adjacency graph of the shape S. The depth of this tree is $O(\mathrm{diameter}(S))$. Root the tree at an arbitrary leaf. Thus, each vertex in the tree has at most three children. Color the vertices with two colors, black and white, alternating per level. For each edge between a white parent and a black child, we assign a white glue to the corresponding tiles' shared edge. For each edge between a black parent and a white child, we assign a black glue to the corresponding tiles' shared edge. All other tile edges receive the null glue. Now a tile has at most three edges of its color connecting to its children, and at most one edge of the opposite color connecting to its parent.

To obtain the sequential assembly, we perform a particular postorder traversal of the tree: at node v, visit its child subtrees in decreasing order of size. To combine at node v, we mix the recursively computed bins for the child subtrees together with the tile corresponding to node v. The bichromatic labeling ensures unique assemblage. The number of intermediate products we need to store is $O(\log n)$, because when we recurse into a second child, its subtree must have size at most $2/3$ of the parent's subtree.

5.2 Scale Factor 2

Although the spanning-tree technique is general, it probably manufactures structurally unsound assemblies. Next we show how to obtain full connectivity of general shapes, while still using only a constant number of glues and tiles.

Theorem 6. *Any simply connected shape has a staged assembly using a scale factor of 2, 8 glues, $O(1)$ tiles, $O(n)$ stages, and $O(n)$ bins. The construction maintains full connectivity.*

Proof. Slice the target shape with horizontal lines to divide the shape into $1 \times k$ strips for various values of k, which scale to $2 \times 2k$ strips. These strips can overlap along horizontal edges but not vertical edges. Define the *strip graph* to have a vertex for each strip and an edge between two strips that overlap along a horizontal edge. Because the shape is simply connected, the strip graph is a tree. Root this tree at an arbitrary strip, defining a parent relation.

A recursive algorithm builds the subtree of the strip graph rooted at an arbitrary strip s. As shown in Figure 4(a), the strip s may attach to the rest of the shape at zero or more places on its top or bottom edge. One of these connections corresponds to the parent of s (unless s is the overall root). As shown in Figure 4(b), our goal is to form each of these attachments using a jigsaw tab/pocket combination, where bottom edges have tabs and top edges have pockets, extending from the rightmost square up to but not including the leftmost square.

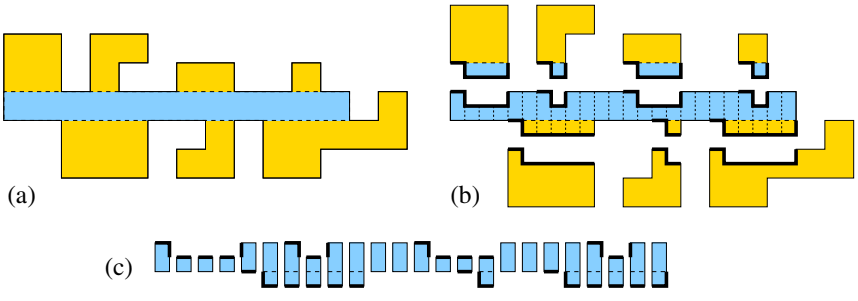

Fig. 4. Constructing a horizontal strip in a factor-2 scaled shape (a), augmented by jigsaw tabs and pockets to attach to adjacent pieces (b), proceeding column-by-column (c)

The horizontal edges of each tab or pocket uses a pair of glues. The unit-length upper horizontal edge uses one glue, and the possibly longer lower horizontal edge uses the other glue. The pockets at the top of strip s use a different glue pair from the tabs at the bottom of strip s. Furthermore, the pocket or tab connecting s to its parent uses a different glue pair from all other pockets and tabs. Thus, there are four different glue pairs (for a total of eight glues). If the depth of s in the rooted tree of the strip graph is even, then we use the first glue pair for the top pockets, the second glue pair for the bottom tabs, except for the connection to the parent which uses either the third or fourth glue pair depending on whether the connection is a top pocket or a bottom tab. If the depth of s is odd, then we reverse the roles of the first two glue pairs with the last two glue pairs. All vertical edges of tabs and pockets use the same glue, 8.

To construct the strip s augmented by tabs and pockets, we proceed sequentially from left to right, as shown in Figure 4(c). The construction uses two bins. At the kth step, the primary bin contains the first $k-1$ columns of the augmented strip. In the secondary bin, we construct the kth column by brute force in one stage using 1–3 tiles and 0–2 distinct internal glues plus the desired glues on the boundary. Because the column specifies only two glues for horizontal edges, at the top and bottom, we can use any two other glues for the internal glues. All of the vertical edges of the column use different glues. If k is odd, the left edges use glues 1–3 and the right edges uses glues 4–6, according to y

coordinate; if k is even, the roles are reversed. (In particular, these glues do not conflict with glue 8 in the tabs and pockets.) The only exception is the first and last columns, which have no glues on their left and right sides, respectively. Now we can add the secondary bin to the primary bin, and the kth column will uniquely attach to the right side of the first $k-1$ columns. In the end, we obtain the augmented strip.

During the building of the strip, we attach children subproblems. Specifically, once we assemble the rightmost column of an attachment to one or two children strips, we recursively assemble those one or two children subtrees in separate bins, and then mix them into s's primary bin. Because the glues on the top and bottom sides of s differ, as do the glues of s's parent, and because of the jigsaw approach, each child we add has a unique place to attach. Therefore we uniquely assemble s's subtree. Applying this construction to the root of the tree, we obtain a unique assembly of the entire shape.

5.3 Simulation of One-Stage Assembly with Logarithmic Scale Factor

In this section, we show how to use a small number of stages to combine a constant number of tile types into a collection of supertiles that can simulate the assembly of an arbitrary set of tiles at temperature $\tau = 1$, given that these tiles only assemble fully connected shapes. In the interest of space, the details of this proof are omitted. Extending this simulation to temperature-2 one-stage systems is an open problem.

Theorem 7. *Consider an arbitrary single stage, single bin tile system with tile set T, all glues of strength at most 1, and that assembles a class of fully connected shapes. There is a temperature-1 staged assembly system that simulates the one-stage assembly of T up to an $O(\log |T|)$ size scale factor using 3 glues, $O(1)$ tiles, $O(|T|)$ bins, and $O(\log \log |T|)$ stages. At the cost of increasing temperature to $\tau = 2$, the construction achieves full connectivity.*

6 Future Directions

There are several open research questions stemming from this work.

One direction is to relax the assumption that, at each stage, all supertiles self-assemble to completion. In practice, it is likely that at least some tiles will fail to reach their terminal assembly before the start of the next stage. Can a staged assembly be robust against such errors, or at least detect these errors by some filtering, or can we bound the error propagation in some probabilistic model.

Another direction is to develop a model of the assembly time required by a mixing operation involving two bins of tiles. Such models exist for (one-stage) *seeded self-assembly*—which starts with a seed tile and places singleton tiles one at a time—but this model fails to capture the more parallel nature of two-handed assembly in which large supertiles can bond together without a seed. Another interesting direction would be to consider nondeterministic assembly in which a tile system is capable of building a large class of distinct shapes. Is it possible to design the system so that certain shapes are assembled with high probability?

Finally, we have focused on two-dimensional constructions in this paper. This focus provides a more direct comparison with previous models, and it is also a case of practical interest, e.g., for manufacturing sieves. Many of our results also generalize to 3D

(or any constant dimension), at the cost of increasing the number of glues and tiles. For example, the spanning-tree model generalizes trivially, and a modification to the jigsaw idea enables many of the other results to carry over. So far, we have not worked out the exact performance measures for these 3D analogs, but we do not expect this to be difficult.

Acknowledgments. We thank M. S. AtKisson and Edward Goldberg for extensive discussions about the bioengineering application.

References

[ACG+02] Adleman, L., Cheng, Q., Goel, A., Huang, M.-D., Kempe, D., de Espanés, P.M., Rothemund, P.W.K.: Combinatorial optimization problems in self-assembly. In: Proceedings of the Thirty-Fourth Annual ACM Symposium on Theory of Computing (electronic), pp. 23–32. ACM Press, New York (2002)

[ACG+05] Aggarwal, G., Cheng, Q., Goldwasser, M.H., Kao, M.-Y., de Espanes, P.M., Schweller, R.T.: Complexities for generalized models of self-assembly. SIAM Journal on Computing 34(6), 1493–1515 (2005)

[ACGH01] Adleman, L., Cheng, Q., Goel, A., Huang, M.-D.: Running time and program size for self-assembled squares. In: Proceedings of the 33rd Annual ACM Symposium on Theory of Computing, pp. 740–748. ACM Press, New York (2001)

[Adl00] Adleman, L.M.: Toward a mathematical theory of self-assembly. Technical Report 00-722, Department of Computer Science, University of Southern California (January 2000)

[BRW05] Barish, R.D., Rothemund, P.W.K., Winfree, E.: Two computational primitives for algorithmic self-assembly: Copying and counting. Nano Letters 5(12), 2586–2592 (2005)

[KS06] Kao, M.-Y., Schweller, R.: Reducing tile complexity for self-assembly through temperature programming. In: Proceedings of the 17th Annual ACM-SIAM Symposium on Discrete Algorithm, pp. 571–580. ACM Press, New York (2006)

[MLRS00] Mao, C., LaBean, T.H., Reif, J.H., Seeman, N.C.: Logical computation using algorithmic self-assembly of DNA triple-crossover molecules. Nature 407, 493–496 (2000)

[PPA+06] Park, S.H., Pistol, C., Ahn, S.J., Reif, J.H., Lebeck, A.R., Dwyer, C., LaBean, T.H.: Finite-size, fully addressable DNA tile lattices formed by hierarchical assembly procedures. Angewandte Chemie 45, 735–739 (2006)

[Rei99] Reif, J.: Local parallel biomolecular computation. In: Proc. DNA-Based Computers, pp. 217–254 (1999)

[Rot06] Rothemund, P.W.K.: Folding DNA to create nanoscale shapes and patterns. Nature 440, 297–302 (2006)

[RPW04] Rothemund, P.W.K., Papadakis, N., Winfree, E.: Algorithmic self-assembly of DNA sierpinski triangles. PLoS Biology 2(12), 424 (2004)

[RW00] Rothemund, P.W.K., Winfree, E.: The program-size complexity of self-assembled squares. In: Proceedings of the 32nd Annual ACM Symposium on Theory of Computing, pp. 459–468. ACM Press, New York (2000)

[See98] Seeman, N.C.: DNA nanotechnology. In: Siegel, R.W., Hu, E., Roco, M.C. (eds.) WTEC Workshop Report on R&D Status and Trends in Nanoparticles, Nanostructured Materials, and Nanodevices in the United States (January 1998)

[SKFM05] Somei, K., Kaneda, S., Fujii, T., Murata, S.: A microfluidic device for dna tile self-assembly. In: DNA, pp. 325–335 (2005)

[SQJ04] Shih, W.M., Quispe, J.D., Joyce, G.F.: A 1.7-kilobase single-stranded DNA that folds into a nanoscale octahedron. Nature 427, 618–621 (2004)

[SW04] Soloveichik, D., Winfree, E.: Complexity of self-assembled shapes. In: Ferretti, C., Mauri, G., Zandron, C. (eds.) DNA Computing. LNCS, vol. 3384, pp. 344–354. Springer, Heidelberg (2005)

[Wan61] Wang, H.: Proving theorems by pattern recognition—II. The Bell System Technical Journal 40(1), 1–41 (1961)

[Win98] Winfree, E.: Algorithmic Self-Assembly of DNA. PhD thesis, California Institute of Technology, Pasadena (1998)

[WLWS98] Winfree, E., Liu, F., Wenzler, L.A., Seeman, N.C.: Design and self-assembly of two-dimensional DNA crystals. Nature 394, 539–544 (1998)