Optimal Self-Assembly of Counters at Temperature Two

Qi Cheng *
University of Oklahoma

Ashish Goel †
Stanford University

Pablo Moisset de Espanés ‡
University of Southern California

Abstract

We present a tile system for assembling a counter of width K and height $N = 2^K$ on top of a seed-row in optimum time $\Theta(N)$. Our counter works at temperature 2. Earlier known constructions either took sub-optimal time or required temperature 3. Using existing techniques, our system for assembling a counter from a seed row can be extended into a system for assembling an $N \times N$ square in optimum time $\Theta(N)$ and with optimum program size $\Theta(\log N / \log \log N)$.

We also present a general technique for analyzing the expected completion time of irreversible self-assemblies.

1 Introduction

Self-assembly is the ubiquitous process by which objects autonomously assemble into complexes. Nature provides many examples: Atoms react to form molecules. Molecules react to form crystals and supramolecules. Cells sometimes coalesce to form organisms. It has been suggested that self-assembly will ultimately become an important technology, enabling the fabrication of great quantities of small complex objects such as computer circuits. DNA self-assembly has received significant attention over the last few years, both by practitioners [13, 14, 11], and by theoreticians [5, 7, 12, 8, 9, 1, 2]. The theoretical results have focused on efficiently assembling structures of a controlled size (the canonical example being assembly of $N \times N$ squares).

*Qi Cheng is at the Department of Computer Science, University of Oklahoma. Research supported by an NSF CAREER award. Email: qcheng@ou.edu.

†Ashish Goel is at the Department of Management Science and Engineering and (by courtesy) Computer Science, Stanford University, Terman 311, Stanford CA 94305. Email: ashishg@stanford.edu. Research supported in part by NSF CAREER Award 0133968 and NSF Award 0323766.

‡Pablo Moisset de Espanés is at the Department of Computer Science, University of Southern California. Email: pmoisset@usc.edu. Research supported in part by a grant from NASA/JPL.

The Tile Assembly Model, originally proposed by Rothemund and Winfree [9], and later extended by Adleman *et al.* [1], provides a useful framework to study the *efficiency* of DNA self-assembly. In this model, a square tile is the basic unit of an assembly. Each tile has a glue on each side; each glue has a label and a strength (typically 1 or 2). A tile can add to a position in an existing assembly if at all the edges where this tile "abuts" the assembly, the glues on the tile and the assembly are the same, and the total strength of these glues is at least equal to a system parameter called the temperature (typically 2). Assembly starts from a single seed crystal and proceeds by repeated accretion of single tiles. The speed of an addition (and hence the time for the entire process to complete) is determined by the concentrations of different tiles in the system. Details are in section 2.

Rothemund and Winfree [9] gave an elegant self-assembling system for constructing squares by self-assembly in this model. Their construction of $N \times N$ squares requires time $\Theta(N \log N)$ and program size $\Theta(\log N)$. Adleman *et al.* [1] presented a new construction for assembling $N \times N$ squares which uses optimal time $\Theta(N)$ and optimal program size $\Theta(\frac{\log N}{\log \log N})$. Both constructions first assemble a roughly $N \times \log N$ rectangle by simulating a binary counter, and then complete the rectangle into a square.

Our Results: We observe that the construction by Rothemund and Winfree [9] works at temperature 2 and is neither optimal with respect to program size nor running time. The running time was determined by the binary counter they used, which performed the increment operations in a sequential fashion. The construction by Adleman *et al.* [1] is optimal in both senses but requires temperature 3. The speedup was due to a counter that uses parallelism. It is easy to combine the techniques used in both constructions to achieve optimal program size at temperature 2, but it is not clear how to make the running time linear in N at the same time. The main difficulty is constructing a binary counter at temperature 2 that has enough parallelism to count from 0 to $N-1$ in $O(N)$ time.

We summarize our results as follows.

1. We show how to assemble a binary counter at temperature 2 in optimal assembly time and program size.
2. We prove our counter is a partial order system, as defined in [2], an interesting class of tile systems.
3. We derive a technique to analize assembly time of self-assembly systems in the Tile Assembly Model and we apply the technique to our problem.

2 Definitions

The Tile Assembly Model: The tile assembly model [9, 1] extends the theoretical model of tiling by Wang [10] to include a mechanism for growth based on the physics of molecular self-assembly. We will present a succinct definition, with minor modifications for ease of explanation.

A tile is an oriented unit square with the north, east, south and west edges labeled from some alphabet Σ of glues. For each tile $t \in T$, the labels of its four edges are denoted $\sigma_N(t)$, $\sigma_E(t)$, $\sigma_S(t)$, and $\sigma_W(t)$. Sometimes we will describe a tile t as the quadruple $(\sigma_N(t), \sigma_E(t), \sigma_S(t), \sigma_W(t))$. Consider the triple $< T, g, \tau >$ where T is a finite set of tiles, $\tau \in \mathbf{Z}_{>0}$ is the *temperature*, and g is the *glue strength* function from $\Sigma \times \Sigma$ to $\mathbf{Z}_{\geq \mathbf{0}}$, where Σ is the set of glues. We assume that $(x \neq y) \Rightarrow g(x, y) = 0$ for all $x, y \in \Sigma$. In [3, 4] the case when the assumption is removed was explored.

Given $p = (x, y), p' = (x', y') \in \mathbf{Z}^2$, we say p and p' are *position adjacent* iff $|x - x'| + |y - y'| = 1$. A *shape* S is a finite subset of $\mathbf{Z}^2$, such that for all pairs of positions (p, p') in S^2, either $p = p'$ or there exists a sequence of positions $p_1, p_2, \cdots, p_n$ such that $p_1 = p$, $p_n = p'$ and for all $1 \leq k < n$, p_k and p_{k+1} are position adjacent. A *configuration* is a partial function from $\mathbf{Z}^2$ to T.

Let $Dom(f)$ denote the domain of a function f. Let C and D be two configurations. Suppose there exist some $t \in T$ and some $(x, y) \in \mathbf{Z}^2$ such that $(x, y) \notin Dom(C)$, $D(x, y) = t$ and $D = C$ except at (x, y). Let $f_{N,C,t}(x, y) = g(\sigma_N(t), \sigma_S(C(x, y+1))$ if $(x, y+1) \in Dom(C)$ and $f_{N,C,t}(x, y) = 0$ otherwise. Informally $f_{N,C,t}(x, y)$ is the strength of the bond between C and the north side of t. Define $f_{S,C,t}(x, y)$, $f_{E,C,t}(x, y)$ and $f_{W,C,t}(x, y)$ similarly. Then we say that tile t is *attachable* to C at position (x, y) iff $f_{N,C,t}(x, y) + f_{S,C,t}(x, y) + f_{E,C,t}(x, y) + f_{W,C,t}(x, y) \geq \tau$, and we write $C \rightarrow_{\mathbf{T}} D$ to denote the transition from C to D in attaching a tile to C at position (x, y). Informally, $C \rightarrow_{\mathbf{T}} D$ iff D can be obtained from C by adding a tile t such that the total strength of interaction between t and C is at least τ.

A *tile system* is a quadruple $\mathbf{T} =< T, s, g, \tau >$, where T, g, τ are as above and s is a special configuration called the "seed", whose domain is a shape. If $Dom(s)$ is a singleton, we will say $\mathbf{T}$ is a *unit seed* tile system. We define the notion of a *derived supertile* of a tile system $\mathbf{T} =< T, s, g, \tau >$ recursively as follows:

1. The seed configuration s is a derived supertile of $\mathbf{T}$, and
2. if $C \rightarrow_{\mathbf{T}} D$ and C is a derived supertile of $\mathbf{T}$, then D is also a derived supertile of $\mathbf{T}$.

Informally, a derived supertile is either just the seed (condition 1 above), or obtained by legal addition of a single tile to another derived supertile (condition 2). We will often omit the word "derived" in the rest of the paper, and use the terms "seed supertile" or just "seed" or s to denote the special supertile in condition 1.

A *terminal supertile* of the tile system $\mathbf{T}$ is a derived supertile A such that there is no supertile B for which $A \rightarrow_{\mathbf{T}} B$. If there is a terminal supertile A such that for any derived supertile B, $B \rightarrow^*_{\mathbf{T}} A$, we say that the tile system *uniquely produces* A. Given a tile system $\mathbf{T}$ which uniquely produces a supertile, we say that the program size complexity of the system is $|T|$ i.e. the number of tile types.

The *shape of a supertile* Γ is $Dom(\Gamma)$. We will use the notation $[\Gamma]$ to denote the shape of Γ. A tile system $\mathbf{T}$ is said to *uniquely produce a shape* W iff it uniquely produces some supertile Γ and $[\Gamma]$ is identical (up to translation) to W.

We will now add the notion of running time to this model. We associate with each tile $t \in T$ a nonnegative probability $P(t)$, such that $\sum_{t \in T} P(t) = 1$. We assume that the tile system has an infinite supply of each tile, and $P(t)$ models the concentration of tile t in the system. Now self-assembly of the tile system corresponds to a continuous time Markov process where the states are in a one-one correspondence with derived supertiles, and the initial state corresponds to the seed s. Suppose a single tile t can be added to supertile B to produce supertile C. Then there is a transition from state B to C in the Markov chain, and the rate of the transition is $P(t)$. Suppose the tile system produces a unique terminal supertile A_T. In the Markov chain, the time for reaching A_T from s is a random variable. The "running time" of the self-assembly process is defined as the expected value of this random variable.

Note that the Markov process modeling the self assembly process is inherently parallel. For details see [1].

3 Counting to N at T=2 in $O(N)$ time.

3.1 The Tile System

For clarity of explanation, we begin by showing a pictorial representation of the tile system and of the supertile it produces for $N = 8$.

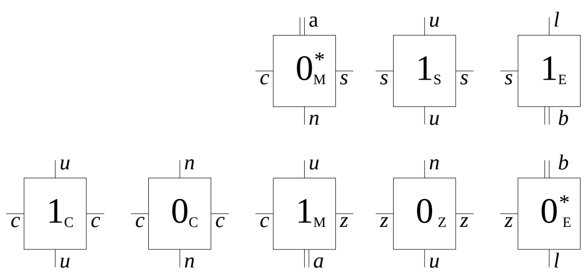

Figure 1: The tiles for the SA-counter. The number of lines jutting from each edge of the tile represent the strength of the bond (1 or 2) and the labels on the edges represents the glues. Some edges do not show any glue, meaning they have a zero strength glue.

We initially assume that $N = 2^K$ for some positive integer K – we will later show how this assumption can be removed. The fully assembled counter is going to be a rectangle, with K tiles in each row. Each row represents a binary number between 0 and $N - 1$, with the row representing i being immediately above the row representing $i - 1$. While each tile is completely specified by its four glues, it is convenient for the purpose of exposition to allow tiles to have labels. Each tile has one main label (either 0 or 1). We will refer to tiles representing the digits 0 and 1 as "0-tiles" and

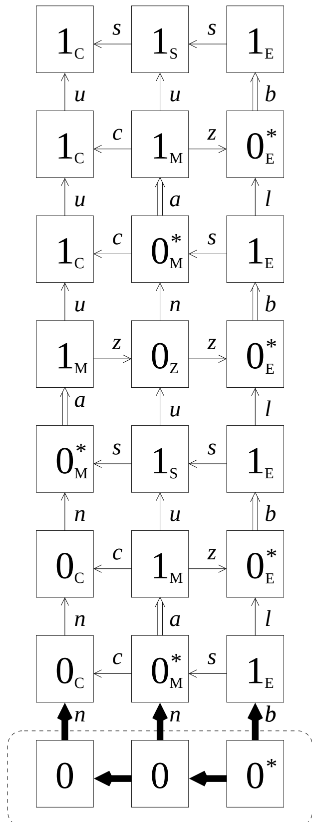

Figure 2: Graph representing the SA-counter supertile. The dashed line encloses the seed row. The thin arrows indicate the order relation. The thick arrows represent the edges added to construct an Equivalent Acyclic graph.

”1-tiles” respectively. The number represented by each row can be obtained by reading the main labels on the tiles in that row, with the most significant bit being the west-most in the row. We will refer to the row representing the number i as the i^{th} row of the counter. We will now describe the self-assembly process which results in the counter being incremented.

The increment operation going from row to row uses 8 different tiles as well as K special tiles that appear in the seed row only (see Figure 1). For all positive i less than N, the assembly of the i^{th} row always starts as the attachment of a 1-tile on top of the east-most 0-tile in the $(i-1)^{st}$ row. Call p_i the position of the first tile being attached in the i^{th} row. All the remaining tiles in the i^{th} row to the east of p_i will be 0-tiles. All the remaining tiles in the i^{th} row to the west of p_i will ”copy” the labels of the tiles immediately below, i.e. if p is to the west of p_i, then a 1-tile will attach to the growing supertile at p iff there is a 1-tile at $p-(0,1)$.

We will assume that the temperature for this counter construction is 2. We now describe the tiles in each of the rows and their glues. In addition to the main label, a 0-tile may have a star symbol. 0^*-tiles denote the east-most 0-tile in each row.

0-tiles: There are the following 0-tiles: 0^*_{E}, 0^*_{M}, 0_{Z}, 0_{C}.

1-tiles: 1_{E}, 1_{M}, 1_{S}, 1_{C}.

The glues and their strengths are depicted pictorially in Figure 1. The label suffixes indicate the position in the row each row a tile may occupy and/or the function in the assembly.

E: The tile can be only in the east-most column of the counter. Only tiles with the E suffix can be in the east-most column of the counter.

M: The tile is either a 0^*-tile or a 1-tile that is directly on top of a 0^*-tile.

Z: Attach a 0-tile to the east. Tiles with the Z suffix can appear only to the east of the 0^*-tile in the previous row and to the west of the 0^*-tile in the same row.

C: Copy to the west the digit in the lower row. All tiles to the west of an M tile in the same row are C tiles. Every C tile has the same main label as the tile immediately to its south.

S: Search in the east to west direction for the position of the first 0-tile in the row immediately below. Tiles with the S suffix can appear only to the east of 0^*-tiles.

We define $\mathbf{T}_{SA}(K)$ as the tile system that has a tile set consisting of the tiles in Figure 1. The glue strength function is as indicated in the same figure. The temperature is 2, and there is a single seed supertile called S_K. S_K is a row of K special tiles. The glues on the north sides of these K tiles represent the initial value of the counter in binary. To represent the number 0, the glue on the north side of the east-most tile in S_K is b, and all the other glues on the north side of S_K are N. We will assume that all tiles which are not in S_K have the same constant probability p_{SA} which is a constant independent of N.

We reproduce a definition from [8] with minor modifications in the terminology to make it consistent with the definitions in this paper.

Deterministic row-column (RC) property: Let $\mathbf{T} = \langle T, S, g, \tau \rangle$ be a unit-seed tile system and let Γ be a terminal supertile of **T**. Let $w_\Gamma = \min\{x | \exists y \text{ s.t. } (x, y) \in [\Gamma]\}$, i.e. the x coordinate of the west-most tile in Γ. Define e_Γ, n_Γ and s_Γ similarly.

We say Γ is a *full assembly* iff for all pairs (p_1, p_2) of positions in $[\Gamma]^2$, if p_1 and p_2 are position adjacent, then the tiles at p_1 and p_2 in Γ have the same glue on their abutting edges and the strength of that glue is non-zero. We say Γ has the *C-property* iff Γ is a full assembly and for all x s.t. $w_\Gamma \leq x < e_\Gamma$ there exists exactly one y such that the strength of the glue between $\Gamma(x, y)$ and $\Gamma(x + 1, y)$ is greater or equal than τ. Similarly, we say Γ has the *R-property* iff Γ is a full assembly and for all y s.t. $s_\Gamma \leq y < n_\Gamma$ there exists exactly one x such that the strength of the glue between $\Gamma(x, y)$ and $\Gamma(x, y + 1)$ is greater or equal than τ. We say Γ has the *RC-property* iff Γ has the R-property and the C-property. We say an attachment of a tile t to a supertile X at a position p is *deterministic* iff t is the only attachable tile to X at p. We say Γ has the *deterministic RC-property* iff Γ has the RC-property and there exists a derivation of Γ such that all attachments in that derivation are deterministic.

The following theorem, also from [8], will help us to prove correctness of the counter.

Theorem 3.1 *If a unit-seed tile system* $\mathbf{T}$ *produces a supertile* Γ *that has the deterministic RC-property, then* $\mathbf{T}$ *uniquely produces* Γ.

Theorem 3.2 *The tile system* $\mathbf{T}_{SA}(K)$ *uniquely produces a supertile which is a* $2^K \times K$ *rectangle.*

Proof: We start from defining a tile system $\mathbf{T}'_{SA}(K)$ by modifying the glues on some tiles in $\mathbf{T}_{SA}(K)$. Consider a unit-seed tile system $\mathbf{T}'_{SA}(K)$ that results from modifying the K special tiles so $\mathbf{T}'$ can assemble S_K starting from a single seed tile at the east-most position *without* attaching tiles that do not belong to the 0^{th} row. This can be easily accomplished by making all east-west bonds in the seed row of strength 2, by using unique glues to form those bonds and by using the east-most tile in S_K as seed in $\mathbf{T}'_{SA}(K)$.

Consider an arbitrary derivation of a supertile of $\mathbf{T}_{SA}(K)$. The derivation can be thought of as a sequence of pairs $(p_1, t_1), (p_2, t_2), \cdots, (p_N, t_N)$, where the tile t_i attaches at position p_i at the i^{th} step. It is clear that the derivation $(\tilde{p}_1, \tilde{t}_1)$, $(\tilde{p}_2, \tilde{t}_2), \cdots,$ $(\tilde{p}_{K-1}, \tilde{t}_{K-1})$, (p_1, t_1), $(p_2, t_2), \cdots, (p_N, t_N)$, were the $\tilde{p_i}$'s are positions in the seed and the $\tilde{t_i}$'s are modified special tiles, is the derivation of a supertile of $\mathbf{T}'_{SA}(K)$. Intuitively, we are saying that after assembling the 0^{th} row, $\mathbf{T}'_{SA}(K)$ behaves exactly like $\mathbf{T}_{SA}(K)$.

Clearly, $\mathbf{T}'_{SA}(K)$ assembles an $N \times K$ rectangle because there exists a derivation that assembles the terminal supertile row by row, i.e. no tile is attached in the i^{th} row before the $(i - 1)^t h$ row is totally assembled. Further, observe that in that derivation all attachments are deterministic. Also observe that the all the east-west bonds in the 0^{th} row use strength 2 glues, and no strength 2 east-west bonds are present other than those in the 0^{th} row, so the assembly has the C-property. Note that the only north-south strength-2 bonds occur between the 0^*-tile in row i and a 1-tile in the $(i + 1)^{st}$ row, and there is exactly one 0^*-tile per row (except the topmost row) so the assembly has the R-property. Therefore, Γ has the deterministic RC property, and by Theorem 3.1, $\mathbf{T}'_{SA}(K)$ uniquely produces a $N \times K$ rectangle.

■

Minor modification of the seed row results in a tile system that uniquely produces a supertile which is a $N \times K$ rectangle for all positive $N \leq 2^K$.

3.2 Analysis

Our counter construction follows the line of the counter by Adleman *et al.* [1], exploiting "parallelism" to speed up the assembly process. The construction of Rothemund and Winfree [9] takes time $O(N \log N)$ in the model of running time described in Section 2. We show that the completion time for the SA-counter is linear in N in Section 4. Intuitively, the stronger $\Theta(N)$ bound on the assembly time is due to the fact that a row may start getting assembled before the row immediately below is completed.

In spite of having many possible derivations of the rectangle from the seed row, this particular tile system has a property that makes the analysis of the running time easier. We reproduce the definition of *partial order systems* introduced in [2], with minor adaptations.

Assume that we are given a tile system $\mathbf{T}$ and a shape S that is uniquely produced by $\mathbf{T}$. Let A denote the supertile of $\mathbf{T}$ that has shape S. Let $A(i, j)$ represent the tile at position (i, j). Consider a derivation of A and let $t_{i,j}$ represent the time when a tile attaches to the growing assembly at position (i, j). Define a partial order $\prec$ on the tile positions in S such that $(i, j) \prec (p, q)$ iff $t_{i,j} \leq t_{p,q}$ for all possible derivations of A using $\mathbf{T}$.

Definition: A tile system $\mathbf{T}$ is said to be a *partial order system* iff it uniquely produces a shape S, and if for all adjacent positions (i, j),(x, y) in S, either $(i, j) \prec (x, y)$, or $(x, y) \prec (i, j)$, or the strength of the glues connecting tiles at positions (i, j) and (x, y) is zero.

Note that we can represent the partial order relation as a DAG $G = (S, E)$, where $(p, q) \in E$ iff $p \prec q$.

Figure 2 depicts the assembly of a counter counting from 0 to 7 in binary. The thin arrows indicate the order relation in the process. For a arrow, the position at the tail must be filled before the position at the head. Intuitively, a long path in the graph suggests long running time because it implies those positions will be filled sequentially.

Lemma 3.3 *The tile system* $\mathbf{T}_{SA}(K)$ *is a partial order system.*

Proof: For all rows (other than the seed row), the first tile to be added to that row is the one directly above the 0^*-tile in the immediately lower row. This is because all tiles with a strength 2 glue on their north edges are 0^*-tiles. Since there is exactly one 0^*-tile per row, we observe that no tile can attach to the assembly if the tile immediately to the south is not already in place. This gives the north to south order relation for all pairs of positions of the form $(p, p + (0, 1))$. Inside each row, we note that for all positions p to the west of the 0^*-tile, an attachment is impossible if the position immediately to the east of p is empty. Similarly, for all positions p to the east of the 0^*-tile, an attachment is impossible if the position immediately to the east of p is empty. This completes the definition of the partial order relation for all pairs of adjacent positions. ■

Let G_N be the DAG representing the partial order relation in an SA-counter supertile that corresponds to counting from 0 to $N - 1$. Figure 2 shows an example of a counter counting from 0 to 7 and a pictorial representation of the partial order relation that corresponds to G_8.

Lemma 3.4 *For all N in $\mathbf{N}$, the length of the longest directed path in G_N is $\Theta(N)$.*

Proof: Let L be length of the longest path in G_N. The longest path in a DAG must end in a sink, and G_N has exactly one, namely the vertex that corresponds to the north-west corner tile in the rectangle. It is easy to see that longest path must start from the vertex corresponding to the south-east corner of the rectangle. Since the longest path will necessarily contain exactly $N-1$ south-north edges, that path will be the one containing the most west-east edges. There is exactly one path containing all of them. It is the path containing all the vertices corresponding to the 0^*-tiles. All paths going from the bottom row to the top row will traverse exactly $N-1$ north to south edges, hence $L = \Omega(N)$. The number of west to east and east to west edges traversed is easily bounded by adding the distance from the east-most zero to the east-most column for all rows.

$L \leq N - 1 + \sum_{j=1}^{K}(K-j) \times 2^j$

and

$N - 1 + \sum_{j=1}^{K}(K-j) \times 2^j = N - 1 + N\sum_{i=0}^{K-1} i2^{K-i} < N(1 + \sum_{i=0}^{\infty} i2^{-i}) = O(N)$

■

Theorem 3.5 *The time complexity for building an SA-counter that counts from 0 to N is $\Theta(N)$.*

We will present the proof of Theorem 3.5 at the end of Section 4 after we develop some more machinery.

4 Running time Analysis.

For the purpose of our analysis, we transform the self-assembly process into another process which we call the "sentinel" process. The sentinel process does not adhere to the model described in Section 2; in fact there does not seem to be an easy implementation of the sentinel process. However, the sentinel process is more amenable to analysis, and the time for this process to complete is an upper bound (in the stochastic domination sense) on the completion time for the self-assembly process.

Given a directed graph $G = (V, E)$ and vertices v, v' in V, we say v is a *predecessor* of v' iff $(v, v') \in E$. Given a vertex v, we define the *predecessors of v in G*, and we denote it as $Pred_G(v)$, to be the set of all vertices v' such that v' is a predecessor of v.

Equivalent Acyclic Graph (EAG): Let $\mathbf{T} = \langle T, S, g, \tau \rangle$ be a tile system that uniquely produces some supertile Γ, and let $G = (V, E)$ be a DAG. G is an *Equivalent Acyclic Graph (EAG) for* $\mathbf{T}$ iff:

1. $V = [\Gamma]$.
2. G has exactly one source.
3. For all edges $(p, p') \in E$, p and p' are position-adjacent.
4. Let σ be the seed supertile. For all $p \in [\Gamma] - [\sigma]$, for all derived supertiles Γ' of **T**, if $Pred_G(p) \subset [\Gamma]$ and $p \notin [\Gamma']$, then $\Gamma(p)$ is attachable to Γ' at p.

Note that for a given **T** there may exist more than one EAG.

Lemma 4.1 *For all tile systems* **T** *that uniquely produce a supertile, there exists an EAG.*

Proof: Let Γ be the terminal supertile produced by **T**, let σ be the seed supertile and let n be the number of elements of $[\Gamma] - [\sigma]$. Consider one arbitrary derivation of Γ and the graph $G = (V, E)$ constructed in the following way: For $1 \leq i \leq n$ call p_i the position where the i^{th} attachment occurs. Define $V = [\Gamma]$, $E_1 = \{(p_i, p_j)|i < j$ and p_i, p_j are position adjacent$\}$ and $E_b = \{(p, p')|p \in [\sigma]$ and $p' \in [\Gamma] - [\sigma]$, and p, p' are position-adjacent$\}$. Define some DAG $G' = ([\sigma], E_s)$ such that $E_b \subset E_s$, G' is connected, for all edges (p, p') in E_s p and p' are position adjacent, and G' has a single source. It is easy to see that G' must exist, since σ is a connected configuration.

To complete the proof, we have to verify that G is an EAG. We start by proving G is a DAG. Because of the way we constructed it, there cannot be a cycle in G'. We observe that all edges in E_1 connect vertices in $[\Gamma] - [\sigma]$, that all edges in E_b go from a vertex in $[\sigma]$ to a vertex in $[\Gamma] - [\sigma]$ and that all edges in E_s connect elements in $[\sigma]$. Therefore, there cannot be a path from a vertex in $[\Gamma] - [\sigma]$ to $[\sigma]$. There are not cycles in G', so if there is a cycle in G it must contain exclusively elements in $[\Gamma] - [\sigma]$. By the definition of E_1 this is clearly impossible, hence G is a DAG. Now we prove that G has exactly one source, i.e. the one in $[\sigma]$. We claim that there cannot be a source in $[\Gamma] - [\sigma]$: Assume there is one and let p_i be that vertex. If there is a vertex $p \in [\sigma]$ such that p_i and p are position-adjacent then, by definition of E_b, p_i has an incoming edge and is not a source. Now consider the case when there is not such a p. Because of the definition of E_1 we know that for all all outgoing edges (p_i, p_j), $i < j$. Therefore, in our derivation the attachment at i happened before the attachment at j. If p_i is a source, it means that the attachment at p_i happened before any of its position-adjacent locations were filled. No derivation can exhibit that behavior.

To verify the last condition of the definition of EAG, we simply observe that E was obtained from an actual derivation. ■

In the previous proof we used a derivation of the supertile to construct a suitable EAG. We say that the EAG was *induced* by the derivation.

Sentinel process: Let $\mathbf{T} = \langle T, S, g, \tau \rangle$ be a tile system that uniquely produces a supertile, let Γ be that supertile, let P be a concentrations function for **T**, let G be a EAG for **T** and let g_s be the source of G. Let S be the set of all subgraphs $G' = (V', E')$ of G such that $g_s \in V'$ and for all edges $(p, p') \in E$, $(p, p') \in E'$ iff $p, p' \in V'$.

We define $S_{\mathbf{T},G,P}$ to be the continuous time Markov process such that:

1. Each state in $S_{\mathbf{T},G,P}$ corresponds to a graph in S.
2. There is a transition from state s' to s'' in $S_{\mathbf{T},G,P}$ iff the corresponding subgraphs $G' = (V', E')$ and $G'' = (V'', E'')$ are such that $V'' - V'$ is a singleton and $E'' - E' = \{(u, v) | v \in V'' - V$ and $u \in Pred_G(v)\}$.
3. The probability associated with a transition from s to s' is $P(v)$, where v is the only element of $V'' - V$.

Note that $S_{\mathbf{T},G,P}$ has exactly one source state, corresponding to the source of G, and exactly one sink state, corresponding to G. We define the completion time for the sentinel process as the time to go from the source state to the sink state. Intuitively, the sentinel process is obtained by modifying the self-assembly process by deleting some transitions in it, i.e. *disallowing* some bonds to form. This elimination of transitions will make the completion time of the sentinel process an upper bound for the running time of the self-assembly process.

Completion time for the sentinel process: Consider a tile system **T** that uniquely produces a supertile, and a concentrations function P for **T**, and let $c = min_i(P(i))$, i.e. the smallest concentration. Let G be an EAG for **T** and let L be the length of the longest directed path in G. Let t be the completion time for $S_{\mathbf{T},G,P}$.

Lemma 4.2 $\mathbf{E}[t] = O(L/c)$. *Further, t has an exponentially decaying tail.*

Proof: Let $P_1, P_2, \ldots P_N$ represent the N directed paths from the seed to any sink in the sentinel graph. At each step in any path, the edge must go to a position-adjacent vertex and there are at most three candidates. Therefore N is at most $3^L \leq e^{2L}$. Let $X_{i,j}$ be an exponential random variable with mean $1/P(\Gamma(i, j))$, i.e. the reciprocal of the probability associated with the tile at position (i, j) in the final supertile. Let all the $X_{i,j}$ be independent. A tile attaches at position (i, j) in the self-assembly $X_{i,j}$ time after this position becomes attachable[1]. Let S_l denote the sum of all $X_{i,j}$ such that position (i, j) lies on path P_l. Then the completion time $t = \max_{l=1}^{N} S_l$. S_l is the sum of at most L mutually independent exponential variables, each with mean less or equal to $1/c$. Hence $\mathbf{E}[S_l] \leq L/c$; let ϕ denote the value L/c. Clearly $\phi = O(L/c)$. Using Chernoff bounds for exponential variables [6], it follows that $\mathbf{Pr}[S_l > \phi \cdot (1 + \delta)] \leq ((1 + \delta)/e^{\delta})^L$. Hence $\mathbf{Pr}[t > \phi(1 + \delta)] \leq N((1 + \delta)/e^{\delta})^L \leq e^{2L}((1 + \delta)/e^{\delta})^L = ((1 + \delta)/e^{\delta-2})^L$. Let us choose $\delta = \delta' + 4$, where $\delta' > 0$. Now

$$\mathbf{Pr}[t > 5\phi(1 + \delta')]$$

[1]This fact follows from the fact that once a tile position becomes attachable it remains attachable till it actually attaches.

$$\leq \textbf{Pr}\left[t > \phi(1 + 4 + \delta')\right]$$
$$\leq ((1 + \delta')/e^{\delta'})^L.$$

This clearly gives an exponential tail bound. Now,

$$\textbf{E}\left[t\right] \leq 5 \cdot \phi(1 + \int_{\delta'=0}^{\infty} ((1 + \delta')/e^{\delta'})^L d\delta'),$$

which is $O(L/c)$ as $\phi = O(L/c)$ and the integral is bounded by a constant for any value of $L \geq 1$. ■

Stochastic dominance: Define $t_{i,j}$ to be the time at which the (i, j) position gets filled in the original self-assembly process and $t'_{i,j}$ to be the time at which it gets filled in the sentinel process. If (i, j) is in the shape of the seed supertile then assume $t_{i,j} = 0$. Note that $t_{i,j}$ and $t'_{i,j}$ are random variables. Let t and t' be the random variables denoting the times at which the original self-assembly process and the sentinel assembly complete.

A real valued random variable A is said to be stochastically dominated by another random variable B, denoted $A \leq_{sd} B$, if for all x, $\textbf{Pr}\left[A > x\right] \leq \textbf{Pr}\left[B > x\right]$.

Lemma 4.3 *For all (i, j) in the shape of the uniquely produced supertile, $t_{i,j} \leq_{sd} t'_{i,j}$*

Proof: Let Γ be the supertile produced by **T**. Let $X_{i,j}$ be an exponential random variable with mean $1/P(\Gamma(i, j))$, i.e. the reciprocal of the probability associated with the tile at position (i, j) in the final supertile. Let all the $X_{i,j}$ be independent. A tile attaches at position (i, j) in the self-assembly $X_{i,j}$ time after this position becomes attachable. We couple the sentinel process and the self-assembly process by setting the values of $X_{i,j}$ to be the same for both processes. Define $a_{i,j}$ and $a'_{i,j}$ to be the times at which tile position (i, j) is attachable in the self-assembly process and the sentinel process, respectively. Let t be the earliest time when a tile gets attached in the sentinel process but is still unattached in the self-assembly process. Let (i, j) be this tile position. Clearly, $a'_{i,j} < t$. Therefore any tiles which had attached in the sentinel process by time $a'_{i,j}$ had also attached in the self-assembly process. Since the sentinel process was formed by *disallowing* certain bonds in the self-assembly process, tile position (i, j) is also attachable in the self-assembly process at time $a'_{i,j}$. Hence $a_{i,j} \leq a'_{i,j}$. But $t_{i,j} = a_{i,j} + X_{i,j}$ and $t'_{i,j} = a'_{i,j} + X_{i,j}$. This implies that $t_{i,j} \leq t'_{i,j}$, which is a contradiction. Since $t_{i,j} \leq t'_{i,j}$ for each coupled experiment, $t_{i,j} \leq_{sd} t'_{i,j}$. ■

Lemma 4.3 in conjunction with the lemma 4.2 now allows us to conclude:

Theorem 4.4 $\textbf{E}[t] = O(L/c)$. *Further, t has an exponentially decaying tail.*

Consider a unit-seed tile system **T** that uniquely produces a supertile, and a constant valued concentrations function C for **T**. Let c be the value of C.

Theorem 4.5 *There exists an EAG G for* **T** *such that the time complexity of* **T** *is* $\Theta(L/c)$.

Proof: Let Γ be the terminal supertile of **T**, and let t be the time complexity of **T**. Consider an experiment, i.e. the assembly of Γ from the seed. Let $\tilde{t}$ be the completion time of the experiment.

Define $p_1 = \textbf{Pr}\left[\tilde{t} \leq 2t\right]$. Clearly, $\textbf{E}[\tilde{t}] = t$ and by Markov's inequality $p_1 \geq 1/2$.

Let $\tilde{G}$ be the EAG induced by the experiment, that can be constructed as we shown in the proof of Theorem 4.1. Let $\tilde{L}$ be the length of the longest directed path in $\tilde{G}$. Define $p_2 = \textbf{Pr}\left[\tilde{L}/(4c) \geq \tilde{t}\right]$. We will now show that $p_2 < 1/4$.

Define $\beta = x_1 + x_2 + \cdots + x_{\tilde{L}}$ where the $x_i's$ are independent exponential random variables with mean $1/c$. Using the fact that exponential variables are the inter-arrival times of Poisson events we get:

$$\textbf{Pr}\left[\beta \leq \tilde{L}/(4c)\right] = \textbf{Pr}\left[\rho(\tilde{L}/4) \geq \tilde{L}\right]$$

where $\rho(\tilde{L}/4)$ is a Poisson variable with mean $\tilde{L}/4$. Using Markov inequality again, we obtain $\textbf{Pr}\left[\rho(\tilde{L}/4) \geq \tilde{L}\right] \leq 1/4$, hence $\textbf{Pr}\left[\beta \leq \tilde{L}/(4c)\right] \leq 1/4$. Since $\beta \leq_{sd} \tilde{t}$ we conclude that $\textbf{Pr}\left[\tilde{L}/(4c) \geq \tilde{t}\right] \leq 1/4$.

Since $p_1 > p_2$, it must be the case that $\textbf{Pr}\left[\tilde{L}/(4c) \leq \tilde{t} \wedge \tilde{t} \leq 2t\right] \neq 0$. Therefore, there must exist a derivation that induces an EAG G such that $L/(4c) \leq 2t$ and hence $t = \Omega(L/c)$, where L is the length of the longest directed path in G. From Theorem 4.3 we know that $t = O(L/c)$, completing the proof. ■

For completeness, we state the following lemma of trivial proof that gives a lower bound on the time-complexity of a partial order system. Let **T** be a partial-order system, let G be the DAG describing the partial order relation and let L be the length of the longest directed path in G.

Lemma 4.6 *For all concentrations functions P for* **T**, *the time complexity of* **T** *is* $\Omega(L)$

Theorem 3.5 can be obtained immediately from Lemma 3.4, Lemma 4.6, and Theorem 4.5. We present the details below.

Proof of Theorem 3.5: Let G be the acyclic graph representing the partial order relation in the counter. Lemma 3.4 states that the longest path in G has $\Theta(N)$ length. The $\Theta(N)$ running time follows from Theorem 4.5. A suitable EAG is G, augmented with all edges of the form $((x+1, y), (x, y))$, where both $(x+1, y)$ and (x, y) are in the seed row, and with edges of the form $((x, y), (x, y+1))$, where (x, y)'s are in the seed row, (see Figure 2). It's easy to see that the length of the longest directed path in the augmented graph is still $\Theta(N)$. To prove the time complexity is $\Omega(N)$, we simply invoke Lemma 4.6.

References

[1] L. Adleman, Q. Cheng, A. Goel, and M. Huang. Running time and program size for self-assembled squares. In *Proceedings of the thirty-third annual ACM symposium on Theory of computing*, pages 740–748. ACM Press, 2001.

[2] L. Adleman, Q. Cheng, A. Goel, M. Huang, D. Kempe, P. Moisset de Espanes, and P. Rothemund. Combinatorial optimization problems in self-assembly. In *Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*, pages 23–32. ACM Press, 2002.

[3] G. Aggarwal, M. Goldwasser, M. Kao, and R. T. Schweller. Complexities for generalized models of self-assembly. In *Proceedings of symposium on discrete algorithms*. ACM Press, 2004.

[4] Q. Cheng and P. Moisset de Espanes. Resolving two open problems in the self-assembly of squares. Technical Report 03-793, University of Southern California, 2003.

[5] M. Lagoudakis and T. LaBean. 2d dna self-assembly for satisfi ability. In *Proceedings of the 5th DIMACS Workshop on DNA Based Computers in DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, volume 54. MIT: Cambridge, 1999.

[6] R. Motwani and P. Raghavan. Randomized Algorithms. Cambridge University Press, 1995.

[7] J. Reif. Local parallel biomolecular computation. In H. Rubin, editor, *Third Annual DIMACS Workshop on DNA Based Computers, DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, 1998.

[8] P. Rothemund. *Theory and Experiments in Algorithmic Self-Assembly*. PhD thesis, University of Southern California, 2001.

[9] Paul W. K. Rothemund and Erik Winfree. The program-size complexity of self-assembled squares (extended abstract). In *Proceedings of the thirty-second annual ACM symposium on Theory of computing*, pages 459–468. ACM Press, 2000.

[10] H. Wang. Proving theorems by pattern recognition ii. Bell Systems Technical Journal, 1961. 40:1-42.

[11] E. Winfree. *Algorithmic Self-Assembly of DNA*. PhD thesis, California Institute of Technology, Pasadena, 1998.

[12] E. Winfree, F. Liu, L. Wenzler, and N. Seeman. Design and self-assembly of two-dimensional dna crystals, 6 pages. *Nature*, (394):539–544, Aug 1998.

[13] E. Winfree, X. Yang, and N. Seeman. Universal computation via self-assembly of dna: Some theory and experiments. In *Proceedings of the Second Annual Meeting on DNA Based Computers*. Princeton University, June 1996.

[14] B. Yurke, A. Turberfi eld, A. Mills Jr, F. Simmel, and J. Neumann. A dna-fuelled molecular machine made of dna. *Nature*, (406):605–608, Aug 2000.