Published in final edited form as:

Nat Comput. 2012 June 1; 11(2): . doi:10.1007/s11047-012-9313-1.

Self assembly of rectangular shapes on concentration programming and probabilistic tile assembly models

Vamsi Kundeti and Sanguthevar Rajasekaran

Department of Computer Science and Engineering, University of Connecticut, Storrs, CT 06269, USA

Vamsi Kundeti: vamsik@engr.uconn.edu; Sanguthevar Rajasekaran: rajasek@engr.uconn.edu

Abstract

Efficient tile sets for self assembling rectilinear shapes is of critical importance in algorithmic self assembly. A lower bound on the tile complexity of any deterministic self assembly system for an n

$\times n$ square is $\Omega(\frac{\log(n)}{\log(\log(n))})$ (inferred from the Kolmogorov complexity). Deterministic self assembly systems with an optimal tile complexity have been designed for squares and related

shapes in the past. However designing $\Theta(\frac{\log(n)}{\log(\log(n))})$ unique tiles specific to a shape is still an intensive task in the laboratory. On the other hand copies of a tile can be made rapidly using PCR (polymerase chain reaction) experiments. This led to the study of self assembly on *tile concentration programming* models. We present two major results in this paper on the concentration programming model. First we show how to self assemble rectangles with a fixed aspect ratio $(\alpha:\beta)$, with high probability, using $\Theta(\alpha + \beta)$ tiles. This result is much stronger than the existing results by Kao et al. (Randomized self-assembly for approximate shapes, LNCS, vol 5125. Springer, Heidelberg, 2008) and Doty (Randomized self-assembly for exact shapes. In: proceedings of the 50th annual IEEE symposium on foundations of computer science (FOCS), IEEE, Atlanta. pp 85–94, 2009)—which can only self assembly squares and rely on tiles which perform binary arithmetic. On the other hand, our result is based on a technique called *staircase sampling*. This technique eliminates the need for sub-tiles which perform binary arithmetic, reduces the constant in the asymptotic bound, and eliminates the need for *approximate frames* (Kao et al. Randomized self-assembly for approximate shapes, LNCS, vol 5125. Springer, Heidelberg, 2008). Our second result applies staircase sampling on the *equimolar concentration programming model* (The tile complexity of linear assemblies. In: proceedings of the 36th international colloquium automata, languages and programming: Part I on ICALP '09, Springer-Verlag, pp 235–253, 2009), to self assemble rectangles (of fixed aspect ratio) with high probability. The tile complexity of our algorithm is $\Theta(\log(n))$ and is optimal on the probabilistic tile assembly model (PTAM)— n being an upper bound on the dimensions of a rectangle.

Keywords

Tile assembly model; Concentration programming; Self assembly algorithms; Probabilistic self assembly model (PTAM); Randomized algorithms

1 Introduction

Self assembly is an autonomous process by which simple nanoscale components assemble into much complex nano structures. Self assembly is ubiquitous in nature in the form of lipid polymers and crystals. Theoretical foundations for understanding how to assemble these nano components into desired shapes is of central importance to nanotechnology. Applications of self assembly span from human health to nanoelectronics. Fundamental limits of optical lithography in circuit fabrication is being addressed by directed self assembly. Gold nanostructures are being used for drug delivery in a controlled fashion to treat cancer. It has been proven (Rothmund et al. 2004) that DNA crossover molecules can help in building stable structures.

The *tile assembly model* introduced by Winfree¹ has gained importance because of its close proximity to the physical self assembly process. This model is based on the theory of Wang's (Wang 1962) tiling of planar shapes. Informally this model has unit square tiles with a colored glue attached to each side. Two tiles can stick to each other along an edge only if they have a glue of the same color along the sticking edge. These tiles cannot be rotated during the self assembly process and they can only be translated. Given a target shape the goal is to design these tiles with colored glues so that they can be assembled into the target shape uniquely. The efficiency of the tileset is determined by the number of unique tiles used in the tileset. Initial work in this area focussed on the design of tilesets which can assemble the target shape exactly (Rothmund and Winfree 2000; Adleman et al. 2001; Aggarwal 2005; Kao and Schweller 2006). Efficient tilesets were designed for self assembling squares by Rothmund et al. (2000) and Adleman et al. (2001). Also, a lower

bound of $\Omega\left(\frac{\log(n)}{\log(\log(n))}\right)$ on the number of unique tiles to self assemble a square has been established by Rothmund et al. (2000).

The tile complexity lower bound is dictated by the Kolmogorov complexity. In all of these tilesets the information about the dimension is encoded in the tiles. As the dimension of the target shape increases the number of unique tiles also increases. The design of a large number of unique tiles is an intensive task. However copies of a tile can be created rapidly using PCR experiments. This observation has led to a new paradigm of *concentration programming* in self assembly systems. In a concentration programming model the information about the dimension is encoded in the relative concentration of the tiles. Thus the number of unique tiles in these systems is reduced considerably. However the price we pay on this model is that the final structure may not have the exact dimensions we are targeting. The difference between the target dimension and the assembled dimension is bounded by a certain probability. Currently there are two variants in concentration programming. In the first type the relative concentrations of the tiles may not be the same. This model, known as the *non-equimolar model*, was used in (Kao and Schweller 2008; Doty 2009) and (Becker et al. 2006). In the second type the relative concentrations of all the tiles are equal. This model, called the *equimolar model*, was studied in (Chandran et al. 2009). Equimolar concentration programming is also referred to as the *Probabilistic Tile Assembly Model* (PTAM).

Becker et al. (2006), studied the problem of self assembling squares on the non-equimolar model. They showed how to self assemble squares and rectangles which are expected to have dimensions close to the target dimensions. The techniques introduced by Becker et al. suffered from high variance on the dimensions achieved. This shortcoming was addressed

¹Phd thesis of Winfree (1998) [<http://resolver.caltech.edu/Calte-CHETD:etd-05192003-110022>].

by Kao and Schweller (2008). Later Doty (2009) independently considered the problem of making the idea of Kao (2008) more precise. Doty [2] showed that the line sampling estimate of n (the dimension of the square) can be made precise by estimating smaller parts of the binary representation of n . However this construction still needs to encode a binary number on the tiles to self assemble a $n \times n$ square. The constructions of both Kao (2008) and Doty (2009) are not simple and involve sub-tilerets to perform binary addition and division—thus have large hidden constants. In contrast one of the major results of our paper is to show how to construct simple tilerets without the need to have sub-tilerets to perform binary addition and division. Our algorithms can self assemble squares and rectangles with dimensions close to the target with high probability. Also, all the existing randomized techniques only address self assembly of squares. In this paper we also give constructions that can self assemble rectangles with a constant aspect ratio (α/β) and squares are a special case of these rectangles. These rectangles have dimensions close to target dimensions with high probability. Our rectangle constructions are applicable in the deterministic context as well. On the other hand, Chandran et al. (2009) have studied the complexity of self-assembling lines on the *PTAM* model. In our second result we show how to self assemble squares and rectangles (with a fixed aspect ratio) on the *PTAM* model with a tile complexity of $\Theta(\log(n))$.

In this paper we use n to denote the target dimension that needs to be assembled and n' to denote the dimension of the final assembled structure.

2 Basics of the tile assembly model

The tile model $t = (N(t), S(t), W(t), E(t)) \in \Sigma^4$ used in the assembly is a square shaped Wang tile with glues/symbols (from alphabet Σ) attached to each of the four sides. $N(t)$, $S(t)$, $W(t)$, $E(t)$ denote symbols attached to the north, south, west and east directions, respectively. Note that these tiles cannot be rotated and hence $(\sigma_n, \sigma_s, \sigma_w, \sigma_e) \neq (\sigma_e, \sigma_w, \sigma_n, \sigma_s)$. A *self assembly system* is characterized by a five tuple $\langle T, s, \tau, G, P \rangle$. $T \subseteq \Sigma^4$ is the *tileret* and is a collection of unique tiles used in the assembly process. The tile s is called the *seed tile* which is fixed at a particular location to initiate the assembly process. The symbol τ is a positive integer that indicates the *temperature* of the model. The function $G: \Sigma \times \Sigma \rightarrow \{0, 1, \dots, \tau\}$ is called the *bond strength* function. For any two symbols $x, y \in \Sigma$ the function $G(x, y)$ is defined as follows. Here ϵ denotes the empty symbol.

$$G(x, y) = \begin{cases} 0 & \text{If } (x \neq y) \vee (x = \epsilon \vee y = \epsilon) \\ k \in \{1, 2, \dots, \tau\} & \text{If } x = y \neq \epsilon \end{cases}$$

Informally the function G associates a glue strength with each symbol in Σ . A *bond* can only be formed between glues which are of the same type/symbol. The function P is the probability distribution associated with a subset of tiles in T . Since we study the self assembly of only planar shapes we consider only self assembly models which have a temperature $\tau = 2$.

Theoretically, we can think of the self assembly as a process that occurs on an infinite 2-dimensional integer grid. We define a null tile ϕ as $(\epsilon, \epsilon, \epsilon, \epsilon)$. The state of the self assembly process at any stage is represented by a mapping $C': \mathbb{Z} \times \mathbb{Z} \rightarrow \{T \cup \phi\}$. If $C'(x, y) = \phi$ it means that there is no tile placed at position (x, y) on the integer grid. We define a set $C := \{(x, y) | C'(x, y) \neq \phi\}$ as the *configuration* of the self assembly system. It is assumed that the seed tile s is placed initially at position $(0, 0)$ and the null tile ϕ is placed on all the grid points except $(0, 0)$. Thus the initial configuration of the self assembly system is $C = \{(0, 0)\}$

and $C'(0,0) = s$. Given a configuration C we say that a tile $t \in T$ is *attachable* to C at (x, y) only if both the following constraints (**C1** and **C2**) hold.

$$\begin{aligned} \text{Let } g_1 &= G(N(t), S(C'(x, y+1))) + G(S(t), N(C'(x, y-1))) \\ \text{Let } g_2 &= G(W(t), E(x-1, y)) + G(E(t), W(x+1, y)) \\ C'(x, y) &= \varphi(\text{C1: position } (x, y) \text{ is empty}) \\ g_1 + g_2 &\geq \tau \\ (\text{C2: neighbour tiles should provide enough strength}) \end{aligned}$$

Informally both these constraints mean that a tile $t \in T$ can be assembled to a configuration C at position (x, y) only if that position is empty and it could receive a *glue strength* of at least τ from the neighbouring tiles if it were to be placed at position (x, y) . If a tile $t \in T$ is *attachable* to C at (x, y) , then the configuration C moves to a new configuration C_1 such that $C_1 = C \cup \{(x, y)\}$ and $C'(x, y) = t$. In the self assembly literature a configuration C is also referred to as a *supertile* C . We can describe the self assembly process with a directed acyclic graph $D_g = (V, E)$, where V is the set of all supertiles and E is the set of directed edges. There is a directed edge (C_1, C_2) if and only if $\exists t \in T$ which is *attachable* to supertile C_1 to get the supertile C_2 . The *in-degree* of the initial supertile C [i.e., the seed tile s at position $(0,0)$] is zero. The self assembly process continues to move from one supertile to another with the help of some *attachable* tile $t \in T$. The assembly process stops at a supertile C_t if $\nexists t \in T$ which is *attachable* to C_t . Such a supertile is called a *terminal* supertile. A tile system $\Gamma = \langle T, s, \tau, G, P \rangle$ is said to produce a shape Y *uniquely* if all the *terminal* supertiles have exactly the same shape (including dimensions) as Y . Deterministic assembly systems focus on the construction of tilesets which can *uniquely* assemble a given shape with the minimum *tile complexity*.

2.1 PTAM and concentration programming

PTAM $\langle T, s, \tau, G, P \rangle$, in contrast to deterministic assembly, can produce terminal supertiles which correspond to multiple shapes. However one of these shapes (close to the target shape) is produced with high probability. Note that the probability distribution function P , in PTAM is uniform. On the other hand the distribution P may not be uniform on the concentration programming model.

2.2 Our results

We summarize our results as follows. In our first result on the concentration programming model we introduce a new sampling technique based on the sum of random variables which follow geometric distribution. We then use this to show how to construct a supertile which can encode a binary number n' on the tiles such that $|n' - n| < \epsilon n$ with high probability (i.e., $\Omega(1 - 1/n^a)$, for any fixed $a > 0$). Next we introduce a new self assembly sampling technique called the *staircase sampling*. We then show how this sampling scheme can be used to self assemble squares and rectangles with constant aspect ratio such that $|n' - n| < \epsilon n$ with high probability.

Our second result in this paper shows how staircase sampling can be used on the PTAM model to self assemble rectangles (with fixed aspect ratios) with high probability. Our result has a tile complexity of $\Theta(\log(n))$ and is optimal on the PTAM.

3 Randomized self assembly with a sum of geometric distributions (on the concentration programming model)

The idea of assigning probabilities to various tile types was considered by Becker et al. (2006). They introduced a concentration programming model tile system consisting of two tiles $A = (\epsilon, \epsilon, \sigma, \sigma)$, $A' = (\epsilon, \epsilon, \sigma, \epsilon)$ and a seed tile $s = (\epsilon, \epsilon, \epsilon, \sigma)$. The glue strength function and the probability distribution of the tiles are defined as follows. $G(\sigma, \sigma) = 2$, $P[A] = p$ and $P[A'] = 1 - p$. The seed tile s is placed at $(0,0)$. The self assembly grows a line from left to right. The tile A helps to increase the length of the line since $G(W(A)) = G(E(A)) = 2$. The tile A' terminates the assembly process since $G(E(A')) = 0$. We can associate a random variable L corresponding to the length of the assembled line. It is clear that $E[L] = 1/p$ since L follows a geometric distribution. So if we choose the concentration of tile A' such that $P[A'] = p = 1/n$, then the expected length of the self assembled line is n . By adding a constant number of tiles to this tileset it is also possible to self assemble a square of expected size $n \times n$. However this line sampling technique suffers from high variances on the dimensions.

Kao et al. (2008) have focused on reducing the variance by introducing an extra tile $A_1 = (\epsilon, \epsilon, \sigma, \sigma)$ into the previous tile set. Note that the tile A_1 has exactly the same definition as A , however it has a different color (say red). Kao et al. associated a random variable R that corresponds to the number of red tiles (i.e., of type A_1) in a self assembled line of length L . It is easy to see that the random variable R follows a binomial distribution. Since R follows a binomial distribution, Kao, et al. bounded the quantity L/R by applying Chernoff bounds. However Kao et al.'s tileset needs tiles for computing the values of L , R and L/R . This means that we need sub-tilesets to perform binary addition and division. One of the contributions of our paper is to show that the same high probability bounds can be derived in a much simpler manner by considering a random variable that is expressed as the sum of multiple geometrically distributed random variables. We now give more details of our constructions. Theorem 1 is already known and gives the Chernoff bounds for the sum of random variables that follow a geometric distribution.

Theorem 1. If $X = \sum_{i=1}^k X_i$ is the sum of k independent geometrically distributed variables then for any $0 < \epsilon < 1$, $P[X \geq (1 + \epsilon)\mu] \leq e^{-\mu\epsilon^2/3}$ and $P[X \leq (1 - \epsilon)\mu] \leq e^{-\mu\epsilon^2/2}$ where $\mu = E[X]$.

Our randomized self assembly technique uses multiple geometric distributions. Consider the line sampling technique introduced earlier which has only two unique tiles A and A' . The sampling line continues to grow with tile A until terminated by tile A' . We now introduce tiles $A_1, A_2, \dots, A_k$ and $A'_0, A'_1, A'_2, \dots, A'_k$. The growth of tile A_i is terminated by tile A'_{i+1} , $1 \leq i \leq k$. The tile A'_0 denotes the seed tile. In contrast to line sampling we now have k independent self assembling stages terminated by tile A'_{k+1} . Figure 1 further clarifies our idea. We now associate a random variable X corresponding to the length of the terminal self assembled line. We also associate the random variables X_i , $1 \leq i \leq k$ corresponding to the length of the self assembly between tiles A'_{i-1} and A'_i . Clearly $X = \sum_{i=1}^k X_i + (k+1)$ and each of the random variables X_i follows a geometric distribution with a mean of $1/p$ where p is the probability of success. We now prove Theorem 2 based on our randomized self assembly technique. Later we will show how to add a constant number of tiles to this sampling technique to actually encode the binary value on to the tiles. Note that n' and n are used to denote the length of the self assembled line with sampling and the length of the target/required line, respectively.

Theorem 2. For any given $\varepsilon \in (0, 1)$ and a positive integer n , the multi geometric sampling technique can self assemble a line of length n' such that $|n' - n| < \varepsilon n$ with a probability of

$$\Omega(1 - 1/n^\alpha) \text{ for any } \alpha \leq \frac{n\varepsilon^2 \log(e/2)}{3 \log(n)}.$$

Proof We have associated a random variable X corresponding to n' . Since

$X = \sum_{i=1}^k X_i + (k+1)$ and each X_i follows a geometric distribution we can apply Chernoff bounds in Theorem 1 as follows.

$$\begin{aligned} \mu &= E[X] = \sum_{i=1}^k E[X_i] + (k+1) = k/p + (k+1) \\ \text{Let } S_1 &\text{ be the event that } X \geq (1+\varepsilon)\mu; \\ \text{Let } S_2 &\text{ be the event that } X \leq (1-\varepsilon)\mu; \\ \rightarrow & P[S_1] \leq e^{-\frac{\mu\varepsilon^2}{3}} && (\text{By Chernoff bounds in Theorem 1}) \\ \rightarrow & P[S_2] \leq e^{-\frac{\mu\varepsilon^2}{2}} && (\text{By Chernoff bounds in Theorem 1}) \\ \rightarrow & P[S_1 \cup S_2] \leq P[S_1] + P[S_2] && (\text{Union rule}) \\ & P[S_1] + P[S_2] \leq 2e^{-\frac{\mu\varepsilon^2}{3}} \\ \rightarrow & P[\overline{S_1} \cap \overline{S_2}] \geq 1 - \frac{2}{e^{(\mu\varepsilon^2)/3}} \\ & \text{Now select } p = \frac{k}{n-(k+1)} \rightarrow \mu = n \\ \rightarrow & P[\overline{S_1} \cap \overline{S_2}] \geq 1 - 2/e^{(n\varepsilon^2)/3} > 1 - 1/n^\alpha \quad (\alpha \leq \frac{n\varepsilon^2 \log(e/2)}{3 \log(n)}) \\ & \text{Note } \overline{S_1} \cap \overline{S_2} = \{X : |X - \mu| < \varepsilon\mu\} \\ \rightarrow & P[|n' - n| < \varepsilon n] = \Omega(1 - 1/n^\alpha). \end{aligned}$$

3.1 Supertile to encode a binary string

We now show how our random sampling technique can be used to self assemble a supertile which can encode a binary string of $\log(n)$ bits whose value is n' . Building a supertile that can encode a given binary string is fundamental to the construction of several shapes. As shown by Rothmund and Winfree (2000), given a supertile encoded with a specific binary string, we can build an $n \times n$ by adding only a constant number of tiles to the tileset.

Deterministically optimal tilesets of size $\Theta(\frac{\log(n)}{\log(\log(n))})$ exist for building $n \times n$ [see (Adleman et al. 2001)]. In Kao and Schweller (2008) gave a tile assembly that can encode any binary value n' on the tiles. In Kao et al.'s construction the length of the sampling line does not have a direct correspondence to n' and requires a sub-tile assembly which performs division. In contrast, our multi-geometric sampling technique can achieve equivalent (proved in Theorem 2) stochastic properties for encoding n' in binary with a much simpler supertile construction. Figure 2 gives the overview of how this encoding is done with just a constant (7) number of tiles. The basic idea behind this binary encoding is to integrate a binary counter to the self assembling sampling line by modifying the binary counter tileset presented in (Rothmund and Winfree 2000). Figure 2 shows the self assembly of a supertile with two geometrically distributed variables when the random variable $X = n'$ takes a value of 13. In this example, $X_1 = 6$ and $X_2 = 4$. The green colored tile in Fig. 2 represents the seed tile. The tiles on the left part of the tileset vary with the number of variables used in the sampling process. If we use k geometrically distributed variables, then this would have $2k + 1$ (in this example $k = 2$) unique tiles. The right part of the tileset is independent of the number of variables and will always have 7 tiles in it. The binary value of n' is encoded on the tiles along the right corner of the supertile from top to bottom. We now summarize the discussion in this section with Theorem 3.

Theorem 3. For any given $\varepsilon \in (0, 1)$ and a positive integer n there exists a tile system with constant tiles to encode a binary string of value n' such that $|n' - n| < \varepsilon n$ with a probability

$$\text{of } \Omega(1 - 1/n^\alpha) \text{ for any } \alpha \leq \frac{n\varepsilon^2 \log(1/2)}{3\log(n)}.$$

4 The technique of staircase sampling

We now introduce our idea of *staircase sampling*. This technique helps in the randomized self assembly of rectangular shapes with a constant aspect ratio (α/β). The idea of randomized self assembly of rectangles was introduced by Becker et al. (2006). Their idea generalizes the line sampling described in Sect. 3 for two dimensions (north, east). They associate two random variables M and N to these sampling lines to self assemble a $m \times n$ rectangle where m and n are the expected height and width, respectively. Since these are only expected dimensions their technique suffers from high variances on these dimensions. Also their technique will not work in self assembling much stronger classes of shapes such as a class of rectangles with constant aspect ratio. Note that the random variables M and N have a geometric distribution. If p_n and p_e are the probabilities of success along north and east, then the probability of producing a $m_1 \times n_1$ rectangle by the self assembly is $P[M = m_1, N = n_1] = (p_n)^{m_1} (p_e)^{n_1}$. This will remain the same even if the aspect ratio is a constant (i.e. $m/n = \alpha/\beta$), since both the random variables are assumed to be independent in their technique.

Let $R(\alpha, \beta) := \{r(x, y) : x/y = \alpha/\beta\}$ denote a class of rectangles with a constant aspect ratio. Given the dimensions of any rectangle $r(n\alpha, n\beta) \in R(\alpha, \beta)$ our aim in this section is to show that we can self assemble $r(n'\alpha, n'\beta) \in R(\alpha, \beta)$ such that n' and n are very close with high probability. The key idea behind our sampling technique is to sample along the staircase in contrast to the existing techniques that sample along a straight line. We now show how to construct the tileset for staircase sampling.

4.1 Discussion on the tileset for staircase sampling

We introduce $\alpha + \beta - 1$ unique tiles corresponding to the aspect ratio α/β of the rectangle we would wish to self assemble. Please see Fig. 3 to follow this discussion. All the symbols in Fig. 3 $\Pi, \Psi, a_1, a_2, \dots, a_{\alpha-1}$ and $b_1, b_2, \dots, b_{\beta-1}$ have a glue strength of 2. The tiles which do not carry any symbol along sides are assumed to have a dummy symbol x which has a glue strength of 1 (i.e., $G(\omega, \omega) = 2$). With this definition of glue strength function G , the tiles $t_1, t_2, \dots, t_\alpha$ can deterministically self assemble into a horizontal line of length α ending with tile t_α . Notice that the tile t_α has a symbol b_1 on the top. Since the symbol b_1 has a glue strength of 2, the tiles $t'_1, t'_2, \dots, t'_{\beta-2}$ deterministically self assemble into a vertical line of height $\beta - 2$. Till this point the self assembly process is deterministic. However after the tile $t'_{\beta-2}$ gets attached into the vertical line, we now have two tiles ($t'_{\beta-1}$ and e_0) which can get attached on top of tile $t'_{\beta-2}$. In the first case if the tile $t'_{\beta-1}$ gets attached on top of tile $t'_{\beta-2}$ then the self assembly can progress further. This is because the tile $t'_{\beta-1}$ has a symbol W on top which has a bond energy of 2. However in the second case if the tile e_0 gets attached on top of tile $t'_{\beta-2}$ the self assembly process will not grow further vertically since the tile e_0 has no symbol on the top. The filler tile f_0 gets assembled into places which can provide a glue strength of 2 and completes the rectangle.

To further illustrate the idea consider the tileset shown in Fig. 4 to self assemble a rectangle with aspect ratio $4/3$ ($\alpha = 4, \beta = 3$). Figure 5 shows the actual assembly process. Notice that the tiles t'_2 and e_0 play a key role in either increasing the width (and height correspondingly) or stopping further growth in the dimension of the rectangle. In general we attach a probability of *success* (p) and *failure* ($1 - p$) to the tiles $t'_{\beta-1}$ and e_0 , respectively. We state the following Theorem 4 relevant to the discussion in this section.

Theorem 4 Given any three integers $\alpha, \beta, n > 1$ the staircase sampling technique can self assemble a rectangle with an expected width and height of $n\alpha$ and $n\beta$, respectively.

4.2 Generalization of staircase sampling with multiple variables

We now apply the multi geometric distribution technique introduced in Sect. 3 to staircase sampling to derive high probability bounds on the dimensions of the terminal self assembled rectangle. We generalize the tileset for staircase sampling in Fig. 3 by introducing an extra subscript for all the symbols and tiles. If we choose to have k geometric distributions, then for $1 \leq i \leq k$, we would have symbols $\Phi_i, \Psi_i, a_{i1}, a_{i2}, \dots, a_{i\alpha}, b_{i1}, b_{i2}, \dots, b_{i\beta-1}$ on the tiles $t_{i1}, t_{i2}, \dots, t_{i\alpha}, t'_{i1}, t'_{i2}, \dots, t'_{i\beta-1}$ similar to the construction in Fig. 3. Also notice that previously we used the tile e_0 (in Fig. 3) to stop growth in the dimensions of the rectangle. However in this case we need to introduce k such tiles ($e_0, e_1, \dots, e_{k-1}$) with an objective of moving from distribution i to distribution $i + 1$. This is similar to the role played by the tiles A'_i in Sect. 3. To accomplish this task we need to modify tiles $e_{j-1}, 1 \leq j \leq k-1$, to carry a symbol of Π_j along their right side. Finally, we introduce random variables $X_i, 1 \leq i \leq k$, to indicate the number of tiles of type $t'_{i\beta-1}$ during the i th stage of the multi geometric sampling. Let

$X := \sum_{i=1}^k X_i$. Clearly if $X = t$ then the final self assembled rectangle will have dimensions $t\alpha \times t\beta$. With this we state the following Theorem 5.

Theorem 5 For any given $\varepsilon \in (0, 1)$, and integers $n, \alpha, \beta > 1$, the multi geometric sampling combined with stair case sampling can self assemble a $n'\alpha \times n'\beta$ rectangle such that $|n' - n|$

$$< \varepsilon n \text{ with a probability } \Omega(1 - 1/n^q) \text{ for any } q \leq \frac{n\varepsilon^2 \log(e/2)}{3 \log(n)}.$$

A direct corollary when $\alpha = \beta$ self assembles a square of dimension n' such that $|n' - n| < \varepsilon n$ with high probability. This is equivalent to what Kao and Schweller derived in 2008 without the need of any supertile that encodes a binary string. Thus we can conclude that combining multiple geometric variables with stair case sampling lets us derive high probability bounds for self assembling rectangles with constant aspect ratios.

5 Self assembly of rectangles on the PTAM model

We now show how our stair case sampling technique can be extended on the PTAM model to self assemble rectangles with fixed aspect ratios ($\alpha:\beta$). The key idea behind our technique is the following tile binding diagram (Fig. 6). This is an extension to the linear tile binding diagram in (Chandran et al. 2009), replacing the linear binding assembly with the staircase component. This construction has n sub tilesets each of which can deterministically assemble a stair. For example the sub tileset $t_0^1, t_1^1, \dots, t_\alpha^1, \dots, t_{\alpha+\beta-1}^1$ self assembles the first stair. At any stage i , the non-determinism part is induced by tiles t_0^{i+1} and t_0^1 (green colored tile, see Fig. 6) which compete to bind with $t_{\alpha+\beta-1}^1$. The self assembly process at the i th stage restarts every time a t_0^1 tile binds to $t_{\alpha+\beta-1}^1$ and moves to stage $i + 1$.

We first state the following Lemma 1 from (Chandran et al. 2009) on a sequence of Bernoulli trials. Let p be the success probability of a Bernoulli trial.

Lemma 1 Let Y be a random variable which indicates the number of Bernoulli trials to get

$$m \text{ consecutive successes. Then } E[Y] = \frac{1 - m^m}{p^m(1-p)}.$$

It is easy to see that every terminal assembly produces a rectangle with fixed aspect ratio $\alpha:\beta$. This is because of the sub tilesets which always assemble a stair with width α and height β . The filler tile (blue, see Fig. 6) fills up all the gaps and completes a rectangle. Now let the dimensions of this rectangle be αLL . The assembly becomes terminal iff we see the following sub tilesets $(t_0^1, \dots, t_{\alpha+\beta-1}^1), (t_0^2, \dots, t_{\alpha+\beta-1}^2), \dots, (t_0^n, \dots, t_{\alpha+\beta-1}^n)$.

Theorem 6 *Let L be the random variable corresponding to the width (after self assembly) of a rectangle with fixed aspect ratio $\alpha:\beta$. Then $E[L] = 2^{n+1} - 2$. Thus, we can self assemble a rectangle with an expected width of $\alpha 2^{n+1} - 2$ and an expected height of $\beta 2^{n+1} - 2$.*

Proof We associate a Bernoulli trial with every stage of the self assembly. We treat the

binding of tile t_0^1 to $t_{\alpha+\beta-1}^1$ as a success (probability being $\frac{1}{2}$). Then by Lemma 1 the expected width of the assembly is $2^{n+1} - 2$.

To make this expected bound a high probability bound we can use a technique similar to the one in Sect. 3, which sums up a constant number of independent and identically distributed random variables (i.e., $L = L_1 + L_2 + L_3 + \dots + L_k$). Even though in our construction the dimension of the rectangle is a power of 2 we can easily add tiles which perform binary arithmetic and extend the result for any arbitrary N . Finally, this tile complexity is optimal on the PTAM, because the result in (Chandran et al. 2009) proves a lower bound of $\Omega(\log(N))$ to self assemble a line on the PTAM.

Theorem 7 *There exists an algorithm on the PTAM to self assemble a rectangle (with a fixed aspect ratio) of dimension N using $\Theta(\log(N))$ tiles with high probability.*

6 Conclusion

In this paper we have presented new constructions for probabilistic self assembly and derived high probability bounds on the dimensions of the terminal supertile. We conclude that probabilistic self assembly with a sum of random variables following geometric distribution is helpful in simpler self assembly constructions in contrast to the existing techniques presented in Kao et al. (2008). We also have introduced a new idea of *staircase sampling* which is useful in assembling rectangles with constant aspect ratios. These shapes cannot be assembled with the existing probabilistic self assembly techniques.

7 Future work

Probabilistic self assembly is an exciting area especially because of the fact that we need only $\Theta(1)$ tiles to self assemble shapes which are close to the target shapes. One future research will include the investigation of how the current techniques to self assemble squares and rectangles can be combined to self assemble Manhattan shapes within the $\Theta(1)$ tile complexity. Another direction of research would be to explore probabilistic constructions of non-rectangular regular polygons.

Acknowledgments

This work has been supported in part by the following grants: NSF 0829916, NIH LM010101, and NIH GM079689.

References

- Adleman, L.; Cheng, Q.; Goel, A.; Huang, M. annual ACM symposium on theory of computing. ACM; New York: 2001. Running time and program size for self-assembled squares; p. 740-748.

- Aggarwal G, Cheng QI, Goldwasser MH, Kao M, De Espanes PM, Schweller RT. Complexities for generalized models of self-assembly. *SIAM J Comput.* 2005; 34(6):1493–1515.
- Becker, F.; Remila, E.; Rapaport, I. Self-assembling classes of shapes with a minimum number of tiles, and in optimal time. *Proceedings of the 26th Conference on Foundations of Software Technology and Theoretical; Springer.* 2006.
- Chandran, H.; Gopalkrishnan, N.; Reif, J. The tile complexity of linear assemblies. *proceedings of the 36th international colloquium automata, languages and programming: Part I; Springer-Verlag; 2009.* p. 235-253.ICALP '09
- Doty, D. Randomized self-assembly for exact shapes. *proceedings of the 50th annual IEEE symposium on foundations of computer science (FOCS), IEEE; Atlanta.* 2009. p. 85-94.
- Kao, M.; Schweller, R. Reducing tile complexity for self-assembly through temperature programming. *Annual ACM-SIAM Symposium on Discrete Algorithms; Miami.* 2006. p. 571-580.
- Kao, M.; Schweller, R. Randomized self-assembly for approximate shapes, LNCS, vol 5125. Springer; Heidelberg: 2008.
- Rothmund, PWK.; Winfree, E. Program-size complexity of self-assembled squares. *ACM symposium on theory of computation (STOC); New York: ACM; 2000.* p. 459-468.
- Rothmund PWK, Papadakis N, Winfree E. Algorithmic self-assembly of dna sierpinski triangles. *PLoS Biol.* 2004; 2(12):e424. [PubMed: 15583715]
- Wang, H. An unsolvable problem on dominoes. 1962. Technical Report BL-30
- Winfree, E. Dissertation (PhD). California Institute of Technology; 1998. Algorithmic self-assembly of DNA. <http://resolver.caltech.edu/CaltechETD:etd-05192003-110022>

Fig. 1.
Randomized self assembly with multiple geometric distributions

Fig. 2. Self assembly of the supertile encoding a binary string on the tiles. The strength function G of a symbol is illustrated by the bars attached to the tiles. For example strength of the symbol c , $G(c, c)$ is indicated by *two bar*

Fig. 3.
General tileset for staircase sampling to self assemble a rectangle of aspect ratio α/β

Fig. 4.
Tile set for self assembling a rectangle with aspect ratio 4/3

Fig. 5.
Self assembly of the rectangle with the tileset in Fig. 4

Fig. 6. Tile binding diagram to self assemble a rectangle ($a2^n \times \beta 2^n$) using exactly $n \times (a + \beta + 1)$ unique tiles. In this example $a = 3, \beta = 4$