Sequential Non-Determinism in Tile Self-Assembly: A General Framework and an Application to Efficient Temperature-1 Self-Assembly of Squares

David Furcy
Computer Science Department, University of Wisconsin Oshkosh, WI, USA

Scott M. Summers[1]
Computer Science Department, University of Wisconsin Oshkosh, WI, USA

Abstract

In this paper, we work in a 2D version of the probabilistic variant of Winfree's abstract Tile Assembly Model defined by Chandran, Gopalkrishnan and Reif (SICOMP 2012) in which attaching tiles are sampled uniformly with replacement. First, we develop a framework called "sequential non-determinism" for analyzing the probabilistic correctness of a non-deterministic, temperature-1 tile assembly system (TAS) in which most (but not all) tile attachments are deterministic and the non-deterministic attachments always occur in a specific order. Our main sequential non-determinism result equates the probabilistic correctness of such a TAS to a finite product of probabilities, each of which (1) corresponds to the probability of the correct type of tile attaching at a point where it is possible for two different types to attach, and (2) ignores all other tile attachments that do not affect the non-deterministic attachment. We then show that sequential non-determinism allows for efficient and geometrically expressive self-assembly. To that end, we constructively prove that for any positive integer N and any real $\delta \in (0,1)$, there exists a TAS that self-assembles into an $N \times N$ square with probability at least $1-\delta$ using only $O\left(\log N + \log \frac{1}{\delta}\right)$ types of tiles. Our bound improves upon the previous state-of-the-art bound for this problem by Cook, Fu and Schweller (SODA 2011).

2012 ACM Subject Classification Theory of computation

Keywords and phrases Self-assembly, tile assembly model, temperature 1, high probability self-assembly

Digital Object Identifier 10.4230/LIPIcs.DNA.32.1

Funding *Scott M. Summers*: This author's research was supported in part by University of Wisconsin Oshkosh Research Sabbatical (S581) during Fall 2023.

Acknowledgements We thank Matthew Patitz for providing helpful insights regarding the implementation of several aspects of our construction. We also thank Dave Doty for providing initial guidance on probabilistic self-assembly. Finally, we are grateful to the anonymous reviewers whose detailed comments helped improve the clarity and technical accuracy of the paper.

1 Introduction

The notion of "self-assembly" is generally said to involve seemingly simple, fundamental components, evolving through instances of local interaction, to a terminal assembly whose complexity is greater than the sum of its parts. Self-assembly is ubiquitous in nature, e.g., atoms self-assemble into molecules, molecules self-assemble into macromolecules, etc.

While self-assembly is ubiquitous throughout the natural world, researchers have recently begun investigating the extent to which the power of self-assembly can be harnessed for the systematic creation of atomically-precise computational, biomedical and mechanical

[1] Corresponding author

32nd International Conference on DNA Computing and Molecular Programming (DNA 32).
Editors: Dominic Scalise and Robert Schweller; Article No. 1; pp. 1:1–1:18
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

devices at the nano-scale. For example, in the early 1980s, Nadrian Seeman introduced a revolutionary experimental technique for controlling nano-scale self-assembly known as "DNA tile self-assembly" [28] thus initiating the scientific study of algorithmic self-assembly. Erik Winfree's abstract Tile Assembly Model (aTAM) [30] is a discrete mathematical model of DNA tile self-assembly, which we describe briefly in the next paragraph.

In the aTAM, a DNA tile is represented as an non-rotatable unit square *tile type*, each side of which has a corresponding *glue* comprising a string *label* and a positive integer *strength*. A *tile* is an instance of a tile type that is *placed* at some integer point in the 2D Cartesian coordinate space. The aTAM restricts the number of tile types in a tile set to be finite but infinitely many tiles of the same type may be placed. If two tiles are placed next to each other and their opposing glues match, then the tiles *bind* with the strength of the glue, thus creating a tile *assembly*, i.e., a mapping of integer points to tile types, of size 2. Before the process of self-assembly begins, a positive integer *temperature* value is chosen, which is usually 1 or 2 and a fixed *seed* tile is placed at a designated integer point defining the initial seed-containing assembly. The tile set, along with the associated temperature and placement of the seed tile make up a *tile assembly system* (TAS). Self-assembly in the aTAM is modeled by an *assembly sequence* of the TAS, which is a series of *tile attachment steps* such that in each step, the assembly sequence *attaches* a tile if the tile can be placed at an unoccupied integer point and bind to one or more tiles of the seed-containing assembly with total strength at least the temperature. If more than one tile may attach in a single step, then one tile is chosen non-deterministically to attach. Each tile attachment step in an assembly sequence *results* in a new assembly one tile larger than the assembly in the previous step, and we say the resulting assembly is *produced* by the TAS. The steps of an assembly sequence continue until a *terminal assembly* is reached to which no further tiles can bind. Figure 1 depicts an example of a *non-deterministic*, temperature-1 TAS $\mathcal{T}_1$ in which two different terminal assemblies are produced.

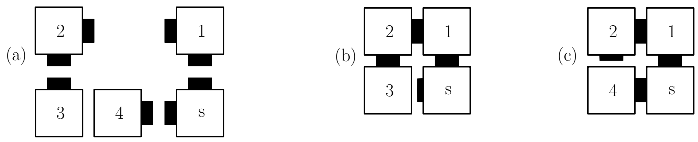

Figure 1 Example of a TAS $\mathcal{T}_1$ with the tile set shown in (a), in which the temperature is 1 and s is the seed tile. Sub-figures (b) and (c) show assembly sequences of $\mathcal{T}_1$ that result in two different terminal assemblies, with α_{1c} denoting the terminal assembly in sub-figure (c).

Observe that temperature-1 self-assembly, which involves a TAS whose temperature is set to 1, does not possess a synchronization mechanism that can be used to enforce cooperative binding, i.e., a tile attaching at a point only after at least two of its neighboring tiles attach. However, temperature-2 self-assembly provides a mechanism to enforce cooperative binding, which can be leveraged to exhibit 1) optimal self-assembly of squares [1, 27] and algorithmically-specified finite shapes [29], 2) efficient self-assembly of rectangles [4], and 3) even an intrinsically universal tile set, i.e., a tile set that can, for an arbitrary input TAS, be configured to carry out a natural simulation of the input TAS but at a higher scale factor [8]. With the lack of a cooperative binding mechanism in temperature-1 self-assembly, it would seem natural to speculate that the computational and geometric expressiveness of temperature-1 self-assembly is fundamentally limited.

Such speculation was formalized back in 2008 by Doty, Patitz and Summers [10] through their pumpability conjecture in which they claimed that the only kinds of patterns that result from temperature-1 self-assembly in which a unique terminal assembly is produced effectively correspond to regular languages. Even before the pumpability conjecture, temperature-1 self-assembly had been shown by Rothemund and Winfree [27] to hinder the efficient self-assembly of shapes when tile assemblies are required to be fully connected, i.e., if two tiles are neighbors, then they bind to each other. Manuch, Stacho and Stoll [21] generalized this result to temperature-1 self-assembly that always results in tile assemblies that merely contain no glue mismatches. Subsequent results showing that temperature-1 self-assembly is neither intrinsically universal [22, 25] nor capable of bounded Turing computation [25] continued to elucidate the supposed weakness of not being able to prevent non-cooperative binding. A major breakthrough by Meunier, Regnault and Woods [24] established a general pumping lemma for temperature-1 self-assembly. Although it did not directly affirm the pumpability conjecture, Meunier and Regnault [23] used it in their proof of the pumpability conjecture.

Yet the lack of enforceable cooperative binding is not solely to blame for the weakness of temperature-1 self-assembly. In fact, generalizations of the aTAM such as [6, 9, 11, 12, 18] enjoy a striking resemblance of computational and geometric expressiveness to that of temperature-2 self-assembly. This is also the case when the aTAM is only mildly generalized, e.g., by allowing: 1) the use of a unique negative glue [26], 2) duple tiles [19], 3) the attachment of cubic tiles in at most one additional plane [2, 5, 13–17], or 4) non-determinism [5] in which multiple terminal assemblies may be produced.

The results in this paper are motivated by the problem of computing the probability that a particular terminal assembly is produced by non-deterministic, temperature-1 self-assembly. To study this problem for a given non-deterministic, temperature-1 TAS $\mathcal{T}$, we use the Markov chain $\mathcal{M}_{\mathcal{T}}$, in which the states are the assembly sequences of $\mathcal{T}$ and there is a transition in $\mathcal{M}_{\mathcal{T}}$ from one assembly sequence to another with a positive probability equal to one over the number of distinct assemblies that can result from attaching a single tile to the result of the former. This is the *tile concentration programming* aTAM subject to the restriction that all tile types have the same concentration (see [7, 20] for a more general and thorough development of the tile concentration programming model). Note that this is also the same as the probabilistic aTAM (PTAM), which was defined by Chandran, Gopalkrishnan, and Reif [3], but extended to 2D, which we will refer to as the 2DPTAM going forward. It is easy to see that $\mathcal{M}_{\mathcal{T}}$ is a rooted tree. Then, the *probability* that a given TAS $\mathcal{T}$ produces a terminal assembly is the probability of $\mathcal{M}_{\mathcal{T}}$ reaching any state that corresponds to an assembly sequence of $\mathcal{T}$ that results in the terminal assembly. For example, the probability that the TAS shown in Figure 1 produces α_{1c} can be easily computed as $\frac{7}{8}$ and the corresponding $\mathcal{M}_{\mathcal{T}_1}$ is shown in Figure 2. In that and all subsequent figures, all edge label values equal to 1 are omitted (but implied). While $\mathcal{M}_{\mathcal{T}}$ can be equivalently formulated in terms of a directed acyclic graph (e.g., [5]), we use a tree to aid in the visualization of some of our results.

Computing the probability that $\mathcal{T}_1$ from Figure 1 produces one of its terminal assemblies is easy. However, consider the TAS shown in Figure 3, which we will denote as $\mathcal{T}_3$, and is the same as $\mathcal{T}_1$ from Figure 1 but with the addition of the A and B tile types, along with corresponding east-facing glues to the 1 and s tiles.

The presence of the A and B tiles in $\mathcal{T}_3$ does not affect, but rather complicates the computation of, the probability that $\mathcal{T}_3$ produces one of its terminal assemblies. To see this, Figure 4 shows $\mathcal{M}_{\mathcal{T}_3}$, from which the probability that $\mathcal{T}_3$ produces the terminal assembly α_{3c} can be derived.

Based on $\mathcal{M}_{\mathcal{T}_3}$ shown in Figure 4, it is easy (although tedious) to see that the probability that $\mathcal{T}_3$ produces α_{3c} is: $\frac{1}{3}\frac{1}{4}\frac{1}{4}\frac{1}{2} + \frac{1}{3}\frac{1}{4}\frac{1}{4}\frac{1}{2} + \cdots + \frac{1}{3}\frac{1}{2}\frac{1}{2} = \frac{7}{8}$. Perhaps unsurprisingly, the above probability is equal to the probability of $\mathcal{T}_1$ producing α_{1c}. After all, the A and B tiles should

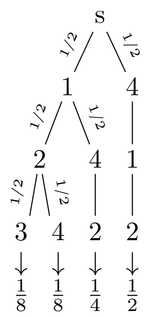

Figure 2 $\mathcal{M}_{\mathcal{T}_1}$ corresponding to the TAS in Figure 1. Each path from the root to a leaf node corresponds to an assembly sequence of $\mathcal{T}_1$. For the sake of simplification, the label of each node is merely the type of the last tile to attach in the assembly sequence to which that node corresponds, rather than the full assembly sequence, which can easily be reconstructed by following (and then reversing) the path back up to the root. Each leaf node corresponds to an assembly sequence that results in a terminal assembly. The number under each leaf node is the probability of the corresponding sequence being produced by $\mathcal{T}_1$.

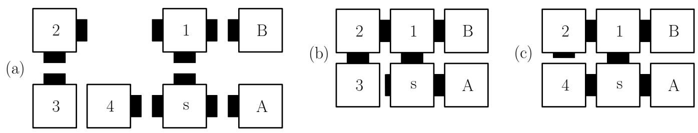

Figure 3 Example of another TAS $\mathcal{T}_3$ with the tile set shown in (a), in which the temperature is 1 and s is the seed tile. Sub-figures (b) and (c) show assembly sequences of $\mathcal{T}_3$ that result in two different terminal assemblies, with α_{3c} denoting the terminal assembly in sub-figure (c).

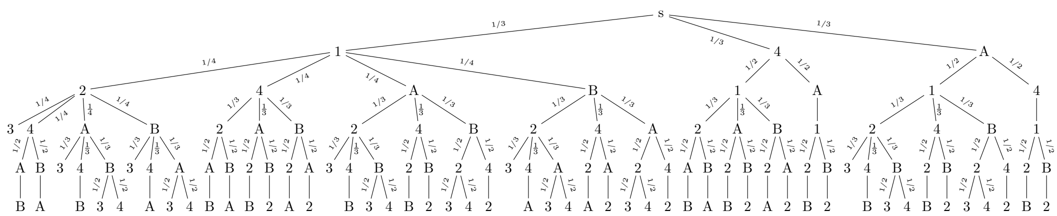

Figure 4 Markov chain $\mathcal{M}_{\mathcal{T}_3}$, where $\mathcal{T}_3$ is the TAS from Figure 3. Note that this tree is not quite complete because it only contains the assembly sequences that result in α_{3c}. All of the sub-trees whose root is labeled with a 3 and that do not grow down to the lowest level of the tree have been pruned below their root node (e.g., the leftmost node in the figure).

not affect the probability of $\mathcal{T}_3$ producing α_{3c}. So $\mathcal{T}_3$ should be equivalent to $\mathcal{T}_1$ from Figure 1 in a probabilistic sense. Nevertheless, the point of this example is to show how analyzing the probabilistic correctness of even a seemingly simple non-deterministic, temperature-1 TAS in the 2DPTAM can become unwieldy very quickly. This is because the probability that the TAS produces some terminal assembly depends on all the tile attachment steps of all the assembly sequences of the TAS that result in the assembly, including tile attachment steps in which tiles attach in such a way that should not affect the probability.

In this paper, we formalize the intuition that tiles that attach in such a way that should not affect the probability of a certain kind of non-deterministic, temperature-1 TAS in fact do not affect the probability. To that end, we develop a first-of-its-kind general framework called "sequential non-determinism" for analyzing the probabilistic correctness of such a

TAS as $\mathcal{T}_3$. In sequential non-determinism, we assume that for a given non-deterministic, temperature-1 TAS and in every step of every assembly sequence thereof, a tile initially attaches via exactly one of its glues and is the only type of tile that can attach at that location and via that glue. We also assume there exist (potentially infinitely many) *points of competition (POC)*, at which one of only two types of tiles may (and may only) be placed via two corresponding competing paths of tiles that originate from a common starting point and are disjoint until potentially meeting at the POC, e.g., the location at which tiles of type 3 or 4 in Figure 3 may attach. Without going into all the technical details, we require that the POCs be ordered, sufficiently spaced out, and that, in every assembly sequence of the TAS, tiles attach at the POCs in the specified order. We assume that each POC is either *essential* or *inessential*, that the number of essential ones is finite, and that each essential POC is associated with a type of tile that is the correct type of tile that should be placed at that POC. Under these assumptions, the TAS produces a unique *correct* terminal assembly (as well as arbitrarily many incorrect ones). Our main sequential non-determinism theorem says that the probability in the 2DPTAM that a given sequentially non-deterministic, temperature-1 TAS produces the unique correct (potentially infinite) terminal assembly is a finite product of probabilities, each of which corresponds to an essential POC having the correct type of tile placed at it. Moreover, our framework allows for the computation of this probability to be carried out while effectively ignoring the tile attachments that should not affect whether the correct type of tile attaches at an essential POC. For example, our main theorem allows us to analyze $\mathcal{T}_3$ as if it were $\mathcal{T}_1$.

We then show that the requirements of sequential non-determinism are not overly restrictive, since sequential non-determinism does not hinder sufficiently geometrically and computationally expressive self-assembly. To that end, we constructively prove that for any real $\delta \in (0,1)$ and any $N \in \mathbb{Z}^+$, there exists a TAS that produces the correct terminal assembly with probability at least $1-\delta$, whose domain is the $N \times N$ square $\{0,1,\ldots,N-1\}^2$ and whose tile set has a size that is $O\left(\log N + \log\frac{1}{\delta}\right)$. Note that, for a fixed δ, our bound is simply $O(\log N)$, which is an improvement over the previous state-of-the-art bound of $O\left(\log^2 N + \log N \log\frac{1}{\delta}\right)$ for this problem by Cook, Fu and Schweller [5].

In addition to the previous bound by Cook, Fu and Schweller, there are some other notable previous results related to probabilistic self-assembly in the aTAM. Working in the general tile concentration programming aTAM, where the concentration of each tile type can be set accordingly to affect the behavior of the TAS, Kao and Schweller [20] showed that for some fixed TAS and any positive integer N, it is possible to set the tile concentrations of the tile set of the TAS so that it produces, with high probability, a terminal assembly whose domain corresponds to an "approximation" of an $N \times N$ square. Technically, they showed that there exists a TAS such that for any $N \in \mathbb{Z}^+$ and for any real $\delta, \varepsilon > 0$, the tile concentrations can be set so that the TAS produces with probability $1-\delta$ a terminal assembly whose domain is $\{0,1,\ldots,N'-1\}^2$ such that $(1-\varepsilon)N \leq N' \leq (1+\varepsilon)N$. Among several other results, Doty [7] proved an "exact" analog of the aforementioned "approximate" result by demonstrating the existence of a TAS with constant size tile set such that, for any $N \in \mathbb{Z}^+$ and for any real $\delta > 0$, the concentrations of the tile types of the TAS can be set so that it produces with probability $1-\delta$ a terminal assembly whose domain is $\{0,1,\ldots,N-1\}^2$. Chandran, Gopalkrishnan, and Reif [3] investigated the size of the smallest tile set such that there is a corresponding singly-seeded, temperature-1, non-deterministic TAS that produces a terminal assembly whose domain is expected to be a *linear assembly*, i.e., an assembly whose domain, for some $N \in \mathbb{Z}^+$, is $\{0,1,\ldots,N-1\} \times \{0\}$. While the previously-mentioned results by Doty, and Kao and Schweller utilize arbitrarily-defined tile concentrations, in several of their results, Chandran, Gopalkrishnan and Reif worked in the PTAM (in 1D) and assumed the same uniform tile concentration programming model that we do in this paper. Among numerous other results, they proved that for any $N \in \mathbb{Z}^+$, there exists a TAS, whose tile set has size $O(\log N)$, that produces a linear terminal assembly with expected but perhaps widely varying length N. Chandran, Gopalkrishnan and Reif followed this result up by showing that there is a TAS whose tile set has size $O\left(\log^3 N\right)$ that does the same thing but can more precisely control the variance of the length of the produced terminal assembly. In this paper, we extend the PTAM to the 2DPTAM and use it to self-assemble squares rather than lines.

2 Preliminaries

In this section, we give definitions for the abstract Tile Assembly Model (aTAM) as well as probabilistic self-assembly in the aTAM, yielding a model that we call the 2DPTAM.

2.1 Formal description of the abstract Tile Assembly Model

Fix an alphabet Σ. Σ^* is the set of finite strings over Σ. Let $\mathbb{Z}$, $\mathbb{Z}^+$, and $\mathbb{N}$ denote the set of integers, positive integers, and nonnegative integers, respectively. Given $V \subseteq \mathbb{Z}^2$, the *full grid graph* of V is the undirected graph $G^{\mathrm{f}}_V = (V, E)$, and for all $\vec{x}, \vec{y} \in V$, $\{\vec{x}, \vec{y}\} \in E \iff \|\vec{x} - \vec{y}\| = 1$.

A *tile type* is a tuple $t \in (\Sigma^* \times \mathbb{N})^4$, that is, a unit square with four sides listed in some standardized order, and each side having a *glue* $g \in \Sigma^* \times \mathbb{N}$ consisting of a finite string *label* and nonnegative integer *strength*. We assume a finite set of tile types, usually denoted T, but an infinite number of copies of each tile type. Intuitively, a *tile* is a copy of a tile type that is placed at a point but technically it is a pair $(\vec{x}, t) \in \mathbb{Z}^2 \times T$.

A *configuration* is a possibly empty arrangement of tile types on the integer lattice $\mathbb{Z}^2$, i.e., a partial function $\alpha : \mathbb{Z}^2 \dashrightarrow T$, or, equivalently, a subset of $\mathbb{Z}^2 \times T$. We say that the tile $(\vec{x}, t)$ is *placed* by α if $\alpha(\vec{x}) = t$. Two adjacent tiles in a configuration *bind* if the glues on their opposing sides are equal (in both label and strength) and have positive strength. Each configuration α induces a *binding graph* G^{b}_α, a grid graph whose vertices are positions occupied by tile types, according to α, with an edge between two vertices if the tile types at those vertices bind.

An *assembly* is a connected, non-empty configuration, i.e., a partial function $\alpha : \mathbb{Z}^2 \dashrightarrow T$ such that $G^{\mathrm{f}}_{\mathrm{dom}\,\alpha}$ is connected and $\mathrm{dom}\,\alpha \neq \emptyset$. The *shape* $S_\alpha \subseteq \mathbb{Z}^2$ of α is $\mathrm{dom}\,\alpha$ and its size is $|\mathrm{dom}\,\alpha|$.

Given $\tau \in \mathbb{Z}^+$, α is *τ-stable* if every cut of G^{b}_α has weight at least τ, where the weight of an edge is the strength of the glue it represents. When τ is clear from the context, we say α is *stable*. Given two assemblies α and β, we say that α and β *agree* if, for all $\vec{x} \in \mathrm{dom}\,\alpha \cap \mathrm{dom}\,\beta$, $\alpha(\vec{x}) = \beta(\vec{x})$. Furthermore, we say that α is a *subassembly* of β, and we write $\alpha \sqsubseteq \beta$, if $S_\alpha \subseteq S_\beta$ and α and β agree.

A *tile assembly system* (TAS) is a triple $\mathcal{T} = (T, \sigma, \tau)$, where T is a tile set, $\sigma : \mathbb{Z}^2 \dashrightarrow T$ is the finite, τ-stable, *seed assembly*, and $\tau \in \mathbb{Z}^+$ is the *temperature*. We say that $\mathcal{T}$ is *singly-seeded* if $|\mathrm{dom}\,\sigma| = 1$. From this point on, unless explicitly stated otherwise, we assume that every TAS is singly-seeded, with $\mathrm{dom}\,\sigma = \{\vec{s}\}$, and has temperature 1.

Given two τ-stable assemblies α, β, we write $\alpha \to^{\mathcal{T}}_1 \beta$ if $\alpha \sqsubseteq \beta$ and $|S_\beta \setminus S_\alpha| = 1$. In this case we say α *$\mathcal{T}$-produces β in one step*. If $\alpha \to^{\mathcal{T}}_1 \beta$, $S_\beta \setminus S_\alpha = \{\vec{p}\}$, and $t = \beta(\vec{p})$, we write $\beta = \alpha + (\vec{p}, t)$ to denote the *tile attachment step* in which a tile of type t *attaches* to α at $\vec{p}$

to produce β. We use the notation $\beta\backslash\alpha$ to denote $(\vec{p},t)$, i.e., the tile that attaches to α to produce β in one step. If $\beta\backslash\alpha$ attaches to α with total strength 1 and the tile in α to which $\beta\backslash\alpha$ binds is located at point $\vec{p}\,'$, we use $\vec{u}_{\beta\backslash\alpha}$ to denote the unit vector $\vec{p}\,'-\vec{p}$.

The $\mathcal{T}$*-frontier* of α is the set $\partial^{\mathcal{T}}\alpha = \bigcup_{\alpha\to_1^{\mathcal{T}}\beta} S_\beta \setminus S_\alpha$, the set of empty locations at which a tile could stably attach to α. When $\mathcal{T}$ is clear from the context, we omit the superscript $\mathcal{T}$ on the $\mathcal{T}$-frontier of α. The *t-frontier* $\partial_t\alpha \subseteq \partial\alpha$ of α is the set $\left\{ \vec{p} \in \partial\alpha \;\middle|\; \alpha \to_1^{\mathcal{T}} \beta \text{ and } \beta(\vec{p}) = t \right\}$.

Let $\mathcal{A}^T$ denote the set of all assemblies of tiles from T, and let $\mathcal{A}^T_{<\infty}$ denote the set of finite assemblies of tiles from T. A sequence of $k \in \mathbb{Z}^+ \cup \{\infty\}$ assemblies $\alpha_1, \alpha_2, \ldots$ over $\mathcal{A}^T$ denoted as $\vec{\alpha} = (\alpha_i \mid 0 \le i-1 < k)$ is a $\mathcal{T}$*-assembly sequence* if, for all $1 \le i-1 < k$, $\alpha_{i-1} \to_1^{\mathcal{T}} \alpha_i$. We use this notation for defining the range of values for i because it allows us to handle both cases for k (i.e., finite or infinite) simultaneously. For integers $0 \le i-1 \le j-1 < k$, $\vec{\alpha}[i \ldots j]$ denotes the subsequence $(\alpha_i, \ldots, \alpha_j)$. When $\vec{\alpha}$ is finite, we say that $\vec{\alpha}$ *terminates* at $\vec{p} \in \mathbb{Z}^2$, where $\{\vec{p}\} = \text{dom } \alpha_1$ if $k=1$, and $\{\vec{p}\} = \text{dom } \alpha_k \backslash \text{dom } \alpha_{k-1}$ otherwise. The *result* of an assembly sequence $\vec{\alpha}$, written as $\text{res}\,(\vec{\alpha})$ is the unique assembly such that $\text{dom res}\,(\vec{\alpha}) = \bigcup_{i=1}^{k} \text{dom } \alpha_i$ and for all $0 \le i-1 < k$, $\alpha_i \sqsubseteq \text{res}\,(\vec{\alpha})$. In other words, $\text{res}\,(\vec{\alpha})$ is the unique limiting assembly, which for a finite sequence, is the final assembly in the sequence. If $\text{res}\,(\vec{\alpha}) = \alpha$ and $\vec{x} \in \text{dom } \alpha$, then $\text{index}_{\vec{\alpha}}\,(\vec{x}) = \min\{\, i \mid \vec{x} \in \text{dom } \alpha_i \,\}$.

If $\alpha = \text{res}\,(\vec{\alpha})$ is finite and, for some $t \in T$, $\vec{x} \in \partial_t\alpha$, then the notation $\vec{\alpha} + (\vec{x},t)$ denotes *appending the tile* $(\vec{x},t)$ *to* $\vec{\alpha}$, and is the assembly sequence $(\alpha_1, \ldots, \alpha_k, \alpha_{k+1})$, where $\alpha_{k+1} = \alpha_k + (\vec{x},t)$.

We write $\alpha \to^{\mathcal{T}} \beta$, and we say α $\mathcal{T}$*-produces* β (in 0 or more steps) if there is a $\mathcal{T}$-assembly sequence $(\alpha_1, \alpha_2, \ldots)$ of length $k = |S_\beta \setminus S_\alpha| + 1$ such that (1) $\alpha = \alpha_1$, (2) $S_\beta = \bigcup_{0 \le i-1 < k} S_{\alpha_i}$, and (3) for all $0 \le i-1 < k$, $\alpha_i \sqsubseteq \beta$. If k is finite then it is routine to verify that $\beta = \alpha_k$. We say α is $\mathcal{T}$*-producible* if $\sigma \to^{\mathcal{T}} \alpha$, and we write $\mathcal{A}[\mathcal{T}]$ to denote the set of $\mathcal{T}$-producible assemblies. We say that $\vec{\alpha}$ is $\mathcal{T}$-producing if $\vec{\alpha} = (\sigma, \ldots)$ and $\text{res}\,(\vec{\alpha}) \in \mathcal{A}[\mathcal{T}]$. An assembly α is $\mathcal{T}$*-terminal* if α is τ-stable and $\partial^{\mathcal{T}}\alpha = \emptyset$. We write $\mathcal{A}_\square[\mathcal{T}] \subseteq \mathcal{A}[\mathcal{T}]$ to denote the set of $\mathcal{T}$-producible, $\mathcal{T}$-terminal assemblies.

2.2 The two-dimensional probabilistic aTAM or 2DPTAM

We model probabilistic self-assembly in the aTAM with non-empty, rooted, directed, weighted tree structures in which all edges point away from the root and, for each node v, S_v denotes the sum of all the numerical weights over all the outgoing edges on v. If v is a leaf node, then $S_v = 0$. We say that a non-empty, rooted, weighted tree $\mathcal{Q}$ is a *sub-probability tree* (SPT) if:

1. all edges of $\mathcal{Q}$ are labeled with some *probability* $p \in (0,1]$, and
2. for each node v in $\mathcal{Q}$, $S_v \le 1$.

If $\mathcal{Q}$ is an SPT and v is a node of $\mathcal{Q}$, then we say v is *normalized* in $\mathcal{Q}$ if $S_v = 1$. If all internal nodes of an SPT $\mathcal{Q}$ are normalized, then $\mathcal{Q}$ is a discrete time Markov chain, or simply Markov chain. For an SPT $\mathcal{Q}$, if v is a node of $\mathcal{Q}$, then we define the *probability* of v, denoted as $\Pr_{\mathcal{Q}}[v]$, as follows:

1. If v is the root node of $\mathcal{Q}$, then $\Pr_{\mathcal{Q}}[v] = 1$.
2. Otherwise, if u is the parent of v and p is the probability of the edge from u to v, then $\Pr_{\mathcal{Q}}[v] = \Pr_{\mathcal{Q}}[u] \cdot p$.

Let $m \in \mathbb{Z}^+ \cup \{\infty\}$, $\pi = (q_i \mid 0 \le i-1 < m)$ be a sequence of nodes in $\mathcal{Q}$ where q_1 is the root of $\mathcal{Q}$, for all $1 \le i-1 < m$, there is an edge from q_{i-1} to q_i in $\mathcal{Q}$, and p_i is the probability on this edge. We say that π is a *maximal path in* $\mathcal{Q}$ if $m = \infty$, or $m \in \mathbb{Z}^+$ and q_m is a leaf node of $\mathcal{Q}$. If $m = 1$, then $\Pr_{\mathcal{Q}}[\pi] = 1$. Otherwise, if $m > 1$, then we define

$\Pr_{\mathcal{Q}}[\pi] = \prod_{\{i \in \mathbb{Z}^+ \mid 1 \leq i-1 < m\}} p_i$. Note that if $m = \infty$, then π is a ray in $\mathcal{Q}$, starting from the root. We define the *probability of SPT* $\mathcal{Q}$ to be $\Pr[\mathcal{Q}] = \sum_{\pi \text{ maximal path in } \mathcal{Q}} \Pr_{\mathcal{Q}}[\pi]$.

If $\mathcal{Q}$ is a rooted tree and u is a node in $\mathcal{Q}$, then $\mathcal{Q}^u$ is the subtree of $\mathcal{Q}$ that contains u, along with all descendants of u in $\mathcal{Q}$. Note that, if $\mathcal{Q}$ is an SPT, then $\mathcal{Q}^u$ is also an SPT. We say that any subtree $\mathcal{Q}'$ of $\mathcal{Q}$ is *full relative to* $\mathcal{Q}$ if, for every internal node v of $\mathcal{Q}'$, all children of v in $\mathcal{Q}$ are children of v in $\mathcal{Q}'$. We say $\mathcal{Q}$ is *finitely branching* (or *locally finite)* if, for every internal node v of $\mathcal{Q}$, there exists $c \in \mathbb{Z}^+$ such that v has c children in $\mathcal{Q}$.

Assume $\pi = (x_i \mid 0 \leq i-1 < m)$ for $m \in \mathbb{Z}^+ \cup \{\infty\}$ denotes a simple path whose elements may be, for example, either points in $\mathbb{Z}^2$ or nodes in an SPT. Then $\mathrm{dom}\, \pi$ denotes the set of elements $\{x_i \mid i \geq 1\}$, $|\pi| = m$ denotes the length of π, $\pi[i]$ denotes x_i for $i \geq 1$, and, for integers $0 \leq i-1 \leq j-1 < m$, $\pi[i \ldots j]$ denotes the subsequence $(x_i, \ldots, x_j)$. Any subsequence $\pi[1 \ldots j]$ for integers $j \geq 1$ is called a *prefix* of π, whereas, assuming π is a finite path, any subsequence $\pi[i \ldots m]$ for integers $1 \leq i \leq m$ is called a *suffix* of π.

We study probabilistic self-assembly in the aTAM using a simplified version of the tile concentration programming model, which we call the 2DPTAM, where tiles are assumed to have uniform concentration (see [7] for a formal development of the general model). We model probabilistic self-assembly of a TAS $\mathcal{T} = (T, \sigma, 1)$ with the SPT $\mathcal{M}_{\mathcal{T}}$, where:

1. the root node of $\mathcal{M}_{\mathcal{T}}$ is the $\mathcal{T}$-assembly sequence (σ),
2. there is a one-to-one correspondence between nodes in $\mathcal{M}_{\mathcal{T}}$ and finite $\mathcal{T}$-producing assembly sequences,
3. there is an edge in $\mathcal{M}_{\mathcal{T}}$ from $\vec{\alpha}$ to $\vec{\beta}$, assuming $\mathrm{res}\left(\vec{\alpha}\right) = \alpha$ and $\mathrm{res}\left(\vec{\beta}\right) = \beta$, if $\alpha \to_1^{\mathcal{T}} \beta$, and
4. the probability of the edge in $\mathcal{M}_{\mathcal{T}}$ from $\vec{\alpha}$ to $\vec{\beta}$, assuming $\mathrm{res}\left(\vec{\alpha}\right) = \alpha$ and $\mathrm{res}\left(\vec{\beta}\right) = \beta$, is $\frac{1}{M_\alpha}$, where $M_\alpha = \left|\left\{\gamma \,\middle|\, \alpha \to_1^{\mathcal{T}} \gamma\right\}\right| > 0$, since $\alpha \to_1^{\mathcal{T}} \beta$.

Note that $\mathcal{M}_{\mathcal{T}}$ is a finitely branching SPT.

We define the probability of a $\mathcal{T}$-producible assembly in terms of the $\mathcal{T}$-producing assembly sequences from which it may result. Since there is a one-to-one correspondence between nodes of $\mathcal{M}_{\mathcal{T}}$ and finite $\mathcal{T}$-producing assembly sequences, for $\alpha \in \mathcal{A}[\mathcal{T}]$ we define the *probability* of α as: $\Pr_{\mathcal{T}}[\alpha] = \sum_{\substack{\vec{\alpha} \text{ is a } \mathcal{T}\text{-producing assembly sequence} \\ \text{such that } \mathrm{res}(\vec{\alpha}) = \alpha}} \Pr_{\mathcal{M}_{\mathcal{T}}}[\vec{\alpha}]$.

We say that $\mathcal{T}$ *strictly self-assembles a shape* $X \subseteq \mathbb{Z}^2$ with probability at least $p \in [0,1]$ if $\sum_{\substack{\alpha \in \mathcal{A}_{\Box}[\mathcal{T}] \text{ such} \\ \text{that } \mathrm{dom}\, \alpha = X}} \Pr_{\mathcal{T}}[\alpha] \geq p$. We say that X *strictly self-assembles* with probability at least p if there exists some TAS that strictly self-assembles it with probability at least p.

3 Defining sequential non-determinism

In this section, we give a formulation of sequential non-determinism.

Going forward, assume $\mathcal{T} = (T, \sigma, 1)$ is a singly-seeded TAS with $\mathrm{dom}\, \sigma = \{\vec{s}\}$.

Note that we use boxes like the one above to state assumptions that hold by default, unless otherwise stated. This practice allows us to significantly shorten the statement of several lemmas that share an increasingly long sequence of assumptions.

▶ **Definition 1.** *Let $\vec{x}, \vec{y} \in \mathbb{Z}^2$ with $\vec{x} \neq \vec{y}$, and π and π' be finite simple paths in $G^{\mathrm{f}}_{\mathbb{Z}^2}$ from $\vec{x}$ to $\vec{y}$. We say that π and π' are* competing for $\vec{y}$, from $\vec{x}$ in $\mathcal{T}$, *if all of the following conditions hold.*

1. $\pi \neq \pi'$
2. $\mathrm{dom}\ \pi \cap \mathrm{dom}\ \pi' = \{\vec{x}, \vec{y}\}$
3. *There exist a path $p \in \{\pi, \pi'\}$ and an assembly $\alpha \in \mathcal{A}[\mathcal{T}]$ such that p is a simple path in G^b_α.*
4. *For all $\alpha \in \mathcal{A}[\mathcal{T}]$, each path $p \in \{\pi, \pi'\}$ and all integers $1 \leq l < |p|$, if p' is any finite simple path from $\vec{s}$ to $p[l]$ in G^b_α, then:*
   a. *if p' goes through $\vec{x} = p[1]$, then $p[1 \dots l]$ is a suffix of p',*
   b. *otherwise p' is a prefix of every finite simple path from $\vec{s}$ to $p[1]$ in G^b_α.*
5. *For all $\alpha \in \mathcal{A}[\mathcal{T}]$, if p' is any simple path from $\vec{s}$ to $\vec{y}$ in G^b_α, then there exists a path $p \in \{\pi, \pi'\}$ such that p is a suffix of p'.*
6. *For all $\alpha \in \mathcal{A}[\mathcal{T}]$ such that $\vec{x} \in \mathrm{dom}\ \alpha$, the following conditions hold.*
   a. *For each path $p \in \{\pi, \pi'\}$, there exists a $\mathcal{T}$-assembly sequence $\vec{\beta} = (\beta_i \mid 1 \leq i \leq k)$ such that $\beta_1 = \{(\vec{x}, \alpha(\vec{x}))\}$ and $\mathrm{dom}\ \beta_k = \mathrm{dom}\ p$.*
   b. *For all $\mathcal{T}$-assembly sequences $\vec{\beta} = (\beta_i \mid 1 \leq i \leq k)$, if $\beta_1 = \{(\vec{x}, \alpha(\vec{x}))\}$ and p is a simple path from $\vec{x}$ to $\vec{y}$ in $G^b_{\beta_k}$, then $p \in \{\pi, \pi'\}$.*

We call p a competing path *if there exist $\vec{x}, \vec{y} \in \mathbb{Z}^2$ and paths π and π' that are competing for $\vec{y}$ from $\vec{x}$ such that $p \in \{\pi, \pi'\}$.*

Conditions 1 and 2 together mean that the two competing paths only share their starting and ending points and that their domains are not identical. Condition 3 means that there exists a $\mathcal{T}$-producing assembly sequence that results in an assembly that places tiles on all the points in π or π'. Condition 4a means that, if the binding graph G of a $\mathcal{T}$-producible assembly contains a simple path p from $\vec{s}$ to any non-ending point on one of the competing paths such that p goes through $\vec{x}$, then p must first go through $\vec{x}$ and follow π or π'. Condition 4b means that, if the binding graph G of a $\mathcal{T}$-producible assembly contains a simple path p from $\vec{s}$ to any non-ending point on one of the competing paths such that p does not go through $\vec{x}$, then p must be a prefix of all paths from $\vec{s}$ to $\vec{x}$ in G. Condition 5 means that all $\mathcal{T}$-assembly sequences that follow a path from $\vec{s}$ to $\vec{y}$ must go through $\vec{x}$ and then follow either π or π'. Condition 6a means that it is possible for either competing path to self-assemble in isolation in a $\mathcal{T}$-assembly sequence starting from any tile that is placed at $\vec{x}$ by some $\mathcal{T}$-producible assembly. Condition 6b means that, if there is a $\mathcal{T}$-assembly sequence that starts from any tile that is placed at $\vec{x}$ by some $\mathcal{T}$-producible assembly and follows a simple path to $\vec{y}$, then this simple path must be one of the two competing paths π or π'.

▶ **Definition 2.** *Let $\vec{y} \in \mathbb{Z}^2$. We call the point $\vec{y}$ a* point of competition*, or* POC *for short, in $\mathcal{T}$, if there exist $\vec{x} \in \mathbb{Z}^2$ and simple paths π and π' in $G^f_{\mathbb{Z}^2}$ that are competing for $\vec{y}$, from $\vec{x}$ in $\mathcal{T}$. We say that π and π' are the* competing paths that correspond to $\vec{y}$*, and $\vec{x}$ is the corresponding starting point.*

Going forward, let $r \in \mathbb{Z}^+$ and assume $\mathcal{T}$ has a non-empty, possibly infinite set P of POCs with $\|P\| \geq r$.

▶ **Definition 3.** *Let $Y = \{\vec{y}_1, \dots, \vec{y}_r\}$ be an r-element subset of P. A* winner function *(for $\mathcal{T}$) is a function of the form $w : Y \to T$. Having defined w, we call the set Y the set of* essential *POCs in $\mathcal{T}$.*

Going forward, assume $Y = \{\vec{y}_1, \ldots, \vec{y}_r\}$ is the r-element set of essential POCs in $\mathcal{T}$ and $S_Y = \{\vec{x}_1, \ldots, \vec{x}_r\}$ is the corresponding set of starting points.

For example, a TAS could be designed to strictly self-assemble an $N \times N$ square with high probability such that its set of essential POCs is a particular subset of $\{0, \ldots, N-1\}^2$. However, this TAS may induce infinite, erroneous $\mathcal{T}$-assembly sequences that attach tiles at infinitely many POCs. The motivation behind Definition 3 is that $\mathcal{T}$ may produce multiple (possibly infinite) terminal assemblies and we use w to determine the correctness of such an assembly in the following sense.

▶ **Definition 4.** *Let $\alpha \in \mathcal{A}[\mathcal{T}]$ and $w : Y \to T$ be a winner function for $\mathcal{T}$. If:*

1. *for all $\vec{p} \in \text{dom }\alpha \cap Y$, $\alpha(\vec{p}) = w(\vec{p})$ and*
2. $\text{dom }\alpha \cap (P \backslash Y) = \emptyset$,

*then we say that α is a w-*correct $\mathcal{T}$-assembly*. If $\vec{\alpha}$ is any $\mathcal{T}$-assembly sequence such that $res(\vec{\alpha}) = \alpha$, then we say that $\vec{\alpha}$ is a w-*correct $\mathcal{T}$-assembly sequence*.*

Condition 1 of this definition means that α may only place correct tiles at the essential POCs included in its domain. Condition 2 means that a tile cannot attach to an inessential POC unless it is attaching to a $\mathcal{T}$-producible assembly that is not w-correct.

▶ **Definition 5.** *Let $\vec{p} \in \mathbb{Z}^2$, $\alpha, \alpha', \beta, \beta' \in \mathcal{A}[\mathcal{T}]$ and $t_\alpha, t_\beta \in T$ such that $\alpha' = \alpha + (\vec{p}, t_\alpha)$ and $\beta' = \beta + (\vec{p}, t_\beta)$ are valid tile attachment steps. We say that $\vec{p}$ is a* directionally deterministic *point in $\mathcal{T}$ if either $\vec{p} = \vec{s}$, or the following two conditions hold:*

1. *If $\vec{p} \in P$, then $t_\alpha \neq t_\beta$ implies at least one of the following conditions hold:*
   - **a.** *$\vec{u}_{\alpha' \backslash \alpha} \neq \vec{u}_{\beta' \backslash \beta}$ or*
   - **b.** *$\alpha\left(\vec{p} + \vec{u}_{\alpha' \backslash \alpha}\right) \neq \beta\left(\vec{p} + \vec{u}_{\beta' \backslash \beta}\right)$.*
2. *If $\vec{p} \in \mathbb{Z}^2 \backslash P$, then $t_\alpha \neq t_\beta$ implies that α and β do not agree.*

Condition 2 of this definition means that, when two $\mathcal{T}$-producible assemblies agree, no more than one tile may attach at any non-POC that belongs to the intersection of their frontiers. Condition 1 says that, if any two distinct tiles ever attach to any $\mathcal{T}$-producible assemblies at any POC $\vec{p}$, then these tiles must attach via glues on different sides or bind to different tile types. The TAS shown in Figure 1 produces two terminal assemblies, respectively placing the 3 tile at the point $\vec{p}$ in the bottom-left corner, as depicted in Figure 1(b), or placing the 4 tile at the same point, as depicted in Figure 1(c). Note that, in this example, the 3 tile binds only via its north glue, whereas the 4 tile binds only via its east glue. Therefore, the point $\vec{p}$ is directionally deterministic in this TAS.

▶ **Definition 6.** *We say that $\mathcal{T}$ is a* directionally deterministic *TAS if all points in the domains of all $\mathcal{T}$-producible assemblies are directionally deterministic.*

The next lemma says that if $\mathcal{T}$ is directionally deterministic and α is a w-correct $\mathcal{T}$-terminal assembly, then any w-correct $\mathcal{T}$-producing assembly sequence must result in a subassembly of α.

▶ **Lemma 7.** *Let $w : Y \to T$ be a winner function for $\mathcal{T}$ and $\alpha \in \mathcal{A}_\Box[\mathcal{T}]$ be w-correct. If $\mathcal{T}$ is directionally deterministic and β is any w-correct, $\mathcal{T}$-producible assembly, then $\beta \sqsubseteq \alpha$.*

▶ **Corollary 8.** *If $\mathcal{T}$ is directionally deterministic and there exists a w-correct assembly $\alpha \in \mathcal{A}_\Box[\mathcal{T}]$, then α is unique.*

Going forward, assume $w : Y \to T$ is a winner function for $\mathcal{T}$.

The following definition is the cornerstone of the framework we introduce in this paper. For the sake of completeness, we explicitly state all of the assumptions that were instantiated in previous "Going forward" boxes.

▶ **Definition 9.** *Assume $\mathcal{T} = (T, \sigma, 1)$ is a singly-seeded TAS with $\mathrm{dom}\ \sigma = \{\vec{s}\}$, $r \in \mathbb{Z}^+$, $\mathcal{T}$ has a non-empty, possibly infinite set P of POCs with $|P| \geq r$, S_P denotes the set $\{\vec{x} \mid \vec{x}$ is the starting point for some $\vec{y} \in P\}$, $Y = \{\vec{y}_1, \ldots \vec{y}_r\}$ is the r-element set of essential POCs in $\mathcal{T}$ with corresponding set of starting points $S_Y = \{\vec{x}_1, \ldots, \vec{x}_r\}$, and $w : Y \to T$ is a winner function for $\mathcal{T}$ such that α is the unique w-correct $\mathcal{T}$-terminal assembly with $Y \subseteq \mathrm{dom}\ \alpha$. We say that $\mathcal{T}$ is w-*sequentially non-deterministic *if all of the following conditions hold:*

1. *$\mathcal{T}$ is directionally deterministic,*
2. *for all $\mathcal{T}$-producing assembly sequences $\vec{\alpha}$, if $Y \subseteq \mathrm{dom}\ res(\vec{\alpha})$, then for all integers $1 \leq i < r$, $index_{\vec{\alpha}}(\vec{y}_i) < index_{\vec{\alpha}}(\vec{y}_{i+1})$, and*
3. *$S_P \cap P = \emptyset$.*

Figure 5 depicts visually and explains informally the assumptions and implications of Definition 9.

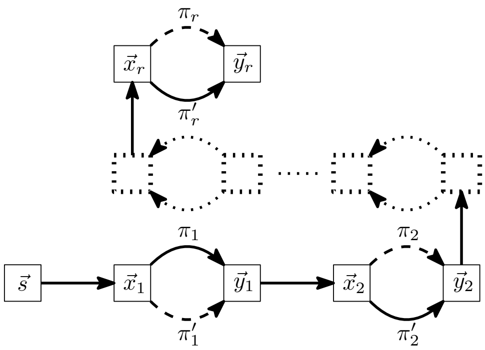

Figure 5 In this example, $\mathcal{T} = (T, \sigma, 1)$ is a w-sequentially non-deterministic TAS with $\mathrm{dom}\ \sigma = \{\vec{s}\}$, $r > 2$, and α is the unique w-correct, $\mathcal{T}$-terminal assembly whose domain contains all of the essential POCs. Note that each arrow represents a sequence of tile attachment steps that follow a simple path. Definition 9 either requires or implies that:
a) [Condition 1] every point at which a tile attaches in the assembly sequence is directionally deterministic,
b) [Assumption in Definition 9] the $\vec{y}_i$ points are all the essential POCs in $\mathcal{T}$,
c) [Condition 2] in all $\mathcal{T}$-assembly sequences whose results place tiles at every $\vec{y}_i$, the tiles always attach at the $\vec{y}_i$ points in order of increasing i values,
d) [Implied] in all $\mathcal{T}$-producing assembly sequences whose results place tiles at every $\vec{y}_i$, tiles also attach at all of the $\vec{x}_i$ points, and tile attachment steps always occur in the order $\vec{x}_1$, $\vec{y}_1$, $\vec{x}_2$, $\vec{y}_2$, …, $\vec{x}_r$, $\vec{y}_r$, and
e) [Condition 3] none of the $\vec{x}_i$ points is a point of competition.
The solid arrows in the figure depict a $\mathcal{T}$-producible assembly that is the result of a $\mathcal{T}$-producing assembly sequence that attaches tiles at points along a simple path in $\mathbb{Z}^2$ and in the following, required order: from $\vec{s}$ (the seed's point) to $\vec{x}_1$, then to $\vec{y}_1$ via π_1, then to $\vec{x}_2$, then to $\vec{y}_2$ via π'_2, …, to $\vec{x}_r$, and finally to $\vec{y}_r$ via π'_r.

Going forward, assume $\mathcal{T}$ is w-sequentially non-deterministic, with corresponding unique w-correct $\mathcal{T}$-terminal assembly α.

In the following, we define the TAS that corresponds to a pair of paths competing for a POC.

▶ **Definition 10.** *For all integers $1 \leq i \leq r$, let $\mathcal{C}_i = (T, \sigma_i, 1)$ be the TAS with $\sigma_i = \{(\vec{x}_i, \alpha(\vec{x}_i))\}$. We say that $\mathcal{C}_i$ is* competition i *in $\mathcal{T}$ associated with POC $\vec{y}_i$.*

Note that Definition 10 defines each competition $\mathcal{C}$ as the TAS $\mathcal{T}$ with a different seed tile. This allows us to reason about the behavior of $\mathcal{C}$ in the absence of any other tiles, such as, e.g., a simple path of tiles from the seed of $\mathcal{T}$ to the starting point of $\mathcal{C}$ that may or may not block one of the competing paths of $\mathcal{C}$. Importantly, Definition 10 does not restrict the set of $\mathcal{C}$-producible assemblies to only those that place no tiles outside the domains of the associated competing paths. $\mathcal{T}$. However, in the following definition, we will characterize a subset of "winning" $\mathcal{C}$-producible assemblies that place tiles on every point in the winning path and possibly a prefix of the losing path. We will later define (see Definition 14) the probability of $\mathcal{C}$ only considering $\mathcal{C}$-producing assembly sequences that not only result in a "winning" $\mathcal{C}$-producible assembly but are also comprised of tile attachment steps that faithfully follow the competing paths.

▶ **Definition 11.** *Let $\vec{y} \in Y$ with corresponding competing paths π and π', $\mathcal{C}$ be the competition in $\mathcal{T}$ associated with $\vec{y}$, and $\gamma \in \mathcal{A}[\mathcal{C}]$ such that:*

1. *π is a path in G^b_γ,*
2. *$\gamma \sqsubseteq \alpha$, and*
3. *every simple path in G^b_γ that starts at $\vec{x}$ is a prefix of either π or π'.*

We say that π is the winning path for $\mathcal{C}$, *π' is the* losing path for $\mathcal{C}$, *and that γ is a* winning assembly *for $\mathcal{C}$. Let $W_\mathcal{C}$ denote the set of all winning assemblies for $\mathcal{C}$. Any $\mathcal{C}$-producing assembly sequence $\vec{\gamma}$ that terminates at $\vec{y}$ and whose result is in $W_\mathcal{C}$ is a* winning assembly sequence *for $\mathcal{C}$.*

Going forward, assume $\vec{y} \in Y$ with corresponding starting point $\vec{x} \in S_Y$, $\vec{\beta}_{\vec{x}}$ is a $\mathcal{T}$-producing sequence that terminates at $\vec{x}$ and whose result $\beta_{\vec{x}}$ is a subassembly of α, and $\mathcal{C}$ is the competition in $\mathcal{T}$ associated with $\vec{y}$ with corresponding winning and losing paths π_{win} and π_{lose}, respectively.

The following definition describes the notion that a competition need not be "fair" relative to a starting assembly.

▶ **Definition 12.** *We say that $\mathcal{C}$ is rigged by $\vec{\beta}_{\vec{x}}$ if there exists an integer $1 < l < |\pi_{lose}|$ such that:*

1. *$\pi_{lose}[l] \in \operatorname{dom} \beta_{\vec{x}}$ and*
2. *every simple path from $\vec{s}$ to $\vec{x}$ in $G^b_{\beta_{\vec{x}}}$ goes through $\pi_{lose}[l]$.*

▶ **Definition 13.** *We say that $\mathcal{C}$ is* rigged in $\mathcal{T}$ *if it is rigged by every $\mathcal{T}$-producing sequence that terminates at $\vec{x}$ and whose result is a subassembly of α.*

We now define the probability of a competition as the sum of the probabilities of all of its winning assembly sequences. The probability of a winning assembly sequence of the competition is defined relative to a modified version of its corresponding SPT where certain nodes are removed to simplify the computation of the probability of the winning assembly sequence.

▶ **Definition 14.** *Let $\vec{y} \in Y$ with corresponding starting point $\vec{x}$ and competing paths π_{win} and π_{lose}. Let $\mathcal{C}$ be the competition in $\mathcal{T}$ associated with $\vec{y}$ and $\mathcal{M}'_{\mathcal{C}}$ be the SPT obtained from $\mathcal{M}_{\mathcal{C}}$ by removing all nodes $\vec{\beta}$ of $\mathcal{M}_{\mathcal{C}}$ such that the binding graph of res $\left(\vec{\beta}\right)$ contains a simple path that starts at $\vec{x}$ and is not a prefix of either π_{win} or π_{lose}. If $\vec{\gamma}$ is any winning assembly sequence for $\mathcal{C}$, we define the* probability of $\vec{\gamma}$ *as* $Pr_{\mathcal{C}}[\vec{\gamma}] = Pr_{\mathcal{M}'_{\mathcal{C}}}[\vec{\gamma}]$, *and the* probability of competition $\mathcal{C}$ *as* $Pr[\mathcal{C}] = \sum_{\vec{\gamma} \text{ winning assembly sequence for } \mathcal{C}} Pr_{\mathcal{C}}[\vec{\gamma}]$.

Intuitively, $Pr[\mathcal{C}]$ is defined so that it can be computed easily, i.e., without having to account for potentially numerous points at which tiles could attach outside of the competing paths. The next lemma first counts the number of winning assembly sequences for a competition that is not rigged that all result in a certain winning assembly and then derives the probability of each such winning assembly sequence.

▶ **Lemma 15.** *Let $\vec{y} \in Y$ with corresponding competing paths π_{win} and π_{lose} of length $l \geq 2$ and $l' \geq 2$, respectively. Let $\mathcal{C}$ be the competition associated with $\vec{y}$ in $\mathcal{T}$. For any integer $0 \leq i \leq l' - 2$, let π_i denote the (possibly empty) simple path $\pi_{lose}[2 \ldots (i+1)]$ and γ denote the winning assembly in $W_{\mathcal{C}}$ whose domain equals* dom $\pi_{win} \cup$ dom π_i. *If $\mathcal{C}$ is not rigged in $\mathcal{T}$, then:*

1. *The number of winning assembly sequences for $\mathcal{C}$ with result γ is $\binom{l+i-2}{l-2}$.*
2. *If $\vec{\gamma}$ is a winning assembly sequence for $\mathcal{C}$, with result γ, then $Pr_{\mathcal{C}}[\vec{\gamma}] = \frac{1}{2^{|\vec{\gamma}|-1}} = \frac{1}{2^{l+i-1}}$.*

▶ **Corollary 16.** $Pr[\mathcal{C}] = \sum_{i=0}^{l'-2} \binom{l+i-2}{l-2} \cdot \left(\frac{1}{2}\right)^{l+i-1}$.

The following is the first main result of this paper, which states that the probability that the w-sequentially non-deterministic TAS $\mathcal{T}$ strictly self-assembles the shape of the unique α is at least the product of the probabilities of all of its competitions, which themselves can be easily computed using our framework.

▶ **Theorem 17.** *If $\mathcal{C}_1, \ldots, \mathcal{C}_r$ are the competitions in $\mathcal{T}$, then $\mathcal{T}$ strictly self-assembles* dom α *with probability at least* $\prod_{i=1}^{r} Pr[\mathcal{C}_i]$.

4 Efficient, sequentially non-deterministic self-assembly of an $N \times N$ square with high probability

In this section, we state and sketch the proof of our main "constructive theorem" namely that, for every $N \in \mathbb{Z}^+$ and real $\delta > 0$, there exists a TAS that strictly self-assembles an $N \times N$ square with probability at least $1 - \delta$ and uses only $O\left(\log N + \log \frac{1}{\delta}\right)$ tile types.

Our construction uses three modified copies of a single template for a zigzag binary counter. This counter template is K tiles wide and n tiles high, with $n \leq N$. It is built out of "gadgets", which are groups of tiles that self-assemble within a fixed region of space to accomplish a specific task, e.g., non-deterministically read (i.e., guess) a single unknown, geometrically-represented binary value. Since our counter assembles in a zigzag pattern, each one of its values is first incremented, going from right to left, and then copied from left to right. Since one row of tiles is used to hardcode the starting value of the counter and all of the copy and increment gadgets are h tiles high, with $h = \left\lceil 2\log N + \log \frac{1}{\delta}\right\rceil + 5$, the number of values taken on by the counter is equal to $m = \left\lfloor \frac{n-1}{2h} \right\rfloor$. Then $e = (n-1) \mod 2h$ is the number of extra tiles (i.e., rows) needed for the height of our counter template to be exactly equal to n. The counter template is P-shaped. The stem of the P is an upside-down L whose domain we denote by $L_e = \{0\} \times \{0, \ldots, e-1\} \cup \{(1, e-1), (2, e-1)\}$ and depict in dark gray in Figure 6. The "loop" of the P is a $K \times (n-e)$ rectangular counter whose domain we denote by $R_{K,n,e} = \{0, \ldots, K-1\} \times \{e, \ldots, n-1\}$ and depict in light gray in Figure 6.

The last parameters we need pertain to the binary representation of the counter's values. We use k to denote the number of bits in the binary expansion of m. Each bit value will be geometrically encoded in a 6-tile-wide gadget corresponding to a competition. Since two tiles are added on each side of the counter's value and each bit is actually encoded by two gadgets, namely the bit's value itself and an additional indicator bit used to distinguish the leftmost and rightmost value bits from the other ones, K is equal to $12k + 4$. Finally, the starting value of our counter is set to $s = 2^{\lfloor \log m \rfloor + 1} - m$.

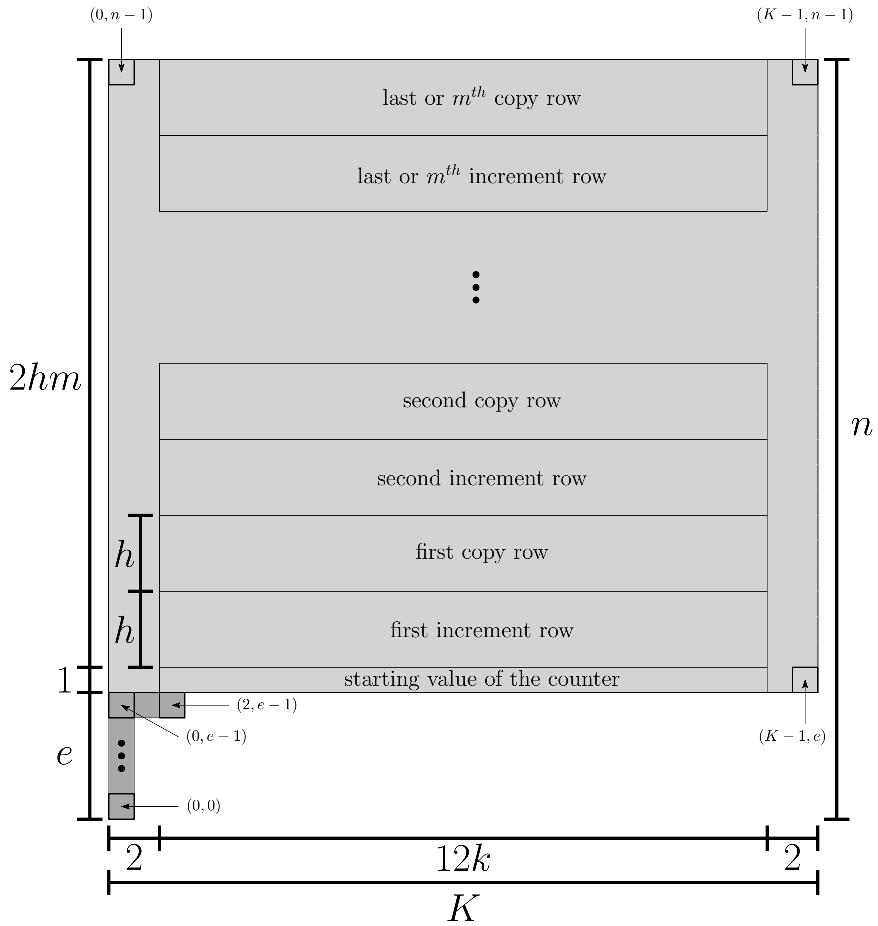

Figure 6 Shape of, and parameters for, our counter template.

To prove the main constructive theorem of this section, we will use the following lemma:

▶ **Lemma 18.** *Let $N \in \mathbb{Z}^+$, $\delta \in \mathbb{R}^+$ such that $0 < \delta < 1$, $h = \lceil 2 \log N + \log \frac{1}{\delta} \rceil + 5$, $n \in \mathbb{Z}^+$ such that $n \leq N$, and $m = \left\lfloor \frac{n-1}{2h} \right\rfloor$. Then there exists a TAS $\mathcal{T} = (T, \sigma, 1)$ satisfying the following conditions:*

1. *If $m > 0$, $s = 2^{\lfloor \log m \rfloor + 1} - m$, k is the number of bits in the binary expansion of m, $K = 12k + 4$, and $e = (n-1) \mod 2h$, then there exists a finite set $P \subseteq \{0, \ldots K-1\} \times \{0, \ldots, n-1\}$, a finite set $Y \subseteq P$ and $w : Y \to T$ such that:*

a. *P is the set of POCs in $\mathcal{T}$, Y is the set of essential POCs in $\mathcal{T}$ and $Y = P$,*
b. *$\mathcal{T}$ is w-sequentially non-deterministic,*
c. *α is the unique w-correct $\mathcal{T}$-terminal assembly and* $\text{dom}\,\alpha = L_e \cup R_{K,n,e}$*,*
d. *in every w-correct $\mathcal{T}$-producing assembly sequence resulting in α, no tile placed at any point on the perimeter of α has a positive-strength glue on its outward-facing side, the last one among α's perimeter points at which a tile is placed is $(K-1, n-1)$, and this point is the only one in* $\text{dom}\,\alpha$ *where any tile of this type is placed.*
e. *for every competition $\mathcal{C}$ of $\mathcal{T}$, if π and π' are the corresponding winning and losing paths, respectively, then $|\pi| = 4$ and $|\pi'| = 4h - 10$ (or vice versa),*

2. *$\mathcal{T}$ strictly self-assembles* $\text{dom}\,\alpha = L_e \cup R_{K,n,e}$ *with probability at least $1 - \delta$, and*
3. $|T| = O\left(\log N + \log \frac{1}{\delta}\right)$.

The reason we set h so high is to allow the P-shaped assembly produced by the TAS in Lemma 18 to be used in a larger construction, e.g., one that strictly self-assembles an $N \times N$ square, which will require a higher value for h in order to guarantee a sufficient probability of correctness. We chose the P-shape of $\text{dom}\,\alpha$ to be able to invoke Lemma 18 multiple times when assembling an $N \times N$ square.

Proof sketch. To prove Lemma 18, we proceed as follows:

1. Prove that the lemma holds in the corner cases when n is too small, i.e., $n < 13$, or when $m = 0$, i.e., when h is too large for a given value of even a sufficiently large n. Assume going forward that $n \geq 13$ and $m > 0$.
2. Define T in terms of "gadgets" of tiles and compute an asymptotic upper bound on its cardinality. The tile types for T will simulate a zigzag binary counter that is initialized with the starting value s, and then alternates between increment steps in which the counter value is incremented by 1 going from right to left, followed by copy steps in which the counter value is copied going from left to right. The counter stops after a copy step in which the copied value is all 0s.
   a. Create one seed gadget of size $O(e)$, which contains the unique seed tile of $\mathcal{T}$.
   b. Create $O(k)$ gadgets, each of size $O(1)$, to initialize the counter.
   c. Create $O(1)$ gadgets, each of size $O(h)$, to increment the counter.
   d. Create $O(1)$ gadgets, each of size $O(h)$, to copy the value of the counter.
   e. Consolidate the outomes of the four previous gadget creation steps to infer that $|T| = O\left(\log N + \log \frac{1}{\delta}\right)$.
3. Define P, Y and w.
4. Prove that $\mathcal{T}$ is w-sequentially non-deterministic such that α is the unique w-correct $\mathcal{T}$-terminal assembly and $\text{dom}\,\alpha = L_e \cup R_{K,n,e}$.
5. Prove that every w-correct $\mathcal{T}$-producing assembly sequence resulting in α attaches the last tile at the upper-right corner of α and that no tile on the perimeter of α has an outward-facing positive-strength glue.
6. Prove that $\mathcal{T}$ strictly self-assembles $\text{dom}\,\alpha$ with probability at least $1 - \delta$.
   a. Use Corollary 16 to prove that, for every competition $\mathcal{C}$ of $\mathcal{T}$, $\Pr[\mathcal{C}] \geq 1 - \frac{\delta}{N^2}$.
   b. Apply Theorem 17 and use Bernoulli's inequality to bound the product of the probabilities of the $2km \leq N^2$ essential competitions of $\mathcal{T}$. ◀

Lemma 18 can be used to prove the following theorem:

▶ **Theorem 19.** *For each $N \in \mathbb{Z}^+$ and real δ such that $0 < \delta < 1$, there exists a TAS $\mathcal{T}_{N,\delta} = (T_{N,\delta}, \sigma_{N,\delta}, 1)$ that strictly self-assembles $\{0, \ldots, N-1\}^2$ with probability at least $1 - \delta$ and such that $|T_{N,\delta}| = O\left(\log N + \log \frac{1}{\delta}\right)$.*

The construction of $\mathcal{T}_{N,\delta}$ follows from three invocations of Lemma 18, a high-level depiction of which is shown in Figure 7.

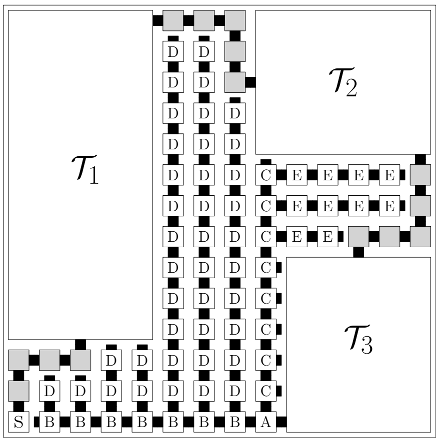

Figure 7 Construction of $\mathcal{T}_{N,\delta}$ via three invocations of Lemma 18 resulting in $\mathcal{T}_1$, $\mathcal{T}_2$ and $\mathcal{T}_3$, respectively. The latter two are rotated clockwise by 90 and 180 degrees, respectively, and then translated appropriately. The 'S' tile in the lower left corner denotes $\sigma_{N,\delta}$. The new tile types A, B, C, D and E suffice to fill in the interior region of the square.

5 Conclusion

In this paper, we introduced the notion of sequential non-determinism, which is a general framework for analyzing the probabilistic correctness of a TAS in the PTAM that has a unique correct terminal assembly and an arbitrary number of (potentially infinite) incorrect assemblies. We then demonstrated that it is a sufficiently general framework to allow the efficient self-assembly of an $N \times N$ square at temperature 1 with high probability, i.e., Theorem 19. Note that Theorem 19 improves upon the previous state-of-the-art bound by Cook, Fu and Schweller [5], but our result is still not optimal from an information-theoretic perspective. It is future work to determine whether such an asymptotic lower bound can be achieved in the 2DPTAM.

References

1 Leonard M. Adleman, Qi Cheng, Ashish Goel, and Ming-Deh A. Huang. Running time and program size for self-assembled squares. In *STOC*, pages 740–748, 2001. `doi:10.1145/380752.380881`.

2 Nathaniel Bryans, Ehsan Chiniforooshan, David Doty, Lila Kari, and Shinnosuke Seki. The power of nondeterminism in self-assembly. *Theory Comput.*, 9:1–29, 2013. `doi:10.4086/TOC.2013.V009A001`.

3 Harish Chandran, Nikhil Gopalkrishnan, and John H. Reif. Tile complexity of linear assemblies. *SIAM J. Comput.*, 41(4):1051–1073, 2012. `doi:10.1137/110822487`.

4 Qi Cheng, Gagan Aggarwal, Michael H. Goldwasser, Ming-Yang Kao, Robert T. Schweller, and Pablo Moisset de Espanés. Complexities for generalized models of self-assembly. *SIAM Journal on Computing*, 34:1493–1515, 2005. `doi:10.1137/S0097539704445202`.

5 Matthew Cook, Yunhui Fu, and Robert Schweller. Temperature 1 self-assembly: Deterministic assembly in 3d and probabilistic assembly in 2d. In *Proceedings of the 22nd Annual ACM-SIAM Symposium on Discrete Algorithms*, 2011.

6 Erik D. Demaine, Martin L. Demaine, Sándor P. Fekete, Mashhood Ishaque, Eynat Rafalin, Robert T. Schweller, and Diane L. Souvaine. Staged self-assembly: nanomanufacture of arbitrary shapes with $O(1)$ glues. *Natural Computing*, 7(3):347–370, 2008. doi:10.1007/s11047-008-9073-0.

7 David Doty. Randomized self-assembly for exact shapes. *SICOMP: SIAM Journal on Computing*, 39(8):3521–3552, September 2010. Preliminary version appeared in FOCS 2009. doi:10.1137/090779152.

8 David Doty, Jack H. Lutz, Matthew J. Patitz, Scott M. Summers, and Damien Woods. Intrinsic universality in self-assembly. In *Proceedings of the 27th International Symposium on Theoretical Aspects of Computer Science*, pages 275–286, 2009.

9 David Doty, Matthew J. Patitz, Dustin Reishus, Robert T. Schweller, and Scott M. Summers. Strong fault-tolerance for self-assembly with fuzzy temperature. In *Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science (FOCS 2010)*, pages 417–426, 2010. doi:10.1109/FOCS.2010.47.

10 David Doty, Matthew J. Patitz, and Scott M. Summers. Limitations of self-assembly at temperature one (extended abstract). In Turlough Neary, Damien Woods, Anthony Karel Seda, and Niall Murphy, editors, *Proceedings International Workshop on The Complexity of Simple Programs, CSP 2008, Cork, Ireland, 6-7th December 2008*, volume 1 of *EPTCS*, pages 67–69, 2008. arXiv:0906.3251.

11 Sándor P. Fekete, Jacob Hendricks, Matthew J. Patitz, Trent A. Rogers, and Robert T. Schweller. Universal computation with arbitrary polyomino tiles in non-cooperative self-assembly. In *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2015, San Diego, CA, USA, January 4-6, 2015*, pages 148–167, 2015. doi:10.1137/1.9781611973730.12.

12 Bin Fu, Matthew J. Patitz, Robert T. Schweller, and Robert Sheline. Self-assembly with geometric tiles. In *Automata, Languages, and Programming – 39th International Colloquium, ICALP 2012, Warwick, UK, July 9-13, 2012, Proceedings, Part I*, pages 714–725, 2012. doi:10.1007/978-3-642-31594-7_60.

13 David Furcy, Samuel Micka, and Scott M. Summers. Optimal program-size complexity for self-assembled squares at temperature 1 in 3D. *Algorithmica*, 77(4):1240–1282, 2017. doi:10.1007/S00453-016-0147-6.

14 David Furcy and Scott M. Summers. Optimal self-assembly of finite shapes at temperature 1 in 3D. *Algorithmica*, 80(6):1909–1963, 2018. doi:10.1007/S00453-016-0260-6.

15 David Furcy, Scott M. Summers, and Hailey Vadnais. Tight bounds on the directed tile complexity of a just-barely 3d $2 \times N$ rectangle at temperature 1. In Daniela Genova and Jarkko Kari, editors, *Unconventional Computation and Natural Computation – 20th International Conference, UCNC 2023, Jacksonville, FL, USA, March 13-17, 2023, Proceedings*, volume 14003 of *Lecture Notes in Computer Science*, pages 79–93. Springer, 2023. doi:10.1007/978-3-031-34034-5_6.

16 David Furcy, Scott M. Summers, and Christian Wendlandt. Self-assembly of and optimal encoding within thin rectangles at temperature-1 in 3D. *Theoretical Computer Science*, 872:55–78, 2021. doi:10.1016/J.TCS.2021.02.001.

17 David Furcy, Scott M. Summers, and Logan Withers. Improved lower and upper bounds on the tile complexity of uniquely self-assembling a thin rectangle non-cooperatively in 3d. *Theory Comput. Syst.*, 67(5):1082–1130, 2023. doi:10.1007/S00224-023-10137-9.

18 Oscar Gilbert, Jacob Hendricks, Matthew J. Patitz, and Trent A. Rogers. Computing in continuous space with self-assembling polygonal tiles (extended abstract). In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016, Arlington, VA, USA, January 10-12, 2016*, pages 937–956, 2016. doi:10.1137/1.9781611974331.CH67.

19 Jacob Hendricks, Matthew J. Patitz, Trent A. Rogers, and Scott M. Summers. The power of duples (in self-assembly): It's not so hip to be square. *Theoretical Computer Science*, 743:148–166, 2018. doi:10.1016/J.TCS.2015.12.008.

20 Ming-Yang Kao and Robert T. Schweller. Randomized self-assembly for approximate shapes. In *Automata, Languages and Programming, 35th International Colloquium, ICALP 2008, Reykjavik, Iceland, July 7-11, 2008, Proceedings, Part I: Tack A: Algorithms, Automata, Complexity, and Games*, volume 5125 of *Lecture Notes in Computer Science*, pages 370–384. Springer, 2008. doi:10.1007/978-3-540-70575-8_31.

21 Ján Manuch, Ladislav Stacho, and Christine Stoll. Two lower bounds for self-assemblies at temperature 1. *Journal of Computational Biology*, 17(6):841–852, 2010. doi:10.1089/CMB.2009.0067.

22 P.-E. Meunier, M. J. Patitz, S. M. Summers, G. Theyssier, A. Winslow, and D. Woods. Intrinsic universality in tile self-assembly requires cooperation. In *Proceedings of the 25th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 752–771, 2014.

23 Pierre-Étienne Meunier and Damien Regnault. Directed non-cooperative tile assembly is decidable. In Matthew R. Lakin and Petr Sulc, editors, *27th International Conference on DNA Computing and Molecular Programming, DNA 27, September 13-16, 2021, Oxford, UK (Virtual Conference)*, volume 205 of *LIPIcs*, pages 6:1–6:21. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2021. doi:10.4230/LIPIcs.DNA.27.6.

24 Pierre-Étienne Meunier, Damien Regnault, and Damien Woods. The program-size complexity of self-assembled paths. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing, STOC 2020, Chicago, IL, USA, June 22-26, 2020*, pages 727–737, 2020. doi:10.1145/3357713.3384263.

25 Pierre-Étienne Meunier and Damien Woods. The non-cooperative tile assembly model is not intrinsically universal or capable of bounded turing machine simulation. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017*, pages 328–341, 2017. doi:10.1145/3055399.3055446.

26 Matthew J. Patitz, Robert T. Schweller, and Scott M. Summers. Exact shapes and Turing universality at temperature 1 with a single negative glue. In *Proceedings of the 17th international conference on DNA computing and molecular programming*, DNA'11, pages 175–189, Berlin, Heidelberg, 2011. Springer-Verlag. doi:10.1007/978-3-642-23638-9_15.

27 Paul W. K. Rothemund and Erik Winfree. The program-size complexity of self-assembled squares (extended abstract). In *STOC '00: Proceedings of the thirty-second annual ACM Symposium on Theory of Computing*, pages 459–468, 2000. doi:10.1145/335305.335358.

28 Nadrian C. Seeman. Nucleic-acid junctions and lattices. *Journal of Theoretical Biology*, 99:237–247, 1982.

29 David Soloveichik and Erik Winfree. Complexity of self-assembled shapes. *SIAM J. Comput.*, 36(6):1544–1569, 2007. doi:10.1137/S0097539704446712.

30 Erik Winfree. *Algorithmic Self-Assembly of DNA*. PhD thesis, California Institute of Technology, June 1998.