5-2014

Exponential replication of patterns in the Signal Tile Assembly Model and experimental non-deterministic assembly of lines in the Probabilistic Tile Assembly Model

Alexandra B. Keenan
University of Texas-Pan American

EXPONENTIAL REPLICATION OF PATTERNS IN THE SIGNAL TILE ASSEMBLY MODEL AND EXPERIMENTAL NON-DETERMINISTIC ASSEMBLY OF LINES IN THE PROBABILISTIC TILE ASSEMBLY MODEL

A Thesis

by

ALEXANDRA B. KEENAN

Submitted to the Graduate School of
The University of Texas-Pan American
In partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

May 2014

Major Subject: Computer Science

EXPONENTIAL REPLICATION OF PATTERNS IN THE SIGNAL TILE ASSEMBLY MODEL AND EXPERIMENTAL NON-DETERMINISTIC ASSEMBLY OF LINES IN THE PROBABILISTIC TILE ASSEMBLY MODEL

A Thesis
by
ALEXANDRA B. KEENAN

COMMITTEE MEMBERS

Dr. Robert Schweller
Chair of Committee

Dr. Zhixiang Chen
Committee Member

Dr. Bin Fu
Committee Member

May 2014

ABSTRACT

Keenan, Alexandra B., Exponential Replication of Patterns in the Signal Tile Assembly Model and Experimental Non-Deterministic Assembly of Lines in the Probabilistic Tile Assembly Model. Master of Science (MS), May, 2014, 51 pp., 2 tables, 20 figures, references, 41 titles.

We introduce the problem of self-replication of rectangular two-dimensional patterns in the practically motivated Signal Tile Assembly Model (STAM) [23], which is an ex- tension of the aTAM.. In the first part of this thesis, we construct an exponential pattern replicator that replicates a two-dimensional input pattern over some fixed alphabet of size φ with O(φ) tile types, O(φ) unique glues, and a signal complexity of O(1). In the second part of this thesis, we use a non-deterministic model of tile assembly to significantly reduce the tile complexity of specified-length linear assemblies, which are a particularly important substructure for building more complicated nanostructures.

ACKNOWLEDGEMENTS

First, I would like to express my sincere gratitude to my advisor, Robert Schweller, for his continuous support of my Master's research and the freedom and encouragement to pursue a variety of projects. Over the past two years, I have benefitted from his selfless dedication to his students, his patience as I grappled with a new field, and his wealth of expertise. I also thank the past and current students in our group: Michael Sherman, Xingsi Zhong, and David Chavez, whom I spent many hours with working on constructions, proofs, and experiments. It was a pleasure.

I must thank Matt Patitz for introducing me to the field and encouraging me to pursue research in self-assembly, and Damian Woods and Jennifer Padilla for help and encouragement in breaking into the experimental side of self-assembly. I also thank Michael Persans for welcoming me into his lab.

My time at UTPA was rounded out by some exceptional courses and faculty. In particular, I would like to thank Bin Fu, an excellent professor, for introducing me to the beautiful world of computational theory. I would also like to thank Laura Grabowski for her unending support and enthusiasm for all of my pursuits.

Lastly, I'd like to thank my family for their incredible love, support and understanding throughout my education. I would be nowhere without them.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER I

INTRODUCTION

In May of 1959, physicist Richard Feynman gave a speech entitled, *There's Plenty of Room at the Bottom* which outlined his vision, 'that we could arrange atoms one-by-one, just as we want them [3].' Much of the nano-technology community that emerged in the subsequent decades after Feynman's conceptual description of the field focused on mechanical devices and silicon-based electronics, and the evolution-honed efficiency of Nature's self-assembling nano machines was neglected. In 1992, Eric Drexler first described the role that biology and biochemistry might play in nano-scale information storage and computation [14]:

> Most experimental research in molecular electronics has focused on the development of molecules that exhibit useful electronic properties in thin films or in microscale aggregates; some proposals, however, have focused on the construction of computational devices in which individual molecules or moieties would serve as signal carrying and switching elements.

In the years since, DNA gained favor as the molecule of choice in molecular computation and nano-technology for its predictable interactions with other DNA molecules, its simple reaction kinetics, and its ability to store vast amounts of information. Seminal works by Seeman and Robinson toward designing a self-assembling nanoscale biochip [28], combined with Drexel's novel approach to molecular computation mark a shift in nano-technology strategy during the late 1980s and early 1990s toward investigating biological and chemical grounds for molecular scale computing.

A few years after Nadrian Seeman and colleagues built the world's first nanoscale structure--a cube made out of DNA [6], Leonard Adleman performed his first DNA computing experiment:

solving an instance of the Hamiltonian Path Problem in which 1014 operations were executed in one second, faster than the super-computers of the day [2]. This seminal paper describes Adleman's first DNA computing experiment which solved an 8-vertex instance of the Hamiltonian Path Problem in which 1014 operations were executed in one second, faster than the super-computers of the day. Soon after this announcement, the media and scientists alike were heralding the end of silicon-based computers in favor of molecular computers. However, the excitement of the earlier, headier days of molecular computation subsided as the inherent limitations of DNA computing came to light.

The goal of most DNA computing laboratories now is not to compete with silicon-based electronics, but to mimic the sorts of information processing which may occur in cells via dynamic biomolecular interactions [40]. Thus, the two branches, DNA nanotechnology and DNA computing evolved symbiotically, with many researchers working in both fields. However, it was Erik Winfree that first established a direct link between the two paradigms by suggesting that that branched DNA molecules could undergo self-assembly in a programmable way, thus performing computation as part of their growth process [39]. It was this idea that led to the combination of computational geometry, DNA nanotechnology, and DNA computing into the rich field known as algorithmic self-assembly. Algorithmic self-assembly has its foundations in Wang tiles, which are a formal system modeled by square tiles with a color on each of their four sides [37]. Two Wang tiles may be arranged side-by-side in two dimensional space such that abutting edges on two adjacent tiles have matching colors. They may tile the plane in this way, creating shapes and/or patterns. This abstract notion can be realized in reality via the use of branched DNA molecules called *DNA tiles* with branches that may attach to other tiles via hybridization to branches on surrounding tiles with complementary nucleotide bases. In this way, structures of predictable tile composition can self-assemble in solution. Winfree realized that the manner in which these DNA tiles assemble is algorithmic and observed that the Chomsky Hierarchy of languages emerges naturally in self-assembling structures. The progression from regular to context-free to recursively enumerable languages is mirrored by the progression of computational power from linear DNA tiles to den-

drimer tiles to two-dimensional DNA self-assembling structures [39]. Linear DNA self-assembly, that is, self-assembly beginning with oligonucleotides or duplex molecules with sticky ends, is capable of generating regular languages. Dendrimer self-assembly is self-assembly of duplexes, hairpins, and 3-armed junctions with sticky ends, all of which generate context-free languages. Two-dimensional self-assembly, DNA self-assembly into a planar lattice structure, is capable of Turing universal computation [39].

I.I Abstract Tile Assembly Model

In order to visualize and generate DNA tile sets, an abstract model of two-dimensional self-assembly was developed by Winfree and is known as the abstract Tile Assembly Model (aTAM). A tile is expressed as a square with labeled edges. Adjacent squares may attach to one another if their abutting edges have matching labels and the strength of the bond is adequate. Whether a bond is adequately strong depends on a parameter called temperature, which is a characteristic of the environment in which the self-assembly is taking place. If the bond strength is equal to or greater than the temperature of the system, then the bond will form. For example, if the temperature of the system is 2, then a single bond may form only if it has strength 2 or greater. This can be extended to include multiple bonds between tiles, where the sum of the strengths of the bonds must be equal to or greater than the temperature of the system [39]. A tile self-assembly system begins with infinitely many copies of certain tiles, each with some specified label, also known as a glue, on each of its four sides. One tile type is different from another tile type if it has some other glue types or configuration of glues. A tile set consists of all the tile types available in the system. In general, tile self-assembly begins with a seed tile, the tile that nucleates the aggregation of other tiles in the tile set. Tiles may transform horizontally or vertically to attach to the seed or growing aggregate. Tiles may not flip or rotate [39]. An interesting phenomenon dubbed *cooperativity*, occurs in two-dimensional assembly systems when the temperature is at least two. This characteristic is manifested when two binding sites of a tile act together when binding to the aggregate. For example, if a tile has one

glue of strength 1 on its East face and one glue of strength 1 on its West face, these two glues may act simultaneously in binding to the aggregate such that the total strength with which the tile is bound to the aggregate is strength two [39]. The most studied consideration for any tile self-assembly system is the tile complexity required to build a given pattern or shape. The tile complexity is often defined differently in different models, but in the aTAM it is defined as the number of tile types in the system. This is of practical importance, because each tile type requires significant experimental characterization and development in the laboratory. Glue complexity is also important and is defined as greatest number of glues on a given side of a given tile. This is also important to consider because each glue requires significant experimentation to identify potential cross-reactions with other glues in the system that could result in a tile binding to the aggregate in an erroneous manner. An obvious upper bound for tile complexity for any tile shape would be the area of the shape. More specifically, a shape with area n would require at most $O(n)$ tile types to construct, with each tile type occupying a unique coordinate within the shape. Indeed, this does describe the tight asymptotic behavior for some cases, as an $nx1$ line does require $O(n)$ tile types. However, because it is possible to build computation into the construction, oftentimes fewer tile types will suffice. For example, an nxn square requires $O(\frac{logn}{loglogn})$ tile types because a base-n log counter may be embedded into the growing assembly [29]. The aTAM at temperature 2 is capable of Turing universal computation, and it is believed the cooperativity phenomenon is responsible for this [20]. It has also been shown that aTAM at temperature 2 is intrinsically universal, that is, there exists a single tile set that can be used to simulate any other arbitrary aTAM system [12]. It has been conjectured that at temperature 1, Turing universal computation in aTAM is impossible, and some progress has been made toward proving this but it is still an open problem [13].

I.II Extensions to the aTAM

Since Winfree's thesis was published, a rich spectrum of research in algorithmic DNA self-assembly has arisen and now includes several models of self-assembly. A subset of the models

most relevant to this thesis will be reviewed in this section. First, the signal tile assembly model (STAM) is the model used in replication of patterns discussed in Chapter II. The experimental work in Chapter III relies heavily on the theoretical work performed by [5] in developing the probabilistic tile assembly model (pTAM) described in Section I.II.I, the concentration programming model described in Section I.II.IV, and the staged assembly model described in Section I.II.II.

I.II.I Probabilistic Tile Assembly Model

The probabilistic tile assembly model (pTAM) utilizes randomized assembly to construct $1xn$ rectangles, or rulers, much more efficiently [5]. These rulers can then be used to guide the assembly of a variety of complex shapes. Recall that in the aTAM, rulers may be assembled with $O(n)$ tile types. This bound is asymptotically tight, which begs the question why the theoretic limit of $O(\frac{logn}{loglogn})$ cannot be reached. One possible answer lies in the deterministic nature of the aTAM. The aTAM is deterministic in the sense that only one tile may attach to a given position in a partially formed assembly. The pTAM is non-deterministic in the sense that it allows for the possibility of more than one tile attaching in a given place to the partially formed assembly. It leverages this possibility to construct rulers with fewer tile types than what is required in the aTAM. In the pTAM, rulers may be assembled with a tile set of $O(logn)$ tile types that have only one glue per side. With two glues per side, it is possible to achieve the theoretic lower bound of $O(logn/loglogn)$ tile types.

To achieve a tile complexity of $O(logn)$ two tile subsets are used that may bind to the growing ruler. One subset is a set of *counter* tiles, of which there are p tiles where each tile may be designated as $x_1, x_2, ..., x_{p-1}, x_p$. The other subset contains *reset* tiles, of which there are also p tiles designated as $r_1, r_2, ..., r_{p-1}, r_p$. If a reset tile binds to the growing assembly, then the assembly is reset back to binding x_1. It is only after all the counter tiles (x_1 through x_p) bind consecutively that the assembly terminates. If we consider a reset tile binding to be *heads* and a counter tile binding to be *tails*, then the aforementioned event occurring would be equivalent to flipping p heads in a row. Thus, the expected length of the line will be much longer than the number of tiles required to build it. In fact the expected length of the line is $2p$, so the tile complexity is $O(logn)$ where n is the length of

the line [5].

It should be noted, however, that rulers developed using this technique have an expected length but this length cannot be guaranteed because of the nature of the non-deterministic self-assembly process in this model. The set of terminal assemblies will have more than one shape using this model with a small probability of formation of each erroneous terminal assembly possibility. The probabilistic model described in [5] describes an elegant solution to the construction of a line in fewer tile types.

Some work has been done using other probabilistic models for the efficient construction of polygons. Most of these approaches have included varying the concentrations of available tile types to achieve exact shapes with high probability [10, 19]. While not a nondeterministic method, it should also be mentioned here that temperature programming, the raising and lowering of the temperature as the self-assembly process proceeds, can also result in in approximations of sufficiently scaled-up shapes with few tile types [33].

I.II.II Staged Self-Assembly

Of all of the tile assembly models developed, the staged self-assembly model is perhaps most conducive to the success of the experimentalist attempting to build complex shapes or scaffolds in the laboratory. In this model, assemblies are constructed in stages, with only a few tiles in the tiles set available to the attach to the assembly at a given time [8]. This is accomplished by starting a seed with only a strategic few of the tile types in the set available to it. When all possible binding instances have occurred, the experimentalist may add the next batch of tile types to the system in order to continue the assembly process. It is also possible to start two separate assemblies in two separate test tubes and then mix them together at some later point in the assembly process to achieve the desired effect. This helps to reduce the tile types and glue types required to construct a given assembly at the expense of the operations that the experimenter must perform during the assembly process. In reality, given a reasonable number of mixing and/or separating reactions, this approach is very practical because it drastically minimizes the time which the experimenter

must spend on the construction of a particular assembly system because the most lengthy process is actually developing tile types. The model is also beneficial because it is difficult to design a large number of glues that attract only in pairs, as there is a limit to the number of sufficiently different ATGC sequences that can be achieved on a short piece of DNA. To achieve efficient staged self-assembly, the goal is to minimize both tile types and experimenter operations. The developers of this model show that any patterned shape can be constructed with a constant number of glues in $O(logn)$ steps [8]. This model was also used to design a shape replicating system in [1]. Given an input shape, the self-assembly system replicates that shape either into a specific number of copies or an unbounded number of copies. This scheme takes advantage of the fact that tiles systems can be synthesized with DNA or RNA, both of which would theoretically behave the same way. The replication scheme utilizes a stage in which RNAse is added to the batch. RNAse is an enzyme that rapidly degrades RNA. Shape replication is achieved by first surrounding the input shape to be replicated with RNA tiles in a series of stages. These RNA tiles are then surrounded by layers of DNA tiles. RNAse is then added to the batch which dissolves the RNA tiles, releasing the input shape and leaving a template through which the shape may be replicated. In the next stage, the RNAse enzyme is disabled and a new layer of RNA tiles lines the negative-shape template. DNA tiles may then fill in the form of the shape. RNAse is added once more, which dissolves the RNA tiles and releases the shape replica. This may be repeated the desired number of times in order to achieve the desired number of replications. Alternatively, a new template may be created for each new replica generated, creating an exponential replication scheme which allows for the unbounded generation of replicates. Under this model, the authors show that genus-0 shapes may be replicated infinitely many times using $O(1)$ tiles types and $O(1)$ stages. Replicating a precise number of copies of a shape requires $O(1)$ tiles types and $O(logn)$ stages.

I.II.III Signal Tile Assembly Model

First described in [23], the Signal Tile Assembly Model (STAM) is a powerful model of tile self-assembly in which activation, via binding, of a glue on an individual tile may turn other glues

either on or off elsewhere on the tile. In this way, signals may be propagated across distances and assemblies may be broken apart (Figure I.1). Additionally, certain glues may only be made available after a desired event occurs, giving control over the assembly sequence that was not achievable in passive tile models such as aTAM and staged self-assembly. One limitation of the STAM is that the signals are asynchronous, meaning that once a signal has been activated, it can only be guaranteed to be propagated at some point in the future. The designer cannot depend on the signal being propagated immediately [23]. This limitation is due to the physical realization of the signals.

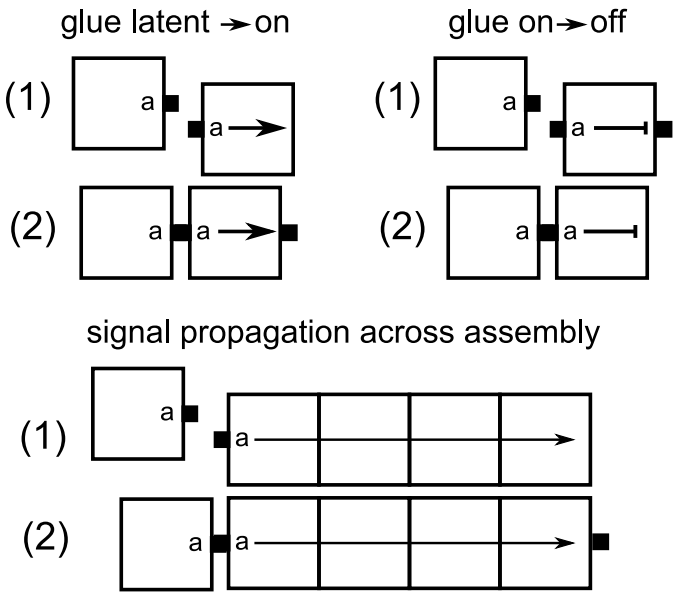

Figure I.1: STAM schematic.

Queued DNA strand displacement reactions on tile monomers provide a plausible physical basis for the signaling cascades used in the STAM. A DNA strand-displacement reaction occurs when two strands with at least partial Watson-Crick complementarity hybridize, displacing one or more pre-hybridized strands in the process [40]. Such reactions can be set up in sequence, resulting in a cascade (Figure I.2).

Using the STAM, it is possible to assemble a $1xn$ line with $O(1)$ tile types and signal complexity $O(logn)$. The authors also presented a Turing machine simulation that operated without

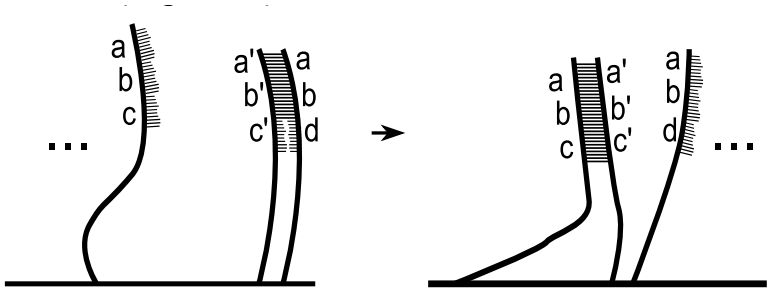

Figure I.2: The DNA strand with nucleotide regions a, b, and c displaces the strand with regions a, b, and d by hybridizing to abc. The strand with nucleotide regions a, b, and d is then free to perpetuate the hybridization cascade.

making a copy of the entire row representing the tape at each step, but which instead uses only a constant number of new tiles per step. This was new in the tile self-assembly model, as all previous constructions kept the tape history in the growing assembly. The model is also capable of strictly assembling a discrete self-similar fractal, something which was shown to be impossible in the aTAM. Specifically, the authors presented a strict construction of the Sierpinski triangle, a self-similar fractal which can only be approximated in the aTAM.

Complexity in the STAM is represented as both the number of tile types and with a parameter known as signal complexity. Signal complexity is measured as the maximum number of signals on a side of any tile in the tile set. Using the STAM, it is possible to assemble a $1xn$ line with a $O(1)$ tile types and signal complexity $O(logn)$. The model developers also presented a Turing machine simulation that operated without making a copy of the entire row representing the tape at each step, but which instead uses only a constant number of new tiles per step. This was new in the tile self-assembly model, as all previous constructions kept the tape history in the growing assembly. The authors also presented a strict construction of the Sierpinski triangle, a self-similar fractal which can only be approximated in the aTAM.

I.II.IV Concentration Programming

Tile concentration programming is a non-deterministic extension to the TAM where more than one tile type is capable of binding to a given position on a growing assembly. Concentration pro-

gramming utilizes differing tile type concentrations to determine probabilities that a given tile will bind to a position where multiple tile types could bind [11, 19]. By programming relative concentrations of tile types, it is possible to probabilistically assemble lines with lengths over some distribution but with expected length L with very few tile types.

I.II.V Single Phase

In its simplest form, construction of a linear assembly with some expected length using concentration programming requires three tile types: a seed S, a growth tile G, and a termination tile T. Assembly initiation begins with the seed tile. Growth is unidirectional with G and T tiles competing to attach to the growing assembly (Figure I.3). The binding of a G tile continues the growth of the line eastward while the binding of a T tile halts line growth and terminates the assembly, meaning no more tiles may attach. When G tile types are assigned a concentration of $1 - p$ and T tiles are assigned a concentration of p, assembly lengths fall over a geometric distribution described by $P(\ell) = (1 - p)^{\ell-2}p$ where the expected assembly length $L = \frac{1}{p}$.

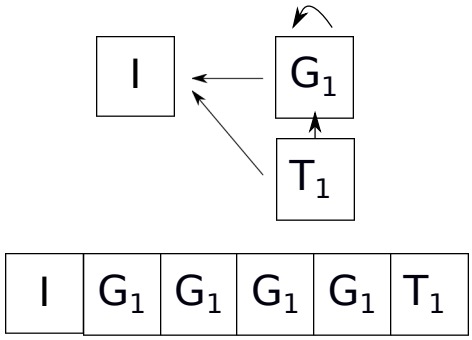

Figure I.3: Top: Tiles involved in assmembly with arrows indicating binding possibilities. Bottom: Possible assembly.

I.II.VI Multiple Phases

To produce linear assemblies with lengths more tightly concentrated around L, multiple *phases* of growth and termination tiles are required: $SG_1T_1G_2T_2...G_kT_k$. This assembly mechanism can generate all strings described by the regular expression

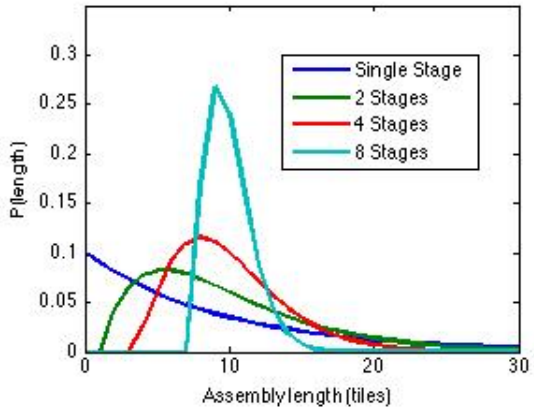

Figure I.4: These distrubtions demonstrate the tightening of assembly lengths around some expected length L with an increasing number of stages. Here, p and r were set for each assembly system to provide for an expected length of 10 tiles.

$IG_{1,1}(G_{1,2}G_{1,1})^*T_1G_{2,1}(G_{2,2}G2,1)^*T_2...G_{n,1}(G_{n,2}G_{n,1})^*T_n.$

Tiles G_i and T_i compete with one another at a given stage i to bind to the growing linear assembly. The binding of T_i signals a transition to stage $i+1$ in which tile G_{i+1} and T_{i+1} would then compete to bind to the assembly. Tile G_i has a concentration of $p-1$ and tile T_i has a concentration p which generates assemblies of expected length $L = \frac{k}{p}$ with lengths falling over the negative binomial distribution $P(\ell|k,p) = \binom{\ell-1}{k-1}p^k(1-p)^{\ell-k-1}$. Increasing k, the number of stages, tightens the distribution around L (Figure I.4).

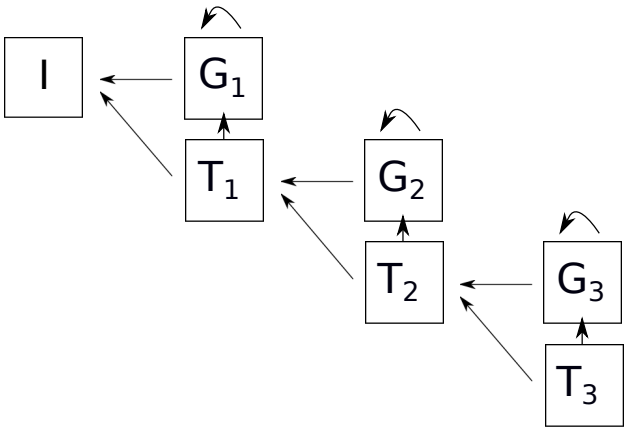

Figure I.5: Tiles involved in a 3-stage linear assembly example with arrows indicating binding possibilities.

I.III Contents and Contributions

A version of Chapter II appears in: Keenan, A, Schweller, R, and Zhong, X. *Exponential Pattern Replication in the Signal Tile Assembly Model*. Proceedings of the 19th International Meeting on DNA Computing (DNA19), Arizona State University, Tempe, AZ, September 22-27, 2013.

Robert Schweller wrote the definitions in this paper and Xingsi Zhong made contributions to the 2D replication constructions. The experiments in Chapter III were designed by me and performed by myself, Robert Schweller, David Chavez, Leslie Sweet and Cameron Chalk.

CHAPTER II

EXPONENTIAL REPLICATION IN THE STAM

Artificial self-replicating systems have been the subject of various investigations since John von Neumann first outlined a detailed conceptual proposal for a non-biological self-replicating system [21]. Following Watson and Crick's seminal paper in 1953 [38], L.S. Penrose used DNA as inspiration in designating the necessary features of a self-replicating system [26]. Gunter von Kiedrowski, who demonstrated the first enzyme-free abiotic replication system in 1986 [36], describes a model that can be used to conceptualize template-directed self-replication [24]. In this model, minimal template-directed self-replicating systems consist of an autocatalytic template molecule, and two or more substrate molecules that bind the template molecule and join together to form another template molecule. To date, simple self-replicating systems have been demonstrated in the laboratory with nucleic acids, peptides, and other small organic molecules [25, 35, 36, 41].

Given that substrate molecules must come together without outside guidance to replicate the template, a template-directed self-replicating system is necessarily a self-assembling system. Self-assembly has long been nature's method of choice for building complex structures, large and small. How did such complex self-assembly schemes come to be? What kind of autonomous chemical systems can self-replicate and evolve? Among evolutionary biologists, there is a perspective that the problem of the origin of life is a problem of information processing. In theoretical computer science, the Tile Assembly Model (TAM) has become the most commonly used model to describe various self-assembly processes [39]. Many model variants have been described since Erik Winfree first introduced the TAM, however models that are most relevant to self-replicating systems are those that allow for assembly breakage. These include the enzyme staged assembly model [1],

the temperature programming model [18], the signal tile assembly model [22, 23], and the use of negative glues [27].

Replication of arbitrary 0-genus shapes has been shown within the staged assembly model with the use of RNAse enzymes [1]. Replication and evolution of combinatorial `genomes' via crystal-like growth and breakage have also been demonstrated in the laboratory using DNA tile monomers [30]. Under this replication mechanism, a DNA crystal ribbon has a sequence of information, or 'genome', in each row. Upon chance breakage, the daughter crystal continues to grow and copy the 'genome' of the mother crystal. It was further shown that the fidelity of the replication process is sufficiently high for Darwinian evolution. Such simple, enzyme-free systems are of particular importance to the study of the origins of life. Several decades ago, Graham Cairns-Smith proposed that life began with clay. Clay is composed of tiny layered crystals that may have a variety of patterns of atoms or molecules in a layer. This pattern can be viewed as a sort of genome that is copied through the layers as the crystal grows. If a portion of the crystal broke off from the parent, this new crystal would continue to copy the same pattern as it grew. If some layer patterns grew faster than others, then those particular patterns would be selected for. As an example, patterns that allowed the crystal to stick to the riverbed better might be selected for because clay crystals grow in an aqueous environment. Crystals that tended to be unable to stick to the riverbed would flow downriver and eventually end up on a sand bank and disintegrating. The clay hypothesis contends that perhaps organic molecules began adhering to the clay and being replicated along with the patterns. Perhaps they conferred some advantage to clay layer reproduction. Perhaps, then, the organic molecules began to behave more autonomously and reproduce themselves. Recently, industry has been using clay crystals as catalysts for a variety of organic reactions, which raises interesting possibilities, as the absence of enzymes in early RNA or DNA worlds has presented a conundrum for scientists studying the origins of life. Schulman et al studied the possibility of the Cairns-Smith clay hypothesis via the use of DNA tile assembly. Because clay crystals grow so slowly, the clay hypothesis was never given much thought by experimentalists because it simply

could not be studied under any reasonable time scale in a laboratory. However, with the advent of self-assembling DNA tiles, the possibility of the crystal pattern replication and breakage model could be studied on a more reasonable time scale. Schulman and Winfree constructed, zig-zag ribbon crystals with a sequence of information in each row. They showed that in theory crystals that can compute and can undergo open-ended evolution as they try to produce more and more complex 'genomes' to take advantage of available growth resources. This mechanism was simple enough to observe in the laboratory. This work suggests that the concept of a self-replicating chemistry is closely related to the concept of a chemistry that can store information and compute. It is only by clearly understanding how chemical systems can transfer and process information that we can hope to understand how self-replication and evolution can occur, and by implication, understand how life might have begun.

A template-directed method of exponential self-replication within the tile assembly system, where the child molecule detaches from and is identical to the parent (as is found in biological systems), has not yet been described. Here, we present a theoretical basis for template-directed exponential self-replication in the practically motivated Signal Tile Assembly Model (STAM), and in doing so partially address an open question presented by Abel and colleagues [1]. Specifically, we consider the problem of self-replication of rectangular two-dimensional patterns in the STAM. The STAM is a powerful model of tile self-assembly in which activation, via binding, of a glue on an individual tile may turn other glues either on or off elsewhere on the tile [23]. In this way, signals may be propagated across distances greater than a single tile and assemblies may be broken apart. DNA strand displacement reactions provide a plausible physical basis for the signaling cascades used in the STAM. DNA strand displacement occurs when two DNA strands with at least partial complementarity hybridize with each other, which can displace pre-hybridized strands. In the STAM, these reactions may be queued to result in a cascade that ultimately turns a glue 'on' by releasing a pre-hybridized strand. Conversely these queued reactions could turn a glue 'off' by binding a free strand, thus making it unavailable to interact with other glues.

An important objective of nanotechnology is to manufacture things inexpensively, thus the prospect of self-replicating materials with useful patterns or functions is enticing. Additionally, an enzyme-free self-replicator that can support and autonomously replicate an information-bearing genome could provide the basis for a model of Darwinian evolution. Because true Darwinian selection necessitates exponential population growth [34], and this rate of growth is also desirable for low-cost manufacturing of nanoscale devices, we approach this problem with the goal of exponential growth in mind.

The Signal Tile Assembly Model of [23] is briefly defined formally in Section II.I, followed by our formal definition of exponential replication. We first present a 2D signal tile system that replicates a linear pattern and then extend this mechanism to present our main result in Section II.III: there exists a single, general purpose 2D signal tile system that exponentially replicates any rectangular 2D pattern (Theorem 1). This is followed by a discussion of replication of linear assemblies in 1D space and 2D assemblies in 3D space, however we omit a detailed analysis of these systems due to space limitations.

II.I Definitions

II.I.I Basic Definitions

Multisets. A multiset is an ordered pair (S, m) where S is a subset of some universe set U and m is a function from U to $\mathbb{N} \bigcup \{\infty\}$ with the property that $m(x) \geq 1$ for all $x \in S$ and $m(x) = 0$ for all $x \notin S$. A multiset models a collection of items in which there are a positive number of copies $m(x)$ of each element x in the collection (called the multiplicity of x). For a multi-set $A = (S, m)$ and $x \in S$, we will use notation $A(x) = m(x)$ to refer to the multiplicity of item x, and $|A| \triangleq \sum_{a \in S} m(a)$ to refer to the *size* of A. For multisets $B = (b, m)$ and $A = (a, n)$, define $B \bigcup A$ to be the multiset $(a \bigcup b, m')$ where $m'(x) = m(x) + n(x)$. If $m(x) \geq n(x)$ for all $x \in U$, then define $B - A$ to be the multiset $(b', m'(x))$ where $b' = \{x \in b | m(x) - n(x) \geq 1\}$ and $m'(x) = m(x) - n(x)$. We use standard set notation $\{a_1, \ldots, a_r\}$ to denote multi-sets with the

multiplicity of an item a being inferred by the number of i such that $a_i = a$.

Patterns. Let ϕ be a set of labels that contains at least one particular label `null` $\in \phi$ which conceptually denotes a blank, non-existent label. Informally, a 2D pattern is defined to be a mapping of 2D coordinates to elements of ϕ. Further, as these patterns will denote patterns on the surface of free floating tile assemblies, we add that patterns are equal up to translation. Formally, a 2D pattern over set ϕ is any set $\{f_{\Delta_x,\Delta_y}(x,y)|\Delta_x,\Delta_y \in \mathbb{Z}\}$ where $f : \mathbb{Z}^2 \to \phi$, and $f_{\Delta_x,\Delta_y}(x,y) = f(x+\Delta_x, y+\Delta_y)$. In this paper we focus on the the class of *rectangular* patterns in which the `null` label occurs at all positions outside of a rectangular box, with positions within the box labeled arbitrarily with non `null` labels.

II.I.II Signal Tile Model

In this section we define the signal tile assembly model (STAM) by defining the concepts of an *active tile* consisting of a unit square with *glue slots* along the faces of the tile, as well as *assemblies* which consist of a collection of active tiles positioned on the integer lattice. We further define a set of three *reactions* (*break* reactions, *combination* reactions, and *glue-flip* reactions) which define how a set of assemblies can change over time. Figure II.1 represents each of these concepts pictorially to help clarify the following technical definitions. Please see [23] for a more detailed presentation of the STAM.

Glue Slots. *Glue slots* are the signal tile equivalent of glues in the standard tile assembly model with the added functionality of being able to be in one of three states, *on*, *off*, or *latent*, as well as having a *queued command* of *on*, *off*, or -, denoting if the glue is queued to be turned on, turned off, or has not been queued to change state. Formally, we denote a glue slot as an ordered triple $(g,s,q) \in \Sigma \times \{on, \mathit{off}, latent\} \times \{on, \mathit{off}, -\}$ where Σ is some given set of labels referred to as the *glue type* alphabet. For a given glue slot $x = (g,s,q)$, we define the *type* of x to be g, the *state* of x to be s, and the *queued action* of x to be q.

Active Tiles. An active tile is a 4-sided unit square with each edge having a sequence of *glue slots* $g_1, \ldots g_r$ for some positive integer r, as well as an additional *label* taken from a set of symbols

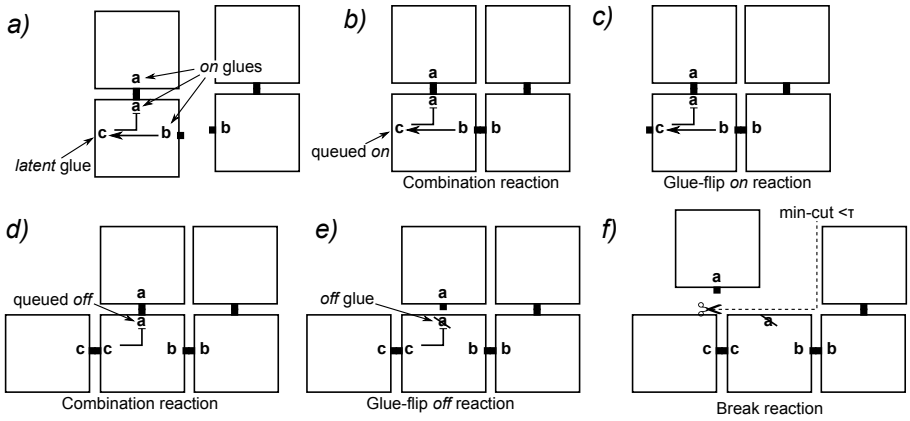

Figure II.1: This sequence (a-f) demonstrates the reaction types, glue states, and queued commands defined in the STAM.

ϕ. For simplicity of the model, we further require that the glue type of each g_i on each tile face is the same (although state and queued commands may be different), and that the glue type of g_i is distinct from the glue type of g_j if $i \neq j$. For an active tile t, let $t_{d,i}$ denote the glue slot g_i on face d of active tile t.

Finally, an active tile t has an associated *signal function* $f_t(d, i)$ which assigns to each glue slot i on each tile side d a corresponding set of triples consisting of a glue slot, a side, and a command, which together denote which glue slots of each tile face should be turned on or off in the event that slot i on face d becomes bonded. Formally, each active tile t has an associated *signal function* f : $\{north, south, east, west\} \times \{1, \ldots r\} \rightarrow \mathcal{P}(\{north, east, south, west\} \times \{1, \ldots r\} \times \{on, off\})$. For the remainder of this paper we will use the term *tile* and *active tile* interchangeably.

Assemblies. An assembly is a set of active tiles whose centers are located at integer coordinates, and no two tiles in the set are at the same location. For an assembly A, define the weighted graph $G_A = (V, E)$ such that $V = A$, and for any pair of tiles $a, b \in V$, the weight of edge (a, b) is defined to be 0 if a and b do not have an overlapping face, and if a and b have overlapping faces d_a and d_b, the weight is defined to be $|\{i : state(a_{d_a,i}) = state(b_{d_b,i}) = on\}|$. That is, the weight of two adjacent tiles is the total number of matching glue types from a and b's overlapping edges that are both in state on. Conceptually, each such pair of equal, on glues represents a bond between a and

b and thus increases the bonding strength between the tiles by 1 unit. For a positive integer τ, an assembly A is said to be τ-*stable* if the min-cut of the bond graph G_A is at least τ. For an assembly A, there is an associated pattern $p(A)$ defined by mapping the labels of each tile to corresponding lattice positions, and mapping the `null` label to lattice positions corresponding to locations not covered by the assembly.

Reactions. A reaction is an ordered pair of multi-sets of assemblies. Conceptually, a reaction (A, B) represents the assemblies of multi-set A replacing themselves with the assemblies in multi-set B. For a reaction $r = (A, B)$, let $r_{\texttt{in}}$ denote the multi-set A, and $r_{\texttt{out}}$ denote the multi-set B. For a set of reactions R, let $R_{\texttt{in}} = \bigcup_{r \in R} r_{\texttt{in}}$ and $R_{\texttt{out}} = \bigcup_{r \in R} r_{\texttt{out}}$.

A reaction (A, B) is said to be *valid* for a given temperature τ if it is either a *break*, *combination*, or *glue-flip* reaction as defined below:

- Break reaction. A reaction $(A = \{a\}, B = \{b_1, b_2\})$ with $|A| = 1$ and $|B| = 2$ is said to be a break reaction if the bond graph of a has a cut of strength less than τ that separates a into assemblies b_1 and b_2.

- Combination reaction. A reaction $(A = \{a_1, a_2\}, B = \{b\})$ with $|A| = 2$ and $|B| = 1$ is said to be a combination reaction if a_1 and a_2 are *combinable* into assembly b (see definition below).

- Glue-flip reaction. A reaction $(A = \{a\}, B = \{b\})$ with $|A| = 1$ and $|B| = 1$ is said to be a glue-flip reaction if assembly b can be obtained from assembly a by changing the state of a single glue slot x in b to either `on` from `latent` if x has queued command `on`, or `off` from `on` or `latent` if x has queued command `off`. Note that transitions among `latent`, `on`, and `off` form an acyclic graph with sink state `off`, implying glues states can be adjusted at most twice. This models the ``fire once'' property of signals.

Two assemblies a_1 and a_2 are said to be *combinable* if a_1 and a_2 can be translated such that a_1 and a_2 have no overlapping tile bodies, but have at least τ `on`, matching glues connecting tiles

from a_1 to tiles from a_2. Given this translated pair of assemblies, consider the product assembly b to be the assemblies a_1 and a_2 merged with the queued commands for each glue slot set according to the specifications of the glue functions for each tile with newly bonded `on` glues along the cut between a_1 and a_2. In this case we say a_1 and a_2 are *combinable* into assembly b. See Figure II.1 for example reactions and [23] for a more detailed presentation of the model.

Batches. A batch is a multi-set of assemblies, ie, a set of assemblies such that each assembly has a positive or infinite multiplicity. A batch B is said to be τ*-transitional* to a batch B' if the application of one of the break, combination, or transition rules at temperature τ can be applied to B to get B'. A *batch sequence* for some temperature τ is any sequence of batches $\langle a_1, \ldots a_r \rangle$ such that a_i is τ-transitional to a_{i+1} for each i from 1 to $r-1$.

Signal Tile System. A signal tile system is an ordered pair (B, τ) where B is a batch referred to as the *initial seed* batch, and τ is a positive integer referred to as the temperature of the system. Any batch B' is said to be *producible* by (B, τ) if there exists a batch sequence $\langle B_1, \ldots, B_r \rangle$ with respect to temperature τ such that $B' = B_r$ and $B = B_1$, i.e., B' is reachable from B by a sequence of τ-transitions.

II.I.III Exponential Replication

Our first primary definition towards the concept of exponential replication defines a transition between batches in which multiple reactions may occur in parallel to complete the transition. By counting the number of such parallelized transitions we are able to define the number of time steps taken for one batch to transform into another, and in turn can define the concept of exponential replication.

However, to avoid reliance on highly unlikely reactions, we parameterize our definition with a positive integer c which dictates that any feasible combination reaction should involve at least one combinate with at least multiplicity c. By doing so, our exponential replication definition will be able to exclude systems that might rely on the highly unlikely combination of low concentration combinates (but will still consider such reactions in a worst-case scenario by requiring the

subsequent monotonicity requirement). The following definition formalizes this concept.

(τ, c)-transitional distance. We say a batch B is (τ, c)-transitional to a batch B', with notation $B \rightarrow_{\tau,c} B'$, if there exists a set of reactions $R = \texttt{COMBO} \bigcup \texttt{BREAK} \bigcup \texttt{FLIP}$, where COMBO, BREAK, and FLIP partition R into the combination, break, and flip type reactions, such that:

1. $B - R_{\texttt{in}}$ is defined and $B' = B - R_{\texttt{in}} + R_{\texttt{out}}$.

2. For each $(\{x, y\}, \{z\}) \in \texttt{COMBO}$, the multiplicity of either x or y in $B - R_{\texttt{in}}$ is at least c.

Further, we use notation $B \rightarrow^{t}_{\tau,c} B'$ if there exists a sequence $\langle B_1, \ldots, B_t \rangle$ such that $B_1 = B$, $B_t = B'$, and $B_i \rightarrow_{\tau,c} B_{i+1}$ for i from 1 to $t - 1$. We define the (τ, c)-transitional distance from B to B' to be the smallest positive integer t such that $B \rightarrow^{t}_{\tau,c} B'$.

Our next primary concept used to define exponential replication is the concept of monotonicity which requires that a sequence of batches (regardless of how likely) has the property that each subsequent batch in the sequence is at least as close (in terms of (τ, c)-transition distance) to becoming an element of a given goal set of batches as any previous batch in the sequence.

Monotonicity. Let B be a batch of assemblies, τ a positive integer, and G a set of (goal) batches. We say B grows monotonically towards G at temperature τ if for all temperature τ batch sequences $\langle B, \ldots, B' \rangle$, if $B \rightarrow^{t}_{\tau,c} g$ for some $g \in G$, then $B' \rightarrow^{t'}_{\tau,c} g'$ for some $g' \in G$ and $t' \leq t$.

Note that g' in the above definition may differ from g. This means that B is not required to grow steadily towards any particular element of G, but simply must make steady progress towards becoming an element of G.

We now apply the concepts of (τ, c)-transition distance and monotonicity to define exponential replication of patterns. Informally, an STAM system is said to replicate the pattern of an assembly a if it is always guaranteed to have a logarithmic (in n) sequence of *feasible* transitions that will create at least n copies of a shape with a's pattern for any integer n. Further, to ensure that the system makes steady progress towards the goal of n copies, we further require the property of *monotonicity* which states that the number of transitions needed to attain the goal of n copies never increases, regardless of the sequence of reactions.

Exponential Replication. Let B_p^n denote the set of all batches which contain an n or higher multiplicity assembly with pattern p. A system $T = (B, \tau)$ exponentially replicates the pattern of assembly a if for all positive integers n and c:

1. $B \bigcup \{a\} \rightarrow_{\tau,c}^{t} B'$ for some $B' \in B_{p(a)}^n$ and $t = O(\texttt{poly}(|a|) \log(cn))$.

2. B grows monotonically towards $B_{p(a)}^n$.

Given the concept of a system replicating a specific assembly, we now denote a system as a general *exponential replicator* if it replicates all patterns given some reasonable format that maps patterns to input assemblies. Let M denote a mapping from rectangular patterns over some alphabet ϕ to assemblies with the property that for any rectangular pattern w over ϕ, it must be that 1) $w = p(M(w))$ (The assembly representing pattern w must actually have pattern w), 2) all tiles in $M(w)$ with the same non-null label are the same active tile up to translation, and 3) the number of tiles in $M(w)$ is at most an additive constant larger than the size of w. Such a mapping is said to be a *valid format mapping* over ϕ. We now define what constitutes an exponential pattern replicator system.

Exponential Replicator. A system $T = (B, \tau)$ is an exponential pattern replicator for patterns over ϕ if there exists a valid format mapping M over ϕ such that for any rectangular pattern w over ϕ, $T = (B, \tau)$ exponentially replicates $M(w)$.

II.II Replication of Linear Patterns

In this section, we focus on the replication of a linear assembly in two-dimensional space as a simplified version of the extended mechanism presented in Section II.III. In this replication scheme, which occurs at temperature 1, some pre-formed linear patterned template assembly R is added to the replicating tile set T to compose the *initial seed batch* (Figure II.2). In general, the mechanism described follows the simple model outlined by von Kiedrowski *et al.* for template-directed self-replication. However, our scheme has a difference in that two types of products are

formed: *terminal replicates* (*tr*) and *non-terminal replicates* (*ntr*). While the pattern of each type of replicate is identical to that of the parent, each replicate type serves a different function. *Non-terminal replicates* may catalyze the formation of more product while *terminal* replicates serve as an inert final product and may not catalyze the formation of more product. Each *non-terminal replicate* may serve as a template for the formation of another *ntr* and a *tr* concurrently.

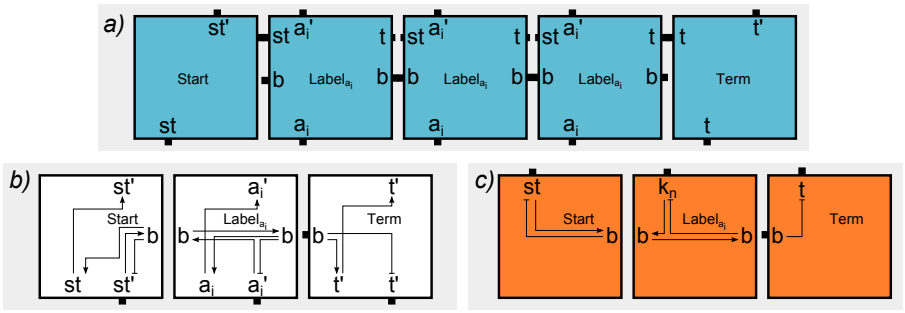

Figure II.2: The *initial seed batch* for replication of a 1xn patterned template. a) General form of template to be replicated R b) Tiles involved in formation of *non-terminal replicates*. c) Tiles involved in formation of *terminal replicates*.

Upon addition of R to the replicating tile set, tiles involved in *non-terminal replicate* formation (white) may attach to the north face glues of the template: a_i', st, and t. Simultaneously, tiles involved in *terminal replicate* formation (orange) may attach to the south face glues of the template: a_i, st, and t (Figure II.3a). Upon binding, b glues are activated on the west face of each $Label_{a_i}$ and $Term$ *ntr* or *tr* tiles. On *ntr* and *tr* $Start$ tiles, b glues are activated on the east face upon binding to the template. Note that the template has no active signals. After binding of the $Start$ tile, a signal is propagated from west to east along the newly forming replicate via the b glues. When a $Label$ tile has bound the b glues on both its east and west faces, it may detach from the parent template (Figure II.3b). Following complete detachment of the replicates from the parent template (Figure II.3c), the *terminal replicate* is inert and may not undergo any further binding events. The *non-terminal replicate*, however, can continue to catalyze the formation of product. The *non-terminal replicate* has south face a_i glues exposed, allowing it to immediately serve as a template for the

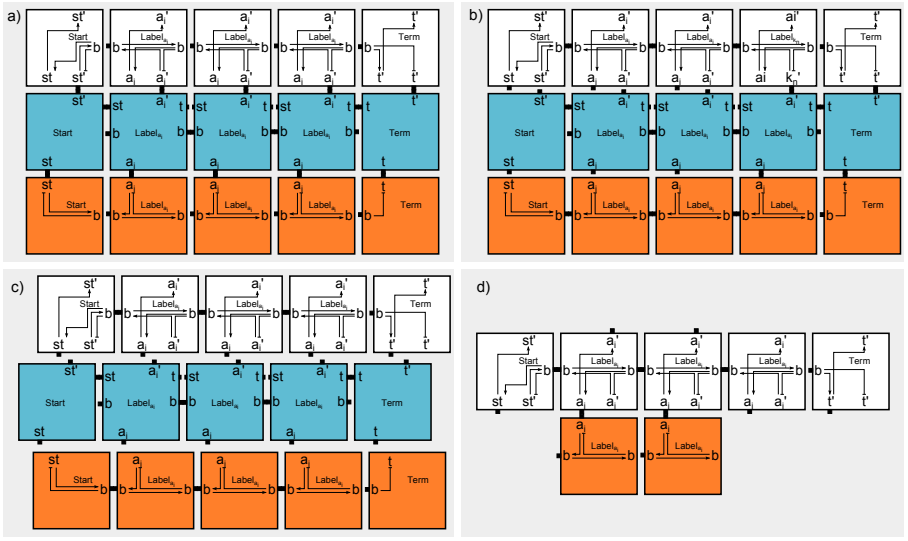

Figure II.3: a) The *ntr* (white) and *tr* (orange) bind to the template assembly (blue) via label glues a_i, a_i', st, and t. b) b glues are activated on the west face of the *ntr* and *tr* $Label$ and $Term$ tiles. The b blue on the $Start$ tile is activated on the east side, which propagates a signal through the newly forming replicate to detach from the parent. c)The newly-formed *terminal replicate* (orange) and *non-terminal replicate* (white) are completely detached from the parent template assembly. d) North face label glues on an *ntr* template assembly are activated only upon binding of *tr* tiles to the south face of the template assembly.

formation of a next-generation *terminal replicate*(Figure II.3d). Note that upon detachment, the north face label glues: a_i', st, and t of the *non-terminal replicate* are latent and are activated only upon the binding of *tr* tile to the south face of the *ntr*. This was designed so that *ntr* aggregates do not form, which would hinder the formation of *terminal replicates*. Following activation of the north face label glues, the *ntr* may serve as a template for the next-generation *ntr*s and *tr*s via the same mechanism as the original template assembly R.

II.III Replication of 2D Patterns in Two Dimensions

We first informally discuss the mechanism for replication of 2D patterns in two dimensions with the tileset shown in Figure II.4. The replication process described here can be summarized in three phases. In the first phase, *template disassembly*, a template R containing some pattern over some alphabet ϕ is combined with the tile set that can replicate R. Initially, an inverted staircase cooperatively grows along the west face of R. The effect of this tile growth is that each row of the original assembly R has a unique number of tiles appended to its west side. These appendages are used in reassembly later in the replication process. As the inverted staircase structure grows, rows of the original template are signaled to detach from each other. In Phase 2, the detached rows of the input assembly are available to serve as templates for the formation of *non-terminal replicates*. As described in Section II.II, two types of replicate products are formed: *terminal replicates* (*tr*) and *non-terminal replicates* (*ntr*). *Non-terminal replicates* may catalyze the formation of more product while *terminal* replicates serve as a final product and may not catalyze the formation of more product. After formation, this first generation of non-terminal replicates detach from the parent and enter Phase 3. In Phase 3, each *non-terminal replicate* may serve as a template for the formation of another *ntr* and a *tr* concurrently. The *tr* detaches from the parent upon completion and assembles, along with other terminal replicates, into a copy of R. Also during Phase 3, when the new *non-terminal replicate* is fully formed, it may detach from the parent and begin producing replicates.

We now present a more detailed synopsis of the replication mechanism. The 12 active tile types which comprise T are depicted in Figure II.4d-f. Note that the input pattern itself is not included in T. The input pattern to be replicated is of the form shown in Figure II.4c, and this, together with T, comprises the initial seed batch. The pattern is mapped onto this input via the composition of the *Label* signal tiles. Figure II.4a shows the tile types for a binary alphabet, while Figure II.4b shows the tile type for some a_i of alphabet ϕ which consists of elements $a_1, a_2, \ldots a_\phi$.

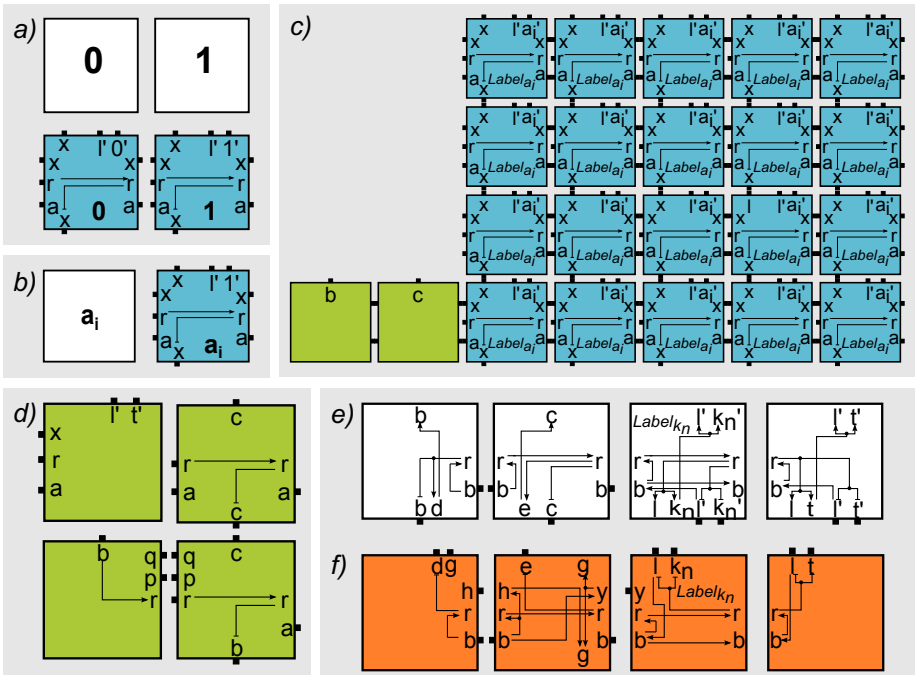

Figure II.4: a) Input assemby tile types for a binary alphabet b) The tile type for some a_i of alphabet ϕ which consists of elements $a_1, a_2, \ldots a_\phi$ c) General form of template to be replicated R d) Tiles involved in inverted staircase construction and disassembly of the original template. e) Tiles involved in formation of *non-terminal replicates*. f) Tiles involved in formation of *terminal replicates*.

Template disassembly and First Generation of Replicates. Upon addition of the template assembly R to the replicating tile set T, an inverted staircase forms on the west side of R (Fig. II.5a). Concurrently, an end cap attaches to the east side of R. Note that while the east-side end caps are attaching to R, it is possible that an *ntr* tile type (white) found in Fig. II.4e may attach to the north

side of an end cap, blocking the attachment of an endcap to a row. This does not adversely affect replication, because given a temperature of 2, the template will still disassemble and the end cap may attach to rows lacking end caps following this event. Also, given that the north face label glues a_i' of the northernmost template row are exposed, it is possible for this row to begin replicating immediately. In fact, this is necessary for the row immediately below the northernmost row to detach. Any row s of R may release the row below it by turning off its south face glues (Fig. II.5b). This can occur only if the row above s has activated the b glue on the westernmost tile of s. A signal is then propagated from west to east in row s via glue r and all south-face glues of s are turned off.

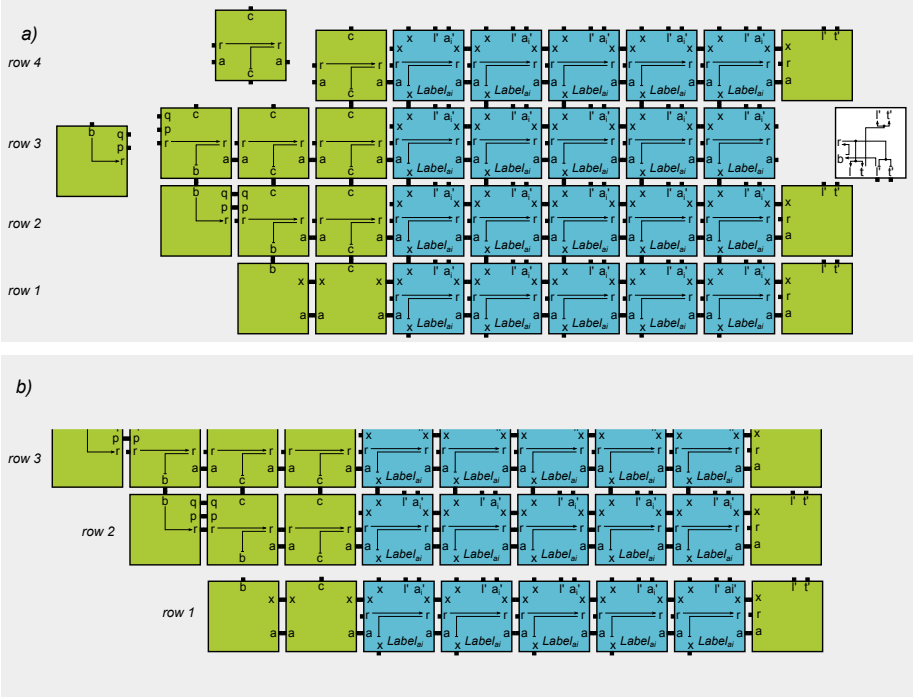

Figure II.5: a) Growth of inverted staircase along the west face of R and a cap on the east face of R. b) Row 1 is released after the b glue is activated on the westernmost tile of Row 2.

Following R disassembly, label glues a_i' are exposed on the north face of each row of the input assembly. Tiles involved in *ntr* formation (white) may attach along the north face of the template row (blue/green) (Fig. II.6a). Following attachment, west face b glues are turned on. Once the westernmost *Label* tile has attached, appendage tiles may cooperatively attach, sending a signal

via b glues from west to east and turning on r glues. (Fig. II.6b). After the westernmost appendage tile has attached, a signal is propagated from west to east via glue r queueing label glues a_i' on the south face of the new *ntr* to turn off, thus detaching the *ntr* from its parent (Fig. II.6c). Label glues a_i are also queued on. These glues serve to generate a terminal replicate (*tr*) on the south face of the *ntr* (Fig. II.6d). Following the detachment of the *ntr* and the parent template, the parent template is available to generate another *ntr*, while the first-generation *ntr* is immediately available to generate a *tr*. Exponential Replication and Reassembly After the formation of the first-generation

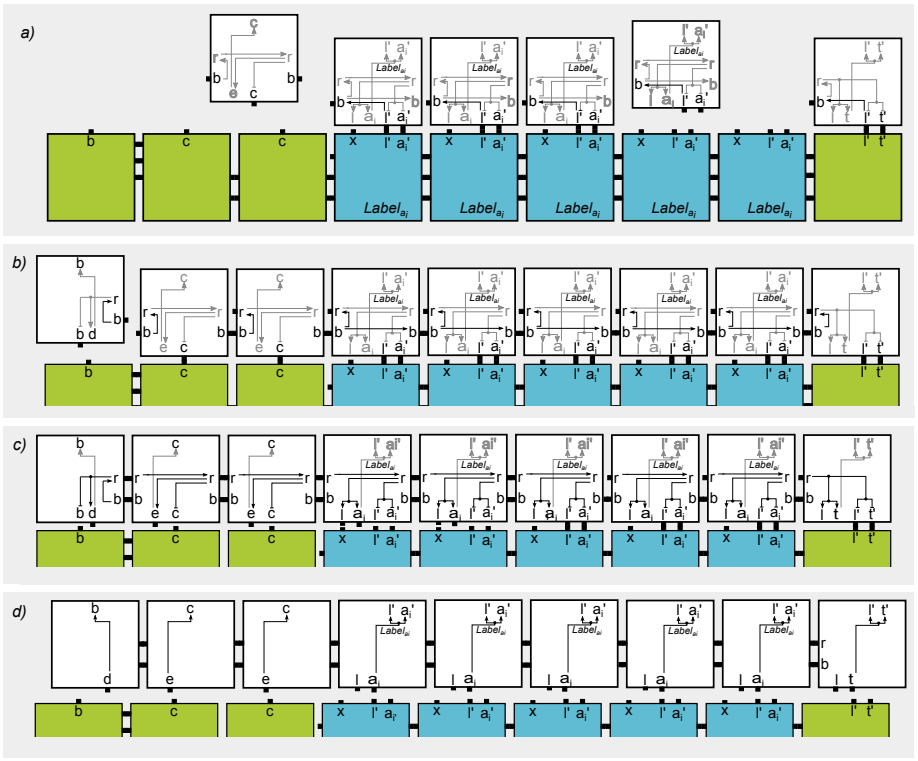

Figure II.6: The above sequence outlines details of the production of *non-terminal replicates*. For clarity, glues turned *off* and signals previously executed are not shown.

ntr, replication is free to proceed exponentially. Glues on the south face of the *ntr* may bind label tiles from the *tr* tile set (Fig. II.7a). Upon binding, b glues are turned *on* on the west face of the *tr* label tiles, allowing for the binding of appendage tiles on the western side of the growing *tr* assembly. Upon binding of the first appendage tile (Fig. II.7b), a signal is propagated through the

tr via glue b and r glues on the west faces of the *tr* tiles. After the next appendage tile binds, the y glue on the tile adjacent to it is activated, which activates two g glues on the north and south faces of the easternmost appendage tile (Fig. II.7c). These g glues will assist in proper reassembly of each row into a correct copy of the template R. Also note that upon binding a *tr* tile, label glues a_i' on the north face of the *ntr* are turned *on*. This allows for synthesis of a new *ntr* on the north side

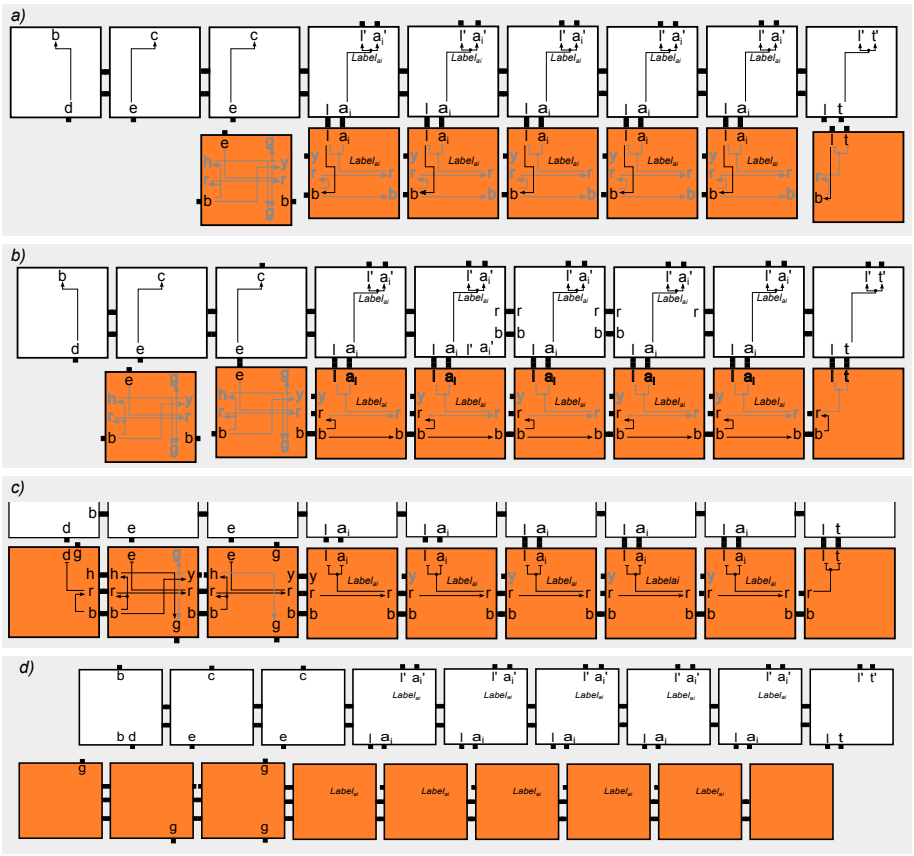

Figure II.7: The above sequence shows details of the formation of *terminal replicates*. For clarity, glues turned *off* and signals executed during template disassembly are not shown.

of the parent *ntr* while a new *tr* is being formed on the south face. The synthesis of a new *ntr* from a parent *ntr* is not described in detail here, as it is very similar to the process described in Figure II.6. Upon attachment of the westernmost appendage tile, a north face g glue of the *tr* is turned *on* as well as a south face g glue on the tile immediately adjacent to it. Additionally, a signal is

propagated from west to east along the *tr* via glue r and the north face glues of the *tr* are turned *off*. The *tr* then detaches from the parent *ntr* (Fig. II.7d) and is available for reassembly into a copy of the original template R while the parent *ntr* is available to produce a new *ntr* on its north face and a new *tr* on its south face. The alignment of g glues enables the proper reassembly of the *terminal replicates* into a copy of R (Fig. II.8).

The detachment of the inverted staircase is not described here. If a signal cascade were designed such that upon the complete assembly of a copy of the original template pattern, the inverted staircase detached, it would be considered a waste product. The number of these waste assemblies would grow proportionally to the number of replicates of R. Similarly, if the replication process were somehow halted, and the copies of R harvested, the *ntr*s might also be considered waste. These, too, would have grown proportionally to the copies of R.

Theorem 1. For any alphabet ϕ, there exists an exponential pattern replicator system $\Gamma = (T, 2)$ for patterns over ϕ. Furthermore, the seed batch T consists of $O(\phi)$ distinct singleton active tile types with a total of $O(\phi)$ unique glues.

Proof. To prove this we argue that the STAM system $(T, 2)$ defined by the tileset T depicted in Figure II.4b-c is an exponential pattern replicator. The valid format mapping M for the system is depicted in Figure II.4a.

We now argue that for any $w \times \ell$ pattern P, the assembly $A_P = M(P)$ derived by applying the format mapping described in Figure II.4a to pattern P is exponentially replicated by $(T, 2)$. First, by Lemma 5, the system satisfies the monotonicity requirement of the exponential replication. We therefore focus on the remaining requirement that for any positive integers n and c, the (τ, c)-transition distance from $T \bigcup A_P$ to some batch with at least n copies of an assembly with pattern P is $O(\log(n + c))$. To show this, we construct a (τ, c)-transitional sequence of batches $\langle T \bigcup \{A_P\}, \ldots, B_{cleanBreak}, \ldots, B_{nReplicateRows}, \ldots, B_{nFinalRows}, \ldots, B_{nPatterns} \rangle$ with the property that batch $B_{cleanBreak}$ contains 1 non-terminal replicate of each row of the initial input assembly A_p, $B_{nReplicaterows}$ contains at least $n+c$ copies of the non-terminal replicate assembly for each row

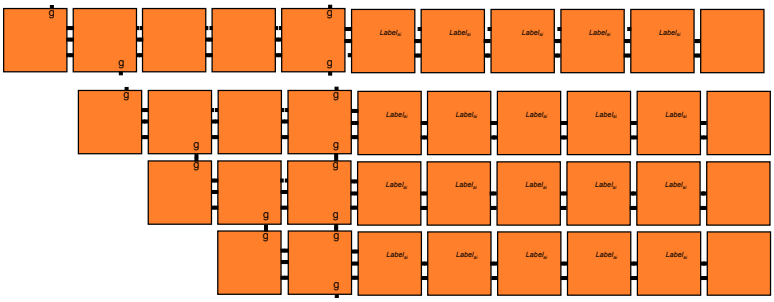

Figure II.8: *Terminal replicates* reassemble into a copy of R.

of the input assembly, $B_{nFinalRows}$ contains at least $n + c$ terminal replicates of each assembly row, and finally $B_{nPatterns}$ contains at least n copies of an assembly with pattern P. The construction for each segment of this sequence is depicted in Lemmas 1, 2, 3, and 4. From these Lemmas we get that the desired sequence of batches can be constructed with length at most $O(\log(n + c))$.

□

Lemma 1. For any $w \times \ell$ rectangular patterned input assembly P, let the batch $B_p = T \bigcup A_p$. For some sequence $\langle B_p, B_{p+1}, \ldots B_{cleanBreak} \rangle$, the (τ, c) transitional distance from B_p to $B_{cleanBreak}$ is $O(\ell + w\ell)$ where $B_{cleanBreak}$ contains at least one non-terminal replicate of each row within the input assembly.

Proof. We first consider the cooperative binding of inverted staircase tiles. The entire inverted staircase forms cooperatively in $x = \frac{(\ell+1)(\ell+2)}{2} - 3$ combination reactions. Therefore, the (τ, c) transitional distance for the formation of the inverted staircase is x. For each row, signals must traverse west-to-east and then east-to-west across the entire row for the row beneath to completely detach. The number of glue-flip reactions required over the entire input assembly P for detachment of all rows from on another is $2((\ell-1)(w-1)+\frac{(\ell+1)(\ell+2)}{2}-3)$. We then end up with input assembly rows with no other tiles attached. It follows from the analysis in case 1 that $O(w + \ell)$ transition steps are sufficient to generate clean non-terminal replicates.

Therefore, from any batch B_k, there exists a $O(w\ell + \ell^2)$ (τ, c) transitional distance to batch $B_{cleanBreak}$, where batch $B_{cleanBreak}$ contains a clean non-terminal replicate of each row of the

input assembly .

□

Lemma 2. Consider that batch B_k contains one clean non-terminal replicate of each input assembly row. Then there exists a (τ, c) transitional distance of $O((\ell + w) \log (n + c))$ for the batch transition sequence $\langle B_k, B_{k+1}, \ldots B_{ReplicateRows} \rangle$ such that $B_{ReplicateRows}$ contains $n + c$ copies of non-terminal replicates of each input assembly row.

Proof. For any newly-assembled *ntr* to generate an identical offspring *ntr*, it must first bind *terminal replicate* tiles in order to activate its north-face glues which serve as templates to bind *ntr* tiles. This is a one-time activation event, after which an *ntr* may generate unbounded copies of identical *ntrs*. For any newly generated *ntr*, at most $w+\ell+2$ combination reactions and $2w+\ell+4$ glue-flip reactions must occur to activate the north-face glues on the parent *ntr*. Once an *ntr* has been activated, *ntr* label tiles and the easternmost cap tile bind in $w + 1$ combination reactions. Following these combination reactions, a series of at most $8w + 7\ell + 4$ glue-flip and combination reactions are required to fully connect the newly-formed *ntr* and detach it from the parent input assembly row. In total, there is a $12w + 9\ell + 11$, or $O(\ell + w)$ (τ, c) transitional distance *ntr* exist for each *ntr* in a batch to be activated and generate an identical *ntr*. Thus, after every $O(\ell + w)$ transitions, the population of *ntrs* doubles. Therefore, there is a $O((\ell + w) \log (n + c))$ (τ, c) transitional distance to achieve $n + c$ copies of a non-terminal replicates of each input assembly row. □

Lemma 3. Consider that batch B_k has n clean non-terminal replicates of each row of the input assembly. There exists a batch sequence $\langle B_k, B_{k+1}, \ldots B_r \rangle$ with a (τ, c) transitional distance of $O(w + \ell)$ from B_k to $B_{FinalRows}$ where batch $B_{FinalRows}$ contains $n + c$ terminal replicates of each row.

Proof. Terminal replicate label tiles, the easternmost capping tile, and the tag tiles must bind with a non-terminal assembly in at most $w+\ell+2$ combination reactions. Following these tile attachments, $8w + 7\ell + 5$ glue-flip and combination reactions are required to fully connect the newly-formed

tr and detach it from the parent *ntr*. Therefore, $9w + 8\ell + 7$ parallel reactions exist to produce n *terminal replicates* from n *non-terminal replicates*. □

Lemma 4. Consider that batch B_k has $n+c$ terminal replicates of each row. There exists a batch sequence $\langle B_k, B_{k+1}, \ldots B_{nPatterns} \rangle$ with a (τ, c) transitional distance of $O(\ell)$ where batch $B_{nPatterns}$ contains n identical assemblies a where $P(a)$ is identical to the input assembly.

Proof. Upon detachment from the parent, terminal replicates have g glues oriented such that they may attach to the correct neighbors within the patterned assembly. For these terminal replicates to combine into assembly that has an identical pattern to the input assembly, ℓ rows must attach to one another in $\ell - 1$ combination reactions. Therefore, $O(\ell)$ parallel batch transitions exist for batch B_k to transition to batch $B_{nPatterns}$. □

Lemma 5. Consider the seed batch B_p where $B_p = T \bigcup A_p$. B_p grows monotonically toward G where G contains at least n copies of an assembly with pattern P

Proof. For any valid batch sequence $\langle B_p, B_1, \ldots, B_k \rangle$, let the (τ, c) transitional distance from batch B_k to B_r be x where B_r contains n identical assemblies a where $P(a)$ is identical to the initial input pattern. For some batch sequence $\langle B_k, B_{k+1}, \ldots, B_\ell \rangle$, let the (τ, c) transition distance from B_ℓ to B_r be y. We transition B_k and B_ℓ to the nearest 'perfect' batches P_k and P_ℓ, respectively, where the ntrs that comprise P_k and P_ℓ have no tiles attached to their north or south faces. This means that during the transition from B_k to P_k or from B_ℓ to P_ℓ, any partially formed trs or ntrs on the ntr templates are completed and detach. Therefore, batches P_k and P_ℓ will be comprised of some number of completed trs (some may be combined with one another to form copies of a or partial copies of a) and ntrs with no other tiles attached. Of these ntrs, there are two classes: passive ntrs and active ntrs. Passive ntrs have no more active signals and may serve as a template for the formation of a tr and an ntr concurrently, which form independently of each other. Active ntrs are those that have just been released from their parent ntr and have not yet served as a template. These must first serve as a template for a tr. Upon formation of this first child tr, the active ntr

will become passive. Because, an ntr must first serve as a template for at least one tr before an ntr, for any perfect batch P_i, the number of trs in the batch must be at least as great as the number of passive ntrs. The (τ, c) transitional distance from $\langle B_i, B_{i+1}, \ldots, P_i \rangle$ is $O(\ell + w)$.

Because B_k can transition into B_ℓ, the numbers of passive ntrs, active ntrs, and trs must each be at least as large in P_ℓ as in P_k when $P_k \neq P_\ell$. Therefore, when $P_k \neq P_\ell$, $x \leq y$, thus satisfying the monotonic growth requirement.

When $P_k = P_\ell$, we consider the transition $\langle B_k, B_{k+1}, \ldots, P_k, \ldots, Pk + 1 \rangle$. Because B_k may transition into B_ℓ, B_ℓ can mimic the path of B_k from $P_k = P_\ell$ to $Pk + 1$ in at most as many steps as B_k.

□

II.IV Future Work

The results of this paper provide several directions for future work. One interesting problem is the replication of shapes in the STAM, or more specifically, patterned shapes. One might imagine a mechanism similar to the one presented in this paper but where the growth of the inverted staircase is preceded by a ``rectangularization" of the shape to be replicated. The replication of a cuboid is conceivable by extending the mechanism of template disassembly and reassembly presented in Section II.III to three dimensions where layers of the cuboid might be separated, replicated, and the replicates reassembled. Precise replication of a certain number of copies could also be possible, as was considered in [1].

Another direction for future work is studying the extent to which staged self-assembly systems can be simulated by non-staged active self-assembly systems such as the signal tile model. In [7] efficient staged algorithms are developed to assemble linear structures, while a signal tile system achieves a similar result in [23]. Shape replication through stages and RNA based tiles are used to replicate general shapes in [1], while this paper and future work suggests similar results may be obtained with signal tiles. Can the complexity of the mixing algorithm of a staged assembly

algorithm be encoded into a signal tile system of similar complexity? As a first step towards such a simulation we might consider the case of 1D assemblies. Can the efficient construction of labeled linear assemblies through staging shown in [8] be efficiently simulated with a signal tile system?

A final direction for future work involves the simulation of the signal tile model through a passive model of self-assembly such as the abstract or two-handed tile assembly model [4]. Recent work has shown how restricted classes of signal tile systems can be simulated by passive 3D systems [**?**]. The ability for signal tile systems to perform *fuel-efficient* computation was shown to be achievable within passive 2D tile assembly given the added power of negative force glues [**?**]. Is it possible to simulate *any* signal tile system with the use of negative glues? Can this be done in 2D?

CHAPTER III

GROWING NON-DETERMINISTIC LINEAR ASSEMBLIES

DNA tile self-assembly describes the process by which small tiles comprised of DNA come together spontaneously via local interactions to create larger, more complex nanostructures. Information about the size, shape and pattern of the structure to be assembled by these tiles is programmed into the tiles themselves so that no outside guidance is needed. A significant challenge in the design of these promising self-assembling systems is minimizing the number of tile types required to uniquely assemble a given construct. This minimization of tile complexity is the most studied problem in DNA tile self-assembly, and is crucial to the ability to program complex molecular self-assembling systems because the total number of tile types largely determines the cost of implementation. A given tile type requires significant characterization before use, and must be tested with other tiles in the set for unintended interactions and hybridizations. Furthermore, there is a fundamental limit to the number of tiles that can be constructed with DNA sequences of a fixed length. Compared to the breadth of theoretical work related to tile complexity reduction for a target assembly, experimental work on the topic is scarce. Therefore, experimentally exploring methods to mitigate this challenge represents a valuable contribution to the basic construction and characterization of a modular DNA self-assembly toolkit.

Broadly, this project aims to test the effectiveness of and improve upon two theoretical models that could significantly reduce the tile complexity of target shapes. In DNA nanoscience, one particularly important construction is the ruler, a linear assembly of a specified length that is composed of unit tiles. These rulers can be used as nanoscale beams and struts, or to define the boundaries of a growing assembly. Rulers can also serve as nucleation sites for more complex structures

because they can encode the binary description for two-dimensional shapes. In deterministic tile self-assembly, tile complexity for any length n line is $O(n)$ which is also the matching upper bound for the size of any tile set in deterministic tile self-assembly. Thus, it becomes impractical if not impossible to form large specified-length linear assemblies in deterministic tile assembly. This project aims to construct large linear assemblies with sub-linear-sized tile sets using non-deterministic tile self-assembly systems.

Current experimental work in DNA tile assembly is deterministic in the sense that only one tile type may attach to a given position in a partially formed assembly. One possibility to address the tile type reduction problem is to allow for non-determinism during the course of the assembly process. The Probabilistic Tile Assembly Model (pTAM) and the concentration programming model are non-deterministic in the sense that they allow for the possibility of more than one tile type attaching in a given place to the partially formed assembly. This non-determinism during assembly simply allows for the construction of a assemblies with some distribution of sizes and/or shapes. Using different techniques, these distributions can be tightened around some expected assembly x while using fewer tile types that would be required to deterministically assemble x. This work presents an advancement in the ability to efficiently build linear assemblies non-deterministically with very few tile types. We attempt to achieve linear assemblies of some expect length L within some tight distribution of assemblies. This first step toward minimizing tile complexity using the methods described in this thesis can later be extended into the self-assembly of complex shapes with greater efficiency.

While there are many DNA tile implementations in use, we propose to begin with the simplest of DNA tile implementations-- the oligomer. These oligos are be programmed to non-deterministically self-assemble into a nicked double helix (our linear assembly) of some expected length L with some tight distribution. Both mechanisms to be used to accomplish this require that assembly begins with some initiator, or seed tile and that the linear assemblies grow unidirectionally. We enforce these constrictions by utilizing a hybrid chain reaction first described by Dirks and Pierce and summa-

rized in Section III.I.

III.I Hybridization Chain Reaction

To enforce the notion that assembly begins at some seed tile and grows unidirectionally to form a linear assembly, we utilize a hybridization chain reaction, first described and implemented by [9]. In this mechanism, a mixture of stable DNA monomeric hairpin tiles assemble into a nicked double helix of alternating fragments only upon exposure to some linear DNA fragment, or seed. A simple example of this process is shown in Figure III.1.

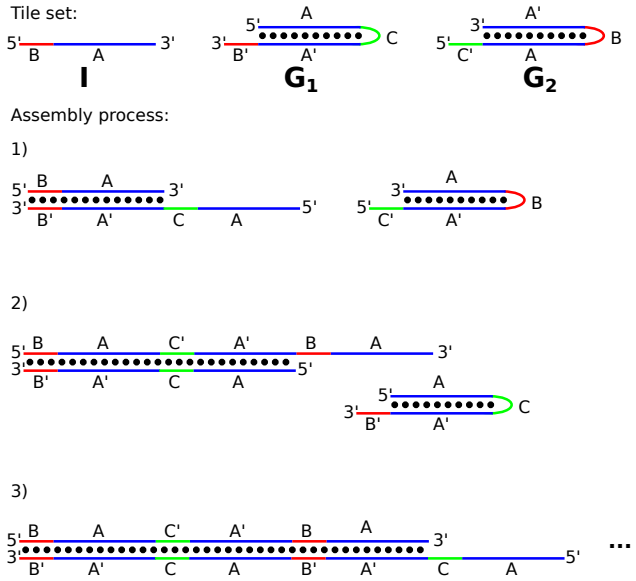

Figure III.1: Basic HCR System. Tile Set: Secondary structure of DNA tiles I, G_1 and G_2. Assembly Process: 1) Initiator tile I binds to the toehold on tile G_1 and forces the hairpin open, exposing region C on tile G_1. 2) The toehold of tile G_2 may then bind region C on tile G_1, forcing tile G_2 open and exposing region B. 3) The toehold of tile G_1 may then bind to region B on tile G_2, forcing tile G_1 open. Steps two and three are repeated as the resulting nicked double helix grows.

III.II Concentration Programming

III.II.I Single Phase

Experimental implementation of this scheme using HCRs requires the use of four distinct hairpin tile types which we designate I, $G_{1,1}$, $G_{1,2}$, and T_1 with glues assigned such that non-deterministic

assembly of any sequence of tiles described by the regular expression $IG_1(G_2G_1)^*T$ is possible (Figure III.2). The seed tile I has some very small concentration compared to p and q, the concentrations of the growth tile $G_{1,2}$ and the terminating tile T_1, respectively. Growth tile G_1 has concentration $p + q$. Please see Section III.IV for technical details on implementation.

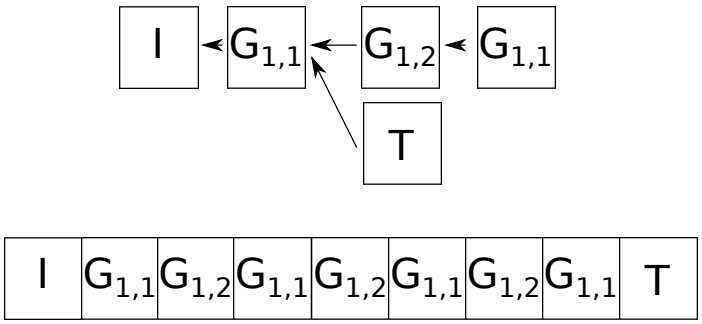

Figure III.2: Top: Tiles involved in assembly with arrows indicating binding possibilities. Bottom: Possible assembly.

III.II.II Multiple Phase

Experimental implementation of this scheme using HCRs requires the use of an initiator tile I, and three tile types for each stage i: two growth tiles $G_{i,1}$ and $G_{i,2}$, and a terminating tile for the stage T_i. The concentration m of tile type I is very small compared to p and q, which are the concentrations of T_i and $G_{i,2}$, respectively. Please see Section III.IV for technical details on implementation.

III.III Randomized Self-Assembly

As an alternative to programming the relative concentrations of tile types to influence the probability of binding, another non-deterministic self-assembly scheme for construction of linear assemblies with very few tile types was described by [5]. In its simplest form, this mechanism involves the use of a seed tile, reset tiles $R_1R_2...R_{n-1}$, and terminating tiles $T_1T_2...T_n$. Each reset tile R_i deterministically binds tile T_1 while each terminating tile T_i where $i < n$ may bind either T_{i+1} or R_{i+1}. The binding of all terminating tiles in a contiguous sequence $T_1T_2...T_n$ results in the termina-

tion of the assembly. This process is akin to flipping a biased coin until n contiguous *heads* appear. The expected length L of the linear assembly (the expected number of Bernoulli trials), before the consecutive binding of $T_1T_2..T_n$ termination tiles (consecutively flipping n heads) is $L = \frac{1-p^m}{p^m q}$ when we assign p as the concentration of T_i and q as the concentration of R_i.

For implementation using HCR, this mechanism would require an initiator tile I, termination tiles $pT_1pT_2....pT_n$, reset tiles $R_1R_2...R_{n-1}$, and spacer tiles $S_1, S_2...S_n$. Reset tiles R_i and spacer tiles S_{i+1} compete to bind to the exposed east side glue of T_i tiles. If spacer tile S_{i+1} binds, then terminating tile pT_{i+1} deterministically binds to the east side of S_{i+1}. If R_i binds, then terminating tile pT_1 deterministically binds to the east side of R_i and the process of accumulating contiguous termination tiles begins again (Figure III.3). This approach was not addressed in this thesis, however tiles were designed for this approach and can be seen in Figure III.5.

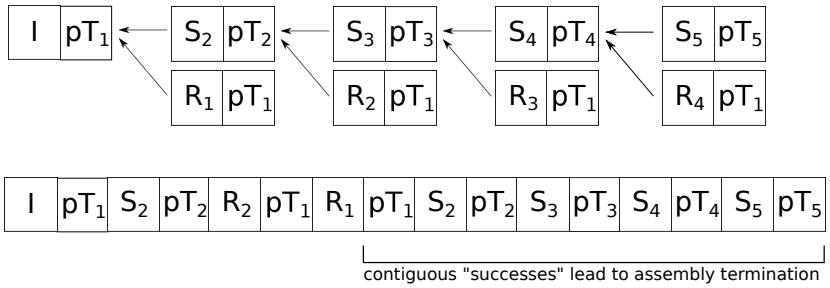

Figure III.3: Pre-attached tiles represent tiles that deterministically bind together during the assembly process. Arrows represent binding possibilities.

III.IV Initial Experiments

III.IV.I Tile Design and Preparation

Please see Figures III.4 and III.5 for tile sequences. These 48-bp sequences were designed to minimize sequence symmetry and to maximize the probability of adopting the target secondary structure at equilibrium. The sequence of the hybridized stem of all hairpin tiles was taken from [9], as were tiles I, $G_{1,1}$ and $G_{1,2}$. The toeholds and hairpin loop sequences of the remaining tiles were generated randomly and then assessed for the desired characteristics (i.e. minimizing non-specific

binding).

Concentration Programming Tile Set

I	5′-AGTCTAGGATTCGGCGTGGGTTAA-3′
$G_{1,1}$	5′-TTAACCCACGCCGAATCCTAGACTCAAAGTAGTCTAGGATTCGGCGTG-3′
$G_{1,2}$	5′-AGTCTAGGATTCGGCGTGGGTTAACACGCCGAATCCTAGACTACTTTG-3′
T_1	5′-AGTCTAGGATTCGGCGTGCCTATTCACGCCGAATCCTAGACTACTTTG-3′
$G_{2,1}$	5′-AATAGGCACGCCGAATCCTAGACTTGAGCAAGTCTAGGATTCGGCGTG-3′
$G_{2,2}$	5′-AGTCTAGGATTCGGCGTGCCTATTCACGCCGAATCCTAGACTTGCTCA-3′
T_2	5′-AGTCTAGGATTCGGCGTGTGTGTCCACGCCGAATCCTAGACTTGCTCA-3′
$G_{3,1}$	5′-GACACACACGCCGAATCCTAGACTCGTTACAGTCTAGGATTCGGCGTG-3′
$G_{3,2}$	5′-AGTCTAGGATTCGGCGTGTGTGTCCACGCCGAATCCTAGACTGTAACG-3′
T_3	5′-AGTCTAGGATTCGGCGTGCAGACTCACGCCGAATCCTAGACTGTAACG-3′

Figure III.4: Tile sequences for the concentration programming implementation with up to 3 stages.

III.IV.II Materials and Methods

HPLC purified oligo sequences I, $G_{1,1}$, $G1,2$, and T_1 were purchased from Integrated DNA Technologies (Coralville, IA) and diluted in water to $100mM$. As per [9], concentrated DNA tile stock solutions were diluted in reaction buffer to a final concentration of $50mM Na_2HPO_4$ and $0.5NaCl(pH6.8)$. Before mixing tile types together, samples were heated to 95 °C for 1 minute and then allowed to cool to room temperature for 1 hour. For reactions $1-6$, which served to replicate experiments from [9], $9\mu L$ of each species was combined. For reactions $7-10$ and $A-H$, $7\mu L$ of each species was combined. For reactions $I-M$, $4\mu L$ if each species was combined. In all cases, tiles were combined in an order such that the tile that the initiator binds was added last. Reactions were incubated at room temperature for 24 hours before analysis of reaction products on a 2% agarose gel.

III.IV.III Results

Replication of the work of Dirks et al. is show in reactions $1-6$ in Figure III.6 and an initial experiment with single-phase concentration programming is shown in reactions $7-10$ in Figure

III.6. Further experiments with single-phase concentration programming are shown in reactions $A - H$ and results from two-phase concentration programming are show in reactions $H - M$ in Figure III.7.

Probabilistic Assembly Tile Set

I	5′-AGTCTAGGATTCGGCGTGGGTTAA-3′
pT1($G_{1,1}$)	5′-TTAACCCACGCCGAATCCTAGACTCAAAGTAGTCTAGGATTCGGCGTG-3′
S2(T_1)	5′-AGTCTAGGATTCGGCGTGCCTATTCACGCCGAATCCTAGACTACTTTG-3′
pT2($G_{2,1}$)	5′-AATAGGCACGCCGAATCCTAGACTTGAGCAAGTCTAGGATTCGGCGTG-3′
S3(T_2)	5′-AGTCTAGGATTCGGCGTGTGTGTCCACGCCGAATCCTAGACTTGCTCA-3′
pT3($G_{3,1}$)	5′-GACACACACGCCGAATCCTAGACTCGTTACAGTCTAGGATTCGGCGTG-3′
S4(T_3)	5′-AGTCTAGGATTCGGCGTGCAGACTCACGCCGAATCCTAGACTGTAACG-3′
pT4	5′-AGTCTGCACGCCGAATCCTAGACTTGACTGAGTCTAGGATTCGGCGTG-3′
S5	5′-AGTCTAGGATTCGGCGTGGGTTCACACGCCGAATCCTAGACTCAGTCA-3′
pT5	5′-TGAACCCACGCCGAATCCTAGACTTTCGTCAGTCTAGGATTCGGCGTG-3′

R1	5′-AGTCTAGGATTCGGCGTGGGTTAACACGCCGAATCCTAGACTACTTTG-3′
R2	5′-AGTCTAGGATTCGGCGTGGGTTAACACGCCGAATCCTAGACTTGCTCA-3′
R3	5′-AGTCTAGGATTCGGCGTGGGTTAACACGCCGAATCCTAGACTGTAACG-3′
R4	5′-AGTCTAGGATTCGGCGTGGGTTAACACGCCGAATCCTAGACTCAGTCA-3′

Figure III.5: Tile sequences for the probabilistic assembly implementation.

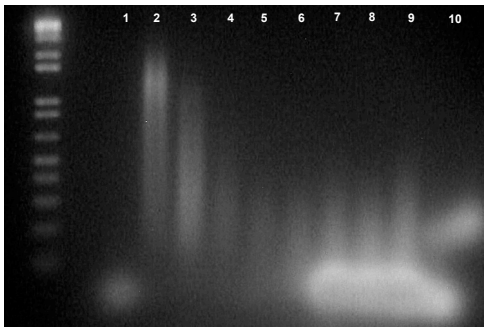

Figure III.6: Results of reactions 1-10. Please see Table III.1 for reaction contents.

Table III.1: Reaction Contents Experiment 1

	I	$G1,1$	$G1,2$	$T1$
Reaction 1	$0.0\mu M$	$1\mu M$	$1\mu M$	$0\mu M$
Reaction 2	$0.1\mu M$	$1\mu M$	$1\mu M$	$0\mu M$
Reaction 3	$0.32\mu M$	$1\mu M$	$1\mu M$	$0\mu M$
Reaction 4	$1.0\mu M$	$1\mu M$	$1\mu M$	$0\mu M$
Reaction 5	$3.2\mu M$	$1\mu M$	$1\mu M$	$0\mu M$
Reaction 6	$10.0\mu M$	$1\mu M$	$1\mu M$	$0\mu M$
Reaction 7	$0\mu M$	$10\mu M$	$10\mu M$	$10\mu M$
Reaction 8	$0.1\mu M$	$10\mu M$	$10\mu M$	$10\mu M$
Reaction 9	$0.1\mu M$	$10\mu M$	$10\mu M$	$5\mu M$
Reaction 10	$0.1\mu M$	$10\mu M$	$5\mu M$	$10\mu M$

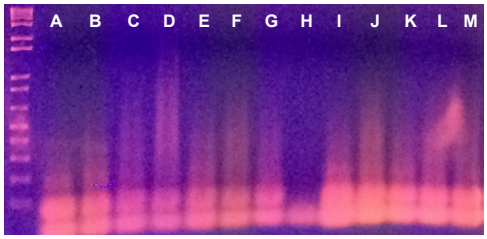

Figure III.7: Results of reactions A-M. Please see Table III.2 for reaction contents.

III.IV.IV Discussion

Replication of the work of Dirks et al. in reactions was successful. In reactions 7, A, and I we saw some nucleation and growth where there should have been none due to a lack of initiator. However, this is not seen in reaction H where the concentrations of tiles $G_{1,1}$, $G_{1,2}$, and T_1 were lower, which suggests that the high concentrations of growth and/or termination tiles initiated unintended growth. This may be mitigated in future experiments by working with lower tile concentrations. However, the ratio of growth and termination tiles to initiator tiles must remain high in order to approximate constant tile concentrations over time.

Experiment 1 indicates that the addition of tile T_1 to the reactions does halt growth, and reactions A-H in Experiment 2 indicate that decreasing the concentration of T_1 results in longer assemblies for single-phase concentration programming. The results for two-phase concentration programming (reactions I-M in Experiment 2) remain inconclusive. Lowering tile concentrations

Table III.2: Reaction Contents Experiment 2

	I	$G1,1$	$G1,2$	$T1$	$G2,1$	$G2,2$	$T2$
Reaction A	$0.0\mu M$	$5\mu M$	$5\mu M$	$5\mu M$	$0\mu M$	$0\mu M$	$0\mu M$
Reaction B	$0.1\mu M$	$5\mu M$	$5\mu M$	$5\mu M$	$0\mu M$	$0\mu M$	$0\mu M$
Reaction C	$0.1\mu M$	$5\mu M$	$5\mu M$	$2.5\mu M$	$0\mu M$	$0\mu M$	$0\mu M$
Reaction D	$0.1\mu M$	$5\mu M$	$5\mu M$	$1\mu M$	$0\mu M$	$0\mu M$	$0\mu M$
Reaction E	$0.1\mu M$	$5\mu M$	$2.5\mu M$	$5\mu M$	$0\mu M$	$0\mu M$	$0\mu M$
Reaction F	$0.1\mu M$	$5\mu M$	$2.5\mu M$	$2.5\mu M$	$0\mu M$	$0\mu M$	$0\mu M$
Reaction G	$0.05\mu M$	$5\mu M$	$5\mu M$	$1\mu M$	$0\mu M$	$0\mu M$	$0\mu M$
Reaction H	$0.0\mu M$	$1\mu M$	$1\mu M$	$1\mu M$	$0\mu M$	$0\mu M$	$0\mu M$
Reaction I	$0.0\mu M$	$5\mu M$	$5\mu M$	$5\mu M$	$5\mu M$	$5\mu M$	$5\mu M$
Reaction J	$0.1\mu M$	$5\mu M$	$5\mu M$	$4\mu M$	$5\mu M$	$5\mu M$	$4\mu M$
Reaction K	$0.1\mu M$	$5\mu M$	$5\mu M$	$3\mu M$	$5\mu M$	$5\mu M$	$3\mu M$
Reaction L	$0.1\mu M$	$5\mu M$	$5\mu M$	$2\mu M$	$5\mu M$	$5\mu M$	$2\mu M$
Reaction M	$0.1\mu M$	$5\mu M$	$5\mu M$	$1\mu M$	$5\mu M$	$5\mu M$	$1\mu M$

so that unintended nucleation and growth does not occur may produce results more concurrent with theoretical results.

III.V Future Work

The results in Figure III.6 and Figure III.7 serve as a starting point in supporting the theoretical work in [19]. Due to time constraints, we were unable to perform further experiments at publication of this thesis. Optimization of experimental conditions is required to alleviate many of the concerns outlined in III.IV.IV. The mathematically predicted behavior of the each relies on the assumption that tile type concentrations are 'fixed' as self-assembly proceeds. In reality, however, concentrations of each tile type will diminish over time. One solution to this is to start with very high tile type concentrations, however this was found to encourage nonspecific tile interactions.

Once experimental conditions are optimized, further work may involve increasing the number of phases in concentration programming, as well as working with the randomized tile assembly scheme outline in Section III.III. Each model of assembly addressed here has potential drawbacks and problems that will need to be addressed. In both non-deterministic models, differences in glue strength (the strength of the hybridization between complementary strands) among different glue

types could skew the probabilities that a given tile will bind over another. Good design of ?glues? can largely alleviate this problem.

Construction of specified length rulers is important to the construction of more complicated nanostructures. Due to the inefficiency of constructing these rulers in the standard tile assembly model, continuation of this work could result in an advancement in the ability to efficiently build these linear assemblies with very few tile types. This first step toward minimizing tile complexity using the methods described here might later be extended into self-assembly of complex shapes with greater efficiency.

REFERENCES

[1] Zachary Abel, Nadia Benbernou, Mirela Damian, Erik Demaine, Martin Demaine, Robin Flatland, Scott Kominers, and Robert Schweller, *Shape replication through self-assembly and RNase enzymes*, SODA 2010: Proceedings of the Twenty-first Annual ACM-SIAM Symposium on Discrete Algorithms (Austin, Texas), Society for Industrial and Applied Mathematics, 2010.

[2] L.M. Adleman, *Molecular computation of solutions to combinatorial problems*, Science **266** (1994), no. 5187, 1021--1024.

[3] American Physical Society, *There's plenty of room at the bottom*, Pasadena, California, California Institute of Technology, December 1959.

[4] Sarah Cannon, Erik D. Demaine, Martin L. Demaine, Sarah Eisenstat, Matthew J. Patitz, Robert Schweller, Scott M. Summers, and Andrew Winslow, *Two hands are better than one (up to constant factors)*, Arxiv preprint arXiv:1201.1650 (2012).

[5] Harish Chandran, Nikhil Gopalkrishnan, and John H. Reif, *The tile complexity of linear assemblies*, 36th International Colloquium on Automata, Languages and Programming, vol. 5555, 2009.

[6] J.H. Chen and N.C. Seeman, *Synthesis from dna of a molecule with the connectivity of a cube*, Nature **350** (1991), no. 6319, 631--633.

[7] 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** (2008), no. 3, 347--370.

[8] Erik D. Demaine, Sarah Eisenstat, Mashhood Ishaque, and Andrew Winslow, *One-dimensional staged self-assembly*, Proceedings of the 17th international conference on DNA computing and molecular programming, DNA'11, 2011, pp. 100--114.

[9] Robert M. Dirks and Niles A. Pierce, *Triggered amplification by hybridization chain reaction*, PNAS **101** (2004), no. 43, 15275--15278.

[10] David Doty, *Randomized self-assembly for exact shapes*, Proceedings of the Fiftieth IEEE Conference on Foundations of Computer Science, 2009.

[11] ———, *Randomized self-assembly for exact shapes*, SIAM Journal on Computing **39** (2010), no. 8, 3521--3552, Preliminary version appeared in FOCS 2009.

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

[13] David Doty, Matthew J. Patitz, and Scott M. Summers, *Limitations of self-assembly at temperature 1*, Proceedings of The Fifteenth International Meeting on DNA Computing and Molecular Programming (Fayetteville, Arkansas, USA, June 8-11, 2009), 2009, pp. 283--294.

[14] Eric Drexler, *Engines of creation: The coming era of nanotechnology*, Anchor Books, New York City, NY, 1987.

[15] H. Chen el al., *Optimizing tile concentrations to minimize errors and time for dna tile self-assembly systems*, DNA Lecture Notes in Computer Science, vol. 6518, Springer, 2010, pp. 13--24.

[16] Kenichi Fujibayashi, David Yu Zhang, Erik Winfree, and Satoshi Murata, *Error suppression mechanisms for dna tile self-assembly and their simulation*, Natural Computing: an international journal **8** (2009), no. 3, 589--612.

[17] M. Goos and P. Orponen, *Synthesizing minimal tile sets for patterned dna self-assembly*, vol. 6518, 2010, pp. 71--82.

[18] Ming-Yang Kao and Robert T. Schweller, *Reducing tile complexity for self-assembly through temperature programming*, SODA 2006: Proceedings of the 17th Annual ACM-SIAM Symposium on Discrete Algorithms, 2006, pp. 571--580.

[19] _____, *Randomized self-assembly for approximate shapes*, International Colloqium on Automata, Languages, and Programming, Lecture Notes in Computer Science, vol. 5125, Springer, 2008, pp. 370--384.

[20] James I. Lathrop, Jack H. Lutz, Matthew J. Patitz, and Scott M. Summers, *Computability and complexity in self-assembly*, Proceedings of The Fourth Conference on Computability in Europe (Athens, Greece, June 15-20, 2008), 2008.

[21] Pierre Marchal, *John von neumann: The founding father of artificial life*, Artificial Life **4** (1998), no. 3, 229--235.

[22] J.E. Padilla, W. Liu, and N.C. Seeman, *Hierarchical self assembly of patterns from the Robinson tilings: DNA tile design in an enhanced tile assembly model*, Natural Computing **online first, 17 August 2011** (2011).

[23] Jennifer E. Padilla, Matthew J. Patitz, Raul Pena, Robert T. Schweller, Nadrian C. Seeman, Robert Sheline, Scott M. Summers, and Xingsi Zhong, *Asynchronous signal passing for tile self-assembly: Fuel efficient computation and efficient assembly of shapes*, Proceedings of the 2013 International Conference on Unconventional Computation and Natural Computation, to appear (2013).

[24] Volker Patzke and Gunter von Kiedrowski, *Self-replicating sytems*, ARKIVOC **5** (2007), 293--310.

[25] Natasha Paul and Gerald F. Joyce, *A self-replicating ligase ribozyme*, PNAS **99** (2002), no. 120, 12733 -- 12740.

[26] Lionel Penrose, *Mechanics of self-reproduction*, Annals of Human Genetics **23** (1958), no. 1, 59--72.

[27] John H. Reif, Sudheer Sahu, and Peng Yin, *Complexity of graph self-assembly in accretive systems and self-destructible systems*, DNA, 2005, pp. 257--274.

[28] B.H. Robinson and N.C. Seeman, *The design of a biochip: a self-assembling molecular-scale memory device*, Protein Engineering **1** (1987), no. 4, 295--300.

[29] Paul W. K. Rothemund and Erik Winfree, *The program-size complexity of self-assembled squares (extended abstract)*, STOC '00: Proceedings of the thirty-second annual ACM Symposium on Theory of Computing (Portland, Oregon, United States), ACM, 2000, pp. 459--468.

[30] Rebecca Schulman, Bernard Yurke, and Erik Winfree, *Robust self-replication of combinatorial information via crystal growth and scission*, PNAS **109** (2012), no. 17, 6405--6410.

[31] David Soloveichik, Matthew Cook, and Erik Winfree, *Combining self-healing and proofreading in self-assembly*, Natural Computing **7** (2008), no. 2, 203--218.

[32] David Soloveichik and Erik Winfree, *Complexity of self-assembled shapes*, DNA 10: 10th International Meeting on DNA Computing (Claudio Ferretti, Giancarlo Mauri, and Claudio Zandron, eds.), Lecture Notes in Computer Science, vol. 3384, Springer, 2004, pp. 344--354.

[33] Scott M. Summers, *Reducing tile complexity for the self-assembly of scaled shapes through temperature programming*, Algorithmica **63** (2012), no. 1, 117--136.

[34] Eors Szathmary and Irina Gladkih, *A self-replicating hexadeoxynucleotide*, Journal of Theoretical Biology **138** (1989), no. 1, 55--58.

[35] T. Tjivikua, P. Ballester, and J. Rebek Jr., *Self-repllicating system*, J. Am. Chem. Soc. **112** (1990), no. 3, 1249--1250.

[36] Gunter von Kiedrowski, *A self-replicating hexadeoxynucleotide*, Angewandte Chemie International Edition in English **25** (1986), no. 10, 932--935.

[37] Hao Wang, *Proving theorems by pattern recognition -- II*, The Bell System Technical Journal **XL** (1961), no. 1, 1--41.

[38] James Watson and Francis Crick, *Molecular structure of nucleic acids*, Nature **171** (1953), 737--738.

[39] Erik Winfree, *Algorithmic self-assembly of DNA*, Ph.D. thesis, California Institute of Technology, June 1998.

[40] D.Y Zhang and G. Seelig, *Dynamic dna nanotechnology using strand-displacement reactions*, Nature Chemistry **3** (2011), no. 2, 103--113.

[41] Wojciech Zielinski and Leslie Orgel, *Autocatalytic synthesis of a tetranucleotide analogue*, Nature **327** (1987), 346--347.

BIOGRAPHICAL SKETCH

Alexandra Barrett Keenan was born in Des Moines, Iowa on 29 December 1986, the daughter of David D. Keenan and Jan M. Barrett. She received the Bachelor of Science degree in Biochemistry from the University of Iowa in 2010 and the Bachelor of Science in Engineering degree in Biomedical Engineering from the University of Iowa in 2013. During her time at the University of Iowa, she worked primarily in tropical disease laboratories at UI and an UC Berkeley, made several research trips to India, and volunteered in Mexico. She taught Physics and Astronomy at Rio Grande City High School in Rio Grande City, Texas from 2010 to 2012 and began her graduate work in Computer Science at the University of Texas-Pan American in 2012. She interned at the NASA Glenn Research Center during the summer of 2013 and began working on the NASA bioastronautics contract at Wyle Integrated Science and Engineering as a Biomedical Engineer in Houston, Texas in 2014.

Permanent address: 1109 Peden Street B
Houston, Texas 77006