Staged Self-assembly and Polyomino Context-Free Grammars

Andrew Winslow

Department of Computer Science, Tufts University
awinslow@cs.tufts.edu

Abstract. Previous work by Demaine et al. (2012) developed a strong connection between smallest context-free grammars and staged self-assembly systems for one-dimensional strings and assemblies. We extend this work to two-dimensional polyominoes and assemblies, comparing staged self-assembly systems to a natural generalization of context-free grammars we call polyomino context-free grammars (PCFGs).

We achieve nearly optimal bounds on the largest ratios of the smallest PCFG and staged self-assembly system for a given polyomino with \( n \) cells. For the ratio of PCFGs over assembly systems, we show that the smallest PCFG can be an \( \Omega(n/\log^3 n) \)-factor larger than the smallest staged assembly system, even when restricted to square polyominoes. For the ratio of assembly systems over PCFGs, we show that the smallest staged assembly system is never more than a \( O(\log n) \)-factor larger than the smallest PCFG and is sometimes an \( \Omega(\log n/\log \log n) \)-factor larger.

1 Introduction

In the mid-1990s, the Ph.D. thesis of Erik Winfree [14] introduced a theoretical model of self-assembling nanoparticles. In this model, which he called the abstract tile assembly model (aTAM), square particles called tiles attach edgewise to each other if their edges share a common glue and the bond strength is sufficient to overcome the kinetic energy or temperature of the system. The products of these systems are assemblies: aggregates of tiles forming via crystal-like growth starting at a seed tile. Surprisingly, these tile systems have been shown to be computationally universal [14,5], self-simulating [8,9], and capable of optimally encoding arbitrary shapes [12,1,13].

In parallel with work on the aTAM, a number of variations on the model have been proposed and investigated. One well-studied variant called the hierarchical [4] or two-handed assembly model (2HAM) [6] eliminates the seed tile and allows tiles and assemblies to attach in arbitrary order. This model was shown to be capable of (theoretically) faster assembly of squares [4] and simulation of aTAM systems [2], including capturing the seed-originated growth dynamics. A generalization of the 2HAM model proposed by Demaine et al. [6] is the staged
assembly model, which allows the assemblies produced by one system to be used as reagents (in place of tiles) for another system, yielding systems divided into sequential assembly stages. They showed that such sequential assembly systems can replace the role of glues in encoding complex assemblies, allowing the construction of arbitrary shapes efficiently while only using a constant number of glue types, a result impossible in the aTAM or 2HAM.

To understand the power of the staged assembly model, Demaine et al. [7] studied the problem of finding the smallest system producing a one-dimensional assembly with a given sequence of labels on its tiles, called a label string. They proved that for systems with a constant number of glue types, this problem is equivalent to the well-studied problem of finding the smallest context-free grammar whose language is the given label string, also called the smallest grammar problem (see [113]). For systems with unlimited glue types, they proved that the ratio of the smallest context-free grammar over the smallest system producing an assembly with a given label string of length $n$ (which they call separation) is $\Omega(\sqrt{n}/\log n)$ and $O((n/\log n)^{2/3})$ in the worst case.

In this paper we consider the two-dimensional version of this problem: finding the smallest staged assembly system producing an assembly with a given label polyomino. For systems with constant glue types and no cooperative bonding, we achieve separation of grammars over these systems of $\Omega(n/(\log \log n)^2)$ for polyominoes with $n$ cells (Sect. 6.1), and $\Omega(n/\log^3 n)$ when restricted to rectangular (Sect. 6.2) or square (Sect. 6.3) polyominoes with a constant number of labels. Adding the restriction that each step of the assembly process produces a single product, we achieve $\Omega(n/\log^3 n)$ separation for general polyominoes with a single label (Sect. 6.1). For the separation of staged assembly systems over grammars, we achieve bounds of $\Omega(\log n/\log \log n)$ (Sect. 4) and, constructively, $O(\log n)$ (Sect. 5). For all of these results, we use a simple definition of context-free grammars on polyominoes that generalizes the deterministic context-free grammars (called $RCFG$s) of [7].

When taken together, these results give a nearly complete picture of how smallest context-free grammars and staged assembly systems compare. For some polyominoes, staged assembly systems are exponentially smaller than context-free grammars ($O(\log n)$ vs. $\Omega(n/\log^3 n)$). On the other hand, given a polyomino and grammar deriving it, one can construct a staged assembly system that is a (nearly optimal) $O(\log n)$-factor larger and produces an assembly with a label polyomino replicating the polyomino.

2 Staged Self-assembly

An instance of the staged tile assembly model is called a staged assembly system or system, abbreviated SAS. A SAS $S = (T, G, \tau, M, B)$ is specified by five parts: a tile set $T$ of square tiles, a glue function $G : \Sigma(G)^2 \rightarrow \{0, 1, \ldots, \tau\}$, a temperature $\tau \in \mathbb{N}$, a directed acyclic mix graph $M = (V, E)$, and a start bin function $B : V_L \rightarrow T$ from the leaf vertices $V_L \subseteq V$ of $M$ with no incoming edges.