Synthesizing Small and Reliable Tile Sets for Patterned DNA Self-assembly

Tuomo Lempiäinen, Eugen Czeizler, and Pekka Orponen

Department of Information and Computer Science
Helsinki Institute for Information Technology HIIT
Aalto University School of Science
P.O. Box 15400, FI-00076 Aalto, Finland
firstname.lastname@aalto.fi

Abstract. We consider the problem of finding, for a given 2D pattern of colored tiles, a minimal set of tile types self-assembling to this pattern in the abstract Tile Assembly Model of Winfree (1998). This Patterned self-Assembly Tile set Synthesis (PATS) problem was first introduced by Ma and Lombardi (2008), and subsequently studied by Göös and Orponen (2011), who presented an exhaustive partition-search branch-and-bound algorithm (briefly PS-BB) for it. However, finding the true minimal tile sets is very time consuming, and PS-BB is not well-suited for finding small but not necessarily minimal solutions. In this paper, we modify the basic partition-search framework by using a heuristic to optimize the order in which the algorithm traverses its search space. We find that by running several parallel instances of the modified algorithm PS-H, the search time for small tile sets can be shortened considerably. We also introduce a method for computing the reliability of a tile set, i.e. the probability of its error-free self-assembly to the target tiling, based on Winfree’s analysis of the kinetic Tile Assembly Model (1998). We present data on the reliability of tile sets found by the algorithms and find that also here PS-H constitutes a significant improvement over PS-BB.

1 Introduction

Self-assembly of nanostructures templated on synthetic DNA has been proposed by several authors as a potentially ground-breaking technology for the manufacture of next-generation circuits, devices, and materials [4,9,14,16]. Also laboratory techniques for synthesizing the requisite 2D DNA template lattices, many based on Rothemund’s [12] DNA origami tiles, have recently been demonstrated by many groups [6,10].

In order to support the manufacture of aperiodic structures, such as electronic circuit designs, these DNA templates need to be addressable. When the template is constructed as a tiling from a family of DNA origami (or other kinds of) tiles, one can view the base tiles as being “colored” according to their different functionalities, and the completed template implementing a desired color pattern. Now a given target pattern can be assembled from many different families of
base tiles, and it is clearly advantageous to try to minimize the number of tile types needed and/or maximize the probability that they self-assemble to the desired pattern, given some model of tiling errors.

The task of minimizing the number of DNA tile types required to implement a given 2D pattern was identified by Ma and Lombardi [8], who formulated it as a combinatorial optimization problem, the *Patterned self-Assembly Tile set Synthesis* (PATS) problem. Ma and Lombardi proposed two greedy heuristics for solving the task, and subsequently Göös and Orponen [3] presented an exhaustive partition-search branch-and-bound algorithm for it. While the search algorithm presented in [3], which we denote here as PS-BB, is somewhat successful in finding minimal tile sets for small patterns, the size of the search space grows so rapidly that it seems to hit a complexity barrier at approximately pattern sizes of $7 \times 7$ tiles. In practice one would of course not need to find an absolutely minimal tile set for a given pattern, but any reasonably small solution set would suffice. However, when the algorithm PS-BB fails to find a minimal solution, it does not seem to yield very good approximate solutions either.

In the present work, we approach the task of finding small but not necessarily minimal tile sets for a given 2D pattern by tailoring the basic partition-search framework of [3] towards this goal. Instead of a systematic branch-and-bound pruning and traversal of the complete search space, we apply a heuristic that attempts to optimize the order of the directions in which the space is explored. The new algorithm, denoted PS-H, is described in more detail below in Sect. 3.

It is well known in the heuristic optimization community [2,7] that when the runtime distribution of a randomized search algorithm has a large variance, it is with high probability more efficient to run several independent short runs ("restarts") of the algorithm than a single long run. Correspondingly, we investigate the efficiency of the PS-H search method for a number of parallel executions ranging from 1 to 32, and note that indeed this number has a significant effect on the success rate of the algorithm in finding small tile sets. Also these results are discussed below in Sect. 3.

Given the inherently stochastic nature of the DNA self-assembly process, it is also of interest to assess the reliability of a given tile set, i.e. the probability of its error-free self-assembly to the desired target tiling. In Sect. 4 we introduce a method for estimating this quantity, based on Winfree’s analysis of the kinetic Tile Assembly Model [15]. We present empirical data on the reliability of tile sets found by the PS-BB and PS-H algorithms and find that also here the PS-H algorithm constitutes a significant improvement over the PS-BB method.

2 The PATS Problem and the PS-BB Algorithm

2.1 The Abstract Tile Assembly Model [11,15]

Let $\mathcal{D} = \{N, E, S, W\}$ be the set of four functions $\mathbb{Z}^2 \rightarrow \mathbb{Z}^2$ corresponding to the four cardinal directions. Let $\Sigma$ be a finite set of *glue types* and $s : \Sigma \times \Sigma \rightarrow \mathbb{N}$ a *glue strength* function such that $s(\sigma, \sigma') > 0$ only if $\sigma = \sigma'$. A *tile type* $t \in \Sigma^4$ is a quadruple $(\sigma_N(t), \sigma_E(t), \sigma_S(t), \sigma_W(t))$ and a (tile) assembly $\mathcal{A}$ is a partial