Yield Optimization Strategies for (DNA) Staged Tile Assembly Systems

Eugen Czeizler and Pekka Orponen

Department of Information and Computer Science
Aalto University P.O. Box 15400, FI-00076 Aalto, Finland
{firstname.lastname}@aalto.fi

Abstract. The staged Tile Assembly Model has been introduced by Demaine et al. 2008 as an enhancement of the previous tile self-assembly model of Winfree. In this framework, the assembly is allowed to be performed in parallel in different test-tubes, and the obtained products are stored and mixed in subsequent assembly stages. Using elegant combinatorial constructions, in has been shown that staged assembly systems possess remarkable advantage in comparison to their abstract counterparts. Because of their parallel nature, one can choose from a multitude of staged assembly strategies for assembling a given target structure. In the current work we analyze these assembly variations from a kinetic perspective, in order to determine and possibly maximize, their final assembly yield. As a pre-requirement for this task, we provide a procedure for associating an analytically tractable mathematical model to a given staged assembly experiment, based on which we can predict the yield concentration of the final assembly product. As a case study, we consider various assembly strategies as well as optimized and non-optimized assembly protocols for generating a size-10 tile assembly.

Keywords: Tile Assembly Model, staged assembly, numerical modelling, yield optimization.

1 Introduction

The abstract Tile Assembly Model (aTAM) has been introduced by Winfree [9] as a custom-made generalization of Wang tile systems, designed for the study of DNA tile self-assembly. The basic components of the aTAM are non-rotatable unit square tiles, uniquely defined by the sets of four glues placed on top of their edges. The glues are part of a finite alphabet and each pair of glues is associated a strength value, determining the stability of the link between two tiles having these glues on the abutting edges. The assembly process starts from a single nucleation point, the seed, and it continues by sequential attachments of tiles until no more tiles can be added to the assembly. All the individual tiles are placed inside a unique assembly "pot", and the assembly process progresses with no external interactions.

A.-H. Dediu et al. (Eds.): TPNC 2013, LNCS 8273, pp. 31–44, 2013.

In order to improve the efficiency of these systems, with respect to assembling more complex structures from a fewer initial number of distinct *tile-types*[1], Demain et al. introduced the staged Tile Assembly Model (sTAM) [2]. In this framework, the assembly is performed in stages and in different *test-tubes* (or *bins*). Each test-tube is initialized with one or several non-interacting tile-types, and in each stage, one or several test-tubes are mixed together according to a predefined scheme. Different tile-types are thus mixed into the same compartment and start interacting. No seed structures are defined in this framework, and thus the reactions are implemented population-wise. The external observer allows the reaction to progress for some time, after which the content of the test-tubes is filtered and only the generated reaction products are used in subsequent stages.

Using elegant combinatorial designs, Demain et al. [2,3] demonstrated how various structures can be assembled efficiently, both in terms of the total number of different tile-types used, and in terms of the tile-interaction complexity, i.e., using only temperature-1 systems[2]. For example, one requires only 3 tile-types and $log(n)$ stages for constructing an n-size ribbon of contiguous tiles, while a similar structure assembled in a "one pot" system, i.e. classical aTAM, requires n distinct tile-types. Similarly, using a constant number of tile-types and only $log(n)$ stages, one can assemble a full $n \times n$ square, whereas in the aTAM framework $O(\log n / \log \log n)$ tile-types are required to assemble an analogous structure.

Because of the parallel design feature, one can choose from a multitude of staged assembly strategies for assembling a given target structure. Moreover, this freedom of choosing between several assembly variants remains valid even when one restricts to those strategies employing a minimum number of assembly stages. In the current work we analyze these assembly variations, as well as possible different implementations of the same assembly strategy, all in terms of their predicted final yield. Our objective is to study possible yield optimization protocols for the target assembly of these system. Considering assembly systems with an abundance of inter-molecular interaction (as is the case of DNA self-assembly systems), putting together larger concentrations of reactants and allowing them more time to react will always generate better yields. Thus, in order to perform a fair comparison between various assembly strategies, we require the total initial reactant concentrations, volume, as well as total time allowed for the reactions, to be constant in all of the compared strategies.

The particular aspects we want to investigate are:

- the yield variations in between different staged assembly schemes (generating the same final structures); and
- the yield variations within the same assembly scheme, when modifying parameters, such as: i) the time allocation for each of the assembly stages (while

[1] A tile-type is a population of identical copies of the same tile.

[2] Temperature-1 systems are highly advantageous as they can be made very resistent to errors, compared to temperature 2 systems [1]. Such systems can be implemented using e.g. DNA-origami techniques [8].

the total time for the experiment remains constant); or ii) the ratio in which certain assemblies are mixed inside the test-tubes (with total volumes of the reactants kept constant).

The first criterion can be seen as a design optimization level, while the second as a protocol optimization level. Moreover, we ask whether there exists a correlation between the two levels. Namely, would a design scheme performing particularly well on some assembly protocol generally generate better yields (than other assembly schemes) independent of the employed protocols?

In order to be able to address such questions, we first provide a methodology of assigning to every staged tile assembly system (sTAS) a numerical model describing the time-evolution of all its components. The employed modelling methodology is based on the principle of mass-action kinetics [5,6] and is implemented using the formulation given by ordinary differential equations (ODE). The modeling methodology is different from the one considered in the kTAM models, [12], as in this case we do not follow the assembly of only one particular structure (starting from a seed tile), but we keep track of all the species available in solution(s). While such an approach is usually untractable for "one pot" systems, we show that it becomes applicable in the case of sTAS. We use the above modelling methodology and, as a case study, we consider the assembly (and yield optimization process) of a size-10 tile assembly structure. For numerical modelling and optimization we have used the open source software COPASI [7].

The paper is organized as follows. The next section contain background information regarding the aTAM and sTAM models. In Section 3 we introduce our kinetic modelling methodology for sTAS and provide a series of pre-normalization requirements for our models. In Section 4 we introduce several yield optimization strategies applicable to staged assembly systems, and as a case study in the next section we consider the staged assembly and yield optimization protocols employed in obtaining size-10 horizontal ribbons of tiles. In the last section we discuss our results and provide some future research directions.

2 Background

In the following, we provide a very brief introduction of the (abstract) Tile Assembly Model, aTAM, and its staged counterpart, sTAM. For a more detailed presentation of these models we refer to [9,2] as well as the recent survey [4].

Let Σ be a finite set of *glues*, and let $s : \Sigma \times \Sigma \to \mathbb{N}$ be a *glue strength* function, i.e., $s(\sigma_1, \sigma_2) = s(\sigma_2, \sigma_1)$ for all $\sigma_1, \sigma_2 \in \Sigma$. A *tile* (or *tile-type*) t is a unit square structure with glues on its four edges; we assume that the tiles can not be either rotated or reflected. Thus, we can represent a tile as the ordered 4-tuple of glues $t = (t_N, t_E, t_S, t_W) \in \Sigma^4$ where the N, S, E, and W subscripts point to the corresponding edge positioning. An *assembly* $\mathcal{A}$ is a partial mapping $\mathcal{A} : \mathbb{Z}^2 \to \Sigma^4$ assigning tiles to locations in the two-dimensional grid, such that the defined structure is connected. A *tile assembly system* (TAS) $\mathcal{T} = (T, \mathcal{S}, s, \tau)$ consists of a finite set T of tile-types, an assembly $\mathcal{S}$ called the *seed assembly*,

a glue strength function s and a *temperature* $\tau \in \mathbb{Z}^+$. By definition, we assume that the seed assembly S is stable and cannot be disassembled[3].

Given a TAS $\mathcal{T} = (T, \mathcal{S}, s, \tau)$ and an assembly $\mathcal{A}$ (such as the seed $\mathcal{S}$), a new tile can be added to $\mathcal{A}$ if it shares a common boundary with tiles that bind it into place with total strength at least τ; we call such a process a *successful tile addition*. We say that a TAS $\mathcal{T}$ *produces* an assembly $\mathcal{A}$ if this assembly is formed by a sequence of successful tile additions starting from the seed assembly $\mathcal{S}$. Moreover, if no other tiles can be further attached to $\mathcal{A}$, we say that the assembly is *terminal*.

The model of staged assembly differs from the classical aTAM by allowing partial assemblies to be formed in parallel in different test-tubes before merging them together. The notion of successful addition is extended from the previous case by allowing the merging of any two assemblies, as long as the sum of the strength of glues placed along the common boundary of the two assemblies is at least the temperature τ of the system. Thus, in this setting, single tiles are seen just as (elementary) assemblies. The above requirement for an assembly is known as *partial connectivity*[4], see [2], as it does not enforce tiles in the assembly to have matching edges with all the neighboring tiles, as long as the matching which bound them into place exceed or are equal with the temperature τ. For the remaining of this paper we assume working in this partial connectivity requirement for assemblies.

Another difference from aTAM comes from the fact that the assembly process is allowed to be performed in parallel in different *test-tubes* (or *bins*) and across several assembly stages. Each tile-type is placed initially in an isolated test-tube; we call these *initial test-tubes*. When the content of two (or several) test-tubes is mixed in a separate bin, the assemblies start interacting and bind to each-other according to their glue interactions. The process is allowed to progress for some time, after which the mixed solution is filtered and only the reaction products, i.e., the terminal assemblies, are stored for further mixing, while the remaining reactants are discarded. The test-tubes are further mixed synchronously during several assembly stages, until the final product is assembled in the unique test-tube of the last assembly stage.

A staged tile assembly system (sTAS) $\mathcal{T}^s = (T, s, \tau, G)$ is defined by the set T of starting tile-types t_i, each placed in marked *initial test-tubes* $T^0_{t_i}$, a glue strength function s, a temperature parameter τ,and an *assembly graph* G (or *mix graph*). The assembly graph is a direct acyclic graph (DAG) describing the different test-tubes and the way these tubes are mixed along a synchronous succession of assembly stages. The nodes of the graph are the various test-tubes (including the initial ones), while a directed edge between two nodes T^{s_i} and T^{s_j} symbolizes that the assembly product of test-tube T^{s_i} (or the corresponding tile-type in case of an initial test-tube) is transferred (either completely if T^{s_i} has no other out-edges or just a fraction of it otherwise) to test-tube T^{s_j}. The

[3] On some experimental implementations of the TAM, the seed assembly is implemented using e.g., DNA origami [10,11]

[4] As opposed to the *full connectivity* requirement.

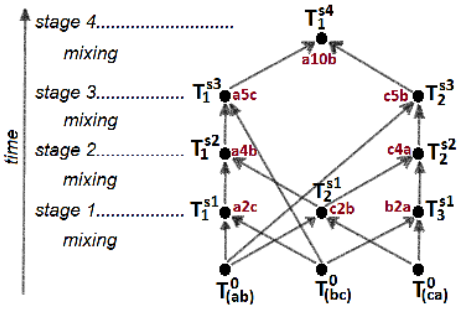

Fig. 1. The annotated assembly graph of an sTAS assembling a size-10 ribbon

final assembly of the sTAS is the assembly product collected at the end of the experiment from the unique test-tube of the last assembly stage.

As an example, in the following we provide an sTAS assembling a size-10 horizontal ribbon of tiles. Since this is a 1D structure, only the glues of the East and West sides of a tile-type are relevant for the assembly process. Thus, a tile is denoted as the pair (x, y) of its West and East glues, respectively. Moreover, a 1D assembly containing $k > 1$ tiles will be denoted as xky, where x (resp. y) is the West (East) glue of the left-most (right-most) tile in the assembly. The temperature τ of the system is 1, and the strength function s is given by $s(x, y)$ is 1 if $x = y$ and 0 otherwise. The sTAM contains 3 initial test-tubes $T^0_{(ab)}, T^0_{(bc)}, T^0_{(ca)}$ for the tile-types $(a, b), (b, c)$, and (c, a), respectively, and employs 4 assembling stages. The assembly-graph from Figure 1 fully describes the design of the sTAS; for ease of understanding we have annotated the graph by providing also the description of the assembly product in each of the test-tubes.

Various 2D assembly structures can be efficiently[5] assembled by appropriate staged assembly systems, even at temperature $\tau = 1$ and using only two reactants per test-tube[6], see e.g. [2]. Although the results of our current research apply to both 1D and 2D assembled structures, in order to simplify the considered mix graph designs and exemplify the applicability of our approach, are going to concentrate over the assembly of 1D ribbons of tiles. Indeed, if more complex 2D assemblies are investigated, the only change comes in the design of the mixing graph. However, the dynamics of the system is preserved, as mixing a size-p assembly (i.e., containing p tiles) with a size-q assembly, always generates a size-$(p + q)$ assembly, assuming the two components are indeed reacting.

Thus, from now on, we represent the tiles as the pairs of glues placed on their West and East edges, respectively, we assume working always at temperature 1, and we use the strength function given by $s(x, y) = 1$ if $x = y$ and 0 otherwise.

[5] Here, we measure the efficiency in terms of the number of different tile-types used

[6] In most of the staged assembly designs from the literature, only two reactants are placed inside a test-tube.

3 Modeling of Staged Tile Assembly Systems

In order to be able to address questions regarding yield optimization of sTAS we need appropriate quantitative tools for estimating and analyzing the corresponding yields. In this section we introduce an adequate mathematical model of the staged assembly process. Using this methodology, for any particular target structure, one can numerically determine the best assembly strategy for it, as well as numerically optimize the parameters of the chosen assembly strategy.

The modelling paradigm that we choose to use is that of ODE, while the formulation of the models is based on the principle of mass-action kinetics. The principle of mass-action, introduced in [5,6], says that the rate of each reaction is proportional to the concentration of reactants. Moreover, this reaction rate gives the measure on which the reactants are consumed and the products are generated. To exemplify, consider the simple reaction $A + B \to A : B$ when an assembly A joins an assembly B and forms an assembly $A : B$. If we denote by $[A](t), [B](t)$, and $[A : B](t)$ the concentrations these assemblies at time t, and by k the kinetic rate constant of the reaction, then the combined measure of consuming and producing each of the reactants is given by the system:

$$\frac{d[A]}{dt} = -k[A] \cdot [B] \qquad \frac{d[B]}{dt} = -k[A] \cdot [B] \qquad \frac{d[A:B]}{dt} = k[A] \cdot [B]$$

We are going to assume (without loss of generality, but with some possible loss of design efficiency) that in each stage of the assembly, we allow to mix the contents of only two test-tubes at a time; most of the sTAS in the literature are nevertheless designed in this way. Moreover, before the mixing procedure, the content of each test-tube is filtered and only the product of the assembly is preserved. Thus, in each test-tube we have only two reactants. As a consequence of this, the chemical reaction system corresponding to each of the test-tubes (each test-tube generates an isolated system) obeys two conservation reactions. Namely, at any time point t we have that

$$[A](t) + [A:B](t) = C_1 \text{ and } [B](t) + [A:B](t) = C_2,$$

for two constants C_1 and C_2, such that $C_1 = [A](0) + [A : B](0)$ and $C_2 = [B](0) + [A : B](0)$[7]. Thus, at any time point t, the concentration of the $[A]$ and $[B]$ species can be derived from the concentration of the $[A:B]$ species. By substituting these into the third differential equation we obtain:

$$\frac{d[A:B]}{dt} = k(C_1 - [A:B])(C_2 - [A:B]) \tag{1}$$

In most cases, such ODE systems derived from corresponding chemical reaction systems are analytically intractable. However, since in the case of sTAS we have that in each test-tube there exist only two reactants interacting and forming a

[7] We denoted by $[A](0), [B](0)$, and $[A:B](0)$ the initial concentration of the species A, B, and $A : B$, respectively, at time $t_0 = 0$.

product (a larger complex), the derived ODE systems can be solved analytically. Namely, equation (1) has the solution:

$$[A{:}B](t) = \frac{-C_1C_2 + [A{:}B](0)C_1 + C_1C_2e^{tk(C_1-C_2)} - [A{:}B](0)e^{tk(C_1-C_2)}}{C_1e^{tk(C_1-C_2)} - [A{:}B](0)e^{tk(C_1-C_2)} - C_2 + [A{:}B](0)} \quad (2)$$

Another particularity of sTAS is that none of the $[A{:}B]$ structures exist before mixing assemblies A and B, that is $[A{:}B](0) = 0$. Thus, equation (2) becomes:

$$[A{:}B]_t = \frac{-C_1C_2 + C_1C_2e^{tk(C_1-C_2)}}{C_1e^{tk(C_1-C_2)} - C_2} = \frac{C_1C_2(e^{tk(C_1-C_2)} - 1)}{C_1e^{tk(C_1-C_2)} - C_2}, \quad (3)$$

where $C_1 = [A](0)$ and $C_2 = [B](0)$. Moreover, if one also assumes that $[A](0) = [B](0) = C$, i.e., the systems is symmetric, then equation (2) becomes

$$[A{:}B]_t = \frac{ktC^2}{1 + Ckt}. \quad (4)$$

Because at each stage of the assembly the initial concentrations for the reactants depend on the concentrations of the products at prior stages, and since equation (1) describing the time-evolution of the product assembly in each test-tube admits an analytic solution, we can provide an analytic formula for the entire system.

An important observation regarding the dynamics of sTAS is that the products obtained in prior stages of the assembly are not further concentrated before mixing them in subsequent stages. Thus, in each stage, the volume of the solution increases, and hence we have to update the concentration of the reactants accordingly (i.e., to decrease these concentrations).

For example, assume the reactants R_1 and R_2 of test-tube T from some stage of the assembly are taken to be fractions of the products P_1 and P_2 of test-tubes $T1$ and $T2$, respectively (from some previous stages). Namely, let

$$Vol_{trans}^{P_1} = r_{T1}^{T} \cdot Vol_{T1} \text{ and } Vol_{trans}^{P_2} = r_{T2}^{T} \cdot Vol_{T2}$$

be the volumes of the fraction of products P_1 and P_2 transferred from $T1$ and $T2$ respectively, to T, where Vol_{T1} (resp. Vol_{T2}) and r_{T1}^{T} (resp. r_{T2}^{T}) denote the volume of test-tube $T1$ (resp. $T2$) and the ratio from this volume which is transferred into T. Then, the initial concentration of reactants R_1 and R_2 in test-tube T is given by

$$[R_1](0) = \frac{[P_1] \cdot Vol_{trans}^{P_1}}{Vol_{trans}^{P_1} + Vol_{trans}^{P_2}}; \text{ and } [R_2](0) = \frac{[P_2] \cdot Vol_{trans}^{P_2}}{Vol_{trans}^{P_1} + Vol_{trans}^{P_2}}, \quad (5)$$

where $[P_1]$ (resp. $[P_2]$) is the concentration of the product P_1 (resp. P_2) at the end of the corresponding stage, and $Vol_{trans}^{P_1} + Vol_{trans}^{P_2} = Vol_T$ is the volume of the test-tube T.

Thus, by keeping track of the volumes of each test-tube and knowing the ratio in which a particular product is split, we can determine the analytic formula of each intermediary (or final) species in the system.

Once such a computational model is derived, it can be estimated numerically for various sets of parameters, e.g., equal time-splits for all stages and/or equal (or proportional) volume-splits of various products. Moreover, the above parameters can be optimized in order to maximize the yield (i.e., concentration) of the final product. Also, using such models, we can compare two or several strategies in determining which provides a better yield, if experiments are performed in similar conditions, i.e., same total time and initial tile concentration.

In order to compare two (or several) assembly strategies we can further simplify the models by making a synchronous pre-normalization of the data. Thus, we are going to assume from now on that the kinetic rate constant of all assembly reactions is equal to 1, and that the concentration of all tile-types in their initial test-tubes, $[mon]$, is also normalized to $[mon] = 100$. Because of the above pre-normalization of the data, the time parameter presents a highly altered behaviour; thus, from now on, we use the notion of *time unit* (*t.u.*) for referring to time variables. Consider for example a system of only two reactants (tile-types) $a1b$ and $b1c$, each having concentration 100 in their initial test-tubes. Assuming these reactants are mixed in equal quantities, their initial concentration in the (mixing) test-tube becomes 50. In these conditions, we observe that the assembly reaction is completed in proportions of approx. 50%, 75%, and 90% only after 0.02, 0.06, and 0.2 t.u., respectively. Thus, our in-silico experiments and numerical analysis will be performed for a total time interval of 0.14–0.25 t.u. per stage, that is 0.42 t.u. for the 3-staged assembly of size-5 ribbons (in Section 4), and 0.9 t.u. for the 4-staged assembly of size-10 ribbons (in Section 5).

4 Yield Optimization Strategies for sTAM

Optimizations at the Assembly Strategy Level. The sTAM framework allows for several assembly strategies to be employed in achieving the same final structure. Moreover, in some cases, each of these strategies, although different in themselves, are all optimal in terms of number of distinct test-tubes or stages they employ. Consider for example the staged assembly process needed for assembling a size-5 ribbon. According to the assembly designs introduced in [2] for constructing size-k 1D ribbons of tiles in optimally possible number of stages, there are four different (staged) assembly strategies for the construction of size-5 ribbons, each employing 3 stages. The assembly graphs of two of these strategies are provided in Figure 2, while the remaining two strategies are symmetric. However, does each of these strategies produce the same amount of final yield (assuming that the initial quantity of resources is proportionally equal in each of the situations)?

In order to compare the previous two strategies, besides the common total time for the experiment ($T_{total} = t_1 + t_2 + t_3 = 0.42$ t.u.) and the similar concentration ($[mon] = 100$) of all tile-types in their initial test-tubes, in both scenarios we are

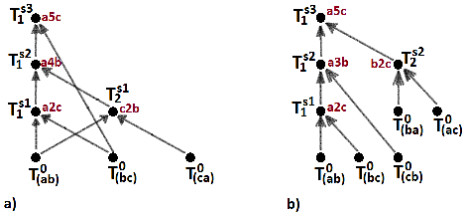

Fig. 2. Two distinct assembly strategies for the same size-5 ribbon, each using only 3 assembling stages

going to use a similar procedure of setting the time- and volume-split parameters, as follows:

- All assembly stages are performed in equal time intervals: $t_1 = t_2 = t_3 = 0.14$;
- Whenever a tile-type (monomer) is a reactant in a test-tube, the introduced quantity of this reactant in the test-tube is exactly one unit volume.

By numerically estimating the associated mathematical models we obtain the concentration of the final product, $[a5c]$, in the assembly scenarios from Figure 2 a) and b) as 54.4% and 56.2% respectively, where 100% would represent the all-maximal value possible for this assembly, e.g. obtained if time would allow the reactions to be fully completed[8].

From the above example, it can be confirmed our initial assumption that different assembly strategies may generate different final yields, despite using the same amount of time and substance resources.

Optimizations at the Experimental Setting Level. Consider now we have chosen a particular assembly strategy, say e.g., assembling the previous 5-tile structure by the scenario in Figure 2 a). A subsequent question concerns the way of allocating the total pool of resources, i.e., substance volume in each test-tube, time allocation for each of the assembly stages, etc., such as to maximize the outcome of the experiment. For example, in the case of the previous example, what would be the best split of the total time of the experiment into three time-periods for the corresponding assembly stages, such as to obtain a maximum amount of 5-tile structures at the end of the final stage? Also, what would be the best way of splitting the amount of tile (bc) in between T_1^{s1} and T_1^{s3}, or similarly the splitting of the amount of tile (ab)? Also, for the cases when an intermediary assembly is used in several reactions from some later stages, what is the optimal way of splitting this product, i.e., its total volume, in between these test-tubes? The above time- and volume-splitting ratios can be subjected to targeted optimization protocols.

[8] In the case of 5-tile ribbons, 100% corresponds to a concentration of 20.

Until now, as suggested by the definition of the sTAM, we have assumed that the entire volume of an intermediate product is transferred to the subsequent stages.However, as suggested by actual lab procedures, we examine also the setting in which only a fraction of the reaction products are re-introduced as reactants. Namely, we force that in every test-tube, the volume of the two reactants sum up to exactly one. In this setting, it becomes even more clear that the ratio in which the two reactants are mixed (each coming into the reaction with possible different initial concentrations) becomes very important in determining a maximum concentration of the product.

5 Case Study: Assembling Size-10 Ribbons

As an yield optimization case study, we are going to consider the process of assembling (in a staged assembly fashion) a size-10 1D horizontal ribbon of tiles. As previously explained, restricting to this 1D structure is not a considerable limitation, since even in the case of assembling 2D complexes, once the mixing graph is designed, the modeling and optimization procedures remain the same.

Since the available pool of possible assembly strategies for a size-10 horizontal line is considerably large, even for the case where we impose using only four stages, we are going to compare only four particular such strategies[9]. We present these strategies (a.k.a. the corresponding mixing graphs) in Figure 3.

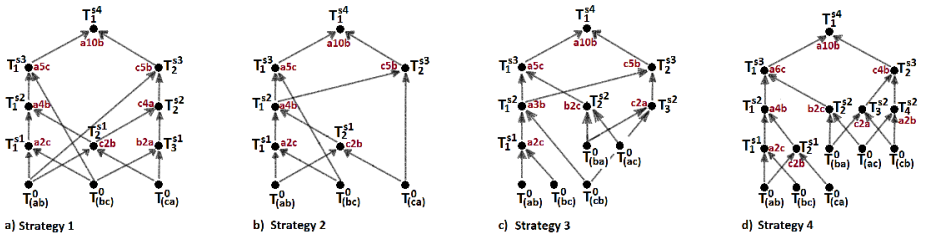

Fig. 3. Four distinct assembly strategies for a size-10 ribbon, each using only 4 assembling stages

In order to illustrate the possible differences between different assembly scenarios as well as between optimized vs. non-optimized assembly protocols, we do not restrict to computing only the optimum values, but provide for comparison a larger pool of parameter setups. Thus, we are comparing all these four assembly strategies, by subjecting each of them to five different setups regarding their time-split and volume-split parameters. Three of these setups are based on combinatorial heuristics, while in two of them we numerically optimize the parameters for maximizing the final yield concentration. In order to compare all of these strategies and setups, we impose some general constraints as follows:

[9] These strategies have been chosen almost at random from the available ones, without a prior knowledge on their behaviour during the optimization process.

i) The total time of the experiment is $T_{total} = 0.9$ t.u.
ii) The concentration of all tile-types in their initial test-tubes is $[mon] = 100$.
iii) (only for three of the setups) The cumulative volume of all tile-types introduced in the various initial test-tubes is 10 unit volumes, where the volume of each tile-type is proportional to the number of times it appears in the final size-10 assembly.

The five setups can be partitioned into two groups:

Group 1: Combinatorially designed setups:

Setup 1: Equal time-splits and proportional volume-splits.

- Equal time intervals for the stages, that is, $t_1 = t_2 = t_3 = t_4 = 0.225$;
- The volume of any species who needs to be partitioned into two or several test-tubes will be done so proportionally to how much the product of these latter test-tubes will contribute to the final assembly.

A combinatorial observation regarding staged assembly systems is that on average the concentration of the reactants is reduced al least by half in each stage. By inspecting equation (3), we observe that if the concentration of both reactants is reduced by half, we obtain the same product-reactant ratio only if we double the time allocated to this stage. Thus, as a possible procedure for improving the overall yield, the time-split parameters from the next setup are in geometric progression.

Setup 2: Time-splits in geometric progression and proportional volume-splits.

- The time intervals are $(t_1; t_2; t_3; t_4) = (0.06; 0.12; 0.24; 0.48)$;
- The volume of products is partitioned proportionally into subsequent test-tubes (as in the case of Setup 1).

Setup 3: Equal time-splits and equal half unit-volumes for all reactants

- Equal time intervals: $t_1 = t_2 = t_3 = t_4 = 0.225$;
- The volume of each of the reactants in a test-tube is set to half unit volume.

This represents a rather "lazy" (or automated) instance of the setting in which the volume of the test-tubes is limited to one unit.

Group 2: Numerically optimized setups:

Setup 4: Optimized time- and volume-splits while using the entire volume of substance. As in the case of Setup 1 and 2, we assume here that we completely use the entire volumes of all intermediary assemblies and of all single tiles.

Table 1. The concentration $[a10b]$ of the size-10 ribbon structure generated in the final stage of the assembly; the results are expressed in their percentage form, where 100% represents the absolute maximal value possible for this assembly, namely 10

Assembly strategy	Group 1			Group 2	
	Setup 1	Setup 2	Setup 3	Setup 4	Setup 5
Strategy 1	41.4%	34.7%	25.0%	44.5%	53.5%
Strategy 2	41.4%	34.7%	25.0%	44.5%	53.5%
Strategy 3	42.6%	42.0%	30.5%	46.9%	49.0%
Strategy 4	42.4%	39.8%	37.3%	49.3%	48.6%

Setup 5: Optimized time- and volume-splits while enforcing unit volumes for all test-tubes As in the case of Setup 3, we enforce that each test-tube contains exactly one unit of mixed reactants.

All four assembly scenarios are subjected to the above setups, and the results are summarized in Table 1. As it can be seen from the selected assembly scenarios, in most cases the differences are relative small. However, there exist both particularly bad and particularly good cases. Namely, the average value of the produced yield is 41.3%, the sample standard deviation is 8.2, the worst case scenario gives a yield percentage of 25%, while the best case scenario provides a yield percentage of 53.5%. It is very interesting to observe that both the best and the worst case scenarios are due to the same assembly strategy, but from different, i.e., non-optimized vs. optimized, parameter setups. Also, it can be observed that for each of the setups, the yield percentage are closed from one assembly scenario to the other, thus suggesting a possible ranking of how good each of these individual setups are. Namely, we are able to say that the worst parameter setting is performed in Setup 3 while if instead of just placing the previous default values we numerically optimize them to maximize the yield, i.e., Setup 5, then we obtain the best possible results from all the considered parameter setups.

6 Conclusions, Discussions, and Further Work

We have investigated yield optimization techniques for staged self-assembly systems. As a first step, we associated (for the first time) a computational model to the staged tile assembly formalism, whose implementation through ODE systems differs considerably from the kinetic counterpart of the regular TAM. This change of modelling methodology can be explained as follows. While in case of abstract TAM the assembly is initiated from a seed structure, and one can thus concentrate over a single assembly product, in case of sTAM the assembly reactions are implemented population wise.

Another important aspect was to determine the possible optimization strategies for our target, the final assembly yield. We were able to identify two levels on which to implement adequate optimization protocols: at the assembly scheme

design level, and at the implementation level. Considering the first level, several assembly strategies are plausible for the same final structure, and some of these assembly schemes may have plausible better chances of maximizing the concentration of the final product. We concentrate here only on those assembly schemes which ensure a minimal number of assembly stages. We conjecture at this level that the best assembly protocols are those in which we minimize the number of mixing of test-tubes from the same stages. The intuition here is that the more advance a stage is, the less concentrated its product, and thus by mixing a test-tube with another one from a lower stage, the concentration of the latter is higher and thus it improves the result of the reaction.

The second optimization level is at the implementation phase, once a particular assembly strategy has been chosen. At this level, the parameters which can be optimized are the time intervals allocated to each of the assembly stages (assuming the total available time is constant) and the proportions in which certain products are split and further mixed in subsequent stages. We believe that equal time-splits are not an optimal choice (unless the experiment involves a low number of stages), but the considered case-study showed that time-splits in geometric progression are also not appropriate (i.e., Setup 2 in Table 1). Although the case study seems to indicates that the optimal time-split parameters are close to an arithmetic progression (Setup 4 and 5 in Table 1, data on time-splits not shown), we believe that further studies are required for providing more intuition regarding a possible combinatorial design approaching the absolute optimum choice.

Regarding volume splits, an undiscriminating equal partitioning seem to be the worst possible choice (Setup 3 in Table 1). On the opposite direction, those partitions of assembly products which take into considerations the amount (or concentration) of substance in each test-tube, and how the product of these test-tubes are further going to be split, tend to have better yields.

Considering the case-study, one particular assembly strategy has generated good outcomes, namely that of requiring the volume of each test-tube to be exactly one. For the future, we plan to concentrate particularly on this strategy, both because it seem to provide the best results, and because it seem to be more tractable from an analytic point of view. Our aims are to provide concrete descriptions of combinatorial parameter-setups and mix-graph designs for which we could provide numeric arguments as why the results of these strategies approach the optimum solutions.

References

1. Cook, M., Fu, Y., Schweller, R.: Temperature 1 self-assembly: deterministic assembly in 3d and probabilistic assembly in 2d. In: Proc. of the 22nd Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2011, pp. 570–589. SIAM (2011)
2. Demaine, E.D., et al.: Staged self-assembly: nanomanufacture of arbitrary shapes with o(1) glues. Natural Computing 7(3), 347–370 (2008)
3. Demaine, E.D., Eisenstat, S., Ishaque, M., Winslow, A.: One-dimensional staged self-assembly. Natural Computing 12(2), 247–258 (2013)

4. Doty, D.: Theory of algorithmic self-assembly. Commun. ACM 55(12), 78–88 (2012)
5. Guldberg, C.M., Waage, P.: Studies concerning affinity. C. M. Forhandlinger: Videnskabs-Selskabet i Christiana (Norwegian Academy of Science and Letters) 35 (1864)
6. Guldberg, C., Waage, P.: Concerning chemical affinity. Erdmanns Journal fr Practische Chemie 127 (1879)
7. Hoops, S., et al.: Copasi - a complex pathway simulator. Bioinformatics 22(24), 3067–3074 (2006)
8. Padilla, J.E., Patitz, M.J., Pena, R., Schweller, R.T., Seeman, N.C., Sheline, R., Summers, S.M., Zhong, X.: Asynchronous signal passing for tile self-assembly: Fuel efficient computation and efficient assembly of shapes. In: Mauri, G., Dennunzio, A., Manzoni, L., Porreca, A.E. (eds.) UCNC 2013. LNCS, vol. 7956, pp. 174–185. Springer, Heidelberg (2013)
9. Rothemund, P.W.K., Winfree, E.: The program-size complexity of self-assembled squares. In: Proc. 32nd Annual ACM Symposium on Theory of Computing (STOC 2000), pp. 459–468. ACM, New York (2000)
10. Schulman, R., Winfree, E.: Synthesis of crystals with a programmable kinetic barrier to nucleation. Proceedings of the National Academy of Sciences 104(39), 15236–15241 (2007)
11. Schulman, R., Yurke, B., Winfree, E.: Robust self-replication of combinatorial information via crystal growth and scission. Proceedings of the National Academy of Sciences 109(17), 6405–6410 (2012)
12. Winfree, E.: Simulations of computing by self-assembly. Technical Report CaltechCSTR:1998.22, California Institute of Technology (1998)