Programmable Self-Assembly
- Unique Structures and Bond Uniqueness

Björn Högberg‡ and Håkan Olin§
Dept. of Physics, Mid Sweden University, SE-851 70 Sundsvall, Sweden

Abstract. An important problem in nanotechnology is to develop a method for assembling complex, aperiodic, structures. While simple self-assembly will not be able to address this problem, programmable self-assembly is powerful enough to be a potential solution for complex assembling tasks. Here, we address the question of how the basic properties of the constituent building blocks are related to the complexity of the resulting assembly. By introducing the parameters unique structures, which gives a measure of the complexity of an assembly, and bond uniqueness, which gives a measure of how the building blocks fit together, we show how to quantify the complexity of a general assembly system and present relations between the parameters. The introduced methods will be helpful when designing assembly systems to be used for direct fabrication of nanosystems or for nano-scaffolds and addressable arrays.

PACS numbers: 89.75.Fb

Submitted to: Nanotechnology

1. Introduction

Probably, the single most important problem in nanotechnology is to develop a method for assembling complex structures, such as a nanochip. Today, in microelectronics different types of lithography is used to define the structures. These methods are referred to as top-down methods. However, the continuing route towards smaller structures using top-down methods is increasingly difficult, which is reflected in fabrication costs that are growing much faster than the electronic market [1].

A bottom-up building approach based on self-assembly has been widely discussed as an alternative method for nanofabrication. In self-assembly the building blocks, that could be atoms, molecules or larger structures, diffuse around and eventually bind to a specific location. The simplest type of self-assembly, such as self-assembling monolayers or growth of nanowires, will merely lead to a non-complex crystal, and will be of little use to assemble a chip. We refer to this simplest type of self-assembly as crystal self-assembly (Fig. 1a). At the other extreme are building blocks that all have a unique address tag that will bind to a corresponding address. Using this unique addressing self-assembly method any kind of complex structure might be build, but a

‡ bjorn.hogberg@miun.se
§ hakan.olin@miun.se
A large number of building blocks are needed (Fig. 1b). Between these two extremes is programmable self-assembly (or algorithmic self-assembly) [2, 3] (Fig. 1c).

A formal model for studies of programmable self-assembly is the Tile Assembly Model [4] where the building blocks called tiles are self-assembled into a square lattice. The model is an extension of Wangs theory of tiling [5], but each of the sides of the square tiles contains a glue that allow binding of another tiles to the sides. The glue is specific so only a tile-side with the same corresponding glue will bind. Growth starts from a seed tile by adding one tile at the time. By designing the tiles in a specific way the growth can be programmed. Programmable self-assembly is powerful, for example, it has been shown to be capable of universal computation [6]. The Tile Assembly Model has been used for investigation of, for example, the minimum number of tiles needed to self-assembly a square of a certain size (program size complexity) [4], the time complexity [7], optimal size and time complexity [8], or whether a given tile system uniquely produce a given shape [9].

Experimentally, DNA double-crossover molecules with four sticky ends, which are analogues to the four sides of the Wang tiles, have been demonstrated to assemble into two-dimensional lattices [10]. Another experimental demonstration, of a much more complex pattern, is the algorithmic assembly of DNA Sierpinski triangles [2]. Macroscopic systems based on millimeter-scale plastic tiles that float on a liquid [11, 12] can also be used for experimental realization of algorithmic self-assembly [3]. Recently, DNA-linked nanoparticles has been suggested as a base for programmable self-assembly [13]. Another path, for doing programmable self-assembly, is to first self-assemble a scaffold of technologically simple building-blocks, then, self-assemble the interesting devices on the addressable sites created by the scaffold assembly. Potential scaffold materials are for example DNA-crystals [10, 14, 15] mentioned above or DNA-linked nanoparticles [16, 17] as well as more advanced biological systems like proteins [18] or viruses [19]. Using metallization of the biomolecules, the scaffold can also act as an electrical contact between the devices assembled upon the scaffold [20]. For reviews on some of these self-assembly systems see [21, 22].

Suppose, that a complex nanosystem, say an electronic chip, is going to be built using programmable self-assembly. From an engineering point of view it is important to know how to optimize the design of the assembly system given an available technology. This paper describes a solution to this problem by introducing two parameters: unique structures which gives a measure of how complex the final structure is and the bond uniqueness parameter - which is related to the specificity of the glue or bonds. Together with other building block parameters, such as the number of different types and their total number, we will give limits as well as trade-off relations between different assembly designs.

2. Classes of Self-Assembly

In the following we use the word tile to denote a general assembly building block. A tile have one or more functional edges with specific types of glues, or bonds. An open bond can be thought of as a binary number of $c_b$ bits. Here, $c_b$ is called the bond complexity. Two open bonds can form a bond if they are bitwise complements of each other (In some assembly systems, if they are bitwise identical).

Depending on the types of open bonds on the self-assembly tiles and on the physical process that makes the tiles assemble, we classify the process according to the type of assembly it can produce.
Figure 1. Classes of self-assembly systems and unique structures (encircled in gray) in generated assembly patterns. (a) Type 1 bonds can connect to type 0 bonds. Each of the tiles bonds can only connect to one other type of tile. This leads to crystal type assembly. The only unique structure is the entire assembly. (b) Each tile has bonds that can connect only to one other type of tile, and each tile has no more than one type of each bond. The resulting assembly type is unique addressing where each tile constitutes a unique structure. (c) Taking the same bond rules as in (a) but adding a third tile creates a new situation. Now each bond can host two different types of tiles. Together with the added criterion that each incoming tile must interact with at least two other tiles in order to assemble, the process is now of the PSA class. This assembly process needs to be nucleated for the assembly to grow. The resulting assembly in (c) has four unique structures of 10 tiles ($S = 4, \eta = 10$).

In Crystal self-assembly (fig. 1(a)) patterns of tiles are reproduced throughout the assembly. Since a crystal self-assembly (CA) repeats itself, the neighborhood of each tile must be identical everywhere in the assembly. This leads to the conclusion that each open bond of a crystal tile can only make a bond with one specific type of tile. This constraint on the assembly process assures that each tile of a certain type has the same type of tile neighbors everywhere.

In a unique addressing assembly (UA) (fig 1(b)), each tile type only occurs once and the tile position is completely defined [4]. Each bond must be specific as to what type of tile it can accept, like in the CA case above. There must also exist a unique way to arrange the tiles so that there are no more open bonds or so that the remaining open bonds are blocked by the assembly itself.

A Programmable Self-Assembly (1(c)) (PSA) system produces assemblies where one can find both repeating and unique patterns of tiles. For this to happen, some, or all, of the open bonds of a PSA tile must be able to bond to more than one type of tiles. Since an assembled PSA tile must, according to the above definition, accept at least two different types of new neighbors the process is not deterministic and thus not programmable; a single PSA tile alone cannot uniquely define the next tile in the assembly. At least one second neighboring, PSA tile needs to be present to uniquely determine the next tile type. One thus introduce the added criterion that a new tile must bind to at least two, already assembled, tiles; i.e. two tiles is the smallest structure that define the type of the third tile to be assembled. We do not know if this is generally true that programmable semi repetitive systems need this assembly criteria – we call it the three-to-tango conjecture. This feature seems to be the single
most important factor for the creation of complex assemblies. In nature, this type of dependent binding has been shown to be an important feature. For example in the self assembly of ribosomes certain proteins do only assemble if the growing ribosome assembly contains a certain pair of previously assembled proteins.[23]

Winfree and Rothemund has made a similar observation in discussions of the need for cooperation in order for the assembly process to be programmable [4]. Klavins et al[24] has proved that no binary grammar can generate a unique stable assembly, however, this proof is valid for self-assembly systems with conformal switching tiles, i.e. building blocks that change state after being assembled.

The number of bonds with already assembled tiles that an incoming tile must make in order for a tile to attach to the assembly is denoted $\tau$ (in the Tile Assembly Model [4] $\tau$ is called the temperature parameter). Following the above discussion $\tau \geq 2$ for any PSA process.

3. Unique Structures and Bond Uniqueness

As noted in the introductory discussion above, what makes self-assembly interesting for nanotechnology is the ability to construct scaffolds with uniquely addressable structures. In the following we will show how it is possible to estimate the number of unique structures that a certain assembly system can produce.

In each finite assembly, one can identify a certain number of unique structures, $S$. A unique structure is a number of connected tiles that is not reproduced anywhere else in the assembly. As an example, the string $ac$ is a unique structure in $aaabacabca$ while $ab$ is not. The minimum number of tiles needed to define a unique structure is denoted $\eta$ or, number of tiles per unique structure. The symbol $N$ will be used to specify the total number of tiles that constitute the assembly in the discussion that follows. The number of tile types will be denoted by $w$.

The patterns in fig. 1 are examples of a few simple cases where finding $S$ and $\eta$ is straightforward. In general it is always possible to divide an assembly into $S$ unique structures of $\eta$ tiles in each structure. A unique structure parser algorithm for finding the values of the pair $\eta$ and $S$ in any finite assembly, can be constructed. Note that unique structures may not overlap.

The pair $(S, \eta)$ is defined to be; the maximum possible $S$, and the $\eta$ that gives this maximum.

Consider the task of assembling a linear structure of $S$ addressable locations. Since each location should be uniquely addressable, the structure constituting the location must be globally unique. One way to proceed is to make exactly $S$ tiles that stick together in only one way (unique addressing, see fig. 2(a)), or to make some number $w$ of unique tiles where $w < S$ (PSA). If one chooses the later strategy one must figure out a clever way to make the $w$ tile types assemble into $S$ unique structures. Two examples of model PSA systems are found in Fig. 2(b) and (c). The prime-tower assembly (fig. 2(b)) counts to a product of two primes and then stops the assembly. This type of process is similar to the vernier process in biology, that is believed to be a length controlling mechanism in linear protein assembly [25]. The counter by Cheng, Goel and Moisset de Espanés (CGM) [8] is an assembly counter that is optimized to use as few tile-types as possible (optimized for low $w$) and still be able to produce arbitrarily long sequences. Each row is $\log_2 S$ tiles wide and constitutes a unique structure and $w = 8$. These two systems will be used as examples of pseudo-one dimensional assembly processes when comparing the values of $S$ and $\eta$ in different
The number of unique structures in a PSA assembly is limited by the number of distinct tile types, $w$, and the number of tiles per unique structure, $\eta$, according to the inequality $S \leq w^\eta$. Taking the logarithm of this, we get: $\eta \log w \geq \log S$. Note that this result gives a lower bound for $\eta$ or $w$, once $S$ and any one of $\eta$ or $w$ is defined.

When the total number of tiles, $N$, is constant, it is possible to draw some conclusions about $S$ and $\eta$ for several types of assembly processes. The simplest cases being unique addressing, where $S$ would be equal to $N$ and $\eta = 1$, and crystal type assembly, where $S$ would be 1 and $\eta = N$. Programmable self assembly lies somewhere in between. First note again that

$$S \leq \frac{N}{\eta}. \quad (1)$$

This defines an upper bound on $S$ for high values of $\eta$. For low values of $\eta$ the number of unique structures is heavily dependent on the number of tile types. The number of unique structures is limited by the number of ways one can combine $w$ tile-types in blocks of $\eta$-tiles in each block:

$$S \leq w^\eta \quad (2)$$

The inequalities (1) and (2) can be exchanged for equalities defining the maximum $S$. 

![Figure 2. A $\eta - w$ diagram where $S$ is held constant. The $w$-axis is logarithmic. The dotted line is $\eta = \log_2 S / \log_2 w$, which is the lower bound for all self-assembly processes and can be achieved by a random assembly. The dashed and dashed-dotted lines are given by the inequality (4) for the $b_\eta$'s given by the model systems in the insets (b) and (c) respectively. (The graph is plotted for $S=1000$, changing the value of $S$ will only change the scale and not the general characteristics.) The insets (a)-(c) show model systems of linear assembly displaying $S$ unique structures. The process in (a) is an example of a one-dimensional UA assembly. (b) shows a modulo-prime counter. Two sets of tiles, $a$’s and $b$’s, consisting of $p_a$ and $p_b$ tiles each ($p_a, b$ are primes). The assembly stops when tiles $a_{p_a}$ and $b_{p_b}$ are next to one another. To the right: physical model of the assembly. Left: a graph displaying the bond structure. In (c) the counter is of the type described in [8] (referred to as CGM-counter in this letter). This is a pseudo binary counter under the tile assembly model [4] that uses 8 distinct tiles. In the described model systems each row constitutes a unique structure.](image-url)
and the corresponding $\eta$. Combining (1) and (2) we get the following result for $S$ that holds for any random assembly structure:

$$S = N \frac{\log w}{\log S}$$  \hspace{1cm} (3)

PSA processes are not random however, so we expect $S(\eta)$ to grow more slowly than in (2) since the surrounding of a tile is more well defined than in the random case. This would decrease the likelihood that a given substructure is unique. The surrounding is not completely determined by one tile, but the choice of neighboring tiles is usually very limited in PSA depending on the bond uniqueness $b_u$ of the assembly system.

The bond uniqueness, $b_u$, for an assembly system is defined as the average number of tile types that each open bond can make a bond to. For example in the assembly system in fig. 1(c) the bond uniqueness is equal to two because each specific open bond can only harbor two different tile types on average. The first tile in a unique structure can be chosen at will from the $w$ available tile types. Once this tile is specified however, the following tile types can only be chosen from a subset of the tile types, containing $b_u$ tiles. So for PSA assemblies (2) should be replaced by the following expression:

$$S \leq w b_u^{\eta - 1}$$  \hspace{1cm} (4)

### 4. Results and Discussion

Combining (1) and (4) we get the following result:

$$S = N \frac{\log b_u}{\log (S b_u / w)}$$  \hspace{1cm} (5)
Equation (5) gives the maximum number of unique structures of a PSA process of $N$ total tiles with $w$ unique tile types and bond uniqueness $b_u$. This relation gives an optimal number of unique structures, in practice, the value of $S$ could be less than this. In fig. 3 a graphical interpretation of (5) is given as well as a comparison with the actual values for the model structures of fig. 2.

Depending on the application, structures that show repetitions if rotated may, or may not, count as unique. If we do not regard structures, that show repetitions while rotated, as unique, then we must divide the total number of unique structures by the rotational symmetry of the lattice. In the case of square tiles the factor is four. If one considers the case of constructing nanosystems using assembly of nanodevices the orientation of the nanodevices will often be important and we should thus take this factor into consideration.

Another important issue that will affect the assembled structure is the relative concentrations of the tile-types. In (5) we assume that the relative concentrations are all equal $1/w$. Any deviation from this will lead to a reduced number of unique structures.

The bond uniqueness parameter, $b_u$, is strongly influencing the structural complexity of the generated assembly patterns. The reduced bond uniqueness, $\tilde{b}_u$ is defined as $\tilde{b}_u = (b_u - 1)/(w - 1)$. $\tilde{b}_u$ is $= 0$ when the assembly is crystal or uniquely addressed and $= 1$ when the assembly is random. For PSA assemblies $0 < \tilde{b}_u < 1$.

Deterministic complexity (introduced by Solomonoff-Kolmogorov-Chaitin [26]) measures the amount of randomness, or information, in a structure. Janson et al. [27] uses the total number of unique substructures/substrings as a deterministic complexity measure, see also references in [27] for earlier work in this field.

Statistical complexity [28] is based on the notion that both randomness and periodicity are equally simple to describe statistically. Following the same path as Crutchfield [28] and Bennet [29] we denote objects as complex (having a high statistical complexity) if the amount of computation required to produce them is large.

Additional computation is defined as the amount of computation needed to correctly place a tile at a growth site, in excess of bond matching computation. To see if an open bond matches another open bond is fundamentally equivalent to performing $c_b$ AND operations. (Where $c_b$ is the bond complexity defined in sect. 2) This computation must be performed in any assembly processes. The additional computation embedded in PSA processes is the ability to select the correct tile given more than one open bond (in principle, this is equivalent to performing nested AND operations). Here, we propose to use the additional computation as a statistical complexity measure. The amount of additional computation needed to select the appropriate tile type to attach (bind) to an assembly is proportional to the number of incorrect tile types that needs to be sorted out times their bond complexity.

The number of tile types that the assembly process needs to take "into consideration" at each assembly step is the bond uniqueness times the $\tau$-parameter, $b_u \tau$. At least $\tau$ out of these $b_u \tau$ tile types will be the same, namely the correct tile type(s) that fits all bonds that needs to be filled. The number of incorrect tile types the process must consider is thus $(b_u - 1)\tau$ or, using the reduced bond uniqueness, $\tilde{b}_u (w - 1)\tau$. This value needs to be corrected for the fact that, as $b_u$ increases, the fraction of incorrect tile types is decreasing. (When $b_u$ reaches $w$ all tile types will fit and there are no incorrect tile types to consider.) The fraction of "available" incorrect tile types is $1 - \tilde{b}_u$ so the maximum additional computation, $C$, is thus given by

$$C = c_b \tau (w - 1) \tilde{b}_u (1 - \tilde{b}_u)$$
where \( w, t > 1, \ c_b \geq 1 \) and \( 0 \leq \tilde{b}_u \leq 1 \). The maximum additional computation is maximal around \( \tilde{b}_u = \frac{1}{2} \) and vanishing for \( \tilde{b}_u = 0 \) and \( \tilde{b}_u = 1 \), see the inset of fig. 3.

Since \( b_u \) and thus \( \tilde{b}_u \) can be approximated from the parameters \( \eta, N \) and \( w \), of an assembly using, (rewritten version of eq (4) ):

\[
b_u = \left( \frac{S}{w} \right) \frac{1}{\eta - 1} \quad (7)
\]

the statistical complexity of an assembly can be be estimated using the unique structure concept.

Looking now at the error rate in self-assembly, how is it related to the bond uniqueness? At thermal equilibrium the probability of forming a certain configuration \( L \) is given by [30]:

\[
P_L = \frac{\exp(-U_L/k_b T)}{\sum_L \exp(-U_L/k_b T)} \quad (8)
\]

where \( U_L \) is the energy of the formation, \( T \) is the temperature and \( k_b \) is Boltzmanns constant. Assuming that the energy of all bonds are equally strong and the binding energy is equal to \( U \). Assuming that the correct binding of a tile will have binding energy \( \tau U \) and any incorrectly bonded tile (only one bond fit) will have energy \( U \). Following the discussion above of the number of incorrect tile types, we get:

\[
P_L \simeq 1 - \tau \tilde{b}_u (w - 1) \exp(U/k_b T) \quad (9)
\]

The complimentary probability is the probability of errors and as shown in (9) it is proportional to \( \tilde{b}_u \). So \( \tilde{b}_u \) not only influences the structural complexity but also the error rate in physical assembly, if one is concerned about the correctness of the assembled structures it is thus advisable to chose a low value for the bond uniqueness.

5. Example of Application

The results in this paper will be helpful when designing nanosystems using self-assembly. Consider an example where one wants to construct a scaffold for nanodevices. Lets say that we have the technology to produce building blocks that measure 25 nm x 25 nm. 36 building blocks together should constitute one unique structure, giving us quadratic unique structures that measure about 150 nm x 150 nm. The scaffold should contain about 1090 unique structures covering an area of about 5 \( \mu \)m x 5 \( \mu \)m, \( N = S\eta = 39240 \). How many tile-types do we need? First we must decide approximately what value of the bond uniqueness we wish to use. A to high value will make the assembly process more random and give a higher error rate, a to low value would mean that an unnecessarily large number of tile-types is needed. Lets try \( \tilde{b}_u = 0.01 \). Equation (5) together with \( b_u = (w - 1)\tilde{b}_u + 1 \) leads to \( w \gtrsim 15 \) and thus \( b_u \gtrsim 1.14 \). Lets assume that the building blocks are 50 nm x 50 nm instead and we still want to squeeze in 1090 unique structures on the same surface, now \( \eta = 9 \) and \( N = 9810 \) which gives \( w \gtrsim 50 \) and \( b_u \gtrsim 1.49 \). Note that we have calculated limiting values, in practice, the values of \( w \) may be larger.
6. Conclusions

We have presented a method for the parametrization of assembly systems derived from their ability to form unique structures. We have also introduced the concept of bond uniqueness and showed how it influences the number of unique structures that a programmable self-assembly system can create. Furthermore, we have argued that programmable self-assembly systems creates embedded, additional computation that is reflected in the complexity of the generated structures. This structural complexity is directly dependent on the reduced bond uniqueness of the system. By using the relations obtained in this paper (mainly eq (5)) for nanotechnology applications it will possible to vary the parameters \( w \) and \( b_u \) to obtain the desired structural complexity and the desired size \( \eta \) of the addressable locations / unique structures. The concepts will also prove helpful when evaluating the theoretical limits of a proposed self-assembly technology at an early stage.

References


