Error Rate Reduction in DNA Self-Assembly by Non-Constant Monomer Concentrations and Profiling

B. Jang, Y-B. Kim, and F. Lombardi
Department of Electrical and Computer Engineering
Northeastern University, Boston, MA 02115
{bjang,ybk,lombardi}@ece.neu.edu

Abstract

This paper proposes a novel technique based on profiling the monomers for reducing the error rate in DNA self-assembly. This technique utilizes the average concentration of the monomers (tiles) for a specific pattern as found by profiling its growth. The validity of profiling and the large difference in the concentrations of the monomers are shown to be applicable to different tile sets. To evaluate the error rate new Markov based models are proposed to account for the different types of bonding (i.e. single, double and triple) in the monomers as modification to the commonly assumed kinetic trap model. A significant error rates reduction is accomplished compared to a scheme with constant concentration as commonly utilized under the kinetic trap model. Simulation results are provided.

1. Introduction

Molecular environments based on PCR-like reactions and DNA self-assembly have been proposed as potential solutions to achieve nano-scale integration [2]. There is substantial evidence that DNA self-assembly is one of the most promising alternatives for manufacturing future chips as current VLSI methodologies are fast reaching the limits of CMOS. In recent years, DNA self-assembly has been studied extensively [1].

Even though DNA self-assembly has potentially many advantages over traditional manufacturing mechanisms, many challenges are still left unsolved; in particular, process robustness is of a major concern, i.e. robustness refers to the tolerance of errors that may occur in the DNA self-assembly process. As the number of basic elements (so-called *monomers*, or *tiles*) (usually assumed to have the same constant concentration) required for self-assembly of molecular ICs is expected to be in magnitudes of many millions, even a modest reduction in the error rate has a significant impact on manufacturing. Hence, an extensive research on schemes for error tolerance has been pursued. Several works have been reported on error tolerance, mostly based on the utilization of massive redundancy in the tiles; *proofreading tile sets* [4] replace each original tile with a $K \times K$ block of tiles [6]. [5] has suggested a different scheme, namely *snake proofreading*.

In this paper, a reduction in error rate is achieved through pattern profiling and the utilization of non-constant monomer concentrations. The concentration of each tile is found based on the average demand in the desired self-assembly during the entire growth process. After profiling the growth patterns, a novel analysis of the bonding process is proposed to assess the error rate by utilizing new models. Extensive simulation results are reported to validate the proposed technique. It is shown that the proposed technique achieves significant error rate reductions with no use of redundancy.

2 Review

In his seed work [1], Winfree has considered the physical process of crystallization to model DNA self-assembly. During crystal growth, each monomer (as basic modular unit) is added at the boundary of the crystal already formed in a sequential fashion. The commonly assumed model for DNA self-assembly is the kinetic Tile Assembly Model (kTAM). In this model, the basic component is the so-called *monomer* (hereafter monomer and tile will be used exchangeably). An *aggregate* is formed by adding each tile to an existing smaller aggregate that initially started from a *seed* tile. In the kinetic Tile Assembly Model (kTAM), the following assumptions [1] are applicable: (1) Monomer concentration is held *constant*. Furthermore, all monomer types are held at the *same concentration*. (2) Aggregates do not interact with each other. So, the only reactions are the addition of monomers to existing aggregates. (3) The forward rate constants of all monomers are identical as in the

978-3-9810801-2-4/DATE07 © 2007 EDAA

hybridization of oligonucleotides. (4) The reverse rate depends exponentially on the number of base-pair bonds that must be broken.

The association and dissociation of tiles are controlled by two parameters: G_{mc} and G_{se}. G_{mc} represents the entropic cost of fixing the location of a monomer and is dependent on the monomer concentration. G_{se} represents the energy that is needed for breaking a single sticky-end bond, i.e. the side of a tile whose strength is one. Both parameters are numbers greater than zero. A further parameter is the *forward rate constant*, k_f. k_f doesn't affect the behavior of the model, but it sets the time units. By defining these parameters, the *rate of association* in kTAM is given by $r_f = k_f \times e^{-G_{mc}}$. The *rate of dissociation* with b sticky-end bonds is expressed as $r_{r,b} = k_f \times e^{-bG_{se}}$.

3 Profiling and Non-Constant Concentration

Consider the growth of any specific patterned shape in DNA self-assembly; differently from the assumption of kTAM, the demand for each monomer or tile is not constant among all monomers and varies with the desired pattern. Different patterns require different monomer concentrations and self-assembly must be provided with this information to efficiently execute, hence in this paper the first step for reducing the error rate consists of profiling a growth pattern using Xgrow as simulation tool. Figure 1 shows different plots of patterns for various tile sets. The growth snapshots have been obtained using Xgrow. The demand of each tile in the set (on the y-axis) is plotted versus the normalized growth time (on the x-axis). As observed in all patterns of Figure 1, not only the demand of each tile is not constant over the normalized growth time but also there is a heavy demand for some tiles than others throughout the growth process. Consider, for example, the *BinaryCounter* pattern in (c) and (d) of Figure 1. The E tile contributes the most, while all other tiles show a limited contribution to growth; moreover maximum demand of E is 1319 at a normalized time of 0.9, while the minimum demand is 754 at a normalized time of 0. The variation (between maximum and minimum values) is approximately 75%.

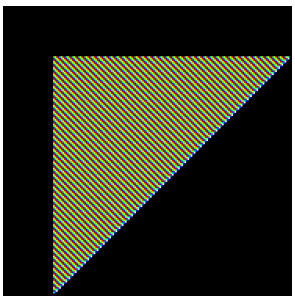
In kTAM, all tiles in the sets have the same (constant) concentration [1] [4]. While the concentration of each tile could be varied over time, control of the aggregate is very complex in laboratory work. However, the demand for a given tile varies with a pattern. To model and establish the relationship between monomer concentrations and error rate some assumptions must be made. (1) The concentration of each tile in tile set is controllable; monomer concentration is set by profiling a pattern and the average demand is utilized for each tile. (2) The total concentration is the same as in the original kTAM. Therefore, if the concentration of some tiles increases, then the same amount of concentration for the other tiles must be decreased. (3) The demand for a specified monomer is established over the entire growth period by profiling the pattern prior to self-assembly for large scale manufacturing.

4 Bonding Models

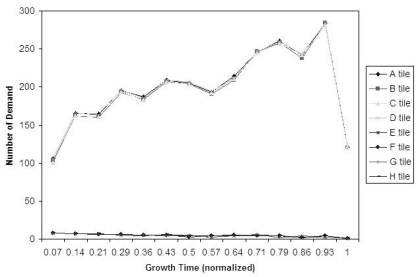
The kinetic trap state model has been used for describing and analyzing errors within kTAM [1]. In the kinetic trap state model, errors are defined to occur when the growth of an aggregate is faster than establishing local equilibrium. So, if an incorrect tile is surrounded by other tiles before it falls off, then it is trapped permanently in the self-assembly, and an error is said to occur. If the concentrations of all monometers are not the same, then the association rate for each tile is different and the traditional single trap model is not sufficient. Moreover, for a more accurate and complete analysis of the error rate in kTAM under constant monomer concentration, the kinetic trap model must be divided into three cases based on the bond condition of the empty tile site as target of the error. Bonding is defined by the number of sides (either 1, 2 or 3) between a tile and an aggregate. There exists no quadruple bonding because it corresponds to the growth completion of a site, Consider the scenario of *single bonding*. In this case by considering the target tile, there are three state levels: each level represents the number of adjacent tiles attached to the considered (target) tile. The state model with respect to the target tile is illustrated in Figure 2; it starts from the empty state (E_s) and a tile is attached to the site after some time, transiting to state C (correct tile is attached) or I (incorrect tile is attached). Finally, the target tile is trapped when one last remaining side is attached by a tile and it transits to the FC or FI state (from the C and I state respectively where F stands for frozen or permanently attached). The association and dissociation rates for each state are shown as labels on each edge of the state diagram. The association rate r^* denotes the approximate rate of growth, $r^* = r_f - r_{r,2}$.

For *double bonding*, the kinetic trap model is shown in Figure 3. For *triple bonding* the kinetic trap model is shown in Figure 4. In this case, there are only three state levels because the first bonding entails three sides of a tile. The tile attaching to the one remaining side of the target tile determines its final state: FC or FI.

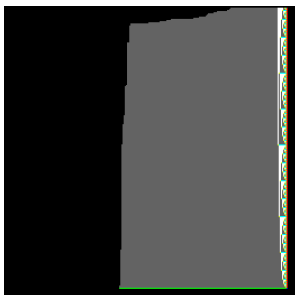
In the proposed non-constant concentration case, the state diagram for each types of a tile is different because the association and dissociation rates for the tiles are changed. As an example, consider in the Sierpinski tile set, the change of the AA tile's concentration while each of other six tiles share the remaining concentration equally. The kinetic trap models for the rule tiles in the single, double, and triple bonding conditions are shown in Figure 5, Figure 6, and Figure 7, respectively. In all cases, the association rates



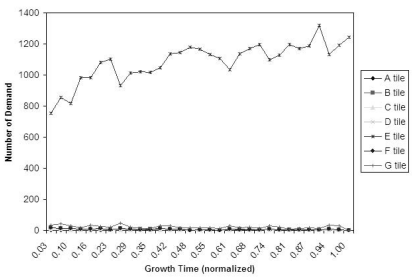
(a)



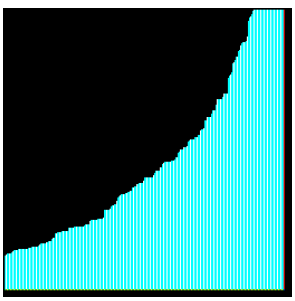
(b)



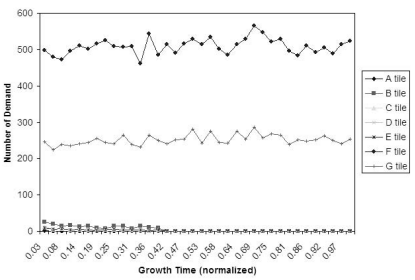
(c)



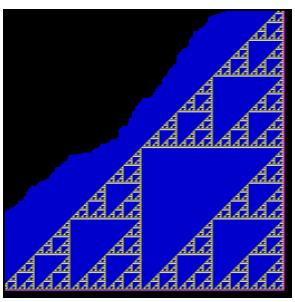
(d)



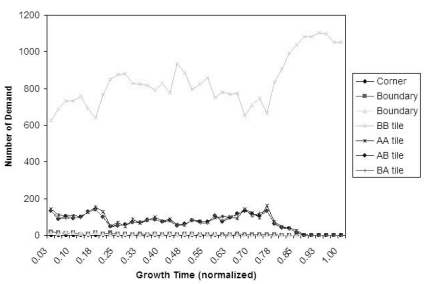
(e)



(f)



(g)



(h)

Figure 1. Profiling of Tile Demand during Growth (a) Snapshot of *Barseed* growth, (b) Tile demand during *Barseed* growth, (c) Snapshot of *BinaryCounter* growth, (d) Tile demand during *BinaryCounter* growth, (e) Snapshot of *Lines2* growth, (f) Tile demand during *Lines2* growth, (g) Snapshot of *Sierpinski* growth, (h) Tile demand during *Sierpinski* growth,

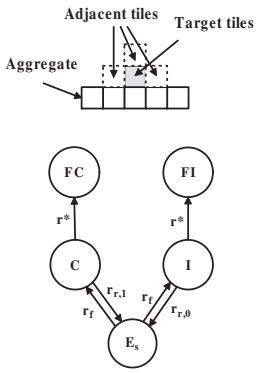


Figure 2. Kinetic trap model for single bonding case under constant monomer concentration: target tile mode

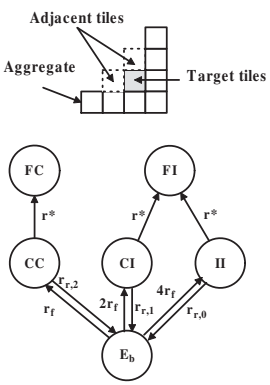


Figure 3. Kinetic trap model for double bonding under constant monomer concentration: target tile model

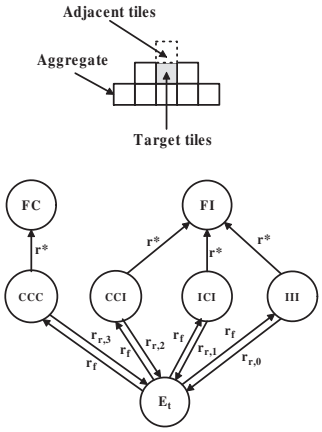


Figure 4. Kinetic trap model for triple bonding under constant monomer concentration: target tile model

for the AA tile and the other tiles are different and denoted as $r_{f,x}$ and $r_{f,y}$, i.e. in the non-constant monomer concentration case, the association and dissociation rates for each tile are changed, resulting in different rate equations. Consider the Sierpinski tile set, and assume that the AA tile has a concentration of 40% and each of the other six tiles has a 10% concentration. The error rates are plotted in Figure 8 for single bonding, Figure 9 for double bonding and Figure 10 for the triple bonding. These results confirm that the single bonding case is not sufficient to describe the kinetic trap model, as observed by the abrupt shape of the plot. This can be intuitively understood as the maximum bonding strength is only one (i.e. a so-called easy dissociation occurs) and the state transition is too simplified to capture the behavior of assembly.

5 Evaluation and Conclusion

The average demand for each tile for various patterns have been calculated from profiling using Xgrow and are summarized in Table 1. The results for the error rate are reported in Table 2. For all patterns, a significant improvement in error tolerance is achieved by non-constant monomer concentration. It is also shown that error tolerance by non-constant monomer concentration is possible for patterns in which smaller tile types account for a large portion of the total demand. For example, approximately a 70% reduction in error rate is achieved for *BinaryCounter* in which one tile type takes nearly 96% of the total demand.

In conclusion, this paper has shown that monomer concentration can be used for reducing error rates during DNA self-assembly by employing no redundancy. A novel model

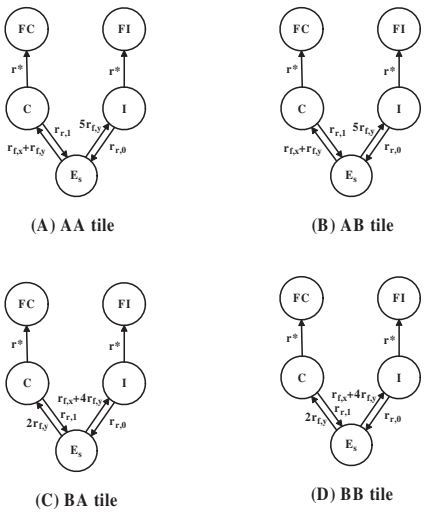


Figure 5. Kinetic trap model for single bonding under non-constant monomer concentration

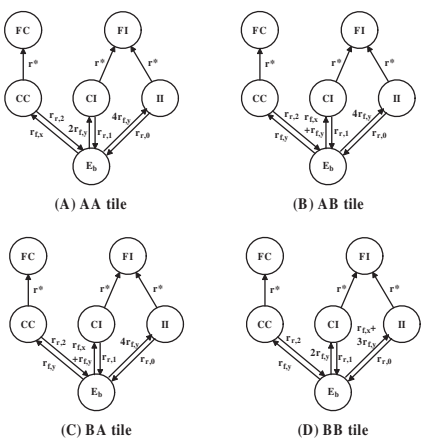


Figure 6. Kinetic trap model for double bonding under non-constant monomer concentration

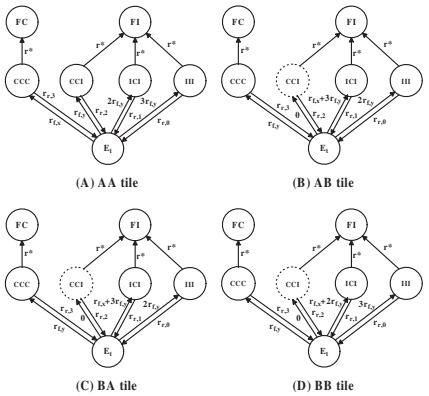


Figure 7. Kinetic trap model for triple bonding under non-constant monomer concentration: the dotted circle denotes a dummy state used for readability

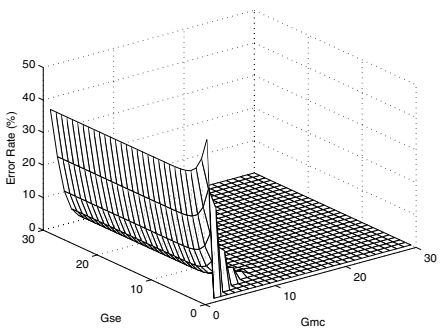


Figure 8. Error Rate of Single Bonding in Non-Constant Monomer Concentration (AA tile = 40%, Other tiles = 10%)

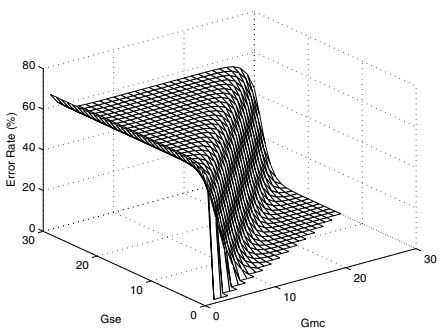


Figure 9. Error Rate of Double Bonding in Non-Constant Monomer Concentration Case (AA tile = 40%, Other tiles = 10%)

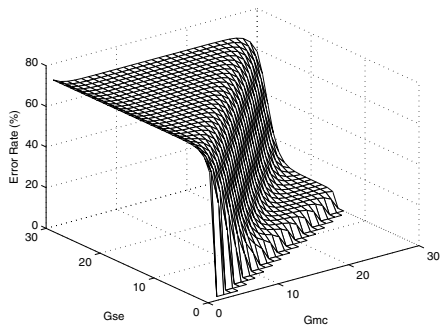


Figure 10. Error Rate of Triple Bonding in Non-Constant Monomer Concentration (AA tile = 40%, Other tiles = 10%)

Table 1. Average demand for each tile in various patterns

Pattern	Tile Types	Average % of demand during growth
Barseed	8	A=24.58% B=24.29% C=24.29% D=24.58% E=0.6% F=0.6% G=0.6% H=0.6%
Binary-Counter	7	A=0.003% B=0.62% C=0.36% D=0.62% E=95.87% F=0.61% G=1.962%
Lines2	7	A=0.004% B=0.68% C=0.2% D=0.2% E=0.2% F=66% G=33%
Sierpinski	7	Seed=0.002% Boundaries (both)=0.94% BB=78.3% AA=7.24% AB=6.74% BA=6.76%

Table 2. Error rates for different schemes in various patterns

Pattern	Constant Concentration	Non-Constant Concentration	Error Reduction
Barseed	0.0672%	0.0658%	2.083%
Binary-Counter	0.1341%	0.0398%	70.321%
Lines2	0.0671%	0.0664%	1.043%
Sierpinski	0.1341%	0.1206%	10.067%

that extends the kinetic Tile Assembly Model (kTAM), has been presented; this new model (as a variant of the kinetic trap model) accounts for the scenario in which there is a different concentration for each monomer. By profiling different patterns, it has been shown that tile demands varies over the growth process, hence monometer concentration is not constant (as commonly assumed in existing models). Bonding analysis of the self-assembly process has been pursued to confirm that monomer concentration has a significant impact on tolerance to errors. The simulation results have shown that monomer concentration is a promising scheme for error tolerance in DNA self-assembly.

References

[1] E. Winfree, "Simulation of Computing by Self-Assembly", Technical Report CS-TR:1998.22, Caltech, 1998

[2] K. Sakamoto, D. Kiga, K. Komiya, H. Gouzu, S. Yokoyama, S. Ikeda, H. Sugiyama, and M. Hagiya, "State transitions by molecules", 4th Int. Meeting on DNA-Based Computing, Baltimore, Penns., June, 1998

[3] E. Winfree et al, "The xgrow simulator", http://www.dna.caltech.edu/Xgrow/xgrow-www.html.

[4] E. Winfree and R. Bekbolatov, "Proofreading tile sets: Error-correction for algorithmic self-assembly", DNA Computers 9, LNCS volume 2943:126-144, 2004

[5] H.-L. Chen and A. Goel, "Error free self-assembly using error prone tiles", DNA Computing 10, LNCS volume 3384:62-75, 2005

[6] D. Soloveichik and E. Winfree, "Complexity of Compact Proofreading for Self-Assembly Patterns", DNA Computing 11, 2006.