

Error Tolerance of DNA Self-Assembly by Monomer Concentration Control

Byunghyun Jang, Yong-Bin Kim and Fabrizio Lombardi

Department of Electrical and Computer Engineering

Northeastern University

Boston, MA 02115

Email: {bjang, ybk, lombardi}@ece.neu.edu

Abstract—This paper proposes the control of monomer concentration as a novel improvement of the kinetic Tile Assembly Model (kTAM) to reduce the error rate in DNA self-assembly. Tolerance to errors in this process is very important for manufacturing highly dense ICs; the proposed technique significantly decreases error rates (i.e. it increases error tolerance) by controlling the concentration of monomers. A stochastic analysis based on a new state model is presented. Error rates reductions of at least 10% are found by evaluating the proposed scheme comparing to a scheme with constant concentration. One of the significant advantages of the proposed scheme is that it doesn't entail an overhead such as increase in size and a slow growth, while still achieving a significant reduction in error rate.

Keywords: DNA self-assembly, error tolerance, molecular manufacturing, tiling.

I. INTRODUCTION

Lithography-based manufacturing of integrated circuits (IC) is fast approaching its limits, as predicted by the Semiconductor Technology Roadmap. High density and smaller feature sizes necessitate novel techniques, such as molecular based implementations. Molecular environments based on PCR-like reactions and DNA self-assembly have been proposed as potential solutions to achieve nano-scale integration. DNA self-assembly relies on a Turing-based universal model developed by Winfree in 1996 [1]; although only few experimental demonstrations and logic operations have been reported and implemented, 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. The possible advantages of DNA self-assembly are usually estimated through simulation by considering chemical thermodynamics and kinetics to establish the basic framework by which implementations should abide. In recent years, DNA self-assembly has been studied extensively. The kinetic Tile Assembly Model (kTAM) was proposed in 1998; kTAM is a realistic model that considers association and dissociation of basic molecular elements (so-called monomers, or tiles) within the original framework provided by the so-called abstract Tile Assembly Model (aTAM) [2]. Several tile implementations such as *AB*, *Bar Code*, and *Sierpinski* based on kTAM have been simulated, mostly using Xgrow [3] as tool.

Even though DNA self-assembly has potentially many advantages over more 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. Several works have been reported on error tolerance; *proofreading tile sets* [4] replace each original tile with a $K \times K$ block of tiles. A new tile set is therefore generated and a reduction in error rate (by order of magnitude compared with the original tile sets [7]) has been reported under specific conditions. [5] has suggested a different scheme, namely *snaked proofreading*. In this scheme the original tile set is replaced by a $2K \times 2K$ block of tiles; however, snake proofreading, uses a block of tiles different from proofreading. It has been proved that snake proofreading can correct many errors (such as for example, growth and nucleation errors) and significantly reduce the error rate in self-assembly. In [6], self-healing tile sets are proposed, where algorithmic self-assembly is able to heal damage to a self-assembled object. They presented block transformations that convert damaged areas into new tile sets. As the number of tiles required for the self-assembly of molecular ICs is expected to be in magnitudes of at least 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. In this paper, monomer concentration is proposed as a new scheme for error tolerance in DNA self-assembly. The concentration of each tile is controlled during growth and depending on the demand of each tile, the effects of this scheme are assessed with respect to tolerance to errors. The error rate is analyzed through a stochastic analysis by utilizing a new state model.

In the next section, the DNA self-assembly model is briefly described for review. Section III describes the proposed scheme of monomer concentration control and analyzes the effects of error tolerance on DNA self-assembly. The state model of the proposed scheme is presented in Section IV. The evaluation of the proposed scheme is given in Section V. Finally, Section VI concludes the paper.

II. DNA SELF-ASSEMBLY

In his seed work, 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. There are two DNA self-assembly models:

- The abstract Tile Assembly Model (aTAM);
- The kinetic Tile Assembly Model (kTAM).

A. The Abstract Tile Assembly Model (aTAM)

In this model, the basic component is the so-called unit square *tile* (also 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. Many types of tile are possible and throughout the literature, it is commonly assumed that the supply of each tile is *unlimited*. However, in this model tiles are rather static during assembly, i.e. a tile can not be rotated or reflected. Each edge of a tile is represented by a label and an associated strength (that is a non-negative number). In this model, a further important parameter is given by the temperature τ . A tile aggregate can grow by addition of a further tile whenever the summed strength of the label matching edges exceeds τ . Figure 1 shows the Sierpinski tile set in the abstract Tile Assembly Model (aTAM).

- (A) shows the tile set (consisting of seven tiles); the tiles are classified as rule tile (four of them), boundary tile (two of them) and seed tile (only one).
- (B) shows an example of an aggregate formed arbitrarily when $\tau = 0$.
- (C) and (D) show examples of aggregates formed when $\tau = 1$ and $\tau = 2$, respectively.

At temperature $\tau = 0$, every possible tile addition, is stable and the resulting aggregate is random. At temperature $\tau = 1$, any addition that has a strength greater than 1 is stable. Its resulting aggregate is mostly random, but it can have any patterned shape, depending on the sequence of the additions. At temperature $\tau = 2$, however, there is an unique addition, so its resulting aggregate has also an unique pattern that depends on the employed tile sets.

B. Kinetic Tile Assembly Model (kTAM)

In practice, DNA self-assembly entails a more complicated process than the simple model described previously. For instance, there are an *infinite number* of different types of monomer (tile) and the bonding strength for each edge is not a discrete integer, but a continuous number. Therefore, a different model is required for a realistic simulation of this process. Winfree suggested the kinetic Tile Assembly Model (kTAM). The assumptions in the kTAM [2] are as follows.

- Monomer concentration is held *constant*. Furthermore, all monomer types are held at the *same concentration*.
- Aggregates do not interact with each other. So, the only reactions are the addition of monomers to existing aggregates.
- The forward rate constants of all monomers are identical as in the hybridization of oligonucleotides.
- The reverse rate depends exponentially on the number of base-pair bonds that must be broken.

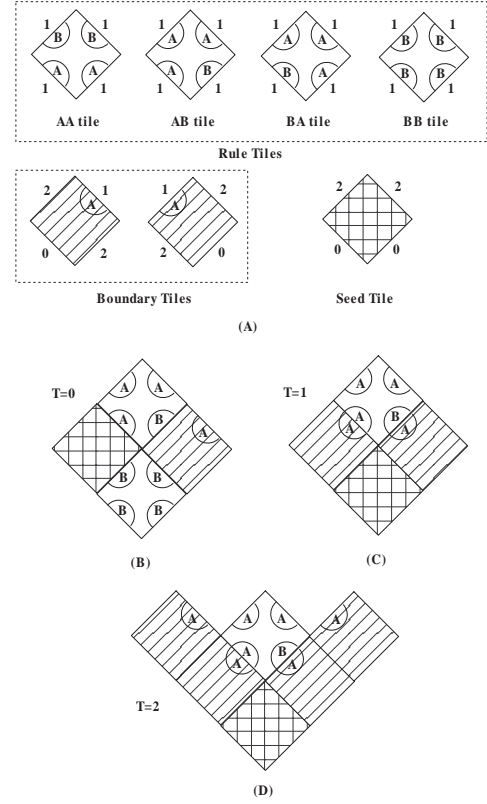


Fig. 1. Sierpinski tile set in aTAM

The association and dissociation of tiles in this model 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 need 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

$$r_f = k_f e^{-G_{mc}} \quad (1)$$

The *rate of dissociation* with b sticky-end bonds is expressed as

$$r_{r,b} = k_f e^{-bG_{se}} \quad (2)$$

Therefore, the association rate of monomers increases exponentially as the monomer concentration increases, while the dissociation rate of monomers decreases exponentially as the sticky-end bonds increase. Figure 2 illustrates the possible association and dissociation of a monomer onto existing aggregates in the kinetic tile assembly model for the Sierpinski tile set.

III. MONOMER CONCENTRATION CONTROL

In this section, monomer concentration control is investigated in detail. Consider the growth of DNA self assembly; it is then obvious that the demand for each tile varies with time.

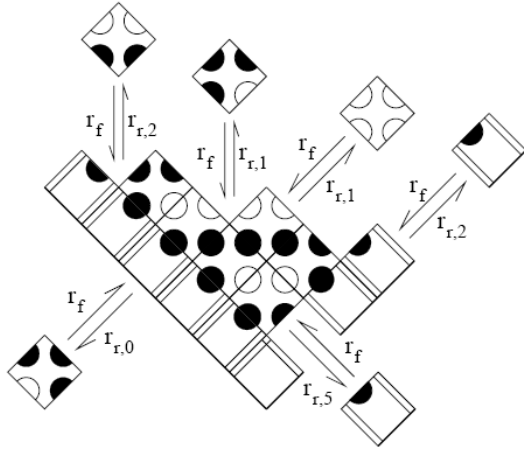


Fig. 2. kTAM ([2])

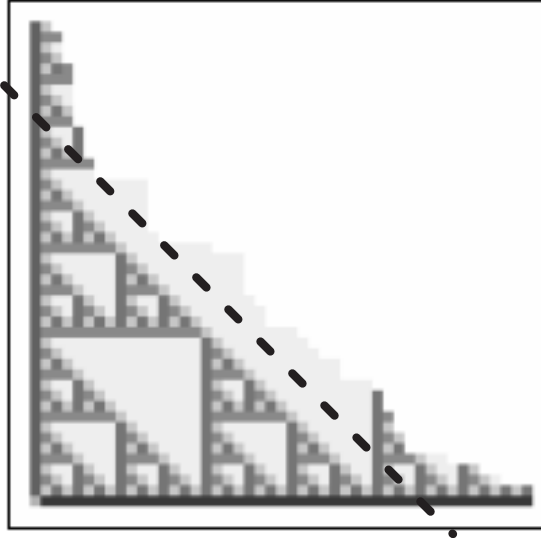


Fig. 3. Snapshot of the growth of the Sierpinski tile set ([2])

As an example, when the large triangle is being generated in Figure 3, the BB tile is needed the most for a correct growth of the aggregate. As the error rate and the speed of assembly depend on the probability that a tile is attached, intuitively if the concentration of the BB tile is increased, then growth will be fast and the error rate will hopefully be low.

The DNA self-assembly rate is a function of the *monomer concentration* (i.e. $r_f = k_f e^{-G_{mc}}$). Moreover, it is commonly assumed that all tiles have the same concentration. If this assumption is *not applicable* and the concentration of each tile can be controlled, a different scenario is therefore possible. For example, assume that any of the seven tiles in the Sierpinski set can be controlled; then, the tile whose concentration is increased, would associate faster than before, while the other tiles would associate slowly. There is also a higher probability that an error would occur if the demand for that tile remains the same. However, as mentioned previously, the demand for

a given tile varies as the aggregate grows. This phenomenon implies that there is a point at which the error rate is at the lowest level, or in general it can be reduced. To model and establish the relationship between monomer concentration and error rate based on the demand of each tile, some assumptions are made within kTAM, as follows.

- The concentration of each tile in the set is controllable at any moment during self-assembly.
- The forward rate of a tile is proportional to the concentration of that tile. This is applicable because the exact relationship between G_{mc} and concentration is not known.
- The total concentration is the same as in the original kTAM. Therefore, if the concentration of a tile increases, then the same amount of concentration of the other tiles will be decreased.
- The demand for a specified monomer is known at any moment during the growth of the aggregate.

IV. STATE MODEL

If the concentration of a monomer is controlled, then the association rate for each tile will be different and the single trap model shown in Figure 4 can not describe the kinetic tile assembly model. Consider the scenario in which the concentration of the AA tile of the Sierpinski tile set (shown in Figure 1) is changed. Let the association rate for the AA tile be higher, therefore the association rate for all other tiles will be lower based on the assumption that the total concentration remains unchanged. In this case, the kinetic trap model for each tile is different as shown in Figure 5.

Consider the probability of a particular tile being present in a site as a function of time. For the model and corresponding state diagram, six cases can be distinguished for a site.

- E is the empty state and every site starts from the empty state.
- C represents the state in which the correct tile is attached and its dissociation rate is $r_{r,2} = k_f e^{-2G_{se}}$.
- The A state means only one boundary matches and its dissociation rate is $r_{r,1} = k_f e^{-G_{se}}$.
- I represents the state in which the incorrect tile is attached and its dissociation rate is $r_{r,0} = k_f$.
- FC and FI are the states that are frozen correctly and incorrectly respectively. In this context, "frozen" indicates the scenario in which at the surrounding of the site another tile is added (the tile currently in place is effectively on a permanent basis [2]).

Once the FC and FI states are reached, the state transition is said to be complete, i.e. a slot is not available for both association and dissociation (unless an undefined event occurs). This kinetic trap model is shown in Figure 4.

In a monomer concentration, the state diagram of each tile is different because the association and dissociation rates for the tiles are also different. As an example, consider the AA tile at 40% of the total concentration, while each of the other six tiles take 10% of the total concentration. The kinetic trap

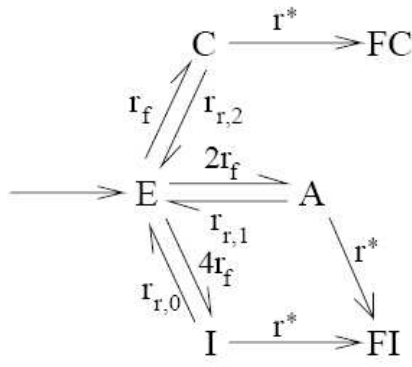


Fig. 4. Model for kinetic trapping at a single growth site ([2])

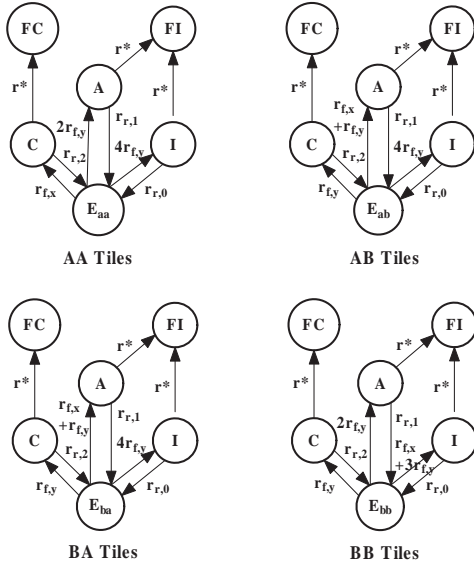


Fig. 5. Proposed kinetic trap model for rule tiles

models for the rule tiles are shown in Figure 5. In this case, the forward rates for the AA tile and the other tiles are different and denoted as $r_{f,x}$ and $r_{f,y}$ respectively. The forward rates from the E state to states C , A , and I for each tile are changed as shown in Figure 5.

Let $p_i(t)$ be the probability that (i) is the state at t seconds after the site has appeared. For example, the rate equations for the AA and AB tiles can be written as

$$\dot{\mathbf{p}}_{AA}(t) = \begin{bmatrix} -r^1 & r_{r,2} & r_{r,1} & r_{r,1} & 0 & 0 \\ r_{f,x} & -r^2 & 0 & 0 & 0 & 0 \\ 2r_{f,y} & 0 & -r^3 & 0 & 0 & 0 \\ 4r_{f,y} & 0 & 0 & -r^4 & 0 & 0 \\ 0 & r^* & 0 & 0 & 0 & 0 \\ 0 & 0 & r^* & r^* & 0 & 0 \end{bmatrix} \begin{bmatrix} p_E(t) \\ p_C(t) \\ p_A(t) \\ p_I(t) \\ p_{FC}(t) \\ p_{FI}(t) \end{bmatrix} \\ \doteq \mathbf{M}\mathbf{p}_{AA}(t)$$

(where $r^1 = r_{f,x} + 6r_{f,y}$, $r^2 = r_{r,2} + r^*$, $r^3 = r_{r,1} + r^*$, and

$r^4 = r_{r,0} + r^*$) and

$$\dot{\mathbf{p}}_{AB}(t) = \begin{bmatrix} -r^1 & r_{r,2} & r_{r,1} & r_{r,1} & 0 & 0 \\ r_{f,y} & -r^2 & 0 & 0 & 0 & 0 \\ r^5 & 0 & -r^3 & 0 & 0 & 0 \\ 4r_{f,y} & 0 & 0 & -r^4 & 0 & 0 \\ 0 & r^* & 0 & 0 & 0 & 0 \\ 0 & 0 & r^* & r^* & 0 & 0 \end{bmatrix} \begin{bmatrix} p_E(t) \\ p_C(t) \\ p_A(t) \\ p_I(t) \\ p_{FC}(t) \\ p_{FI}(t) \end{bmatrix} \\ \doteq \mathbf{M}\mathbf{p}_{AB}(t)$$

where $r^1 = r_{f,x} + 6r_{f,y}$, $r^2 = r_{r,2} + r^*$, $r^3 = r_{r,1} + r^*$, $r^4 = r_{r,0} + r^*$, and $r^5 = r_{f,x} + r_{f,y}$. As FC and FI are both sink states, then the probability that growth occurs with no error, is given by $\mathbf{p}_{FC}(\infty)$, while the probability that growth has at least an error is $\mathbf{p}_{FI}(\infty)$.

V. EVALUATION

As mentioned previously, the demand for monomers must be known a-priori to determine the most appropriate concentration for the lowest error rate. Once the demand for monomers is known for a particular application, the error rate can be calculated. An evaluation of the proposed scheme has been pursued and results are presented.

- The plots of the error rate versus the AA tile concentration when $G_{mc} = 13, 14, 15$ and $G_{se} = 8$ are shown in Figures 6, 8, and 10 respectively.
- The plots of the resulting error rate versus the concentration of other tiles (different from AA) for $G_{mc} = 13, 14, 15$ and $G_{se} = 8$ are shown in Figures 6, 8, and 10.

The horizontal straight line in each graph indicates the error rate for the scenario in which the concentration is constant for all tiles (as assumed in previous models found in the literature). The error rate changes as the concentration of tiles varies; the smallest error rates by varying the concentration of the AA tile are 1.19%, 0.444%, and 0.162% (for constant concentration, the error rates are 1.32%, 0.49%, and 0.18%). Reductions of 10.9%, 10.3% and 11.1% have occurred as result of controlling the concentration of the tiles; note that the number of tiles involved in DNA self-assembly of molecular chips is expected to be in the order of at least many millions, so a reduction of 10% is a significant figure. Also, the error rate is the lowest at approximately 27% for the concentration of the AA tile and 12.2% for the concentration of the other tiles.

VI. CONCLUSION

This paper has proposed monomer concentration control for reducing error rates during DNA self-assembly. A novel model that extends the kinetic Tile Assembly Model (kTAM) has been presented; this new model accounts for the scenario in which there is a different concentration for each monomer and it is controllable. The state diagram of this process has been provided. The results have shown that monomer concentration is a promising scheme for error tolerance in DNA self-assembly. Its most significant advantage is that it doesn't incur in overhead (such as those encountered in previous schemes based on proofreading [4] and snake tiling [5]).

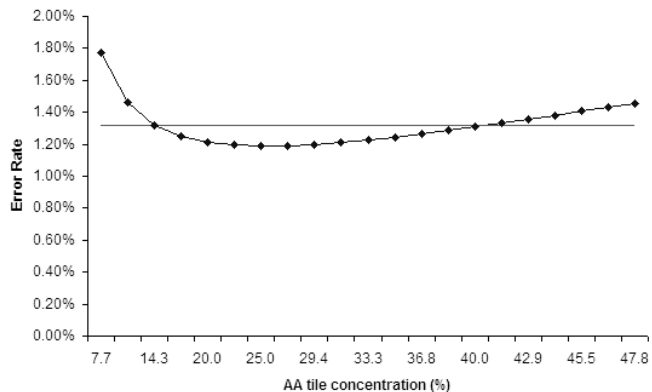


Fig. 6. Error Rate vs. AA Tile Concentration ($G_{mc} = 13$ and $G_{se} = 8$)

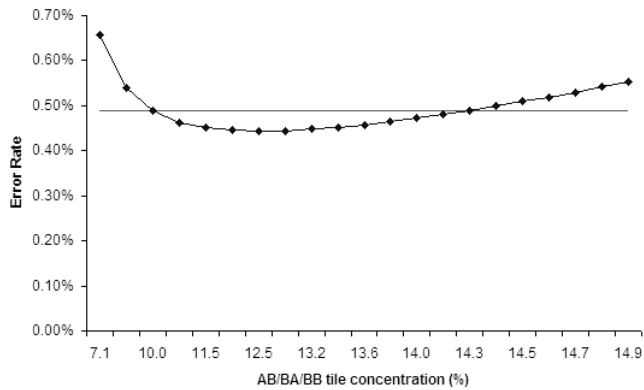


Fig. 9. Error Rate vs. AB/BA/BB Tile Concentration ($G_{mc} = 14$ and $G_{se} = 8$)

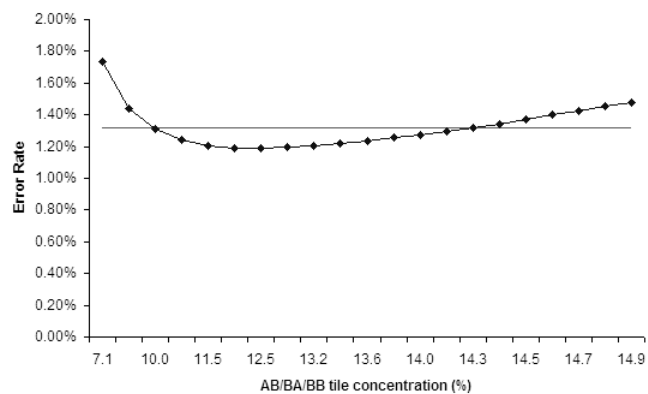


Fig. 7. Error Rate vs. AB/BA/BB Tile Concentration ($G_{mc} = 13$ and $G_{se} = 8$)

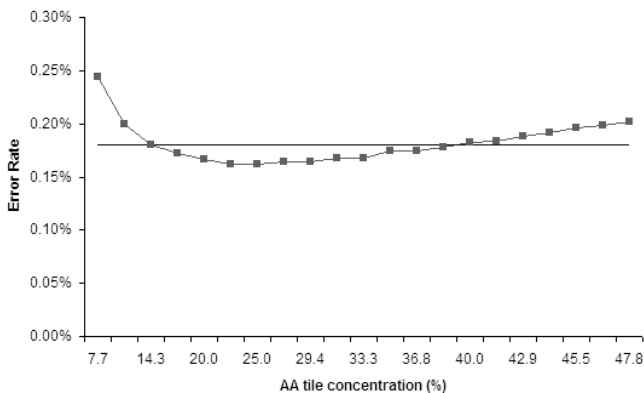


Fig. 10. Error Rate vs. AA Tile Concentration ($G_{mc} = 15$ and $G_{se} = 8$)

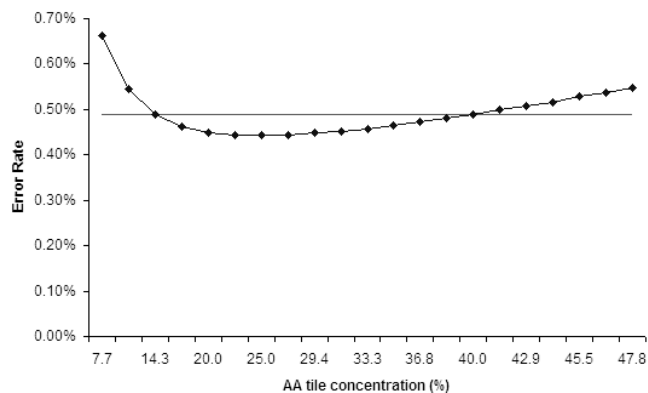


Fig. 8. Error Rate vs. AA Tile Concentration ($G_{mc} = 14$ and $G_{se} = 8$)

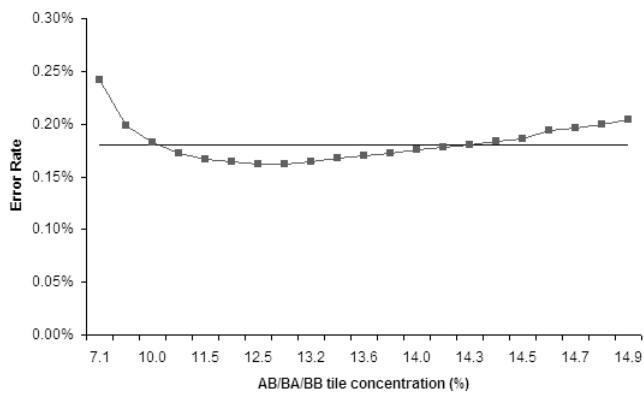


Fig. 11. Error Rate vs. AB/BA/BB Tile Concentration ($G_{mc} = 15$ and $G_{se} = 8$)

REFERENCES

- [1] E. Winfree, "Algorithmic Self-Assembly of DNA", Ph.D thesis, California Institute of technology, Pasadena, 1998
- [2] E. Winfree, "Simulation of Computing by Self-Assembly", Technical Report CS-TR:1998.22, Caltech, 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] E. Winfree, "Self-Healing Tile Sets", Nanotechnology: Science and Computation, pages 55-78, 2006.
- [7] D. Soloveichik and E. Winfree, "Complexity of Compact Proofreading for Self-Assembly Patterns", DNA Computing 11, 2006.