

Probabilistic Models for Damage and Self-Repair in DNA Self-Assembly

U. Majumder*

* Department of Computer Science, Duke University, Durham NC, USA, urmim@cs.duke.edu

ABSTRACT

Since its inception, the focus of DNA self-assembly based nanostructures has mostly been on one-time assembly. However, DNA nanostructures are very fragile and prone to damage. Knowing the extent of damage that can occur under various physical conditions can be useful in making robust designs for self-assembled nanostructures. Thus in this paper, we present simple models for estimating the extent of damage in DNA nanostructures due to various external forces. We note that these models have not been validated against experimental data and are only meant to serve as a basis for designing DNA nanostructures that are robust to external damage. We conclude with a discussion on computing the probability of repair of a damaged nanostructure.

Keywords: dna, self-assembly, damage, self-repair, probability

1 Introduction

DNA self-assembly is an autonomous phenomenon where components (single strands or DNA “tiles”¹) organize themselves into stable superstructures. In recent years, DNA based self-assembled nanostructures have gone from conceptual design to experimental reality [2]. They can, however, be very fragile. For instance, observe the DNA nanotube in Figure 1 which has been partially opened up by an AFM tip during imaging.

Winfree [1] briefly investigated if a hole in a DNA lattice will repair correctly in a kinetic simulation model. However, to the best of our knowledge, there has been no work on either modeling the extent of damage in DNA nanostructures or any experimental measurement of it. Short of an experimental demonstration, we felt that a realistic model for lattice damage and self-repair would enable us to better evaluate the current designs for self-repairing tile set. We study the extent of damage due to external mechanical forces as well as intense radiation. Furthermore, we compute the probability of self-repair at equilibrium.

¹DNA tiles are made of several individual DNA strands with single strand overhangs (also called “sticky ends”) that allow them to assemble into a lattice.

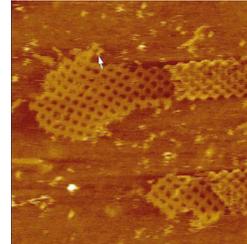


Figure 1: *Damage in DNA nanostructures: broken DNA nanotubes due to repeated AFM scanning.*

2 Mechanical Damage Model

Suppose an AFM tip strikes a tile on a lattice composed of several DNA tiles. If we assume that our lattice is a rigid surface then the effect of the hitting force on a neighboring tile can be modeled as a function of its distance from the source of the impact [3]. In this model, the probability that a tile will fall off is given by the probability that a shock wave from the hitting force propagates to this tile along some path (Probabilistic Damage Model: see Section 2.1). However, if we assume that the lattice is flexible, we can use a mass spring model for the lattice and hence calculate the effect of the hitting force on the lattice [5] (Flexible Lattice Model: see Section 2.2). One important observation that we should make here is that we do not know yet whether these models are appropriate, since they have not been verified experimentally.

2.1 Probabilistic Damage Model

We assume that the force F_1 on a tile located at a distance of r from the tile receiving the impulse is proportional to $\frac{1}{\sqrt{r}}$ [3]. In our model, r is the Manhattan distance of the tile under consideration from the source of the impact. F_2 is the resistive force from the sticky end connections of the tiles. For simplicity, we assume F_2 to be the same for all the tiles. For any tile with $F_1 > F_2$, the probability that a tile gets knocked off the lattice is greater than zero so long as the shock wave has reached it from the origin of damage. To estimate the fraction of the lattice damaged, we first compute the probability of a damage path of length i . A *damage path* is defined as a path that originates in the tile which is

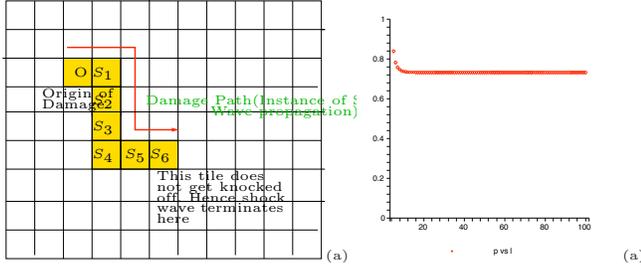


Figure 2: (a) Instance of shock wave propagation and creation of damage path, (b) Estimation of p

directly hit by the tip (say O) and meanders outwards through its successors $\langle S_1, S_2, S_3, \dots, S_{i-1} \rangle$ and stops at S_i . In the damage path each tile $O, S_1, S_2, S_3, \dots, S_{i-1}$ is knocked off the lattice except for S_i [Figure2(a)]. Let us denote the probability of the damage path that stops at S_i by $P(i)$. Observe that S_i is located at a Manhattan distance of i from O . Since F_1 is proportional to $\frac{1}{\sqrt{i}}$, the probability that a tile will fall off, given that at least one of its neighbors is already knocked off, is given by $\frac{p}{\sqrt{i}}$ where p ($0 < p < 1$) is the normalization factor and can be evaluated from the probability distribution for a damage path. Then,

$$P(i) = \frac{p^i}{(\sqrt{i}-1)!} \left(1 - \frac{p}{\sqrt{i}}\right) \quad (1)$$

This is because each tile on this path falls with a probability $\frac{p}{\sqrt{j}}$ where $j = 1, 2, \dots, i-1$ and the shock wave stops at S_i with probability $(1 - \frac{p}{\sqrt{i}})$. Furthermore, $P(i)$ is a probability mass function and hence, if the maximum length of our damage paths is restricted to l we have

$$\sum_{i=1}^l P(i) = 1 \quad (2)$$

Now we can estimate the expected fractional damage size $D(n, l)$ for a lattice L with n tiles, by summing the probabilities of a damage path from O to each S_i (maximum damage path length being l) in the lattice. For ease of computation, we achieve this by calculating the expected number of tiles knocked off at a Manhattan distance of i from O , $\forall i, 1 \dots l$.

$$D(n, l) = \frac{(1-p) + \sum_{i=1}^l 4i \times P(i)}{n} \quad (3)$$

The first term accounts for the event when the damage path probabilistically stops at O and $4i$ is the number of tiles located at a Manhattan distance i from O .

Simulation Results Based on equation 1 and 2, we can solve for p analytically. However, for large systems,

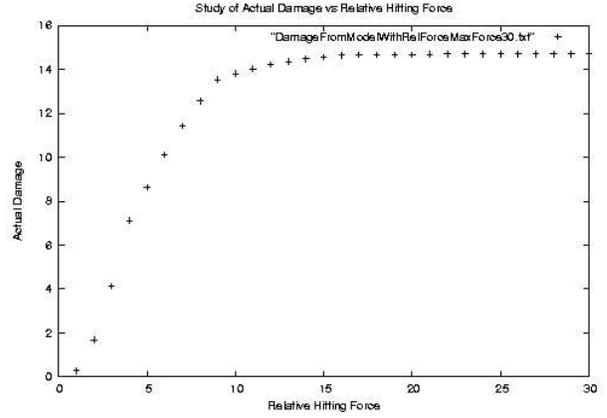


Figure 3: Plot of Actual Damage vs relative hitting force of the AFM. The plot reveals the pseudo-geometric nature of the probability distribution of damage.

the problem is too hard and hence we use Monte Carlo simulation to evaluate p [Figure2(b)]. Now our l is determined by the Manhattan distance where $F_1(r) = F_2$. Thus if $F_1(l) = \frac{c}{\sqrt{l}}$, where c is a constant, then

$$l = \left(\frac{c}{F_2}\right)^2 \quad (4)$$

We call l the *relative hitting force* in our simulation plot [Figure 3] since it approximately measures the ratio between the original impulse ($F_1(0)$) and the resistive binding force between a pair of tiles (F_2). We study the effect of relative hitting force on the lattice in isolation. As the plot in Figure2(b) reveals that the value of p stabilizes to 0.7316 beyond a relative hitting force of 10. Even the drop from an initial value of 0.87 to 0.75 occurs before the relative hitting force even reach a value of 5. Hence for all practical purposes we consider the value of p to be 0.73. Furthermore, the plot in Figure 3 verifies the pseudo-geometric probability distribution of a damage path described in equation 3. Evidently, the amount of actual damage reaches nearly a constant value beyond a relative hitting force of 10.

2.2 Flexible Lattice Model

This section considers a different model which views the rectangular DNA lattice of size $m \times n$ as a simple mass spring system similar to cloth dynamics in computer graphics [5]. In this model, each tile is positioned at grid point (i, j) , $i = 1, 2, \dots, m$ and $j = 1, 2, \dots, n$. For simplicity, assume that the external mechanical impulse F hits the lattice at a single tile location. The internal tension of the spring linking tile $T_{i,j}$ with each of its neighbor $T_{k,l}$ is given by

$$F_{int_{i,j}} = - \sum_{(k,l) \in R} K_{i,j,k,l} \left[l_{i,j,k,l} - l_{i,j,k,l}^0 \frac{l_{i,j,k,l}}{\|l_{i,j,k,l}\|} \right] \quad (5)$$

where R is the set of neighbors of $T_{i,j}$. $l_{i,j,k,l} = \overrightarrow{T_{i,j}T_{k,l}}$, $l_{i,j,k,l}^0$ is the natural length of the spring linking tiles and

$K_{i,j,k,l}$ is the stiffness of that spring. With DNA the actual value of $K_{i,j,k,l}$ is dependent on the type of bases that are involved in the sticky ends for the pair of tiles. We make some simplifying assumptions: the stiffness K and the unextended spring length l^0 are the same for every pair of adjacent tiles and known to us *a priori* for the purposes of simulation.

2.3 Simulation Algorithm

Since it is difficult to obtain a closed form solution for the expected number of tiles that are knocked off the lattice because of the impulse, we outline a simple simulation algorithm for estimating the number of tiles removed when an impulse hits the lattice. We assume that the lattice is a connected graph where the state of the tiles is updated every Δt time. To that extent, we start a breadth first search from the tile receiving the original impulse.

The tile under consideration moves for Δt time in the direction of the hitting force, thus extending the springs connecting it to its neighbors. If the extension is beyond a threshold value, the tile snaps off the lattice. In the next level of the breadth first search, a component of the spring force pulls each of the neighbor away from the lattice while a reaction force pushes the tile under consideration to restore its original configuration, if it has not fallen out of the lattice. In the following level, we examine spring extensions for the neighbors of the neighbors of the tile that received the initial impulse. The process is repeated recursively until we reach the boundary. We repeat the breadth first search on the tiles until the velocity of all the tiles are either zero or the whole lattice has fallen apart. The algorithm returns the number of tiles that are knocked off. A cartoon of the lattice before and after an impact is shown in Figure 4.

Some of the important assumptions we make are (1) the velocity of each tile remains constant during the Δt interval, (2) a damping force models collisions between tiles and individual water molecules.

The model approximately captures the overall behavior of a free floating lattice when acted upon by an external impulse. It, however, may be very sensitive to the parameter values such as the spring stiffness constant and the fracture threshold. We also note that this model does not consider collisions between tiles, which may change their relative velocity and, hence, affect the total number of tiles that actually are knocked off the lattice.

3 Thermal Damage Model

Suppose a lattice is being irradiated by a powerful electron beam that raises the local temperature of a part of the lattice beyond the melting temperature of

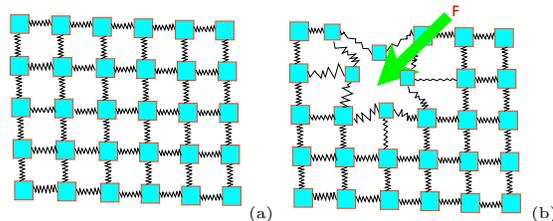


Figure 4: *Cartoon for Mechanical Damage in a Flexible Lattice: (a)Original Lattice in solution, (b)Damaged Lattice due to an external force. The tile that received the immediate impact F (not shown in the figure) is knocked off since one of the springs that connect it to its four neighbors extended beyond the threshold value while some of the neighbors are displaced (The affected tiles have smaller size to indicate that they are below the rest of the tiles in the lattice). However, the magnitude of the force is relatively low and the shock wave from the original impact dies long before it reaches the boundary as is evident from the relatively unchanged configuration of the tiles on the boundary.*

the tiles. The increased temperature will alter the dissociation rate of tiles in the region, since it is temperature dependent. This section develops a model for computing number of tiles removed due to this local temperature increase.

In Winfree's kATM model [4], if there are m empty sites adjacent to the aggregate, then the net "on rate" is given as:

$$k_{on} = m\hat{k}e^{-G_{mc}} \quad (6)$$

where G_{mc} is the entropic cost of fixing the location of a monomer unit while $\hat{k} = 20k_f$, k_f being the forward rate constant. For all occupied sites (i, j) within the aggregate, the net "off rate" is

$$k_{off} = \sum_b k_{off,b}, k_{off,b} = \sum_{ij \text{ s.t } b_{ij}=b} \hat{k}_f e^{-b_{ij}G_{se}} \quad (7)$$

where b_{ij} is the total strength for matching labels and $G_{se} = (\frac{4000K}{T} - 11s)$ is the free energy cost of breaking a single sticky end bond, with s being the sticky end length of the oligonucleotide and T being the temperature. In general, for a tile with b matches at a site with the forward rate of association as r_f and the rate of dissociation of a tile with b matches as $r_{r,b}$, we have

$$\frac{r_f}{r_{r,b}} = e^{bG_{se} - G_{mc}} \quad (8)$$

Since G_{se} is inversely proportional to T , if T increases such that $\frac{r_f}{r_{r,b}} < 1$, a hole is created. As an example of how temperature changes lead to reduced lattice, Figure 5 shows the effects of globally increasing temperature by $1.9^\circ C$ after the lattice is fully formed. Here starting with

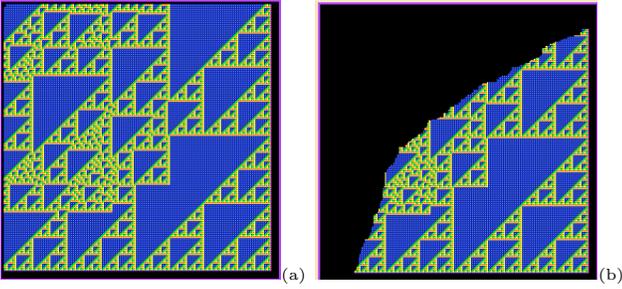


Figure 5: a) Original Lattice, (b) Damaged Lattice due to increased temperature

a Sierpinski Triangle patterned lattice with 62992 tiles assembled at a temperature of $35.969^{\circ}C$ with $G_{mc} = 19$ and $G_{se} = 9.7$, when heated to $37.892^{\circ}C$ ($G_{mc} = 19$, $G_{se} = 9.3$) for 2×10^5 sec reduces to 34585 tiles.

Although $\frac{r_f}{r_r, b} < 1$ gives us the condition for damage, we are more interested in designing a concrete model to compute the extent of damage. For this purpose we first compute the net rate of both binding and dissociation events k_{any} as

$$k_{any} = k_{on} + k_{off}. \quad (9)$$

In equilibrium, all the rates stabilize and the probability of any event E occurring is given by a Poisson distribution. Furthermore, the probability of an off event O happening (given that some event has happened), is given by $O \sim Binomial(N, k_{off})$ where N is the total number of events. If n is the size of the damage, we can compute the probability $P(O = n | E = i)$ of a hole of size n , given that a total of i events has occurred. Then the probability of dissociating n tiles is $P(O = n)$ where

$$P(O = n) = \sum_{i=0}^{\infty} P(O = n | E = i) P(E = i) \quad (10)$$

and the expected size of the damage in that case is

$$E(O) = \sum_{j=0}^M j P(O = j) \quad (11)$$

where M is the size of the initial aggregate. This equation is not easy to solve analytically but we can obtain an estimate for $E(O)$ with Monte Carlo integration. A more accurate estimate can be obtained if we do not assume that the rates k_{on} or k_{off} to be constant. According to kTAM, these rates vary with number of empty sites, concentration and total number of bonds which change as events occur.

4 Self-repair Model

Given a damage model, it is also useful to estimate the likelihood of the lattice reconstructing itself, a phenomenon we call *self-healing*. This can also be modeled

probabilistically. We estimate the probability of self-repair using techniques from Ref. [4]. We use the same notation as in Section 3.

Suppose an error-free lattice has a hole of size n to repair. We assume that self-healing is error-free too. Then using the principles of detailed balance, at equilibrium, it has been shown in Ref. [4] that an aggregate A formed by the addition of any sequence of n tiles T_1, T_2, \dots, T_n with a total strength $b_A = \sum_{i=1}^n b_i$. will obey the following equation:

$$\frac{[A]}{[T]} = e^{-((n-1)G_{mc} - b_A G_{se})} \quad (12)$$

where T denotes one of T_1, T_2, \dots, T_n . Furthermore, we will use A_m to denote a lattice with m errors (and A_0 is the error-free self-healed lattice). For small m , there can be $\binom{2n}{m}$ suboptimal aggregates for each perfect aggregate of size n and, hence, the probability of an errorless aggregate A_0 of size n , $P(A_0 = n)$, can be derived as :

$$P(A_0 = n) \sim \frac{[A_0]}{\sum_{m=0}^{\infty} \binom{2n}{m} [A_m]} \sim 1 - 2ne^{-G_{se}} \quad (13)$$

Hence, the probability that a damage of size n will get self-healed is $1 - 2ne^{-G_{se}}$.

5 Summary

This paper presents damage and self-repair models for self-assembled DNA nanostructures that would allow us to design error-resilient nanostructures in the future. As a part of future work, we intend to validate our models with experimental data.

REFERENCES

- [1] Erik Winfree and Renat Bekbolatov Proofreading tile sets: Error correction for algorithmic self-assembly *DNA 9*, LNCS 2943, pp:126-144.
- [2] J Xu and T. H LaBean and S.L Craig DNA Structures and their Applications in Nanotechnology *ed:A Ciferri*, CRC Press -Taylor & Francis Group, Boca Raton.
- [3] Hertzberg R Deformation and Fracture Mechanics of Engineering Materials *John Wiley and Sons*, NY 1996.
- [4] E Winfree Simulations of Computing by Self-Assembly *Caltech CS Report* 1998.22.
- [5] X Provot Deformation Constraints in a Mass Spring Model to Describe Rigid Cloth Behavior *In Proc. Graphics Interface*, 95, pgs:147-154, 1995.