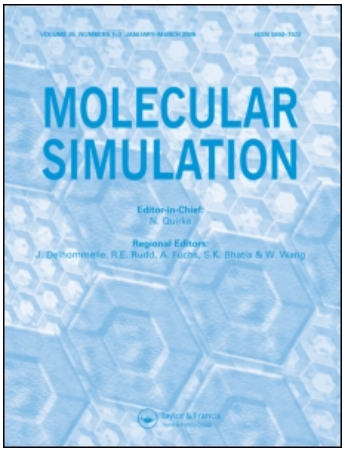

Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

Probabilistic models for damage and self-repair in DNA self-assembly

U. Majumder[a]
[a] Department of Computer Science, Duke University, Durham, NC, USA

First published on: 30 September 2009

To cite this Article Majumder, U.(2010) 'Probabilistic models for damage and self-repair in DNA self-assembly', Molecular Simulation, 36: 4, 255 — 266, First published on: 30 September 2009 (iFirst)
To link to this Article: DOI: 10.1080/08927020903277995
URL: http://dx.doi.org/10.1080/08927020903277995

Molecular Simulation
Vol. 36, No. 4, April 2010, 255–266

Taylor & Francis
Taylor & Francis Group

Probabilistic models for damage and self-repair in DNA self-assembly

U. Majumder*

Department of Computer Science, Duke University, Durham, NC, USA

(*Received 31 October 2008; final version received 18 August 2009*)

To date, research on 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 designing robust self-assembled nanostructures. This paper presents simple models for estimating the extent of damage in DNA nanostructures due to various external forces, with emphasis on mechanical and thermal forces. We note that these models have not been validated against experimental data. As such, they 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 assuming no change in the design of the self-assembling components.

Keywords: DNA; self-assembly; damage; self-repair; probability

1. Background

1.1 Damage and repair of DNA in nature

Damage in cellular DNA can occur as a result of a large number of normal metabolic activities, such as from free radicals produced during cellular respiration, as well as a variety of environmental factors, such as UV radiation and carcinogenic chemicals [1,2]. This damage can result in lesions that may prevent the cell from being able to transcribe its genes (resulting in a loss of functionality of the cell) or affect the daughter cells that are produced following mitosis. With respect to its structure, DNA damage can be in the form of (1) *covalent modification* of its bases (for instance, deamination of a nucleotide), (2) *base mismatches*, (3) *nicks* in the backbone (primarily caused by ionising radiation) and (4) *crosslinks* (covalent links between bases on the same strand or different strands). Fortunately, cells possess a number of DNA repair processes that are capable of limiting the extent of such damage. At the base level, damaged or inappropriate bases can be repaired by several mechanisms including (1) *direct chemical reversal* of the damage and (2) *excision repair* (where incorrect bases are locally removed and resynthesised). Nicks in the backbone can be healed using various DNA repair enzymes. At the cellular level, DNA repair processes include (1) *apoptosis*, a process by which the cell destroys itself, (2) *senescence*, a process whereby the cell becomes irreversibly dormant, and (3) *uncontrolled cellular division*, often resulting in the growth of tumours. Failure to repair DNA produces a mutation. In fact, the effectiveness of the DNA repair mechanisms in a species is directly related to the average and maximum life expectancy of that species.

1.2 Fragility of DNA nanostructures

DNA self-assembly is an autonomous phenomenon where components (single strands or DNA tiles, comprising of several single strands of DNA) organise themselves into stable superstructures. In recent years, DNA-based self-assembled nanostructures have gone from conceptual design to experimental reality [3], with potential applications in nanoelectronics [4] and nanomedicine [5].

Synthetic DNA used for building DNA self-assembled nanostructures is vulnerable to the same damaging forces as natural DNA. Synthetic DNA is frequently exposed to damaging forces during synthesis, resulting in improperly synthesised strands. These strands are typically removed using purification techniques such as PAGE before strands are mixed. Any damage that happens at the base level after the purification process (e.g. during the formation and characterisation of superstructures) is beyond our control. In fact, this paper does *not* focus on mitigating the damage at the DNA base level, but rather at the self-assembled superstructure level.

The end goal of DNA self-assembly is to have a 'good' yield of a desired nanostructure. Most DNA self-assembled superstructures (e.g. lattices) are composed of substructures (e.g. tiles), held together by sticky ends that are only a few bases long. Under the influence of external forces, such as mechanical and thermal, the lattices can easily dissociate into smaller lattices. This paper attempts to estimate the extent of such damage as a function of the

*Email: urmim@cs.duke.edu

ISSN 0892-7022 print/ISSN 1029-0435 online

DOI: 10.1080/08927020903277995
http://www.informaworld.com

applied force. The analysis will help us to characterise the sticky ends that can withstand the applied force and, consequently, guarantee higher yields of desired nanostructures. However, DNA nanostructures are inherently fragile, due to the nature of DNA. If damage cannot be completely eliminated, then it is useful to estimate the extent of a structure's ability to self-repair under the same physical conditions.

1.3 External forces that can affect a DNA nanostructure

A variety of forces can affect DNA nanostructures. One such force is the force from a cantilever tip while imaging with an atomic force microscope (AFM). We typically use the AFM in tapping mode in fluid to characterise most DNA nanostructures (Figure 1 (left)). AFM provides two modes. In contact mode, the AFM drags the tip across the surface as it images. This mode can greatly damage fragile nanostructures such as DNA lattices. The other mode, called tapping mode, images nanostructures by driving the tip to oscillate with an amplitude of 100–200 nm using a piezoelectric oscillator. As the tip strikes the surface, the AFM infers the height. Repeated tapping, however, can greatly damage the nanostructures being imaged. When the tip comes close to the surface, different kinds of forces act on the tip including van der Waals forces, dipole–dipole interaction, electrostatic forces and others, causing the amplitude of this oscillation to decrease. Using suitable scanning parameters, however, it is possible to leave the nanostructures unaffected for several hours of scanning. Nevertheless, the sample eventually disintegrates from repeated scanning, as shown in Figure 2.

Mechanical damage is also possible, even when the sample is not on a rigid surface. For instance, in solution, large lattices are often destroyed during sample preparation by actions such as pipetting. In general, the larger the size of the nanostructure, the more susceptible it is to denaturation due to shear forces applied while pipetting the sample for characterisation. It is possible to avoid such

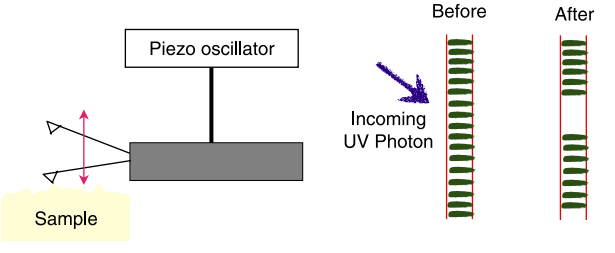

Figure 1. Left: cartoon of the AFM cantilever in tapping mode. Repeated tapping on the sample surface can eventually destroy lattices. Right: radiation-induced damage in DNA double helix.

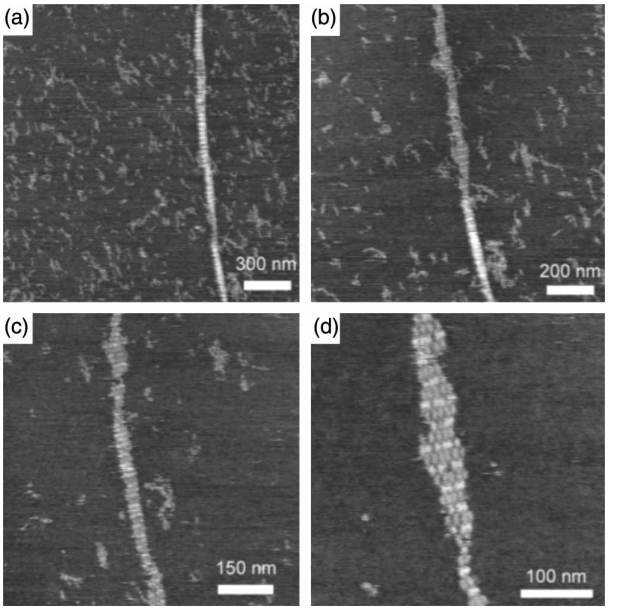

Figure 2. A series of AFM images captured by repeatedly imaging and zooming in on the same DNA nanotube, which appears mostly tube-like in (a), with increasing wear-and-tear through (b) and (c) until a section of unfolded tube becomes a single layer at lattice in (d). Used with permission from [28].

damage to a large extent by mixing or transferring the sample slowly, but it is not possible to mitigate this damage completely.

We have already discussed how radiation can damage DNA at the base level (Figure 1 (right)). At the superstructure level, however, intense radiation may create nicks in the backbone of the hybridised sticky ends, breaking the lattice into two or more smaller lattices. Radiation can also increase the temperature of part of the lattice, possibly denaturing the sticky ends that hold the tiles together. If the whole sample is exposed to radiation, there can be global temperature increases as well. This can lead to global denaturation.

1.4 Robust design of a DNA nanostructure

One of the goals of this paper is to improve the robustness of DNA nanostructure designs. By robust designs, we are not referring to changes in tile design or the tile set to improve its self-repairability. Since we are interested primarily in lattices, one of our goals is to quantify the relationship between the bond strength resistance to damage. The hope is that better designed sticky ends, both in terms of base sequences and length, would allow much higher yields. Given the importance of sequence design to DNA self-assembly in general, there already exist many tools for this purpose. See [6] for a survey of such tools. In particular, Pistol et al. [7] proposed a new thermodynamic optimisation tool, specifically for designing tile arm and

sticky end sequences, based on the tile design. The tool performs an exhaustive parallel thermodynamic search algorithm that reduces non-specific interactions during the assembly process. Furthermore, this algorithm evaluates each possible tile arm and sticky end sequence against all other candidate sequences and already chosen tile sequences, computing their mutual interaction. Empowered with such sequence design tools, we believe that a better understanding of the extent of damage would allow us to design better sticky ends.

1.5 Need for a damage and self-repair model for DNA nanostructures

As observed in Sections 1.2 and 1.3, DNA nanostructures can be very fragile. Thus, a model for estimating the extent of damage due to external forces would be very useful in improving both the designs of the overall nanostructures and total yields. Existing work in this area is very sparse. Winfree and Bekbolatov [8] applied a kinetic simulation model to investigate if a hole in a DNA lattice will repair correctly (Section 1.7). However, to the best of our knowledge, there has been no work modelling either the extent of damage in DNA nanostructures or in characterisation of damage. 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 sets. We study the extent of damage due to external mechanical and radiation forces. We further compute the probability of self-repair at equilibrium. Specifically, we study the effects of a mechanical force on a DNA lattice when it is on a rigid surface (for instance, when the sample is being imaged with an AFM). We also study the damage caused by mechanical forces on large DNA nanostructures in solution. We, moreover, consider damage caused by thermal forces, such as when part of the lattice is irradiated leading to local rise in temperature. Once we have estimated the extent of damage due to external forces, we attempt to calculate the extent of repair without any change in tile design or physical parameters. This calculation is useful for a more accurate estimate of the final yields of desired products. Note that, in each of the above-mentioned cases, the force may also denature the individual tiles. However, we are interested only in the size of the lattices and, hence, we ignore such destruction.

1.6 Correlation between the model and real damage and self-repair in DNA nanostructures

1.6.1 Correlation between the mechanical damage model and damage in reality

For our model of damage caused by a mechanical impact on a DNA lattice (both rigid and flexible), we assume that only one tile receives the initial impact. This assumption is not unreasonable since the size of a substructure or a tile can vary widely from one self-assembly experiment to another. Tile design can be controlled precisely and, hence, we can safely assume that it is possible to design a tile with dimensions such that the initial impact is received by one tile. Nevertheless, the key difference between the rigid lattice model and the flexible lattice model is that, in the rigid model, the lattice cannot move and, hence, the impact received by tiles in the lattice drops off pseudogeometrically from the tile that receives the initial impulse. In the flexible lattice model, however, since the sample is in solution, when a tile is hit from above, it can bend downwards. In other words, in this model, the lattice behaves like a mass–spring system, where the tiles are the masses and the interconnections between them represent the springs.

1.6.1.1 Rigid damage model vs. damage in reality. Our rigid damage model uses the example of an AFM tip hitting the sample adsorbed on a mica surface. Since our DNA lattices assemble in a buffer containing Mg^{2+} cations, once adsorbed onto mica, the interactions between the cations (Mg^{2+}) and counterions effectively bind the DNA to the mica surface, making it rigid. This phenomenon occurs at a fixed fractional surface density of the cations, which is the operating conditions for our experiments.[1] However, at higher ionic strength, the binding is weakened by the screening effect of the ions [9]. Since the lattice is rigid and the AFM tip is small, it is reasonable to assume that a single tile is struck.

1.6.1.2 Flexible damage model vs. damage in reality. For a flexible lattice, the assumption that the force hits only a single tile is much more non-obvious. However, we contend that large lattices are nearly stationary in solution, thus making it much more plausible to hit a single tile. The free-solution mobility of DNA in TAE buffer has been found to be $(3.75 \pm 0.04) \times 10^{-4}\,\mathrm{cm^2\,V^{-1}\,s^{-1}}$ at 25°C, independent of DNA concentration, sample size, electric field strength and capillary coating. The free-solution mobility was found to be independent of DNA molecular weight from approximately 400 bp to 48.5 kbp [10]. A single cross tile[2] [11] contains around 200 bp and hence falls in this range. However, cross tiles have been shown capable of forming micrometre-sized lattices. Supposing the typical size of such a lattice is $5\,\mu\mathrm{m} \times 5\,\mu\mathrm{m}$ and each cross tile is about $20\,\mathrm{nm} \times 20\,\mathrm{nm}$, then there are 6.25×10^4 cross tiles in such a lattice which amounts to about 13×10^6 kbp. To the best of our knowledge, there is no study on the solution mobility of DNA complexes of that size. However, performing everyday experiments

(such as a 5% native PAGE) on such a lattice shows that the lattice does not move quickly through the gel. When observed with fluorescence microscopy, researchers often observe that the lattices move quite slowly as well. These observations lead us to believe that it is possible for a localised area to receive the initial impulse. Intuitively, it seems reasonable for large lattices to have a low free-solution mobility. In our model, this localised area corresponds to a single tile. Another simplifying assumption we make is that the lattice is initially planar and only assumes a three-dimensional shape after the impact.

1.6.2 Correlation between the thermal damage model, self-repair model and damage and self-repair in reality

DNA damage induced by exposure to radiation is well studied [12,13]. In fact, the major kinds of heat-induced damage (e.g. UV exposure during gel analysis) to DNA have been shown to depend in their extent of oxygen content in solution [14]. These works primarily consider bases and backbone damage, whereas our focus is at the superstructure level. We model the effect of the increased temperature as an increased probability of dissociation of the interconnections among the affected tiles. Because of the increased temperature in the affected region, the binding energy of the sticky ends decreases, increasing the dissociation rate. The thermal damage model uses the same parameters as Winfree's kinetic Tile Assembly Model (kTAM) [15]. The model for self-repair also uses the same binding energy (G_{se}) and tile concentration (G_{mc}) parameters from Winfree's kTAM.

1.7 Previous work on damage of DNA nanostructures and their self-repair

As mentioned previously, Winfree and Bekbolatov [8] briefly investigated if a hole in a DNA lattice will repair correctly in the kTAM. In XGROW [15], it is possible to puncture a hole in a growing pattern and observe whether the pattern would regrow correctly. Figure 3 shows the growth of a Sierpinski pattern using 2×2 proofreading tiles [8] and $G_{mc} = 17$ and $G_{se} = 8.6$. Because of the choice of kinetic parameters, the initial growth of the lattice is error-free. However, when we puncture a hole, the pattern grows back with 36 mismatches. Puncturing again causes it to grow back with 10 mismatches in the kTAM.

1.8 Overview of the models

In this section, we present a brief overview of the models before describing the details in the next few sections. Suppose an AFM tip strikes a tile on a lattice composed of several DNA tiles. Before AFM imaging, DNA nanostructures are bound to mica. Hence, we can safely assume that our lattice is a rigid surface. Since the lattice is rigid, the effect of the hitting force on a neighbouring tile can be modelled as a function of its distance from the source of the impact [16]. In this model, which we call *rigid lattice* model, the probability that a tile will dissociate is given by the probability that a shock wave from the tile that receives the initial impact propagates to this tile along any path (Section 2). However, if we assume that the lattice is flexible (i.e. the lattice is in solution), we can model the lattice as a mass–spring system and, hence, directly calculate the effect of the hitting force on the lattice [17]. This model is called the *flexible lattice model* (Section 3). Finally, the *thermal damage model* and the *self-repair model* based on the kTAM are presented in Sections 4 and 5, respectively. One important observation that we should make here is that we are yet to verify these models against experimental data.

In the following sections, we describe the detailed models of damage and self-repair.

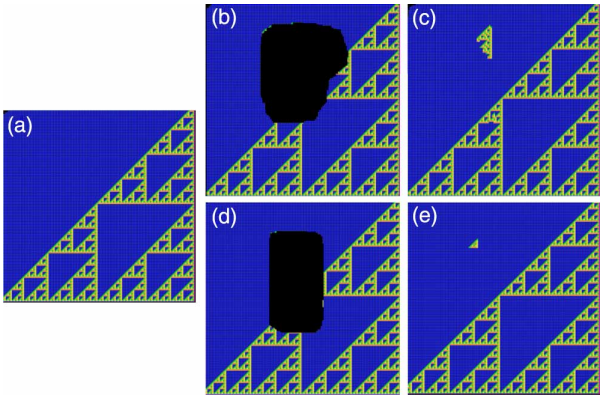

Figure 3. Study of damage and self-repair in XGROW: (a) growth of a Sierpinski pattern using 2×2 proofreading tiles and $G_{mc} = 17$ and $G_{se} = 8.6$, (b) punctured pattern, (c) healed pattern with same kinetic parameters (36 mismatches), (d) pattern punctured a second time and (e) healed pattern with 10 mismatches.

2. Model for mechanical impact on a rigid DNA lattice

2.1 The model

In the *rigid lattice model*, we assume that the force $F_1(t)$ on a tile t located at a distance of r from the impacted tile t^* is proportional to $1/\sqrt{r}$ [16]. In our model, r is the Manhattan distance between t and t^*. $F_2(t)$ is the resistive force from the sticky end connections of the tile t. For simplicity, we assume $F_2(t)$ to be the same for all the tiles t. One open question, how can we model damage when $F_2(t) \neq F_2(t')$ for two tiles t and t'. For any tile t with $F_1(t) > F_2(t)$, the probability that t is knocked off the lattice is greater than

zero so long as the shock wave has propagated from t^* to t. One of the primary assumptions in this model is that we only consider non-cyclic paths for the shock waves.

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 t^* (Figure 4) and meanders outwards through its successors $\langle t_1, t_2, t_3, \ldots, t_{i-1}\rangle$ and stops at t_i. In the damage path, each tile $t^*, t_1, t_2, \ldots, t_{i-1}$ is knocked off the lattice except for t_i. Let us denote the probability of the damage path that stops at t_i by $P(i)$. Observe that t_i is located at a Manhattan distance of i from t^* since i tiles are in the path. Since $F_1(t_i)$ is proportional to $1/\sqrt{i}$, the probability that t_i will fall off, given that at least one of its neighbours is already knocked off, is given by $\mathbf{c}/\sqrt{i}$, where $\mathbf{c}$ $(0 < \mathbf{c} < 1)$ is a normalisation factor and can be evaluated from the probability distribution for a damage path. Since each knock off event is assumed to be independent,

$$P(i) = \frac{\mathbf{c}^i}{\sqrt{(i-1)!}}\left(1 - \frac{\mathbf{c}}{\sqrt{i}}\right). \tag{1}$$

This is because each tile t_j on this path falls with a probability $\mathbf{c}/\sqrt{j}$, where $j = 1, 2, \ldots, i-1$, and the shock wave stops at t_i with probability $(1 - (\mathbf{c}/\sqrt{i}))$. Furthermore, $P(i)$ is a probability mass function and, hence, if the maximum length of our damage paths is restricted to l (assuming $l < n$ and $l < m$, where $m \times n$ is the number of tiles in the original rectangular lattice), we have

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

We will discuss in Section 2.3 how to obtain the value of l. We can now 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 t^* to each t_i (maximum damage path length being l) in the lattice.

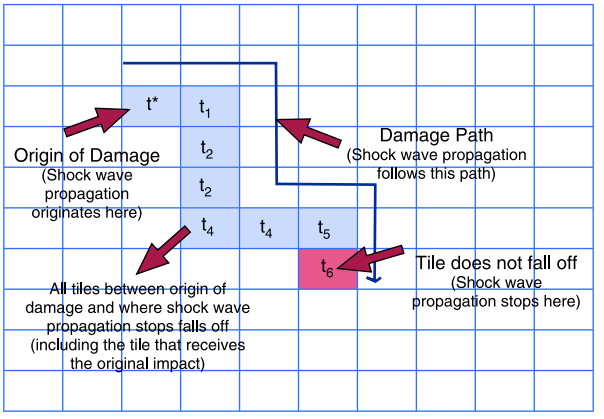

Figure 4. Instance of shock wave propagation and creation of damage path.

For ease of computation, we achieve this by calculating the expected number of tiles knocked off at a Manhattan distance of i from t^*, $\forall i, 1, \ldots, l,\ \ 1 \le i \le l$,

$$D(n, l) = \frac{(1 - \mathbf{c}) + \sum_{i=1}^{l} 4i \times P(i)}{n}. \tag{3}$$

The first term accounts for the event when the damage path probabilistically stops at t^* and $4i$ is the number of tiles located at a Manhattan distance i from O.[3]

2.2 Simulation algorithm

We evaluated our rigid lattice model using a simulation algorithm. The general idea of the algorithm is as follows: (1) we first initialise an array that contains the information whether a tile exists in location (i, j) in the lattice, (2) we apply the impulse to a randomly chosen tile, (3) tiles dissociate with probability $P(k)$, where k is the Manhattan distance from the origin of damage, and (4) if we succeed in knocking off a tile, we recursively knock off its neighbours based on the same probability distribution.

The pseudocode of this simulation algorithm is described below.[4] Assume that we are given the size of the rectangular lattice $X[1 \ldots m][1 \ldots n]$ and the value of $\mathbf{c}$ (obtained empirically from the value of relative hitting force, as discussed in Section 2.3):

```
Rigid-Lattice-Mechanical-Damage(X[][], c)
  ∀i, j   X[i][j] = 1.
  Randomly choose tile t receiving initial impact.
  Compute neighbour set N of t.
  KNOCK-OFF(X, t, N, c, 0).
  Output X
```

Below is the pseudocode for recursively knocking tiles off the lattice, given the lattice $X[m][n]$, tile t, its neighbour set N, initial probability of knock off (same as $\mathbf{c}$) and the tile's distance from the origin of damage d.

```
KNOCK-OFF(X[][], t, N(), c, d)
  if d > l
  return
  q = a random number between 0 and 1.
  Compute r = P(d, c).
  if q > r
    then return
    else
      (i, j) =index of t
      X[i][j] = 0
      for n ∈ N
        do
          compute neighbour set N' of n
          KNOCK-OFF(X, n, N', c, d + 1)
```

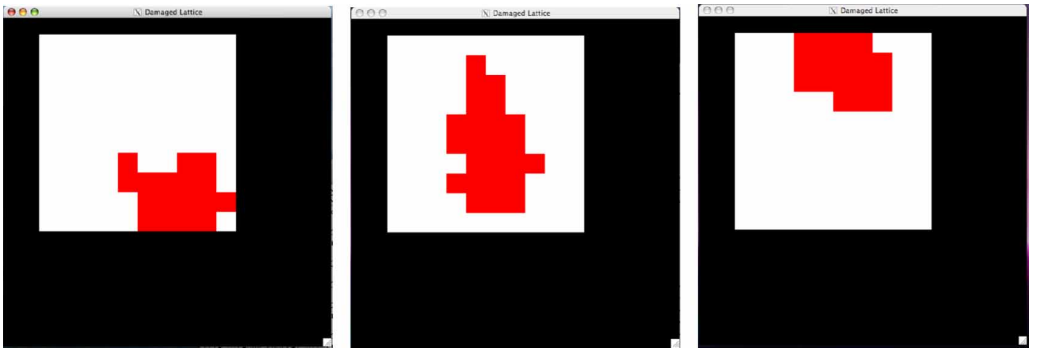

Figure 5. Snapshots of damaged lattice comprising of 100 tiles when hit by a mechanical force [initial probability of damage (same as **c**) = 0.87].

2.3 *Simulation results*

When we convert the pseudocode in Section 2.2 into actual code in C and OpenGL, the output from the simulation looks like the lattice in Figure 5. We started our simulation with a rectangular lattice of size 10×10 and $\mathbf{c} = 0.87$. The white region indicates what remains of the lattice after a randomly chosen tile is struck while the red region indicates the extent of the damage.

Based on Equations (1) and (2), we can solve for **c** analytically. However, for large systems, the problem is too hard and hence we use Monte Carlo simulation to evaluate **c** (Figure 6). Note that our l is determined by the Manhattan distance, where $F_1(t_l) = F_2(t_l)$. Thus, if $F_1(t_l) = F_1(t^*)/\sqrt{l}$, where $F_1(t^*)$ is the initial impact on tile t^*, then

$$l = \left(\frac{F_1(t^*)}{F_2(t_l)}\right)^2. \tag{4}$$

We call l the *relative hitting force* in our simulation plot (Figure 7) since it approximately measures the ratio between the original impulse ($F_1(t^*)$) and the resistive binding force for any tile t ($F_2(t)$) in the lattice. We study the effect of relative hitting force on the lattice in isolation. Given l, it is possible to compute **c**, the normalisation factor in the probability distribution of the damage. We calculate **c** via computer simulation. As the plot in Figure 6 reveals, the value of **c** asymptotically approaches 0.7316 beyond a relative hitting force of 10. The value of **c** drops from an initial value of 0.87–0.75 before l even reaches 5. Hence, for all practical purposes, we consider the value of **c** to be 0.73. Furthermore, the plot in Figure 7 verifies the pseudogeometric probability distribution of a damage path described in Equation (3).

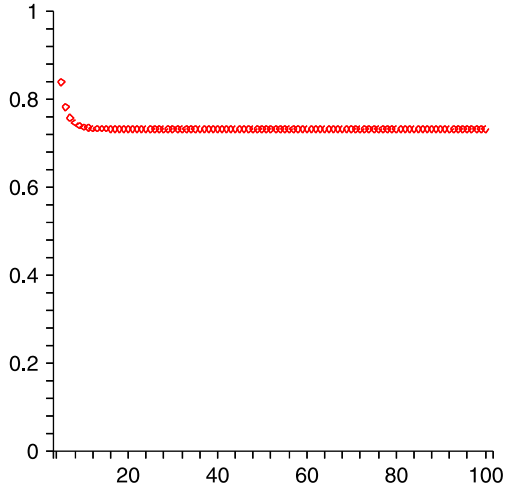

Figure 6. Estimation of **c**: values of **c** as a function of l.

3. Model for damage due to mechanical impact on a flexible lattice

3.1 *The 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 model in computer graphics [17]. We refer to this model as the *flexible damage model* (Figure 8). It is particularly applicable when the lattice is floating in solution. In this model, each tile is positioned at grid point (i, j), $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$. For simplicity, we 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 neighbour $T_{k,l}$ is given by

$$F\_\text{int}_{i,j} = -\sum_{(k,l)\in R} K_{i,j,k,l}\left[l_{i,j,k,l} - l^0_{i,j,k,l}\frac{l_{i,j,k,l}}{\|l_{i,j,k,l}\|}\right], \tag{5}$$

where R is the set of neighbours of $T_{i,j}$. $l_{i,j,k,l} = \overrightarrow{T_{i,j}T_{k,l}}$, $l^0_{i,j,k,l}$ is the natural length of the spring linking tiles at (i, j) to tiles at (k,l) and $K_{i,j,k,l}$ is the spring constant of that spring. With DNA, the actual value of $K_{i,j,k,l}$ is highly dependent on the specific bases in the sticky ends binding the pair of tiles. We make some further simplifying assumptions. The stiffness K and the unextended spring

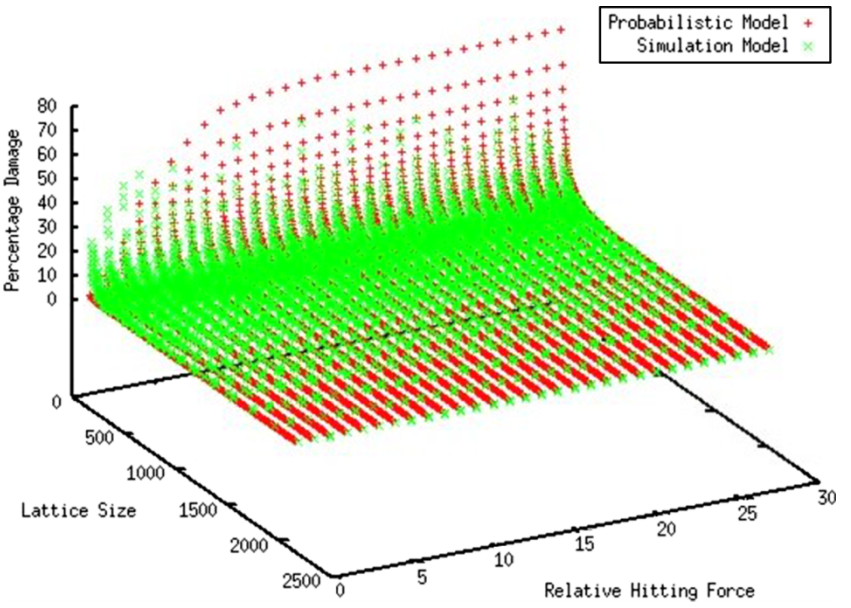

Figure 7. Plot of percentage damage vs. relative hitting force of the AFM and lattice size both calculated analytically as well as via simulation. The plot reveals the pseudogeometric nature of the probability distribution of damage. Furthermore, the analytical and simulation values seem to be in good agreement.

length l^0 are the same for every pair of adjacent tiles and known to us a *priori*.

3.2 *Simulation algorithm*

As with the rigid lattice model, we evaluated our flexible lattice model in simulation. Our simulation uses Open Dynamics Engine (ODE) [18], a free industrial quality library for simulating articulated, rigid body dynamics in virtual reality environments. An articulated structure is created when rigid bodies are connected together with joints of various kinds. Our simulation uses various features of ODE including built-in collision detection features, stable integrator (so that simulation errors do not grow out of control) and hard contacts (meaning that when two bodies collide they do not penetrate). The equations of motion are derived from a Lagrange-multiplier[5] velocity-based model. ODE uses a first-order integrator to compute positions and velocities and the standard 'big matrix method' for stepping. In our case, we model our tiles as rectangular-shaped rigid bodies connected by slider joints. A slider joint allows one object to slide (relative to another object). This joint is modelled as a spring connecting two objects and the library provides methods to model the spring joint.

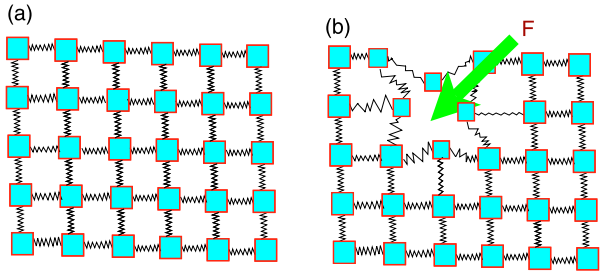

Figure 8. Cartoon for mechanical damage in a flexible lattice: (a) original lattice in solution and (b) damaged lattice due to an external force. The tile that received the immediate impact F (not shown in the figure) is knocked off since the springs that connect it to its four neighbours extended beyond the threshold value. Some of the neighbours are displaced in the process as well. The affected tiles have smaller size to indicate that they are below the rest of the tiles in the lattice. However, in this example, 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 general idea of the algorithm is as follows: (1) we create tiles as rigid bodies and sticky end connections as slider joints and initialise their positions, (2) next, we apply the external impulse on a randomly chosen tile and (3) we simulate the system for a fixed amount of time applying a water-damping force and spring force on each tile. Whenever the extension between a tile and its neighbour is beyond a predefined threshold value, we destroy the slider joint between them (in other words the tiles dissociate). (4) Finally, we output the extent of damage.

Below is the simulation algorithm described in the form of a pseudocode. We assume that several values are given to us, such as *ext_force*, the value of the external

force that is applied to the lattice, *threshold_dist*, the fixed distance for tiles to remain associated, *size_lattice*, the size of the lattice, *sim_time*, the total simulation time and *time step*, the length of each time step. Furthermore, we define a neighbour of a tile as a tile bound to the former through sticky end hybridisation.

```
Flexible-Lattice-Mechanical-Damage()
  //initialisation step
  for i ← 1 to size_lattice
    do
        create tile[i] as a rigid body
        initialise position of tile[i]
  for i ← 1 to size_lattice
    do
      N[i] =indices of neighbours of tile[i]
      for each n ∈ N[i]
        do
          create binding as slider joint
          s(i, n) between tile[i] and tile[n]
          initialise position of s(i, n)

  //impact step
  choose a random tile index r
  apply ext_force to tile[r]
  steps = 0
  while steps · time_step ≤ sim_time
    do
      SIMULATE-LATTICE(tile, N, s)
      steps = steps + 1
  damage = CALCULATE-DAMAGE(tile, N)
  output damage
```

Next, we describe the *Simulate-Lattice* and *Calculate-Damage* routines, which simulate the lattice dynamics for a time step and compute the final amount of damage, respectively.

```
SIMULATE-LATTICE(tile[], N[], s())
  for i ← 1 to size_lattice
    do
      apply water damping force to tile[i]
      based on its velocity
      APPLY-SPRING-FORCE(tile, N[i], s)
```

The next pseudocode is the most crucial part of the algorithm, since it is here we determine the extent of the damage based on the extension of the hybridised sticky ends between a tile and its neighbour.

```
APPLY-SPRING-FORCE(tile[], N[i], s())
  for each n ∈ N
    do
      if extension(s(i, n))>threshold_dist
        then
          remove s(i,n)
          update N[i]
        else
          apply spring force to tile[i]
```

The following pseudocode utilises a depth-first strategy to find the largest set of connected tiles. In general, this corresponds to the largest part of the remaining lattice and the damage can be easily computed from this value.

```
CALCULATE-DAMAGE(tile[], N')
  for i ← 1 to size_lattice
    do
      rt[i] = count the number of tiles
      reachable from tile[i]

  max_rem_lat = max{ rt[1], ..., rt[size_lattice]}
  damage = size_lattice − max_rem_lat
  return damage
```

3.3 Choice of simulation parameters

The lattice that we simulated originated from the self-assembly of DNA cross tiles [11]. Typically, a cross tile is about $20 \times 20 \times 0.34\,\text{nm}^3$ and comprises of around 400 nucleotides, with a mass of around $132 \times 10^3\,\text{D} = 2.19 \times 10^{-22}\,\text{kg}$. This provides us with the dimension and the mass of each rigid body in the simulation. Cross tiles have been used to create a wide variety of DNA nanostructures over a wide length scale from small completely addressable lattices (in nanometre scale) [19] to large periodic structures in the millimetre scale [20]. However, most of the lattices observed under an atomic force microscopy lie in the micrometre range. Hence, we simulate lattices ranging in sizes from 200 to 4000 tiles. Force spectroscopy and other methods that can manipulate objects at submicrometre scale tells us that the range of force these structures experience is very small, generally in the 100 s of pN [21]. On the other hand, the threshold distance is a function of the length of sticky end, given that we use the worm-like chain (WLC) model [22]. This model depicts that we can extend a double-stranded DNA to 1.7 times its contour length before it unwinds [23]. Assuming a sticky end length of four bases, this threshold distance evaluates to 0.95 nm. Furthermore, since a pair of adjacent tiles is held by only 4 bp, we exert a force in the femto Newton range. Additionally, assuming the stretch modulus S of double-stranded DNA to be 650 pN [24] and the unextended sticky end length L to be 1.36 nm, the spring constant $K = S/L = 0.365\,\text{N/m}$. The damping constant, C, of water can be obtained from Stokes' law [25]. Hence, $C = 6\pi\eta r$ for small objects with radius in the micrometre range. Our lattice falls in that category and, hence, we can use this equation. Assuming the dynamic viscosity η of water to be $10^{-3}\,\text{Pa s}$ and diameter $2r$ of a

typical DNA lattice made of cross tiles to be 5 μm, $C = 47.1 \times 10^{-9}\,\mathrm{N\,s\,m^{-1}}$.

3.4 Simulation results

A snapshot of the ODE simulation with 25 tiles is shown in Figure 9. This figure shows the initial lattice as well as its gradual disintegration. Moreover, we run the simulation over a force range of 0.1–250 fN, for lattice sizes ranging between 200 and 4000. We observe a linear increase in the extent of damage as we increase the force and lattice size until we reach a critical force value for every lattice size when the whole lattice disintegrates (Figure 10).

4. Thermal damage model

4.1 The model

Mechanical force is hardly the only force capable of destroying a DNA nanostructure. Suppose, for example, a lattice is irradiated by a powerful electron beam that raises the local temperature of a part of the lattice beyond the melting temperature of tiles. The increased temperature will alter the dissociation rate of tiles in the region, since the rate is highly temperature dependent. This section develops a model for computing the number of tiles removed due to this local temperature increase.

According to Winfree's kTAM [15], recall that if there are m empty sites adjacent to the aggregate, then the net 'on rate' is given as

$$k_{\mathrm{on}} = m\hat{k}\,\mathrm{e}^{-G_{\mathrm{mc}}}, \tag{6}$$

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

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

where b_{ij} is the total strength for matching labels and $G_{\mathrm{se}} = ((4000K/T) - 11)s$ 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

$$\frac{r_f}{r_{r,b}} = \mathrm{e}^{bG_{\mathrm{se}} - G_{\mathrm{mc}}}. \tag{8}$$

Since G_{se} is inversely proportional to T, if T increases such that $(r_f/r_{r,b}) < 1$, a hole will gradually form. As an example of how temperature changes lead to reduced lattices, Figure 11 shows a simulation of globally increasing the temperature by 1.9°C after the lattice is fully formed.[6] Here, starting with a Sierpinski triangle patterned lattice with 62,992 tiles assembled at a temperature of 35.969°C with $G_{\mathrm{mc}} = 19$ and $G_{\mathrm{se}} = 9.7$, heating to 37.982°C ($G_{\mathrm{mc}} = 19$, $G_{\mathrm{se}} = 9.3$) for 2×10^5 s reduces the lattice to 34,585 tiles. The above simulation has been performed with XGROW [15].

Although $(r_f/r_{r,b}) < 1$ gives us a condition for damage from temperature changes, we are primarily interested in designing a concrete model to compute the extent of damage. Hence, if we increase the temperature of a $n \times n$ completely internal region within a $m \times m$-sized lattice $(n < m)$ from T to T', then the new binding energy parameter G'_{se} is $((4000K/T') - 11)s$ for the $n \times n$ region. Suppose all the tiles in the lattice have equal binding strength. Binding sites for affected region are internal by assumption (i.e. not on the lattice boundary). Hence, each tile belonging to this region has four neighbours and there are n^2 such tiles. The number of tiles in the non-affected region is $2(m^2 - n^2)$, with $2(m + n) - 4$ tiles with three binding sites, four corner tiles with two binding sites and while the remaining with four binding sites. Consequently, the off rate in the affected region is

$$k_{\mathrm{off}_{\mathrm{aff}}} = \hat{k}_f n^2\,\mathrm{e}^{-4G'_{\mathrm{se}}}, \tag{9}$$

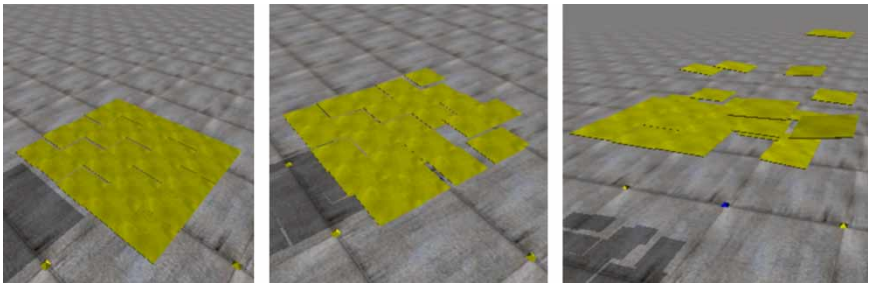

Figure 9. ODE simulation of mechanical damage in a flexible lattice: three screenshots showing the intact lattice that subsequently falls apart. This figure is best viewed in colour.

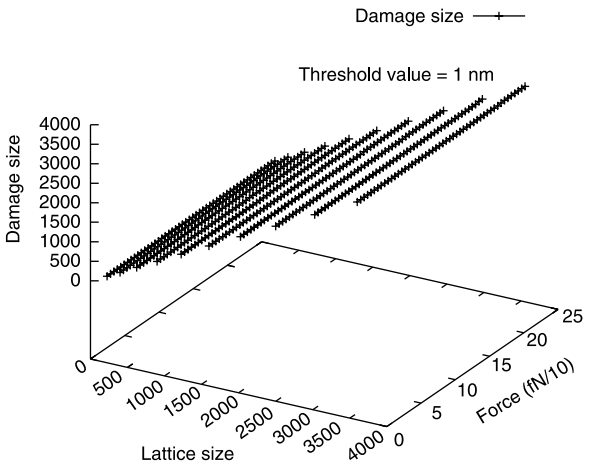

Figure 10. Damage versus flexible lattice size and exerted external force in fN/10.

and the off rate in the unaffected region is

$$k_{\text{off}_{\text{unaff}}} = \hat{k}_{\text{f}}\big((m^2 - n^2 - 2m - 2n - 4)\text{e}^{-4G_{\text{se}}} + 4\,\text{e}^{-2G_{\text{se}}} + \big(2m + 2n - 4\big)\text{e}^{-3G_{\text{se}}}\big). \tag{10}$$

Next, we compute the net rate of both binding and dissociation events k_{any} as

$$k_{\text{any}} = k_{\text{on}} + k_{\text{off}_{\text{unaff}}} + k_{\text{off}_{\text{aff}}}. \tag{11}$$

In equilibrium, all the rates stabilise and the probability of any event E occurring is given by a Poisson distribution. Hence, the probability of a total of i on and off events from both affected and unaffected regions happening is

$$P(E = i) = \frac{k_{\text{any}}^{i} e^{-k_{\text{any}}}}{i!}. \tag{12}$$

Furthermore, the probability of an off event O from the affected region happening (given that some event has happened), is given by $O \sim \text{Multinomial}(i, k_{\text{on}}, k_{\text{off}_{\text{unaff}}}, k_{\text{off}_{\text{aff}}})$, where N is the total number of events. Hence, the probability that n ($n \leq i$) tiles dissociate from the affected site (given that i events have occurred) is given by$P(O = n|E = i) = P$,

$$P = \sum_{j=0}^{i-n} \frac{i!}{n!j!(i-n-j)!}(x)^n(y)^j(z)^{i-n-j}, \tag{13}$$

where

$$x = \frac{k_{\text{off}_{\text{aff}}}}{k_{\text{any}}}, \quad y = \frac{k_{\text{off}_{\text{unaff}}}}{k_{\text{any}}} \quad \text{and} \quad z = \frac{k_{\text{on}}}{k_{\text{any}}}.$$

Then, the probability of n tiles dissociating is $P(O = n)$, where

$$P(O = n) = \sum_{i=n}^{\infty} P(O = n|E = i)P(E = i), \tag{14}$$

and the expected size of the damage is given by

$$E(O) = \sum_{j=0}^{M} jP(O = j), \tag{15}$$

where $M = n \times n$ is the size of the affected region. This equation is not easy to solve analytically but we can obtain an estimate for $E(O)$ using Monte Carlo integration.

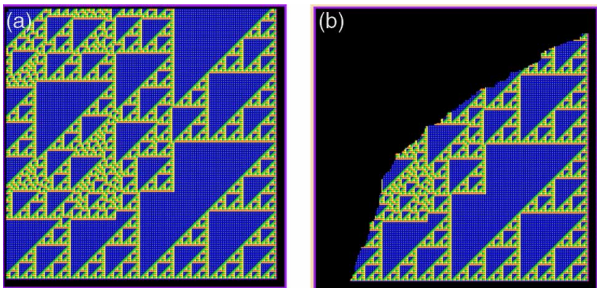

Figure 11. (a) Original lattice and (b) damaged lattice due to increased temperature (simulations performed with XGROW, a tile assembly simulator [15]).

4.2 Simulation algorithm

We first describe the basic algorithm in words. For every damage size ranging from 1 to the size of the affected region, we generate a random number between 0 and 1 and decide if any of the two off events or an on event will take place at this moment. If it does say that an off event will take place, then we generate another random number between 0 and 1 and decide whether that event will be knocking off a tile from the affected region. If it is, then we keep track of how many such events occur for each damage size. Finally, we compute the average size of the damage given those frequency values.

The simulation algorithm used to calculate the extent of damage is described in the form of a pseudocode below. We assume that we are given the size of lattice M, the size of irradiated region X, the increased temperature T', G_{se}, G_{mc}, the total number of simulation steps *sim_steps* that and a very large number (which we call *INFTY*) that we use to simulate a very large number of events. Furthermore, assume that the *rand*() function computes a random number between 0 and 1, *dam* is an array that stores the frequency for each irradiated region length and $p(j,i)$ denotes the probability $P(O = j|E = i)$.

SIMULATE-THERMAL-DAMAGE(M, X, T)
  $\forall i$ initialize $dam[i] = 0$
  Calculate k_{on}, $k_{\text{off}_{\text{aff}}}$, $k_{\text{off}_{\text{unaff}}}$ and k_{any}.

for $j \leftarrow 1$ **to** X
  do for $i \leftarrow j$ **to** *INFTY*
    do for *steps* $\leftarrow 1$ **to** *sim_steps*
      do $q = rand()$
        if $q \leq P(E = i)$
          then
            $r = rand()$
            if $r \leq p(j, i)$
              then
                $dam[j]$
                $= dam[j]$
                $+1$

compute $av\_dam = \frac{\sum_{i=0}^{X} i \cdot dam[i]}{\sum_{i=0}^{X} dam[i]}$
output *av_dam*

4.3 Simulation results

We computed the average damage for a lattice comprising 2500 tiles, where the intense radiation affected only a central region of 100 tiles. The radiation locally increased the temperature of that region from 33 to 65°C. For the various parameters, we used $G_{mc} = 19$, number of simulation steps *sim_steps* $= 1000$, total number of events $= 1000$, sticky end length $= 5$, $\hat{k}_f = 600,000$; we obtained $k_{on} = 0.672336$, $k_{off_{aff}} = 7.583902$, $k_{off_{unaff}} = 0.009040$ and $k_{any} = 8.265278$. Hence, the on probability p_{on}, the off probability for the affected region $p_{off_{aff}}$ and the off probability for the unaffected region $p_{off_{unaff}}$ are $p_{on} = 0.081345$, $p_{off_{aff}} = 0.917562$ and $p_{off_{unaff}} = 0.001094$, respectively. The expected damage was 5 out of 100 affected tiles. We can also calculate damage as a function of the size of the affected region and increased temperature. A more accurate estimate can be obtained if we do not assume that the rates k_{on} or k_{off} are constant. In reality, these rates vary based on the number of empty sites, concentration and total number of bonds, all of which change as events occur.

5. Self-repair model

As mentioned earlier, Winfree and Bekbolatov [8] investigated if a hole in a lattice would repair correctly in the kinetic simulation model. There is also some work by the same group on how to convert a given tile set into a self-healing tile set. Essentially, with a new redundancy-based tile design, if the remaining tiles in a damaged lattice remain connected, then the lattice will self-heal [26]. Unfortunately, this technique increases the tile set size by a multiplicative factor. Majumder et al. [27] proposes a compact self-healing tile set capable of repairing any damage to computations which are *reversible*. In this paper, however, we are concerned with the extent of repair without any change in design. In other words, given a damage model, we would like to estimate the likelihood of the lattice reconstructing itself. This can also be modelled probabilistically. We estimate the probability of self-repair using the kTAM just as we did with the thermal damage model [15]. Hence, we use the same notation as in Section 4; thus, the free parameters in the model are the size of the lattice (number of tiles), binding strength of individual tile type, the tile concentration parameter (G_{mc}) and the free energy of binding parameter (G_{se}).

Suppose an error-free lattice has a hole of size n to repair. Then, using the principles of detailed balance, it has been shown in [15] that, at equilibrium, an aggregate A formed by the addition of any sequence of n tiles $T_1, T_2, \ldots, T_n$, with a total strength $b_A = \sum_{i=1}^{n} b_i$, will obey the following equation:

$$\frac{[A]}{[T]} = e^{-((n-1)G_{mc} - b_A G_{se})}, \tag{16}$$

where T denotes one of $T_1, T_2, \ldots, T_n$. Furthermore, we use the notation 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 error-less aggregate A_0 of size n, $P(A_0 = n)$, can be derived as:

$$P(A_0 = n) \sim \frac{[A_0]}{\sum_m \binom{2n}{m} [A_m]} \sim 1 - 2ne^{-G_{se}}. \tag{17}$$

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

6. Summary

This paper presents the damage and self-repair models for self-assembled DNA nanostructures. These models will help to characterise damage as well as design robust nanostructures in the future. These models apply when a DNA lattice is hit by either mechanical force or exposed to intense radiation. For mechanical damage, we proposed models for when the lattice is rigid (e.g. when the lattice is bound to mica) and when the lattice is flexible (e.g. in solution). In both cases, we observed that damage increases with force, until a critical value of force is reached when the whole lattice dissociates. Furthermore, we showed how to calculate the extent of damage (in a kinetic setting) due to local increases in temperature. Finally, we also estimated the extent of repair that may take place in the damaged lattice without any change in tile design, assuming that assembly follows the kTAM. As a part of future work, we intend to validate our models with experimental data.

Notes

1. In our laboratory, we typically use a Mg^{2+} concentration of 12.5 mM in our TAE buffers.
2. This is the tile we frequently use in our own laboratory.
3. We assume in this equation that none of the boundary tiles receive the initial impact.
4. In the computer simulation of the rigid lattice model, we allow both tiles on the boundary to be hit as well as cyclic shock wave propagation paths.
5. The method of Lagrange multipliers provides a strategy for finding the maximum/minimum of a function subject to constraints.
6. We assume no net force from the electron beam.

References

[1] H. Lodish, A. Berk, P. Matsudaira, C.A. Kaiser, M. Krieger, M.P. Scott, S.L. Zipursky, and J. Darnell, *Molecular Biology of the Cell*, 5 ed., WH Freeman, New York, NY, 2004, pp. 1–693.

[2] W.S. Browner, A.J. Kahn, E. Ziv, A.P. Reiner, J. Oshima, R.M. Cawthon, W.C. Hsueh, and S.R. Cummings, *The genetics of human longevity*, Am. J. Med. 117(11) (2004), pp. 851–860.

[3] J. Xu, T.H. LaBean, and S.L. Craig, *DNA Structures are their Applications in Nanotechnology*, CRC Press/Taylor & Francis Group, Boca Raton, 2005.

[4] M.N. Stojanovic, D. Stefanovic, T.H. LaBean, and H. Yan, *Computing with nucleic acids*, in *Bioelectronics: From Theory to Applications*, I. Willner and E. Katz, eds., Wiley-VCH, Weinheim, 2005, pp. 427–455.

[5] Y. Benenson, T. Paz-Elizur, R. Adar, E. Keinan, Z. Livneh, and E. Shapiro, *Programmable and autonomous computing machine made of biomolecules*, Nature 414 (2001), pp. 430–434.

[6] A. Brenneman and A. Condon, *Strand design for bio-molecular computation*, Theoret. Comput. Sci. 287(1) (2002), pp. 39–58.

[7] C. Pistol, A.R. LeBeck, and C. Dwyer, *Design automation for DNA self-assembled nanostructures*, Proc. 43rd Design Automation Conf. (DAC 06), ACM Press, 2006, pp. 919–924.

[8] E. Winfree and R. Bekbolatov, *Proofreading tile sets: Error correction for algorithmic self-assembly*, DNA 9, LNCS 2943 (2003), pp. 126–144.

[9] D. Pastre, O. Piétrement, S. Fusil, F. Landousy, J. Jeusset, M. David, L. Hamon, E.L. Cam, and A. Zozime, *Adsorption of DNA to mica mediated by divalent counterions: A theoretical and experimental study*, Biophys. J. 85(4) (2003), pp. 2507–2518.

[10] N.C. Stellwagen, C. Gelfi, and P.G. Righetti, *The free solution mobility of DNA*, Biopolymers 42(6) (1997), pp. 687–703.

[11] H. Yan, S.H. Park, G. Finkelstein, J.H. Reif, and T.H. LaBean, *DNA-templated self-assembly of protein arrays and highly conductive nanowire*, Science 301(5641) (2003), pp. 1882–1884.

[12] R. Roots, G. Kraft, and E. Grooschalk, *The formation of radiation induced DNA breaks: The ratio of double-strand breaks to single strand breaks*, Int. J. Radiat. Oncol. Biol. Phys. 11 (1985), pp. 259–265.

[13] M.O. Bradley and K.W. Kohn, *X-ray induced DNA double strand break production and repair in mammalian cells as measured by neutral filter elution*, Nucleic Acids Res. 7 (1979), pp. 793–804.

[14] A.V. Chernikov, S.V. Gudkov, I.N. Shtarkman, and V.I. Bruskov, *Oxygen effect in heat induced DNA damage*, Mol. Biophys. 52(2) (2007), pp. 185–190.

[15] E. Winfree, *Simulation of computing by self-assembly*, Tech. Rep. 1998.22, Caltech, 1998.

[16] R. Hertzberg, *Deformation and Fracture Mechanics of Engineering Materials*, John Wiley and Sons, New York, NY, 1996.

[17] X. Provot, *Deformation constraints in a mass spring model to describe rigid cloth behavior*, Proc. Graph. Interf. 95 (1995), pp. 147–154.

[18] R. Smith, *Open dynamics engine*. Available at http://www.ode.org

[19] S.H. Park, C. Pistol, S.J. Ahn, J.H. Reif, A.R. Lebeck, C. Dwyer, and T.H. LaBean, *Finite-size, fully-addressable DNA tile lattices formed by hierarchical assembly procedures*, Angew. Chem. 45 (2006), pp. 735–739.

[20] Y. He, Y. Tian, Y. Chen, Z. Deng, A.E. Ribbe, and C. Mao, *Sequence symmetry as a tool for designing DNA nanostructures*, Angew. Chem. Int. Ed. 44 (2005), pp. 6694–6696.

[21] M. Salomo, K. Kegler, G. Gutsche, M. Struhalla, J. Reinmuth, W. Skokow, U. Hahn, and E. Kremer, *The elastic properties of single double stranded DNA chains of different lengths as measured with optical tweezers*, Colloid Polym. Sci. 284 (2006), pp. 1325–1331.

[22] J.F. Marko and E.D. Siggia, *Stretching DNA*, Macromolecules 28 (1995), pp. 8759–8770.

[23] P. Lai and Z. Zhou, *Stretching a double-stranded DNA: Elongation by untwisting*, Chin. J. Phys. 45(5) (2002), pp. 465–470.

[24] S.B. Smith, Y. Cui, and C. Bustamante, *Overstretching B-DNA: The elastic response of individual double-stranded and single-stranded DNA molecules*, Science 271 (1996), pp. 795–799.

[25] G.K. Batchelor, *An Introduction to Fluid Dynamics*, Cambridge University Press, Cambridge, 1967.

[26] E. Winfree, *Self-healing tile set*, *Nanotechnology: Science and Computation*, Springer, Berlin, 2006, pp. 55–78.

[27] U. Majumder, S. Sahu, T. LaBean, and J.H. Reif, *Design and simulation of self-repairing DNA lattices*, LNCS, also in the Proc. DNA 12 (2006), 4287 (2007), pp. 195–214.

[28] D. Liu, S.H. Park, J.H. Reif, and T.H. LaBean, *DNA nanotubes self-assembled from tx tiles as templates for conductive nanowires*, Proc. Natl Acad. Sci 101 (2004), pp. 717–722.