

THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF PHYSICS

TOPOLOGICAL FEATURES IN CELLULOSE MICROFIBRILS STUDIED USING
COARSE-GRAINED MOLECULAR DYNAMICS

FADIL ABBAS ALI
SPRING 2015

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Physics
with honors in Physics

Reviewed and approved* by the following:

Vincent Crespi
Professor of Physics
Thesis Supervisor

Richard Robinett
Professor of Physics
Honors Adviser

* Signatures are on file in the Schreyer Honors College.

ABSTRACT

Cellulose is the most abundant organic polymer in the biosphere, yet its growth mechanism remains misunderstood. Compact and crystalline fibrils typically consist of anywhere in from eighteen to thirty individual glucan chains. After these strands have coalesced, interesting macroscopic features emerge, including twists and bends. This project seeks to determine how these features are related to the fibril's overall stability, by implementing a coarse-grained model to see how their prevalence and behavior varies with fibril diameter. This model captures the anisotropic intermolecular interactions by modeling glucose rings as beads and glycosidic linkages as springs. These simulations revealed interesting qualitative information regarding a fibril's stiffness, stimulating curiosity of how persistence length varies with fibril diameter. An interesting feature noticed from these simulations is that fibril bends may occur at regions where the helicity along the fibril changes abruptly; we are searching to characterize the correlations between these two phenomena. By understanding how these phenomena vary with fibril diameter, we can achieve a better understanding of how the preferred macroscopic structure affects the cellulose fibrils' stiffness, which gives insight on which structures yield greater stability.

TABLE OF CONTENTS

LIST OF FIGURES	iii
ACKNOWLEDGEMENTS	v
Chapter 1 Background Information	1
1.1 Coarse-Grained Model and its Implementation	2
Chapter 2 Algorithms.....	7
2.1 Cellulose Backbone.....	7
2.2 Plane Projections	11
2.3 Degree of Helicity	13
2.4 Fibril Curvature.....	15
2.5 Persistence Length	15
Chapter 3 Procedure.....	20
Chapter 4 Data and Discussion	22
Chapter 5 Conclusion and Future Work	28
BIBLIOGRAPHY.....	30

LIST OF FIGURES

Figure 1 Chain separation R_{ij} along with its projections on the involved chains. The arrows in this figure display the local directions of each chain.	4
Figure 2 A step-by-step illustration of how each bead's region is determined. When evaluating which beads of a particular strand is below a given plane, each bead on that strand is evaluated using the plane equation until a positive value results. The bead with the positive value will be marked and evaluation for beads on the next strand begin. The process repeats. Evaluation for the next adjacent plane begins with the beads marked on all of the strands from the previous iterations.	9
Figure 3 A twelve stranded microfibril with its backbone.....	11
Figure 4 For the twenty-four stranded unbounded system: 9A-Raw helicity data. 9B- Raw curvature data. 9C- Moving time average of helicity data. 9D- Moving time average of curvature data.....	16
Figure 5 Microfibrils with one end bounded. Fibril strand number from left to right: 12, 18, 24, and 36.	22
Figure 6 Unbounded fibrils. From left to right, the fibrils have 12 strands, 18 strands and 24 strands.	23
Figure 7 For the twelve stranded unbounded system: 7A-Raw helicity data. 7B- Raw curvature data. 7C- Moving time average of helicity data. 7D- Moving time average of curvature data.....	25
Figure 8 For the eighteen stranded unbounded system: 8A-Raw helicity data. 8B- Raw curvature data. 8C- Moving time average of helicity data. 8D- Moving time average of curvature data.....	26
Figure 9 For the twenty-four stranded unbounded system: 9A-Raw helicity data. 9B- Raw curvature data. 9C- Moving time average of helicity data. 9D- Moving time average of curvature data.....	27

ACKNOWLEDGEMENTS

I'd like to start off with a special thanks to Dr. Vincent Crespi for taking me under his wing towards the end of my sophomore year. Throughout the course of this project, he has been a great mentor in both life and academics. His advising helped shaped the thinking needed to not only to make achievements in this project, but will also shape the success I have as a scientist.

I'd also like to thanks Majid Nili for developing the coarse-grained model. This model working was the basis of this project, and therefore this project could not have been done without his involvement in Dr. Crespi's research group. I'd also like to thank him for walking me through his code.

I'd like to also thank Amelia Winner. She was a good lab partner to work with in my first year with this group and has turned out to be one of the greatest friends.

I'd also like to thank Paul Lammert for helping me find an expression to determine the persistence length of cellulose microfibrils.

Chapter 1

Background Information

Cellulose is among the most common organic compounds on the planet. Despite this, many details regarding its growth mechanism remain unknown. Cellulose is the main compound found in a plant's cell wall-the rigid polysaccharide layer that provides plant cells with structural support [2]. It contributes to the cell wall in fulfilling its purpose of plant rigidity because of its compact crystalline structure. What makes cellulose stiffer than other carbohydrate polymers is the beta linkages between adjacent glucose molecules. These linkages make the chain appear “flat”, allowing glucose chains to pile up on top of each other in a highly ordered crystalline arrangement, which therefore causes there to be more hydrogen bonds between adjacent glucan chains. These rope-like structures will interact with other cellulose molecules, strengthening the cell wall [1].

These polymers are constructed by a symmetric six-lobed rosette structure called the cellulose synthesis complex (CSC) [3]. Each lobe has a set of protein machinery that polymerize glucose molecules. The number of active sites in the CSC is unknown. The chains protrude into the extracellular matrix, where-once they are long enough- would coalesce into a crystalline structure, and continue expanding. We will refer to the region in between the CSC and where the strands form a compact structure as the “coalescence zone.” As these strands coalesce, defects occur in the crystalline registry. Such defects in the topology include instances where neighboring strands would swap locations along the length of the fibril, where they change their neighboring relationships with one another. Another interesting phenomenon is that these fibrils

have a bends and macroscopic twists in their lattice arrangement. In this project, we examine the fibrils' degree of helicity and the "curvature" and how they vary with fibril diameter. We also compared these phenomena to fibrils not bounded to the rosette.

Our cellulose investigations relied a so-called called a coarse-grained model. In this model, the glucan chains are modeled as beads connected by Hookean springs that represent glycosidic linkages. Each bead represents an individual monomer subunit of the glucan chain. Anisotropic bead-bead Lennar-Jones interactions interactions-pertaining to glucose's uneven width along its axes are captured. However, glucose's chirality is sacrificed in the model, since the linear chain of coarse-grained beads that results is too simple to contain any chiral centers.

With varying fibril diameter, a possible correlation between bends along the fibril and change of helicity along the fibril was noticed. We then investigated to see if a relationship exists between these two occurrences. We want to see how these properties influence the stiffness of the fibril, therefore we worked to see how the persistence length varies with fibril diameter.

1.1 Coarse-Grained Model and its Implementation

The work done in this project revolved around our coarse-grained model, so first one must address "what" is a coarse-grained model and why is it useful for studying cellulose. Cellulose, as a polymer, is composed of repeating individual monomer units, in this case a repeating sequence of glucose rings which are interconnected across opposing (so-called 1,4) carbon atoms, to yield a linear chain. The major step in developing a coarse-grained model is

identifying which components in the material you are studying should be collected as a single grain. Considering cellulose is an assembly of glucose monomers, each glucose ring was represented by a coarse-grained bead. This sacrificed the sub-molecular influences of glucose on chain-chain interactions.

The second major step in coarse-grained simulation is modeling the interaction between beads. As mentioned earlier, the glycosidic bonds are represented as bending and extensional springs. The extensional stiffness of these springs were determined by:

$$\kappa_{ext} = \frac{EA_0}{l} \quad (1)$$

with A_0 being a strand's cross sectional area, E being the strand's Young's modulus and l being a bond's equilibrium length. The bond bending stiffness, κ_{bond} , could be described in terms of a chain's persistence length, P , as

$$\kappa_{bond} = \frac{k_B T P}{l} \quad (2)$$

where k_B and T are the Boltzmann constant and the absolute temperature respectively.

Next is to determine the inter-chain interactions. How this is modeled will influence several key physical characteristics. The equilibrium distance between beads of neighboring chains will have a direct impact on the unit cell vector components of the fibril. Modeling inter-chain interactions influences the energy needed for one chain to slide axially against another. Therefore, the inter-chain interactions modeled will also have an impact on the fibril's shear stiffness, which affects the persistence length. The energy of the anisotropic bead-bead interactions of neighboring chains was based off of a modified Lennard-Jones potential, which was generalized to be capable of describing anisotropic interactions through the introduction of two distinct length scales, parallel and perpendicular to the chain axis:

$$E_{LJ}(i, j) = 4\epsilon_0 \left[\left(\left(\frac{R_{ij,\parallel}^2}{\sigma_{\parallel}^2} \right) + \left(\frac{R_{ij,\perp}^2}{\sigma_{\perp}^2} \right) \right)^{-6} - f(\theta_i, \theta_j) \left(\left(\left(\frac{R_{ij,\parallel}^2}{\sigma_{\parallel}^2} \right) + \left(\frac{R_{ij,\perp}^2}{\sigma_{\perp}^2} \right) \right)^{-3} \right)^2 \right] \quad (3)$$

Where,

$$R_{ij,\perp} = \frac{R_{ijj} + R_{iji}}{2} \quad (4)$$

$$R_{ij,\parallel} = \frac{\sqrt{R_{ij}^2 - R_{ij,j}^2} + \sqrt{R_{ij}^2 - R_{ij,i}^2}}{2} \quad (5)$$

Where R_{ij} is the separation of two strands i and j [Figure1], σ represents the distance between the beads on the two strands where the potential energy is at a minimum and ϵ_0 is a factor that influences the depths of the potential well between the two beads.

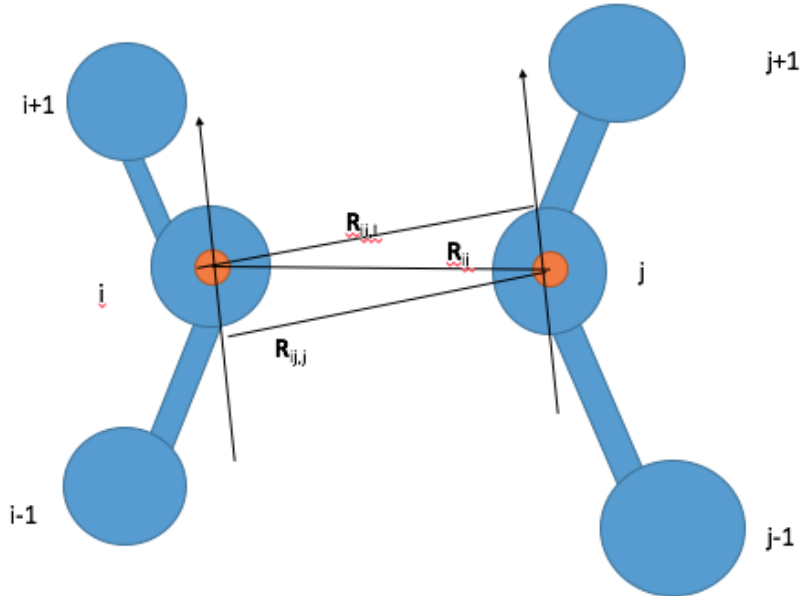


Figure 1 Chain separation R_{ij} along with its projections on the involved chains. The arrows in this figure display the local directions of each chain.

The second term of equation five deals with the anisotropy incorporated to account for glucose's varying thickness. How the anisotropy is incorporated is that each bead is assigned a particular direction perpendicular to the chain's local direction at that position and its magnitude is expressed as,

$$f(\theta_i, \theta_j) = 1.0 + \alpha[1.0 + \cos(2\theta_i) \cos(2\theta_j)] \quad (6)$$

where α sets the anisotropy's strength, θ_i is the angle between the bead i 's director and $\mathbf{R}_{ij,j}$ and θ_j between j 's director and $\mathbf{R}_{ij,i}$.

Solvent effects on the polymers are included by taking into account the single-chain stiffness in solvent and the free energy resulting from inter-chain interactions, and the Brownian forces on each bead. The kinetics from the solvent, impacting a force in a random direction on each bead can be described as:

$$\mathbf{F}_{rand} = \sqrt{\frac{6k_B T}{\mu dt}} \mathbf{U}(i) \quad (7)$$

where dt is the time step, $\mathbf{U}(i)$ is the random unit vector describing the direction the bead will move, and μ is glucose's translational mobility in water. For residues with velocities $\mathbf{V}(i)$, the drag force is:

$$\mathbf{F}_{drag} = -\frac{1}{\mu} \mathbf{V}(i) \quad (8)$$

All of these forces apply a torque on each bead, for the i^{th} bead being expressed as:

$$\boldsymbol{\tau}(i) = \boldsymbol{\tau}_{Lj}(i) + \boldsymbol{\tau}_{rand}(i) + \boldsymbol{\tau}_{drag}(i) \quad (9)$$

The torque sourced from the Brownian forces and drag force can be expressed in terms of the moment of inertia (I) and bead mass (M) as:

$$\boldsymbol{\tau}_{rand}(i) = \sqrt{\frac{2Ik_BT}{\mu M dt}} \mathbf{S}(i) \quad (10)$$

$$\boldsymbol{\tau}_{drag}(i) = -\frac{I}{\mu M} \boldsymbol{\omega}(i) \quad (11)$$

where $\mathbf{S}(i)$ is a unit vector that randomly lines along or against the residue's angular velocity, $\boldsymbol{\omega}(i)$. The moment of inertia is estimated assuming the beads are spherical shells with a radius of 2.5 Å.

Chapter 2

Algorithms

The coarse-grained simulation outputs data files for each time point; each data file being a “snapshot”, holding the Cartesian coordinates of each bead on each strand (labeled by their strand number and bead number). These snapshots were compiled in molecular dynamics visualization and analysis software to see the strands interacting, displaying qualitative features of the fibril. The information from the data files will be used for our data analysis to answer questions regarding fibril stiffness, fibril helicity, fibril curvature and topological defects. Once we get the data files needed from the simulation, we need a means of quantifying the helicity and the curvature and a means of calculating the persistence length.

Before we could get started on the data analysis for the overall geometrical characteristics of a (multi-chained) microfibril, we needed a fibril “backbone”, a set of discrete points arranged to represent a fibril’s general path path – in effect, a coarse-graining of the already coarse-grained model. Each point is paired with a normal vector that points tangent to the curve’s path. Each point-vector pair represents a plane that “slices” the fibril. The cross section of each fibril through each plane is used to help quantify helicity along the fibril, and the normal vector directions are used to quantify fibril curvature and stiffness.

2.1 Cellulose Backbone

The algorithm for the backbone slices the fibril with planes, averages the coordinates of the beads in between each of those planes, and creates new planes from those averaged “center of mass” points. The vector paired with each center of mass points from the center of the region before it and the center of the region after it. For the center of mass at the end of the fibril, the points to it from the point before it. For the first point in the fibril, the normal vector will be the vector pointing from it to the second point.

We started by averaging the first third, second third and last third of the beads of each strand to obtain a very coarse three-point representation suitable for further refinement. We then assigned a normal vector to each point. The center of mass paired with each normal vector provided planes to divide the fibril into a set of regions. A bead is assigned to be in a particular region χ if a nonpositive value is evaluated when using the plane equation for the bead’s coordinates and the χ^{th} normal vector:

$$(N_x, N_y, N_z) \cdot ((a - x_0), (b - y_0), (c - z_0)) = 0 \quad (12)$$

N_x, N_y , and N_z refer to the χ^{th} planes’ normal vector, (x_0, y_0, z_0) refer to the center of mass’s coordinates, and (a, b, c) refer to the bead’s coordinates. The algorithm to assigning each bead to a region is as follows:

Each iteration through the strands, through each plane, starts from the last bead investigated on that strand (the first strand where for the plane’s normal vector evaluated, a positive number was the result). For the plane $t=1$, you’ll start from the first bead from each strand.

- Iterate through each plane (from $t = 1$ to total number of planes).
 - Iterate through all strands in the fibril (from $i = 1$ to total number of strands).

- Increment along the i^{th} strand until the plane equation evaluates a positive value (from j = first bead at strand i to first bead that evaluates positive for plane equation).
- Record that bead number for strand i . That'll be the bead to start from on strand i in the next plane iteration.

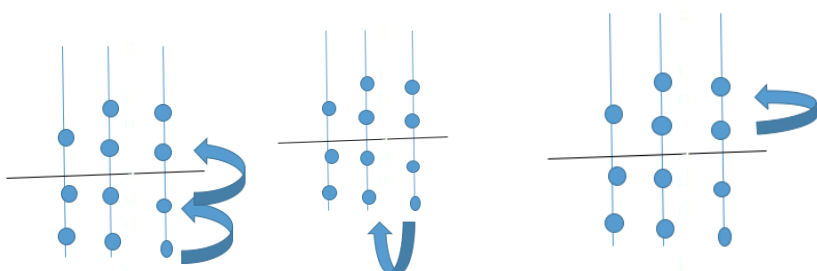


Figure 2 A step-by-step illustration of how each bead's region is determined. When evaluating which beads of a particular strand is below a given plane, each bead on that strand is evaluated using the plane equation until a positive value results. The bead with the positive value will be marked and evaluation for beads on the next strand begin. The process repeats. Evaluation for the next adjacent plane begins with the beads marked on all of the strands from the previous iterations.

Once regions are assigned to all beads, the coordinates for all of the beads in each region are averaged to generate new center of mass coordinates. The information (coordinates and normal vectors) for previous center of mass coordinates are forgotten. The ending conditions for these iterations in creating a finer backbone depends on the total number of points on the backbone and the number of beads in each region. Since the backbone is constructed from characteristics averaged over the cross-section of the fibril, it is not well-grounded to generate backbone points that are much more closely space than roughly the radius of the fibril. Having center of masses densely packed along the backbone removes its usefulness of generating a fine path of the fibril. Before any iteration begins, there is a limit defined where each region cannot be further divided if it has fewer beads than that value. Once a region has fewer beads than the limit, then the center of masses and the normal vectors of the two planes one both sides of the

region will be preserved in further iterations. This focuses divisions on fibril regions that are relatively dense with beads. A limit is also imposed on the total number of center of masses the backbone can have. Once all regions have fewer beads than the defined limit, or the limiting number of center of masses has reached, the iterative refinement ends. Because the normal vectors for beads with relatively low bead numbers are preserved throughout the iteration, they aren't exactly tangent to the final backbone arrangement. After the iteration ends, each normal vector along the fibril is modified to be tangent according to the characteristics outlined in sub-chapter one. The backbone's purpose is so that we can quantify the evolution of the fibril's overall shape over time. Therefore, we must determine the backbones over a set of data files from the same simulation. The algorithm for developing the general path over a set of data points can be summarized as:

- Iterate over a set of time points
 - Iterate until you exceed the maximum number of backbone points or until all regions have less beads than the limiting amount.
 - Determine normal vectors.
 - Identify regions for beads.
 - Determine the number of beads in each region.
 - Average coordinates in each region if they have more beads than the limiting amount.
 - Preserve center of masses and normal vectors for low bead regions.
 - Save the center of mass coordinates, the normal vectors paired with them, the regions each bead is assigned to, and the number of beads in each region

For each time point, the number of center of masses may change and the spacing between any two center of masses could change – these variations must be handled appropriately when making comparisons between different moments in time, or when constructing time-averaged quantities.

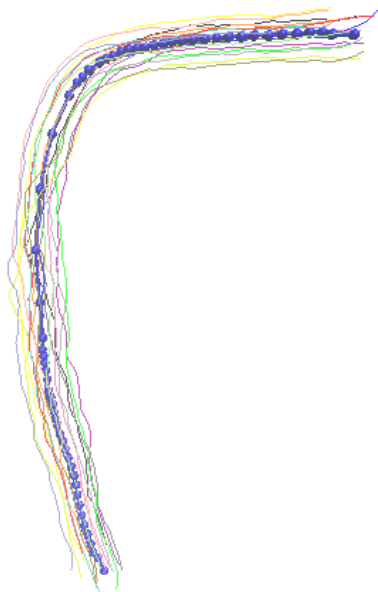


Figure 3 A twelve stranded microfilament with its backbone.

2.2 Plane Projections

Now that we have a backbone and a set of tangent vectors for each discrete point, we can determine the projections of strands onto each plane that they pass through. Because we have each bead's region assignments, we can iterate through all beads of each strand and see how each bead's assigned region compares to beads immediately adjacent to it. If two adjacent beads are in different regions, then we know there is at least one plane in between them. Also, we must

take into account the possibility that some strands may “bend” or “U-turn” in a fibril, i.e. some strands might have more than one projections onto some planes. The T^{th} plane has the T^{th} region directly behind it. To determine strand’s projection on each plane, we must first see if the strand cross the plane, then identify the closest bead from that strand to the plane at that particular crossing point. We could perform the following operation to obtain a bead’s distance from a plane:

$$W_x = a - X_0 \quad (14 \ a), \quad W_y = b - Y_0 \quad (14 \ b), \quad W_z = c - Z_0 \quad (14 \ c)$$

$$\text{Distance to plane} = \frac{|(N_x, N_y, N_z) \cdot (W_x, W_y, W_z)|}{|(N_x, N_y, N_z)|} \quad (15)$$

Where (X_0, Y_0, Z_0) , are the coordinates for the center of mass point, (a, b, c) are the coordinates for the bead, and (N_x, N_y, N_z) is the vector paired with the center of mass. The following calculation is performed to determine the bead’s projection on a plane:

$$D = \frac{-(N_x, N_y, N_z) \cdot (W_x, W_y, W_z)}{|(N_x, N_y, N_z)|^2} \quad (16)$$

$$\text{Proj}_x = a + N_x * D \quad (17 \ a), \quad \text{Proj}_y = a + N_y * D \quad (17 \ b), \quad \text{Proj}_z = a + N_z * D \quad (17 \ c)$$

Where $(\text{Proj}_x, \text{Proj}_y, \text{Proj}_z)$ are the projection coordinates of the bead. The algorithm for determining the projection of each strand for all possible planes is as follows:

- Iterate through each strand ($i = 1, \text{number of strands}$).
 - Iterate through each bead on the i^{th} strand ($j = 1, \text{number of beads}$).
 - Check to see if the assigned regions for adjacent beads are the same.
 - Then determine which planes you must check for that cross between the two adjacent beads. The possible planes in between are planes from regions whose values vary from the region value of the j^{th} bead and the region of the $(j+1)^{\text{th}}$ bead.

- Iterate through all planes that cross between the two beads.
 - For each plane, calculate the shortest distance to each of the beads to determine which is closest.
 - Calculate the projection of that bead on the plane.

2.3 Degree of Helicity

Now that we have the plane projections and the backbone points, we can begin collecting quantitative analysis of the fibril's overall structural properties and evolution over time. The first feature of interest is the fibril's helicity along its length. The determination of the fibril's helicity can be broken down into several steps. First, we define helicity: the helicity in a given a region is defined as the fibril's twist (in radians) over the distance between the two planes on both ends of the region. The fibril's twist over a region is defined as the average twist of all strand between two planes surrounding the region. Determining a strand's twist between any two adjacent planes T and (T -1) first requires tilting these two planes so that their normal vectors point in the same direction. To minimize the overall rotation angles, plane T was rotated to (T-1)'s direction (rather than bringing both planes into a common fixed reference orientation). For all strands that intersected both planes, a vector \mathbf{v}_1 was calculated from (T-1)'s center of mass to the strand's intersection at plane (T-1). A vector \mathbf{v}_2 was calculated which points from the rotated center of mass of plane T to the rotated intersection of the strand on plane T. With the vectors being calculated on planes pointing on the same axis, the angle between them could be accurately measured. After determining the twist between any two fibrils, that value would be divided by

the distance between the two planes' center of masses, resulting in the twist per unit length for that particular region.

The algorithm for rotating each plane to the plane immediately before it is as follows:

- Iterate from planes $T=2$ to $T = \text{final plane}$.
 - Determine a rotation matrix that rotates vectors by that angle (Mathematica has a rotation matrix function that outputs the appropriate rotation matrix for a tilt between two vectors inputted). This will be a 3×3 rotation matrix.
 - Multiply each strand intersection for plane T and the plane T 's center of mass coordinates by the rotation matrix.

With the information given from the rotated planes, the algorithm for calculating fibril helicity for each region is as follows:

- Iterate from planes $T=2$ to $T = \text{final}$.
 - Iterate through all strands ($i = 1$ to $i = \text{total number of strands}$).
 - If the strand intersects planes T and $T-1$, then:
 - Calculate the vector pointing from center of mass $T-1$ to strand i 's intersection at plane $T-1$ (\mathbf{v}_1).
 - Calculate the vector pointing from the rotated center of mass T to strand i 's rotated intersection with plane t (\mathbf{v}_2).
 - Calculate the angle between those two vectors.
 - The angle between those two vectors is designated a positive value if $(\mathbf{v}_1 \times \mathbf{v}_2) \cdot \mathbf{N}_{T-1}$ evaluates to being positive and negative otherwise.
 - Average the angles calculated for all strands that intersected planes T and $T-1$.

- Helicity will be that average angle over the distance between center of masses of planes T and T-1.
- As a result, there will be a sequence with the helicity values along the microfibril axis, one for each pair of adjacent planes.

2.4 Fibril Curvature

Fibril curvature was defined as the angle between two adjacent normal vectors over the distance between their center of mass points. The concept behind determining the curvature along a fibril's region was similar to determining the helicity. This information sheds light on how the fibril's general shape fluctuates along its length. When compared plotted along time, this information gives a quantitative picture of how the fibril fluctuates throughout the simulation.

We'd iterate through each set of adjacent center of mass points along the backbone and determine the angle in between them and divide that value by the distance separating them. That way we'd have the tilt per unit length along the fibril. We'd have a sequence with the curvature values along the microfibril axis, each value for a pair of adjacent center of masses.

2.5 Persistence Length

The persistence length of a fibril measures the distance along the fibril that one must travel before the direction in which the fibril is pointed changes substantially. As such, it quantifies the stiffness of the fibril under a given set of conditions. In our case, we examine the

persistence length of the fibril backbone. Imagine an N- bead polymer with beads equally spaced apart [Figure 4]. The total energy of that polymer can be expressed as:

$$H(\theta) = \sum_{i=1}^{N-1} K \times \cos(\theta) \quad (18)$$

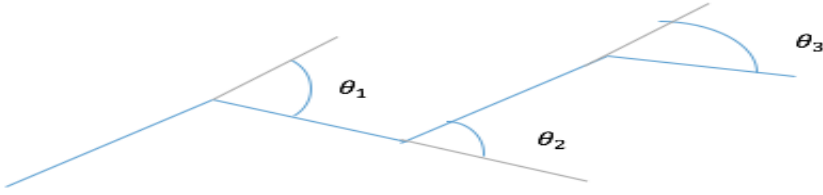


Figure 4 For the twenty-four stranded unbounded system: 9A-Raw helicity data. 9B- Raw curvature data. 9C- Moving time average of helicity data. 9D- Moving time average of curvature data

Where θ is the summation of all angles. The partition function, Z, describing the polymer's states of different angle arrangements can be expressed as:

$$Z = \int e^{-\beta H(\theta)} d\theta = \prod_{i=1}^{N-1} e^{\beta K \cos(\theta_i)} d\theta_i = (\int e^{\beta K \cos(\theta)} d\theta)^{N-1} \quad (19)$$

According to Boltzmann's law, the probability the polymer will be found in a state where all angels will sum up to θ is expressed as

$$P(\theta) = \frac{e^{-\beta H(\theta)}}{Z} = \frac{\prod_{i=1}^{N-1} e^{\beta K \cos(\theta_i)}}{Z} \quad (20)$$

If the fibril is very stiff, then

$$\langle \cos(\theta) \rangle \cong \langle 1 - \frac{\theta^2}{2} \rangle = 1 - \langle \frac{\theta^2}{2} \rangle \quad (21)$$

With this θ approximation, for any value x, its expectation value can be expressed as:

$$\langle x \rangle = \frac{\int_{-\infty}^{\infty} x e^{\beta k(1-\frac{\theta^2}{2})} d\theta}{\int_{-\infty}^{\infty} e^{\beta k(1-\frac{\theta^2}{2})} d\theta} = \frac{\int_{-\infty}^{\infty} x e^{K(1-\frac{\theta^2}{2})} d\theta}{\int_{-\infty}^{\infty} e^{K(1-\frac{\theta^2}{2})} d\theta} \quad (22)$$

where, $K = \beta k$. To solve for an expression for $\langle \cos(\theta) \rangle$, an expression is needed for $\langle \frac{\theta^2}{2} \rangle$,

$$\begin{aligned} \langle \frac{\theta^2}{2} \rangle &= \frac{\int_{-\infty}^{\infty} \frac{\theta^2}{2} e^{-K(\frac{\theta^2}{2})} d\theta}{\int_{-\infty}^{\infty} e^{-K(\frac{\theta^2}{2})} d\theta} \\ &= \frac{\int_{-\infty}^{\infty} \frac{-d}{dK} e^{-K(\frac{\theta^2}{2})} d\theta}{\int_{-\infty}^{\infty} e^{-K(\frac{\theta^2}{2})} d\theta} \\ &= \frac{-d}{dK} \left[\ln \left(\int_{-\infty}^{\infty} e^{-K(\frac{\theta^2}{2})} d\theta \right) \right] \quad (23) \end{aligned}$$

If you set $y = \sqrt{K}\theta$, then:

$$\langle \frac{\theta^2}{2} \rangle = \frac{-d}{dK} \left[\ln \left(\frac{1}{\sqrt{K}} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy \right) \right] = \frac{1}{2K} = \frac{1}{2\beta k} \quad (24)$$

Therefore,

$$\langle \cos \theta_i \rangle \cong 1 - \frac{1}{2\beta k} \quad (25)$$

For $\theta_i \ll 1$ and therefore, $\beta k \gg 1$. If you refer to Figure 4, you can see that

$$\mathbf{a}_1 \cdot \mathbf{a}_2 = |a|^2 \cos \theta_i \quad (26)$$

$$\langle \mathbf{a}_1 \cdot \mathbf{a}_2 \rangle = 1 - \frac{1}{2\beta k} \quad (27)$$

Let's investigate the dot products for any of the N discrete normal vectors in the polymer.

For any two adjacent tangent vectors evaluated from equation 26 can equal $c \times e^{-\frac{l}{\lambda}}$, where l is the distance between two adjacent points in the polymer, λ is the persistence length and c is a

proportionality constant between the exponential and $\langle \mathbf{a}_1 \cdot \mathbf{a}_2 \rangle$. We wish to evaluate the dot product between the first and third tangent vectors:

$$\langle \mathbf{a}_1 \cdot \mathbf{a}_3 \rangle = \langle [(\mathbf{a}_3 \cdot \mathbf{a}_2)\mathbf{a}_2 + \mathbf{a}_{3\perp 2}] \cdot \mathbf{a}_1 \rangle \quad (28)$$

Where $(\mathbf{a}_3 \cdot \mathbf{a}_2)$ is \mathbf{a}_3 's component along \mathbf{a}_2 and $\mathbf{a}_{3\perp 2}$ is \mathbf{a}_3 's component along \mathbf{a}_2 . This expression further evaluates to:

$$\langle \mathbf{a}_3 \cdot \mathbf{a}_2 \rangle \langle \mathbf{a}_1 \cdot \mathbf{a}_2 \rangle + \langle \mathbf{a}_{3\perp 2} \cdot \mathbf{a}_1 \rangle \quad (29)$$

Where,

$$\langle \mathbf{a}_{3\perp 2} \cdot \mathbf{a}_1 \rangle = \langle \mathbf{a}_{3\perp 2} \cdot (\mathbf{a}_2(\mathbf{a}_2 \cdot \mathbf{a}_1) + \mathbf{a}_{1\perp 2}) \rangle = \langle 0 + \mathbf{a}_{3\perp 2} \cdot \mathbf{a}_{1\perp 2} \rangle \quad (30)$$

This polymer is free to rotate around its vertices in all three dimensions, making $\langle \mathbf{a}_{3\perp 2} \cdot \mathbf{a}_{1\perp 2} \rangle = 0$.

Because vectors \mathbf{a}_2 and \mathbf{a}_3 are adjacent unit vectors a distance a apart,

$$\langle \mathbf{a}_3 \cdot \mathbf{a}_2 \rangle = c \times e^{-\frac{l}{\lambda}} \quad (31),$$

$$\langle \mathbf{a}_1 \cdot \mathbf{a}_3 \rangle = c^2 e^{-\frac{2l}{\lambda}} \quad (32)$$

A relationship can be shown for any n^{th} tangent vector,

$$\langle \mathbf{a}_1 \cdot \mathbf{a}_n \rangle = c^{n-1} e^{-\frac{(n-1)l}{\lambda}} \quad (33)$$

Where $(n-1)l$ is the distance between \mathbf{a}_1 and \mathbf{a}_n . Intuitively speaking, equation 33 implies,

$$\langle \mathbf{t}(s) \cdot \mathbf{t}(s+L) \rangle = \text{constant} \times e^{-\frac{L}{\lambda}} \quad (34)$$

Where $\mathbf{t}(s)$ and $\mathbf{t}(s+L)$ are polymer tangent vectors at locations s and $s+L$ along the fibril.

The persistence length can then be calculated by the following linear expression:

$$\ln[< \mathbf{t}(s) \cdot \mathbf{t}(s + L) >] = \ln[constant] + -\frac{L}{\lambda} \quad (35)$$

When implementing this on the fibril backbones, the expectation value of the tangent vector dot products is calculated as the time averaged of those expressions.

Chapter 3

Procedure

With the time available, we were able to obtain results regarding helicity, and curvature of the of twelve, eighteen, and twenty-four stranded microfibrils. Overall, this project aims to determine physical properties of cellulose microfibrils using a coarse-grained model. A post-doctoral researcher (Majid Nili) designed the coarse-grained model in FORTRAN, representing glucose chains interacting at room temperature with each other in water over a period of time. The program would output data files after every 100ps of the simulation.

We ran simulations to determine the fibril behavior both near and away from a region called the coalescence zone of the fibril for the scenarios where each lobe in the cellulose synthesis complex produces one strand, two strands, three strands four strands. The coalescence zone is a region outside of the cell in between where the strands form into a compact fibril and where they protrude from the CSC. To get a big picture of the behavior of the fibril, we could have had each strand contain several hundred beads. However, that would be too computationally expensive. A good compromise was to run a simulation for each system holding 100 beads per strand for two different settings: one in which the fibril is bounded at one end to the cellulose synthesis complex and the other where both ends of the fibril are free.

With the latter, the fibrils were aligned vertically along the Z-axis, symmetric around the rosette. At the first instant of the simulation, a radial guiding force would bring the strands towards the center to begin coalescence. Over a period of time, the coalescence zone would zip down the z axis. Eventually, the fibril would reach a steady state.

For the former, the strands' initial positions had to be different. With neither end being bounded, the fibril was being modeled with little influence from the coalescence zone. The strands were separated in a lattice arrangement where their separation was large enough so that the fibril would not immediately "snap" into that orientation, but small enough where no external forces were necessary to coalesce the strands. We ran these simulations until the fibril settled in a steady state. For each time point we'd generate the fibril's backbone to begin calculations.

For each simulation we'd investigate the following features: the length of each fibril, the degree of helicity along the fibril's length, the curvature along the fibril's length, its persistence length and topological defects in the crystallinity. Unfortunately, for this thesis, investigating the topological defects and the persistence length algorithms haven't been fully implemented yet.

Using the methods detailed in chapter 2, we would plot the data of helicity and curvature along the fibrils' length over a set of time. For each simulation, we'd generate two sets of three-dimensional graphs. One would have helicity plotted along the z-axis and the other would have curvature (both $\text{rad}/\text{\AA}$), with time and length along the fibril being the other two axes on both graphs. Each point would capture any slight fluctuation from region to region, making the graphs look jagged [Figures 7, 8 and 9]. To get a clear trend over time and along fibril lengths on helicity and curvature, we implemented a moving time average over the set of data.

Chapter 4

Data and Discussion

Before we begin discussing the data, it must be pointed out that the more number of strands in a system, the more time is required to for the fibril to reach steady state. With that said, for fibrils with both strands unbounded, we were only able to generate results for 12, 18 and 24 stranded systems.

What prompted us to investigate helicity was seeing how bounded 12 stranded fibrils behaved in steady state compared to higher stranded systems- 18, 24 and 36 strands. All of these simulations were done with fifty beads per strand. What we saw in higher stranded systems was that the coalescence zones would zip down the fibril quickly, with the strands being lined up vertically against one another. Twelve stranded systems were different, where we saw the lattice twist around its axis [Figure 5].

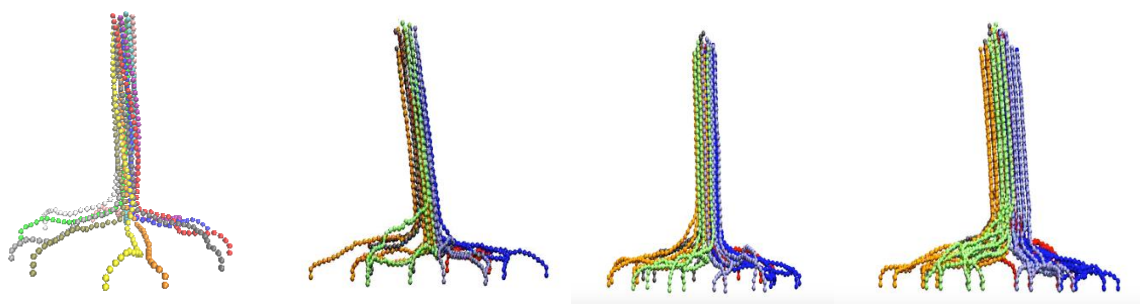


Figure 5 Microfibrils with one end bounded. Fibril strand number from left to right: 12, 18, 24, and 36.

We were curious if we'd see similar behavior for fibril systems with minimal influence from the coalescence zone, so we decided to model strand systems with both ends unbounded. What we found was that with both ends free, all fibrils had noticeable helicity [Figure 6].

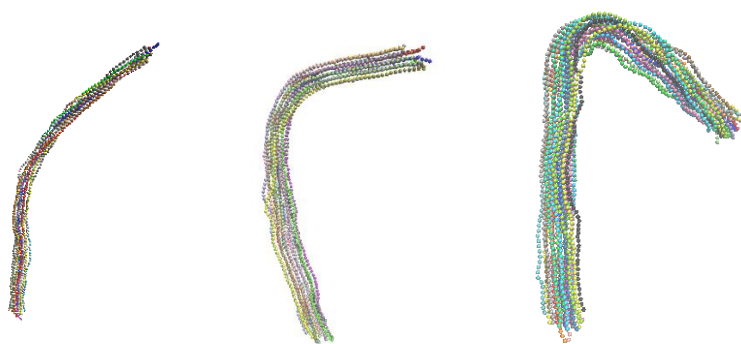


Figure 6 Unbounded fibrils. From left to right, the fibrils have 12 strands, 18 strands and 24 strands.

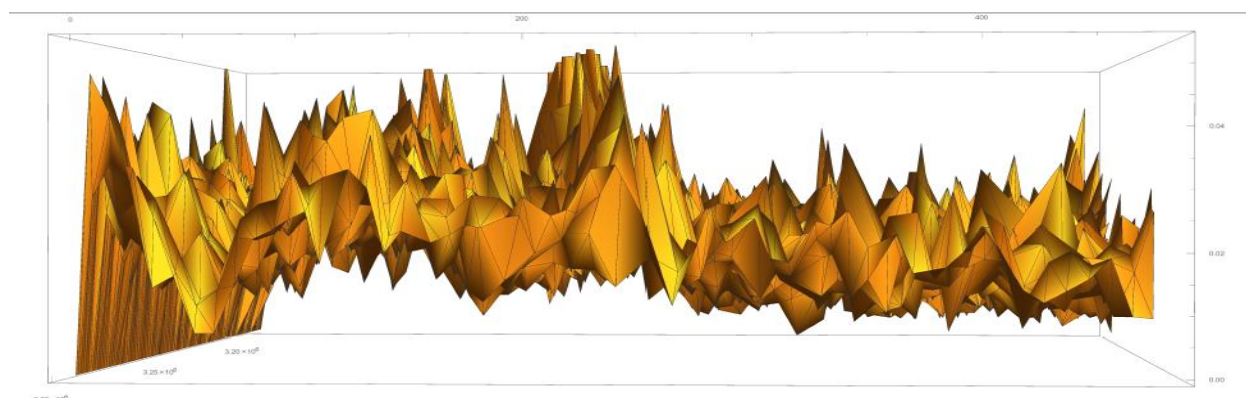
For these fibrils unbounded at both ends, we reached steady state for the 12-stranded, 18-stranded and 24-stranded systems. Because we saw virtually no helicity for the higher stranded systems for bounded fibrils, we were curious to determine how fibril helicity varies with fibril diameter. We also noticed that these fibrils bent. Not knowing exactly what caused these bends, we were curious to see if they occurred at regions with varying helicity.

Below are the plots for curvature and helicity for unbounded 12, 18 and 24 stranded systems with 100 beads per strand, for both smoothed and jagged [Figures 7, 8 and 9]. Implementing the moving time average detailed in chapter three helped with the plots' lucidity. Figure 5 shows that the freed ends behave sporadically. So when investigating a fibril's helicity and curvature, we focused on the region away from the ends.

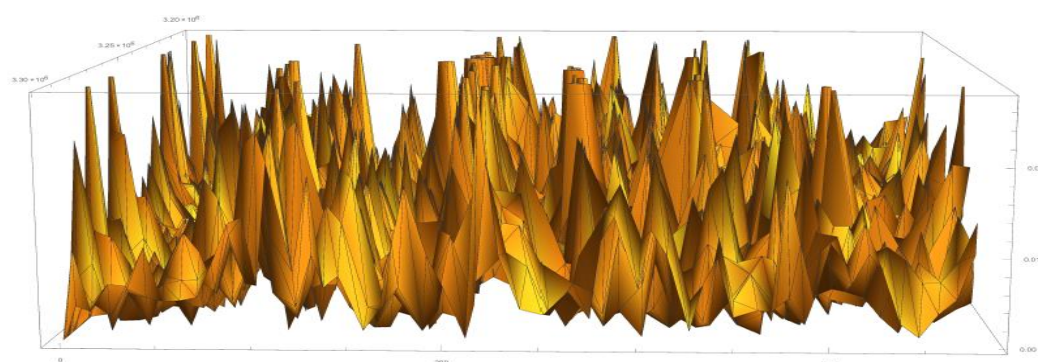
When comparing the helicity and curvature graphs, we noticed that regions of noticeable changes in helicity overlapped with changes in the curvature. This indicates a possible relationship between the two phenomena, where one possibly causes the other. When looking at the helicity plots at regions with with no significant bends (the flat regions on the helicity plots), and comparing those regions for the different stranded systems, we were able to get an idea of how the helicity varied with fibril diameter. Looking at the 12 stranded system, we saw the

helicity being approximately $0.017 \text{ rad}/\text{\AA}$, while for 18 and 24 stranded systems, their flat regions hovered slightly over $0.01 \text{ rad}/\text{\AA}$.

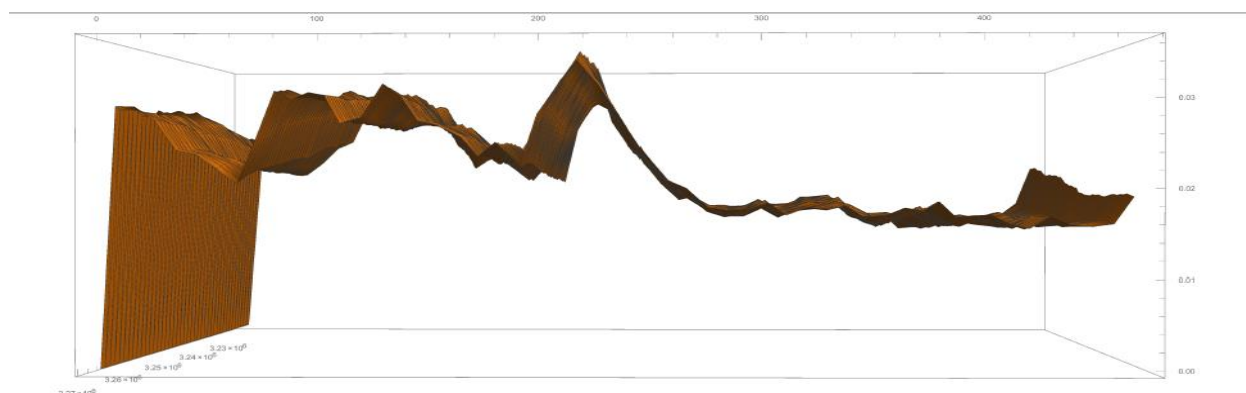
A



B



C



D

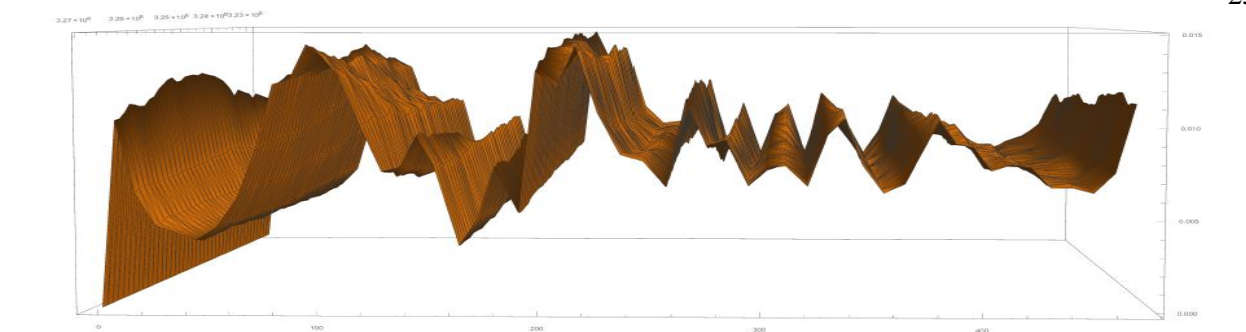
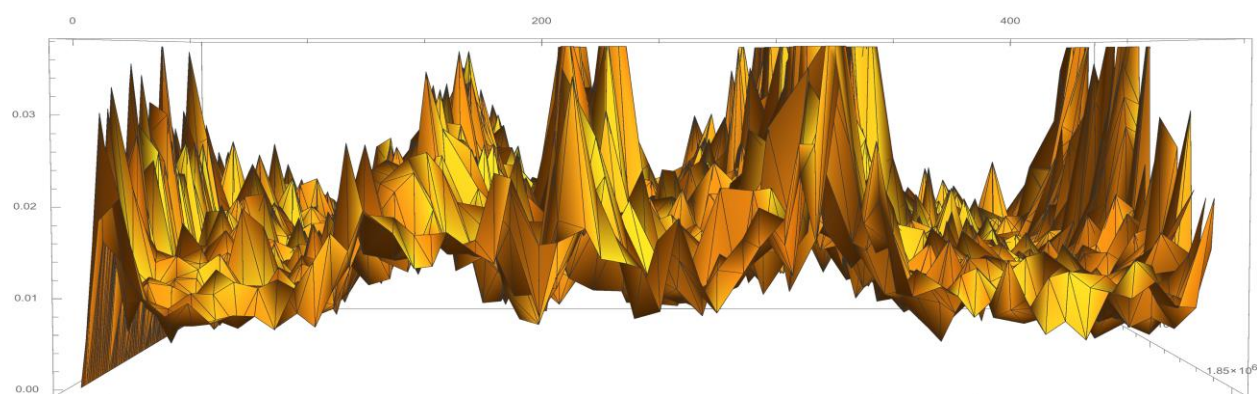
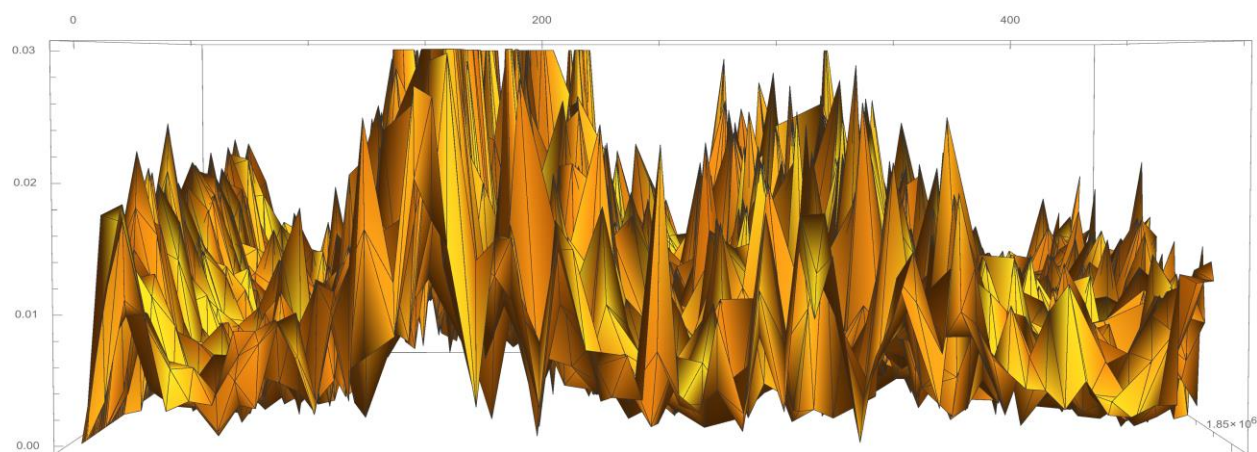


Figure 7 For the twelve stranded unbounded system: 7A-Raw helicity data. 7B- Raw curvature data. 7C- Moving time average of helicity data. 7D- Moving time average of curvature data.

A



B



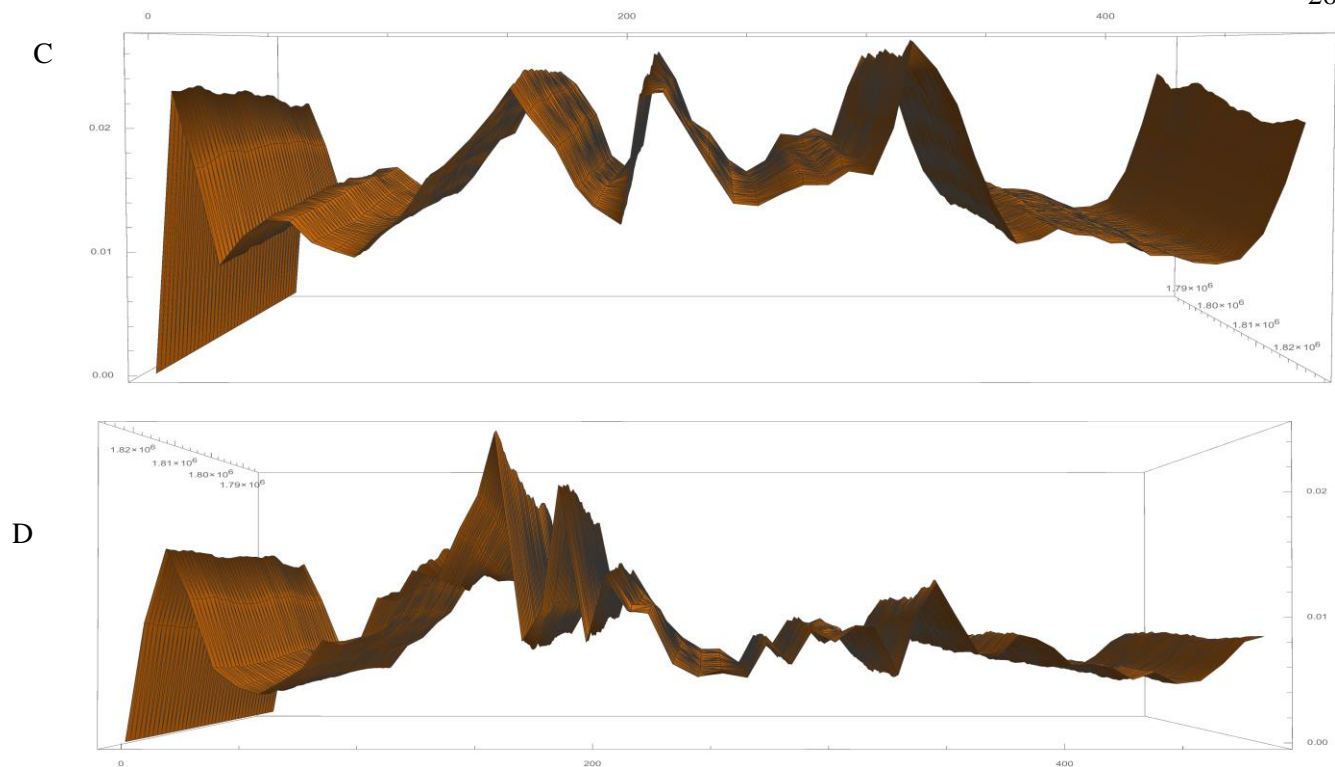
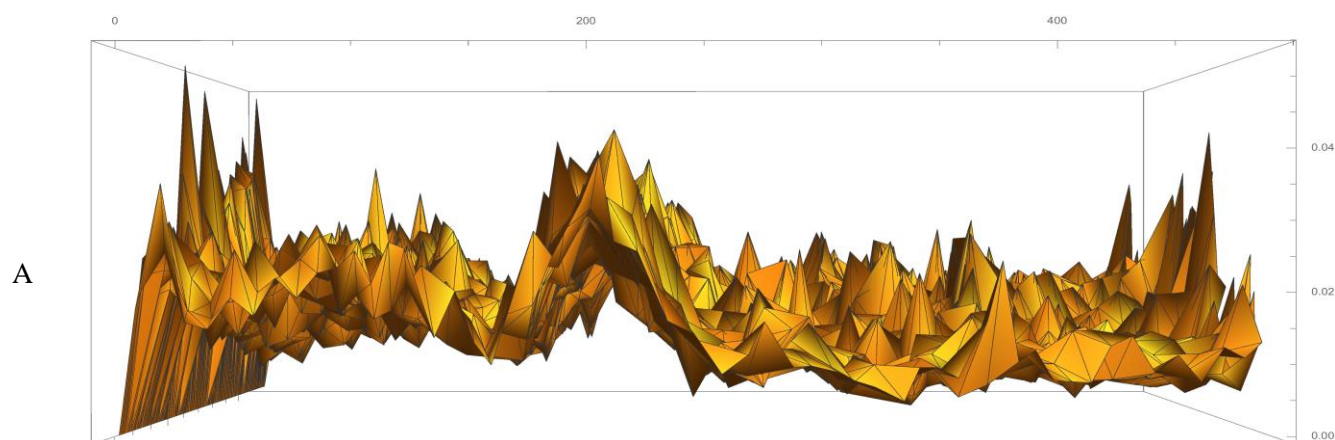


Figure 8 For the eighteen stranded unbounded system: 8A- Raw helicity data. 8B- Raw curvature data. 8C- Moving time average of helicity data. 8D- Moving time average of curvature data.



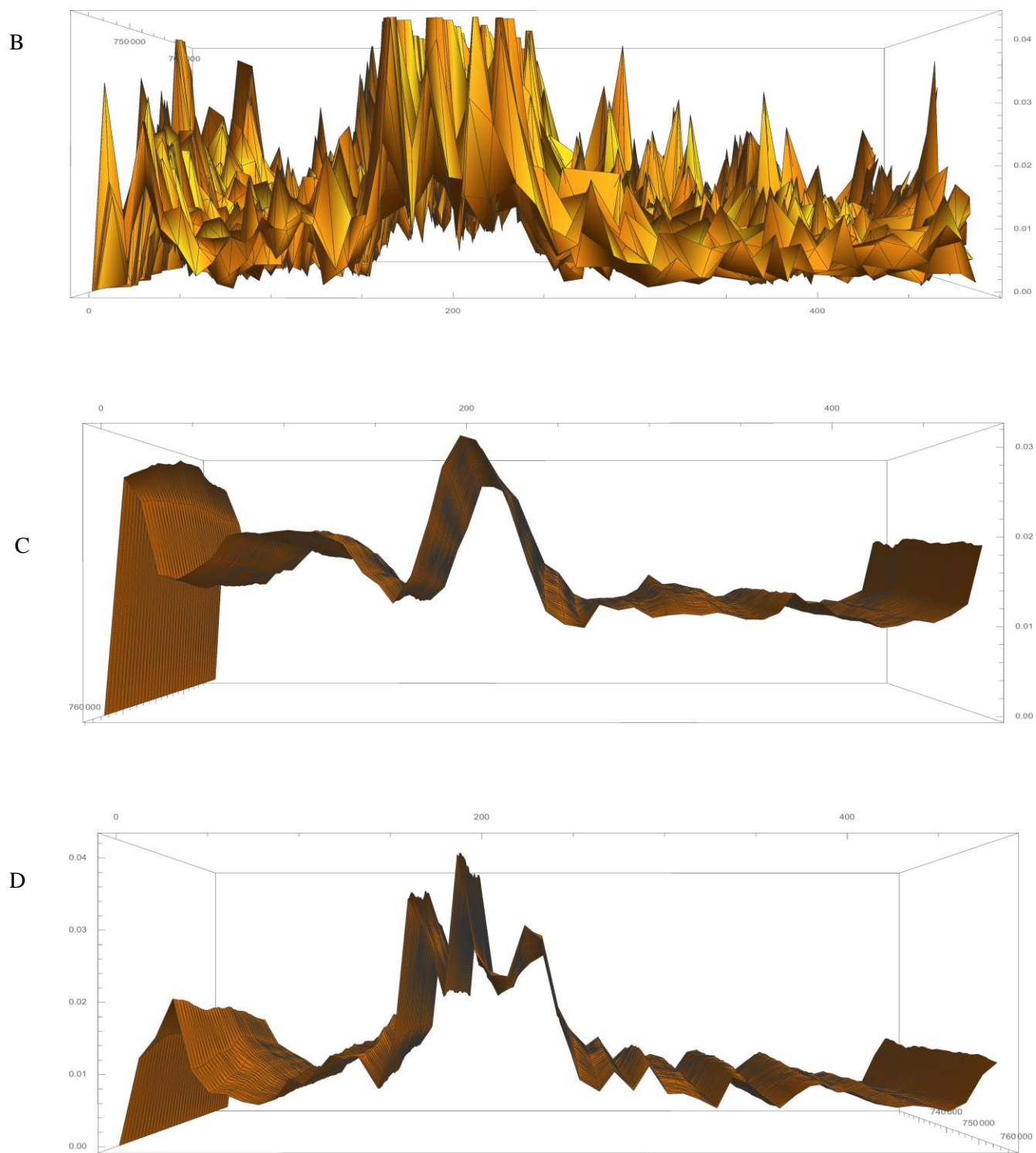


Figure 9 For the twenty-four stranded unbounded system: 9A-Raw helicity data. 9B- Raw curvature data. 9C- Moving time average of helicity data. 9D- Moving time average of curvature data.

Chapter 5

Conclusion and Future Work

With regards to seeing the relationship between fibril helicity and fibril radius, more investigation is required. Other than the fact that results need to be collected for 30 and 36 stranded systems, fibrils with 100 beads/strand are short enough in length where the fluctuation from the freed ends still holds an influential impact on the fibrils' shapes. To minimize the freed ends' influences, the same simulations should be run on fibrils with strands sizes with at least twice as many beads. Also, quantitative analyses need to be performed on fibrils with one end bounded. Developing a fibril backbone is a step which is crucial for the data analysis. For these fibrils, the backbone structure near the coalescence zone is undisciplined. Some sort of algorithm must be implemented where the coalescence zone's region in the backbone is removed. It'd also be interesting to see how bounded fibrils with strand length on the order of ~200 beads per strand behaved and see if they transform into helical structures. Analyses need to be made on all of these fronts before a solid conclusion could be made regarding fibril helicity and fibril diameter and the relationship between change of helicity and curvature.

The implementation of calculating persistence length hasn't been completed yet. Persistence length calculations must be performed to see a relationship between fibril stiffness and and fibril diameter. Also, seeing how helicity impacts a fibril's stiffness would help us answer the question if helicity improves a fibril's stability.

BIBLIOGRAPHY

1. Jane B. Reece, Lisa A. Urry, Michael L. Cain, Steven A. Wasserman, "Campbell Biology", (Benjamin Cummings, 2010).
2. M. Santiago Cintrón, G. P. Johnson, and A. D. French, "Young's modulus calculations for cellulose I β by MM3 and quantum mechanics," *Cellulose*, vol. 18, pp. 505-516, Feb. 2011.
3. S. C. Mueller and R. M. Brown, "Evidence for an intra-membrane component microfibril synthesizing complex in higher plants," *Journal of Cell Biology*, vol. 84, no. February, pp. 315-326, 1980.

Fadil Abbas Ali

faa5115@psu.edu

Local Address:
502 Simmons Hall
University Park, PA 16802

Permanent Address:
354 Ann Circle
Indiana, PA 15701

Cell: 724-762-8386

Home: 724-349-2932

EDUCATION

Bachelor of Science in Physics (Medical Option)

Minors in Biomedical Engineering and Math

The Pennsylvania State University, University Park, PA

Schreyer Honors College

Graduation: May 2016

RELEVANT COURSES

Nuclear Physics Applications in Medicine

Biomedical Applications of Microfluidics

Intermediate Programming

Organic Chemistry

Intermediate Electricity and Magnetism

Theoretical Mechanics

Introduction to Quantum Mechanics I

Anatomy and Physiology

RESEARCH EXPERIENCE

Undergraduate Research Assistant

August 2014- Present

Penn State University Department of Physics

University Park, PA

Investigating how glucan chains coalesce into a compact crystalline microfibril by simulating the process in a coarse-grained model.

Designed Fortran program that arranges coarse-grain cellulose strands into a lattice.

Developing Fortran scripts to investigate the behavior of deformities in cellulose strands.

Summer Research Assistant

June 2015- August 2015

University of Pennsylvania School of Medicine

Philadelphia, PA

Investigated the potential of assessing lung inflammation using FDG-PET.

Extrapolated lung FDG uptake at various time points.

Compared uptake values with various parameters including pulmonary function test results and air sacs volume in the lungs.

REU Research Assistant

May 2014- July 2014

Idaho State University Department of Chemistry

Pocatello, ID

Investigated the potential that germanium-sulfide thin films and tin incorporated germanium-sulfide thin films had as a dielectric for resistive RAM (ReRAM).

Wrote a LabView program that implements advanced sweeps on the ReRAM devices.

Presented research at the ACS Northwestern Regional Meeting (June 24, 2016).

HONORS/AWARDS

Bert Elsbach Honors Scholarship in Physics

Lindquist Family Trustee Scholarship

Teas Scholarship in Science

SOFTWARE

C/C++

Fortran

Unix Shell

MATLAB

PROFICIENCIES

LabView

HTML5

Mathematica

Microsoft Office

ACTIVITIES

Member of Sigma Pi Sigma- Physics Honors Society.