Monte Carlo Study

Of

Periodically Driven DNA

Abhijeet Bhatta

DEPARTMENT OF PHYSICS

NATIONAL INSTITUTE OF TECHNOLOGY AGARTALA – 799046, INDIA

MAY, 2024

Monte Carlo Study
of
Periodically Driven DNA

A Thesis Submitted to

National Institute of Technology Agartala

For The Award of The Degree

of

B.Tech.

in

Engineering Physics

by:

Abhijeet Bhatta (20DTPH024)

Under the Supervision of:

Dr. M Suman Kalyan

DEPARTMENT OF PHYSICS

NATIONAL INSTITUTE OF TECHNOLOGY AGARTALA – 799046, INDIA

MAY, 2024

APPROVAL SHEET

This thesis entitled "*Monte Carlo study of Periodically Driven DNA*" by "*Abhijeet Bhatta*" is approved for the degree of "*Bachelor of Technology, Engineering Physics*".

Examiners:

……..

……..

Supervisor:

……..

……..

Chairman:

……..

……..

Place:

……..

Date:

……..

DECLARATION

"I affirm that this project report is an authentic record of my own work, written in my own words, under the supervision of *Dr. M Suman Kalyan*. Whenever other ideas or words have been employed, I have provided appropriate citations and references. I further declare that there has been no falsification of any idea, fact, or source in this report."

Abhijeet Bhatta

20DTPH024

ACKNOWLEDGEMENT

I am deeply grateful to the Department of Physics at NIT Agartala for the provision of facilities and opportunities essential for the execution of this project. I extend my sincere appreciation to my project supervisor, *Dr. Suman Kalyan* , for his invaluable guidance and unwavering support throughout all phases of the project. His constructive feedback, meticulous corrections, and insightful suggestions have profoundly enriched my learning experience and contributed to the successful realization of the project.

CERTIFICATE

I certify that the research presented in this thesis, titled "*Monte Carlo Study of Periodically driven DNA*" was conducted under my supervision by *Abhijeet Bhatta*. Furthermore, I confirm that this work has not been submitted for publication elsewhere.

Signature of Supervisor:

……

CONTENTS

CHAPTER 1

GENERAL DISCUSSION

1.1 Introduction to DNA

Deoxyribonucleic acid (DNA) is a macromolecule which is an integral part of all living organisms. DNA plays a vital role in various metabolic processes which regulate the proper functioning of a living being. Understanding the properties and behaviors of DNA under a variety of conditions can have numerous applications from gene manipulations for healthcare to DNA based storage and computing of data.

The structure of a double stranded(ds) DNA consists of two individual anti-parallel strands of equal length which twist around each other to form a double helical structure. The basic building blocks of a single stranded(ss) DNA are called nucleotides. These nucleotides consist of nucleobases, Adenine (A), Guanine (G), Cytosine (C) and Thymine (T) along with sugar and phosphate groups. In a strand, sugar group of one nucleotide is bonded with phosphate group of another nucleotide forming a sugar-phosphate backbone. Each of the nucleobases form hydrogen bonds with another nucleobase of the other strand creating a zipped structure. Adenosine of one strand forms a bond with Thymine of the other strand and similarly Guanine forms a bond with Cytosine.

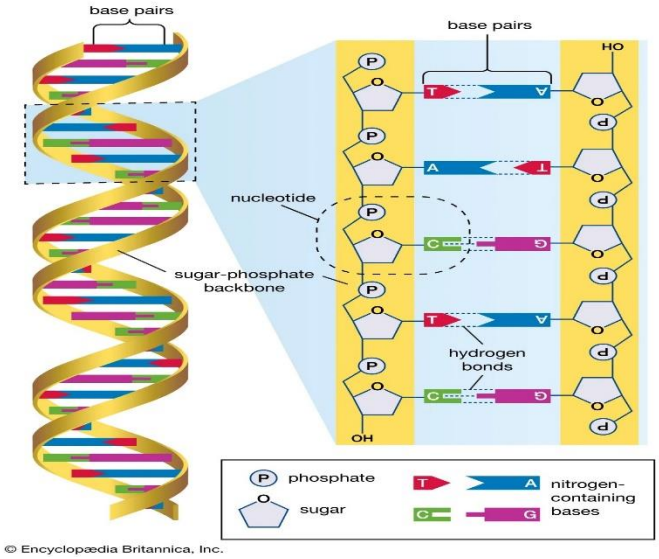


Figure 1.1 Schematic diagram of double stranded DNA helix with base pairs and sugar-phosphorous backbone

1.2 Denaturation

The process of separation of double stranded DNA into two single strands is known as denaturation. The mechanism involved in the denaturation of DNA is a prerequisite for understanding processes like replication and transcription. There are two broad ways of causing denaturation.

1.2.1 Thermal Melting

Denaturation caused by the increase in temperature is called DNA melting. The higher the temperature of the system higher the energy it contains which cause breakage of the bonds between the strands making them separate. The opening of DNA occurs at temperatures in the range of 85-90°C. Such a drastic temperature difference in temperature isn't possible in living bodies, hence the separation must happen by some other means inside organism.

1.2.2 Force induced Unzipping

Denaturation can also be done by applying a certain critical force on the strands of the DNA molecule. The mechanism of the opening of dsDNA is quite complex and is initiated by helicases, DNA and RNA polymerase, etc., which exert a force of the order of piconewtons and as a result DNA unwinds. It is now possible to unzip the two strands of a DNA using techniques like optical tweezers, magnetic tweezers, etc.

Unzipping due to a constant force can be studied easily under equilibrium conditions, however any living system is far from equilibrium. Additionally, helicases are adenosine triphosphate (ATP) driven molecular motors. The periodic hydrolysis of ATP to adenosine diphosphate (ADP) can generate a periodic force causing push and pull kind of motion. Hence it is useful to study DNA unzipping subject to a periodic force as it is more in line with biological processes.

1.3 Monte Carlo Methods

Monte Carlo methods are a heuristic approach of solving problems. Instead of providing exact analytic solutions to a problem these methods make the use of random numbers and random sampling from a model to infer characteristics and parameters with lesser amount computation. The tradeoff to these methods is in the accuracy of the calculated parameters as they have an error value associated with them. This error appears in light of the fact the Monte Carlo methods done solve a system exactly, rather uses statistics to infer data from random samplings. In Statistical Mechanics usually the random sampling is performed using the Boltzmann Distribution.

$$p_i \propto \exp\left(-\frac{\varepsilon_i}{kT}\right)$$

where p_i is the probability of the system being in state i, ε_i is the energy of that state, and a constant kT of the distribution is the product of the Boltzmann constant k and thermodynamic temperature T.

1.3.1 Metropolis Algorithm

The Metropolis algorithm is a Monte Carlo algorithm introduced by Nicolas Metropolis for the simulation of hard-sphere gasses.

Let 'S' be the set of all microstates C_i of a closed system:

$$S = \{C_1, C_2, C_3, \ldots\ldots\ldots\ldots C_n\}$$

In the Metropolis algorithm we generate a Markov Chain of microstates, $C_0 \to C_1 \to \ldots C_n \to C_{n+1} \to \ldots$

Since the progression of microstates form a Markov Chain, the microstate n+1 in the progression only depends on the n^{th} state i.e. the state immediately before it. Past states have no influence over the upcoming state.

We now define the probability of the system to go from state C_j to state C_i in a Monte Carlo step

$$P(C_{k+1} = C_i | C_k = C_j) = W_{i,j} \quad \forall k\,,$$

Let P (C_i , n) be the probability of the system to be in the state C_i at a discrete time n. To achieve equilibrium condition, we take n → ∞. Hence, we have:

$$\pi(C_i) = \pi_i = \lim_{n\to\infty} P(C_i, n) \quad \forall\, i.$$

Where $\pi(C_i)$ denotes the probability of the system to be in state C_i at equilibrium. In Statistical Mechanics this probability $\pi(C_i)$ is defined to be the Boltzmann distribution:

$$\pi(C_i) = \frac{\exp\left(-\frac{\varepsilon_i}{kT}\right)}{\sum_{j=1}^{M} \exp\left(-\frac{\varepsilon_j}{kT}\right)}$$

The principle of detailed balance states that each elementary process is in equilibrium with its reverse process. Hence the detailed balance condition for a reversible Markov Chain model like ours can be expressed as:

$$W_{i,j}\pi_j = W_{j,i}\pi_i \quad \forall\, i, j.$$

$$\frac{W_{i,j}}{W_{j,i}} = \frac{\pi_i}{\pi_j} = \exp\left\{ -\beta\left[E(C_i) - E(C_j)\right]\right\} = \exp\left[-\beta\Delta E\right].$$

Where $\beta = \frac{1}{k_B T}$

An important feature of the Metropolis algorithm is that we can generate the Markov chain without the information of the Partition function of the system, all we need is the Boltzmann weights which are easy to calculate. The ratio of the Boltzmann weights of the present and the next state gives us the acceptance probability of having the next state in the Markov Chain.

CHAPTER 2

MODEL AND IMPLEMENTATION

In order to examine the properties of dsDNA unzipping we must first define an environment where we can simulate a dsDNA structure. Such an environment is called a Model. A model is defined by a certain set of rules which any structure must obey. A good choice of the model enables us to correlate the results of the simulation to the experimental results of the real world with good accuracy. There are numerous models that can be used to simulate dsDNA unzipping, for e.g. the (1+1) dimensional square lattice model, bead-spring off-lattice models etc.

In our study we will be using the 2-dimensional four site Bond-Fluctuation Lattice Model. We decide to use this model as it isn't as computational demanding as the spring-bead in off-lattice models and also not as restrictive in degrees of freedom as the (1+1) dimensional square lattice model.

2.1 Lattice Model of DNA

In Lattice models each monomer is restricted to move along a lattice. Lattice models give a quick and computationally simple way of simulation polymers and their interactions. Lattice models can also be used to simulate the dsDNA wherein each nucleotide can be represented as a point or group of points on a lattice and various properties such as self-avoidance, zipped/unzipped configurations etc. can be simulated under the rules of the defined model. In lattice models, in spite of restricting monomers on lattice sites, a polymer has substantial conformational freedom. It is also observed that lattice models of polymers exhibit qualitatively similar behaviour as real polymers and give results in good agreement with experiments.

2.1.1 Four Site Lattice Model of dsDNA

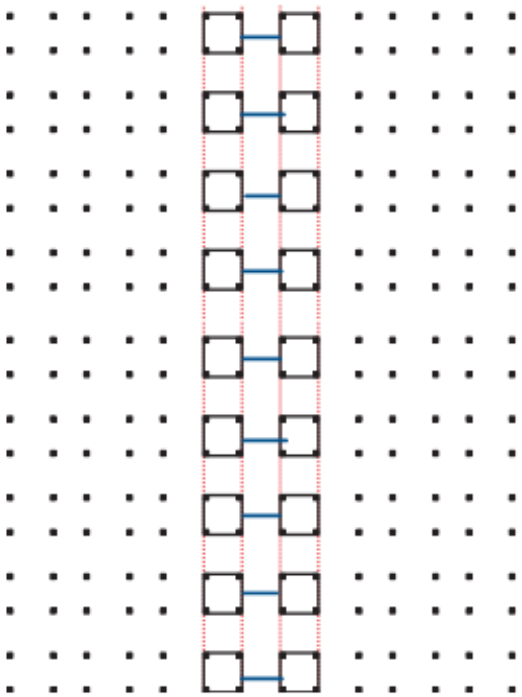


Figure 2.1 Schematic diagram of double stranded DNA on a 4-site lattice model. Each box of 4 lattice sites represents a nucleotide. Vertical red lines denote the bonds between the nucleotides of the same strand. Horizontal blue lines represent the bonds of the nucleobases between two strands.

As depicted in Fig 2.1 the 4-site lattice model gives a simplified representation of the dsDNA structure. Each monomer is represented with a box occupying 4 lattice sites at its corners. Excluded volume effect/hard core repulsion can be easily implemented by demanding no lattice site can belong simultaneously to two or more monomers. Other properties like the self-avoidance of a single polymer strand can be implemented b setting restrictions for the lengths red bonds i.e. the bonds between monomers of the same strand. Zipping and unzipping can be studied by examining bonds between the monomers of the two strands represented as blue lines between boxes. We designate such pairs as native contacts and assign an interaction energy $-\varepsilon_b$ for each such contact. Without loss of generality, we can set $\varepsilon_b = 1$. Non-native contacts carry zero energy. We apply a force perpendicular to the end monomer of one of the strands. The energy due to this applied force

$is \quad \varepsilon f = -f \cdot x$, where x is end-to-end distance i.e. distance between the end monomers of each strand.

Hence the final energy expression for a given state is:

$$E = -\gamma\epsilon_b - \vec{f} \cdot \vec{x},$$

where γ is the number of native contacts in a particular configuration.

We now can use this energy expression for carrying out our Metropolis Algorithm.

2.1.2 Bond Fluctuation Model

In the Bond Fluctuation Model (BFM) the bond lengths between two monomers are allowed to fluctuate between a set of permitted values. These fluctuations provide a us a more realistic description of the DNA mechanisms and hence are also more in line with behaviours observed in continuum models of the dsDNA.

In our study we have used the Bond Fluctuation Model on the 4-site lattice model of dsDNA. In the 4-site lattice model the minimum distance between two monomers is 2, and the maximum length must be less than 4. This constraint ensures that the configurations emerging from the BFM are self-avoiding. Hence the set of allowed bond lengths are:

Allowed Bond Lengths: {2, √5, √8, 3, √10, √13}

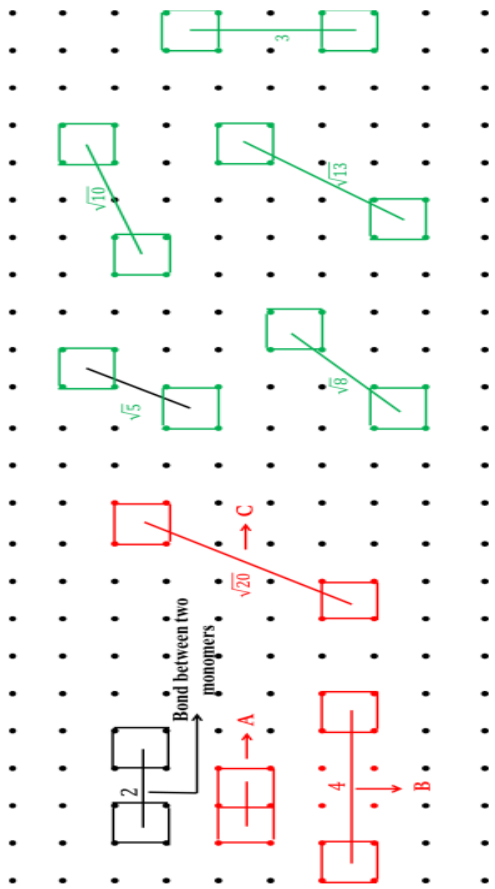


Figure 2.2 Permitted bond lengths are shown in Green, illegal bond lengths are shown in Red.

Handling 4-sites every time a move needs to be made can be a little hectic as we must select all four corners at once and update their positions. A simple workaround to this issue is to just consider the centre of mass of a monomer and treat it like the entire monomer's positions. Hence, we can convert the 4-site lattice model to a single site model which is considerably easier to work with. All the above-mentioned conditions regarding bond lengths remain the same ensuring self-avoidance and hard-core repulsion/excluded volume effect. We now have a framework on which we can start performing our Metropolis Algorithms.

2.2 Metropolis Algorithm on Bond Fluctuation Model of DNA

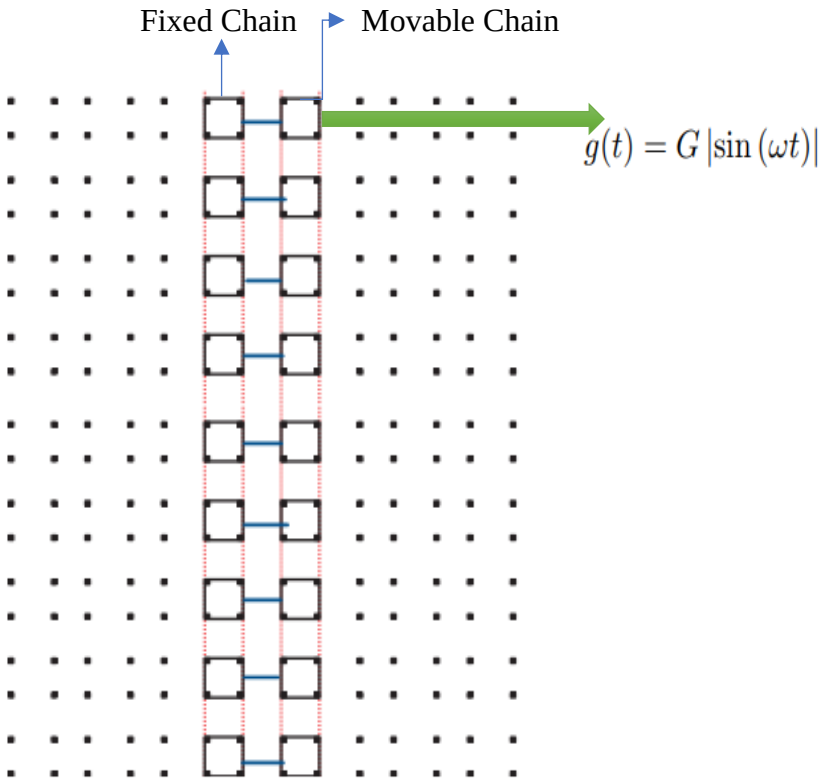


Figure 2.3 Figure shows the left chain is made fixed, the right is free to move and force g(t) acts on the end monomer of the free chain.

As shown in Fig 2.3 we use the single site representation of the BFM to have two strands of equal length of N monomers. The left strand is made immovable and the right strand is free to move in the lattice. The first monomers i.e. the roots of each of the strands are anchored. An external time-dependent periodic force g(t) shown in green acts on the N^{th} monomer of the right strand.

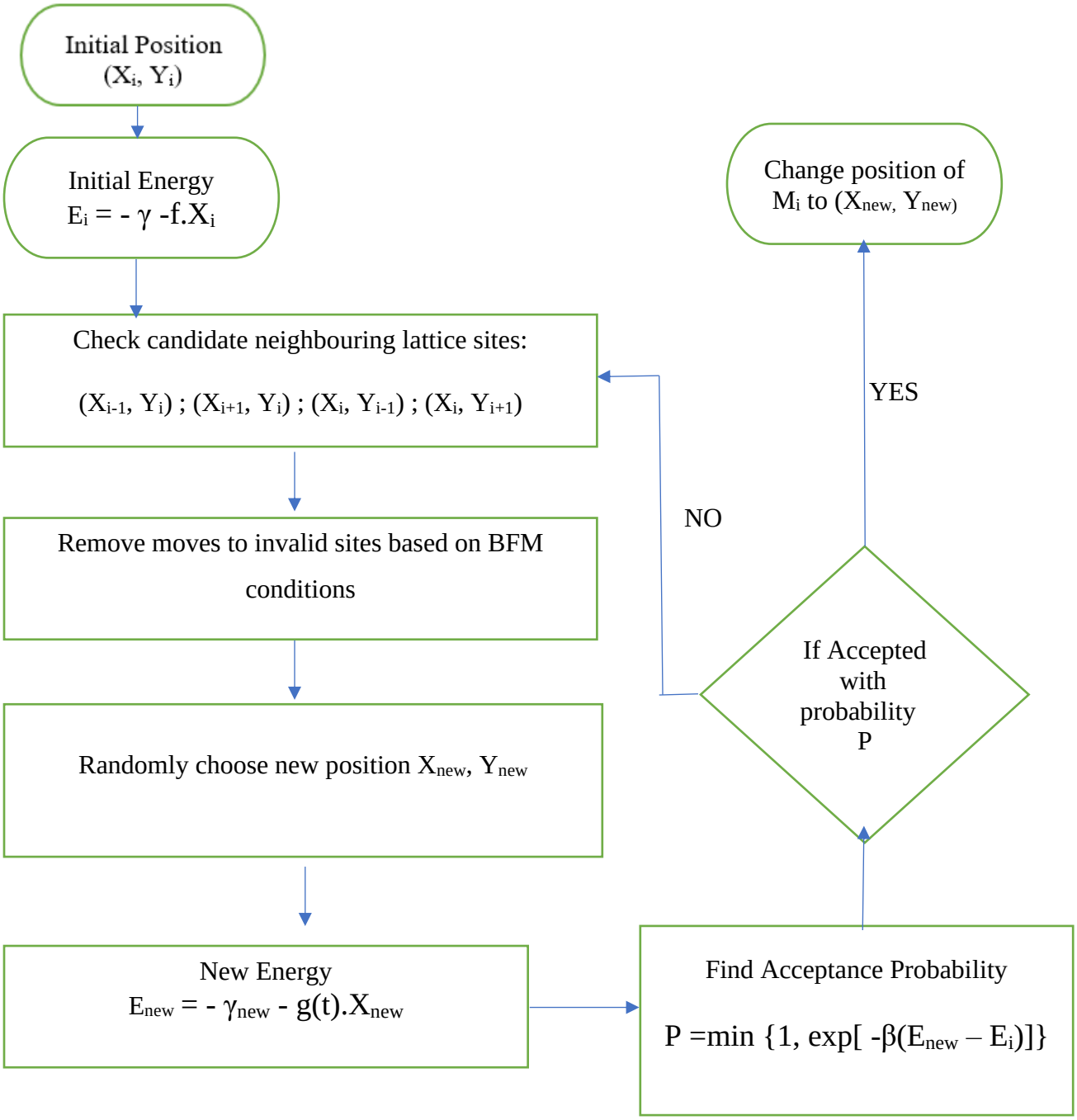
$$g(t) = G\,|\sin(\omega t)|$$

with angular frequency ω and amplitude G acts along the transverse direction (x direction) at the free end.

2.2.1 Monte Carlo Step

A Monte Carlo step is performed on a particular monomer M_i. The Monte Carlo step decides where and with what probability does the position of M_i change based on factors like external force, native contacts and temperature of the system.

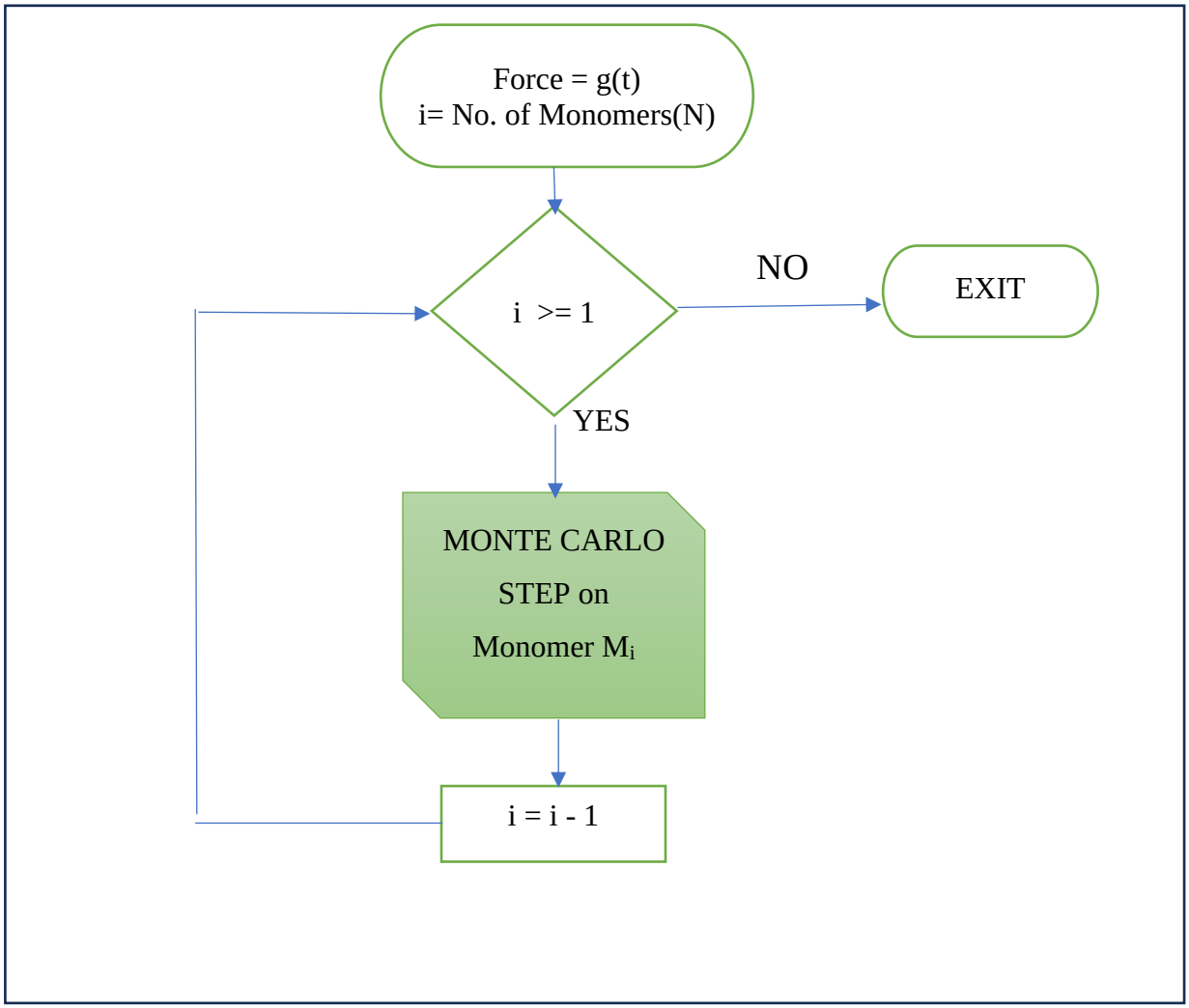
FLOW CHART OF A MONTE CARLO STEP



2.2.2 Monte Carlo Sweep

The force g(t) changes and acts on the free end at discrete time steps. For each time step a Monte Carlo sweep is performed on the system. The configuration change due to the sweep is equivalent to the response of the system due to the force. The following is the Flow Chart of a MC sweep for a system of N monomer strands.

FLOW CHART OF A MONTE CARLO SWEEP



2.3 Parameters of Interest

Now that we have the model and the algorithm to simulate the action of an external periodic force on DNA, we must measure the following parameters to infer useful information about the dynamics of DNA.

2.3.1 Distance between end monomers x(t)

The distance between the end monomers x(t) is time dependent. It changes as the external force g(t) changes with time. The values of x(t) also depend on various other factors such as length of the chain, temperature of the system, angular frequency ω and amplitude G of periodic force.

$$g(t) = G\,|\sin(\omega t)|$$

In the Metropolis Simulation the force g(t) initially starts at t=0 with g(0). Time period of the force is $\frac{\pi}{\omega}$ since the force takes the absolute value of a sinusoidal curve.

For each value of 't' a Monte Carlo Sweep is performed and the horizontal distance between the end monomers of each strand is recorded as x(t). Thereafter 't' is incremented by 1 and the process is repeated. The process ends after 10^4 cycles (time periods).

Since the force is periodic in nature, we obtain the extension x(g) as a function of force g from the time series x(t), and we average it over 10^4 cycles to obtain the average extension **<x(g)>**

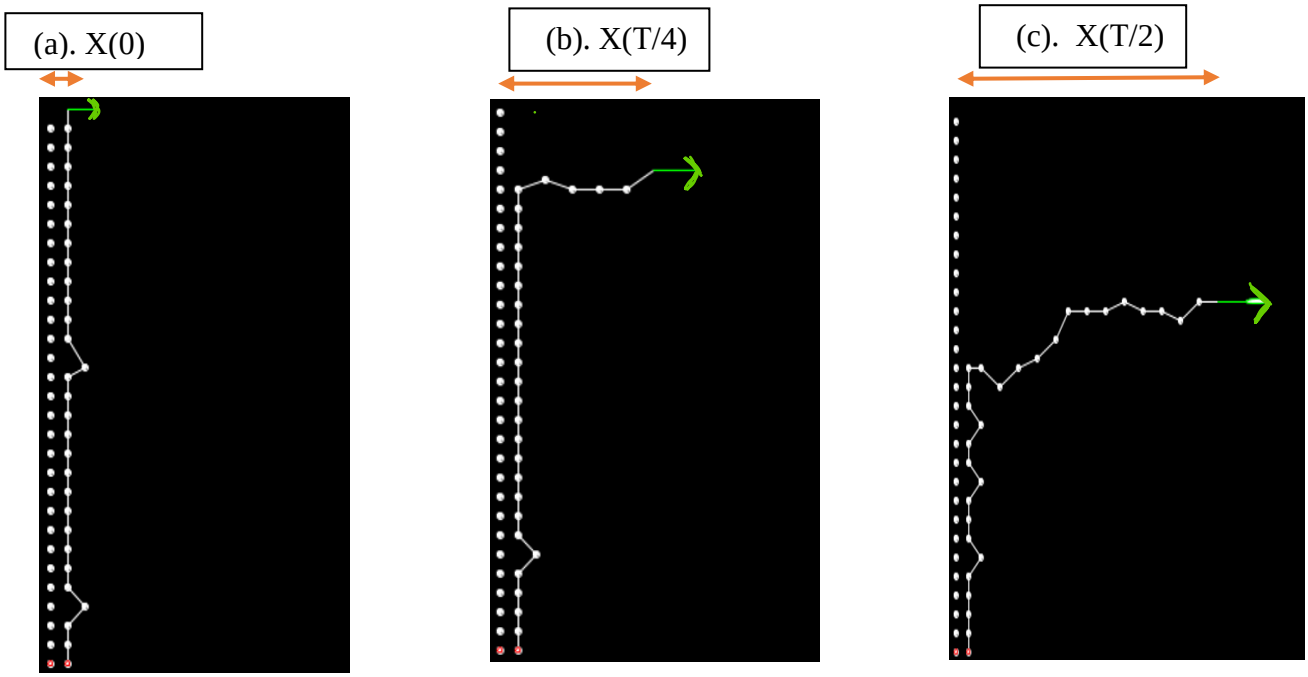


Figure 2.4 (a). Shows the distance between the end monomers at t=0; (b). Shows the distance between the end monomers at t = T/4 (T: time period); (c). Shows the distance between the end monomers at t = T/2 (T: time period). [X(T/2) > X(T/4) > X(0)]

As shown in Fig 2.4 the distance x(t) increases as the force on the monomer increases i.e. in the first half time-period (forward path). [$T = \frac{\pi}{\omega}$]. This shows the unzipping of the two strands of the DNA. In the second half of the T (backward path) when the force goes from its peak back to zero the DNA strand reverts back to its initial zipped state.

2.3.2 Hysteresis loop and its Area

If the force amplitude G is not very small, and the frequency ω of the periodic force is sufficiently high to avoid equilibration of the DNA, the average extension, <x(g)>, for the forward and the backward paths is not the same and we see a hysteresis loop.

A_{loop} depends upon the frequency ω and the amplitude G of the oscillating force and also serves as dynamical order parameter. The value of A_{loop} depends greatly on the angular frequency ω and the amplitude G. A_{loop} also depends on the chain length of the DNA system. The area of the hysteresis loop, A_{loop}, defined by: $$A_{loop} = \oint \langle x(g) \rangle dg$$

CHAPTER 3

RESULTS AND DISCUSSIONS

The results discussed have been obtained by performing the Metropolis algorithm on the dsDNA structure represented in the Bond Fluctuation Model. We have run the simulation for a periodic force g(t) such tha $g(t) = G\,|\sin(\omega t)|$

Throughout the results:

G = 1.5 (Amplitude)

k_bT = 0.18 (Reduced Temperature)

The change of the Area of the Hysteresis loop with varying the angular frequencies and varying chain lengths is studied in the sections below.

3.1 Hysteresis curves at different frequencies

For very low frequencies, the force changes very slowly, the DNA gets enough time to relax to this change and it remains in equilibrium, so the loop area Aloop is very small. Upon increasing ω, the change in the force in unit time increases, and there is some lag in the response of the DNA to this change. This is depicted by an increase in the area of the hysteresis loop. The increase in A_{loop} does not continue indefinitely with an increase in ω, there is a frequency $\omega^*(G)$ at which A_{loop} is maximum and it starts decreasing upon increasing ω above $\omega^*(G)$. For G = 1.5, A_{loop} decreases monotonically as ω increases above $\omega^*(G)$. The area decreased again at high frequencies because the force changes too quickly for the system to respond to the impulse and hence the amount of extension experienced by the DNA becomes limited.

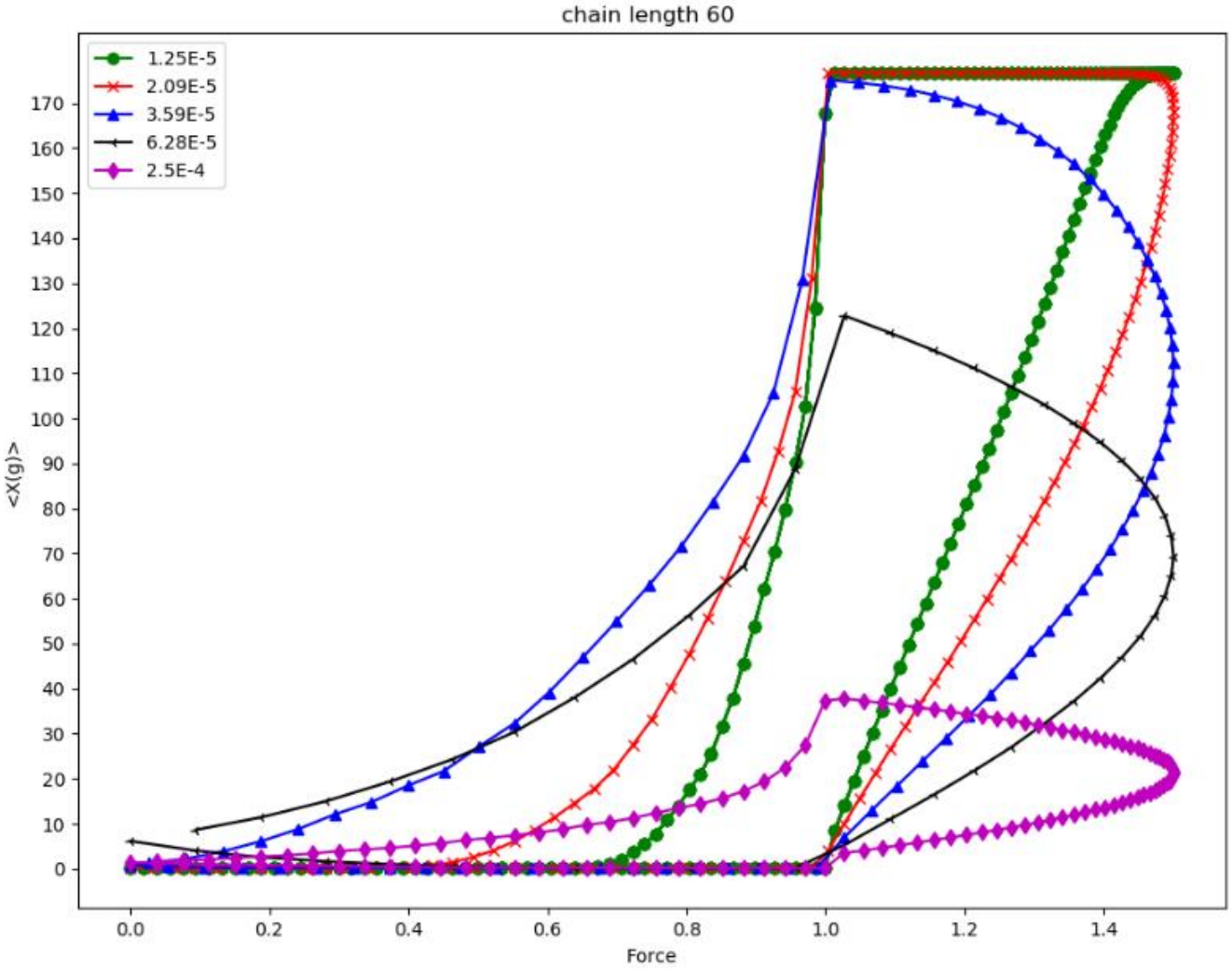


Figure 3.1 Shows the Hysteresis loops for various frequencies for a chain length of 60 monomers. Area of the loop first increases till the frequency $\omega^{*}(G)$ = 3.59E-5 shown in the blue curve and then again decreases as the frequency increases.

3.1.1 Hysteresis area vs Angular frequency data for Chain length 60

Angular freq.	Aloop (60 length)
0.0012566	5.2
0.0006283	8.8
0.000418867	12.38
0.00031415	16.36
0.00025132	19.91
0.000167547	29.41
0.00012566	38.04
0.000100528	47.88
8.37733E-05	55.53
0.00006283	71.69
0.000050264	88.26
4.18867E-05	98.74
3.59029E-05	99.5
0.000031415	95.72
2.79244E-05	90.88
0.000025132	86
2.28473E-05	81.39
2.09433E-05	76.56
1.93323E-05	72.76
1.79514E-05	68.79
1.67547E-05	65.64
1.57075E-05	62.22
1.47835E-05	59.63
1.39622E-05	56.64
1.32274E-05	54.13
0.000012566	52.15
1.19676E-05	49.91
1.14236E-05	48
1.0927E-05	45.93
1.04717E-05	44.3

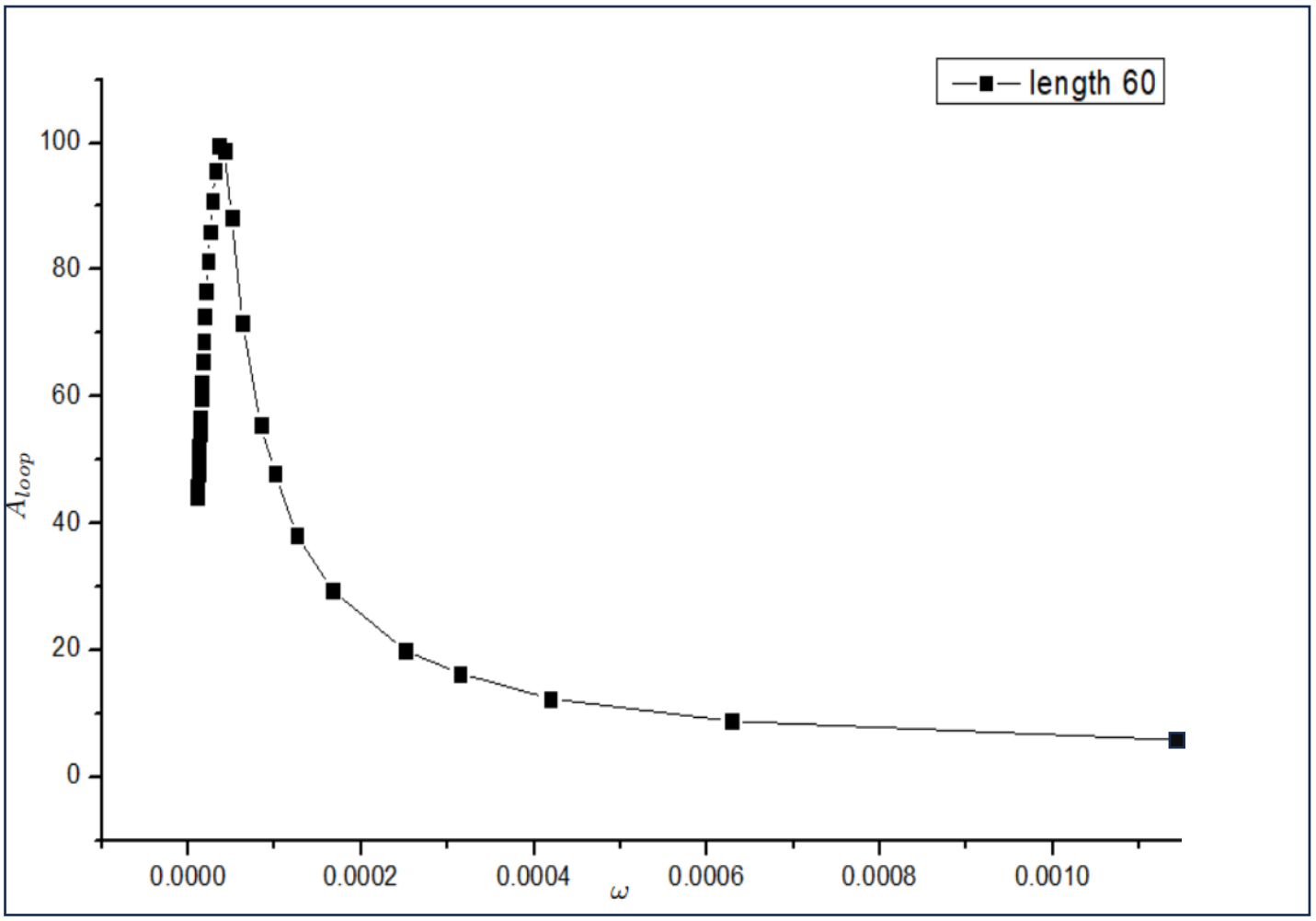


Figure 3.2 ω vs A_{loop} plot for the chain length 60 at a force amplitude G=1.5

3.1.2 Hysteresis area vs Angular frequency data for Chain length 40

Angular freq.	Aloop(40 length)
0.0012566	4.95
0.0006283	8.84
0.000418867	12.57
0.00031415	16.65
0.00025132	19.97
0.000167547	29.036
0.00012566	37.75
0.000100528	46.75
8.37733E-05	54.48
0.00006283	62.38
5.91341E-05	62.83
5.58489E-05	61.43
0.000050264	59.68
4.18867E-05	55.24
3.59029E-05	50.45
0.000031415	46.01
2.79244E-05	42.45
0.000025132	39.18
2.28473E-05	36.45
2.09433E-05	33.88
1.93323E-05	32.04
1.79514E-05	29.67
1.67547E-05	28.19
1.57075E-05	26.53
1.47835E-05	25
1.39622E-05	23.95
1.32274E-05	22.75

3.1.3 Hysteresis area vs Angular frequency data for Chain length 80

Angular freq.	Aloop (80 length
0.0012566	5.16
0.0006283	8.73
0.000418867	12.52
0.00031415	16.43
0.00025132	19.8
0.000167547	28.88
0.00012566	37.45
0.000100528	46.51
8.37733E-05	55.48
0.00006283	72.9
0.000050264	88.27
4.18867E-05	104.93
3.59029E-05	120.93
0.000031415	133.53
2.79244E-05	138.69
0.000025132	137.25
2.28473E-05	133.94
2.09433E-05	128.9
1.93323E-05	124.4
1.79514E-05	119.87
1.67547E-05	114.57
1.57075E-05	109.68
1.47835E-05	105.9
1.39622E-05	101.22
1.32274E-05	97.49
0.000012566	93.99
1.19676E-05	90.237
1.14236E-05	87.12
1.0927E-05	84.37
1.04717E-05	81.52
1.00528E-05	78.67

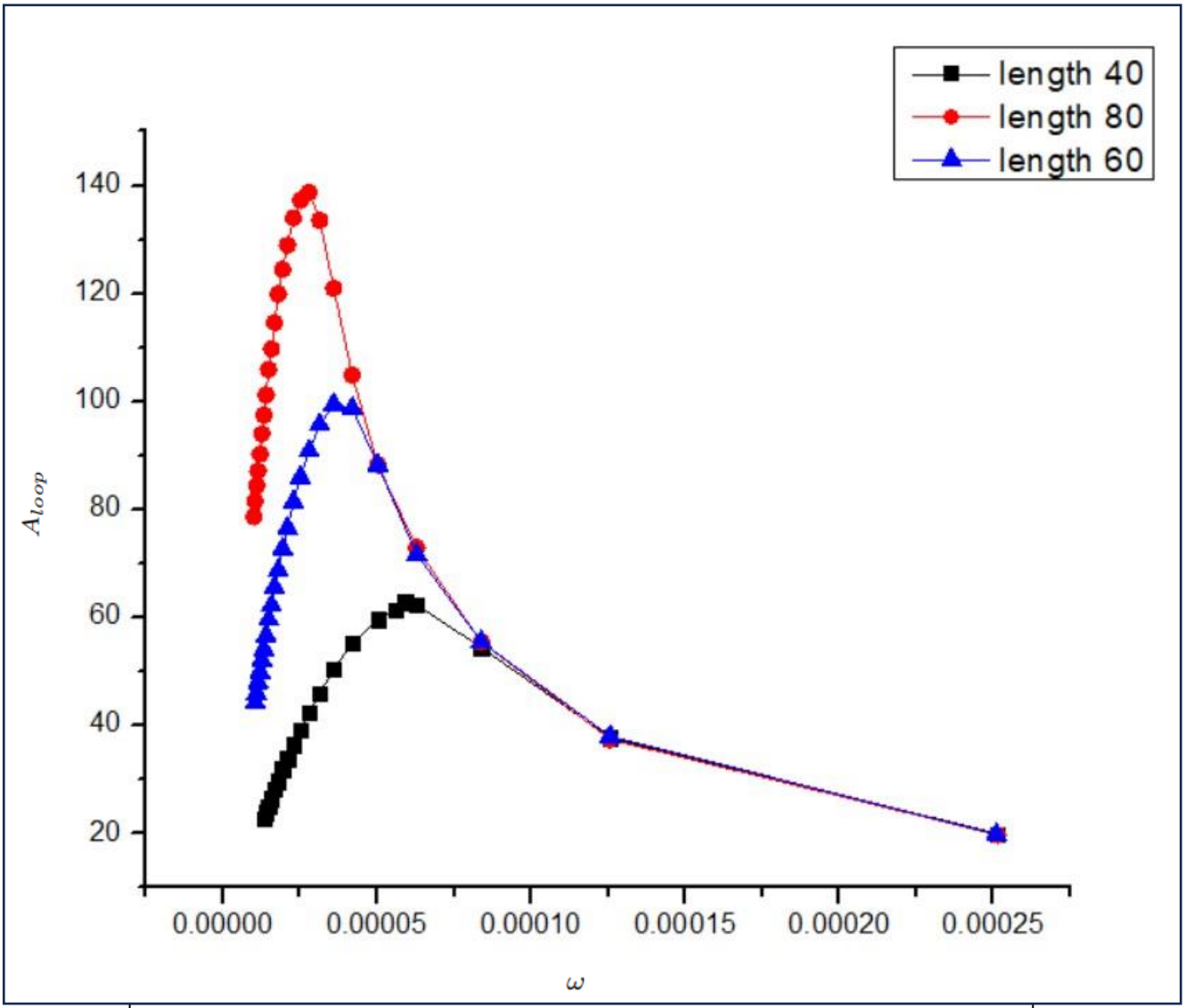


Figure 3.3 ω vs A_{loop} plot for the chain lengths 40, 60, 80 at a force amplitude. As the length increases the peak frequency decreases.

From Fig 3.3, it is clear that $\omega^*(G)$, the frequency at which the loop area is maximum, decreases as the length of the DNA is increased. In the thermodynamic limit $N \to \infty$, we get $\omega * (G) \to 0$. This suggests the scaling form for the loop area Aloop is:

$$A_{loop} = N^{\delta} \mathcal{G}\left(\omega N^{z}\right)$$

where δ and z are critical exponents. And $\mathcal{G}\left(\omega N^{z}\right)$ is a function which depends on the force amplitude G.

3.2 Estimation of critical exponents δ and z

For a given amplitude G we can estimate the δ and z values by plotting:

$$A_{loop}/N^{\delta} \text{ vs } \omega N^{z}$$

We then tune the values of δ and z till we observe a data collapse over many chain lengths over the entire frequency domain.

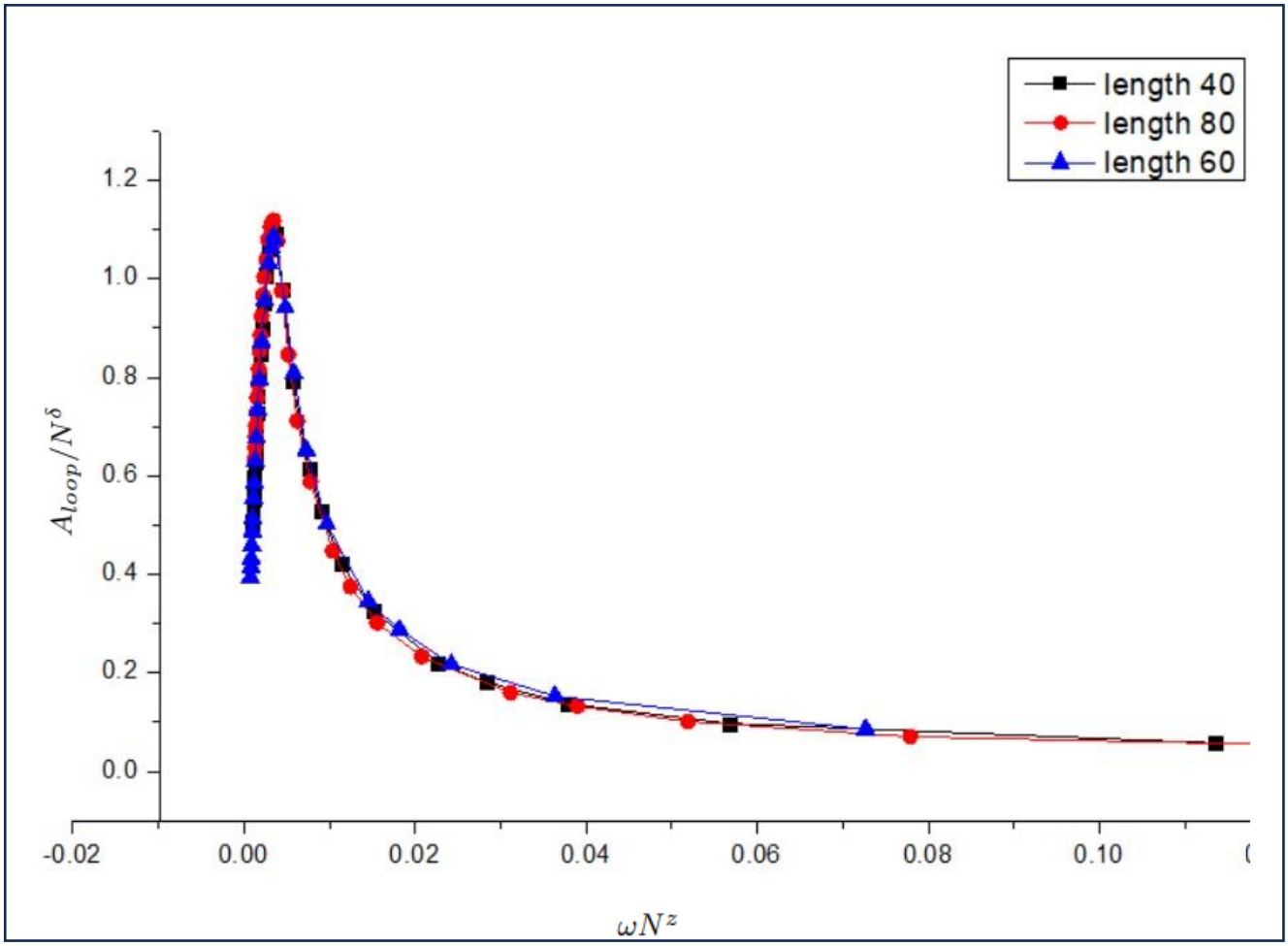


Figure 3.4 showing the collapse of data for the three chain lengths for δ =1 and z = 1

We observe a good data collapse for chain lengths 40,60 and 80 for the value of δ =1 and z = 1

CHAPTER 4

CONCLUSION

Understanding the dynamics of DNA has long been a very active area of research which has attracted a variety of minds to properly model its structure and hence infer useful information about its working. Although studying such a complicated system and all its intricacies is a hefty task and requires loads of computational as well as human hours.

We have demonstrated in this thesis that simplified models without all the minor intricacies can still be used to study some of the properties of the DNA. We have focused our effort in the direction of zipping and unzipping of DNA double strand due to the influence of a periodically driven force. Such interactions happen inside every cell of every living organism. Monte Carlo methods especially the Metropolis Algorithm allows us to build simple and not so computationally intensive models to simulate the dynamics of complicated systems.

Using the metropolis algorithms we have shown the behaviour of the DNA strands in presence of an external periodic force. This revealed that the DNA's extension with the force follows a Hysteresis curve. We have correlated the value of the Area of such a curve with the change in frequencies and lengths of the strands. In doing so we found the Critical exponents δ and z to be in the vicinity of 1

In the Future using the same approach we intend to find out the relation of frequency and of the Force amplitude with respect to the Area under of the Hysteresis.

REFERENCES

1. S. Kumar and G. Mishra, Phys. Rev. Lett. 110, 258102 (2013)
2. G. Mishra, P. Sadhukhan, S. M. Bhattacharjee, and S. Kumar, Phys. Rev. E 87, 022718 (2013)
3. Unzipping DNA by a periodic force: Hysteresis loop area and its scaling Rajeev Kapri∗ arXiv:1409.8392v2 [cond-mat.soft] 14 Jan 2015
4. Hysteresis loop area scaling exponents in DNA unzipping by a periodic force: A Langevin dynamics simulation study Rajeev Kapri arXiv:2104.07170v2 [cond-mat.soft] 3 Aug 2021