Numerical Simulations of Ideal Chain Model of Polymer using the Freely Jointed Chain (FJC)

Ahan Palsole and Eric Vidal

1 Introduction

Physics of polymers is an interdisciplinary field that aims to apply chemistry, maths, and statistical mechanics principles to explain the behavior of this special kind of macromolecule. Polymers are constituted by the repetition of small organic molecules called monomers bonded to each other along chains. For a deeper understanding of macromolecules and polymers, one can consult part I of Cantor et al. book[1].

In this intricate problem, a large ensemble of factors influences polymer properties, e.g. solvent medium, temperature, molecule length, tacticity, steric effects, monomer-monomer or non-covalent interactions, and Brownian motion. This leads to a focus on several features such as cross-bonding, hydrogen bonding, Van-der-Waal's forces, and phase transitions. Even more interesting features can be studied as Chen et al.[2] did for understanding and improving polymeric materials. Hence, several models have been proposed to describe the different sorts of polymer characteristics that can be classified as discrete or continuous. A broader explanation of the polymer properties and classification can be found in the first two chapters of Rubinstein's book[5]. In Saleh's paper[6] different models are studied with a historical scope as well. It is particularly interesting because the case studied is the single polymer mechanics across different force regimes using the model of the entropic string, the freely jointed chain (FJC) model, the worm-like chain (WLC), and the blob models.

1.1 The freely jointed chain (FJC)

In the present work, the FJC model is used to study the numerical simulations of some ideal chain dynamics. This model is contained in the discrete category, therefore, the polymer is treated as a set of discrete monomers connected by rigid bonds. In regards to the FJC model, the bonds are assumed to have constant length and no preferred orientation in space, plus monomer-monomer interactions are assumed to be zero. More realistic models can be constructed by adding constraints on the backbone angles while allowing the torsion angle of the molecule to vary freely, or introducing elasticity to the bonds.

In what notation concerns, each monomer is labeled with an index, i, and the position of each monomer, i, relative to the previous monomer, $i-1$, is defined using the bond vector, $\vec{r}_i$. The position of each monomer, i, relative to some arbitrary origin can then be defined as, $\vec{R}_i = \vec{R}_0 + \sum_{k=1}^{i} \vec{r}_k$, being $\vec{R}_0$ the position of the first monomer, usually set as the origin, $\vec{R}_0 = \vec{0}$.

1.2 Metrics

Since the reference frame is established, several metrics can be defined to obtain the analytical expressions that parameterize the polymer. The end-to-end distance, $\vec{Q} = \sum_{i=1}^{n} \vec{r}_i = \vec{R}_n - \vec{R}_0$, is the most basic measure of extent since it is simply the sum of all the n bond vectors (also referred as N). Note that the chain follows a random walk process, so the mean of $\vec{Q}$ is zero, $\langle \vec{Q} \rangle = 0$. However, there is a magnitude more useful for measuring the extent of the polymer which is the mean of the squared end-to-end distance,

$$\langle Q^2 \rangle = nb^2 + \sum_{i \neq j}^{n} b^2 \langle \cos\theta_{ij} \rangle \tag{1}$$

where b is the bond vector length and the bond angle, θ_{ij} is the angle between monomers i and j. In the case of the FJC model, it follows a random walk process so $\langle \cos\theta_{ij} \rangle = 0$. At this point, it is good to note that there are $n+1$ monomers, $\vec{R}_i$, n bond vectors, $\vec{r}_i$, and $n-1$ bond angles, θ_{ij}.

The radius of gyration, $R_g^2 = \frac{1}{N}\sum_{i=0}^{n}(\vec{R}_i - \vec{R}_{CM})^2$, is also used to quantify the size of a folded polymer. It is defined as the average of the distances of each monomer relative to the molecule's center of mass, $\vec{R}_{CM} = \frac{1}{N}\sum_{i=0}^{n}\vec{R}_i$. When we compute its expression analytically we get to $R_g^2 = \frac{n^3 b^2}{6(n+1)^2}$, which ends up to be,

$$\langle R_g^2 \rangle = \frac{nb^2}{6}, \tag{2}$$

for a large $n \to \infty$.

Lastly, another interesting metric is the probability distribution of Q, in which the *sinc* function can be approximated by a Gaussian, and thus obtain,

$$P(Q) = 4\pi Q^2 \left(\frac{3}{2\pi n b^2}\right)^{3/2} e^{-\frac{3Q^2}{2nb^2}}, \tag{3}$$

again for the limit of large n. However, there are two singular behaviors, for one bond which is trivial with $Q = b$, and the one corresponding to two bonds. The probability distribution for $N = 2$ is affected only by the uniform distribution of one bond angle, and the probability density can be simplified as, $P(Q) = \frac{Q}{2b^2}$.

1.3 Experimental study

Experimentally, the polymer extent and entropic elasticity are usually studied either by Small Angle Scattering (SAS) experiments using visible light, x-rays (SAXS), neutrons (SANS), or Atomic Force Microscopy (AFM) techniques. Specifically, we consider SAS methods as the size of the substrate under study is significantly larger than the wavelength of the incident radiation. Then, we can define the structure factor of the polymer,

$$I(k) = \sum_{i=0}^{n}\sum_{j=0}^{n} avg\left(\frac{sin(k|R_i - R_j|)}{k|R_i - R_j|}\right), \tag{4}$$

which is a function of the incident wavenumber, k. In SAS experiments, the information about the larger-scale structure of the molecule is usually found in the central part of the scatter pattern which corresponds to small k. This leads us to the Guinier

approximation which is applied for $(k\,R_g) \ll 1$,

$$I(k) \approx (n+1)^2 \left(1 - \frac{k^2 R_g^2}{3}\right), \tag{5}$$

therefore we can deduce the R_g from this expression.

Computation of the Structure Function is straightforward following equation (4), where the length of the molecule can be iterated to calculate the distance between each monomer and every other monomer of the molecule, analogous to an autocorrelation function. When light from a sufficiently distant source is incident on the sample, it can be treated as a plane wave $e^{i\vec{k}\vec{x}}$, demonstrating that SAS is a method of physically calculating the power spectral density of the molecule's spatial arrangement.

1.4 Goal

In the first instance, we have to generate multiple configurations of an FJC polymer in order to compare with the theoretical results and evaluate the validity of the approximations. For this purpose, we first study the polymer structure generated, the end-to-end distance, and the gyration radius to verify the model and the simulation are correct.

Once the simulation is done, the mentioned metrics are computed as well as how the error decreases with the number of conformations for $\langle Q^2 \rangle$. Moreover, the Metropolis Monte Carlo algorithm is used to simulate the expansion of the polymer when a force is applied along the x-axis, F_x. In this line, the extension in x, Q_x is studied in function of the number of steps, considered as time, t, and in function of F_x.

2 Methods

For the sake of the study, different conformations of the polymer must be produced via numerical simulations using the fundamentals of the FJC model and random walk theory. On the one hand, the bond length is set to 3, $b = 3.0$, in arbitrary units, and this is the reason why the magnitudes will not have units throughout the paper. Whereas on the other hand, the parameters, n, number of monomers, and T, number of conformations, will vary depending on the matter of study.

The process undertaken to generate the random configurations consisted of saving in a *.xyz* file a uniformly distributed set of positions $x_{t,i}, y_{t,i}, z_{t,i} \in [-1,1]^3$, with indexes for the $t = 1,2,\ldots,T$ conformations and the $i = 0,1,\ldots,n$ coordinates. To ensure the uniform distribution, firstly the bond vectors are computed beforehand, and the ones with norm above 1 are discarded until n are accepted, then normalized, and finally, the $n+1$ coordinates are saved. So the final result is computed and saved in a file containing the polymer in the FJC model as an ensemble of coordinates. From these coordinates, one can generate all the normalized bond vectors uniformly distributed in a sphere of radius 1. This numerical simulation pretends to compute the stochastic fluctuations in the polymer with time. We can see two examples of the conformations generated with $N = 100$ bonds in Fig. 1.

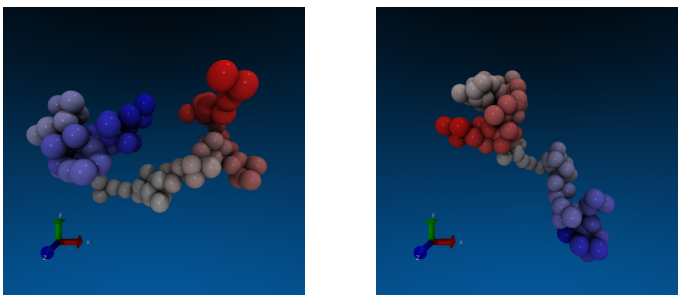

Fig. 1 Two different polymer structures represented with VMD from FJC model simulations for $N = 100$ and $b = 3.0$.

2.1 Time series and Statistical Tools

The simulated polymer conformations must be Markovian chains so that each time step is completely random with no dependence (or memory) on the previous ones. This fact delivers a uniform sampling across all possible conformation as we can see in the time series of $R_g = \sqrt{\vec{R}_g^2}$ and $Q = \sqrt{\vec{Q}^2}$ represented in 2 to show the correctness of this expected behavior. As we can see the distribution is random and, in fact, the mean of $\vec{Q}$ would tend to zero.

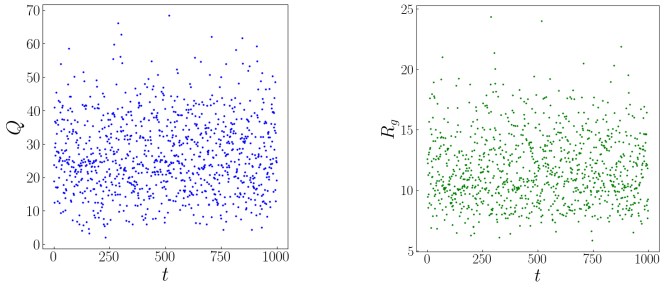

Fig. 2 Representation of Q and R_g of the polymer for $N = 100$ bonds and $T = 1000$ conformations with index of the conformation, t, that can be interpreted as time steps.

The next step is analyzing the validity of the theoretical equations (1) and (2) that conform to the basis of the FJC model. Therefore, these magnitudes are computed for 1000 conformations and different numbers of bonds ranging from 10 to 1000, with the bond length constant it has already been stated. Hence, the simulations should conform to a linear plot in function of the number of bonds N. Indeed, this is what happens and it is shown in Fig. 3. Apart from that a linear regression can be performed but it is not shown in the figure as it superposes the theoretical value, due to its accuracy.

Lastly, it is interesting to study how the error of the studied quantities $\langle Q^2 \rangle$ and $\langle R_g^2 \rangle$ evolves with the total number of conformations averaged to get the simulation value compared to the theoretical value. What is performed in this case is the relative error of these values, ε_{rel}, and plotted in function of T in Fig. 4.

The relative error for $\langle Q^2 \rangle$ decreases as the number of conformations studied increases so there are more values to compute the statistics. However, in the case of $\langle R_g^2 \rangle$ relative error does not behave this way, as it contains a mean over all the conformations for the position of the center of mass. So, the fewer conformations the less accurate the position of the center of mass, but it

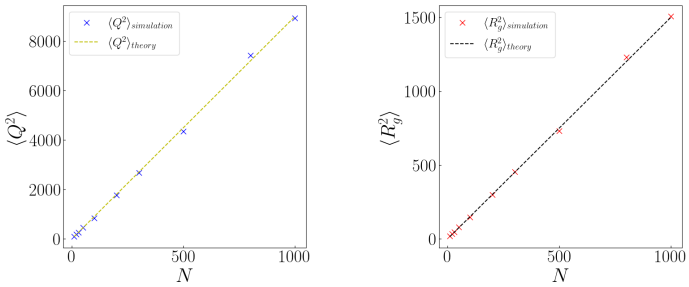

Fig. 3 Representation of $\langle Q^2 \rangle$ and $\langle R_g^2 \rangle$ over $T = 1000$ polymer conformations with respect to the number of bonds, N, ranging from 10 to 1000 compared to the model as a function of the number of molecules and (right) error in calculated Q^2 values as a function of the number of configurations averaged over

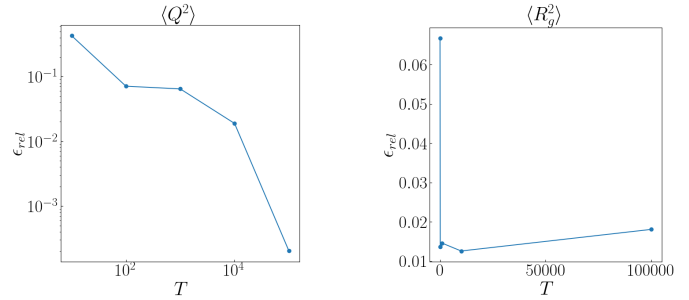

Fig. 4 Representation of the relative error, ε_{rel}, for $\langle Q^2 \rangle$ and $\langle R_g^2 \rangle$ with $N = 100$ in function of the number of conformations ranging the powers of 10, from 10 to 100000.

might induce that the less accurate positions lead to a gyration radius close to the theoretical one. Still for $T \leftarrow \infty$ the numerical and theoretical results would be equal.

2.2 Summary of Data Produced

As an example, a summary of the data produced for the polymer with $N = 100$ bonds and $T = 1000$ of Fig. 2 is gathered in Table 1. From quantities Q and R_g, it is obtained the equivalent simulations to equations (1) and (2), as what is represented in Fig. 3 which were compared with these actual theoretical values.

Number of bonds: N	100
Number of conformations: T	1000
Bond Length: b	3.0
Contour length: ξ	300
Theoretical $\langle Q^2 \rangle$	900
Numerical $\langle Q^2 \rangle$	842
Theoretical $\langle R_g^2 \rangle$	150
Numerical $\langle R_g^2 \rangle$	148

Table 1 It shows the parameters used to create a database of a simulation, the numerical $\langle Q^2 \rangle$ and $\langle R_g^2 \rangle$ derived from it, and its actual theoretical values.

It is worth highlighting that the numerical values are near the theoretical ones as expected which assures that the simulation is properly done and that the fundamental basis of the FJC model works.

2.3 Metropolis Monte Carlo (MC) algorithm

The elastic behavior of the molecule is studied using a Metropolis MC process. Analytically, the entropy of a molecule acts as a driving force toward closed conformations and coagulation of a polymer. The magnitude of the force-extension in the x-axis,

$$|\vec{Q} \cdot \vec{u}_x| = nb \left(\coth\left(\frac{Fb}{k_B T}\right) - \frac{k_B T}{Fb} \right), \tag{6}$$

depends on the product of the force and bond length, Fb, divided by the Boltzmann constant times the temperature, which is set to 1 for the simulation, $k_B T = 1$.

For the Metropolis MC algorithm performance, a random polymer is first generated, and a random bond is proposed as a new bond in a random position as well. If the new bond has a lower or equal potential energy, $V = \vec{F} \cdot \vec{Q}$, so $\Delta V = V_{new} - V_{old} \leq 0$ it is accepted. Otherwise, a new random uniform number, ζ, is computed and if it fulfills the condition, $\zeta < \exp\left[\frac{-\Delta V}{k_B T}\right]$, the new bond is accepted.

For this work, the simulation is run until 1000 new bonds are accepted force in the x-axis ranges from values between 0.01 to 10. To determine the polymer extension in the x-axis, Q_x, for each force, an average of its value is done for the last 10% (in this study case 100) of conformations, which can be seen as time steps.

3 Results and Discussion

For comparing the theoretical probability density distribution, (3), with the simulated data, it is presented as a histogram in Fig. 5. The histogram of the simulated data is normalized so the density is obtained for each bin, ensuring that the normalization condition of a probability density function is fulfilled.

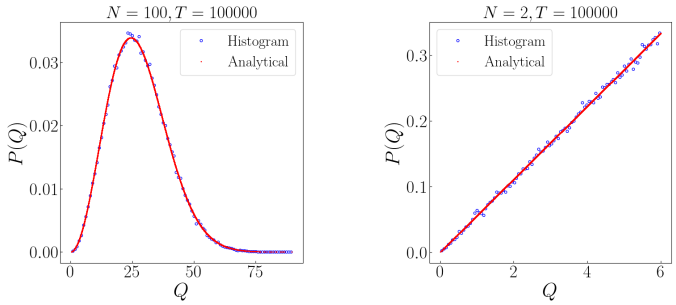

Fig. 5 Representation of the probability density function of Q, $P(Q)$, in function of Q. Both have computed 100000 conformations, the plot at the left used 100 bonds which exhibit the general behavior, and on the right plot, it is 2 bonds which correspond to the singular behavior.

The PDF of Q accurately represents the statistical distribution of the values of Q, with larger n values, when averaged over a larger number of conformations as shown in Fig. 5, as well as, the singular behavior of the PDF for the case two bonds.

In what the Guinier approximation (5) concerns, there is good convergence between theory and numerical results for its valid regime of $kRg \ll 1$ as it can be seen in Fig. 6. After that, the Guinier approximation draws a parabola that rapidly decreases with k, meanwhile, the simulation deviates from this tendency as in opposition decelerates this decrease with k. This difference

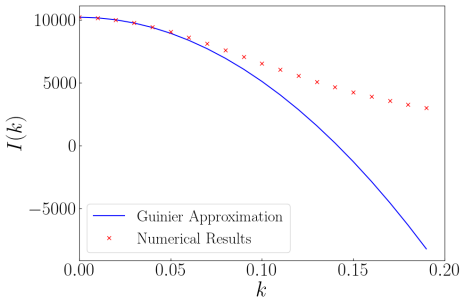

Fig. 6 Representation of the simulated structure factor and the Guinier approximation theoretical values, $I(k)$, in function of k into the interval $[0, 0.2]$, over 100 conformations and for $N = 100$.

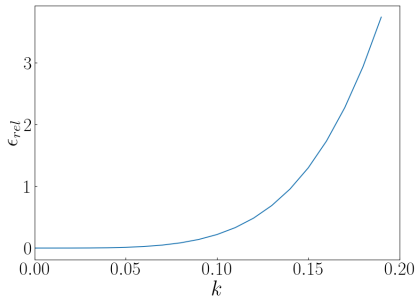

Fig. 7 Representation of the relative error, ε_{rel}, between Guinier approximation and simulated structure factor, $I(k)$ results for $N = 100$ bonds and $T = 100$ conformations.

can be easily noticed in Fig. 7 which represents the relative error between these two quantities in the function of k. Furthermore, it can be seen that beyond $k = 0.05$, the error increases rapidly.

On top of that, it is possible to estimate R_g^2 from equation (5) by doing a linear regression in function of k^2, for the valid region. The range used for k values based on prior observations is set to be below 0.04, where the error between theory and practicals is lower. Once, everything is done the resulting value of R_g^2 is 142 ± 1 (a. u.). It seems to be an acceptable value as the relative error with the theoretical, $R_g^2 = 150$ (a. u.), has a relative error of 5.6%, although the values are incompatible. This result could be improved by sampling more in this region but it would lastly depend on the results of the experimental instrument and the error would increase.

Overall, there is a common occurrence across most of the results that the convergence of the FJC model with numerical results is better when a larger number of conformations, T is taken into account. Therefore, the analytical results of the FJC do not hold much accuracy, unless they are averaged over a large enough sample space to compute the statistics. This serves as a reminder that the statistical results of the FJC model are more useful for describing the general behaviors and tendencies of a macromolecule. When it comes to solving specific cases such as protein folding it is needed to precisely take into account factors, such as monomer interactions and solvent effects, so different models should be used.

3.1 Study of polymer extension when a force is applied

The Monte-Carlo simulation is used to study the entropic elasticity of the molecule so it produces a force-extension, $F(Q_x)$, curve as shown in Fig. 8. The extension-force curve obtained agrees strongly with the theoretical expression (6). One can observe how the molecule begins in a randomly folded state and quickly tends towards a convergence value of extension along the direction of the force where the entropic elasticity and external force are balanced (6). For larger forces, this equilibrium value lies closer to the contour length of the molecule, ζ, a result that can be found in both theory and simulation.

This behavior predicted by the FJC has also been observed for PMAA chains experimentally[4]. When subjected to a force, the extension of a molecule reduces the strength of monomer interactions, and this leads to the physical system being more similar to the FJC. This result, however, is true only for certain domains, since for some molecules an external force can introduce significant deviations from entropic elasticity by inducing phase transitions. The FJC model also does not account for molecules that naturally lie in a curved stable state, causing the theory to conflict with some experimental results[8 7].

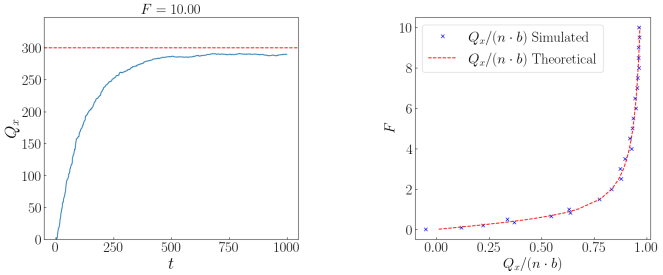

Fig. 8 Representation of the Metropolis Monte-Carlo algorithm simulations for the extension along the x-axis, Q_x. For the plot at left, Q_x is represented across acceptances that can be seen as time steps, t. Whereas, the plot at right represents the Force-Extension curve where the abscissas are actually a mean of the last 10% Q_x values for each force, F.

4 Conclusion

The measures of extent predicted by the FJC model are generally more accurate when computed in larger sample spaces, where a greater number of conformations is taken into account. For chains longer than 10 bonds, the linear trends predicted by the model (1), (2) are followed closely by the data, shown in Fig. 3.

The probability distribution follows a similar behavior as the probability density function accurately describes the distribution of $\sqrt{Q^2}$ for the polymer. The Guinier approximation appears to deviate from the theory more than the other measures, but this happens for larger k values as expected since its derivation assumes that $R_g k << 1$. Thus, the accuracy of the theory is acceptable within this bound.

While there is a good overall convergence of the FJC with numerical data, there is one point for criticism of the model, its fundamental assumption of polymer bonds having no preference for orientation. It is not realistic as the inter-monomer and envi-

ronmental interactions are strong driving effects in phenomena, e.g., protein folding, and DNA translation and replication. The simplification of these phenomena helps provide an analytically and computationally easier model. The addition of constraints to the models developed using simplifications can help them more closely approach reality. Additionally, the present model predicts experimentally measurable quantities like the structure factor, entropic elasticity, and characteristic rigidity with good accuracy, when after introducing constraints on bond angles.

One direction in which the generation of molecules in the FJC model can be improved is by modifying the random walk to be self-avoiding since monomers cannot overlap in reality. While this comes at a significantly heavier computational cost, it is a problem that quantum computing is well suited for, and Micheletti et. al.[3] have developed a formalism for handling this on quantum annealing machines.

Notes and references

1 Cantor, C. R. and Schimmel, P. R. (1980). Biophysical chemistry - part i: The conformation of biological, macromolecules. *Journal of Polymer Science: Polymer Letters Edition*, 18(9):643–644.

2 Chen, G., Xian, W., Wang, Q., and Li, Y. (2021). Molecular simulation-guided and physics-informed mechanistic modeling of multifunctional polymers. *Acta Mechanica Sinica*, 37(5):725–745.

3 Micheletti, C., Hauke, P., and Faccioli, P. (2021). Polymer Physics by Quantum Computing. *Phys. Rev. Lett.*, 127(8):80501.

4 Ortiz, C. and Hadziioannou, G. (1999). Entropic elasticity of single polymer chains of poly(methacrylic acid) measured by atomic force microscopy. *Macromolecules*, 32(3):780–787.

5 Rubinstein, M. and Colby, R. H. (2003). *Polymer Physics*. Oxford University Press.

6 Saleh, O. A. (2015). Perspective: Single polymer mechanics across the force regimes. *The Journal of Chemical Physics*, 142(19):194902.

7 Smith, S. B., Cui, Y., and Bustamante, C. (1996). Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. *Science*, 271(5250):795–799.

8 Smith, S. B., Finzi, L., and Bustamante, C. (1992). Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads. *Science*, 258(5085):1122–1126.