National Technical University of Athens
Interdisciplinary Interdepartmental Postgraduate Program (IIPP)
"Materials Science and Technology"

Generation of polymeric melt models with predefined molecular-weight distributions employing Monte Carlo atomistic simulations

Master's thesis

by

DIMITRIOS - PARASKEVAS E. GERAKINIS
Engineer of Applied Mathematical and Physical Sciences

Supervisor: Doros N. Theodorou
Professor, N.T.U.A.

Athens, July 2022

National Technical University of Athens
Interdisciplinary Interdepartmental Postgraduate Program (IIPP)
"Materials Science and Technology"

Generation of polymeric melt models with predefined molecular-weight distributions employing Monte Carlo atomistic simulations

Master's thesis

by

DIMITRIOS - PARASKEVAS E. GERAKINIS
Engineer of Applied Mathematical and Physical Sciences

Supervisor: Doros N. Theodorou
Professor, N.T.U.A.

Approved by a three-member committee on July 1st, 2022.

(Signature)	*(Signature)*	*(Signature)*
.....	.....	.....
Doros N. Theodorou	Evagelia Kontou	Costas A. Charitidis
Professor, N.T.U.A.	Professor, N.T.U.A.	Professor, N.T.U.A.

Athens, July 2022

(Signature)

.....

Dimitrios - Paraskevas E. Gerakinis
Engineer of Applied Mathematical and Physical Sciences, N.T.U.A.

National Technical University of Athens
Interdisciplinary Interdepartmental Postgraduate Program (IIPP)
"Materials Science and Technology"

Acknowledgements

I prepared the present thesis as member of the Computational Materials Science and Engineering Group (Co.M.S.E.) in School of Chemical Engineering at N.T.U.A. Now that it is completed I want to thank those who assisted me bringing it to fruition.

First of all, I wish to express my most heartful gratitude to my supervisor, Group leader and Professor, Dr. Doros Theodorou for entrusting me to be involved in such an interesting and rich field. Also, his constant and openhandedly offered support, the trust he showed to have in me and his immense knowledge and experience helped me to feel secure and never to lose heart, even at times of long holdups. His solid grasp on even the most intimidatingly abstruse concepts was deeply inspiring. I must also thank him for assisting me with completing the essential for this thesis proof of the new Metropolis criterion. That proof is his work!

My sincere gratitude also goes to the post-doctoral researcher, member of the Group, Dr. Stefanos Anogiannakis for his constant help and advice and our enlightening conversations. Thanks to our way-too-extended correspondence and his insightful notes I successfully managed to navigate my way through labyrinthine directory paths and never-ending code lines.

Abstract

Possessing control over adjusting at will the molecular-weight/length/degree-of-polymerization distribution of a polymeric material is extremely important in applications. This is manifested de facto by the numerous contemporary scientific articles proposing experimental processes for controlling this distribution. Imposing a desired distribution on the system permits tuning the polymer's physical and rheological properties, from mechanical strength and processability to many aspects of block copolymer microphase behavior, thus allowing one to address different specification demands. Yet, the respective literature of computational techniques unfortunately cannot be considered as rich as the technical one.

Our article of reference was: "*Variable Connectivity Method for the Atomistic Monte Carlo Simulation of Polydisperse Polymer Melts*" by Pant and Theodorou (1995), Macromolecules, 28(21). Drawing inspiration from their themodynamic analysis of the $[N_{\rm ch}nPT\boldsymbol{\mu}^*]$ ensemble, that can describe polydisperse polymeric systems, we attempted –on the basis of a Monte Carlo atomistic model employing connectivity-altering moves– to develop a new Metropolis criterion that would allow us to force the system meet selected arbitrary degree-of-polymerization distributions.

The program we started from could already realize the uniform/flat distribution within a range of acceptable degree-of-polymerization values. The criterion we succesfully developed permitted us to realize further the: (i) truncated Gaussian, (ii) truncated bimodal Gaussian, (iii) truncated double Gaussian and (iv) Schulz-Zimm distribution. More importantly, in addition to these analytical distributions we also successfully rendered simulations with tabulated, non-analytical distributions, provided as sets of points through files.

The new criterion is not expressed in terms of ratios of Boltzmann-like factors of chemical energies, derived from the relative chemical potential functions proposed by the article of reference. Instead, it is much simpler, less time consuming computationally, and it only needs knowledge of the unnormalized target distribution. We implemented it in the simulation code and obtained excellent results from all tests we performed on melts of linear polyethylene.

Finally, following the spirit of the article of reference, we introduce the respective chemical potentials of the newly proposed distributions just for thoroughness. The case of the tabulated distributions is not included in this as unnecessary and not easily realizable.

Keywords

Monte Carlo, MC, Markov Chain Monte Carlo, MCMC, atomistic simulation, molecular-weight distribution, Metropolis criterion, polymers, polyethylene, PE, polydispersity.

Περίληψη

Η δυνατότητα ελέγχου και ρύθμισης της κατανομής μοριακού βάρους/μήκους/βαθμού πολυμερισμού ενός πολυμερικού υλικού είναι εξαιρετικά σημαντική για τις εφαρμογές. Αυτό δηλώνεται de facto και από τον μεγάλο αριθμό των προσφάτως δημοσιευμένων επιστημονικών άρθρων που προτείνουν πειραματικές διαδικασίες για τον έλεγχο αυτής της κατανομής. Η ρύθμιση μιας επιθυμητής κατανομής για το σύστημα μας επιτρέπει την ρύθμιση των φυσικών και ρεολογικών χαρακτηριστικών του πολυμερούς, από την μηχανική αντοχή και επεξεργασιμότητά του έως πολλά στοιχεία για την συμπεριφορά της μικροφάσης συμπολυμερούς block, επιτρέποντας έτσι την ικανοποίηση διαφόρων απαιτήσεων περί προδιαγραφών. Κι όμως, η αντίστοιχη βιβλιογραφία υπολογιστικών τεχνικών δυστυχώς δεν μπορεί να θεωρηθεί το ίδιο πλούσια.

Το άρθρο αναφοράς μας ήταν το: "*Variable Connectivity Method for the Atomistic Monte Carlo Simulation of Polydisperse Polymer Melts*" των Pant και Theodorou (1995), Macromolecules, 28(21). Αντλώντας έμπνευση από την θερμοδυναμική ανάλυσή τους πάνω στο στατιστικό σύνολο $[N_{\mathrm{ch}} nPT\boldsymbol{\mu}^*]$, το οποίο δύναται να περιγράψει πολυδιεσπαρμένα, πολυμερικά συστήματα, επιχειρήσαμε –στη βάση ενός ατομιστικού μοντέλου Monte Carlo το οποίο χρησιμοποιεί κινήσεις μεταβολής της συνδεσιμότητας– να αναπτύξουμε ένα νέο κριτήριο Metropolis που θα μας επέτρεπε να υποχρεώσουμε το σύστημα να ακολουθήσει την οποιαδήποτε επιλεγμένη κατανομή βαθμών πολυμερισμού.

Το αρχικό μας πρόγραμμα μπορούσε από πριν να υλοποιήσει την ομοιόμορφη κατανομή εντός ενός εύρους αποδεκτών βαθμών πολυμερισμού. Το κριτήριο που αναπτύξαμε μας επέτρεψε επιτυχώς να υλοποιήσουμε περαιτέρω: (1) την αποκομμένη Gaussian, (2) την αποκομμένη bimodal Gaussian, (3) την αποκομμένη double Gaussian και (4) την Schulz-Zimm κατανομή. Πέραν αυτών των αναλυτικών κατανομών, πραγματοποιήσαμε, το ίδιο επιτυχημένα, προσομοιώσεις με πινακοποιημένες, μη αναλυτικές κατανομές που μας δόθηκαν ως σημεία μέσω αρχείων.

Το νέο κριτήριο δεν εκφράζεται με λόγους όρων Boltzmann χημικών ενεργειών που θα μπορούσαν να εξαχθούν από τα σχετικά χημικά δυναμικά που προτείνονται στο άρθρο αναφοράς. Αντ' αυτού, είναι πολύ πιο απλό, καταναλώνει λιγότερο υπολογιστικό χρόνο και χρειάζεται μονάχα γνώση της μη κανονικοποιημένης, επιθυμητής κατανομής. Το εντάξαμε στον κώδικα των προσομοιώσεων και πήραμε εξαιρετικά αποτελέσματα

από όλους τους ελέγχους που πραγματοποιήσαμε σε τήγματα γραμμικού πολυαιθυλενίου.

Στο τέλος, ακολουθώντας το πνεύμα του άρθρου αναφοράς, παρουσιάζουμε, μονάχα για πληρότητα, τα αντίστοιχα χημικά δυναμικά των υλοποιημένων κατανομών. Τα χημικά δυναμικά των πινακοποιημένων κατανομών δεν περιλαμβάνονται, καθώς η εύρεσή τους θεωρείται περιττή και όχι εύκολα υπολογίσιμη.

Λέξεις Κλειδιά

Monte Carlo, MC, Monte Carlo Αλυσίδα Markov, MCMC, ατομιστικές προσομοιώσεις, κατανομή μοριακών βαρών, κριτήριο Metropolis, πολυμερή, πολυαιθυλένιο, PE, πολυδιασπορά.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Polymers are chain molecules composed of many repeating, chemical units, called monomers, that are covalently bound to their neighboring units. Depending on the complexity of their repeat units, polymers can range from polyethylene (PE) formed by simple methylene (CH_2) monomers to the deoxyribonucleic acid (DNA) double helix consisting of very complex nucleotides to proteins consisting of aminoacids, with the name of the polymer stemming from the repeat unit. Additionally to the sequence of the monomers in the polymer, their molecular microstructure also determines the physical properties of the material, e.g., simple linear as well as long and short side-branched molecules tend to be thermoplastics; they are soft and can be melted and recast almost indefinitely. Examples of thermoplastics are poly(methyl methacrylate) (PMMA), which is also called acrylic glass or plexiglass, poly(propylene) (PP), and poly(ethylene) (PE). Polyethylene and isotactic polypropylene, depending on the density, tend to solidify as semi-crystalline thermoplastics due to their simple monomer constitution, regular stereostructure, and mostly linear architecture.

A polymeric liquid consisting of only one type of chain molecules is known as a polymer melt. The static and dynamic properties of a polymer melt are different from those of a liquid of small molecules. An important feature of polymer chains in a melt is that they can adopt a large number of conformations. Information about different conformations of chain molecules in the melt can be obtained from the study of overall chain dimensions, quantified by the radius of gyration and the end-to-end distance.

It is important to gain theoretical insight into structure-property relations in the material. Theoretical approaches with atomic precision provide knowledge of the microscopic structure of the material. Furthermore, since polymerization can be a complicated and sluggish process, theoretical calculations offer a rapid and more efficient way to generate different architectures and molar mass distributions,

and predict their consequences on properties beforehand. Especially tiny changes in the degree of polymerization of the chain or in the monomer can be applied without much effort by modifications in the theoretical approach, in contrast to a completely new polymerization experiment. Often we need results under conditions which are challenging to obtain in an experimental setup. In a theoretical approach, we only need models, which are valid under those conditions.

Theoretical approaches can be divided into phenomenological models, theoretical models and simulations. Phenomenological models are based on experimental results and are fitted to them. These models have the advantage of being fairly easy to handle numerically. Valid models can be deduced from experimental results and provide a good description of phenomena. This, of course, means that a phenomenological model is only valid for the property onto which it is adjusted [18]. Theoretical models are derived from fundamental principles of the molecular and continuum sciences, aided by the introduction of judicious approximations. These approximations allow casting the model as a set of algebraic, differential and/or integral equations that are relatively easy to solve, analytically or numerically. The nature of the material described enters the model in the form of parameters. These parameters are often fitted to experimental data. Theoretical models are very powerful in describing broad categories of materials with little computational expense and provide valuable understanding of the physics behind observed properties.

A simple theoretical model for the conformation of polymer chains in a bulk melt is the Freely-Jointed Chain (FJC) model which represents the chain as a random walk in three-dimensional space. The description of a chain can be improved further by invoking models of Freely-Rotating Chains (FRC) with fixed bond lengths and bond angles. The worm-like chain model also takes the "persistence" length into account. The "persistence" length is defined as the length scale over which the polymer is flexible. Shorter polymers tend to behave like rigid rods. All of these models have the disadvantage of not being able to predict any bulk behavior, since they all neglect inter-molecular interactions between chains. Theoretical models of multichain systems that include inter-molecular interactions have been developed (e.g., models leading to equations of state, such as Sanchez-Lacombe and the Statistical Associating Fluid Theory (SAFT), the Polymer Reference Interaction Site Model (PRISM), etc.) Due to the approximations they invoke, however, they do not provide fully satisfactory or consistent predictions for properties.

To overcome this downside, one needs to do computer simulations. These are devided in the broad categories of molecular dynamics (MD) and Monte Carlo (MC) simulations, with the former being a simulation method to analyze the dynamics of atoms and molecules. This is done by calculating the trajectory of the particles in

space by numerically solving Newton's equations of motion. Forces between particles are calculated using interatomic potentials and force fields. The forces are used to compute the accelerations, which are used for the calculation of the velocities and finally the new positions of all particles. This is repeated for every time step to obtain (semi)continuous trajectories of the particles. At every time step one can apply spacial boundary conditions, temperature and pressure control, if needed.

Since the atoms and molecules travel along the real trajectories, slow dynamics require tremendous computing effort to sample the relevant part of the configuration space with MD simulations. Especially long chain polymers intertwine like spaghetti and therefore often need a very long time (e.g., ms to min) to equilibrate in the melt state.

These results need overlong or even unaffordable simulation times to reach statistical convergence. To decrease the simulation effort, we can reduce the number of simulated molecules. This can work in simulations of short chains. Long chains, on the other hand, consist of so many atoms, that even few molecules result in a large simulation (e.g., protein folding of one polymer). In this case, another possibility to decrease simulation time is to use coarse-grained models and combine atoms into so-called pseudo-atoms or united atoms. This results in fewer particles and smoother effective interaction potentials, and therefore in shorter simulation times. United atoms can range between one methyl group in a PE chain to one pseudo-atom per nucleotide in a double helix biopolymer. Any interatomic potentials and force fields need to be adapted to the pseudo-atom. Since MD simulations offer the only possibility to predict dynamical properties, it is sometimes required to run them despite long computing efforts. If dynamics is not of interest, however, MC simulations can reach statistical convergence much faster. In turn, well-equilibrated configurations arrived at by MC can serve as starting points for dynamical simulations.

The essence of the MC algorithm is to create a Markov chain of states, which should be distributed according to a prescribed equilibrium probability density, dictated by an equilibrium ensemble of statistical mechanics. If a system of given mass, spatial extent, and temperature is considered, the probability density is given by the canonical ensemble. The probability of finding the system in a microscopic state is proportional of the Boltzmann factor $e^{-\beta H}$, with β being $1/(k_b T)$, T the absolute temperature and k_B Boltzmann's constant, and H being the Hamiltonian (total energy as a function of the microscopic degrees of freedom). A Markov chain describes a sequence of transitions between states, realized in a long succession of Monte Carlo steps. Which state the system will adopt in each step depends on the state attained in the previous step. Transitions between states are attempted, and accepted or rejected according to transition probabilities which must satisfy certain

requirements. The acceptance probability depends on the equilibrium probabilities of the old and the new state and on the probabilities of attempting the transition and inverse. The attempted move can be as simple as displacing one atom or molecule, but may also be more complex and may be specially tailored to the system, e.g., an intra-molecular move aiming at rearranging the internal conformation of a chain. In simulations of long polymer chains, it is possible to design connectivity-altering moves, which cut and reconnect chains to massively decrease the simulation time required for equilibration, due to much more efficient sampling. Additionally, this opens the possibility to simulate polydisperse systems, since we can implement transitions which shorten or lengthen chains. Simulating in a semigrand canonical ensemble by introducing a chemical potential depending on the chain length, allows us to adjust the chain length distribution [13, 20].

Generally, MC simulations can use the same interatomic and external potentials as MD simulations. The same is true for modifications to increase simulation efficiency such as coarse-grained models with united atoms. Since these interatomic potentials and force fields only depend on the specific atoms or pseudo-atoms in case of coarse-grained models, different polymers can easily be formed and simulated, rendering highly valuable MC simulations for calculating numerous properties. Additionally, with access to a high-performance-computing cluster, simulations can be efficiently carried out for a large number of conditions simultaneously.

Since the configurations created in the MC simulation do not correspond to real time dynamics but are just used for statistical averaging, this method is not suited for directly calculating a dynamic property. Combining MC and MD simulations can yield advantages for both methods, e.g., fast statistical convergance to equilibrium and subsequent dynamic properties. In addition, MC can be used for the efficient calculation of free energy profiles with respect to selected (usually slowly evolving) variables or "order parameters"; these are essential in the calculation of rate constants for dynamical processes that occur as (sequences of) infrequent events. For static calculations and for equilibration before dynamical simulations, MC simulations are the best possible method, but their computational complexity can be significantly higher than for MD simulations, because of more sophisticated moves in MC than in MD, which just solves an equation of motion. The complexity of polymers which needs to be simulated dictates the complexity of the MC computer code. In one of the simplest cases, we can restrict polymers to be linear without any branches. If we additionally simplify a monomer to one single united atom, we can store a molecule in one array. Even using such simple models, an MC simulation program may be very complex [18].

Simulating polymers is a challenging task, as evidenced by the literature. Poly-

mers are long chain molecules which demonstrate relevant phenomena across many time and length scales. Consequently, the dynamical behavior of polymers is substantially different from that of a simple Newtonian liquid, exhibiting both liquidlike and solidlike characteristics. Due to their complex nature, it has been necessary to invoke coarse-grained approximations. Coarse-graining a polymer screens out the short time and length scales by representing multiple light (i.e. hydrogen) and a few heavier atoms (i.e. carbon) into one effective unit. For polymer melts, where the most important phenomena are usually of an entropic origin, these coarse-graining techniques and associated assumptions are found to be valid and necessary [16].

In this direction, with the huge development of computers nowadays, computer simulation techniques have become valuable tools of fundamental and basic research in polymer science.

Even a single chain exhibits a much more complicated structure than simple atomic liquids; from the scale of a single chemical bond ($\sim 1\,\text{Å}$) to the persistence statistical length ($\sim 10\,\text{Å}$) to the coil radius of the chain ($\sim 100\,\text{Å}$). Intra-molecular correlations and local packing of chains in the bulk exhibit features on the length scale of bond lengths and atomic radii. The statistical, Kuhn segment length of a typical synthetic randomly coiled polymer is on the order of 10 Å and can be considerably larger for very stiff polymers.

The spread of time scales characterizing polymeric dynamics is also wide. While in simple fluids at states far from phase transitions all fluctuations decay on time scales of a few picoseconds ($10^{-12}\,\text{sec}$), the behavior of polymers is substantially different, since motions occur on many and very different time scales. For example fast motions such as vibrations of C−C bonds may take times of the order of $10^{-14}\,\text{sec}$ or even less, while the longest relaxation time, i.e. the time required for a chain to diffuse by a length commensurate with its size and thus forget its previous conformation, scales with chain length N to $N3.4$ in long-chain melts and is typically in the ms-min range.

Computer simulations can play a significant role providing essentially exact results for problems in statistical mechanics, that otherwise would be soluble only by approximate theories. From this point of view, computer simulations offer the possibility of directly testing theories. From another point of view, the results of computer simulations may also be compared to those of lab experiments. This has a dual character. On one hand, it is a test of the model used in the computer simulations. If simulation results do not agree with experiment, the model may be improved by adjusting its parameters (e.g., force field parameters). On the other hand, if the model is proved to be a good one, the simulations can offer insights to the experimentalists and assist in the interpretation of the new results [6]. This

dual role of the simulation, as a bridge between models and theoretical predictions on one hand, and between models and experimental results on the other, is clearly illustrated in Figure 1.1. Due to their connecting role, simulation techniques are often called "computer experiments".

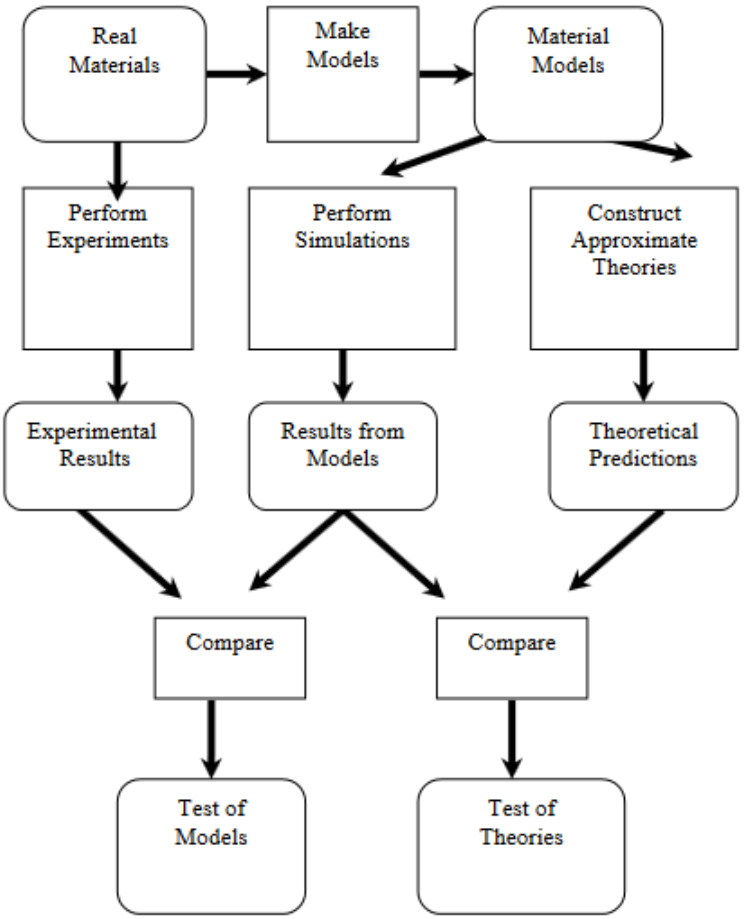

Figure 1.1: Connection between experiment, simulation and theory [2, 6]

Chapter 2

Theory

2.1 Polymer Physics

Polymers are long molecules composed of many elementary units called monomers. The monomers of the same polymeric chain are connected with each other with covalent bonds. The number of monomers that constitute the chain is referred to as the "degree of polymerization". This number is typically quite large, ranging from 10^2 up to 10^6 and some polymers of length up to 10^{10} can be found in the natural world [22]. One important feature of a polymeric system is the architecture of the chains constituting it; the branching architecture: linear, star, Cayley trees, etc. plays an important role in controlling the properties of the system.

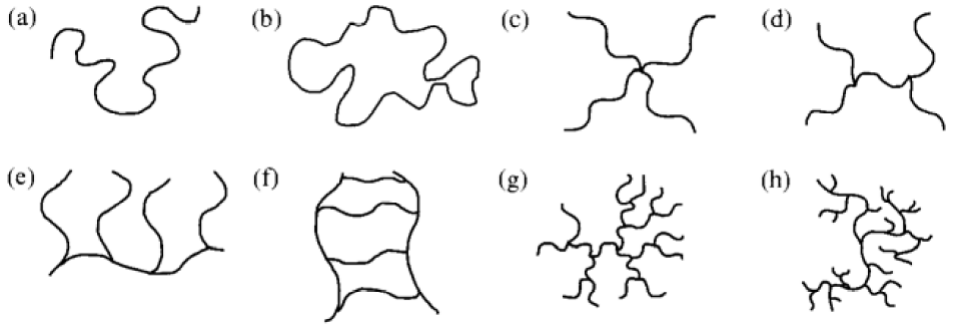

Figure 2.1: Different types of polymeric architectures: (a) linear, (b) ring, (c) star, (d) H, (e) comb, (f) ladder, (g) dendrimer, (h) randomly branched [17]

Some of the polymer's chemical characteristics, including the molecular structure, the degree of polymerization and chemical composition are fixed after the polymerization or after being synthesized. The conformation of the chains, determined by the relative positions of their monomers, continuously changes in the course of thermal fluctuations. The conformation of a linear chain comprised of m bonds

can be described by a set of $(m+1)$ position vectors $\boldsymbol{R}_k$, $k=1,\ldots,(m+1)$ of its backbone atoms (or its monomers' centers of mass). Alternatively, its conformation can be represented by a set of bond vectors $\boldsymbol{r}_i$, where:

$$\boldsymbol{r}_i = \boldsymbol{R}_{i+1} - \boldsymbol{R}_i, \quad i = 1, 2, \ldots, m \tag{2.1}$$

The end-to-end vector of the chain is defined as the resultant vector of all m bond vectors or as the difference between the two terminal position vectors of the chain. Whether we use one or the other definition depends on whether we know the bond vectors or the position vectors:

$$\boldsymbol{R}_\mathrm{e} = \sum_{i=1}^{m} \boldsymbol{r}_i = \boldsymbol{R}_{m+1} - \boldsymbol{R}_1 \tag{2.2}$$

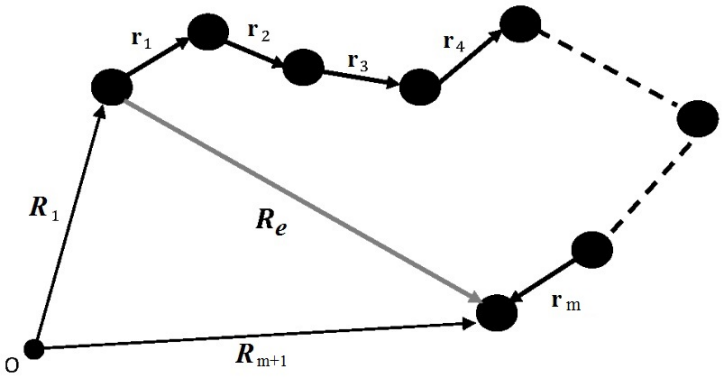

Figure 2.2: Bond, position and end-to-end vectors of a linear polymeric chain [1]

The average end-to-end vector of an isotropic, monodisperse collection of chains of m backbone bonds is zero, since there is no preferred direction in this ensemble. The simplest non-zero average we use instead is the mean-square end-to-end distance:

$$\left\langle \boldsymbol{R}_\mathrm{e}^2 \right\rangle \coloneqq \langle \boldsymbol{R}_\mathrm{e} \cdot \boldsymbol{R}_\mathrm{e} \rangle = \left\langle \left(\sum_{i=1}^{m} \boldsymbol{r}_i \right) \cdot \left(\sum_{j=1}^{m} \boldsymbol{r}_j \right) \right\rangle = \sum_{i=1}^{m} \sum_{j=1}^{m} \langle \boldsymbol{r}_i \cdot \boldsymbol{r}_j \rangle = \left\langle (\boldsymbol{R}_{m+1} - \boldsymbol{R}_1)^2 \right\rangle \tag{2.3}$$

where $\langle \boldsymbol{R}_\mathrm{e}^2 \rangle$ is often considered as a measure of the size of a linear chain. This quantity is not well defined for polymers with other architectures, since sometimes there are either too many ends, e.g., dendrimer polymers, or no ends at all, e.g., ring polymers. An alternative way to characterize the size of a polymer with any topology is to evaluate the radius of gyration, since all objects possess this quantity. The square of the radius of gyration is defined as the average square distance between atoms (or monomers) in a given conformation and the polymer's center of mass, R_cm:

$$\boldsymbol{R}_{\mathrm{g}}^2 = \frac{1}{m+1}\sum_{k=1}^{m+1}\left(\boldsymbol{R}_k - \boldsymbol{R}_{\mathrm{cm}}\right)^2 \tag{2.4}$$

The position vector of the center of mass of the polymeric chain is the mean-number resultant vector of all its monomer position vectors:

$$\boldsymbol{R}_{\mathrm{cm}} = \frac{1}{m+1}\sum_{k=1}^{m+1}\boldsymbol{R}_k \tag{2.5}$$

By substituting $\boldsymbol{R}_{\mathrm{cm}}$ in eq.(2.4) for eq.(2.5) we get another definition for the square of radius of gyration:

$$\boldsymbol{R}_{\mathrm{g}}^2 = \frac{1}{m+1}\sum_{k=1}^{m+1}\left(\boldsymbol{R}_k - \frac{1}{m+1}\sum_{j=1}^{m+1}\boldsymbol{R}_j\right)^2 = \ldots = \frac{1}{(m+1)^2}\sum_{k=1}^{m+1}\sum_{j=k}^{m+1}\left(\boldsymbol{R}_k - \boldsymbol{R}_j\right)^2 \tag{2.6}$$

For polymers, the square radius of gyration is usually averaged over the ensemble of allowed conformations giving the mean-square radius of gyration:

$$\left\langle \boldsymbol{R}_{\mathrm{g}}^2 \right\rangle = \frac{1}{m+1}\sum_{k=1}^{m+1}\left\langle \left(\boldsymbol{R}_k - \boldsymbol{R}_{\mathrm{cm}}\right)^2 \right\rangle = \frac{1}{(m+1)^2}\sum_{k=1}^{m+1}\sum_{j=k}^{m+1}\left\langle \left(\boldsymbol{R}_k - \boldsymbol{R}_j\right)^2 \right\rangle \tag{2.7}$$

A very important value is the ratio of the mean-square end-to-end distance to the mean-square radius of gyration which for an ideal, linear and very long chain or an ensemble of such chains is:

$$\frac{\left\langle \boldsymbol{R}_{\mathrm{e}}^2 \right\rangle}{\left\langle \boldsymbol{R}_{\mathrm{g}}^2 \right\rangle} = 6 \tag{2.8}$$

Another equally important quantity is Flory's characteristic ratio, C_m, which in the case of a single chain of m backbone bonds is equal to:

$$C_m = \frac{1}{m}\sum_{i=1}^{m} C_i' = \frac{1}{m\ell^2}\sum_{i=1}^{m}\sum_{j=1}^{m}\left\langle \boldsymbol{r}_i \cdot \boldsymbol{r}_j \right\rangle \tag{2.9}$$

The main property of ideal chains is that $\left\langle \boldsymbol{R}_{\mathrm{e}}^2 \right\rangle$ is proportional to the product of the number of bonds, m and the square of the bond length, ℓ^2. An infinite chain has a C_i' value for all i given by C_∞. A real chain has cutoff in the sum of C_i' at finite j that results in a smaller C_i'. This effect is more pronounced near chain ends [17].

It is common practice one to calculate the values of function C_K in order to create the graph of $C_K(K)$ (like Figure 2.3), where $K = 1, \ldots, m$:

$$\left\langle \boldsymbol{R}_{\mathrm{e}}^2 \right\rangle = C_K K \ell^2 \Rightarrow C_K = \frac{1}{K\ell^2} \left\langle (\boldsymbol{R}_{i+K} - \boldsymbol{R}_i)^2 \right\rangle_{\forall i:(i+K)\in(1,m+1]} \tag{2.10}$$

This is done by taking under consideration in the calculations phantom subsystems of the system: that is sub-chains inside the existing chains, consisting only of two monomers each ($k = 1$), three monomers ($k = 2$), etc. Combining this practice with the computation of a running average over all conformations realized until that point by the system, leads to employing an enormous ensemble of bond values which gets bigger with every new conformation, eliminating the error fluctuations and returning a very smooth curve for K from 1 up to m.

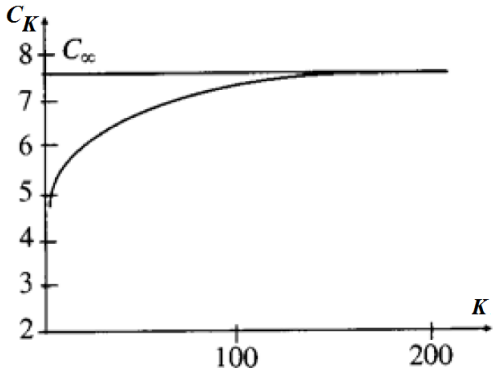

Figure 2.3: Flory's characteristic ratio C_K saturates at C_∞ for long chains [17]

2.2 Polydispersity in Polymers

Polymers are found to have numerous applications in our modern society. Polyethylene, for example, is used in a wide variety of applications; including grocery bags, containers, clothing, transportation, toys, home appliances, engineering plastics and medical applications, among others. Despite its utility, PE is one of the simplest polymers (from a chemical perspective); made of monomers composed of only carbon and hydrogen atoms and its long-chain consisting of k repeating units (ethylenes) is represented as $[-\mathrm{CH_2CH_2}-]_k$ [10, 19].

One distinguishing characteristic of polymers is the presence of very high molecular lengths in their samples, that is the previous k index can be a very big number. Another feature of polymers that differentiates them from other chemical substances is that, unlike small molecules, they do not possess a unique molecular weight, but exhibit a molecular weight distribution (MWD). That is, the range of molecular

weights of the molecular species present in a polymer may cover many orders of magnitude, ranging from a handful of monomers and oligomers up to molecules containing hundreds of thousands of atoms. The entire polymer sample is made up of individual molecules that have a distribution of degrees of polymerization, determined by the particular synthesis method used. If all polymers in a given sample have the same number of monomers, the sample is monodisperse [17]. Such perfection is common in natural processes of, let's say, protein synthesis conducted in human cells, albeit very difficult to accomplish (except for a few oligomers) in the case of synthetic polymers made in labs and the industry. This thesis pertains to the latter.

All polymers comprise molecules that exhibit chemical and physical distributions of many variables; these include molecular weight, branching, steric defects, molecular configuration and preferential chain orientation. The properties and characteristics that we exploit in polymers are controlled by the overall balance of these distributions. There is no such thing as a "pure" polymer [14].

Polymer scientists are well aware of the importance of MWD and that is made clear by the throng of contemporarily published, relevant articles. They typically use a set of shorthand values to describe a polymer's MWD. These values are a series of averages that are increasingly weighted to the longer chains of the distribution. One of the most common molecular weight averages that we encounter is the number average.

The number average molecular weight (M_n) and the number average degree of polymerization (N_n, which interests us in this thesis) are calculated from:

$$M_\mathrm{n} = \frac{\sum M_k N_k}{\sum N_k}, \quad N_\mathrm{n} = \frac{\sum k N_k}{\sum N_k} \tag{2.11}$$

where M_k is the molecular weight of each chain with degree of polymerization k and N_k is the number of chains in fraction k [14].

The polydispersity of a sample is described by its MWD, although its presence can be qualitatively recognized also in the graph of the molecule length distribution. A distribution is shown as n_N, the number fraction (or mole fraction) of molecules containing k monomers each, plotted as a function of molecular weight $M_\mathrm{k} = M_\mathrm{mon}k$ of the molecules (see Figure 2.4) [17], of molecular length or of degree of polymerization k.

The manipulation of a polymer's properties without altering its chemical composition is a major challenge in polymer chemistry, materials science and engineering. Although variables such as molecular weight, dispersity, chemical structure and branching are routinely used to control the physical properties and architecture of polymers, the breadth and shape of the MWD also have profound effect on their

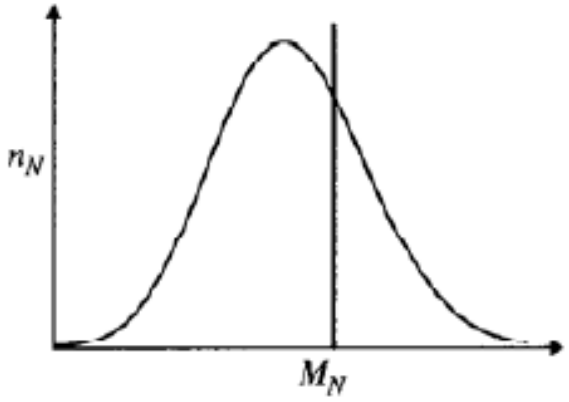

Figure 2.4: Monodisperse and polydisperse molecular weight distributions [17]

properties [5]. The MWD has been described using a variety of probability functions directly measured or modelled mathematically. For example, the Flory-Schulz distribution is a single-parameter function that is typically used to describe the MWDs of polymers obtained from polycondensation reactions. It is useful for modelling step-growth polymerization, which favours short chains and MWDs tailing towards high molar mass. For anionic polymerization processes, in which initiation is much faster than propagation, and termination reactions are nonexistent, the distribution of chain lengths is narrow and is accurately described by the Poisson distribution. Additionally, the shapes of distributions afforded through controlled radical polymerization (such as atom-transfer radical polymerization) can be defined by the Schulz-Zimm distribution (a two-parameter function), because the initiation kinetics and termination events lead to buildup of low-molar-mass polymer chains. Furthermore, Gaussian or logarithmic normal distributions have been used to describe functions that are symmetric about the number-average molar mass [5].

The width and shape of MWDs can significantly affect the properties of polymeric materials and thus are crucial parameters to control. Dispersity (Đ, formerly referred to as "polydispersity index" or PDI) is a measure of the width of a MWD and describes the heterogeneity (or uniformity) of the various chain molecular weights within a polymeric material. While dispersity is a crude parameter for estimating the uniformity of polymer chain lengths, the shape of MWDs provide much more information about the range of chain lengths. For example, two polymeric samples with the same dispersity can have symmetric or asymmetric MWDs, (skewed towards either high or low molecular weight), which can deeply affect material properties [23].

The production of polymers is widespread and long-standing in industrial processes; different applications need different specifications for the polymers. To meet these demands, many performance indices based on polydispersity and average

molecular weight have been proposed. Under these indices, the control objective is commonly dealt with as an optimization problem. Although satisfactory control of performance could be obtained in these studies, these indices cannot fully characterize the entire MWD of polymers. However, the shape of the MWD of polymers is essential in many polymerization processes, such as paints and paper coatings, and the average molecular weight and PDI is unable to reflect the characteristics of the MWD when the distribution is non-Gaussian [24].

MWDs have a substantial impact on a diverse set of polymer physical and rheological properties, from mechanical strength and processability to many aspects of block copolymer microphase behavior. Rheological and morphological properties of polymers exhibit a significant dependence on the shape and symmetry of their component MWDs. The impact of the MWD shape on such a diverse set of polymer properties clearly demonstrates that modulation of the entire distribution of chain sizes is a promising avenue for fine-tuning the function of polymeric materials without the need to change their chemical structure [3]. This correlation is general across all polymers and has motivated the development of many synthetic and processing techniques. Much of the experimental work regarding MWDs has primarily focused on the development of polymerization methods to access polymers with narrow MWDs (referred to as controlled polymerization). Controlled polymerizations have revolutionized the synthesis of advance materials, however, the control they offer does not directly provide tunability for broad MWDs which are advantageous for many applications. In fact, broad distributions remain a staple in industry [21].

2.3 Monte Carlo Simulations

2.3.1 Bayesian statistics

Bayesian statistics, named after Thomas Bayes (1701–1761), refers to a class of methods widely used in the fields of probability and statistics. Bayesian statistics is based on the belief that the uncertainty of unknowns should be quantified by probabilities. One of the most important opinions of Bayesian statistics is that "probability is orderly opinion, and the inference from data is nothing other than the revision of such opinion in the light of relevant new information" as quoted by Edwards, Lindman and Savage (1963) [4] in a paper of theirs. This message delivers a key idea that Bayesian statistics involves updating our knowledge based on the new up-to-date information.

The Bayesian framework is based on a simple rule of probability, which is known as Bayesian formula:

$$P(A|B) = \frac{P(B|A)\ P(A)}{P(B)} \tag{2.12}$$

where A and B are events of interest with probabilities $P(A)$ and $P(B)$, respectively. The conditional probability $P(A|B)$ indicates the probability of observing event A supposing that event B has already been observed.

Bayes' theorem interprets equation (2.12) as follows: event A represents a hypothesis and event B is the observed evidence; the Bayesian rule establishes the relationship between the prior probability of the hypothesis $P(A)$ and the posterior probability of the hypothesis $P(A|B)$ after evidence B is given. Probabilities $P(B|A)$ and $P(B)$ are called "likelihood" and "evidence", respectively. Using such interpretation, the Bayes' theorem can be rewritten as:

$$\text{Posterior probability} = \frac{\text{Likelihood}}{\text{Evidence}} \times \text{Prior probability} \tag{2.13}$$

Prior information is important in Bayesian statistics. It indicates the uncertainty about the quantity of interest before we perform any analysis of the observation. Prior knowledge is usually obtained from previous knowledge, such as experiments, but it could also be derived from a justified belief or intuition. A strong prior provides specific information about the quantity. In such cases, it makes the posterior mainly dominated by the prior and less affected by the data. The stronger the prior is, the more data would be needed for decreasing the impact of the prior.

However, sometimes, we expect a non-informative prior that has less influence on the analysis. To set up a weak prior belief, we usually adopt a prior that is rather flat over the parameter space, e.g., uniform distribution. The good news is that the influence of the prior will decrease as the number of observations increases. Note that this is the case even when we use a very informative prior. The large data set will gradually correct the belief and eventually give us the right posterior even if the prior available is significantly wrong [22].

2.3.2 Markov chain Monte Carlo

Under the term Monte Carlo is encapsulated a plethora of methods that employ generators of (pseudo)random numbers to provide answers for physical and mathematical problems, to make predictions and/or simulations. Monte Carlo was invented by Stanislaw Ulam in the late 1940s. Combined with the Metropolis algorithm, which is the original version of Markov Chain Monte Carlo (MCMC) methods, it is developed slowly until Hastings extended it to a more general case in 1970.

Let $\boldsymbol{\rho}^{\text{eq}}$ be the probability density in the equilibrium ensemble where the MC simulation is carried out, and let us assume that in a given MC step a move is

attempted from an old/prior state to a new/trial state. Let us also write as $a(\text{old} \to \text{new})$ the corresponding stochastic, underlying matrix $\boldsymbol{a}$ of attempt probabilities, which is usually symmetric, i.e. it satisfies:

$$a(\text{old} \to \text{new}) = a(\text{new} \to \text{old}) \tag{2.14}$$

"Stochastic" means that the sum of elements in a column equals 1. Then, applying a Metropolis criterion, the new state can be accepted with probability:

$$p_{\text{acc}}(\text{old} \to \text{new}) = \min\left[1, \frac{\rho^{\text{eq}}(\text{new})}{\rho^{\text{eq}}(\text{old})}\right] \tag{2.15}$$

with $\rho^{\text{eq}}(\text{old})$ and $\rho^{\text{eq}}(\text{new})$ denoting the probabilities of the system to be in the old state or in new one, respectively. The corresponding Markov chain of states generated this way samples asymptotically the probability density distribution $\boldsymbol{\rho}^{\text{eq}}$.

The method of succeeding this, i.e., commencing from a prior probability density distribution and after a number of trials lead it to converge to a preselected posterior one, was developed by Metropolis *et al.* (1953) [12].

Definition. *A Markov chain is a stochastic process, each trial of which meets the next two and only conditions:*

- *The outcome of each trial belongs to a discrete, finite set of possible outcomes, called "state space", and*
- *The outcome of each trial depends solely on the outcome of the preceding trial, in other words the process has a one-step memory of its state*

The power of the MC method in simulating complex physical systems lies in its flexibility in choosing freely the matrix $\boldsymbol{a}$ as long as the requirement $a(\text{old} \to \text{new}) = a(\text{new} \to \text{old})$ is satisfied. One could even experiment with applying totally unphysical moves for the system at hand, as far as the molecular architecture of the constituent molecules and the internal geometry were not ruined. This could speed up the rate with which the system travels around configuration space to a considerable extent. That is particularly advantageous in the case of dense phases of macromolecules that are known to be characterized by a very broad distribution of relaxation times that render their direct MD simulation a formidable task.

An interesting point in the Metropolis MC scheme is that one can replace eq.(2.15) with the following more general condition:

$$\rho^{\text{eq}}(\text{old})\ a(\text{old} \to \text{new})\ p_{\text{acc}}(\text{old} \to \text{new}) = \rho^{\text{eq}}(\text{new})\ a(\text{new} \to \text{old})\ p_{\text{acc}}(\text{new} \to \text{old}) \tag{2.16}$$

also known as "detailed balance" or condition of "microscopic reversibility". Then the Metropolis criterion, eq.(2.16), becomes the Metropolis-Hastings criterion, which is applicable even when the matrix of attempt probabilities (underlying matrix of the Markov chain) is not symmetric:

$$p_{\text{acc}}(\text{old} \to \text{new}) = \min\left[1, \frac{\rho^{\text{eq}}(\text{new})\ a(\text{new} \to \text{old})}{\rho^{\text{eq}}(\text{old})\ a(\text{old} \to \text{new})}\right] \tag{2.17}$$

If, in addition, the elementary move is designed in a set of coordinates that differs from the configuration-space coordinates where the equilibrium probability density $\boldsymbol{\rho}^{\text{eq}}$ was defined, then the previous equation must be modified to account for the Jacobian of transformation $\boldsymbol{J}$ from one coordinate system to the other:

$$p_{\text{acc}}(\text{old} \to \text{new}) = \min\left[1, \frac{\rho^{\text{eq}}(\text{new})\ J(\text{new})\ a(\text{new} \to \text{old})}{\rho^{\text{eq}}(\text{old})\ J(\text{old})\ a(\text{old} \to \text{new})}\right] \tag{2.18}$$

In the simple case of eq.(2.15) (symmetric attempt probability matrix, no Jacobians of transformation), from the elements of the stochastic matrix $\boldsymbol{a}$, the transition probability matrix $\boldsymbol{\pi}$ is defined next according to:

$$\pi_{no} \coloneqq \pi(\text{old} \to \text{new}) = \begin{cases} a(o \to n) & \text{if} \quad \rho^{\text{eq}}(n) \geq \rho^{\text{eq}}(o),\ n \neq o \\ a(o \to n)\dfrac{\rho^{\text{eq}}(n)}{\rho^{\text{eq}}(o)} & \text{if} \quad \rho^{\text{eq}}(n) < \rho^{\text{eq}}(o),\ n \neq o \end{cases} \tag{2.19}$$

$$\pi_{oo} = 1 - \sum_{n \neq o} \pi_{no}$$

and by definition, is also stochastic. Then, as the MC iterations progress, the row vector $\boldsymbol{\rho}_{\text{t}} = (\rho_{\text{t}}(1), \rho_{\text{t}}(2), \ldots, \rho_{\text{t}}(o), \ldots, \rho_t(n), \ldots)$, containing the a priori probabilities of all states after step t of the simulation, converges to the desired equilibrium distribution $\lim_{t\to\infty} \boldsymbol{\rho}_{\text{t}} = \boldsymbol{\rho}^{\text{eq}}$ satisfying:

$$\boldsymbol{\rho}^{\text{eq}} = \boldsymbol{\pi} \cdot \boldsymbol{\rho}^{\text{eq}} \tag{2.20}$$

Eq.(2.20) implies that the equilibrium distribution $\boldsymbol{\rho}^{\text{eq}}$ is an eigenvevtor of the transition probability matrix $\boldsymbol{\pi}$, with the corresponding eigenvalue equal to one.

Due to difficulties associated with chain connectivity, conformational stiffness and excluded volume interactions, early MC simulations were performed on lattice models in discrete space, but soon implementations in continuous space appeared. Today, we can categorize MC moves developed for polymers in a coarse manner into three groups [11]: simple, complex and advanced. The latter is not going to concern us in the context of this thesis.

Simple MC moves include: pivot, reptation, atom identity exchange, concerted rotation (ConRot), intra-chain and inter-chain concerted rotation, endmer rotation, libration or flip, dimer flip, configurational bias, extended configuratonal bias and volume fluctuation [11]. These moves apply only to a single polymer chain; moreover, reptation, end-mer rotation, configurational bias and generalized reptation apply only at chain ends. With these simple moves, only chains up to about 70 monomers (or UAs) long can be equilibrated within reasonable computer time. Special mention should be made of the reptation move, perhaps the very first MC move devised for polymer chains. It is realized by cutting out a monomer from one end of the chain and appending it to the other end. Thus, the move simulates the diffusive (slithering-snake) motion of the chain. In the configurational bias move, on the other hand, one cuts out an entire segment composed of many monomers at one end of the chain and regrows it monomer by monomer in a biased way so as to avoid overlaps with the monomers on the same or nearby chains. The bias introduced in the construction of the new configuration is removed at the final stage of the move when the acceptance criterion is applied, by appropriately modifying it. As far as the local conformation at the interior of a chain is concerned, this is typically equilibrated with the help of the ConRot move, which involves the concerted rotation of an internal segment of the chain, of size equal to 5 monomers.

Complex MC moves induce extensive reconfigurations of large internal sections within one or two polymer chains simultaneously, resulting in drastic repositioning jumps in the configuration space. Initially they appeared in the form of chain breaking moves in MC simulations of polymers on lattices, later they gave rise to the so-called family of "chain-connectivity altering" moves. Following some rigorous schemes, these moves effect large conformational changes at the level of the end-to-end distance of one or two chains at the same time often, however, at the expense of introducing polydispersity in the system. The family of complex MC moves includes: directed internal bridging, directed end-bridging, end-bridging (EB), self end-bridging, fusion-scission, double bridging (DB) and intra-molecular double re-bridging (IDB). Introduction of these moves almost revolutionized the field of the molecular simulation of polymers.

From a technical point of view, and in order not to violate the crucial condition of microscopic reversibility, in a variable connectivity MC move proper care should be taken to: (i) evaluate all possible geometric solutions to the underlying bridging problem, (ii) incorporate appropriate Jacobians in the acceptance criterion (because the solution of the geometric problem is typically carried out in the space of generalized coordinates) and (iii) address both the forward and the reverse problem pertinent to the move [11].

2.4 Connectivity-Altering Moves

2.4.1 End-Bridging move

ConRot is very useful in inducing local configurational rearrangements in atomistic models of chain polymers. It is not sufficient, however, for equilibrating structural features at large length scales, such as the end-to-end distance, $\sqrt{\langle R_e^2 \rangle}$, and the radius of gyration, $\sqrt{\left\langle R_g^2 \right\rangle}$, in systems of long chains.

The need to equilibrate large length-scale structural features led to the development of a "bolder", inter-molecular move based on the bridging construction, which is termed "end-bridging". As shown in Figure 2.5, this move involves two chains. The triplet of skeletal atoms $\{i, j, k\}$ is excised from the chain above, generating two new chain ends, h and l. Next, the bridging construction is invoked to generate a new triplet of atoms $\{i', j', k'\}$, which connects the end p of the chain below to atom h of the above chain. The end result is a dramatic change in connectivity: the chain below (the "attacking" chain) has appended h and the subsequent part of the chain above and left through the newly constructed trimer $\{i', j', k'\}$. The "victim" chain on the other hand, has been shortened, now terminating at l. Although the scheme is illustrated with a trimer, one could clearly have excised a larger internal segment and repositioned it as a bridge of predetermined internal geometry. The bridging construction guarantees that the detailed atomistic geometry of the chains is preserved.

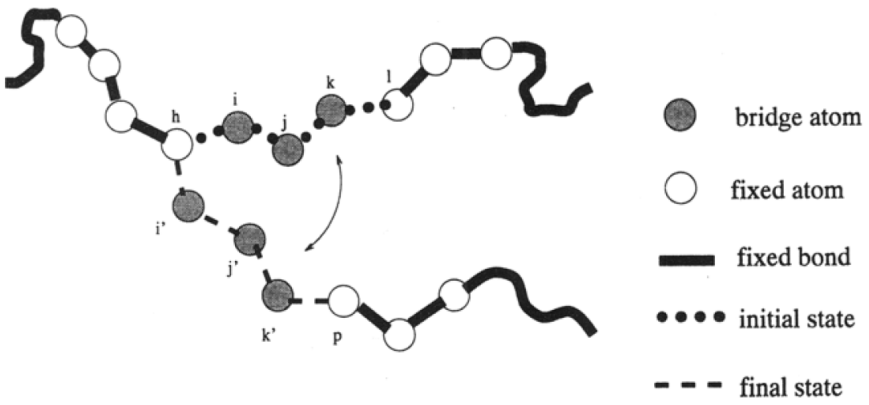

Figure 2.5: Schematic representation of the connectivity-altering End-Bridging move [13]

2.4.2 Double-Bridging move

The End-Bridging move requires a finite degree of polydispersity in order to function. While this is not necessarily a shortcoming in modelling industrial polymers, which are typically polydisperse, an ability to equilibrate strictly monodisperse chain systems is desirable from the point of view of comparing simulation predictions against theory or against experiments on anionically synthesised model systems.

Two decades ago, Karayiannis *et al.* [8] developed two new connectivity-altering moves that are not subject to these shortcomings. The moves are based on the construction of two bridging chain sections (usually trimers) among four properly chosen internal skeletal atoms belonging to one or two chains in the system. The moves have been termed "double bridging" and "intra-molecular double rebridging".

Figure 2.6 presents a schematic of the DB move. An internal skeletal atom i of chain i_{ch} attacks an internal skeletal atom j of chain j_{ch}. The trimer $\{j_a, j_b, j_c\}$ adjacent to j is excited and a new trimer bridge $\{j'_{\mathrm{a}}, j'_{\mathrm{b}}, j'_{\mathrm{c}}\}$ is formed, connecting i and j. At the same time, the internal skeletal atom j_2, which is adjacent to the excised trimer on chain j_{ch}, attacks internal skeletal atom i_2, four skeletal atoms away from i on i_{ch}; the trimer $\{i_{\mathrm{a}}, i_{\mathrm{b}}, i_{\mathrm{c}}\}$ is excised and a new trimer bridge $\{i'_{\mathrm{a}}, i'_{\mathrm{b}}, i'_{\mathrm{c}}\}$ is built between j_2 and i_2.

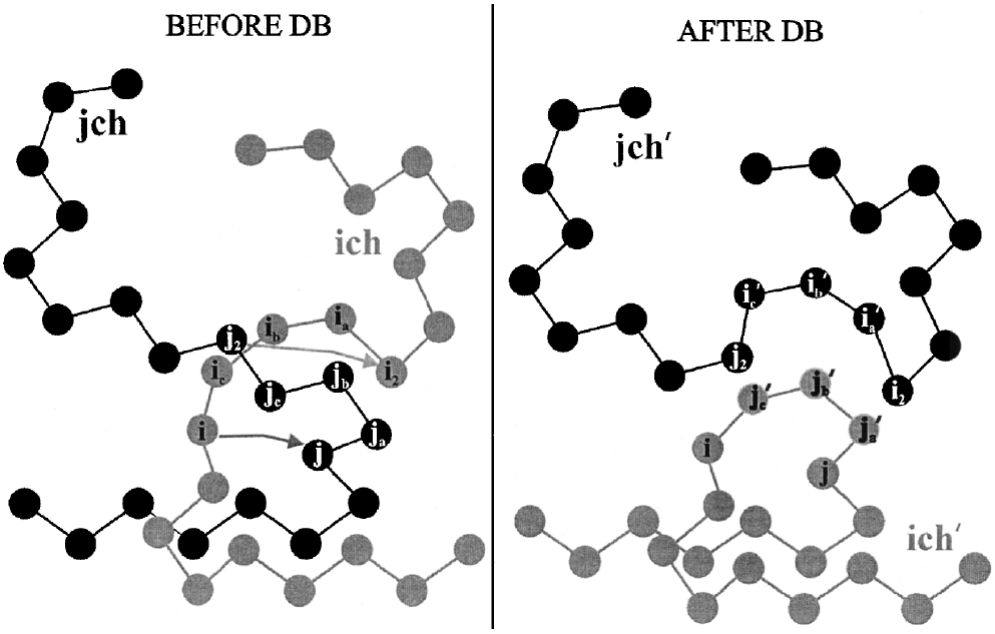

Figure 2.6: Schematic representation of the connectivity-altering Double-Bridging move [8]

The chains obtained after the move, i'_{ch} and j'_{ch}, have completely different conformations from those of i_{ch} and j_{ch}. Given two skeletal atoms i and j along the

backbones of chains i_{ch} and j_{ch}, respectively, there are in general four different ways in which DB can be performed, depending on which two of the four trimers adjacent to j and i are excised. In a monodisperse system, if i and j are approximately positioned relative to the ends of i_{ch} and j_{ch}, one of these four DB moves preserves monodispersity.

2.5 Semigrand $[N_{\mathbf{ch}}nPT\mu_1^*\mu_2^*\ldots\mu_M^*]$ ensemble

Clearly, an end-bridging move modifies the degrees of polymerization of the two participating chains. Thus, the MC simulation must be cast in an ensemble that allows chains of various degrees of polymerization to be present, i.e. some polydispersity in the polymer. An appropriate "semigrand" ensemble was formulated by Pant and Thedorou (1995) [13]. They considered the polymer as a mixture of various chain species. Starting from the differential expression for the Helmholtz energy of such a mixture and introducing variable substitutions and Legendre transformations, they arrived at a partition function, describing the mixture in terms of the following thermodynamic variables: the total number of chains in the system, N_{ch}, the total number of monomers (or UAs) in the system, n, the temperature, T, the pressure, P, and a set of relative chemical potentials, μ_k^*, for all chain species k but two (k_1, k_2), which are taken as reference species and are assigned relative chemical potentials of 0. Each relative chemical potential, μ_k^*, is defined in terms of the actual chemical potentials of the species as (see Pant and Theodorou (1995), eq.(13)):

$$\mu_k^* = \mu_k - \left(\frac{k-k_1}{k_2-k_1}\right)\mu_{k_2} - \left(\frac{k-k_2}{k_1-k_2}\right)\mu_{k_1} \tag{2.21}$$

with k, k_1, k_2 being the number of monomer units in each chain of species k, k_1 and k_2, respectively. The probability density of the $[N_{\mathrm{ch}}nPT\boldsymbol{\mu}^*]$ ensemble in configuration space is (see Pant and Theodorou (1995), eq.(22)):

$$\rho^{N_{\mathrm{ch}}nPT\boldsymbol{\mu}^*}(V, \boldsymbol{r}^n; \text{connectivity}) = \text{const } \exp\left[\beta \sum_{\substack{k=1 \\ k\neq k_1,k_2}}^{M} \mu_k^* N_k - \beta PV - \beta\mathscr{V}(\boldsymbol{r}^n; \text{connectivity})\right] \tag{2.22}$$

where N_k stands for the number of molecules of species k, and M is the maximum number of species that may be present, V is the system volume and $\boldsymbol{r}^n$ is the vector of coordinates of all atoms, specifying the microscopic configuration of the system.

The EB move is initiated by randomly selecting the "attacking" chain end. One of the N_{bridge} bridgeable neighbours of the attacking end is then selected at random. As mentioned above, the bridging construction involved in EB generally has multiple solutions. All solutions are determined and one of them is chosen after overlap screening and weighting by the Boltzmann factor of the torsional potential, as in the case of ConRot. The probability of accepting such an EB move from the initial configuration, o, to the final configuration, n, is then (see Theodorou (2002), eq.(3.15)):

$$p_{\text{acc}}(o \to n) = \min\left[1, \frac{\dfrac{1}{N_{\text{bridge}}(n \to o)} a(n \to o)\rho^{N_{\text{ch}}nPT\boldsymbol{\mu}^*}(n)J(n)}{\dfrac{1}{N_{\text{bridge}}(o \to n)} a(o \to n)\rho^{N_{\text{ch}}nPT\boldsymbol{\mu}^*}(o)J(o)}\right] \tag{2.23}$$

As in the case of ConRot, the inverse bridging problem, starting from the final configuration, o, must be solved. In the acceptance criterion of eq.(2.23), the terms involving N_{bridge} account for the random selection of one of the bridgeable neighbours of the attacking chain end when the move is attempted in the forward (denominator) and in the reverse (numerator) direction. The $\boldsymbol{a}$ terms account for the bias in attempting the move according to a particular solution among all solutions of the geometric bridging problem, in the forward (denominator) or reverse (numerator) direction. The quantities symbolised by $\boldsymbol{\rho}$ are the equilibrium probability densities of the ensemble being simulated in the destination (numerator) and in the origin (denominator) state, while the quantities symbolised by $\boldsymbol{J}$ are the Jacobians of transformation from the generalized coordinates used in the bridging construction to Cartesian coordinates in the destination (numerator) and origin (denominator) states.

By construction, the number-average chain degree of polymerization of the polymer is $X \coloneqq N_n = n/N_{\text{ch}}$. Under a mild approximation concerning the configurational integral of the polydisperse system (please see below), the profile of relative chemical potentials $\boldsymbol{\mu}^*$ controls the degree-of-polymerization distribution at equilibrium.

Setting relative chemical potentials to $-\infty$ for all chain species shorter than $k_{\min}$ and equal to zero for $k_{\min}$ and longer, produces a truncated Flory number distribution function, also referenced as "the most probable" [13], which is zero below $k_{\min}$ and falls exponentially with k above $k_{\min}$.

$$\mu^*(k) = \begin{cases} -\infty & \text{for} \quad k < k_{\min} \\ 0 & \text{for} \quad k \geq k_{\min},\ k \neq k_1, k_2 \text{ with } k_{\min} \leq k_1 < k_2 \end{cases} \tag{2.24}$$

Setting relative chemical potentials to zero for all chain lengths within a symmetric window centered at X and to $-\infty$ outside that window produces a flat, uniform number distribution within the window.

$$\mu^*(k) = \begin{cases} -\infty & \text{for} \quad k < X - \ell \text{ or } k > X + \ell \\ 0 & \text{for} \quad X - \ell \ \leq k \leq X + \ell, \ k \neq k_1, k_2, \text{ with } X - \ell \ \leq k_1 < k_2 \leq X + \ell \end{cases} \tag{2.25}$$

On the other hand, setting relative chemical potentials equal to a parabolic function of the chain length with maximum at X within a symmetric window centered at X and to $-\infty$ outside that window, produces a truncated Gaussian number distribution within the window [20].

$$\mu^*(k) = \begin{cases} \psi \left[\ell^2 - (X-k)^2\right] & \text{for} \quad X - \ell < k < X + \ell \\ -\infty & \text{for} \quad k < X - \ell \text{ or } k > X + \ell \end{cases} \tag{2.26}$$

where the variance of the distribution is $\sigma^2 = 1/(2\beta\psi)$ and $\beta = 1/(k_{\mathrm{B}}T)$.

Simulations in the $[N_{\mathrm{ch}} n P T \boldsymbol{\mu}^*]$ ensemble yield a degree-of-polymerization distribution that depends on the spectrum of chemical potentials, $\boldsymbol{\mu}^*$. The relationship between the chemical potential spectrum and the chain degree-of-polymerization distribution is greatly simplified in cases where the configurational integral (see Pant and Theodorou (1995), eq.(11)):

$$Z(V, T, N_1, N_2, \ldots, N_M) = \int \exp\left[-\beta \mathscr{V}(\boldsymbol{r}^n)\right] \, d^{3n}r \tag{2.27}$$

(where N_i stands for the number of molecules of species M) depends on the monomer density and number-averaged molecular weight, but not on the details of the distribution, i.e. if (see Pant and Theodorou (1995), eq.(23)):

$$Z(V, T, N_1, N_2, \ldots, N_M) = Z(V, T, N_{\mathrm{ch}}, n) \tag{2.28}$$

When this condition is satisfied, the degree-of-polymerization distribution is related to the chemical potential spectrum $\boldsymbol{\mu}^*$ merely by combinatoric considerations and does not depend on the energetics of the system. Condition (2.28) would hold exactly in systems of noninteracting unperturbed chains (no inter-molecular, no nonbonded intra-molecular interactions), where Z can be reduced to a product of Z's of individual chains, each depending exponentially on the chain's degree of polymerization. One would expect that this condition would also hold in systems with realistic potentials, provided short-chain species were absent, under which circumstance the properties of the melt would depend only on the monomer density and the density of chain ends. The extent to which this analysis is valid depends on the extent to

which the configurational integral $Z(V,T,N_1,N_2,\ldots,N_M)$ varies with the specifics of the chain degree-of-polymerization distribution within the bounds imposed upon its fluctuations.

Wherein $N_{\rm ch}$ and n are large, a maximum term approximation on $\ln\tilde{Y}$ subject to the constraints: (i) $\sum_k \langle N_k\rangle = N_{\rm ch}$ and (ii) $\sum_k k\,\langle N_k\rangle = n$, may be invoked to map the distribution of chemical potentials onto the corresponding distribution of degrees of polymerization. $\tilde{Y}$ is a partition function related to Y, $Y = -k_{\rm B}T\ln\tilde{Y}$, where Y is the natural thermodynamic potential for the variables of interest $[N_{\rm ch}nPT\mu^*]$ (see Pant and Theodorou (1995), eq.(16)):

$$Y = A - \sum_{\substack{k=1\\ k\neq k_1,k_2}}^{M} \mu_k^* N_k + PV \tag{2.29}$$

and A is the Helmholtz free energy.

Using Stirling's approximation for the logarithms of factorial terms, this yields $(M-2)$ equations of the form (see Pant and Theodorou (1995), eq.(24)):

$$\ln\left(\langle N_k\rangle\right) = \beta\mu^* + \xi + \psi k, \quad (k = 1,\ldots,M,\ k \neq k_1, k_2) \tag{2.30}$$

where ξ and ψ are Lagrange multipliers corresponding to the two constraints (i) and (ii). These equations simplify to the relation of equation (2.33).

Analogously, the reference species are governied by the following relations (see Pant and Theodorou (1995), eq.(26-27)):

$$\langle N_i\rangle = cy^i \tag{2.31}$$

$$\langle N_j\rangle = cy^j \tag{2.32}$$

where $c = e^{\xi}$ and $y = e^{\psi}$. The populations of all the other species are controled through the relation (see Pant and Theodorou (1995), eq.(25)):

$$\langle N_k\rangle = cy^k \exp\left(\beta\mu_k^*\right), \quad (k = 1,\ \ldots,\ m,\ k \neq k_1, k_2) \tag{2.33}$$

where N_k is the number of chains that consist of k monomers (or UAs). The M equations above are substituted into the constraints (i) and (ii) to solve for the constants c and y, thereby yielding the entire distribution [13].

Chapter 3

Simulation Details

3.1 Molecular model

The molecular model employed in this thesis was initially created by P. V. Krishna Pant, T. Boone, V. G. Mavrantzas, N. C. Karayiannis and A. E. Giannousaki and has received many additions and modifications. The application of the program described in this thesis was restricted to the simulation of linear PE, without any branches. Since hydrogen atoms are small compared to carbon atoms, it adopts the "united atoms" (UAs) representation according which methylene (CH_2) is treated as the same Lennard-Jones (L-J) interacting site as the methyl group (CH_3). These UAs can be placed at the positions of the carbon atoms instead of the center of mass of eah of these reacted monomers. Despite this approach made for simplification purposes (fewer degrees of freedom, lower computation time and MC complexity) the system is still considered “atomistic” in contrast to other “coarse-grained” models that consider whole parts of polymers as centers of interaction. It provides accurate results in PE simulations when the sample is studied as a melt [7].

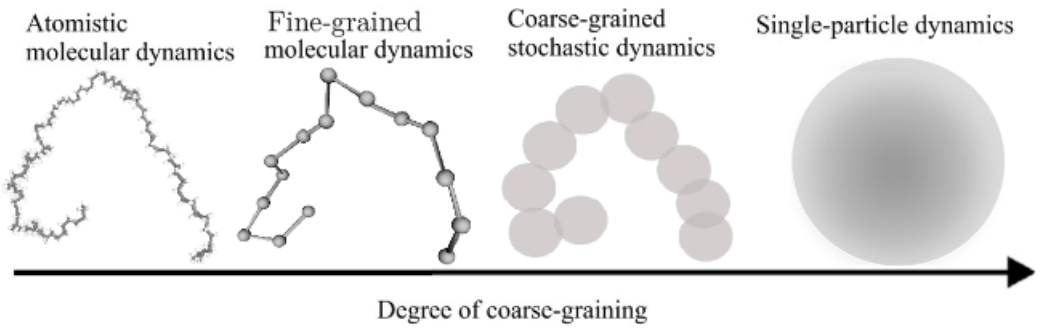

Figure 3.1: Schematic representation of different levels of coarse-graining [22]

In this model parameters are used from the TraPPE (Transferable Potentials for Phase Equilibria) model of Martin and Siepmann, and from the rotational potential of Toxvaerd. The intra-molecular interactions, that is the nonbonded interactions between atoms in the same molecule seperated by three or more bonds and the inter-molecular interactions, that is the nonbonded interactions between atoms of different molecules, are described by the same L-J potential:

$$\mathscr{V}_{\mathrm{LJ}}(r) = 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right] \tag{3.1}$$

where r is the distance of seperation between two interacting sites, σ is the Van der Waals radius and ϵ is the energy well depth. The values of the parameters σ and ϵ define exactly the interactions between two sites and they are different for every pair of interacting centers. Here, there is no distinction between methylene-methylene, methylene-(methyl group) and (methyl group)-(methyl group) interactions, and the exact same parameters are used. The results in this case remain quite accurate, firstly because the extra hydrogen atom does not cause great difference in the values of the parameters and secondly because most of the times, especially inside systems with long chains, the majority of interactions are between CH_2.

Generally, the length, $\ell_{ij} \coloneqq ||\boldsymbol{r}_{ij}|| \coloneqq r_{ij}$, of a bond fluctuates around a value called "reference bond length", ℓ_0. There are many analytical expressions modeling the potential energy curve for a typical bond, but those not being amenable to efficient computation or those requiring the calculation of too many parameters are left aside when it comes to molecular simulations. An example of a stretching potential would be the Morse potential: $\mathscr{V}_{\mathrm{bond}}^{\mathrm{Morse}}(\ell_{ij})/k_{\mathrm{B}} = D_{\mathrm{e}}\{1 - \exp\left[-\alpha(\ell_{ij} - \ell_0)\right]\}^2$ which requires three parameters $(D_{\mathrm{e}}, \alpha, \ell_0)$, to be specified for each bond. Instead, Hooke's law formula or harmonic potential is often used:

$$\frac{\mathscr{V}_{\mathrm{bond}}(r_{ij})}{k_{\mathrm{B}}} = \frac{1}{2} k_\ell \left(r_{ij} - \ell_0\right)^2 \tag{3.2}$$

where k_{B} is Boltzmann's constant and k_ℓ is the stretching constant of the bond. Consideing the length of each bond to remain the same we can ignore the contribution from the fluctuation of the bonds to the total energy of the system. As far as our model is concerned, all bond lengths were maintained constant throughout the simulation.

The deviation of bond angles, θ, from their refrence values is also frequently described as governed by a Hooke's law, such as the Van der Ploeg and Barendsen harmonic bending potential:

$$\frac{\mathscr{V}_{\mathrm{bend}}(\theta_{ijk})}{k_{\mathrm{B}}} = \frac{1}{2} k_\theta \left(\theta_{ijk} - \theta_0\right)^2 \tag{3.3}$$

where θ_0 is the reference value of the bond angle and k_θ is the stretching constant of the angle.

The dihedral angles, ϕ, formed between two planes defined by two sequential and partially overlapped triplets of carbon atoms, $(\boldsymbol{r}_i, \boldsymbol{r}_j, \boldsymbol{r}_k)$ and $(\boldsymbol{r}_j, \boldsymbol{r}_k, \boldsymbol{r}_l)$, are usually sampled using a nine-term, sum-of-cosines torsional potential:

$$\frac{\mathscr{V}_{\text{tor}}(\phi_{ijkl})}{k_{\text{B}}} = \sum_{m=0}^{8} C_i \cos^m(\phi_{ijkl}) \tag{3.4}$$

where C_i are usually referred as "height barriers" and are fitting parameters, and torsional angle ϕ is calculated as:

$$\cos(\phi_{ijkl}) = \frac{\boldsymbol{r}_{ij} \times \boldsymbol{r}_{jk}}{\|\boldsymbol{r}_{ij} \times \boldsymbol{r}_{jk}\|} \cdot \frac{\boldsymbol{r}_{jk} \times \boldsymbol{r}_{kl}}{\|\boldsymbol{r}_{jk} \times \boldsymbol{r}_{kl}\|} \tag{3.5}$$

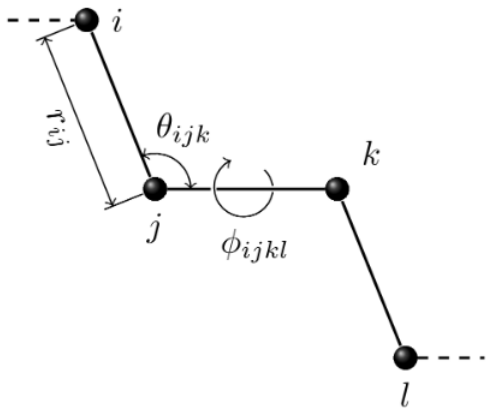

Figure 3.2: Schematic representation of bond $\boldsymbol{r}_{ij}$, bond angle θ_{ijk} and torsional angle ϕ_{ijkl} [18]

Finally, the total potential energy of the system is calculated through a relatively simple four-component formula of the intra- and inter-molecular interactions. One functional form of such a potential is:

$$\begin{aligned}\mathscr{V}(\boldsymbol{r}^n) = &\sum_{\text{bonds}} \frac{1}{2} k_\ell (\ell_{ij} - \ell_0)^2 + \sum_{\text{angles}} \frac{1}{2} k_\theta (\theta_{ijk} - \theta_0)^2 + \sum_{\text{torsions}} \sum_{m=0}^{8} C_i \cos^m(\phi_{ijkl}) \\ &+ \sum_{\text{i=1}}^{n} \sum_{\text{j=i+1}}^{n} \left(4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] \right)\end{aligned} \tag{3.6}$$

where we add the inter-molecular terms: bond fluctuation potential, the angle bend potential, the torsional potential and the L-J potential. A standard 12-6 L-J potential is used in our model for all pairs of intramolecular neighbors seperated by

more than three bonds along the same chain and all intermolecular ones. Usually standard, combining rules like the ones of Lorentz-Berthelot are used for the computation of the energy well depth and collision diameter of the potential, in order to describe nonbonded interactions between sites of different kind:

$$\epsilon_{ij} = \sqrt{\epsilon_{ii}\epsilon_{jj}}, \quad \sigma_{ij} = \frac{\sigma_{ii} + \sigma_{jj}}{2} \tag{3.7}$$

Table 3.1: Parameters of the interaction potentials in our molecular model [25]

Interaction Type	Interaction Potential	Parameters
Non Bonded	Lennard-Jones $U_{LJ}(r_{ij}) = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]$ $i, j = CH_2, CH_3$	$\frac{\varepsilon_{ij}}{k_B} = 46\text{K}$ $\sigma_{ij} = 3.95\text{Å}$
Bonded	Bond length (constant)	$l = 1.54\text{Å}$
	Bond angle bend $\frac{U_{bend}(\theta)}{k_B} = \frac{1}{2}k_\theta(\theta - \theta_0)^2$	$k_\theta = 62500\text{Krad}^{-2}$ $\theta_0 = 114°$
	Dihedral angle torsion $\frac{U_{tor}(\varphi)}{k_B} = \sum_{i=0}^{8} c_i \cos^i\varphi$	$c_0 = 1001\text{K}$, $c_1 = 2130\text{K}$ $c_2 = -303\text{K}$, $c_3 = -3612\text{K}$ $c_4 = 2227\text{K}$, $c_5 = 1966\text{K}$ $c_6 = -4489\text{K}$, $c_7 = -1736\text{K}$ $c_8 = 2817\text{K}$

In order to save some computational time, the L-J potential is truncated and it is calculated only for distances up to $r = r_c := r_2 = 2.2\sigma$. This is also justified from a physical perspective since the L-J potential decays rapidly for large r. In order to smooth the discontinuitiy that arises due to the application of the cut-off distance a quintic spline is employed for distances between $r_1 = 1.45\sigma$ and r_2 (see Μόρφης (2012), eq.(3.5)):

$$\mathscr{V}(r) = \begin{cases} \mathscr{V}_{\text{LJ}}(r) & \text{for} \quad r < r_1 \\ \epsilon(1-\xi)^3\left[\frac{\mathscr{V}_1}{\epsilon} + \left(3\frac{\mathscr{V}_1}{\epsilon} + \Delta\frac{\mathscr{V}_1'}{\epsilon/\sigma}\right)\xi + \left(6\frac{\mathscr{V}_1}{\epsilon} + 3\Delta\frac{\mathscr{V}_1'}{\epsilon/\sigma} + \frac{\Delta^2}{2}\frac{\mathscr{V}_1''}{\epsilon^2/\sigma^2}\right)\xi^2\right] & \text{for} \quad r_1 \le r < r_2 \\ 0 & \text{for} \quad r \ge r_2 \end{cases} \tag{3.8}$$

where $\xi = (r - r_1)/(r_2 - r_1)$, $\Delta = (r_2 - r_1)/\sigma$ and $\mathscr{V}_1, \mathscr{V}_1', \mathscr{V}_1''$ are: the value of the L-J potential function, its first derivative and its second derivative, respectively,

for $r = r_1$. The contribution from the tail correction in the internal energy of the system is calculated via integration:

$$\mathscr{V}_{\mathrm{LJ}}^{\mathrm{tail}} = 2\pi n \rho_n \int_{r_1}^{\infty} [\mathscr{V}_{\mathrm{LJ}}(r) - \mathscr{V}(r)]\, g(r) r^2 \, dr \tag{3.9}$$

where n is the total number of UAs in the system, ρ_n is the number density and $g(r)$ is the pair correlation function, which we could consider as unit for distances $r > r_1$. We can also replace $\mathscr{V}(r)$ with 0 for values of r greater than r_c. Even eq.(3.1) is an approximation since it ignores the contributions from more populous groups of particles, e.g. triads, quartets to the nonbonded energy.

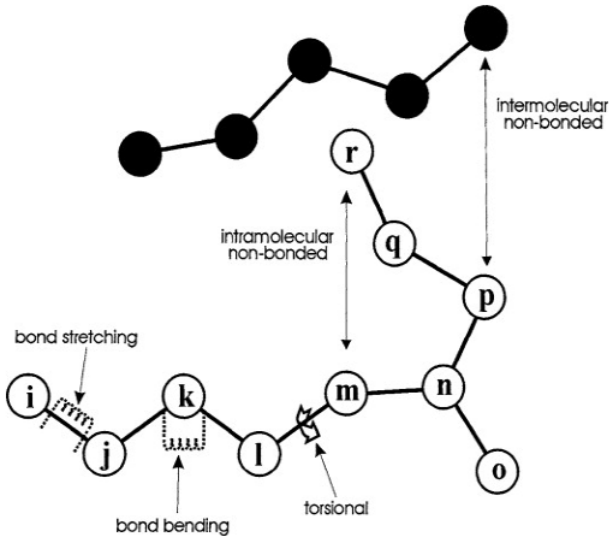

Figure 3.3: Schematic representation of polymer segments belonging to different chains in the UA description along with typical examples of interactions [9]

3.2 Periodic boundaries

Knowing the structure of a system and the way all its constituent particles interact with each other could be enough to theoretically start a simulation, but practically it is not. Not until we manage to get rid of the boundary effects.

In small systems with few hundred particles, boundary effects become dominant. If we start talking about bigger and bigger systems, the percentage of those particles found in a mono-layer at the boundaries gets smaller, but not negligible (for a system of 10^6 particles this percentage remains 6%) and at the same time the computational time reaches prohibitive heights. Simulations with realistic or bulk polymeric systems consisting of order 10^{23} particles per cm are impossible to run.

In this case we simulate a small portion of the system: few particles (the number is chosen in a trade-off between simulation time and accuracy of the results) in a restricted, spatial region (e.g., cube with edge L). If we do not impose periodic boundary conditions on this simulation box, then boundary or surface effects will arise. Particles close to the boundaries will experience a different environment and interactions from those sited inside the bulk, altering the values of the predicted properties. This can be avoided by setting periodic or Born-von Karman boundary conditions which make it possible to approximate an infinite system by using a small part surrounded by translated copies of itself in all directions.

So when a particle exits from the box, a copy of that particle enters into the box from the opposite face. Likewise, when a chain conformation exits the boundaries of the box, an image of the chain enters through the opposite face.

Non-bonded interactions around a particle are explicitly calculated only within a spherical region of radius r_c centered at the particle. The distance between two particles is considered equivalent to the distance between the "minimum image of the one particle with respect to the other", as it is represented in Figure 3.4. If the minimum image convention is met: $L/2 \geq r_c$ with L being the box edge length and r_c the cutoff distance, then for each particle there is only one image for every other particle that could interact with the first one. Besides that, it is also mandatory the value of L to be much bigger than the biggest length scale we can come across in the system. For polymeric melts L should ideally also be much bigger than the radius of gyration of the longest chain present.

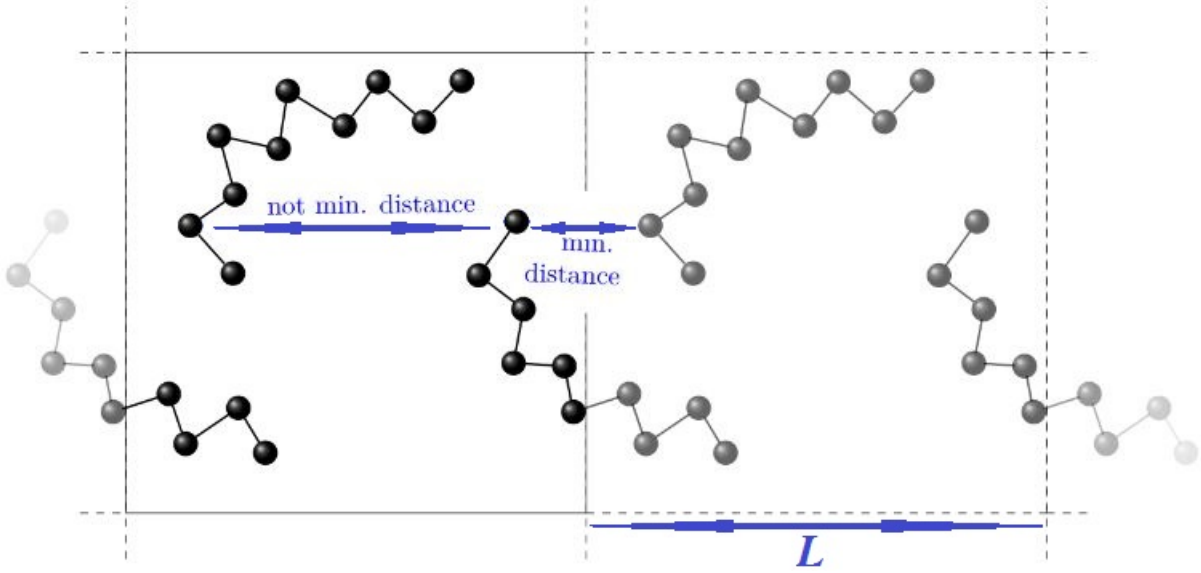

Figure 3.4: Simulation box with periodic boundaries [18]

Chapter 4

New Criterion and Algorithm

Up till now, the importance and even necessity of succeeding in developing a method that would advance computational polymeric models into having control over the shape of the molecular-weight/length/degree-of-polymerization distribution of the simulated system has been made clear. We should look for a method that is simple in application, efficient (not too computation-time consuming) and preferably one that would allow the user to predefine the distribution they desire to be met. Until we succeed for more sophisticated distributions than those currently available; delta (monodispersity) and uniform, the divergence between the systems resulting from our simulations and the real materials, will limit the utility of our simulations. Only one article by Zhang, Pu and Ren (2022) [24] concerning this task (using a novel control algorithm based on the moment-generating function on the polymerization process of styrene), has come to our attention.

In this thesis we use as our main reference the article of Pant and Theodorou (1995) [13]. There, as extensively described in Section 2.5, we find a thermodynamic study of the ensemble portaying a polydisperse system and the introduction of some reduced chemical potential functions. These functions could be used in the Metropolis criterion of a MC simulation enabling us to regulate the distribution of the degrees of polymerization in a system. Pant and Theodorou (1995) [13] provided three profiles of reduced chemical potentials; for the uniform distribution, the most probable distribution and the truncated Gaussian distribution. This thesis contributes another three reduced chemical potentials: for the truncated bimodal Gaussian, the truncated double-Gaussian, and finally the Schulz-Zimm distribution.

The criterion we derive and test in this thesis for the regulation of the chain length distribution in the simulated systems is not based on employing the reduced chemical potentials. Instead, it is much simpler in its form, it does not demand from the user to know at all the reduced chemical potential from which a distribution would be derived and therefore it allows to readily enrich our model's arsenal with

any new distribution. An example that emphatically showcases as de facto the truth of the previous expression: "any distribution", is that our newly proposed criterion successfully manages to sample also tabulated and even non-analytically expressed distributions.

4.1 MWD Metropolis Criterion

The program used in this thesis performs MC atomistic simulations, so a Metropolis criterion is implemented, such as the one in eq.(2.18). Due to the simplicity of the MWDs used so far, there was no need to pay too much attention to the values of the third term of eq.(2.18), both in numerator and denominator. But now, this need arises due to the non-trivial probability density values of the new distributions considered.

$$\frac{\rho^{N_{\rm ch}nPT\boldsymbol{\mu}^*}(\text{new})}{\rho^{N_{\rm ch}nPT\boldsymbol{\mu}^*}(\text{old})} = \frac{\exp\left[\beta \sum\limits_{\substack{k=1\\k\neq k_1,k_2}}^{M} \mu_k^* N_{k,\text{new}} - \beta P V_{\text{new}} - \beta \mathscr{V}(\boldsymbol{r}^n_{\text{new}}; \text{connectivity})\right]}{\exp\left[\beta \sum\limits_{\substack{k=1\\k\neq k_1,k_2}}^{M} \mu_k^* N_{k,\text{old}} - \beta P V_{\text{old}} - \beta \mathscr{V}(\boldsymbol{r}^n_{\text{old}}; \text{connectivity})\right]} \tag{4.1}$$

It is convenient to recast eq.(4.1) for the ratio of probability densities of two configurations involved in a connectivity-altering move, which enters the acceptance criterion of the move, in terms of the target number distribution of chain lengths. For this we need to replace the relative chemical potentials of various chain species appearing in eq.(4.1) in terms of the corresponding mole fractions, or probabilities of these species at equilibrium. Let:

$$p_k^{\text{target}} = \frac{\langle N_k \rangle}{N_{\rm ch}} \tag{4.2}$$

be the equilibrium probability of chain species k, equal to the ensemble averaged number of chains $\langle N_k \rangle$ at equilibrium divided by the total number of chains $N_{\rm ch}$. In actual applications, p_k^{target} is prescribed by the chain length distribution we wish to simulate. This distribution may be completely arbitrary, e.g., given in tabular form.

Pant and Theodorou (1995) [13] explored an approximation for the configurational integral of the polydisperse polymer at given volume and temperature. Specifically, they considered that the configurational integral of the polydisperse multichain system depends on the monomer density and on the number average

chain length, but not on the details of the chain length distribution (see Pant and Theodorou 1995, eq.(23), reproduced here as eq.(2.28)).

Under this mild assumption, they were able to relate the equilibrium (limiting) populations of the chain species to the corresponding chemical potentials through the following equations (see Pant and Theodorou 1995, eqs.(25-27)): eq.(2.33), eq.(2.31) and eq.(2.33).

For a given profile of relative chemical potentials, the parameters c and y in these equations can be determined using the chain and segment population balance equations:

$$\sum_{k=1}^{M} \langle N_k \rangle = N_{\text{ch}} \tag{4.3}$$

$$\sum_{k=1}^{M} k \langle N_k \rangle = n \tag{4.4}$$

Eq.(2.33), eq.(2.31), eq.(2.33), eq.(4.3) and eq.(4.4) constitute a system of M+2 equations in the M+2 unknowns: $\langle N_1 \rangle$, $\langle N_1 \rangle$, $\ldots$, $\langle N_M \rangle$, c, y.

From eq.(2.33), eq.(2.31) and eq.(2.33) we obtain:

$$\exp(\beta \mu_k^*) = \frac{\langle N_k \rangle}{c y^k}, \ (k = 1, 2, \ldots, M) \tag{4.5}$$

Interestingly, eq.(4.5) holds for all k, including the reference species k_1 and k_2, which are assigned $\mu_{k_1}^* = \mu_{k_2}^* = 0$.

We now return to the ratio of probability densities of two configurations, "new" and "old", appearing in eq.(4.1). The factor of that ratio that contains relative chemical potentials can be rewritten as follows:

$$\frac{\exp\left[\beta \sum\limits_{\substack{k=1 \\ k \neq k_1, k_2}}^{M} \mu_k^* N_k^{\text{new}}\right]}{\exp\left[\beta \sum\limits_{\substack{k=1 \\ k \neq k_1, k_2}}^{M} \mu_k^* N_k^{\text{old}}\right]} = \frac{\exp\left[\beta \sum\limits_{k=1}^{M} \mu_k^* N_k^{\text{new}}\right]}{\exp\left[\beta \sum\limits_{k=1}^{M} \mu_k^* N_k^{\text{old}}\right]} = \frac{\prod\limits_{k=1}^{M} \left[\exp(\beta \mu_k^*)\right]^{N_k^{\text{new}}}}{\prod\limits_{k=1}^{M} \left[\exp(\beta \mu_k^*)\right]^{N_k^{\text{old}}}} =$$

$$= \prod_{k=1}^{M} \left[\exp(\beta \mu_k^*)\right]^{N_k^{\text{new}} - N_k^{\text{old}}} \tag{4.6}$$

In the summations of eq.(4.6) we have removed the restriction $k \neq k_1, \ k \neq k_2$ by virtue of the fact that $\mu_{k_1}^* = \mu_{k_2}^* = 0$. A connectivity-altering move alters the populations of at most four chain species in the system. Let k_i^{old} and k_j^{old} be

the species in the old configuration which engage in a connectivity-altering move, producing species k_i^{new} and k_j^{new}. Some or all of the lengths k_i^{old}, k_j^{old}, k_i^{new}, k_j^{new} may equal each other. Whatever their values may be, by virtue of the segment balance they have to satisfy the constraint:

$$k_i^{\text{old}} + k_j^{\text{old}} = k_i^{\text{new}} + k_j^{\text{new}} \tag{4.7}$$

The ratio of eq.(4.6) becomes:

$$\frac{\exp\left[\beta \sum\limits_{\substack{k=1 \\ k\neq k_1,k_2}}^{M} \mu_k^* N_k^{\text{new}}\right]}{\exp\left[\beta \sum\limits_{\substack{k=1 \\ k\neq k_1,k_2}}^{M} \mu_k^* N_k^{\text{old}}\right]} = \frac{\exp\left(\beta\mu^*_{k_i^{\text{new}}}\right)\exp\left(\beta\mu^*_{k_j^{\text{new}}}\right)}{\exp\left(\beta\mu^*_{k_i^{\text{old}}}\right)\exp\left(\beta\mu^*_{k_j^{\text{old}}}\right)} \tag{4.8}$$

Using eq.(4.5) in eq.(4.8) we obtain:

$$\frac{\exp\left[\beta \sum\limits_{\substack{k=1 \\ k\neq k_1,k_2}}^{M} \mu_k^* N_k^{\text{new}}\right]}{\exp\left[\beta \sum\limits_{\substack{k=1 \\ k\neq k_1,k_2}}^{M} \mu_k^* N_k^{\text{old}}\right]} = \frac{\dfrac{\left\langle N_{k_i^{\text{new}}}\right\rangle}{cy^{k_i^{\text{new}}}}\dfrac{\left\langle N_{k_j^{\text{new}}}\right\rangle}{cy^{k_j^{\text{new}}}}}{\dfrac{\left\langle N_{k_i^{\text{old}}}\right\rangle}{cy^{k_i^{\text{old}}}}\dfrac{\left\langle N_{k_j^{\text{old}}}\right\rangle}{cy^{k_j^{\text{old}}}}} = \frac{\left\langle N_{k_i^{\text{new}}}\right\rangle\left\langle N_{k_j^{\text{new}}}\right\rangle}{\left\langle N_{k_i^{\text{old}}}\right\rangle\left\langle N_{k_j^{\text{old}}}\right\rangle} y^{k_i^{\text{old}}+k_j^{\text{old}}-k_i^{\text{new}}-k_j^{\text{new}}} \tag{4.9}$$

In view of eq.(4.7), eq.(4.9) gives:

$$\frac{\exp\left[\beta \sum\limits_{\substack{k=1 \\ k\neq k_1,k_2}}^{M} \mu_k^* N_k^{\text{new}}\right]}{\exp\left[\beta \sum\limits_{\substack{k=1 \\ k\neq k_1,k_2}}^{M} \mu_k^* N_k^{\text{old}}\right]} = \frac{\left\langle N_{k_i^{\text{new}}}\right\rangle\left\langle N_{k_j^{\text{new}}}\right\rangle}{\left\langle N_{k_i^{\text{old}}}\right\rangle\left\langle N_{k_j^{\text{old}}}\right\rangle} \tag{4.10}$$

Invoking eq.(4.2) within eq.(4.10) we can write, equivalently:

$$\frac{\exp\left[\beta \sum\limits_{\substack{k=1 \\ k\neq k_1,k_2}}^{M} \mu_k^* N_k^{\text{new}}\right]}{\exp\left[\beta \sum\limits_{\substack{k=1 \\ k\neq k_1,k_2}}^{M} \mu_k^* N_k^{\text{old}}\right]} = \frac{p_{k_i^{\text{new}}}^{\text{target}}\; p_{k_j^{\text{new}}}^{\text{target}}}{p_{k_i^{\text{old}}}^{\text{target}}\; p_{k_j^{\text{old}}}^{\text{target}}} \tag{4.11}$$

Thus, in the ratio of probability densities of the new and old configurations involved in a connectivity-altering move, the factor that has to do with the relative chemical potentials acquires a transparent interpretation. It is simply the product of target probabilities of the species being created over the product of target probabilities of the species being destroyed during the move. Note that, in order to evaluate this factor, we need only know the target probabilities up to a multiplicative factor. In other words, our target number distribution of chain lengths need not be normalized.

By combining eq.(4.1) with eq.(4.11) we obtain the ratio of full probability densities:

$$\frac{\rho^{N_{\rm ch}nPT\boldsymbol{\mu}^*}(\text{new})}{\rho^{N_{\rm ch}nPT\boldsymbol{\mu}^*}(\text{old})} = \frac{p_{k_i^{\rm new}}^{\rm target}\, p_{k_j^{\rm new}}^{\rm target}}{p_{k_i^{\rm old}}^{\rm target}\, p_{k_j^{\rm old}}^{\rm target}} \exp\left[-\beta P\left(V_{\rm new} - V_{\rm old}\right)\right] \times \\ \times \exp\left\{-\beta\left[\mathscr{V}(\boldsymbol{r}_{\rm new}^n; \text{connectivity}) - \mathscr{V}(\boldsymbol{r}_{\rm old}^n; \text{connectivity})\right]\right\} \quad (4.12)$$

Given that connectivity-altering moves are typically attempted at constant volume, the ratio simplifies to:

$$\frac{\rho^{N_{\rm ch}nPT\boldsymbol{\mu}^*}(\text{new})}{\rho^{N_{\rm ch}nPT\boldsymbol{\mu}^*}(\text{old})} = \frac{p_{k_1^{\rm new}}^{\rm target}\, p_{k_2^{\rm new}}^{\rm target}}{p_{k_1^{\rm new}}^{\rm target}\, p_{k_2^{\rm new}}^{\rm target}} \exp\left\{-\beta\left[\mathscr{V}(\boldsymbol{r}_{\rm new}^n; \text{connectivity}) - \mathscr{V}(\boldsymbol{r}_{\rm old}^n; \text{connectivity})\right]\right\} \quad (4.13)$$

The acceptance criterion of a connectivity-altering move from configuration "old" to configuration "new" is like eq.(2.23) (see DNT [20] in Nielaba, Mareschal, Ciccotti (Eds) (2002), eq.(3.15)):

$$p_{\rm acc}(\text{old} \to \text{new}) = \min\left[1, \frac{\dfrac{1}{N_{\rm bridge}(\text{new} \to \text{old})} a(\text{new} \to \text{old}) J(\text{new}) \rho^{N_{\rm ch}nPT\boldsymbol{\mu}^*}(\text{new})}{\dfrac{1}{N_{\rm bridge}(\text{old} \to \text{new})} a(\text{old} \to \text{new}) J(\text{old}) \rho^{N_{\rm ch}nPT\boldsymbol{\mu}^*}(\text{old})}\right]$$

with eq.(4.13) used within eq.(2.23).

4.2 New Relative Chemical Potentials

This thesis also introduces three new reduced chemical potential profiles for the enrichment of the three-membered list proposed by Pant and Theodorou (1995) [13]. They are for the implementation of three analytical distributions: the truncated multimodal Gaussian, the truncated double-Gaussian and the Schulz-Zimm distribution, assuming the condition of eq.(2.28) holds true.

For a multimodal number distribution to be produced consisting of M overlapping Gaussian distributions with variances equal to $\sigma_i^2 = 1/(2\beta\psi_i)$ (where $\beta = 1/(k_\mathrm{B}T)$) and centered at degree of polymerization X_i, respectively, the reduced chemical potential must be set to:

$$\mu^*(k) = \begin{cases} \dfrac{1}{\beta} \ln \left\{ \displaystyle\sum_{i=1}^{M} \exp\left[\beta\psi_i(\ell^2 - (X_i - k)^2)\right] \right\} & \text{for} \quad X - \ell < k < X + \ell \\ -\infty & \text{for} \quad k < X - \ell \text{ or } k > X + \ell \end{cases} \tag{4.14}$$

For a double Gaussian or two-piece normal or binormal or Fechner distribution to be produced with left-part variance equal to $\sigma_\mathrm{L}^2 = 1/(2\beta\psi_\mathrm{L})$ and right-part variance equal to $\sigma_\mathrm{R}^2 = 1/(2\beta\psi_\mathrm{R})$ centered at degree of polymerization X, the reduced chemical potential must be set to:

$$\mu^*(k) = \begin{cases} \psi_\mathrm{L}(\ell^2 - (X - k)^2)\theta(X - k) + \psi_\mathrm{R}(\ell^2 - (X - k)^2)\theta(k - X), & \\ \qquad\qquad\qquad\qquad \text{for} \quad X - \ell < k < X + \ell & \\ -\infty \quad \text{for} \quad k < X - \ell \text{ or } k > X + \ell & \end{cases} \tag{4.15}$$

where θ is the Heaviside step function.

Finally, for the Schulz-Zimm distribution parametrized in respect with N_n and polydispersity index λ, $f_\mathrm{SZ}(k;\lambda,N_\mathrm{n}) = \lambda^\lambda k^{\lambda-1}\exp(-\lambda k/N_\mathrm{n})/(\Gamma(\lambda)N_\mathrm{n}^\lambda)$ (see Qi *et al.* (2016), eq.(2)), the reduced chemical potential must be set to:

$$\mu^*(K(k)) = \begin{cases} \dfrac{1}{\beta} \ln \left[K(k)^{\lambda-1} \exp\left(-\lambda \dfrac{K(k)}{N_\mathrm{n}}\right) \right] & \text{for} \quad X - \ell < k < X + \ell \\ -\infty & \text{for} \quad k < X - \ell \text{ or } k > X + \ell \end{cases} \tag{4.16}$$

where $K(k) = k - (X - \ell)$ and thus we have also displaced μ^* towards longer lengths by $(X - \ell)$, for $\mu^*(X - \ell) = 0$ to hold true.

4.3 Description of the Algorithm Followed

First and foremost, we need to create the initial configurations of the systems we are going to simulate. This is done by employing the Amorphous Builder software and setting as input parameters the number and length of the chains constituting the systems, the temperature equal to 450 K, the pressure equal to 1 atm and the density equal to $0.766\,\mathrm{g\,cm^{-1}}$. The initial, polymeric systems we obtain are monodisperse. The Amorphous Builder algorithm is based on the biased formation of chains

with respect to the step-by-step increase of the total energy of the system during the insertion of new pairs of atoms. The ultimate goal is the minimization of the ovelapping of the atoms to be achieved and at the same time the right distributions of bond angles and torsional agles to be followed, as they are expected by the respective potentials used, $\mathscr{V}_{\text{bend}}$ and $\mathscr{V}_{\text{tor}}$, for the given temperature [25].

Once the systems are ready and all the necessary for our simulations data are saved in .DAT files we run a tailored script on them in order to turn them into a file that is acceptable and readable by our main MC program. In this file is where the user explicitly sets the needed parameters for their chosen MWD to be implemented. These are six parameters in total:

1. The first, ϕ, is for the calculation of the half width, ℓ, of the window. The window of degrees of polymerization inside which are found all the accessible degree-of-polymerization values is centered on the number-average degree of polymerization of the system, X. For the symmetric distributions it remains upon the one the chains had as a monodisperse system. The half width of the window is calculated as follows: $\ell = \phi\ X$. If it has the value zero, the system remains monodisperse. This variable must always be provided!

2. The second is to declare the chosen MWD (uniform, most probable, Gaussian, bimodal Gaussian, double Gaussian or tabular). This variable must always be provided! (In the case of monodispersity its value has no importance).

3. The third, α_{l}, is for the calculation of the (left, in the case of a double-peak MWD) standard deviation, when it is needed to be calculated, except the case of the Schulz-Zimm distribution where it is the value of parameter λ. The (left) standard deviation is calculated with respect to the half width of the window as follows: $\sigma_{\text{l}} = \alpha_{\text{l}}\ \ell = \alpha_{\text{l}}\ \phi\ X$. The values of this and of all the following variables have no importance if the monodisperse, the uniform or the tabular distributions have been selected.

4. The forth, X_{l}, is to set the position of the (left, in the case of a double-peak MWD) peak of the distribution. The user needs only to pay attention for X_{l} to lie inside the range of the window. In the case of the Schulz-Zimm distribution it is the value of parameter N_{n}.

5. The fifth, α_{r} and the sixth, X_{r}, are for the calculations of the right (in the case of a double-peak MWD) standard deviation and peak of the distribution, respectively, if needed to be calculated.

The default values are set by the script during the file transformation process and they are set for monodispersity. Once this step is done too, we proceed with running the simulations of interest.

The program reads from the input file and saves the data either in 1D variables or in lists (arrays) and also creates all the lists needed to be shared between the routines. As part of this "preparation" stage the list that contains the values of the chosen probability distribution function is created (except in the case of the monodispersity). An "operator" routine checks what MWD is selected. If it is one of the analytical ones then it calculates the values of the respective probability distribution function and saves them inside the list. It does not bother calculating the normalization factor since ultimately these values are going to be replaced in the ratios of eq.(2.23). If the tabular distribution is selected, then it reads in the probability distribution values right from a TMWD.txt file.

If the probability values are to be calculated, a "calculator" routine does the job and afterwards a "reviser" routine is called. The reviser checks that the number-average degree of polymerization of the system is exactly upon the degree of polmerization of the chains in monodispersity. If it is not, it displaces the whole distribution by just the right step for these two values to agree. This has been prooved de facto to be a very important procedure, since, if there is not an agreement the desired distribution can not be met through the simulation. Implications arise that make the system's distribution be trapped in wrong shapes.

The point of our main interest is when the new term of the Metropolis criterion is put to the test. This is exactly where the program's functionality improves comparing to what it was capable of doing before and this is when a connectivity-altering move is tried. The connectivity-altering move, as we have already mentioned, picks randomly two chains with initial lengths, $l_{1,\text{old}}$ and $l_{2,\text{old}}$, and afterwards the chains have two new, proposed lengths, $l_{1,\text{new}}$ and $l_{2,\text{new}}$. A "checker" routine is then called that calculates the ratio of eq.(4.8), based on the values contained in the probability density function list, $p^{\text{target}}_{l_{1,\text{old}}}$, $p^{\text{target}}_{l_{2,\text{old}}}$, $p^{\text{target}}_{l_{1,\text{new}}}$ and $p^{\text{target}}_{l_{2,\text{new}}}$. This result is then used for the advance of the existent Metropolis criterion, through the inclusion of the term: $p^{\text{target}}_{l_{1,\text{new}}} p^{\text{target}}_{l_{2,\text{new}}} / \left(p^{\text{target}}_{l_{1,\text{old}}} p^{\text{target}}_{l_{2,\text{old}}} \right)$. If the new lengths are accepted, then the program udates all the directly affected lists from the new conformation of the system and the simulation moves on to the next iteration step.

In the next chapter, Chapter 5, we display the results we gathered from various simulations we ran, testing all the newly "acquired" distributions.

Chapter 5

Simulation Results

In all our simulations the initial systems employed were prepared following the steps described in Section 4.3 of Chapter 4. All the systems were initially constructed as monodisperse, with 15 chains each of which consisted of 1000 UA units. After running the simulations, fluctuations were induced in the lengths of the chains, due to the connectivity-altering moves, resulting in polydisperse systems following the uniform, most-probable, Gaussian, bimodal Gaussian, double Gaussian and Schulz-Zimm distributions.

Besides the implementation of these analytical expressions for distributions, we managed also to realize tabulated distributions. These are distributions that are not provided through their analytical expressions (which may be unkown) and could for example have come as results from experimental processes. Each such distribution was given to us as a set of points ("*degree of polymerization*", PDF("*degree of polymerization*")).

In all following cases, the simulations ran for about 150 million iterations.

5.1 Uniform Distribution

For the realizaion of the following uniform distribution we set parameter ϕ equal to 0.4, so the half width of the simulation window is $\ell = 400$ UAs. The value of the uniform/flat distribution (Number Fraction) is by definition constant in this case, independent of the Degree of Polymerization and equal to:

$$U(k;\phi) = \frac{1}{(2\ell+1)} = \frac{1}{(2\phi X+1)} \cong 0.0012, \quad k \in [k_{min}, k_{max}] \tag{5.1}$$

since 801 is the total number of all chain lengths accesible by the chains in the range $[600, 1400]$.

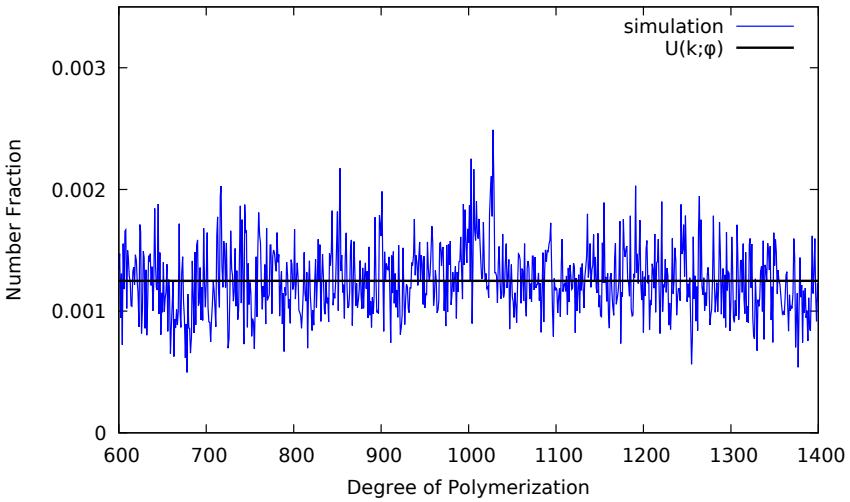

Figure 5.1: A uniform chain length distribution: black line, theoretical result; blue line, results from an $[N_{\rm ch}nPT\boldsymbol{\mu}^*]$ simulation. The average chain length is 1000 and the allowed range of chain lengths is 600–1400 and $N_{ch} = 15$.

Figure 5.2 shows us that Flory's characteristic ratio and the ratio of mean-square end-to-end distance to the mean-square radius of gyration for this system have approached the vicinities of the values 8.2 and 6, respectivery, as expected. The calculation of the ensemble average was calculated as a "running" average, i.e. we considered an ensemble which was updated every 10000 iterations with the new instance of the system. This way, the error bars are getting constantly smaller as the size of our sample is increased. Thus, any deviations from the expected values (please see below) can be interpreted as result of an unfinished simulation.

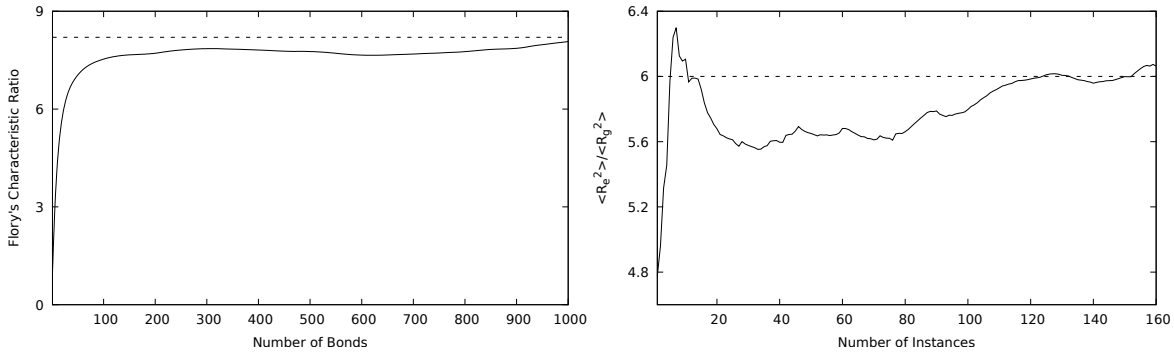

Figure 5.2: *(Left)*: Flory's characteristic ratio C_k, eq.(2.10), saturates at 8.2 for large number of bonds and *(Right)*: evolution of the ratio of eq.(2.8) as more instances from the simulation are used. 160 instances have been used.

5.2 Most-probable distribution

For the realizaion of the following most probable distribution we set parameter ϕ equal to 0.4. In this case, the half width on the left of the simulation window is $\ell = 400$ UAs and the border is at $k_{min} = X - \ell$. On the right, the distribution is theoretically free, without a border at some finite value of degree of polymerization. Technically, the necessary restriction is set at a very high value. Let us suppose that we start with a monodisperse system with degree of polymerization X for each chain. The values of the most probable distribution (Number Fraction) are by definition calculated by the formula:

$$f_{\mathrm{MP}}(k;\phi) = \frac{1}{(\phi X + 1)} \left[\frac{\phi X}{\phi X + 1}\right]^{([k-(1-\phi)X]}, \quad k \geq k_{min} \tag{5.2}$$

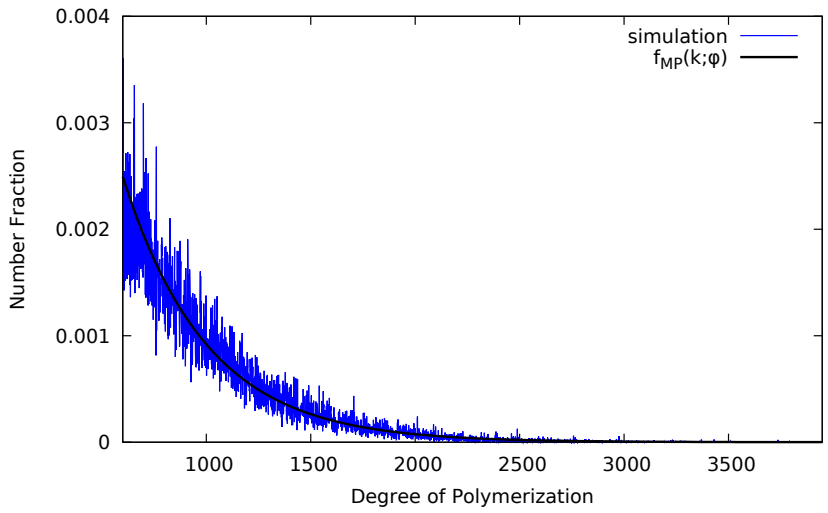

Figure 5.3: A most probable chain length distribution: black line, theoretical result; blue line, results from an $[N_{\mathrm{ch}} nPT\boldsymbol{\mu}^*]$ simulation. $N_{\mathrm{ch}} = 15$.

In contrast to the case of the uniform distribution, in Figure 5.4 we see that the system with the most probable distribution has Flory's ratio qualitatively good, but quantitatively overestimated. As far as the $\langle R_{\mathrm{e}}^2\rangle/\langle R_{\mathrm{g}}^2\rangle$ is concerned, it needs more iterations to reach the value 6. After 108 instances the relative error is about 2%.

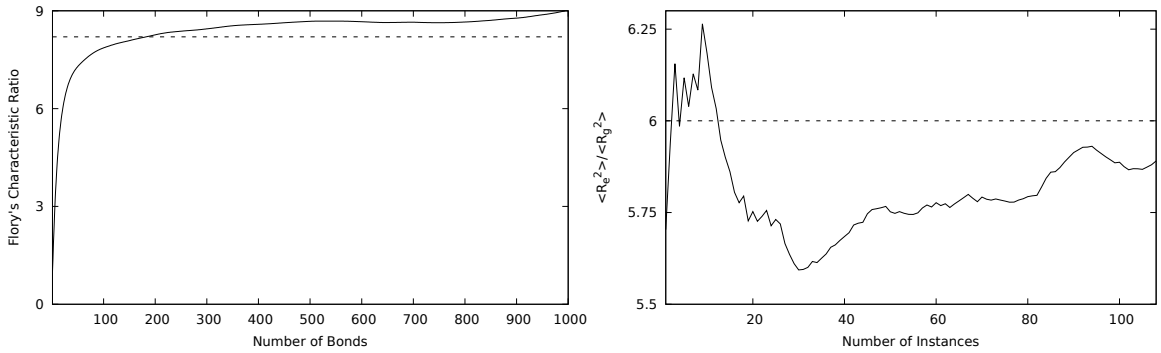

Figure 5.4: *(Left)*: Flory's characteristic ratio C_k, eq.(2.10), saturates at 8.2 for bigger number of bonds and *(Right)*: evolution of the ratio of eq.(2.8) as more instances from the simulation are used. 108 instances have been used.

5.3 Gaussian distribution

We tried our new criterion on two Gaussian distributions with different parameters. One Gaussian could be considered as "narrow" and the other as "wide".

For the realization of the "narrow" Gaussian distribution we set parameter ϕ equal to 0.4, α_l equal to 0.285714285 and X_l was left equal to 1000. This particular value of α_l was chosen because we had decided that ℓ would be equal to 3.5σ (that is why we call it "narrow"). So $\sigma = \alpha\ \phi\ X_l = 114.29$ and the 99.78% of the Gaussian bell is found inside the window.

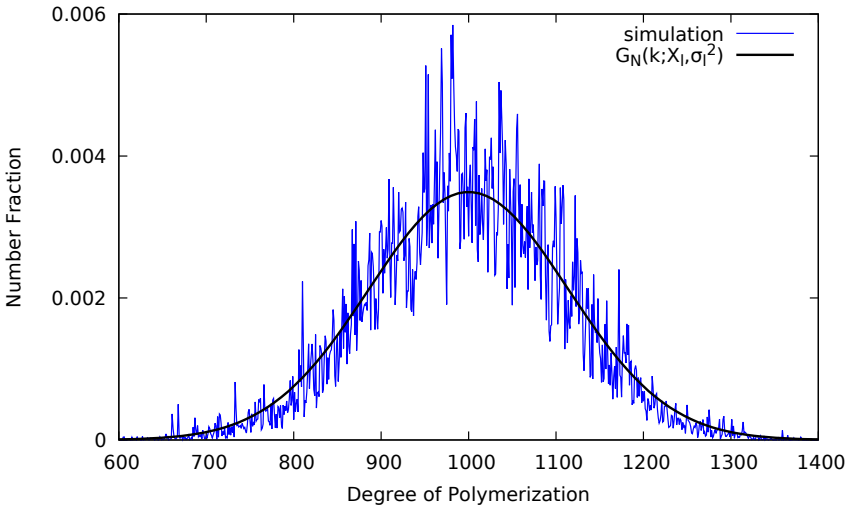

Figure 5.5: A narrow Gaussian chain length distribution: black line, theoretical result; blue line, results from an $[N_{\text{ch}}nPT\boldsymbol{\mu}^*]$ simulation. The average chain length is 1000 and the allowed range of chain lengths is 600–1400 with $N_{ch} = 15$.

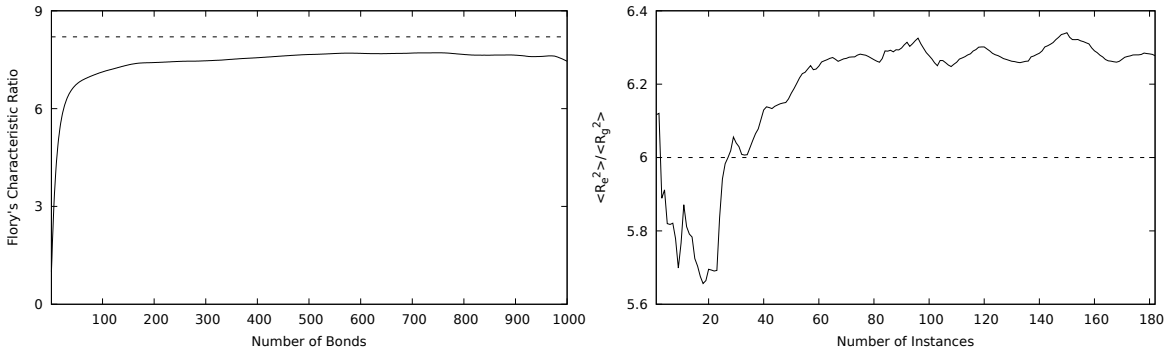

Figure 5.6: *(Left)*: Flory's characteristic ratio C_k, eq.(2.10), tries to saturate at 8.2 for large number of bonds and *(Right)*: evolution of the ratio of eq.(2.8) as more instances from the simulation are used. 182 instances have been used.

The values of the Gaussian function (Number Fraction) are calculated by the formula:

$$G(k; X_l, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(k - X_l)^2}{2\sigma^2}\right], \quad k \in [k_{\min}, k_{max}] \tag{5.3}$$

The two curves in Figure 5.6 are quite off the target values. This is a very good indication that the simulation of this system needed to run more iterations.

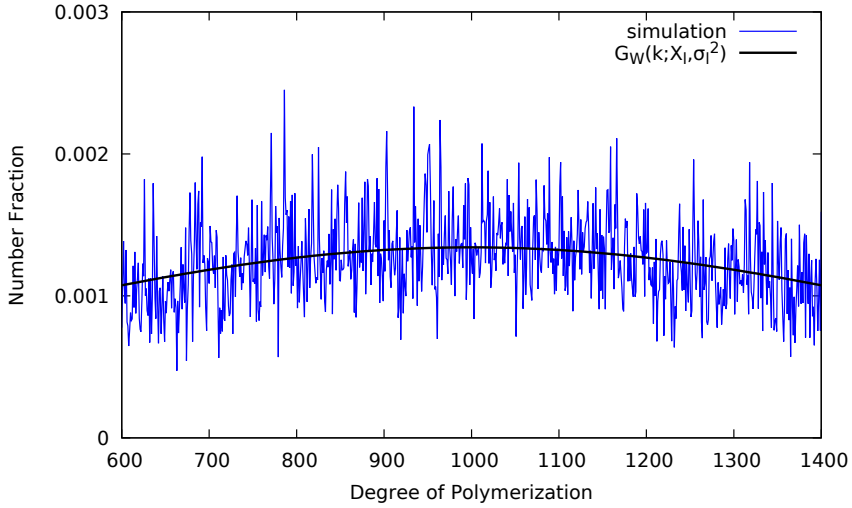

Figure 5.7: A wide Gaussian chain length distribution: black line, theoretical result; blue line, results from an $[N_{\text{ch}} nPT\boldsymbol{\mu}^*]$ simulation. The average chain length is 1000 and the allowed range of chain lengths is 600–1400 with $N_{\text{ch}} = 15$.

For the realization of the "wide" Gaussian distribution we set parameter ϕ equal to 0.4, α_l equal to 1.5 and X_l was also left equal to 1000. So $\sigma = 600$ (that is why we call it "wide") and the 49.08% of the Gaussian bell is found inside the window. The values of the distribution (Number Fraction) are calculated by the formula (5.3).

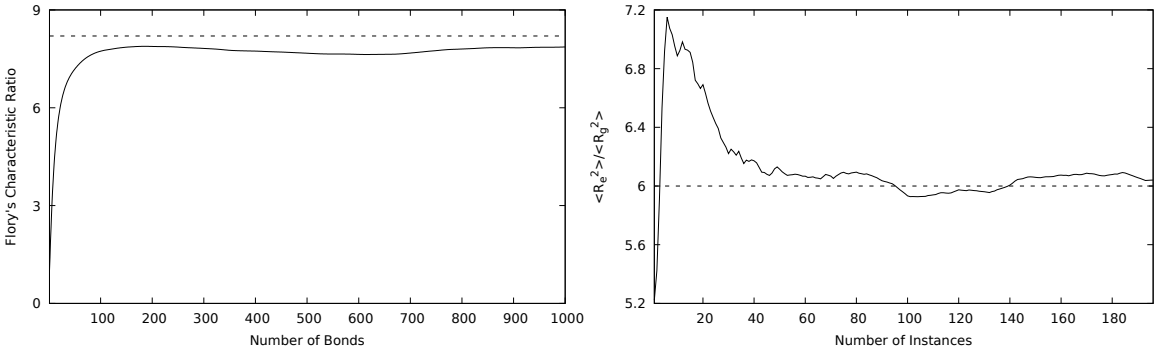

Figure 5.8: *(Left)*: Flory's characteristic ratio C_k, eq.(2.10), saturates at 8.2 for large number of bonds and *(Right)*: evolution of the ratio of eq.(2.8) as more instances from the simulation are used. 196 instances have been used.

In contrast to the case of the narrow Gaussian, the system with the wide Gaussian distribution seems to have approached the theoretical targets of its conformation. Probably its broaden, accessible degree-of-polymerization range played an important role to that.

The Gaussian (or Normal) distribution is given by a function with a domain ranging from $-\infty$ to $+\infty$. It is a well known fact, though, that computers are incapable to handling the concept of infinity. That makes the implementation of truncated Gaussians a necessity rather than an option. That means we use the Gaussian function to calculate the values of probability, and by imposing degree-of-polymerization value restraints we get a truncated distribution.

5.4 Bimodal Gaussian distribution

For the realization of the following bimodal Gaussian or "camel-like" distribution we set parameter ϕ equal to 0.4, α_l equal to 0.15, X_l equal to 800, α_r equal to 0.375 and X_r equal to 1250. In this case we have a doubly peaked distribution, thus we have to provide the positions of two peaks and two standard deviations for the two Gaussians. We used different values for the standard deviations on purpose: to avoid a symmetric distribution and have the chance to show the power of the program for the implementation of the most general cases. In the asymmetric case, we expect the resultant distribution to be defined in a displaced integral. So $\sigma_1 = 60$

and $\sigma_{\mathrm{r}} = 150$, the left Gaussian is very narrow, while the right one is more wide. The values of the bimodal Gaussian distribution (Number Fraction) are calculated by the formula:

$$G_{\mathrm{bimodal}}(k; X_l, \sigma_l^2, X_r, \sigma_r^2) = G(k; X_l, \sigma_l^2) + G(k; X_r, \sigma_r^2), \quad k \in [k_{\min}, k_{\max}] \tag{5.4}$$

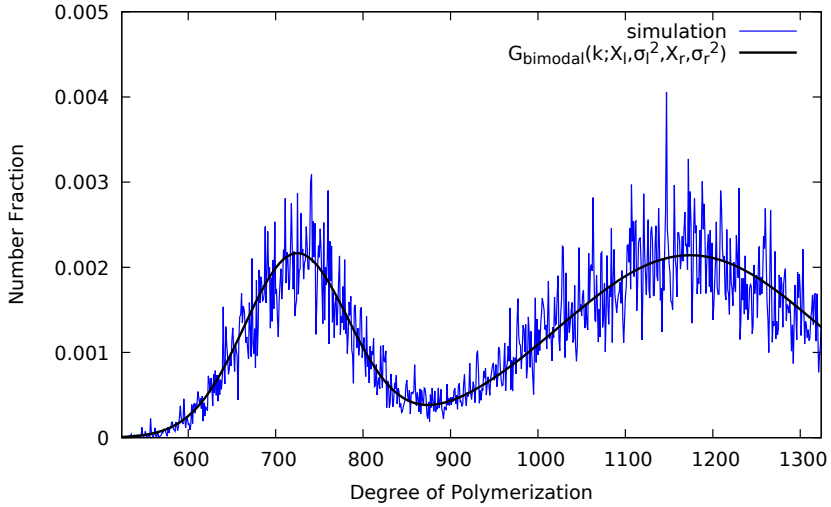

Figure 5.9: A bimodal Gaussian chain length distribution: black line, theoretical result; blue line, results from an $[N_{\mathrm{ch}}nPT\boldsymbol{\mu}^*]$ simulation. The average chain length is 1000 and the allowed range of chain lengths is 524–1324 with $N_{\mathrm{ch}} = 15$.

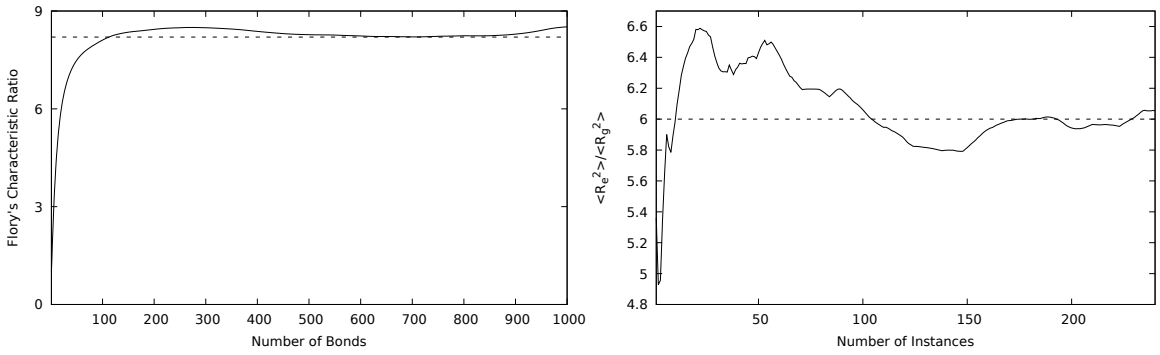

Figure 5.10: *(Left)*: Flory's characteristic ratio C_k, eq.(2.10), tries to saturate at 8.2 for large number of bonds and *(Right)*: evolution of the ratio of eq.(2.8) as more instances from the simulation are used. 240 instances have been used.

It seems from Figure 5.10 that Flory's characteristic ratio and $\langle R_{\mathrm{e}}^2\rangle/\langle R_{\mathrm{g}}^2\rangle$ values have reached successfully the vicinity of the theoretically expected values.

We also realized another bimodal Gaussian (see Figure 5.11) that is symmetric, with narrower peaks and closer together, in order for its tails to become almost zero inside the simulation window. We also set parameter ϕ equal to 0.4, but this time $\alpha_{\rm l} = \alpha_{\rm r}$ equal to 0.175, $X_{\rm l}$ equal to 815 and $X_{\rm r}$ equal to 1150.

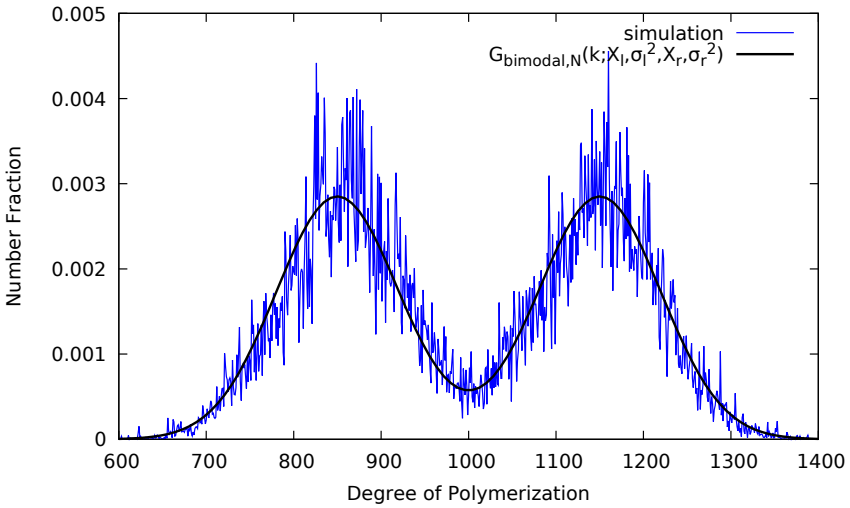

Figure 5.11: A bimodal with narrow Gaussians chain length distribution: black line, theoretical result; blue line, results from an $[N_{\rm ch}nPT\boldsymbol{\mu}^*]$ simulation. The average chain length is 1000 and the allowed range of chain lengths is 600–1400 with $N_{\rm ch} = 15$.

In Figure 5.12 we see that Flory's characteristic ratio value has approached nicely the value 8.2. Quantity $\langle R_{\rm e}^2\rangle/\langle R_{\rm g}^2\rangle$, also, has approached the vicinity of 6, a tendency to "fall" and bonce towards there can be observed.

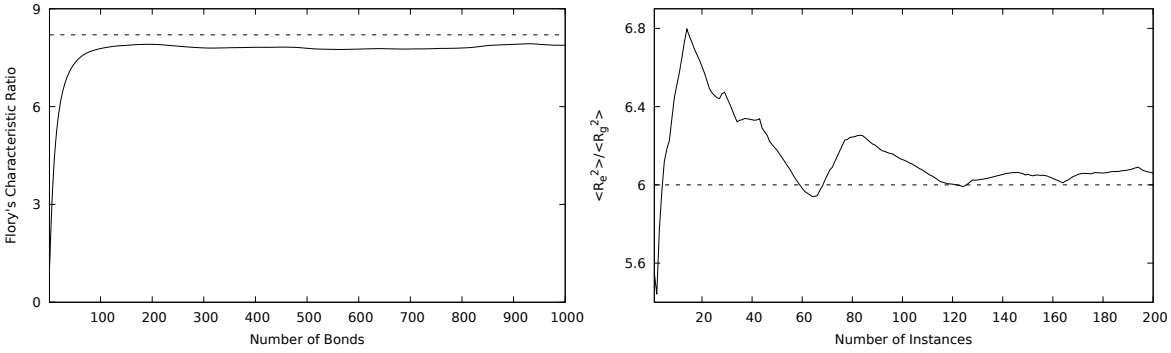

Figure 5.12: *(Left)*: Flory's characteristic ratio C_k, eq.(2.10), saturates at 8.2 for large number of bonds and *(Right)*: evolution of the ratio of eq.(2.8) as more instances from the simulation are used. 200 instances have been used.

5.5 Double Gaussian or Binormal distribution

For the realization of the following double Gaussian or two-piece normal or binormal or Fechner distribution we set parameter ϕ equal to 0.4, α_l equal to 0.2, α_r equal to 0.4 and X_l was left equal to 1000. In this case we have a chimeric Gaussian distribution consisting of two other Gaussians. This is another case of a displaced asymmetric distribution that is not placed in the range $[600, 1400]$ (see Figure 5.13).

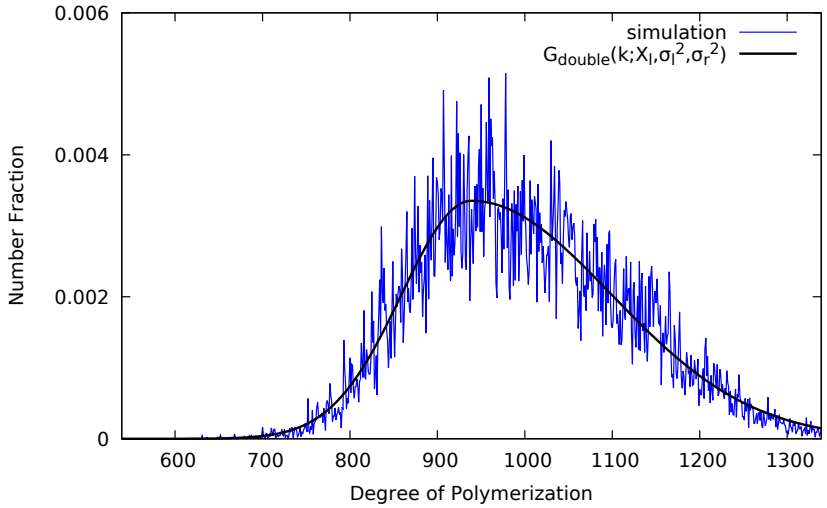

Figure 5.13: A double Gaussian chain length distribution: black line, theoretical result; blue line, results from an $[N_{\text{ch}}nPT\boldsymbol{\mu}^*]$ simulation. The average chain length is 1000 and the allowed range of chain lengths is 539–1339 with $N_{\text{ch}} = 15$.

The distribution is formed by taking the left part of a Gaussian bell with parameters $(X_l,\ \sigma_l^2)$ and the right part of a Gaussian bell with parameters $(X_l,\ \sigma_r^2)$. So $\sigma_l = 160$ and $\sigma_r = 80$. The values of the double Gaussian distribution (Number Fraction) are calculated by the formula:

$$G_{\text{double}}(k; X_l, \sigma_l^2, \sigma_r^2) = \begin{cases} G(k; X_l, \sigma_l^2)/A & \text{for} \quad k \leq X_l \\ G(k; X_l, \sigma_r^2)/A & \text{for} \quad k > X_l \end{cases} \tag{5.5}$$

where the G functions are calculated by the formula (5.3) and $A = 2/(\sqrt{2\pi}(\sigma_l + \sigma_r))$ is the normalization factor.

We can see in Figure 5.14 that in the case of this system, ratio $\langle R_{\text{e}}^2\rangle/\langle R_{\text{g}}^2\rangle$ is very close to the desired theoretical value, while Flory's characteristic ratio is qualitatively but not yet quantitatively there.

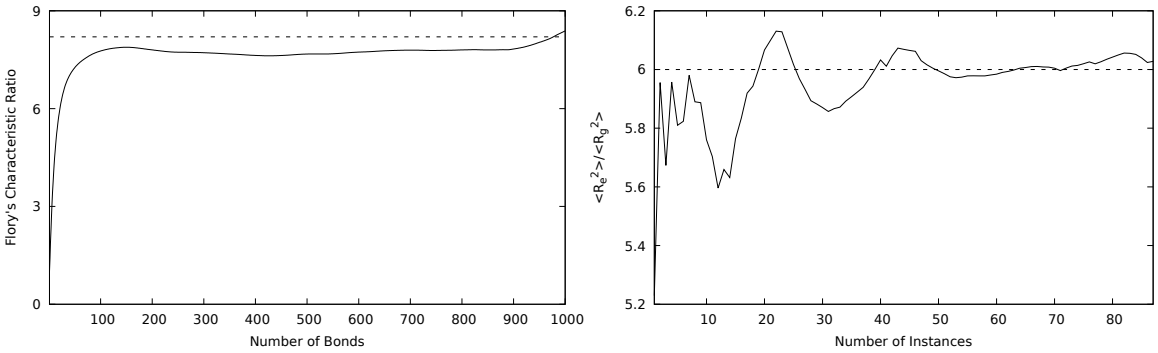

Figure 5.14: *(Left)*: Flory's characteristic ratio C_k, eq.(2.10), saturates at 8.2 for large number of bonds and *(Right)*: evolution of the ratio of eq.(2.8) as more instances from the simulation are used. 87 instances have been used.

5.6 Schulz-Zimm distribution

For the realization of the Schulz-Zimm distribution we used the following formula (see Qi *et al.* (2016), eq.(2)):

$$f_{\mathrm{SZ}}(k;\alpha_l,X_l)=\frac{\alpha_l^{\alpha_l}k^{(\alpha_l-1)}}{\Gamma(\alpha_l)X_l^{\alpha_l}}\exp\left(-\alpha_l\frac{k}{X_l}\right),\quad k\in[k_{\min},k_{\max}] \tag{5.6}$$

where the Schulz-Zimm two-parameter function is expressed in terms of $\alpha_l \coloneqq \lambda$ which is a parameter related to the polydispersity index, $X_l \coloneqq N_{\mathrm{n}}$ is the number-average degree of polymerization and Γ is the Gamma function.

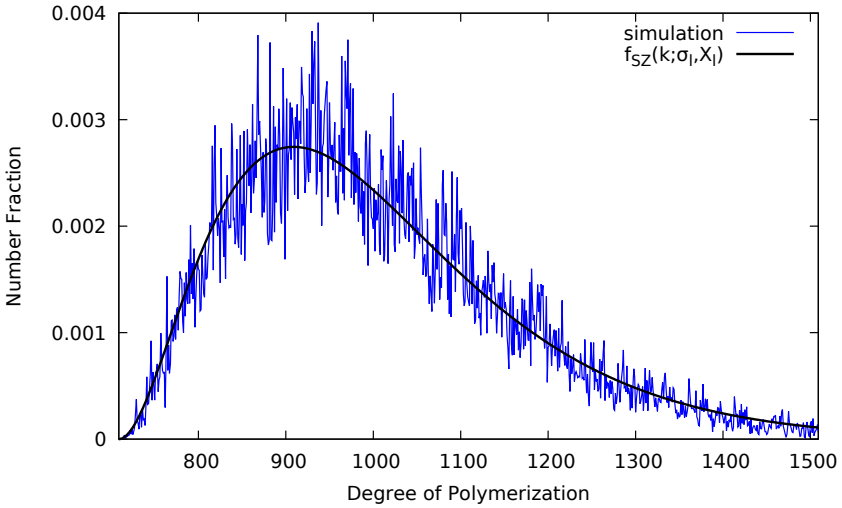

Figure 5.15: A Schulz-Zimm chain length distribution: black line, theoretical result; blue line, results from an $[N_{\mathrm{ch}}nPT\boldsymbol{\mu}^*]$ simulation. The average chain length is 1000 and the allowed range of chain lengths is 709–1509 with $N_{\mathrm{ch}} = 15$.

The relation between λ and the PDI is: $N_{\mathrm{w}}/N_{\mathrm{n}} \coloneqq \mathrm{PDI} = 1 + 1/\lambda$ [15] (where N_w is the weight-averaged degree of polymerization). We set parameter α_{l} equal to 3, parameter X_{l} was set equal to 800 and ϕ equal to 0.4.

From Figure 5.16 we understand that this system needed more iteration steps for its conformation characteristic values to reach their theoretically expected values. Flory's ratio tries to saturate at value 8.2 and $\langle R_{\mathrm{e}}^2 \rangle / \langle R_{\mathrm{g}}^2 \rangle$ seems to reveal a tendency to climb up to value 6.

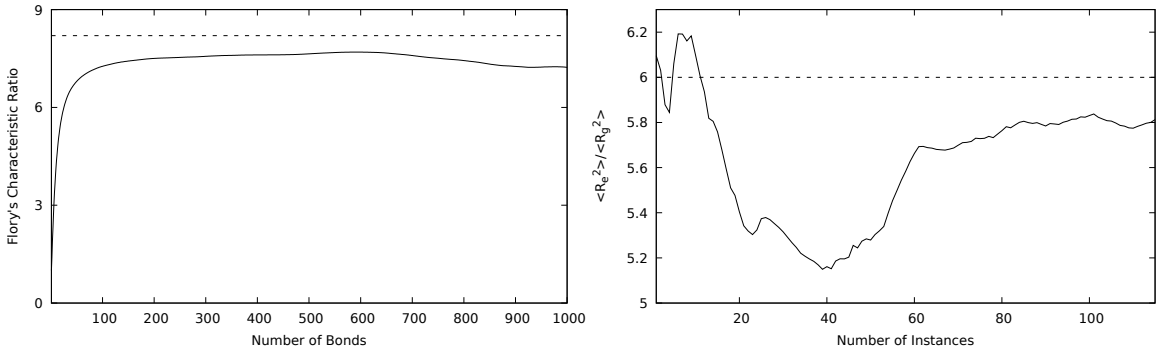

Figure 5.16: *(Left)*: Flory's characteristic ratio C_k, eq.(2.10), saturates at 8.2 for bigger number of bonds and *(Right)*: evolution of the ratio of eq.(2.8) as more instances from the simulation are used. 115 instances have been used.

5.7 Tabular random distribution

Finally, just to present an example of how good the new criterion can be implemented for the realization of a random tabulated distribution, we created a non-analytical random distribution, with extremely large first-derivative values at certain positions, step behaviour and even zeroing for k around 1350.

As it can be seen in Figure 5.17 the distribution of the simulated system has converged quite accurately and with further iteration steps the fluctuations are expected to become even smaller. The one needle peak that is observed can be ignored by virtue of the semi-continuous integral of the simulation window.

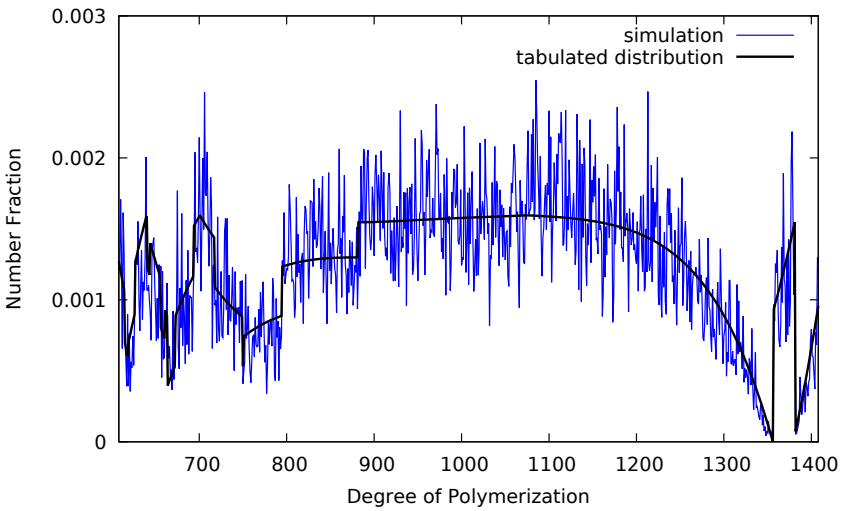

Figure 5.17: Random tabulated chain length distribution: black line, theoretical result; blue line, results from an $[N_{\mathrm{ch}}nPT\boldsymbol{\mu}^*]$ simulation. The average chain length is 1000 and the allowed range of chain lengths is 608–1408 with $N_{\mathrm{ch}} = 15$.

Chapter 6

Conclusions and Future Work

In this chapter we shall summarize the contributions of this thesis and conclude with reccomendations for future work.

6.1 Conclusions

This thesis has successfully prepared the Metropolis criterion needed for MC simulations employing connectivity-altering moves to impose preselected degree-of-polymerization distributions on polymer melt systems. Our reference piece of literature was an article written by Pant and Theodorou (1995) [13] in which, as we have multiple times and extensively described in Section 2.5 and Chapter 4, a thermodynamical analysis of the semigrand $[N_{\text{ch}}nPT\boldsymbol{\mu}^*]$ ensemble is presented which can describe the polydisperse systems of our simulations.

The direct advancement of the existing Metropolis criterion of our model, taking inspiration from the relative chemical potentials of the article, would require the calculation of Boltzmann-like factors of the chemical energy of both the old and the proposed/new conformations. Instead, we developed and propose a simpler criterion, which we prove in Chapter 4, that uses the products of target probabilities of the created and the destroyed species. The simplicity of this criterion arises from the facts that: (a) we do not have to know in advance the relative chemical potential from which the desired distribution results (in certain cases this could be a formidable task), (b) we do not have to calculate the values of relative chemical potentials (which could end up to be quite computation time consuming), (c) we have to know only the values of the target distribution we wish to realize and (d) the target distribution of interest, although it is expressed as a PDF, does not have to be normalized by virtue of the ratio inside the Metropolis criterion. The final program was also used with a tabulated distribution that succesfully approximated, providing a de facto example of the "power" of our criterion.

The second contribution of this thesis is a list of three relative chemical potentials, in the same philosophy as the ones proposed by Pant and Theodorou (1995) [13] (see eq.(28), eq.(31), eq.(34)), for three new analytically expressed distributions: truncated bimodal Gaussian, truncated double Gaussian and Schulz-Zimm distributions, which we presented in Section 4.2. They are intended to provide a more complete approach to the newly realized distributions, adopting the approach used in our article of reference [13].

In conclusion, this new criterion seems to open an efficient way towards the simulation of more realistic polymeric materials, with respect to the length/molecular-weight/degree-of-polymerization distribution of the sample. It would not be a surprise if the new final program were to assist with the simulation and study of polymers of industrial interest or promising for everyday applications. These are cases where the insertion and control of polydispersity in the simulated systems could ultimately be proved enlightening.

6.2 Future Work

- First and foremost the next step would be to implement the criterion in a parallel program that would also enable us to test bigger, more complicated and more entangled systems with respect to the size of the polymeric chains or their number.

- Besides the graphs of Flory's characteristic ratio and the evolution of ratio $\langle \boldsymbol{R}_e^2 \rangle / \left\langle \boldsymbol{R}_g^2 \right\rangle$ with respect to the simulation step, presented in Chapter 5, there must be conducted further and more systematic study on the conformational, volumetric, thermal and even dynamical characteristics and crystalization mechanisms of the resultant polydisperse systems, in order to be ensured that the proposed Metropolis criterion does not spoil them.

- Analysis of the entanglements in polydisperse systems with different molecular weight distributions (in conjuction with the CReTA approach) would be an interesting possibility.

Bibliography

[1] N. Afrasiabian. The journey of a single polymer chain to a nanopore, 2022. URL `https://ir.lib.uwo.ca/etd/6966`.

[2] M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids*. Oxford University Press, 1991.

[3] S. Domanskyi, D. T. Gentekos, V. Privman, and B. P. Fors. Predictive design of polymer molecular weight distributions in anionic polymerization. *Polym. Chem.*, 11:326–336, 2020. doi: 10.1039/c9py00074g.

[4] W. Edwards, H. Lindman, and L. J. Savage. Bayesian Statistical Inference for Psychological Research. 1963.

[5] D. T. Gentekos, R. J. Sifri, and B. P. Fors. Controlling polymer properties through the shape of the molecular-weight distribution. *Nat. Rev. Mater.*, 4: 761–774, 2019. doi: 10.1038/s41578-019-0138-8.

[6] V. A. Harmandaris. *Atomistic Molecular Dynamics Simulations of Polymer Melt Viscoelasticity*. PhD thesis, University of Patras, 2001.

[7] N. C. Karayiannis, A. E. Giannousaki, V. G. Mavrantzas, and D. N. Theodorou. Atomistic Monte Carlo simulation of strictly monodisperse long polyethylene melts through a generalized chain bridging algorithm. *Journal of Chemical Physics*, 117(11):5465–5479. doi: 10.1063/1.1499480.

[8] N. C. Karayiannis, V. G. Mavrantzas, and D. N. Theodorou. A Novel Monte Carlo Scheme for the Rapid Equilibration of Atomistic Model Polymer Systems of Precisely Defined Molecular Architecture. *Physical Review Letters*, 88(10): 105503–105506, 2002. doi: 10.1063/1.1499480.

[9] N. C. Karayiannis, V. G. Mavrantzas, D. Mouratides, E. Chiotellis, and C. Kiparissides. Chapter 6 - Atomistic Molecular Dynamics simulation of short–chain branched polyethylene melts. In M. Laso and E. Perpète, editors, *Multiscale Modelling of Polymer Properties*, volume 22 of *Computer*

Aided Chemical Engineering, pages 333–357. Elsevier, 2006. doi: 10.1016/S1570-7946(06)80016-7.

[10] A. Kumar and R. K. Gupta. *Fundamentals of Polymer Engineering*. Marcel Dekker, 2nd edition, 2003.

[11] V. G. Mavrantzas. Using Monte Carlo to Simulate Complex Polymer Systems: Recent Progress and Outlook. *Frontiers in Physics*, 9, 2021. doi: 10.3389/fphy.2021.661367.

[12] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, 21(6): 1087–1092, June 1953.

[13] P. V. K. Pant and D. N. Theodorou. Variable Connectivity Method for the Atomistic Monte Carlo Simulation of Polydisperse Polymer Melts. *Macromolecules*, 28(21):7224–7234, 1995.

[14] A. J. Peacock and A. Calhoun. *Polymer Chemistry: Properties and Applications*. Hanser, 2006.

[15] S. Qi, L. I. Klushin, A. M. Skvortsov, and F. Schmid. Polydisperse polymer brushes: internal structure, critical behavior, and interaction with flow. *Macromolecules*, 49(24):9665–9683, 2016. doi: 10.1021/acs.macromol.6b02026.

[16] N. A. Rorrer. *Efficient Simulation of Polymer Melt Viscoelasticity by Dynamic Monte Carlo Simulation*. PhD thesis, Colorado School of Mines, 2015.

[17] M. Rubinstein and R. Colby. *Polymer Physics*. Oxford University Press, 2003.

[18] D. Schöberl. Monte Carlo Simulations of Polyethylene Melts. Master's thesis, Johannes Kepler University Linz, 2020.

[19] J. B. P. Soares and T. F. L. McKenna. *Polyolefin Reaction Engineering*. WILEY-VCH, 1st edition, 2012.

[20] D. N. Theodorou. *Variable-Connectivity Monte Carlo Algorithms for the Atomistic Simulation of Long-Chain Polymer Systems*, pages 67–127. Springer Berlin Heidelberg, 2002. doi: 10.1007/3-540-45837-9_3.

[21] D. J. Walsh, D. A. Schinski, R. A. Schneider, and D. Guironnet. General route to design polymer molecular weight distributions through flow chemistry. *Nature Communications*, 11(1), 2020. doi: 10.1038/s41467-020-16874-6.

[22] C. Wang. *Applications of Monte Carlo Methods in Studying Polymer Dynamics*. PhD thesis, University of Reading, June 2018.

[23] R. Whitfield, N. P. Truong, D. Messmer, K. Parkatzidis, M. Rolland, and A. Anastasaki. Tailoring polymer dispersity and shape of molecular weight distributions: methods and applications. *Chem Sci.*, 10(38):8724–8734, 2019. doi: 10.1039/c9sc03546j.

[24] J. Zhang, J. Pu, and M. Ren. Molecular Weight Distribution Control for Polymerization Processes Based on the Moment-Generating Function. *Entropy*, 24 (499), 2022. doi: 10.3390/e24040499.

[25] Α. Μόρφης. Υπολογιστικά αποτελεσματικές μοριακές προσομοιώσεις πολυμερικών τηγμάτων. Master's thesis, Εθνικό Μετσόβιο Πολυτεχνείο, 2012.

CURRICULUM VITAE

Personal/Contact Information

Name	**Dimitrios-Paraskevas E. Gerakinis**
Date of Birth	August 9, 1997
Place of Birth	Athens, Greece
Address	Mikras Asias 62, Athens, 115 27, Greece
Nationality	Greek
e-mail	dpgerakinis@gmail.com
LinkedIn	www.linkedin.com/in/dimitris-gerakinis

Education

10.2020 – 07.2022	Graduate Studies in Science and Technology of Materials, School of Chemical Engineering, National Technical University of Athens
10.2015 – 09.2020	Undergraduate Studies in Applied Physical and Applied Mathematical Sciences, School of Applied Mathematical and Physical Sciences, National Technical University of Athens Thesis Title :: “Photoluminescence spectroscopy measurements in pure and Br-doped perovskite CsPbBr3”, Supervisor :: Dr. Athanasios Kontos Diploma Grade :: 8.49

Experience

08.2020 – 0.9.2020	Internship in the laboratory of “Molecular Thermodynamics and Modeling of Materials” of the National Center for Scientific Research "Demokritos”, on Computational Materials Science Supervisor :: Dr. Niki Vergadou

Workshops/Symposiums/Masterclasses

15.09 – 22.10.2021	Regeneration Academy on Data Science powered by Papastratos
19.12 – 20.12.2020	Workshop on “Computational Materials Science” co-organized by the Department of Materials Science and

	Technology, University of Crete and the Hellenic Society for the Science and Technology of Condensed Matter
23.11 – 28.11.2020	"Global Quantum Programming Workshop" organized by Qworld
11.2018	Education Symposium on "Laser applications in medicine. Clinical implementations and secure use" organized by the National Kapodistrian University of Athens
10.07 – 13.07.2018	6th Hellenic Forum for Science Technology and Innovation organized by the National Center for Scientific Research "Demokritos"
02.07 – 13.07.2018	53rd Summer School organized by the National Center for Scientific Research "Demokritos"

Presentations at Conferences

16.12.2017	**Gerakinis D. P.** "Hume's crystal glass; the problem of induction and Popper's response", 2nd Greek Undergraduate Philosophy Conference, School of Fine Arts, Athens, December 15th-17th, 2017 (oral presentation)

Distinctions

12.2015	2nd prize in the prose category of the 1st Literature Competition of the Municipality of Byron in Athens with the short story: "Byron's daughter"
12.2013	1st place with High School's team in the "Physics Laboratory" part of the preliminary phase of the "Hellenic Student Competition for the 12th European Union Olympiad – EUSO 2014"

Languages

Greek	Mother tongue
English	Fluent (C2 level)
French	Good (B2 level)

Software Skills

Operating Systems	Linux, Windows

Programming	Fortran, Python, Java, Matlab, T-SQL
Knowledgeable in	gnuplot, Emacs, vi, bash shell environment, LibreOffice, Power BI, Jupyter Notebook, LaTeX, data cleaning, data analysis, data visualization, ML modeling for predictions

Interests/Hobbies

Sports	karate (4th kyu), running
Musical instruments	piano, recorder
Hobbies	writing short stories and bonsai stories
Other	Member of the Municipality of Byron's Youthful Company for Literature (since 2015) Member of the Debating club of Athens University of Economics and Business (since 10/2018)

Volunteer Work

11.2018 – **today**	Part of "TED Translators" community
09.2018	"Researchers' Night NTUA": aiding in breaking the news of the event to the public and guiding the participants throught the event venues

Athens, July 2022