

Excess Entropy and Dynamics in Coarse-Grained Models

Zur Erlangung des Grades eines Doktors der Naturwissenschaften (Dr. rer. nat.)
Genehmigte Dissertation von Gustavo Garcia Rondina aus Marília, Brasilien
Tag der Einreichung: 29. April 2022, Tag der Prüfung: 20. Juni 2022

1. Gutachten: Prof. Dr. Florian Müller-Plathe
2. Gutachten: Prof. Dr. Nico van der Vegt
Darmstadt, Technische Universität Darmstadt



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Chemistry Department

Eduard-Zintl-Institute for
Inorganic and Physical
Chemistry

Theoretical Physical
Chemistry

Excess Entropy and Dynamics in Coarse-Grained Models

Accepted doctoral thesis by Gustavo Garcia Rondina

Date of submission: 29. April 2022

Date of thesis defense: 20. Juni 2022

Darmstadt, Technische Universität Darmstadt

Bitte zitieren Sie dieses Dokument als:

URN: urn:nbn:de:tuda-tuprints-212218

URL: <http://tuprints.ulb.tu-darmstadt.de/21221>

Jahr der Veröffentlichung auf TUprints: 2022

Dieses Dokument wird bereitgestellt von tuprints,
E-Publishing-Service der TU Darmstadt

<http://tuprints.ulb.tu-darmstadt.de>

tuprints@ulb.tu-darmstadt.de

Die Veröffentlichung steht unter folgender Creative Commons Lizenz:

Namensnennung – Weitergabe unter gleichen Bedingungen 4.0 International

<https://creativecommons.org/licenses/by-sa/4.0/>

This work is licensed under a Creative Commons License:

Attribution–ShareAlike 4.0 International

<https://creativecommons.org/licenses/by-sa/4.0/>

In memory of Hamilton Rondina and Michael Böhm

Erklärungen laut Promotionsordnung

§ 8 Abs. 1 lit. c PromO

Ich versichere hiermit, dass die elektronische Version meiner Dissertation mit der schriftlichen Version übereinstimmt.

§ 8 Abs. 1 lit. d PromO

Ich versichere hiermit, dass zu einem vorherigen Zeitpunkt noch keine Promotion versucht wurde. In diesem Fall sind nähere Angaben über Zeitpunkt, Hochschule, Dissertationsthema und Ergebnis dieses Versuchs mitzuteilen.

§ 9 Abs. 1 PromO

Ich versichere hiermit, dass die vorliegende Dissertation selbstständig und nur unter Verwendung der angegebenen Quellen verfasst wurde.

§ 9 Abs. 2 PromO

Die Arbeit hat bisher noch nicht zu Prüfungszwecken gedient.

Darmstadt, 29. April 2022

Gustavo Garcia Rondina

Acknowledgements

Starting a Ph.D. is not a decision to be taken lightly. It affects all parts of one's life. And, as often happens with life itself, the Ph.D. may not go as expected. Some ideas don't work out. Not all efforts pay off. Life happens in-between. But it grants more than just knowledge on the subject matter that was studied. Curiosity, wonder, awe. But also resilience, fortitude, tenacity, grit. And come good or bad times, it can be said with certainty that no one has ever gotten through their doctoral studies alone. As with children, it takes a village. And it is this village with all the people in my life who helped me get through this that I would like to take a moment to appreciate and acknowledge.

First and foremost, I would like to thank my advisor, Prof. Florian Müller-Plathe. We met over ten years ago when he gave me my first opportunity in Germany as a HiWi in his group. Since then, he has, time and again, been there for me. With invaluable scientific advice, professional guidance, fruitful discussions, and unlimited support in every matter I ever brought to his attention. I am sincerely grateful for the insights, wit, patience, and support.

My wife Tatiana has been the pillar of my strength. She has shown love, companionship, unconditional support, and encouragement throughout these years. She refused to stop believing in me, not for a second, even when I was near the end of my hope. Without her presence in my life, this would have never ended. I am humbled and wholeheartedly grateful for her commitment, patience, and simply being there for me when needed.

I am especially grateful for having had Prof. Michael Böhm in my life as a friend and mentor. He was the kindest person I have ever met, and everyone who met Michael was made better by knowing him. One of my greatest joys was being able to help him surpass 333 publications, even if with just the one paper that we co-authored together. Michael, thank you.

I have immense gratitude to my parents, Mara and Hamilton, for all their sacrifices, large and small, and for having always supported my career choices. My father remains my greatest role model of hard work, honesty, and compassion. His legacy lives on, and even though he was not physically by my side during my studies, I have no doubt that he was with me all this time. I am grateful to him and my mother for all the opportunities that they gave me and which allowed me to be where I am now.

I am very thankful to my aunt Marcia. She unknowingly instilled in me the desire to pursue a path of ideas and learning. I can still picture her, working on her thesis many years ago, sitting behind an old typewriter and surrounded by piles of books. That image became my reference for intellectual work and has remained with me ever since. My parents-in-law, Zelia and Heraldo, have also contributed to making this journey lighter. Heraldo is one of the most thoughtful and selfless individuals anyone has ever met. I cannot thank him enough for everything he has done and continues to do. My appreciation goes to my aunt Hioshie as well. She has been there for my wife and me, always with gentle and supportive words.

My gratitude is also extended to the members of the Müller-Plathe group. Dana Voss deserves special mention. She is always generously assisting with all kinds of problems that everyone brings her, and it is impossible to not be cheered up after passing by Dana's office. I also thank Prof. Hans-Jürgen Bär for the opportunity to work with the clusters, the trust he placed in me, and the many delicious cakes from Mrs. Bär. The colleagues with whom I shared several hours of interesting conversation and many laughs are highly appreciated - Andreas Kalogirou, Kenkoh Endoh, and Jurek Schneider. Our time together was sadly too short and I will cherish it forever. I am also grateful to the former students I had the pleasure of supervising: Pascal Lerner, Andreas Lehr, Yash Jain, Linsey Nowack, and also all the students in the course BCOM. I am confident that I learned as much (and probably more) from them as they did from me. My sincere thanks to all colleagues and research collaborators, including Prof. Hossein Eslami, Frédéric Leroy, Tinashe Ndor, Armin Shayeghi, Prof. Markus Arndt, Evangelos Voyiatzis, Mahshid Rahimi, Michael Langeloth, Haoxue Han, Prof. Jayant K. Singh, Prof. Hong Liu, Shengyuan Liu, Sarah Fegan, Yunfeng Mao, Vikram Reddy Ardham, Saurav Prasad, Shubhadip Das, Zhenghao Wu, Melissa Meinel, Prof. Hari Krishna Chilukoti, and Prof. Donatas Surblys.

I am also grateful to Prof. Nico van der Vegt for being a reviewer of this thesis and for the many interactions with him and his group, especially in the joint weekly seminars. I thank Prof. Vera Krewald and Prof. Markus Biesalski for kindly accepting to be part of my thesis committee.

I gratefully acknowledge the financial support from the Brazilian National Council for Scientific and Technological Development (CNPq), and from the German Research Foundation (DFG). The compute time in the JURECA supercomputer granted by the Jülich Supercomputing Centre (JSC) is highly valued. The work of Marei Peischl and collaborators in preparing the L^AT_EX templates that made typesetting this thesis very convenient is gratefully acknowledged.

Zusammenfassung

Coarse-Graining-Methoden sind wichtige Technik für Multiskalensimulationen, die in den letzten drei Jahrzehnten weiterentwickelt und angewendet wurde. Mit diesen Methoden können wesentlich größere Modelle als in rein atomistischen Systemen und für Zeitskalen simuliert werden, die mit All-Atom-Modellen nicht erreicht werden können. Diese Erweiterungen ergeben sich aus der Entfernung von unnötigen Freiheitsgraden für die Untersuchung des konkreten Problems. Dadurch wird die molekulare Reibung aus dem System entfernt, und die Flächen der potentiellen Energie werden weicher und weniger kompliziert, wodurch die Dynamik insgesamt stark beschleunigt wird. Diese Beschleunigung ist wünschenswert, um größere Längen- und Zeitskalen zu erreichen. Sie führt jedoch auch zu einer Beeinträchtigung der aus den Simulationen berechneten dynamischen Eigenschaften und macht es schwierig, diese Größen mit experimentellen Daten zu vergleichen. Zum Beispiel steigen die Selbstdiffusionskoeffizienten von Coarse-Grained-Systemen um bis zu drei Größenordnungen im Vergleich zu den entsprechenden atomistischen Systemen oder experimentellen Werten. Auch andere dynamische Eigenschaften wie die Scherviskosität oder lokale Relaxationszeiten werden stark beeinflusst. Eine falsche Systemdynamik ist wissenschaftlich unbefriedigend, und sie verhindert auch die Anwendung von Multiskalenkonzepten für die Berechnung von technologisch wichtigen dynamischen und Transporteigenschaften. Beispiele hierfür sind rheologische Eigenschaften, die bei der Extrusion und dem Spritzgießen von Polymerschmelzen eine wichtige Rolle spielen. Diese Arbeit untersucht, ob die künstlich beschleunigte Dynamik von Coarse-Grained-Modellen in Bezug auf die Variation der Exzessentropie zwischen Kugel-Feder-Modellen von Polymeren bei verschiedenen Modellierungsaufösungen quantifiziert werden kann. Die Beschleunigung der Dynamik von Coarse-Grained-Modellen, und damit die Entfernung von Monomer-Reibung, wird als Unterschiede der Exzessentropie verschiedener Coarse-Grained-Auflösungen ausgedrückt. Wir beginnen mit einem einfachen Kugel-Feder-Modell von unverschlaufte Polymerketten in einer Schmelze. Daraus entwickeln wir zwei weniger detaillierte Modelle nach einem systematischen Ansatz mit Coarse-Graining-Techniken. Die korrekte Entropie wird für die drei Modelle durch strenge thermodynamische Integration berechnet. Es wird auch eine Approximation verwendet, die nur die Anteile der Paarkorrelationen an der Exzessentropie enthält. Die Unterschiede der Exzessentropie über die Auflösungen der Modelle werden dann mit der Beschleunigung der Dynamik korreliert. Bemerkenswert ist, dass die Unterschiede in der Exzessentropie zwischen den gröberen und den feineren Modellen signifikant mit dem Logarithmus des Quotienten der dynamischen Eigenschaften korrelieren, die die Beschleunigung bei Änderung der Modellauflösung quantifizieren. Die Ergebnisse zeigen, dass es möglich ist, die durch die coarse-graining getriebene Beschleunigung in Form von Unterschieden in der Exzessentropie für unverschlaufte Polymere zu verstehen. Diese Korrelationen eröffnen neue Möglichkeiten für einfache empirische Methoden zur a posteriori Korrektur der Coarse-Grained-Dynamik.

Abstract

Coarse-graining modeling is a leading technique in multiscale simulations that has been continuously developed and applied in the past three decades. With coarse-graining, it is possible to simulate significantly larger systems than in purely atomistic systems and for time scales that cannot be reached with all-atom models. Such enhancements arise from removing unnecessary degrees of freedom to study the problem at hand. Consequently, molecular friction is removed from the system, and the potential energy surfaces become smoother and less intricate, leading to a significant acceleration of the overall dynamics. Even though this acceleration is favorable and desirable in achieving larger length scales and longer time scales, it inevitably affects the dynamical properties extracted from the simulations and thus makes it difficult to compare these quantities with experimental data. For example, self-diffusion coefficients of coarse-grained systems increase by up to three orders of magnitude with respect to the corresponding atomistic systems or experimental values. Other dynamical properties such as shear viscosities or local relaxation times are also significantly influenced. The lack of accurate dynamics is scientifically unsatisfactory. It also precludes the application of multiscale concepts to the predictive calculation of technologically important dynamical and transport properties. Examples include rheological properties that play a crucial role in extrusion and injection molding of polymer melts. This thesis addresses whether the artificially accelerated dynamics of coarse-grained models can be quantified in terms of the variation of excess entropy between bead-spring polymer models at different modeling resolutions. The acceleration of the dynamics of coarse-grained models, and thus the decrease of monomer friction, is cast in terms of excess entropy differences of different coarse-grained resolutions. Starting with an elementary bead-spring model of unentangled chains in a melt state, we derive two less-detailed models following a systematic approach using bottom-up coarse-graining techniques. The exact entropy is calculated for the three models using a rigorous approach based on thermodynamic integration and an approximation that only includes contributions of pair correlations to the excess entropy. The excess entropy differences across the resolutions are then correlated with the rate of dynamics acceleration. Among the findings of this project, it is remarkable that the excess entropy differences between the coarser and finer models correlate significantly with the logarithm of the ratio of dynamical properties quantifying the acceleration upon changing the model resolution. The results indicate that it is possible to understand the acceleration driven by coarse-graining in terms of excess entropy differences for unentangled polymers. These correlations may open new possibilities for simple empirical methods to perform *a posteriori* corrections of the coarse-grained dynamics.

Contents

Acknowledgements	vii
1. Introduction	1
1.1. Motivation	3
1.2. Objectives	5
1.3. Structure of the Thesis	5
2. Theoretical Framework	7
2.1. Molecular Dynamics Simulations	8
2.1.1. Integrating the Equations of Motion	8
2.1.2. Ensembles and Thermostats	11
2.1.3. Periodic Boundary Conditions	13
2.1.4. Interaction Potentials	15
2.1.5. Measurements from Simulations	18
2.2. Dynamics in Simulations	25
2.2.1. Self-Diffusion	25
2.2.2. Shear viscosity	26
2.2.3. Bond Reorientation Relaxation	29
2.3. Multiscale Modeling	29
2.3.1. Coarse-Grained Modeling	30
2.3.2. Deriving Coarse-Grained Potentials	33
2.4. Estimating Entropies in Simulations	37
2.4.1. Thermodynamic Integration	38
2.4.2. Two-body Approximation	42
2.5. Excess Entropy Scaling	44
2.5.1. Excess Entropy	44
2.5.2. Scaling Relations	45
2.5.3. Rosenfeld Scaling	46
2.5.4. Dzugutov Scaling	46
3. Models and Methodology	49
3.1. Models	50
3.1.1. Unit System	51
3.1.2. Thermodynamic State Points	52
3.2. Overview of the Workflow	53
3.2.1. Computational Details	55
3.3. System Preparation	55
3.4. Coarse-Graining Procedure	56
3.4.1. Bonded Potentials	57
3.4.2. Non-bonded Potentials	60

3.5. Dynamical Properties	60
3.5.1. Self-Diffusion Coefficient	63
3.5.2. Viscosity	65
3.5.3. Bond Reorientation Time	70
3.6. Excess Entropy Estimation	72
3.6.1. Thermodynamic Integration	72
3.6.2. Two-Body Approximation	83
3.7. Accelerated Dynamics and Excess Entropy Scaling	84
4. Coarse-Graining Effects on Dynamics and Entropy	87
4.1. Dynamics in Coarse-Grained Systems	88
4.2. Entropy in Coarse-Grained Systems	93
4.2.1. Two-body Approximation to the Excess Entropy	96
5. Dynamics and Excess Entropy Differences	101
5.1. Note on Self-Diffusion and Viscosity	102
5.2. Correlations Between Coarse-Grained and Fine-Grained Systems	103
5.3. Correlations Between Coarse-Grained Systems	106
5.4. Correlations Based on the Two-Body Entropy	107
5.5. Correlations Involving Renormalized Dynamical Properties	109
6. Conclusions and Outlook	113
References	117
A. Supplement for Theory and Methodology	131
A.1. Example of Mapping in Coarse-Graining	131
A.2. Absolute Entropy of the FJC	133
B. Building Lennard-Jones Chains	135
C. Correlation Involving the Self-Diffusion Coefficients	141
D. List of Publications	143

List of Abbreviations

ACF	Auto-correlation function
BI	Boltzmann inversion
CG	Coarse-grained
DBI	Direct Boltzmann inversion
DOF	Degree of freedom
EMC	Enhanced Monte Carlo
FG	Fine-grained
GK	Green-Kubo
FJC	Freely-jointed chain
IBI	Iterative Boltzmann inversion
IG	Ideal gas
LSC	Linear-scaling soft core
LJ	Lennard-Jones
LJC	Lennard-Jones chain
MD	Molecular dynamics
mKWW	Modified Kohlrausch-Williams-Watts
MSD	Mean-square displacement
PBC	Periodic boundary conditions
PMF	Potential of mean force
RDF	Radial distribution function
RNEMD	Reverse non-equilibrium molecular dynamics
TI	Thermodynamic integration

1. Introduction

Coarse-graining methods have played an increasingly relevant role in computational studies of macromolecular systems for the past two decades [1, 2]. One of the reasons is that fully atomistic simulations where all atoms are taken into account have inherent practical limits in terms of system size and simulated time. On a more fundamental level, another reason is that many problems are naturally multiscale in which different phenomena are happening at several time and length scales [3]. Chain molecules such as polymers and proteins are quintessential examples of such multiscale systems. By coarse-graining out the details (or degrees of freedom) that are not important in a model it becomes possible to target scale-specific behavior and to simulate larger systems for longer times. The basic idea in systematic coarse-graining is to group together many atoms and assign a single interaction site to this group. The interactions between such sites are derived from data supplied by simulations performed in a fine-grained (FG) reference model, which describes the system with higher resolution of details. The removal of certain degrees of freedom (DOF), however, effectively smoothes the potential energy surface, allowing longer time steps, and removes molecular friction from the system, leading to significantly accelerated dynamics in the coarse-grained (CG) system in comparison with the reference model [4]. Calculation of static or structural properties of systems with inherent slow long-time dynamics, mainly polymer melts, greatly benefit from this acceleration brought about by coarse-graining. However, the spurious acceleration inevitably affects dynamical properties extracted from the CG simulations thus making the comparison of such quantities to experimental data a challenge [5].

CG describes a molecular system in terms of interaction sites, also referred to as beads or superatoms, representing a group of atoms of a higher-resolution model. Physical properties are assigned to each site, such as mass and charge. These sites interact with each other through carefully crafted potentials designed to reproduce structural or thermodynamic properties of interest of the higher-resolution model.

One consequence of removing DOF is that it is possible to simulate more interaction sites than it would in an all-atom model considering the same computational expense. Moreover, removing DOF from the model smoothes out any associated potential energy surfaces, significantly decreasing friction and creating an artificial acceleration of the system dynamics. On the one hand, this is highly attractive, as events and phenomena that would take prohibitive amounts of time to simulate using a model with slow dynamics now become attainable due to the faster dynamics of the corresponding CG model. On the other hand, however, utilizing the coarse-graining model to predict dynamics becomes challenging, as in principle, it is not clear *a priori* by how much the model accelerated.

Many studies proposed different alternative solutions to correct the CG dynamics [6]. If the level of coarse-graining is not too severe [7] such that the accelerated dynamics remain realistic, and if the resulting potentials are not too soft so that atoms or bonds do not cross each other, then the most straightforward way to recover the natural dynamics is through dynamic mapping by performing rescaling of time with a single scaling factor, based on the assumption that the same factor accelerates all dynamical processes. Even though this assumption may not be entirely justified theoretically, for low to moderate degrees of coarse-graining it has been successfully applied [8, 9]. The scaling factors must retain information about the missing friction, either explicitly through the direct estimation of the friction coefficients or implicitly by considering properties related to the monomeric friction, such as the mean-squared displacement (MSD) of the monomers obtained from atomistic simulations. Harmandaris and Kremer [10] expanded this rescaling approach by applying a hierarchical method to derive time scaling parameters based on the effective bead frictions obtained from simulations at different degrees of resolution. One of the problems with this dynamic mapping approach is the necessity of computationally costly atomistic simulations in tandem with the CG simulations, something that is not always desirable nor feasible. Furthermore, the scaling factors are dependent on the thermodynamic conditions for which they were derived [11]. Another approach of completely different nature was taken by Meinel and Müller-Plathe [12], who proposed a method to predict the rate of acceleration of a CG model based on the loss of surface roughness caused by coarse-graining. The idea of these authors is to use the difference between the molecular surface of the CG and FG systems and correlate this to the acceleration of the dynamics as measured by the ratio of self-diffusion coefficients at the different resolutions. The method has been successfully applied to low-molecular-weight hydrocarbon liquids at ambient conditions.

Guenza and co-workers [13–15] took a different path and, starting from the Ornstein-Zernike equation, derived a first-principle analytical approach to rescale the dynamics by taking into account the averaged internal DOF and the enhanced dissipation due to coarse-graining. In their approach, which is considered transferable and general, the equations of motion are solved at different levels of representation. There is no need to perform atomistic simulations to parameterize the potentials or build calibration curves such as those used in dynamic mapping. The method, however, becomes quite involved when CG schemes more sophisticated than simple isotropic models are required. Markutsya and Lamm [16] proposed another approach in the form of a method based on Langevin dynamics with hydrodynamic interactions to derive CG potentials that retain the frictional contribution. When more drastic coarse-graining is considered, e.g., by grouping tens of monomers of a polymer chain in a single interaction bead, or even the whole chain in a soft blob, additional care must be taken so that the dynamics is qualitatively realistic, in particular for entangled polymers [17]. In this scenario, the interactions are dominated by friction and stochastic forces, which must be properly accounted for in the equations of motion. In a recent contribution, Xia et al. [18] employed the energy renormalization coarse-graining method [19] to obtain temperature-transferable coarse-graining potentials that preserve the atomistic dynamics in CG models of glass-forming polymers over a wide range of temperatures. The methodology is based on renormalizing the cohesive portion of the interacting potentials so that errors in dynamic properties related to the configurational entropy are minimized, in particular by exploiting the relationship between short- and long-time dynamical properties. In the field of dynamically consistent systematic coarse-graining there has been a continuous development of methods based on the Mori-Zwanzig theory over the last few years. These methods aim at preserving, rather than recovering, the correct CG dynamics following a more fundamental statistical mechanical route. In this formalism an exact equation of motion of the CG model is expressed as a generalized Langevin equation and the missing friction due to the removal of DOF is explicitly accounted for, which is then solved with different degrees of approximations. Reviews of the state-of-the-art can be found in Refs. [20, 21].

1.1. Motivation

One of the effects of coarse-graining on the dynamics, and directly related to averaging out DOF and grouping several interactions sites into a single bead, is the reduction of the dimensionality of the configuration space. The number of available microstates decreases with increasing level of coarse-graining [22]. The

elimination of a significant number of microstates contributes to a loss of entropy of the system [23, 24], which in turn becomes dependent on the level of coarse-graining. The fact that dynamics and entropy are both governed by model resolution and thereby linked has motivated, in part, this project. Another motivation was the connection of excess entropy and dynamics by the so-called excess entropy scaling relations, first discovered by Rosenfeld [25] over four decades ago in the context of simple liquids. Even if these relations are empirical and only approximate [26], as verified not only in simulations [27] but also in experiments [28–31], they hold for a large number of different fluids.

Excess entropy scaling can be linked to coarse-graining by yet another concept, the relative entropy [32, 33]. In the context of coarse-graining, this quantity measures the loss of information brought about by replacing a FG probability distribution with a CG one derived from it, and thus it is a useful metric of the quality of a given CG model. Moreover, minimizing the relative entropy with respect to model parameters leads to an optimal CG model, providing a systematic and precise manner to derive coarse-graining potentials that minimize errors in macroscopic quantities and retain enough detail to represent the desired property. In model combinations for which the relative entropy is at a minimum, this quantity is equal to the difference between the excess entropy of the two systems, i.e., CG and reference one. Within this approach it was possible to establish a theoretical framework that connects relative entropy to the CG dynamics through excess entropy scaling relations [27, 34].

Previous studies have considered the connection between entropic quantities and the acceleration of CG models for liquids of small molecules. The definition of the acceleration of a CG model with respect to its reference FG model is usually based on ratios of dynamical properties obtained from the two models. Such ratios have also been interpreted as a measure of the coarse-graining errors in the dynamical properties. Armstrong et al. [35] showed that the logarithm of the ratio of properly renormalized self-diffusion coefficients of atomistic and CG liquid water correlates well with the difference in the two-body entropy between the two resolutions. Shell [34] considered the ratio of the self-diffusion coefficients for a liquid of LJ particles as the reference model and a corresponding CG soft-sphere representation with the same number of DOF. The ratio was then connected to the relative entropy of the optimized CG soft-sphere model, arguing that the correlation is possible due to the ties of both dynamics and relative entropy with excess entropies. In their studies, Armstrong et al. and Shell argued that the renormalization rules (e.g., Rosenfeld's) play a key role in connecting the acceleration (or error) of CG models with excess entropy

differences. However, it is not clear to what extent the renormalization rules are required for such purpose, or even if those connections are valid for more complex scenarios of coarse-graining and systems other than liquids of small molecules. In particular, to our best knowledge, no studies address the connection between excess entropy differences and the dynamics in polymer models at different resolutions.

1.2. Objectives

In this thesis we analyze the relation between, on the one hand, the ratios of different dynamical properties characterizing both global and local dynamics of CG polymer models in melt conditions at three levels of coarse-graining and, on the other hand, the respective excess entropy differences between the levels. We start from a generic bead-spring polymer model and systematically derive CG models of two levels of resolution from scratch, obtaining a set of bonded and non-bonded CG potentials for each thermodynamic state point considered. This strategy allows for a consistent set of models and makes comparisons more reliable. Furthermore, we took special care to obtain excess entropies consistently for all resolutions without employing too many approximations - instead, we obtained them from absolute entropies calculated with methods that are, in principle, exact. We also consider how our findings are affected if the excess entropy is approximated by the two-body term of the many-body entropy expansion. Finally, the effect of dynamics reduction rules in excess entropy scaling was also considered.

1.3. Structure of the Thesis

This thesis is organized as follows. Chapter 2 presents a review of the theory relevant to simulation methods, analysis tools and interpretation of results. In Chapter 3 the actual methodology utilized in this work is described, with special attention on how the entropic quantities were obtained. This chapter is also where the first results are presented and discussed. Results are further presented in Chapter 4, where the effects of coarse-graining in entropy and dynamics are discussed. The main results of the thesis are in Chapter 5, which are the correlations between excess entropy differences and the acceleration of the CG dynamics. A brief summary and outlook are presented in Chapter 6. In Appendix D the reader will find a list of articles published during the PhD project, including the main publication related to this thesis [36].

2. Theoretical Framework

In this section a review of the theoretical concepts relevant for this work is presented. It starts with a short introduction to molecular dynamics, followed by an overview of multiscale modeling and then by a review of dynamical processes of interest and a few methods to assess them in simulations. A brief review of multiscale modeling and coarse-graining techniques is also presented. Finally, we discuss how to estimate entropies from simulations and introduce the topic of excess entropy scaling. Computational details, simulation protocols, and applications of these theoretical concepts are left for Chapter 3.

2.1. Molecular Dynamics Simulations

Molecular dynamics (MD) simulations have been used to study polymeric systems for nearly half a century [37–41]. It was developed as a technique based on the classical mechanical and thermodynamic formalism and even though versions based on quantum mechanical methods do exist [42–47], for a large class of problems, in particular in polymer physics and biophysics it is sufficient (and often mandatory) to stay within the classical realm. In a nutshell, MD consists in a step-wise integration of Newton’s second law so that in each step the positions and momenta of the atoms ¹ are updated taking into account the classical forces acting upon them, e.g., due to short- and long-range effects from other atoms, or due to an externally applied field, etc. Therefore the system dynamics evolves in time according to Newtonian dynamics and this time evolution is recorded as a trajectory, which is one of the outputs of the simulation, as well as thermodynamical information of the system. Statistical mechanical tools are then unleashed on the collected data and micro- and macroscopic quantities of interest are estimated, such as diffusivity rates, viscosities, structure factors, heat transfer descriptors, and so on. Each part of this cursory overview has, of course, several layers of theoretical and technical details [48–51]. These are briefly presented over the remaining of this section.

2.1.1. Integrating the Equations of Motion

In order to know the positions and momenta of all atoms in a classical system with N atoms at a given time t we must integrate $3N$ second-order coupled ordinary differential equations resulting from Newton’s second law, i.e.,

$$\mathbf{F}_i = -\frac{\partial U}{\partial \mathbf{r}_i} = m_i \frac{\partial^2 \mathbf{r}_i}{\partial t^2}, \quad (2.1)$$

where m_i and \mathbf{r}_i are the mass and Cartesian coordinates of atom i and \mathbf{F}_i is the total force acting on it, $U = U(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$ the potential energy of the system, and $i = 1, 2, \dots, N$. These coupled equations describe the dynamics of this many-body system and have no analytical solution for $n > 3$. However, they can be simultaneously numerically integrated using various algorithms based on Taylor expansions and finite-differences. These integration techniques constitute one of the building blocks of MD simulations.

One of the most basic techniques for integrating Newton’s equations of motion is Verlet’s algorithm [52]

¹The term *atoms* is used here but the objects in MD simulations need not to correspond to real atoms, thus *particles* may also be used.

in which two Taylor expansions on t are carried out for the atomic positions up to third order, one for $t - \Delta t$ and another for $t + \Delta t$, where Δt is the so called integration time step length. These expansions are

$$\mathbf{r}_i(t - \Delta t) = \mathbf{r}_i(t) - \frac{\partial \mathbf{r}_i}{\partial t}(t) \Delta t + \frac{1}{2} \frac{\partial^2 \mathbf{r}_i}{\partial t^2}(t) \Delta t^2 - \frac{1}{6} \frac{\partial^3 \mathbf{r}_i}{\partial t^3}(t) \Delta t^3 + \mathcal{O}(\Delta t^4), \quad (2.2)$$

and

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \frac{\partial \mathbf{r}_i}{\partial t}(t) \Delta t + \frac{1}{2} \frac{\partial^2 \mathbf{r}_i}{\partial t^2}(t) \Delta t^2 + \frac{1}{6} \frac{\partial^3 \mathbf{r}_i}{\partial t^3}(t) \Delta t^3 + \mathcal{O}(\Delta t^4), \quad (2.3)$$

and they can be added and rearranged to yield

$$\mathbf{r}_i(t + \Delta t) = 2\mathbf{r}_i(t) - \mathbf{r}_i(t - \Delta t) - \frac{1}{m_i} \frac{\partial U}{\partial \mathbf{r}_i} \Delta t^2, \quad (2.4)$$

where Newton's second law from Equation (2.1) has been used. The standard Verlet's algorithm as given by Equation (2.4) allows one to know the position of an atom at the instant $t + \Delta t$ if its position at $t - \Delta t$ is known, in addition to its current position and potential energy, with an error of the order of Δt^4 . One clear advantage of this algorithm is that it is fully time-reversible, a second advantage is that it preserves volumes in the phase space if all forces are conservative. Moreover, the computational implementation of Verlet's algorithm is simple and constitutes an "embarrassingly parallel" problem that can take advantages of massively parallel hardware. In general, MD is very well suited for parallel computing.

One problem with Equation (2.4) is that the atomic velocities are not directly obtained from it. Even though they are not strictly necessary to compute the temporal evolution of the system, velocities are required elsewhere in the MD framework, e.g., to compute kinetic energies, temperature, etc. This can be remedied to a certain degree by subtracting Equation (2.2) from Equation (2.3), yielding the velocities computed as a finite-difference of the positions in the previous and next steps,

$$\mathbf{v}_i(t) = \frac{\mathbf{r}_i(t + \Delta t) - \mathbf{r}_i(t - \Delta t)}{2\Delta t}, \quad (2.5)$$

with an error of $\mathcal{O}(\Delta t^2)$. One should notice in Section 2.1.1 that the velocity is known for t when the computation is already at $t + \Delta t$, thus in a step behind.

An algorithm that improves the availability of the velocities is the leapfrog scheme, in which the velocities

and positions are obtained at interleaved time steps and expressed as

$$\mathbf{v}_i\left(t + \frac{\Delta t}{2}\right) = \mathbf{v}_i\left(t - \frac{\Delta t}{2}\right) + \mathbf{a}_i(t)\Delta t, \quad (2.6)$$

and

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \mathbf{v}_i\left(t + \frac{\Delta t}{2}\right) \Delta t, \quad (2.7)$$

respectively. The fact that the velocities are needed to calculate the new positions in Equation (2.6), differently from the Verlet algorithm in Equation (2.4), makes the leapfrog more convenient to couple with an external thermal bath. However, since positions and velocities are not available in the same time step, calculating kinetic properties at the same time step as the positions, e.g., the kinetic energy, requires averaging the velocities from steps $t - \frac{\Delta t}{2}$ and $t + \frac{\Delta t}{2}$.

Another alternative algorithm to standard Verlet and leapfrog is the related velocity-Verlet algorithm [53] which makes position and velocity available at the same time step. In the velocity-Verlet algorithm the positions are updated as

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \mathbf{v}(t)\Delta t + \frac{1}{2} \frac{\mathbf{F}_i(t)}{m_i} \Delta t^2, \quad (2.8)$$

and once the updated positions are obtained, the forces are updated with respect to them as

$$\mathbf{F}_i(t + \Delta t) = \mathbf{F}_i(\mathbf{r}_i(t + \Delta t)), \quad (2.9)$$

the velocities are then updated by averaging the updated and current acceleration (i.e, forces),

$$\mathbf{v}_i(t + \Delta t) = \mathbf{v}_i(t) + \frac{1}{2} \left(\frac{\mathbf{F}_i(t)}{m_i} + \frac{\mathbf{F}_i(t + \Delta t)}{m_i} \right) \Delta t, \quad (2.10)$$

where the error is of the same order of the standard Verlet. One may notice that this averaging of the forces is equivalent to breaking up the velocity update in two steps, first for half a time step using the current acceleration,

$$\mathbf{v}_i\left(t + \frac{\Delta t}{2}\right) = \mathbf{v}_i(t) + \frac{\Delta t}{2} \frac{\mathbf{F}_i(t)}{m_i} \quad (2.11)$$

then for the full time step using the updated acceleration from Equation (2.9),

$$\mathbf{v}_i(t) = \mathbf{v}_i\left(t + \frac{\Delta t}{2}\right) + \frac{\Delta t}{2} \frac{\mathbf{F}_i(t + \Delta t)}{m_i}, \quad (2.12)$$

with Equation (2.12) being equivalent to Equation (2.10).

Determining the initial positions and velocities for an MD simulation can be done in several ways. The initial configurations may be created based on atomic models built from experimental data, as it is generally done for proteins, e.g., using X-ray crystallography, nuclear magnetic resonance spectroscopy, electron microscopy, etc. The initial coordinates may also be generated using simple lattice models with predefined molecules placed at the lattice sites. Complex macromolecules may be generated via clever algorithms, Monte Carlo techniques, or even by simulating real processes, e.g., polymerization. The initial velocities must be generated in a way that they are compatible with the desired temperature and such that the net linear momentum of the system is zero. Often it is enough to perform a random sampling from a Maxwell-Boltzmann distribution to initialize the velocities around a target temperature, and then perform an equilibration simulation for long enough using a time step length of appropriate size and, when used, sensible thermostat or barostat configurations such that the energy is properly distributed to the internal DOF in the system.

2.1.2. Ensembles and Thermostats

A thermodynamic ensemble is a collection of an infinitely large number of replicas of a system that are identical in the thermodynamic sense on the macroscopic level. On the microscopic level, however, the positions and momenta of the particles in each replica is distinct. Each replica corresponds to a point in the phase space of that system and each point is identified as a microscopic state of the system that is compatible with the imposed macroscopic conditions. By fixing certain macroscopic parameters, such as total number of particles, volume, total energy, or temperature, it is possible to employ convenient statistical mechanical tools to calculate different thermodynamic functions of the system. For example, in the canonical ensemble the system is in thermal equilibrium with a heat bath at a controlled (fixed) temperature and the number of particles, the volume of the system, i.e., parameters N , V , and T are fixed. In this ensemble energy is exchanged with the heat bath so that different microscopic states of the system will have different energy profiles. In MD the heat bath is implemented as a thermostat, for which several

algorithms exist.

One such thermostat algorithm, and the simplest one available, is the velocity scaling method. The temperature along of the MD run is kept fixed by multiplying the particle velocities at the end of each time step by a factor $\lambda = \sqrt{T_0/T(\tau)}$, where T_0 is the target temperature and $T(t)$ is temperature at time step t . The system temperature should then converge to T_0 , however at an uncontrolled rate and without any regards to violation of energy equipartitioning which may lead to spurious artifacts [54]. A similar but smoother approach in which the rate of convergence may be controlled is implemented by the Berendsen thermostat [55],

$$\lambda = \left[1 + \frac{\Delta t}{\tau_T} \left(\frac{T_0}{T(t)} - 1 \right) \right]^{1/2}, \quad (2.13)$$

where τ_T is a time dampening factor and Δt the time step length, thus setting $\tau_T = 1$ recovers the simple velocity rescaling algorithm. The Berendsen thermostat is still subject to violating the equipartition theorem and does not correctly capturing energy fluctuations, which are problems also seen in the standard velocity rescaling. While the Berendsen thermostat is efficient in making a system reach a target temperature, it does not always yield measures compatible with the canonical ensemble, in particular regarding energy fluctuations [56].

An alternative method proposed Nose [57, 58] and Hoover [59] is based on the idea of introducing a fictitious dynamical variable that adds or removes friction from the system in order to slow down or accelerate the particles so that the target temperature is maintained. In the Nose-Hoover framework, Newton's second law for a particle i is written as

$$m_i \frac{d^2 \mathbf{r}_i}{dt^2} = \mathbf{F}_i - \zeta m_i \mathbf{v}_i, \quad (2.14)$$

where $\zeta m_i \mathbf{v}_i$ is a dissipative force and ζ is a time-dependent friction coefficient whose derivative with respect to time is

$$\frac{d\zeta}{dt}(t) = \frac{1}{Q} \left[\sum_{i=1}^N \frac{1}{2} m_i \mathbf{v}_i^2 - \left(\frac{3N+1}{2} \right) k_B T_0 \right], \quad (2.15)$$

where Q is the mass-equivalent of the fictitious particle and controls the rate at which friction is added or removed from the system, effectively acting as a temperature damping parameter; T_0 the target temperature. At steady state we have $d\zeta/dt = 0$ and kinetic energy equal to $\frac{3}{2}(N+1)k_B T$, which is compatible with the equipartition theorem and in which the additional DOF in $3N+1$ corresponds to the fictitious particle.

The velocity Verlet algorithm can be adapted to handle the fictitious particle introduced by the Nose-Hoover thermostat as follows. The update of the positions must account for the dissipative force added in Equation (2.14), so Equation (2.8) is rewritten as

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \mathbf{v}_i(t)\Delta t + \frac{1}{2} \left(\frac{\mathbf{F}_i(t)}{m_i} - \zeta(t)\mathbf{v}_i(t) \right) \Delta t^2, \quad (2.16)$$

and, accordingly, the half time step update of the velocity from Equation (2.11) becomes

$$\mathbf{v}_i \left(t + \frac{\Delta t}{2} \right) = \mathbf{v}_i(t) + \frac{\Delta t}{2} \left(\frac{\mathbf{F}_i(t)}{m_i} - \zeta \mathbf{v}_i(t) \right). \quad (2.17)$$

A half timestep update is performed in the friction coefficient using the current velocities,

$$\zeta \left(t + \frac{\Delta t}{2} \right) = \zeta(t) + \frac{\Delta t}{2Q} \left[\sum_{i=1}^N \frac{1}{2} m_i \mathbf{v}_i(t)^2 - \left(\frac{3N+1}{2} \right) k_B T_0 \right], \quad (2.18)$$

followed by a full time step update using the half time step velocities,

$$\zeta(t + \Delta t) = \zeta \left(t + \frac{\Delta t}{2} \right) + \frac{\Delta t}{2Q} \left[\sum_{i=1}^N \frac{1}{2} m_i \mathbf{v}_i(t + \Delta t/2)^2 - \left(\frac{3N+1}{2} \right) k_B T_0 \right], \quad (2.19)$$

finally, the new velocities are updated as

$$\mathbf{v}_i(t + \Delta t) = \frac{\left[\mathbf{v}_i(t + \Delta t/2) + \frac{\Delta t}{2} \frac{\mathbf{F}_i(t + \Delta t)}{m_i} \right]}{1 + \frac{\Delta t}{2} \zeta(t + \Delta t)}. \quad (2.20)$$

2.1.3. Periodic Boundary Conditions

The size of systems simulated with MD may be as large as $\mathcal{O}(10^9)$ particles, but often a few orders of magnitude less than that. Even at a billion particles this is evidently a very small number if compared to macroscopic systems. In simulations of such small systems it is almost always the case that finite-size and surface effects will have an influence if no further actions are taken. For example, simulating an isolated box of particles will hardly yield any useful information about the bulk behavior since a significant amount of the particles will be located close to the box walls. This problem is solved in MD by using periodic boundary conditions (PBC). Applying PBC in MD means replicating the simulation box in all spatial

directions such that it is adjacent on all sides to identical copies of itself, as schematically represented in Figure 2.1. The replicated boxes are images of the original box and the particles therein are images of the original particles. In practice only one box is simulated and this concept is implemented by making each particle leaving the box on one side have its image entering the box on the other side with the same velocity, such that each particle is simulated only once but it appears as if all images are also being simulated. The number of atoms in the box and the total momentum are conserved.

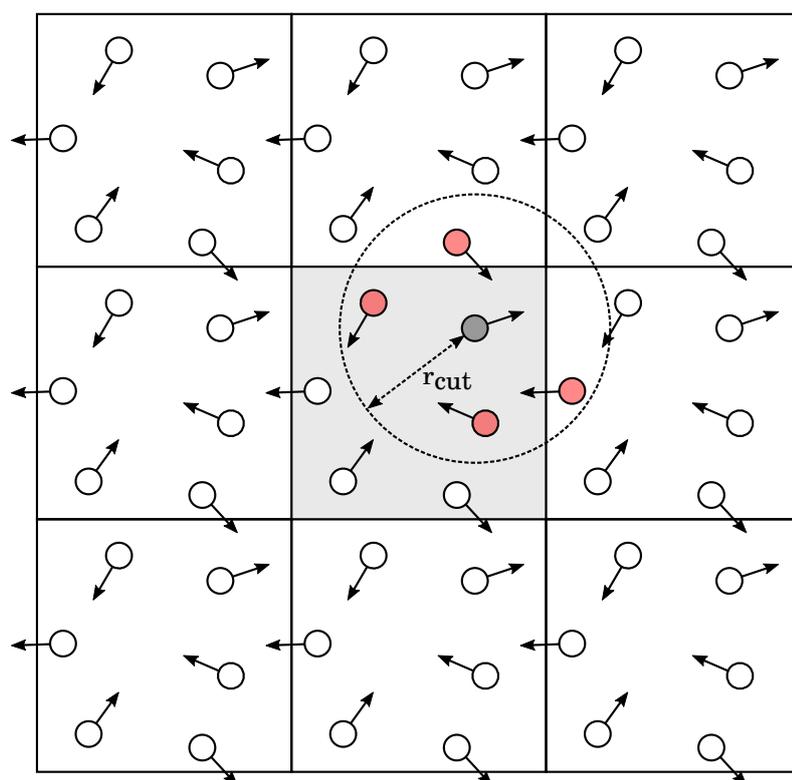


Figure 2.1.: Two-dimensional view of PBC. The central simulation box is highlighted. The gray particle interacts only with the red particles as given by the minimal image convention and within the cutoff distance defined by r_{cut} .

The caveat is that in the PBC framework each particle must interact only with the nearest image of another particle, as per the minimum image convention, otherwise there would be the risk of multiple interactions with the same particle. Moreover, although strictly not necessary for PBC, a radial cutoff distance, r_{cut} , is considered and only images within r_{cut} from any given particle are taken into account when calculating non-bonded interactions of that particle. Without this cutoff distance the particle would

interact with all other particles (or images of them), but the cutoff is convenient in order to reduce the computational costs. This cutoff distance has an effect on the MD machinery itself, i.e., when building neighbor lists only particles within the cutoff need to be accounted for, but, importantly, also in the physics of the system as interaction potentials must be truncated to zero at this distance. For dispersion interactions such as van der Waals this is not very problematic, but truncating many-body long-range potentials such as electrostatic interactions will lead to incorrect results. In this case more involved methods must be considered in which accurate approximations are computed. Widely used methods are those based on Ewald summations and particle meshes [60].

2.1.4. Interaction Potentials

The potential energy term in Equation (2.1), and from which the forces on each particle are derived, characterizes the interactions that are taking place in the system. These interactions are modeled classically and can be bonded or non-bonded. Bonded interactions are always intramolecular and include 2-body bond stretching, 3-body angle bending, improper angles and dihedral angle torsions, both of which are 4-body type of interactions. Non-bonded interactions are many-body and may be intra- or intermolecular, i.e., a particle may interact with other particles in the same molecule, and may be short- or long-ranged. In general, this classical potential energy, often referred as *force field* in the field of MD, is written as

$$U_{\text{total}} = U_{\text{bond}} + U_{\text{angle}} + U_{\text{improper}} + U_{\text{torsion}} + U_{\text{non-bonded}}, \quad (2.21)$$

where each term is explained below. The potential energy format show above is a general one and may be adapted to the problem at hand. Not all of the contributions may be necessary in all models.

Bonded Potentials

Commonly used forms for the bonded terms contributing to the total potential energy in Equation (2.21) are shown below. These may be further refined or replaced entirely depending on how suitable they are for the chosen model. The coefficients that appear in these potentials are external parameters that must be empirically adjusted. When the particles are supposed to model real atoms, these coefficients are fitted to experimental data or results from accurate quantum chemical calculations. If the particles are beads of a

CG model, for example, then in bottom-up multiscale approaches the coefficients are typically obtained from reference atomistic simulations.

Bond Stretching

The bond stretching interaction between two bonded particles may be described by a harmonic potential of the form,

$$U_{\text{bond}} = \sum_{i,j} \frac{1}{2} K_{r,ij} (r_{ij} - r_0)^2, \quad (2.22)$$

where the summation runs over all bonds defined by directly connected particles with indexes i and j , i.e., $i-j$, and $r_{ij} = |\mathbf{r}_j - \mathbf{r}_i|$ is the distance between particles i and j , with r_0 being the bond equilibrium distance such that $U_{\text{bond}}(r_0) = 0$. The coefficient $K_{r,ij}$ is the force constant for this harmonic potential determining the stiffness of the stretch and may depend on the types of the particles involved in the bond.

Angle Bending

The angle bending interaction between three bonded particles may also be described by a harmonic potential,

$$U_{\text{angle}} = \sum_{i,j,k} \frac{1}{2} K_{\theta,ijk} (\theta_{ijk} - \theta_0)^2, \quad (2.23)$$

where the summation runs over all sets of angles defined by bonded particles with indexes i , j , and k , i.e., $i-j-k$, and θ_{ijk} is the angle $\angle ijk$, with θ_0 being the equilibrium angle. The coefficient $K_{\theta,ijk}$ is the force constant of the harmonic potential.

Torsions

Considering a chain of four bonded particles, i.e., $i-j-k-l$, a dihedral angle is defined as the angle between the planes ijk and ijl . The potential torsional energy associated with the rotation of this angle is usually described by periodic functions, such as

$$U_{\text{torsion}} = \sum_{i,j,k,l} K_{\phi,ijkl} [1 + \cos(n\phi_{ijkl} - \delta)], \quad (2.24)$$

where the summation runs over all 4-body bonded interactions, ϕ_{ijkl} is the angle between the two planes, n is the angle periodicity and indicates the number of potential minima as the bond is rotated over a span

of 360° , δ is the phase shift, which acts as an equilibrium angle for $n = 0$. As with the other potentials seen above, $K_{\phi,ijkl}$ is the force constant that may depend on the types of the four particles involved in the interaction.

Improper Angles

The improper angle defined in a similar way as the torsion dihedral angle described above in terms of two planes defined by four particles. The main purpose of this potential is to restrict molecular deformation, for example, by keeping aromatic rings planar, or preventing that a tetrahedral molecule flips over to its mirror image. In many cases this structural constraint is achieved by prescribing a single minimum to the potential and one way to achieve this is by using a harmonic potential such as

$$U_{\text{improper}} = \sum_{i,j,k,l} K_{\xi,ijkl} (\xi_{ijkl} - \xi_0)^2 \quad (2.25)$$

where the summation is again over all four particles relevant for the geometry, ξ_{ijkl} is the angle between the planes ijk and ijl (but others are also possible), ξ_0 is the equilibrium angle, and $K_{\xi,ijkl}$ the force constant.

Non-bonded Potentials

Non-bonded interactions are often separated in short- and long-range contributions, usually given by the Lennard-Jones (LJ) and the Coulomb potential, respectively,

$$U_{\text{non-bonded}} = U_{\text{LJ}} + U_{\text{Coul}}. \quad (2.26)$$

It is possible to also include other non-bonded potentials that can more realistically model the target system, such as embedded-atom potentials involving many-body interactions for metallic systems [61, 62], or Tersoff potentials to simulate organic nanofillers and carbon-carbon interactions in graphene [63, 64].

Short Range

The short range contribution to the non-bonded interaction includes repulsive and attractive interactions between two particles, and the energy is dependent on the distance separating the particles, decaying to zero as the distance is increased, and having a very repulsive character at close separation and weaker

attraction at larger distances, typically with a potential well that defines a minimum at an equilibrium distance. This may model van der Waals interactions, for example, and a common approach is to use the 12-6 LJ potential

$$U_{\text{LJ}} = \sum_{i < j} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right], \quad (2.27)$$

where the summation runs over non-bonded particles i and j , $r_{ij} = |\mathbf{r}_j - \mathbf{r}_i|$ is the distance between the pairs of particles, ϵ_{ij} the depth the potential well and corresponds to the value of the potential energy where the minimum is located, and σ_{ij} is the distance at which the potential is zero and which can be interpreted as a collision distance or particle radius. In practice the LJ potential is shifted so that $U_{\text{LJ}}(r_{\text{cut}}) = 0$, i.e., it decays to zero at the cutoff distance. Moreover, the summation has to be carried out considering the minimum image convention within the PBC framework. It is also common to exclude first, second, and third bonded neighbors from the summation when calculating the LJ non-bonded contribution given that the interaction with these bonded particles may already be accounted with the various bonded potentials.

Long Range

The electrostatic potential describes the long-range interaction between charged particles and is given by Coulomb's law,

$$U_{\text{Coul}} = \sum_{i < j} \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{\epsilon_{ij} r_{ij}} \quad (2.28)$$

where the summation runs over all non-bonded particles, q_i and q_j are the charges of particles i and j , respectively; r_{ij} is the distance between the two particles, ϵ_0 the permittivity of the vacuum, ϵ_{ij} the dielectric constant. Prefactors may be required when computing electrostatic interactions between non-bonded particles that are taking part in bonded interactions. In practice the form given by Equation (2.28) is not used but rather an Ewald summation method [60].

2.1.5. Measurements from Simulations

In order to measure any observable quantity in a MD simulation it must be possible to express this quantity as a function of positions and velocities, as the record of the system evolution in the form of a trajectory is the basic outcome of a MD simulation. For example, temperature is a fundamental macroscopic observable that can be defined based on the kinetic energy following the equipartition theorem and measured at any

time step t during an equilibrium simulation as,

$$T(t) = \frac{2}{3Nk_B} \sum_{i=1}^N \frac{\mathbf{p}_i^2(t)}{2m_i}, \quad (2.29)$$

where N is the total number of particles, \mathbf{p}_i the linear momentum of particle i , m_i its mass, and k_B is Boltzmann's constant.

As introduced in Section 2.1.2, an ensemble is a set of replicas of a system each with its N particles having a distinct set of positions and momenta. It is useful to establish some notation and also few statistical mechanical statements of interest in the context of the canonical ensemble. Let $\mathbf{r}^N \equiv \{\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N\}$ and $\mathbf{p}^N \equiv \{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_N\}$ represent the positions and the conjugate momenta of all N particles of a system of fixed volume V and at temperature T which is maintained by exchanging energy with an external heat bath. The Hamiltonian of such a system may be written as

$$H(\mathbf{r}^N, \mathbf{p}^N) = U(\mathbf{r}^N) + K(\mathbf{p}^N), \quad (2.30)$$

where U and K are the potential and kinetic energy of the system, respectively. In the canonical ensemble, the probability distribution for a given microstate characterized by $\{\mathbf{r}^N, \mathbf{p}^N\}$ is

$$P(\mathbf{r}^N, \mathbf{p}^N) = \frac{1}{h^{3N} N!} \frac{\exp(-H(\mathbf{r}^N, \mathbf{p}^N)/k_B T)}{Z}, \quad (2.31)$$

where h is Planck's constant, k_B is Boltzmann's constant, and Z is the canonical partition function,

$$Z = \frac{1}{h^{3N} N!} \int \int \exp(-H(\mathbf{r}^N, \mathbf{p}^N)/k_B T) d\mathbf{r}^N d\mathbf{p}^N \quad (2.32)$$

where the double integral sign serves as a reminder that the integration has to be carried out over all coordinates and momenta, i.e., it is a $6N$ -dimensional integral. A given system property A that depends on the positions and momenta may be measured by calculating its average over the ensemble weighted by the probability distribution of the corresponding states,

$$\langle A \rangle_{\text{ensemble}} = \int \int A(\mathbf{r}^N, \mathbf{p}^N) P(\mathbf{r}^N, \mathbf{p}^N) d\mathbf{r}^N d\mathbf{p}^N. \quad (2.33)$$

In MD the system dynamics evolves in time and thus any average computed from a finite system trajectory is a time average and not strictly an ensemble average. However, if careful steps are taken in preparing and running the simulation, e.g., by using a proper thermostat that guarantees that snapshots of the simulation taken at regular intervals can be considered as samples taken from a canonical ensemble, or that the thermodynamic parameters of the simulation are compatible with the phase space region that one wants to explore; then the ergodic hypothesis is often assumed for sufficiently long trajectories, which means that sufficiently long MD runs will sample enough of the available phase space such that a time average is equal to an ensemble average. Therefore, instead of using Equation (2.33), averages of a property A in MD are calculated as

$$\langle A \rangle = \frac{1}{M} \sum_{\tau=1}^M A_{\tau}(\mathbf{r}^N, \mathbf{p}^N), \quad (2.34)$$

where M is the number of snapshots taken from the sufficiently long trajectory and A_{τ} is the instantaneous value of property A computed from the coordinates and momenta in snapshot τ .

Auto-Correlation Functions

Time correlation functions, and in particular auto-correlation functions (ACF), are quantities that play an important role in understanding time-dependent processes near and at equilibrium. The ACF of a time-dependent equilibrium property $A(t)$ that depends on quantities of a simulation at a time t is defined as

$$C_A(t) = \langle A(t)A(0) \rangle, \quad (2.35)$$

where the angle brackets indicate an ensemble average. In the context of equilibrium MD this notation is understood to have the following explicit meaning [49],

$$C_A(t) = \langle A(t)A(0) \rangle = \lim_{\Omega \rightarrow \infty} \frac{1}{\Omega} \int_0^{\Omega} A(t_0 + t)A(t_0) dt_0, \quad (2.36)$$

where several time origins Ω are considered and averaged over, not only $t = 0$. Since the system is at equilibrium it does not matter which point in a trajectory is taken as the origin, but only the time delay between the two measurements of A . What $C_A(t)$ is measuring is how the value of A at $t_0 + t$ is correlated with its value at a prior point in time, t_0 . It is often useful to normalize the time ACF as $C_A(t)/C_A(0)$, such that it never exceeds unity at $t = 0$.

Table 2.1.: Example of the computation of the time ACF $C_A(\tau)$ for a property A following the discrete approach with multiple origins of Equation (2.38) for a trajectory that is 5 time steps long.

τ	N_0	τ_0	Terms contributing to $C_A(\tau)$
1	5	0, 1, 2, 3, 4	$A(1)A(0), A(2)A(1), A(3)A(2), A(4)A(3), A(5)A(4)$
2	4	0, 1, 2, 3	$A(2)A(0), A(3)A(1), A(4)A(2), A(5)A(3)$
3	3	0, 1, 2	$A(3)A(0), A(4)A(1), A(5)A(2)$
4	2	0, 1	$A(4)A(0), A(5)A(1)$
5	1	0	$A(5)A(0)$

If the properties of interest are those of individual particles and if these particles are identical, then the averaging over all N particles can be included in the calculation of the ACF to improve statistics, thus Equation (2.36) becomes,

$$C_A(t) = \langle A_i(t)A_i(0) \rangle = \lim_{\Omega \rightarrow \infty} \frac{1}{\Omega} \frac{1}{N} \int_0^\Omega \sum_{i=1}^N A_i(t_0 + t)A_i(t_0) dt_0, \quad (2.37)$$

In practice the trajectory from a MD simulation is discrete with each time step² τ spaced from each other by Δt , thus the real time is $t = \tau \Delta t$. The integration in the time ACF of Equation (2.36) then becomes a summation [48],

$$C_A(\tau) = \langle A(\tau)A(0) \rangle = \frac{1}{N_0} \sum_{\tau_0=0}^{N_0} A(\tau_0 + \tau)A(\tau_0), \quad (2.38)$$

with the particle averaged equivalent being

$$C_A(\tau) = \langle A_i(\tau)A_i(0) \rangle = \frac{1}{N_0} \frac{1}{N} \sum_{\tau_0=0}^{N_0} \sum_{i=1}^N A_i(\tau_0 + \tau)A_i(\tau_0), \quad (2.39)$$

where N_0 is the number of origins available for τ and care must be taken when evaluating the summation so that $A(\tau_0 + \tau)$ is never evaluated past the end of the trajectory. One consequence of this approach is the clear disadvantage that different points of the ACF are averaged differently and those at the beginning of the trajectory have more observations contributing to the average than those at the end. This can be clearly seen in Table 2.1, where a 5-step trajectory is considered. For each moving time origin, τ_0 , there is “less trajectory” left than the previous thus there are less terms available to contribute to the multiple-origin average.

²Throughout this thesis we use t and τ to refer to time units and time steps, respectively, occasionally with subscripts for further specification.

An alternative approach is based on a sliding window [65]. The complete trajectory of length τ_L time steps is parsed through overlapping blocks, i.e., the sliding window, each of duration $\tau_w < \tau_L$. The first step of each window is taken as an origin and the time ACF calculation is only performed inside the window, and then averaged over all windows. Thus this approach is also a multi-origin one. An example is given in Figure 2.2 for the case of a trajectory of length $\tau_L = 20$ time steps and a window of length $\tau_w = 5$, in which for illustrative purposes a spacing of $\tau_s = 3$ time steps is considered between each origin.

This sliding window method is useful beyond the computation of ACFs, e.g., to compute MSDs, and in such cases it is also desirable to have spacing between the different origins in order to avoid correlations between the observations contributing to the averages. The number of origins for this sliding window

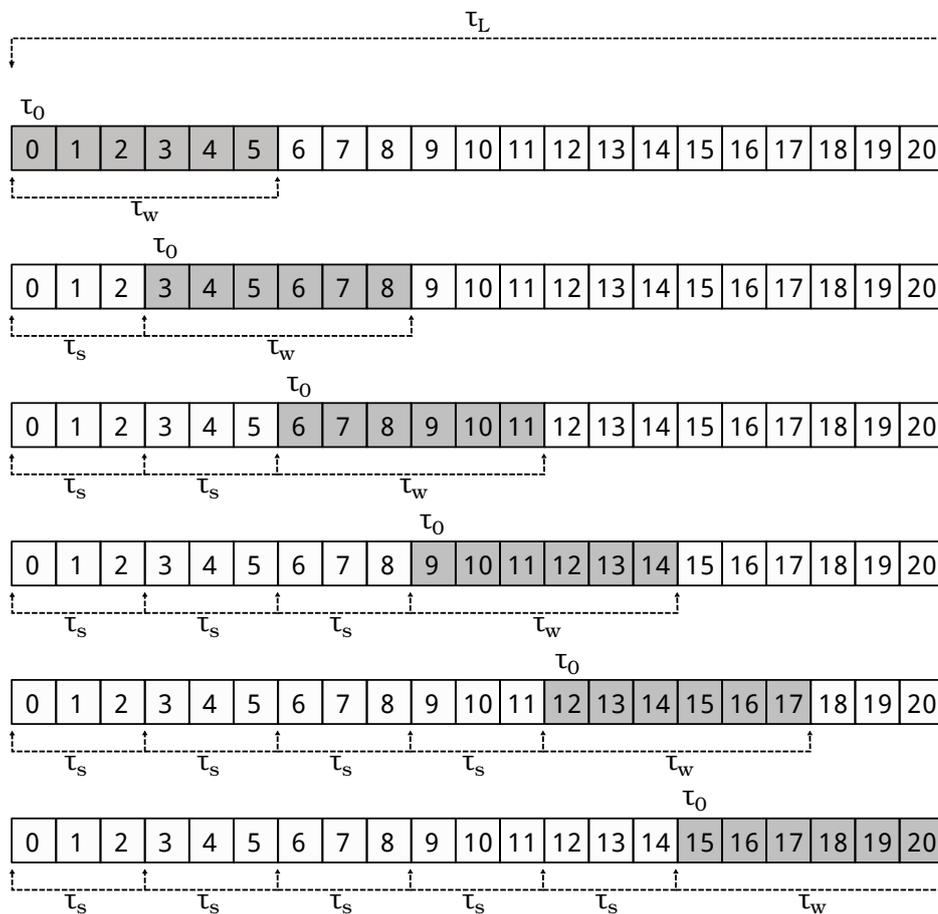


Figure 2.2.: Sliding window method to compute a time ACF for a trajectory of duration $\tau_L = 20$ time steps using a window of length $\tau_w = 5$ time steps with $N_0 = 6$ origins τ_0 spaced in $\tau_s = 3$ time steps. Active window is represented as shaded blocks, each row representing a different origin.

scheme calculated as,

$$N_0 = 1 + \left[\frac{\tau_L - \tau_w}{\tau_s} \right], \quad (2.40)$$

which yields $N_0 = 6$ for the example in Figure 2.2 and Table 2.2, where $\tau_L = 20$, $\tau_w = 5$, and $\tau_s = 3$ time steps. One drawback of this method is that a trajectory that is τ_L long will yield an ACF that only goes up to τ_w , however there is certainty that all points of the ACF have been equally averaged.

Table 2.2.: Example of the computation of the time ACF $C_A(\tau)$ for a property A using the multi-origin sliding-window approach with window duration of $\tau_w = 5$ time steps, origins spaced by $\tau_s = 3$ time steps, number of origins $n_0 = 6$, and a total trajectory duration $\tau_L = 20$ time steps.

τ	Terms contributing to $C_A(\tau)$
1	$A(1)A(0), A(4)A(3), A(7)A(6), A(10)A(9), A(13)A(12), A(16)A(15)$
2	$A(2)A(0), A(5)A(3), A(6)A(6), A(11)A(9), A(14)A(12), A(17)A(15)$
3	$A(3)A(0), A(6)A(3), A(8)A(6), A(12)A(9), A(15)A(12), A(18)A(15)$
4	$A(4)A(0), A(7)A(3), A(9)A(6), A(13)A(9), A(16)A(12), A(19)A(15)$
5	$A(5)A(0), A(8)A(3), A(10)A(6), A(14)A(9), A(17)A(12), A(20)A(15)$

Radial Distribution Function

A static property that is of great interest in this work is the pair correlation function which offers an insight into the local structure of a fluid by assessing how the atoms are organized around one another. A rigorous statistical mechanical definition of the pair correlation function in the canonical ensemble for a system of N atoms at temperature T , volume V and density $\rho = N/V$ is

$$g(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{Z} \frac{N(N-1)}{\rho^2} \int \dots \int \exp(-U(\mathbf{r}^N)/k_B T) d\mathbf{r}_3 d\mathbf{r}_4 \dots d\mathbf{r}_N \quad (2.41)$$

where Z is the canonical partition function and \mathbf{r}_1 and \mathbf{r}_2 are the coordinates of two arbitrary atoms. Thus $g(\mathbf{r}_1, \mathbf{r}_2)$ is obtained by integrating the configurational distribution over the position of all atoms except for two of them [48]. This definition is hardly useful in MD simulations, with the following being an equivalent albeit more pragmatic definition [49]

$$g(\mathbf{r}) = \frac{1}{\rho N} \left\langle \sum_{i=1}^N \sum_{j \neq i}^N \delta(\mathbf{r} - \mathbf{r}_{ij}) \right\rangle \quad (2.42)$$

where δ is Dirac's delta function, \mathbf{r} is an arbitrary point and $\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i$. It is often of interest to understand how the structural arrangements of atoms depend on their distance $r_{ij} = |\mathbf{r}_{ij}|$, ignoring orientational dependencies, in particular for homogeneous uniform systems. Thus dropping the vector notation, and also noticing that $r_{ij} = r_{ji}$ thus only half of the $N(N - 1)$ terms in Equation (2.42) are unique, it can be rewritten as

$$g(r) = \frac{2}{\rho N} \left\langle \sum_{i=1}^N \sum_{j<i}^N \delta(r - r_{ij}) \right\rangle \quad (2.43)$$

which is the radial version of the pair distribution function and thus referred to as the radial distribution function (RDF). After some manipulation it is possible to obtain a more practical expression that can be directly used in an MD simulation,

$$g(r) = \frac{\langle N(r, \Delta r) \rangle}{\frac{1}{2} N \rho V(r, \Delta r)}, \quad (2.44)$$

where $N(r, \Delta r)$ is the average number of particles in a spherical shell of radius r and thickness Δr and $V(r, \Delta r)$ is the volume of this shell. Thus the actual algorithm consists in visiting all particles and, from each, counting how many particles are located in concentric spherical shells of thickness Δr and radius varying from $0 \rightarrow r$ and centered on the reference particle that is currently being visited. A schematic representation of this calculation is shown in Figure 2.3.

The RDF at each r can be interpreted as the ratio of the local density $\rho(r)$ to the system density ρ , and $g(r) \rightarrow 1$ as $r \rightarrow \infty$ in the bulk density limit. It provides useful information about density variations within the system and thus yields information about the local structure. As the probability is normalized to the bulk density, $g = 1$ for regions where particle distribution is uniform, $g < 1$ for regions where there's a depletion of particles, and $g > 1$ where there's an accumulation of particle. In MD it is not generally possible (nor needed) to calculate the RDF for $r > \frac{1}{2}L$, where L is box length, since PBC are used and thus subscribing a sphere with a radius larger than $\frac{1}{2}L$ in the simulation box is not possible considering a cubic box. It is worth pointing out that $g(r)$ is an especially important descriptor in the context of bottom-up coarse-graining as it is one of the target properties that one aims to reproduce in a CG system when building up the model. Another reason behind the importance of $g(r)$ for this work is that the two-body approximation to the excess entropy is derived from it.

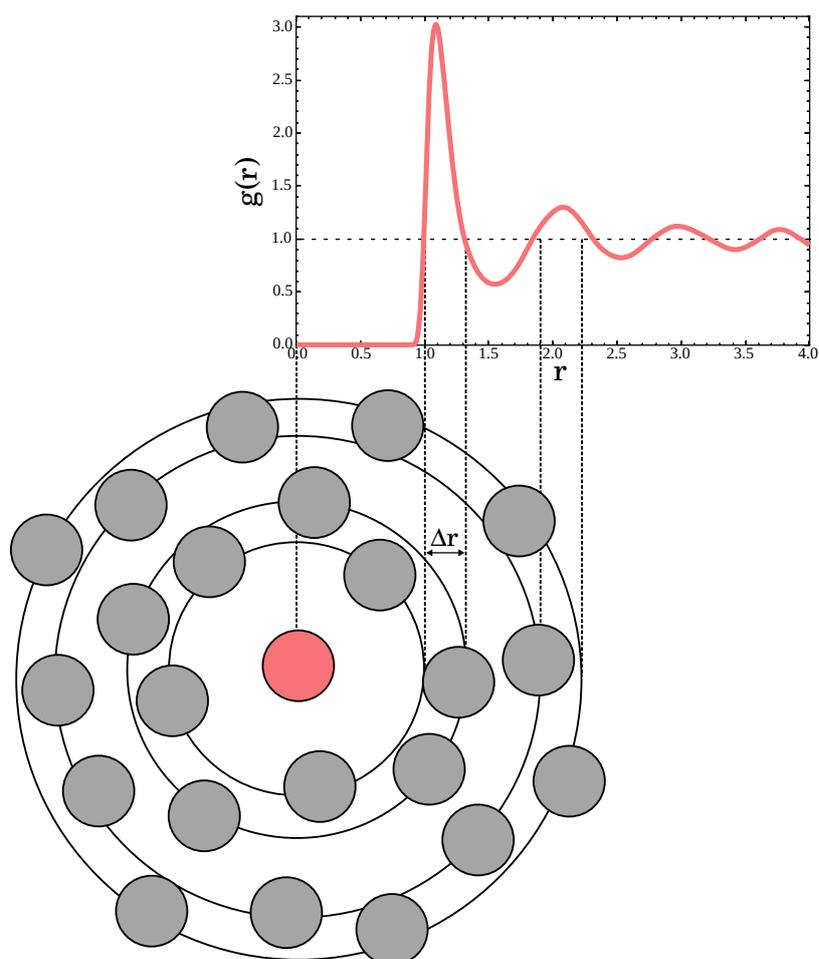


Figure 2.3.: Pictorial representation of the calculation of the RDF, $g(r)$. In the actual calculation each peak corresponds to several bins of width Δr .

2.2. Dynamics in Simulations

This section covers a few dynamical processes of interest for this work from the perspective of MD simulations.

2.2.1. Self-Diffusion

Molecules and atoms in a fluid in thermodynamic equilibrium are subject to random kinetic processes such as translational motions due to thermal fluctuations. This transport process is known as self-diffusion and the rate at which the particles move is related to the system viscosity, its temperature, size and mass of the particles, and is quantified by the self-diffusion coefficient, D . Obtaining this coefficient in MD simulations means tracking the movement of the particles of interest. One method is to use the Green-Kubo (GK)

relation and the velocity ACF [48, 49, 51],

$$D = \frac{1}{3} \int_0^{\infty} \langle \mathbf{v}_i(t) \cdot \mathbf{v}_i(0) \rangle dt, \quad (2.45)$$

where $\mathbf{v}_i(t)$ is the velocity of particle i at time t . This method is alternative to the so called Einstein relation,

$$D = \frac{1}{6} \lim_{t \rightarrow \infty} \frac{d}{dt} \left\langle \left| \frac{1}{N} \sum_{i=1}^N \mathbf{r}_i(t) - \mathbf{r}_i(0) \right|^2 \right\rangle, \quad (2.46)$$

where $\mathbf{r}_i(t)$ is the position of particle i at time t . The term inside the angle brackets is called mean square-displacement (MSD), which in this case is averaged over all particles (cf. Section 2.1.5). It can be seen from Equation (2.46) that the self-diffusion coefficient is related to the slope of the MSD curve, a fact of practical importance as discussed in Section 3.5.1. The limit $t \rightarrow \infty$ is interpreted as the limit in which the long time asymptotic behavior has been reached, characterized by the slope of the MSD curve being proportional to t , i.e., the logarithmic slope is unity. Both methods are entirely equivalent and require similar computational efforts as the integral of the velocity ACF usually converges well. It must be noted that the formalism above considers the self-diffusion of particles. In a molecular fluid what should be strictly tracked is the center-of-mass of each molecule, however for a fluid of identical molecules in equilibrium the MSD of the the center-of-mass of these molecules will match the MSD of their constituent atoms in the long time diffusive regime. This is further illustrated in Chapter 3.

2.2.2. Shear viscosity

The shear viscosity characterizes the resistance that a fluid poses to shear deformation. It is a dynamical property that can be obtained from well-established and relatively straightforward experimental techniques [66, 67]. For this reason it is an interesting quantity to be calculated from simulations as it can be readily compared with experimental results, enabling the assessment of the accuracy of several components of a simulation technique. The shear viscosity of a fluid in equilibrium is related to pressure or momentum fluctuations and can be obtained from these [68, 69]. A well-established method that exploits this relationship to calculate the viscosity in equilibrium MD simulations uses the GK formalism and consists in first obtaining an ensemble average of the ACF of components of the pressure tensor and then integrating it in time. For example, by picking the xz component of the pressure tensor the viscosity may be calculated

as

$$\eta = \frac{V}{k_B T} \int_0^\infty \langle \varsigma_{xz}(t) \varsigma_{xz}(0) \rangle dt, \quad (2.47)$$

where ς is the pressure tensor, and other off-diagonal components may also be used and then averaged. Accurate forces are required for such methods based on pressure fluctuations given that the pressure tensor is determined from the virial,

$$\varsigma_{\alpha\beta}(t) = \frac{1}{N} \left[\sum_{i=1}^N v_{i,\alpha}(t) v_{i,\beta}(t) + \sum_{i=1}^N \sum_{\substack{j=1 \\ i>j}}^N F_{ij,\alpha}(t) (r_{i,\beta}(t) - r_{j,\beta}(t)) \right], \quad (2.48)$$

where $r_{j,\alpha}(t)$ and $v_{i,\alpha}(t)$ are the α component of the position and velocity of particle i at time t , and $F_{ij,\alpha}(t)$ is the α component of the force on particle i due to particle j . This method will converge as the ACF theoretically decays to zero, however this particular correlation function has a very long tail which leads to noise accumulation and makes the convergence difficult, rather fluctuating than converging to a constant value. This is in contrast with the velocity ACF appearing under the integration sign in Equation (2.45), as that ACF can be improved by averaging over all the particles in each time frame. This is not the case with the ACF of the pressure tensor, i.e., more particles will not automatically translate to a better averaged ACF. Even though there is no way to cheat physics and avoid the slow converging nature of this equilibrium-based quantity, it is possible to extract more information from the pressure tensor in order to improve statistics [70, 71].

It has been shown that for systems that are statistically isotropic the convergence can be improved by utilizing the symmetrized portion of the pressure tensor [70], which also includes two independent diagonal components for a total of six components, maximizing the information that can be extracted from the pressure tensor [72, 73]. The generalized GK equation becomes

$$\eta = \frac{V}{10k_B T} \int_0^\infty \left\langle \sum_{\alpha\beta} P_{\alpha\beta}(t) P_{\alpha\beta}(0) \right\rangle dt, \quad (2.49)$$

where α and β identify the tensor components so that $\alpha = \beta$ for diagonal components and $\alpha \neq \beta$ for

off-diagonal components, and $P_{\alpha\beta}$ is the symmetrized portion of the pressure tensor, defined as

$$P_{\alpha\beta} = \frac{1}{2}(\varsigma_{\alpha\beta} + \varsigma_{\beta\alpha}) - \frac{1}{3}\delta_{\alpha\beta} \left(\sum_{\gamma} \varsigma_{\gamma\gamma} \right), \quad (2.50)$$

where $\delta_{\alpha\beta}$ is the Kronecker delta which yields $\delta_{\alpha\beta} = 0$ for $\alpha \neq \beta$. In practice multiple time origins are used to calculate the ensemble average of the ACF, as discussed in Section 2.1.5. Nonetheless, even when using the approach from Equation (2.49), obtaining the viscosity from equilibrium fluctuations is still challenging due to the convergence difficulties of the ACF integral. In particular, identifying the plateau in the $\eta \times t$ plot that corresponds to the converged value is far from trivial. In order to overcome these difficulties, in this work we have employed the method of time decomposition proposed by Zhang et al. [73], as discussed in Section 3.5.2.

It should be brought to attention that non-equilibrium methods to calculate the viscosity are popular alternatives that circumvent the convergence issues faced by equilibrium methods based on the GK integral or the Einstein formalism, not discussed here. It has been shown that both class of methods are able to achieve comparable accuracy in the results [72], but that is not the only factor that should be taken into account when picking any particular method. For simple systems both approaches are similar in terms of computational efforts, whereas equilibrium methods are more costly for more complex systems since long enough simulations are necessary for good statistics and reliable results. Even so, the workflow for equilibrium methods is straightforward and do not involve special adaptations or even different simulations altogether, as it is often the case with non-equilibrium methods where the more involved algorithms commonly contain adjustable parameters that require some degree manual optimization.

One method that merits a mention here is the reverse non-equilibrium molecular dynamics (RNEMD) method of Müller-Plathe [74, 75] in which a momentum flux is imposed in the system by swapping velocities of particles in one direction, and the resulting force is measured to determine the transport coefficient connecting the flux and the force, i.e., the shear viscosity. The method is deemed *reverse* in the sense that it follows the opposite route of experiments and other equilibrium methods, in which usually a force is applied and then the flux is measured. Among the advantages of this method are the relatively simple implementation and fast convergence. In this work we have applied the GK approach to obtain the results that were used in the excess entropy scaling, however the RNEMD method was used to check the

accuracy of the equilibrium results.

2.2.3. Bond Reorientation Relaxation

Another metric of the chain dynamics that was selected for this work was the bond reorientation time. This property is a descriptor of one of the most fundamental relaxation processes in a molecular system, namely how long bonds take to lose memory of their past orientation, and which is also directly measurable in experiments [76]. As this relaxation involves the local mobility of the chain backbone [77–79] this property is a suitable metric to assess the local segmental dynamics of the chains, as opposed to the viscosity or the self-diffusion coefficient that can only provide a global view into the system dynamics. The bond reorientation time can be obtained from the normalized bond orientation ACF, defined here as

$$\Phi_b(t) = \frac{1}{N_c} \frac{1}{N_m - 1} \sum_{n=1}^{N_c} \sum_{m=1}^{N_m-1} \langle \mathbf{b}_{n,m}(t) \cdot \mathbf{b}_{n,m}(0) \rangle, \quad (2.51)$$

where the angle brackets mean an ensemble average as usual, N_c is the number of chains, N_m is the number of monomers per chain, and $\mathbf{b}_{n,m}(t)$ is the time-dependent unit vector connecting the m -th monomer of chain n with the $(m + 1)$ -th,

$$\mathbf{b}_{n,m}(t) = \frac{\mathbf{r}_{n,m+1}(t) - \mathbf{r}_{n,m}(t)}{|\mathbf{r}_{n,m+1}(t) - \mathbf{r}_{n,m}(t)|}. \quad (2.52)$$

The bond reorientation time is simply

$$\vartheta = \int_0^{\infty} \Phi_b(t) dt. \quad (2.53)$$

2.3. Multiscale Modeling

Many important problems that one faces when studying material properties and the governing physics behind them, in particular in soft matter science, are related to phenomena that span over a wide range of length- and time-scales. For polymeric systems in particular these differences in scales are evident not only in structural features, but also in dynamical such as relaxation times. For example, bond lengths are of the order of 1 Å and are relaxed femtoseconds, while whole chains can extend over several nanometers and have whole-chain relaxation times of the order of microseconds, or even longer if highly entangled melts are considered [80]. The structure-property relations are also multiscale in nature, e.g., the relationship

between macroscopic viscoelastic properties of polymers and their underlying microscopic structure. A transition from viscous to elastic behavior, for instance, could be achieved through either local or global modifications to the polymer chains, or both. Replacing certain chemical repeating units changes the polymer locally, while having longer chains and more entanglements affect it globally. Both, however, could independently lead to a significant increase in viscosity. Therefore, understanding the melt viscosity would require investigations on a detailed level focusing on the material and the chemistry, and on a coarser level, taking into account the more general behavior due to the connectivity and topology of the macromolecules [81]. Such problems require a combined approach and are hardly treatable by a single methodology. In particular, realizing the time- and length-scales necessary to investigate phenomena in the meso- and macroscopic scale is not only impossible with purely atomistic simulations, in which all atoms are explicitly treated, but also undesirable as it involves a level of details that are well above what is important to the phenomena under investigation.

The methodology employed must be able to cover the scales at multiple levels of resolution and thus the approach must be a multiscale one that combines hierarchies to address the problem at different levels [82]. These methodological hierarchies are usually defined in terms of individual models that share common properties and are linked with each other in order to attain consistency with the underlying basic problem at hand. This consistency is achieved by adjusting the parameters of each model based on reference information concerning the structure, energy, forces, thermodynamics, etc [83]. In particular, thermodynamic and structural consistency are important in models at different scales that will be used together in order to avoid non-physical behavior and also to prevent a shift of focus from the original problem.

2.3.1. Coarse-Grained Modeling

The core principle of multiscale modeling based on bottom-up approaches, which are the type of approach employed in this work, is to obtain information from a reference model at a higher resolution (i.e., more detailed) and create another model at a lower resolution (i.e., less detailed). As the lower resolution model is simpler it can be applied to larger length scale and simulated longer. Moreover, the model itself will typically be operating at a larger scale automatically with longer internal lengths and characteristic scales. The DOF (i.e., the “details”) of the high-resolution model that are considered unimportant to investigate

the certain problem or property in a different scale are removed, and the output of calculations performed with that high-resolution model are used as input to create the lower-resolution model. For example, one may perform expensive quantum chemical calculations to investigate the interactions between a few water molecules, usually less than 10^2 , where the electrons are explicitly accounted for and from these calculations derive a potential energy surface that will be used as a force-field in classical atomistic MD simulations, where the electrons are considered only implicitly but the atoms are explicitly treated, and in which it is possible to simulate 10^8 water molecules over a wide range of thermodynamic conditions and derive many properties of liquid water. These properties may, in turn, be used to define models at a yet lower resolution continuum-based model such as fluid dynamics models where atoms are now treated only implicitly in order to study the macroscopic flow of water [81].

Coarse-graining is about averaging out features of a model that are deemed less important and focusing on details that are essential for the desired investigation. Such approach provides a clear computational advantage, i.e., only relevant DOF need to be simulated. There is still a conceptual advantage to coarse-graining as well brought about by the increased awareness in the development phase of the model and the need to clearly understand what can and cannot be cast aside. The derivation of the CG potentials is one of the most challenging and laborious task when developing a model. One class of methods for deriving such potentials are the structured-based bottom-up approaches, in which structural information of the high-resolution system are used to construct the CG potential. Within this approach the goal is to build interaction potentials that allow matching structural properties of the FG base system when simulating the CG one.

A typical workflow of the coarse-graining procedure in the bottom-up approach starts by simulating the system of interest at a higher, more detailed resolution. This FG reference system can be an all atom (or atomistic) model, or even an united atom model, for example. This reference simulation will yield a trajectory that needs to be analyzed in order to obtain reference structural information. Such analysis is conducted by employing a specific mapping which essentially defines the CG interaction sites, or CG beads, in terms of the original coordinates, e.g., atomic coordinates of the atomistic model. Once the reference information is at hand, a coarse-graining method is employed to obtain the CG potentials for the bonded and non-bonded interaction of the CG beads. The method itself defines the criteria according to which the CG potential energy surfaces are built.

Before diving into method employed in this work for deriving CG potentials it is convenient to provide a few definitions to clarify the statement of coarse-graining modeling. One central aspect of coarse-graining is the mapping of the FG high-resolution system into a CG scheme, which essentially prescribes how the beads of the CG system will be represented. In the case of the FG model with an all-atom simulation the representation of the system is clear, i.e., each atom corresponds to an interaction site of the system. In coarse-graining there are many possibilities regarding how to define the mapping and, consequently, the representation. The definition of how the mapping procedure must take place strongly influences the range of applicability of the model with respect to the properties of interest, and also to how transferable the resulting model is. An optimal mapping will keep features that play a large enough role in the phenomenon of interest while eliminating sufficient details in order to provide a significant efficiency gain and allow bridging over larger and longer scales, typically integrating out high-frequency fluctuations that are weakly coupled to the slower, global motions [24]. It is worth pointing out that the accuracy of a coarse-graining model doesn't necessarily improve with increasing resolution of the mapping [84], however it is possible to systematically optimize the mapping in order to improve specific features of interest [33, 34, 85].

It is not surprising that a well-defined procedure, e.g., an operator, is needed to transform the actual configuration of the FG model into the configuration of the CG model. Such operator takes as input the spatial coordinates of the FG model and outputs the coordinates of the CG model. In addition to this configuration mapping, another aspect of the mapping that must be considered is the conceptual mapping which does not deal with any particular configuration per se but rather with how one system is defined in terms of another. For example, taking a 3-atom water molecule as a FG model, and a single-bead system as corresponding CG one, the conceptual mapping would not prescribe mathematically how each CG bead is spatially defined within a box of water molecules. It rather specifies that each water molecule is mapped onto a single CG bead, that this CG bead has no internal structure and is not bonded to any other bead, what the pseudo-chemical type of the bead within the CG framework is and how it must interact with other beads. Thus the conceptual mapping refers to definitions of the CG model that go beyond prescribing the definition of beads in terms of an atomistic reference.

A short note on nomenclature: the high-resolution model shall be referred to as fine-grained (FG) or atomistic model in the remaining of this text, even though its particles may not exactly correspond to actual atoms, while the low-resolution model shall be referred to as the CG model and its particles as CG beads.

2.3.2. Deriving Coarse-Grained Potentials

The derivation of the CG potentials starts with the definition of the model, which means defining the CG beads. In general the CG beads are constructed such that their coordinates have a well-defined physical meaning in terms of those of the FG model. Let the Cartesian coordinates and momenta of the N_{FG} particles of the reference FG model be denoted as

$$\mathbf{r}^{N_{\text{FG}}} = \{\mathbf{r}_1, \dots, \mathbf{r}_{N_{\text{FG}}}\}, \quad (2.54)$$

$$\mathbf{p}^{N_{\text{FG}}} = \{\mathbf{p}_1, \dots, \mathbf{p}_{N_{\text{FG}}}\}, \quad (2.55)$$

and let the coordinates and momenta of the $N_{\text{CG}} < N_{\text{AA}}$ beads of a CG model based on the FG one be denoted as

$$\mathbf{R}^{N_{\text{CG}}} = \{\mathbf{R}_1, \dots, \mathbf{R}_{N_{\text{CG}}}\}, \quad (2.56)$$

$$\mathbf{P}^{N_{\text{CG}}} = \{\mathbf{P}_1, \dots, \mathbf{P}_{N_{\text{CG}}}\}. \quad (2.57)$$

The link between the two representations is defined by a linear mapping operator that transforms the FG configuration into the CG one. This operator is defined by a matrix Ω with dimension $N_{\text{CG}} \times N_{\text{FG}}$, where each i -th line prescribes how the i -th CG bead should be constructed. In this notation the CG coordinates are defined as

$$\mathbf{R}_i = \sum_{j=1}^{N_{\text{FG}}} \Omega_{ij} \mathbf{r}_j, \quad (2.58)$$

and the momenta as

$$\mathbf{P}_i = m_{i,\text{CG}} \sum_{j=1}^{N_{\text{FG}}} \frac{\Omega_{ij}}{m_{j,\text{FG}}} \mathbf{p}_j \quad (2.59)$$

where $m_{j,\text{FG}}$ and $m_{i,\text{CG}}$ are the mass of the j -th FG atom and i -th CG bead, respectively. The CG model will generate an equilibrium distribution of momenta that is consistent with the underlying FG model if each atom is part of a single CG bead, and if the mass of the i -th CG bead is defined in terms of the mapping operator as

$$m_{i,\text{CG}} = \left(\sum_{j \in \mathcal{S}_i} \frac{\Omega_{ij}^2}{m_{j,\text{FG}}} \right)^{-1}. \quad (2.60)$$

where S_i is the set of FG atoms that belong into the i -th CG bead,

$$S_i = \{j \mid \Omega_{ij} \neq 0\}, \quad (2.61)$$

as demonstrated by Noid et al. [86]. An illustration of this formalism is given in Appendix A.1.

Ideally one would like to exactly obtain the effective many-body potential resolved at any point in space, and forces derived from it, that would result if a given high-resolution system were mapped to a lower-resolution one. This potential of mean force (PMF) is the key quantity in bottom-up approaches and depends explicitly on the FG model and the CG mapping scheme, and implicitly on the thermodynamic conditions and the volume of the CG system. In practice it is impossible to obtain the PMF but for the simplest systems. Even if that was not the case, the PMF for complex systems would likely be too intricate to be computationally useful in actual simulations. Determining an approximation to the PMF, i.e., the CG interacting potentials, that can be computed in a straightforward manner during a simulation and yields results that are accurate enough to understand the phenomena of interest is the central challenge in bottom-up coarse-graining. The method employed in this work to obtain CG potentials is Boltzmann Inversion, both in the direct and iterative formulations. This approach is attractive from a first-principle point of view given that there is no need for empirical information as the underlying reference model can be constructed entirely from first-principles (e.g., quantum mechanical models). In practice, however, bottom-up models are built based on classical atomistic models that are themselves empirical approximation to a more fundamental model [1].

Direct Boltzmann Inversion

Direct Boltzmann inversion (DBI), or simply Boltzmann inversion (BI), is one of the most straightforward approaches for deriving coarse-graining potentials from a FG reference model. This method essentially yields a PMF for a given DOF, and is mostly used for bonded potentials, e.g., bond stretching, angle bending, dihedral torsions, etc. DBI is based on the idea that in the canonical ensemble the partition function of independent DOF can be factorized, that is, for an independent DOF q its probability distribution is proportional to the Boltzmann factor,

$$P(q) \propto \exp\left(\frac{U(q)}{k_B T}\right) \quad (2.62)$$

where $U(q)$ is the PMF for this DOF. This proportionality could be an equality if the right-hand side were normalized by the canonical partition function, however for the purposes of obtaining a CG potential for q this normalization is unimportant as it would only result in an additive constant. The inversion of Equation (2.62) results in an expression for the PMF,

$$U(q) = -k_B T \ln P(q) + \text{const} , \quad (2.63)$$

where the constant can be ignored, as discussed above. One important assumption here is that q is independent of other DOF, as $U(q)$ as constructed in a way that correlations between different DOF is ignored. This assumption holds well for bonds within a linear chain as they are generally quite stiff and are little unaffected by other DOF. For angles and torsions, or even bonds in more complex conformations such as rings, this assumption may be violated and thus correlation analysis are recommended. If the decoupling of DOF is not possible, e.g., by choosing a carefully crafted mapping scheme, then iterative schemes are usually needed to obtain the bonded potentials.

The DOF q may be a bond length (r), a bending angle (θ), a torsion angle (ϕ), or even a distance coordinate between non-bonded sites, although this latter case is better treated by the iterative versions of the method. The probability distribution, $P(q)$, is built by sampling the DOF q in a simulation from the FG model and building a histogram, e.g., $H(r)$ would be a histogram of bond lengths related to a specific bond type sampled in the FG simulation. One caveat is that the atomistic trajectories have to be mapped onto CG trajectories using the mapping operator previously defined before any analysis. The histograms are then calculated from these mapped trajectories, yielding target distributions that the CG model must ultimately reproduce. Moreover, the histograms have to be rescaled in order to obtain volume normalized distribution functions, as follows,

$$\begin{aligned} P(r) &= \frac{H(r)}{4\pi r^2} , \\ P(\theta) &= \frac{H(\theta)}{\sin \theta} , \\ P(\phi) &= H(\phi) . \end{aligned} \quad (2.64)$$

Iterative Boltzmann Inversion

The Iterative Boltzmann Inversion (IBI) method is an extension of the DBI method in which the Boltzmann-inverted CG potential is self-consistently refined using an iterative scheme. IBI is often used to obtain effective CG pair-potentials for non-bonded interactions or for bonded potentials for correlated DOF, however in this work it has been used only for the former kind of potential. The aim of IBI is to reproduce the target distribution function of the FG model using the CG model and the respective potential derived for it. The starting point is the PMF guess derived from inverting the distribution function in the same fashion as done in DBI, i.e., $U_0(q) = -k_B T \ln P(q)$. This is then iteratively updated with

$$U_{i+1}(q) = U_i(q) + \alpha \Delta U_i(q), \quad (2.65)$$

where $\alpha < 1$ is a system-dependent scaling factor that helps stabilizing the scheme, and the potential update is

$$\Delta U_i(q) = -k_B T \ln \left[\frac{P_i(q)}{P_t(q)} \right], \quad (2.66)$$

where $P_i(q)$ is the distribution obtained in step i in a MD simulation that uses the potential U_i , and $P_t(q)$ is the target distribution of the FG system and which the CG model aims to reproduce. Convergence is reached when the distribution function $P_i(q)$ matches the target distribution within a given tolerance. It should be noted that each IBI step requires an independent MD simulation, which, however short, must be properly equilibrated with the current potential.

For the case of non-bonded pair-wise interactions the DOF is the distance between two particles, $q = r$, and the distribution function in question is the RDF, $g(r)$. The resulting non-bonded CG potential is able to reproduce the structural features of the FG model as they are embedded in the RDF. One should note that this resulting potential is only relevant for the state point for which it was derived given that $g(r)$ is generally dependent on temperature and density. The pressure of the CG system simulated with the IBI-derived potential does not necessarily match the pressure of the FG system, as in IBI only structural properties are target and there was no attempt to optimize the CG potential for thermodynamic properties. One possible route to fix this problem is to add a attractive linear tail term to the CG potential defined as

$$\Delta u(r) = A k_B T \left(1 - \frac{r}{r_{\text{cut}}} \right), \quad (2.67)$$

such that $\Delta u(0) = A$ and $\Delta u(r_{\text{cut}}) = 0$, where r_{cut} is the potential cutoff and the pre-factor A must be tuned for the system. It was originally [87] proposed as $A = -0.1$ but it may have other forms, in particular it can also depend on the difference between the current and target pressure [88, 89]. This pressure correction is added to the iterative scheme as

$$U_{i+1}(r) = U_i(r) + \alpha \Delta U_i(r) + \Delta u(r). \quad (2.68)$$

This adjustment should be carried out after a round IBI yielding a converged non-pressure corrected CG potential, which should then be used as the initial guess and re-optimized with the pressure-corrected IBI.

Other methods

Several other systematic methods are available in the literature to obtain potentials for CG systems from reference data obtained in simulations of high-resolution systems. One is Inverse Monte Carlo [90, 91], which works similarly to IBI by starting with an initial guess for the CG potential derived from a known RDF and then updates it until convergence, but following a Monte Carlo scheme rather than using MD simulations in the iterative steps. Another method that merits mention is Force-Matching [92], which is not structure-based as IBI or IMC but rather targets at reproducing the many-body PMF of the reference system in the CG one. The Conditional Reversible Work method [93, 94] derives physically meaningful CG pair potentials by calculating the interaction free energy at varying distances between groups of atoms in the FG system that will be represented by individual beads in the CG model. In the Relative Entropy method the error between the CG and reference configuration distribution probabilities is minimized, and it also provides a way of achieving optimal mapping [33, 34, 85]. Comprehensive reviews of the ecosystem of CG methods can be found in Refs. [1, 2, 24, 81, 89, 95].

2.4. Estimating Entropies in Simulations

The calculation of entropic quantities from simulations of materials still remains a challenging task [96–98]. A fundamental reason is that the rigorous evaluation of the entropy would mean exploring the whole phase space available to the system under investigation, a task that would require infinitely long simulations [99, 100]. A more practical reason in the simulation context is that, different from properties such as

energy, pressure or temperature, entropy cannot be expressed as averages of functions of the trajectory (i.e., momenta and coordinates) yielded by a MD simulation [51, 101]. In the statistical mechanical sense each microstate available to a system has a calculable and well-defined energy such that, for example, an ensemble average is possible to determine the internal energy of the system. This is not the case for entropy, which is rather a collective property to which all available microstates contribute, i.e., determining absolute values for entropy, and also entropy differences, from MD simulations require the evaluation of the complete partition function, which means comprehensive sampling of all DOF.

In this work we have employed two different methods that to estimate entropies from MD simulations. The first of these methods is theoretically exact and yields the most accurate estimate that is considered in this thesis, which is the method known as thermodynamic integration (TI). Even though it is in principle exact, the computational effort and sheer number of simulations required to obtain useful results is significant and even then numerical fluctuations can be a problem. Therefore it is also worthwhile to consider as well alternative methodologies that, even if not as accurate, can be realized without as much effort. In this work we also consider a two-body approximation to the excess entropy which can be easily computed from the pair-distribution function obtained from simulations. Both methods are discussed in more details in what follows.

2.4.1. Thermodynamic Integration

As with real-life experiments, it is not possible do directly determine exact absolute entropies from simulations, but only their differences. Using TI one is able to obtain the entropy difference between states a and b which is in principle exact, as long the absolute entropy of the state taken as reference is known, and as long as it is possible to define a reversible path over which the Hamiltonian of the system can be changed from a to b . In real-world experiments the integration is restricted to a few thermodynamic variables and limited choices of reversible paths, if these exist at all. In a simulation, however, the path between the reference state and the state of interest does not need to be physical: it can be entirely theoretical and built as conveniently as possible by adding nonphysical parameters that couple the states to the path. Moreover, reversibility is achieved by taking short enough intervals over the path of choice. This transition path can be practically defined by adding a coupling parameter to the Hamiltonian of the system in such a way that the nature of the system can be controlled by varying this parameter [99, 102].

The connection between the Helmholtz free energy F , which is one of the key quantities that can be obtained from the canonical ensemble, and the canonical partition function is expressed as

$$F = -k_B T \ln Z. \quad (2.69)$$

where Z is the canonical partition function, as introduced Section 2.1.5. Since the Helmholtz free energy is a thermodynamic potential that can be expressed as a function of temperature, volume and composition, it can be used to obtain other thermodynamic functions, in particular the entropy, e.g., via the thermodynamic identity

$$S = - \left(\frac{\partial F}{\partial T} \right)_{N,V}, \quad (2.70)$$

which, after substituting Equation (2.69) and carrying out the partial derivative with respect to the temperature, leads to

$$S = k_B \ln Z + \frac{k_B T}{Z} \frac{\partial Z}{\partial T} \quad (2.71)$$

$$= k_B \ln Z + \frac{\langle H \rangle}{T}, \quad (2.72)$$

which by equating expectation value of the Hamiltonian with the system internal energy, $\langle H \rangle = E$, and using Equation (2.69), the equation above recovers the thermodynamic definition of the Helmholtz free energy $F \equiv E - TS$, thus

$$S = \frac{E - F}{T}, \quad (2.73)$$

and therefore the entropy difference between two states a and b of system is

$$\Delta S_{ba} = \frac{\Delta E_{ba} - \Delta F_{ba}}{T}. \quad (2.74)$$

The value for ΔE_{ba} can be directly estimated from two sets of MD simulations, one for each of the states. As the total energy depends on the position and momenta for all the particles per Equation (2.30), it is indeed a quantity that is readily available in a simulation and can be saved every so often and then averaged over each run and over all origins in each set. The same is not true for the free energy difference ΔF_{ba} . As seen above, Equation (2.69) is the expression that connects F to the microscopic description

through Z , thus it is a function of an integral of the whole phase space and thus cannot be estimated in a straightforward manner from a MD simulation. In this work we use TI to determine this free energy difference, however other methods are also available, such as perturbation theory methods, umbrella sampling, weighted histogram methods, and single- and multi-state Bennett Acceptance Ratio [103].

In order to calculate ΔF_{ba} using TI it is convenient to make the Hamiltonian explicitly dependent on the state. This can be done by adding a coupling parameter λ to the Hamiltonian [104],

$$H(\lambda) = H(\mathbf{r}^N, \mathbf{p}^N; \lambda) \quad (2.75)$$

such that $\lambda = \lambda_a$ and $\lambda = \lambda_b$ for states a and b , respectively. This λ -dependence of the Hamiltonian also makes the canonical partition function and the Helmholtz free energy λ -dependent. In this context, the difference ΔF_{ba} can be written as an integral

$$\Delta F_{ba} = F_b - F_a = \int_{\lambda_a}^{\lambda_b} \frac{\partial F}{\partial \lambda} d\lambda, \quad (2.76)$$

where $F_a = F(\lambda_a)$ and $F_b = F(\lambda_b)$. The λ derivative of F can be obtained by differentiating both sides of Equation (2.69) with respect to λ , yielding

$$\left(\frac{\partial F}{\partial \lambda} \right)_{N,V,T} = \left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_{\lambda} \quad (2.77)$$

where $\langle \dots \rangle_{\lambda}$ means an ensemble average for a particular value of λ , i.e., at a well-defined state. It is worth calling attention to the fact that typically the λ -dependence is added only to the potential energy term of the Hamiltonian, i.e., $H(\mathbf{r}^N, \mathbf{p}^N; \lambda) = K(\mathbf{p}^N) + U(\mathbf{r}^N; \lambda)$, thus $\partial H / \partial \lambda = \partial U / \partial \lambda$ and hence the free energy difference in this TI framework is given as

$$\Delta F_{ba} = F_b - F_a = \int_{\lambda_a}^{\lambda_b} \left\langle \frac{\partial U(\lambda)}{\partial \lambda} \right\rangle_{\lambda} d\lambda, \quad (2.78)$$

and the entropy difference

$$\Delta S_{ba} = S_b - S_a = \frac{1}{T} \left[\left(\langle E \rangle_{\lambda_b} - \langle E \rangle_{\lambda_a} \right) - \int_{\lambda_a}^{\lambda_b} \left\langle \frac{\partial U(\lambda)}{\partial \lambda} \right\rangle_{\lambda} d\lambda \right]. \quad (2.79)$$

The practical evaluation of the integral in Equation (2.79) requires the specification of an integration path prescribing how the Hamiltonian changes from $H(\lambda_a)$ to $H(\lambda_b)$. This pathway does not need to reflect a real physical or chemical process but the change must be reversible and thus carried out slowly. Since numerical integration schemes are discrete, a set of intermediate states in the path must be selected, each identified by a distinct λ -point, and an independent equilibrium MD simulation is required for each of these points. How large the set must be depends on the Hamiltonian but at least 10 – 20 are needed in order to avoid large biases, in particular if $\partial H/\partial\lambda$ has a complex profile. The need to perform several MD simulations for the intermediate states entail a significant computational cost and thus it is one of the drawbacks of TI.

One aspect that was not mentioned so far is how to transform the Hamiltonian to include the coupling parameter. Such task is system-dependent and typically involves modifying the potential energy contribution of the Hamiltonian in order to accommodate the state dependence. One example is switching off dispersion interactions at $\lambda = 0$ and recovering the full interactions at $\lambda = 1$, and then calculating the free energy or entropy difference between these two states. A certain care must be taken in order to avoid singularities, in particular when the potential energy is described by pair-potentials such as LJ. Moreover, if the potential energy is tabulated as it is often the case for CG potentials obtained from bottom-up approaches then the introduction of the λ -dependence is more involved than with analytical potentials.

It is also possible to follow a more direct TI approach and obtain an expression for the entropy in the canonical ensemble. The λ -dependence S may be expressed by going back to Equations (2.69) and (2.70) and considering that H and thus Z and F depend now on λ , thus leading to the λ -dependent of Equation (2.72),

$$S(\lambda) = k_B \ln Z(\lambda) + \frac{\langle H(\lambda) \rangle}{T} \quad (2.80)$$

$$= k_B \ln Z(\lambda) + \frac{1}{T} \frac{\int \int H(\mathbf{r}^N, \mathbf{p}^N; \lambda) \exp(-H(\mathbf{r}^N, \mathbf{p}^N; \lambda)) d\mathbf{r}^N d\mathbf{p}^N}{\int \int \exp(-H(\mathbf{r}^N, \mathbf{p}^N; \lambda)/k_B T) d\mathbf{r}^N d\mathbf{p}^N}, \quad (2.81)$$

which can be differentiated with respect to λ to yield

$$\left(\frac{\partial S}{\partial \lambda} \right)_{N,V,T} = \frac{1}{k_B T^2} \left\{ \left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_\lambda \langle H(\lambda) \rangle_\lambda - \left\langle \frac{\partial H(\lambda)}{\partial \lambda} H(\lambda) \right\rangle_\lambda \right\}, \quad (2.82)$$

which leads to the following expression for the entropy difference [99]

$$\Delta S_{ba} = S_b - S_a = \frac{1}{k_B T^2} \int_{\lambda_a}^{\lambda_b} \left\{ \left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_{\lambda} \langle H(\lambda) \rangle_{\lambda} - \left\langle \frac{\partial H(\lambda)}{\partial \lambda} H(\lambda) \right\rangle_{\lambda} \right\} d\lambda, \quad (2.83)$$

where the integration needs to be carried out over a path $H(\lambda_a) \rightarrow H(\lambda_b)$ in an analogous way to the case of the integration in Equation (2.79). However, this formulation is very prone to large fluctuations due to the correlations between the Hamiltonian and its λ -derivative, and thus for this work it was not used and only included here for completeness.

2.4.2. Two-body Approximation

Calculating entropic quantities with TI is a laborious task, despite the accurate results that can be achieved if one is willing to invest sufficient computational effort. Hence it is only natural to seek for alternative, less involved, and importantly less costly methods to obtain entropies and their differences from MD simulations. One possibility arises from the fact that the entropy of a system of N particles with constant volume V and temperature T can be expanded in terms of canonical ensemble multi-body spatial correlation functions

$$S = S_1 + S_2 + S_3 + \dots, \quad (2.84)$$

where S_n is the contribution to the total entropy of the n -body correlations. This expansion has been proposed and developed by several authors and is directly linked to the connection between the entropy and the internal structure of the system and besides prescribing an easier route for entropy estimates, it also grants useful insights regarding the connection of structure and thermodynamics [35]. The first term of the series, S_1 , captures the contribution of monoatomic, non-interacting (and thus uncorrelated) particles and is simply given by the entropy of an ideal gas (IG) of N particles at the same density and temperature as the actual system. The series in Equation (2.84) may evidently be expanded up to N terms, however in practice it is usually unfeasible, and even impossible, to calculate the n -body correlation functions for $n > 3$, and even for $n = 3$ this is already a substantial challenge [105].

The expansion of entropy up until $n = 3$ has been derived by Baranyai and Evans [96] in a form that

can be stably used in numerical computations in the canonical ensemble and is written as

$$\begin{aligned}
s = s_{\text{ig}} &- \frac{1}{2}\rho \int [g_2(\mathbf{r}) \ln g_2(\mathbf{r}) - g_2(\mathbf{r}) + 1] d\mathbf{r} \\
&- \frac{1}{6}\rho^2 \int \int g_3(\mathbf{r}, \mathbf{s}, \mathbf{t}) \ln \delta g_3(\mathbf{r}, \mathbf{s}, \mathbf{t}) d\mathbf{r}d\mathbf{s} \\
&+ \frac{1}{6}\rho^2 \int \int [g_3(\mathbf{r}, \mathbf{s}, \mathbf{t}) - g_2(\mathbf{r})g_2(\mathbf{s}) - g_2(\mathbf{s})g_2(\mathbf{t}) - g_2(\mathbf{r})g_2(\mathbf{t}) + g_2(\mathbf{r}) + g_2(\mathbf{s}) + g_2(\mathbf{t}) - 1] d\mathbf{r}d\mathbf{s},
\end{aligned}$$

where the per-particle notation $s \equiv S/Nk_B$ has been adopted, s_{ig} is the total entropy of the IG, g_2 is the pair correlation function, g_3 is the triplet correlation function. The g_2 and g_3 correlations are expressed as functions of the relative distances involving three particles, $\mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1$, $\mathbf{s} = \mathbf{r}_3 - \mathbf{r}_1$, $\mathbf{t} = \mathbf{r} - \mathbf{s}$, where \mathbf{r}_1 , \mathbf{r}_2 , and \mathbf{r}_3 are the positions of the three particles [106]. The term δg_3 represents the irreducible triplet correlation that cannot be factored into pure 2-body correlations and stems from

$$g_3(\mathbf{r}, \mathbf{s}, \mathbf{t}) = g_2(\mathbf{r})g_2(\mathbf{s})g_2(\mathbf{t})\delta g_3(\mathbf{r}, \mathbf{s}, \mathbf{t}). \quad (2.85)$$

The importance of each correlation function g_N in Section 2.4.2 is expected to decrease with increasing N , while the complexity of calculation increases. This is one reason why the expansion is often evaluated up to the 2-body contributions [107]. Another reason is that much of the structural information of a molecular fluids is available by in the pair correlation function, and even though the 3-body correlations are not negligible [108], the evaluation of s_2 in the context of a MD simulation is simplified by the fact that the pair correlation function readily extracted from the trajectory in the form of the RDF, $g(r)$, i.e.,

$$s_2 = \frac{S_2}{Nk_B} = -2\pi\rho \int [g(r) \ln g(r) - g(r) + 1] r^2 dr. \quad (2.86)$$

where $r = |\mathbf{r}_2 - \mathbf{r}_1|$, thus making it a very convenient route to the entropy. The formulation given above is known as the translational portion of the two-body approximation of the configurational excess entropy, as the orientational contribution has been integrated out. For monoatomic LJ fluids this two-body contribution has been shown to account for at least 85% of the total excess entropy, which is the entropy due to multi-particle correlations [107]. The excess entropy is defined and discussed in Section 2.5.

One aspect of the multi-particle expansion of the entropy given in Section 2.4.2 that is often overlooked in the literature is that the integrals of the correlations must be evaluated for the whole system, i.e., the

complete minimum image cell and not for just sub-region within a predefined cutoff. Clearly the images of the simulation box should not be included, as they do not contribute to the entropy, but if a cubic box is used then care must be taken so that the corners of the box are also accounted for when computing correlations, in particular the pair correlation functions which is of relevance for the two-body term [96, 108].

2.5. Excess Entropy Scaling

Connecting microscopic transport properties of a system with its thermodynamic parameters has been the focus of several studies [27]. The first theoretical attempts at deriving correlations were mostly empirical and numerous experimental and computational investigations that followed aimed at providing evidences for such correlations and also justifying them on more solid theoretical grounds. One of the main topics revolves around expressing transport properties as single-valued functions of excess entropy [26, 109, 110], since the validity of such correlations would make it easier to derive transport coefficients, and especially the scaling thereof in the context of coarse-graining simulations, if excess entropy is known. Excess entropy, in turn, can be computed with varying degrees of accuracy and computational effort, as mentioned in Section 2.4.

2.5.1. Excess Entropy

The total entropy, S , of a classical fluid at temperature T and density ρ can be expressed as the sum of the total entropy of the corresponding IG contribution, S_{ig} , at the same temperature T and with the same density ρ , and an excess term, S_{exc} , which lowers the IG entropy due to the introduction of interactions and structural correlations,

$$S = S_{\text{ig}} + S_{\text{exc}}. \quad (2.87)$$

where the excess term accounts, e.g., for the effect of pair and higher-order correlations [111]. It should be noted that the excess entropy as defined above is different of an alternative and often-used definition in the field of glass-forming liquids, where it is defined as the difference between crystalline or glass material entropy the fluid entropy [112]. The results and discussions in this work only apply to the definition of Equation (2.87). Methods of estimating S_{exc} are discussed in Chapter 2.

2.5.2. Scaling Relations

The mechanics behind excess entropy scaling is as follows: a reduction factor for the dynamical property X of interest is introduced, which for instance can be the self-diffusion coefficient D or the viscosity η , such that this factor normalizes the transport property rendering it dimensionless. This reduced quantity then should be written as a single-valued function of the excess entropy, e.g., using a scaling relation given by an exponential functional form. In other words, the first step is determining a reduction factor ξ such that it is possible to write $\tilde{X} = \xi X$, where \tilde{X} is the reduced dimensionless transport coefficient. This term, in turn, is then expressed as a function of the excess entropy, $\tilde{X} = \tilde{X}(S_{\text{exc}})$. The main difficulty is that the reduction factors, as ξ in this example, by which transport properties can be transformed into reduced forms are not unique. They can be designed empirically or from first-principles, in terms of macroscopic thermodynamic parameters such as temperature and density, or microscopic properties, such as characteristic lengths and times, etc. The same arguments are valid for the corresponding scaling laws that relate the reduced properties to excess entropy, e.g., they can be microscopic- or macroscopic-inspired.

One of the first to propose scaling relations connecting transport properties with excess entropy in the context of fluids was Rosenfeld [26, 113], whose empirical approach was followed by several studies that aimed to rationalize it on more sound theoretical arguments [26, 114, 115]. Among the systems to which the Rosenfeld approach was applied are ionic liquids [116, 117], hydrocarbons [118, 119] and fluids with low molecular weight [29]. Another scaling relation that attracted significant interest and achieved considerable success in providing useful relations linking dynamics and thermodynamics of fluids is the Dzugutov approach [109]. This method provides scaling laws which relate transport properties not to the full excess entropy of the system, but rather to the two-body contribution of the excess entropy, which as seen in Section 2.4.2 can be calculated without significant hurdles in simulations [120]. A few authors applied excess entropy scaling laws to LJ chain systems which are of interest in the context of this work. Goel et al. [121] focused on homogeneous chains formed by up to 10 monomers and modeled by freely rotating linear strings with, due to their short lengths, no impact of entanglements on the overall dynamics. Of more immediate interest to this project are the works from Voyiatzis et al. [78, 122, 123], in which the application of entropy scaling relations to the dynamics of LJ chains with lengths up to 70 monomers was validated for the first time.

2.5.3. Rosenfeld Scaling

The excess entropy scaling relations introduced by Rosenfeld [25, 26, 113] paved the way for the development of several other scaling methods. Rosenfeld's approach introduced a reduction factors for dynamical properties based on the mean free path in a fluid, $l = \rho^{-1/3}$, and the mean thermal velocity, $v_T = \sqrt{k_B T/m}$. The Rosenfeld-reduced expressions for the self-diffusion, viscosity, and bond reorientation time are, respectively,

$$\tilde{D} = D \frac{\rho^{1/3}}{\sqrt{k_B T/m}}, \quad (2.88)$$

$$\tilde{\eta} = \eta \frac{\rho^{-2/3}}{\sqrt{m k_B T}}, \quad (2.89)$$

and

$$\tilde{\vartheta} = \vartheta \frac{\sqrt{k_B T/m}}{\rho^{-1/3}}. \quad (2.90)$$

The relation between a Rosenfeld-reduced dynamical property X and the excess entropy is given by the scaling relation

$$\tilde{X} = A_R \exp(\alpha_R s_{\text{exc}}), \quad (2.91)$$

where the per-particle normalized excess entropy is $s_{\text{exc}} = S_{\text{exc}}/Nk_B$, and A_R and α_R are coefficients that must be fitted to data from simulations or experiments and that depend on the nature of the system, such as chain length in the case of polymers, and to the transport property, such as self-diffusion coefficient or viscosity. This exponential scaling can be more conveniently rewritten as a linear relation,

$$\ln(\tilde{X}) = \ln(A_R) + \alpha_R s_{\text{exc}}. \quad (2.92)$$

2.5.4. Dzugutov Scaling

The scaling relation proposed by Dzugutov [109] was derived for simple atomic and molecular systems and differently from Rosenfeld's macroscopic view, it takes a microscopic approach based on the notion that different effects are influencing the microscopic dynamics at different time scales. Local density fluctuations and cage effects play an important role at short time scales, with the dynamics at these time scales being dominated by the collisions of atoms with their surrounding cage at a rate is given by the Chapman-Enskog

theory [124, 125] and equal to

$$\Gamma_E = 4 \sigma^2 g(\sigma) \rho \sqrt{\pi k_B T / m}, \quad (2.93)$$

where σ corresponds to the position of the first maximum of the RDF, $g(r)$. Taking σ , Γ_E^{-1} , and m as natural units of length, time, and mass, respectively, the Dzugutov-reduced self-diffusion coefficient, viscosity, and bond reorientation time are written as

$$\tilde{D} = \frac{D}{\sigma^2 \Gamma_E}, \quad (2.94)$$

$$\tilde{\eta} = \eta \frac{\sigma}{m \Gamma_E}. \quad (2.95)$$

and

$$\tilde{\vartheta} = \frac{\vartheta}{\Gamma_E^{-1}}, \quad (2.96)$$

respectively.

At longer time scales the dynamics is dominated by the rate at which the cage structures are relaxed, which is proportional to the number of accessible configurations and thus to the system's entropy. Assuming that the two-body entropy, S_2 , contributes the most to the excess entropy, Dzugutov proposed the following excess entropy scaling relation for a dynamical property X ,

$$\tilde{X} = A_D \exp(\alpha_D s_2), \quad (2.97)$$

where the per-particle normalized two-body entropy is $s_2 = S_2 / N k_B$, and A_D and α_D are fit coefficients and which Dzugutov originally suggested $A_D = 0.049$ and $\alpha_D = 1$ for monoatomic liquid systems. As with the Rosenfeld scaling relation, a linear form can be convenient,

$$\ln(\tilde{X}) = \ln(A_D) + \alpha_D s_2. \quad (2.98)$$

3. Models and Methodology

The theoretical base for this work was covered in Chapter 2. The present chapter establishes how the theory was applied to enable this project and details of the strategy taken. Details pertaining to the methods employed and how they were adapted are discussed herein, including parameters used, computational details of the simulations, analysis tools and algorithms, validation of the results, and comparison with alternative methods. A general overview of the workflow is also presented. Furthermore, the models that were used and developed in this work are described, as well as the steps needed to realize them. Selected results are presented herein in the context of discussing particular aspects of the methodology.

3.1. Models

The goal of this work is to investigate the relations between excess entropy differences and the dynamical acceleration in polymeric systems represented by models at different resolutions. In order to focus on the properties and avoid artifacts introduced by the models, the underlying model adopted as the FG model is a system of Lennard-Jones chains (LJC) representing a melt of bead-spring polymers, each chain with fixed length of $N_m = 24$ monomers. This system is referred to as LJC due to the fact that non-bonded interactions between intra- and interchain monomers are described by the 12-6 LJ potential, i.e., the monomers are so called LJ atoms. This non-bonded potential is used to describe the interactions between all monomers that are not connected via a harmonic spring, which means that even second neighbors in the same chain are accounted for when evaluating the potential, as are monomers belonging to other chains, and only the first bonded neighbor is excluded. A distance cutoff of $r_{\text{cut}} = 5$ reduced length units (cf. Section 3.1.1) is used when building neighbor lists for force and energy calculations. The FG model is shown in Figure 3.1.

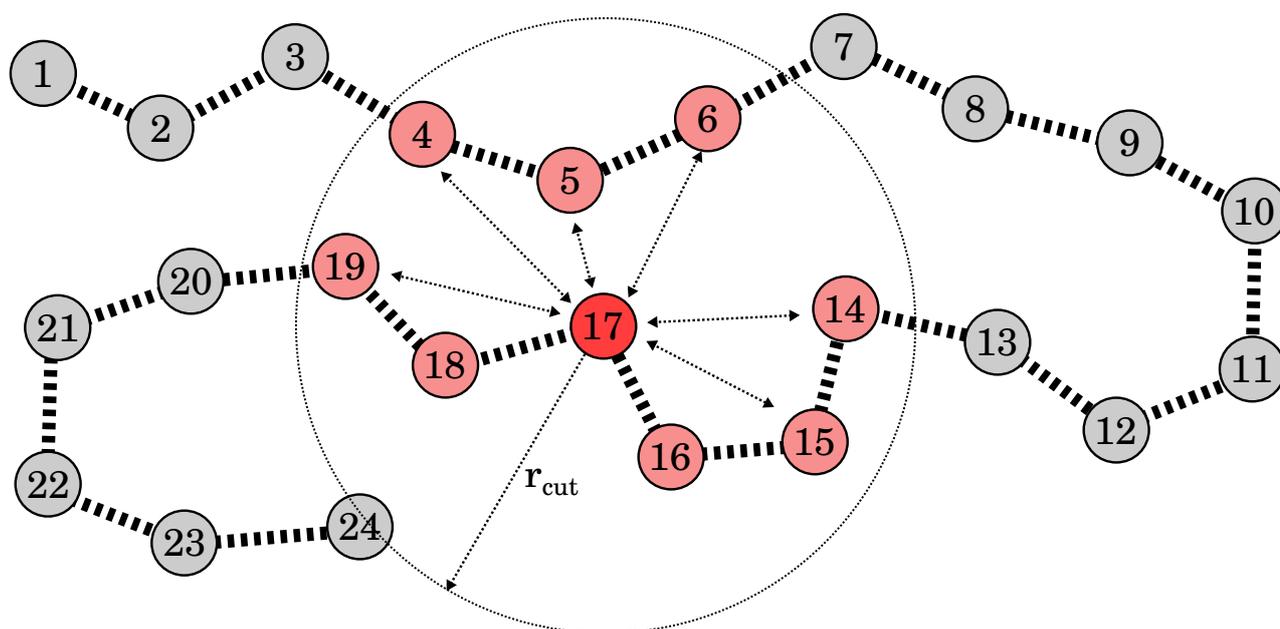


Figure 3.1.: FG model of a 24-monomer polymer chain. The bonds between the monomers are modeled by harmonic springs and all but first neighbors interact via pair-wise non-bonded LJ potential, as depicted for monomer 17 and its non-bonded neighbors within the potential cutoff distance, namely monomers 4, 5, 6, 14, 15 and 19, which are those included in the LJ interaction. Non-bonded interaction is also applied to pairs of monomers belonging to different chains (not depicted).

The bonds between the monomers within each LJC are modeled as harmonic springs where the equilibrium bond length and the spring constant are set to 1 and 3000 in reduced LJ units, respectively. This seemingly very simple model and the corresponding choice of parameters are popular in the literature [121, 122, 126], and yields a system that is generic enough to capture the chain dynamics at the length considered. It is worth pointing out that the entanglement length of such bead-spring polymer models at similar thermodynamic conditions as considered in this work ranges from 35 to 85 monomers [127, 128], thus this work deals only with unentangled models.

The LJC model was used as the FG reference for two CG models at different, less detailed resolutions. In order to maintain a workflow consistency in all models and have the same baseline for all results, the CG models were derived from scratch for this work. The two CG models were created by following two different mapping schemes. One of the CG models is called here CG2 and employs a 2:1 mapping scheme, which means that every CG bead of CG2 corresponds to two adjacent monomers of the FG¹. The other CG model is called CG3 and employs a 3:1 mapping scheme. Therefore the lengths of chains in the CG models are reduced from 24 to 12 in the CG2 system and 8 in the CG3. The mass of the FG monomers is set to unity in the reduced unit system, $m_{\text{FG}} = 1$, and the mass of the CG beads are $m_{\text{CG2}} = 2$ and $m_{\text{CG3}} = 3$, thus preserving the original mass of the FG model. A third CG model was attempted using 4:1 mapping, however the resulting CG potentials had a too soft character to be useful, allowing CG beads and bonds to cross each other. The topology mapping from FG to CG2 and CG3 followed a simple isotropic scheme with each CG bead centered on the center of mass of the reference group. The three models are depicted in Figure 3.2.

3.1.1. Unit System

The unit system adopted in this work is based on reduced LJ units [72]. The energy and length parameters of the LJ potential are set to unity, i.e., $\epsilon = 1$ and $\sigma = 1$, respectively, as is the mass of the FG monomers, $m_{\text{FG}} = 1$. All quantities are reported in reduced units based on the parameters of the LJ potential used for the FG system, i.e., length, energy, volume, number density, temperature, and time are expressed in units of σ , ϵ , σ^3 , σ^{-3} , ϵ/k_{B} , $\sqrt{m\sigma^2/\epsilon}$, respectively, where k_{B} is the Boltzmann constant, and $m = m_{\text{FG}}$. For brevity, the units are omitted altogether in all quantities reported in this thesis.

¹The terms “monomers” and “beads” are used interchangeably to refer to the repeating units of the model chains at any resolution.

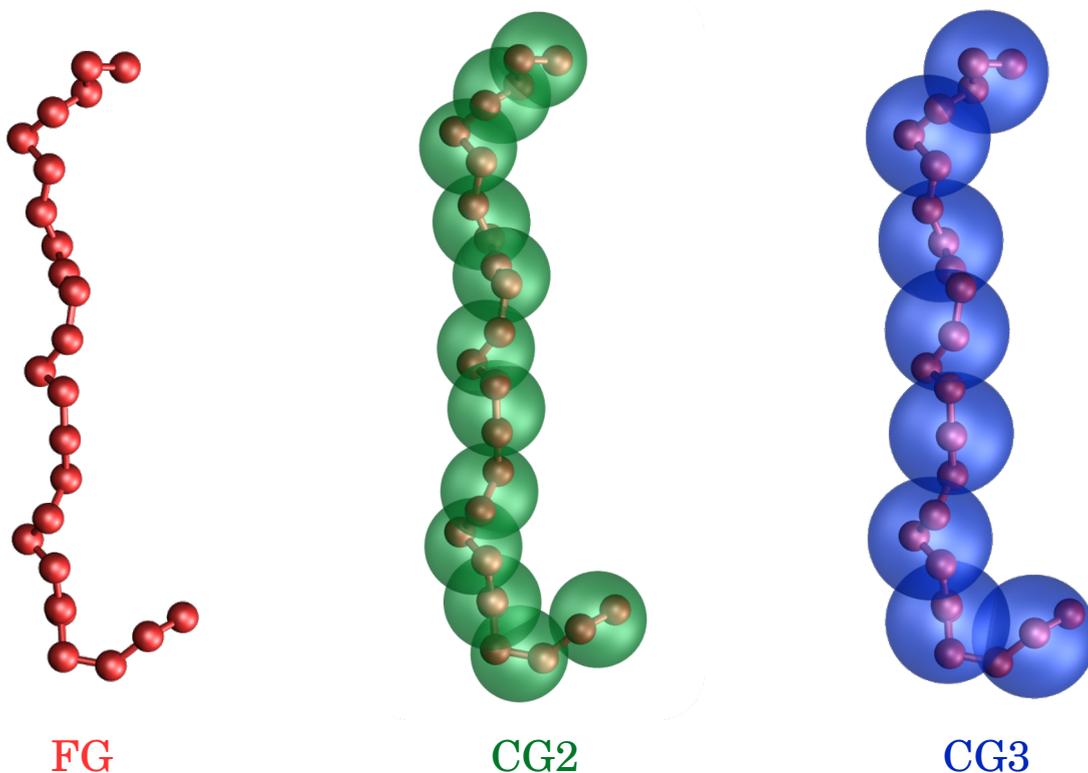


Figure 3.2.: Polymer models employed in this work. From left to right, FG LJC of length 24 monomers, CG model of length 12 beads with 2:1 mapping scheme (CG2) based on the FG model, and CG model of length 8 beads with 3:1 mapping scheme (CG3) based on the FG model.

3.1.2. Thermodynamic State Points

An important aspect of this investigation lies in the thermodynamic conditions under which the models are simulated, i.e., which state points are used. On the one hand, it is essential that the physical state of the system is the same for all state points considered; on the other hand, it is desirable that a reasonably broad range for excess entropy values are obtained. Both requirements cannot be guaranteed in a straightforward way *a priori*, however there is enough information in the literature available to inform de decision [121, 122, 126, 129]. Based on this information two liquid densities were selected: $\rho = 0.8$ and 0.9 , where $\rho \equiv N/V$, where N is the total number of FG monomers in the simulation box ², and V the volume of the box. For each density, all simulations were performed at various reduced temperatures in the range

²The system occupies the whole simulation box in a single homogeneous phase.

2.0–4.2, for a total of 19 thermodynamic state points.

3.2. Overview of the Workflow

An overview of the workflow carried out to obtain the data for the results discussed in this thesis is presented below. The order of presentation is not necessarily chronological and some of the tasks were performed in a different order than shown, while a few evidently need to precede others. This is a summary of the simulations and main tasks, presented in a single location for an easy overview. Detailed information pertaining to post-processing and method-specific analysis of the simulation data are available in the following Sections.

Preparation of the FG system

- Independent construction of the FG system for each state point using a step-wise bond-growth procedure coupled with short Monte Carlo pre-equilibrations.
- Thermalization of the FG systems using the Langevin thermostat.
- Equilibration of the FG systems using the Nosé-Hoover thermostat.

Derivation of the CG potentials

- Independent initialization of the CG systems based on mapping the corresponding FG trajectories.
- Derivation of bonded CG potentials from bond-length distributions obtained from the reference CG-mapped trajectories using DBI.
- Derivation of non-bonded CG potentials using IBI with pressure corrections, and starting from an initial guess based on the RDF calculated from the reference CG-mapped trajectories.
- Post-processing and refinement of the tabulated CG potentials so that they could be used with LAMMPS [130, 131].
- Thermalization of the CG systems using the Langevin thermostat.
- Equilibration of the CG systems using the Nosé-Hoover thermostat.

Calculation of dynamical and structural properties

-
- Production runs of the MD simulations for data collection for all FG and CG systems yielding thermodynamic and energetic data, and trajectories with coordinates and velocities.
 - Calculation of the self-diffusion coefficient using the MSD derived from the trajectories.
 - Calculation of the bond reorientation times from the bond orientation ACF derived from the trajectories.
 - Calculation of the viscosity from the symmetrized pressure tensor data using the equilibrium GK formalism.
 - Validation of the equilibrium viscosity with results from RNEMD simulations.

Calculation of entropic quantities

- Estimation of excess entropy using TI.
 - Analytical determination of the absolute entropy of the FJC model for each resolution and thermodynamic state point, yielding S_{FJC} .
 - Derivation of intermediate bonded potentials for different λ -points for calculating the entropy difference between the rigid FJC system and the flexible FJC system.
 - Derivation of intermediate non-bonded potentials for different λ -points for calculating the entropy difference between the non-interacting flexible FJC model and the actual interacting FG and CG systems.
 - TI production runs for different λ -points yielding energetic data for the entropy difference between rigid and flexible FJC systems.
 - TI production runs for different λ -points yielding energetic data for the entropy difference between non-interacting and interacting systems.
 - Integration of data obtained from the production runs, yielding ΔS_{b} and ΔS_{nb} .
 - Calculation of excess entropy from entropic data obtained in previous steps.
- Calculation of high-accuracy RDFs from the trajectories resulting from the production runs.
- Estimation of the two-body excess entropy from these high-accuracy RDFs.

Evaluation of entropy scaling relations

Apply the renormalization relations to the dynamical properties.

Correlate the renormalized dynamics properties and changes thereof across resolutions with excess entropies and changes thereof across resolutions.

3.2.1. Computational Details

All the MD simulations performed for this thesis employed a cubic simulation box using PBC and containing $N_c = 500$ chains, with the box dimensions adjusted to yield the desired number densities. The equations of motion were integrated using the velocity-Verlet algorithm and a time step length $\Delta t = 0.001$, which was kept even for the CG simulations for which it is generally possible to increase the time step as the fastest DOF are coarse-grained out of the model, and also due to smoother potential energy surfaces [132]. However, to avoid any artifacts in our CG simulations [133] we have kept the time step equal across all resolutions. The simulations were realized in the canonical ensemble and the Nosé-Hoover thermostat was employed to maintain constant-temperature conditions with the temperature damping parameter of the thermostat set to $100\Delta t$. The MD simulations were performed with versions from 2016 and 2017 of the LAMMPS package [130, 131] partially on the Biby and Sniffa HPC clusters of the Theoretical Physical Chemistry Group, and partially on the JURECA supercomputer of the Juëlich Research Center [134].

3.3. System Preparation

The FG system was initialized independently for each of the 19 pairs of density and temperature. The initialization was performed using a bond growth method coupled with Monte Carlo relaxation as implemented in the package Enhanced Monte Carlo (EMC) [122, 135]. In this strategy we start with an empty cubic box with edge size defined by the prescribed density, and with a single chain-starting monomer randomly placed in the box. This will be the first monomer of a new chain and will not necessarily correspond to a terminal monomer. Another monomer is added at a bonding distance from the previous monomer and then a few thousand pre-relaxation Monte Carlo steps are performed in a limited region that includes these two monomers plus a portion of their environment. Then another monomer is added after the other until a complete chain has been built. Then a new chain-starting monomer is added and placed randomly in a free region and the process is repeated. This procedure continues until the target number of chains

has been created, which in our case was $N_c = 500$ chains. Then the whole box is finally relaxed with a Metropolis Monte Carlo at the desired temperature for a few million steps.

This step-wise procedure has a few interesting advantages. One of them is that the chains can be built and pre-relaxed using the actual interaction potentials that will be used in the MD simulations, which means that excluded-volume effects are already taken into account at the building phase. Moreover, the Monte Carlo relaxation is carried out with a Metropolis algorithm, i.e., there's a temperature parameter and the possible conformations are sampled from a Boltzmann distribution. This favors building up a topology that is already reasonably close to that of an equilibrated system. One drawback of the method is that, as the building procedure progresses and the system density increases, it becomes more and more difficult to find a place for a new bead. For the FG systems used here this did not result in prohibitive long Monte Carlo simulations, but this could very well be the case for realistic polymer melts of high molecular weight. An example input file for the EMC package can be found in Appendix B as it can be useful to the reader.

The bond-growth MC initialization was carried out independently from scratch for all 19 state points. After initialized, each of these FG systems were subject to geometry minimization with the Polak-Ribière variant of the conjugate-gradient algorithm [136, 137], with energy and force tolerances of 10^{-1} and 10^{-6} reduced units, respectively, and a maximum number of minimization steps equal to 10^6 . The minimized systems were then thermalized using MD simulations in the canonical ensemble using a Langevin thermostat for 10^7 steps until the target temperature for each system was reached, followed by making the aggregate total linear and angular momentum of the system equal to zero. Then an equilibration phase followed, now using the Nosé-Hoover thermostat with a time step length of $\Delta t = 0.001$ for 50×10^6 steps for system. This resulted in 19 well equilibrated simulation boxes of the FG system, each with $N_c = 500$ chains at a particular state point defined by a density and temperature pair.

3.4. Coarse-Graining Procedure

The FG LJC system was CG in two lower-resolution models referred herein as CG2 and CG3 and using 2:1 and 3:1 mapping schemes, respectively. In order to avoid transferability issues due to state point dependency in the resulting CG potentials, in this project we followed the pragmatic but laborious route of independently deriving CG potentials from scratch for the two CG resolutions for each of the 19

selected thermodynamic state points. The derivation was performed separately for bonded and non-bonded potentials, as described in remaining of this section. One requirement that is common for both bonded and non-bonded potentials is the availability of a reference FG trajectory mapped to the CG representations, referred herein as CG-mapped trajectories. These are obtained by taking the trajectory generated by the FG production simulations and then applying the CG mapping scheme to them. The resulting reference CG-mapped trajectories convey then the correct FG dynamics but in the CG representation and can be used to calculate the target probability distributions that the CG potentials should ultimately reproduce.

As described in more details in Section 3.1, the LJC used in this project have no angle bending or dihedral interactions, only bond stretching and non-bonded. The bonded CG potentials are derived using DBI, while the non-bonded CG potentials are derived using iterative version of the method. The resulting potentials are numerical rather than analytical and must be post-processed in order to be useful in MD simulations. This processing includes fixing poorly sampled regions using interpolation or extrapolation schemes, especially at the edges of the domain, fitting the potential to splines in order to calculate its derivative needed for the force, re-sampling at a finer grid, etc. The derivation of the CG potentials was performed with the VOTCA toolkit [88].

3.4.1. Bonded Potentials

Deriving the bonded potentials essentially means calculating the volume-normalized bond-length distributions, P_r , of the reference CG-mapped trajectories and then Boltzmann-inverting these distributions to obtain the PMF that will be used as bonded potentials. However, in order to truly capture the bond behavior and exclude contributions introduced by the non-bonded potentials, the reference CG-mapped trajectory used to obtain P_r were not the one obtained from the production run of the actual complete reference FG system. Instead, a modified version of the FG system was simulated in which the non-bonded LJ interactions were switched off, i.e., an IG of bead-spring chains. This modified FG system was created by taking a single frame of the equilibrated trajectory of the complete FG system and equilibrating it at the same temperature but with LJ interactions removed. Once equilibrium was reached, a production run was performed for 20×10^6 steps and the CG mapping schemes were applied to the resulting trajectory to obtain the reference CG-mapped trajectories.

The bonded CG potentials were obtained by then calculating the bond-length distributions of these

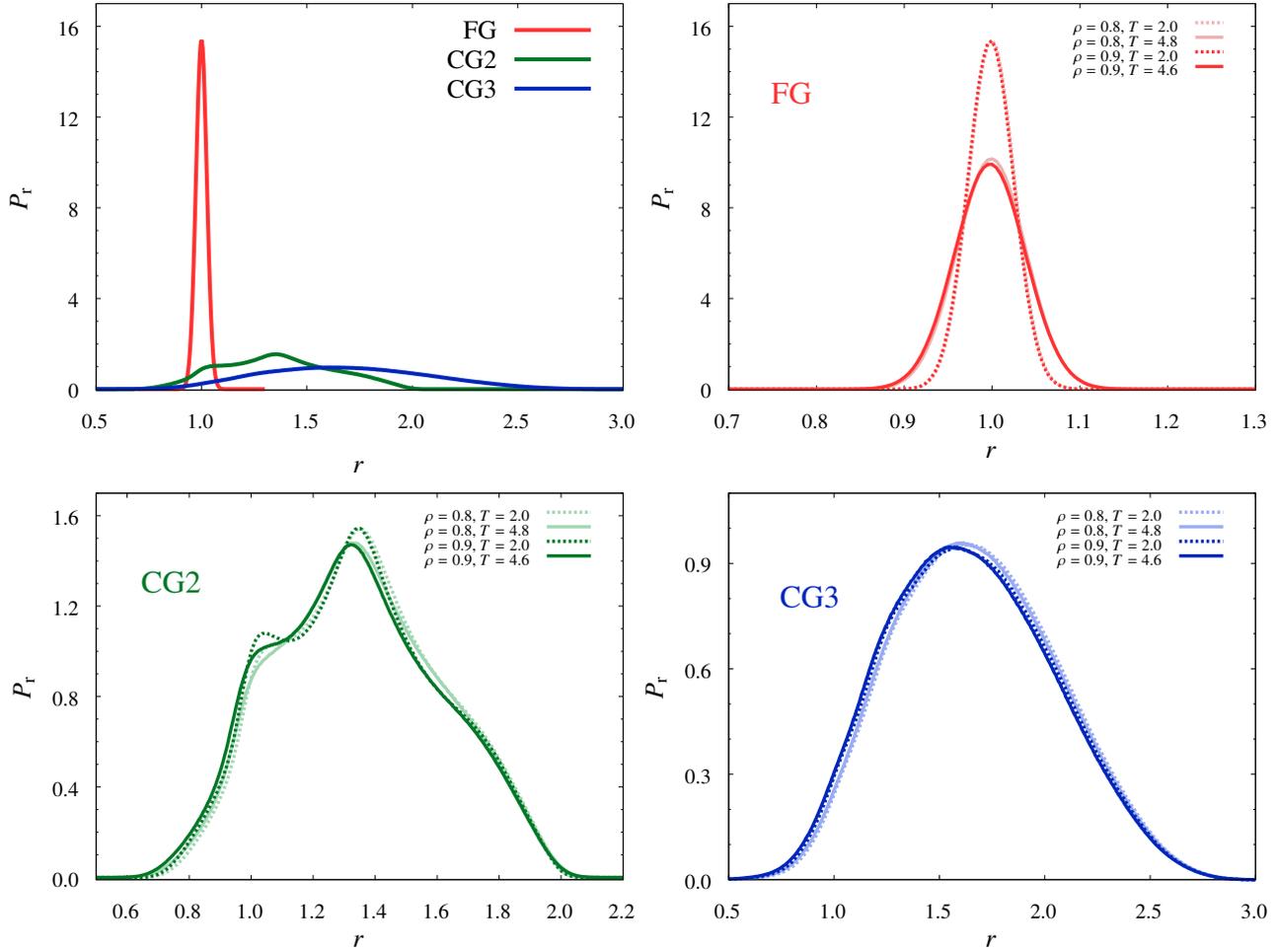


Figure 3.3.: Bond length distributions for the FG and CG systems. A comparison among the three resolutions is shown for $\rho = 0.8$ and $T = 2.0$ on the top left panel. The top right and bottom panels show how the distributions change for different state points for each resolution, in which lighter shades correspond to the lower density and the dotted lines to the lower temperatures.

reference CG-mapped trajectories for CG1 and CG2, for all temperatures and densities, as shown in Figure 3.3 for selected state points along with the corresponding FG distributions for comparison. It can be seen that the bonds in the CG systems are significantly softer than in the FG resolution and can stretch over a wider range of lengths, e.g., while for FG the bonds are relatively stiff and their stretch deviates little the equilibrium length which is 1, for CG3 there are bonds that stretch well above twice of that. For the state points shown in Figure 3.3 the equilibrium bond length for FG is reflected in the distribution centered around 1, while for CG2 and CG3 it is approximately 1.4 and 1.6, respectively. In Figure 3.3 it is also shown how the distributions vary for the different state points. For FG it can be seen that there is little

difference between densities $\rho = 0.8$ and $\rho = 0.9$, but a larger difference for temperatures, with the higher temperatures favoring a wide spread of the distribution and consequently a shorter height. For the CG2 distributions it does not seem that varying density and temperature have a clear systematic effect, while in the case of CG3 the distributions the higher density are slightly shifted towards shorter bond lengths.

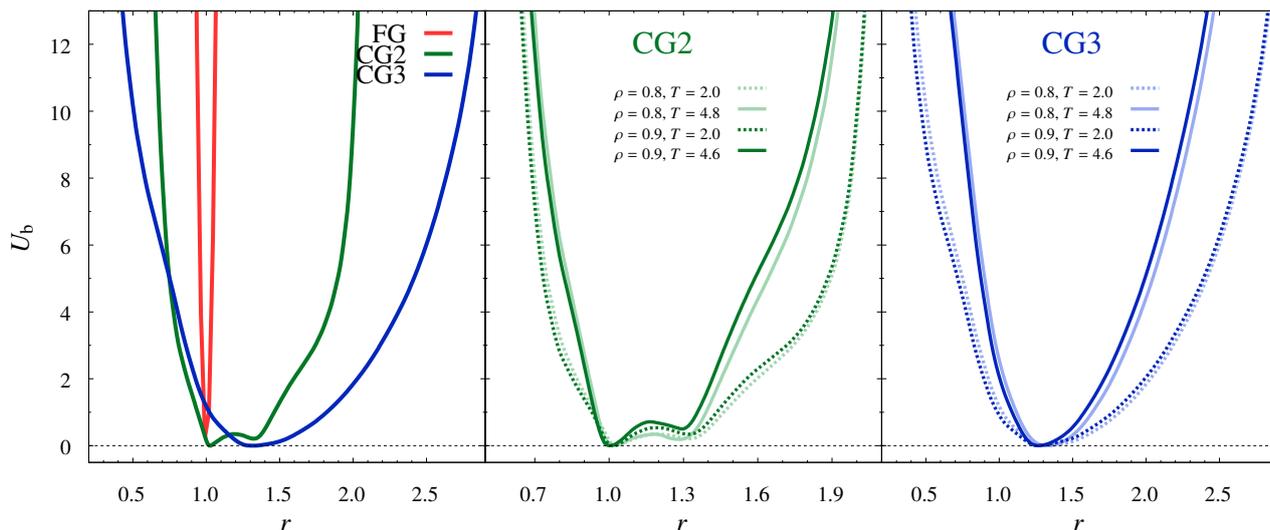


Figure 3.4.: Bonded potentials for the FG and CG systems. A comparison among the three resolutions is shown for $\rho = 0.8$ and $T = 2.0$ on left panel. The middle and right panels show how the potentials change for different state points for each CG resolution, in which lighter shades correspond to the lower density and the dotted lines to the lower temperatures.

The DBI as stated in Equation (2.63) is applied to the bond-length distributions for the two CG resolutions using the VOTCA toolkit, which already takes care of the correct histogram normalization and as a result the raw bonded potentials in numerical form are obtained. These potentials then need to go through some processing before being ready to use with LAMMPS, which includes making sure that the potential is energetically inaccessible where $P_r = 0$ (i.e., at the edges of the distributions), fitting over any discontinuities that could have resulted from poorly sampled regions, and interpolating with cubic splines in order to achieve a satisfactory grid density so that the tabulated points are not too far apart from each other. The resulting bonded CG potentials for a few selected state points can be seen in Figure 3.4.

3.4.2. Non-bonded Potentials

The non-bonded CG potentials were derived using IBI, which required first obtaining the target structure from the reference FG systems. This target are the RDFs calculated from the reference CG-mapped trajectories obtained from the FG production runs, and are shown for a few state points in Figure 3.5. One should note that these RDFs were calculated excluding the first bonded neighbor of each monomer, i.e., only monomers interacting via the non-bonded potential were included. As it can be seen, the state point dependence is significantly more pronounced for the FG system, with higher temperatures leading to lower peaks for the same respective densities. The dependence on the thermodynamic conditions seem to become weaker as the coarse-graining resolution is increased.

Even though there seems to be only a small influence of the thermodynamic state points on the RDFs, these are amplified in the final non-bonded potentials, as shown in Figure 3.6, where the pressure-corrected potentials obtained with IBI are shown. Correct pressures are important not only because they ensure that the same state points are simulated in the CG systems, but also to have reliable viscosity calculations, as in the GK formalism the viscosity is calculated from the pressure tensor (cf. Section 3.5). The non-bonded CG potentials obtained with this procedure were mostly repulsive after the pressure correction, as illustrated in Figure 3.6 for a specific state point. Even though this is the case, they reproduce the structure of the CG-mapped FG trajectories. This repulsive character may be interpreted as a dominant excluded volume contribution [138]. For each state point and each CG resolution, the canonical MD simulations performed in each IBI iteration consisted of 2×10^6 equilibration steps and 8×10^6 sampling steps. The convergence for the IBI procedure is relatively fast, typically under 100 IBI steps, as illustrated in Figure 3.7.

3.5. Dynamical Properties

The production MD runs to generate data for the analysis of the dynamical descriptors were performed for $t_L = 400 \times 10^3$ time units for each state point at each resolution, i.e., the complete production trajectories were $\tau_L = 400 \times 10^6$ time steps long, given the time step length $\Delta t = 0.001$. The independent components of the pressure tensor as well as thermodynamic information such as pressure, temperature and energies were saved at every time step, and the positions and velocities were saved at every 10^3 steps.

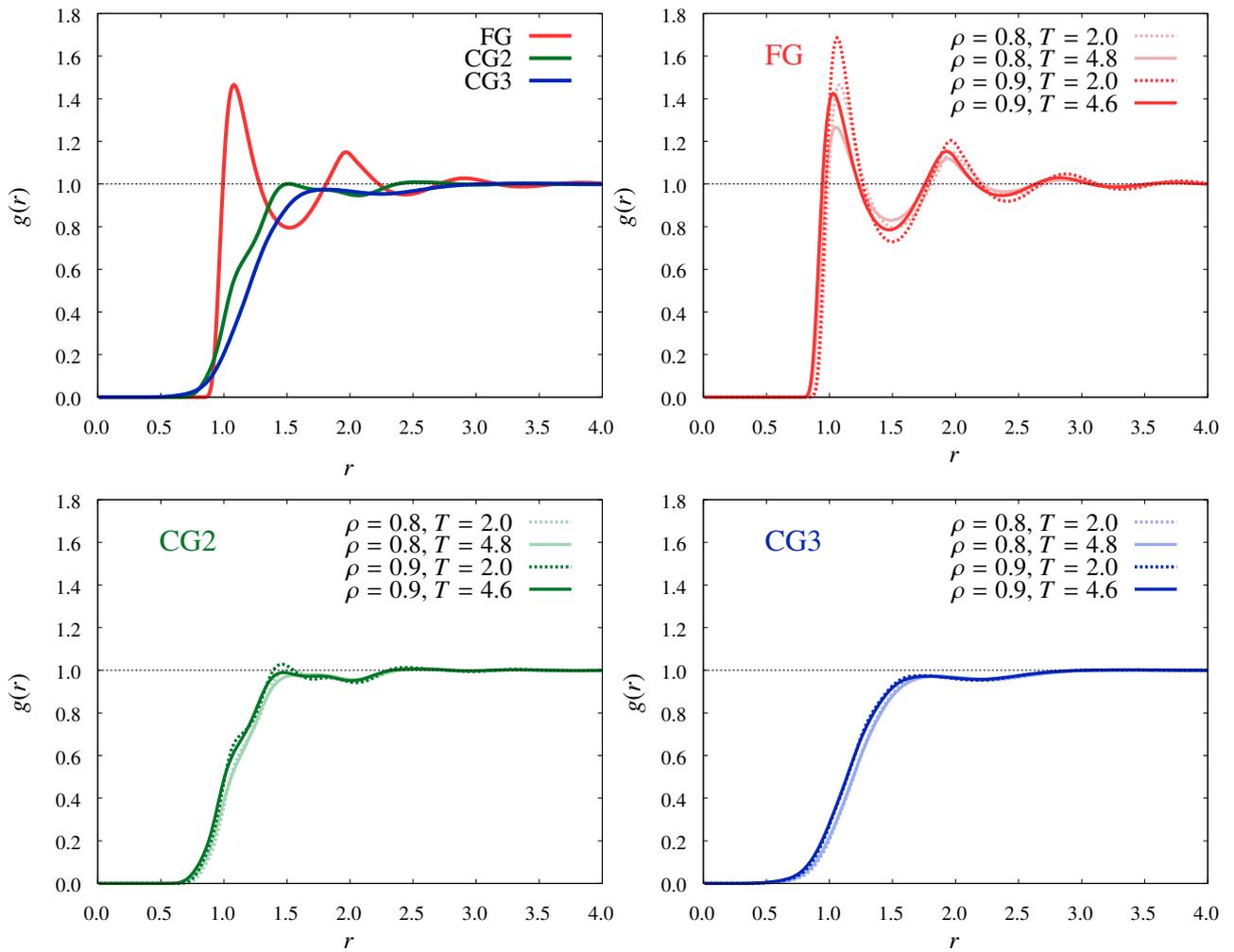


Figure 3.5.: RDF for the FG and CG systems. A comparison among the three resolutions is shown for $\rho = 0.8$ and $T = 2.0$ on the top left panel. The top right and bottom panels show how the distributions change for different state points for each resolution, in which lighter shades correspond to the lower density and the dotted lines to the lower temperatures.

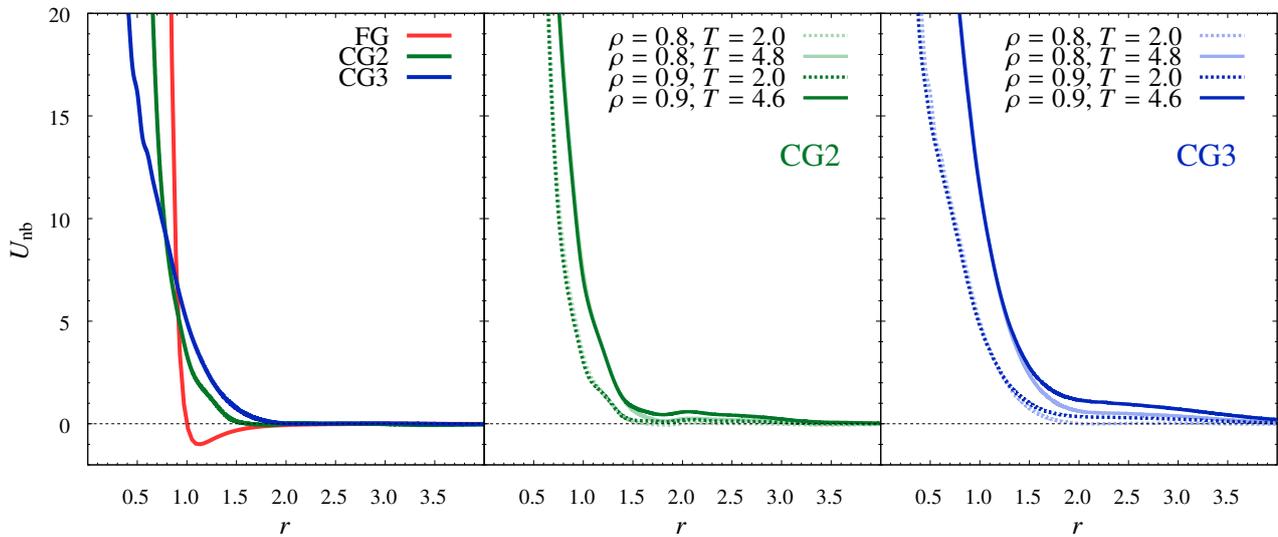


Figure 3.6.: Pressure-corrected non-bonded potentials for the FG and CG systems. A comparison among the three resolutions is shown for $\rho = 0.8$ and $T = 2.0$ on left panel. The middle and right panels show how the potentials change for different state points for each CG resolution, in which lighter shades correspond to the lower density and the dotted lines to the lower temperatures.

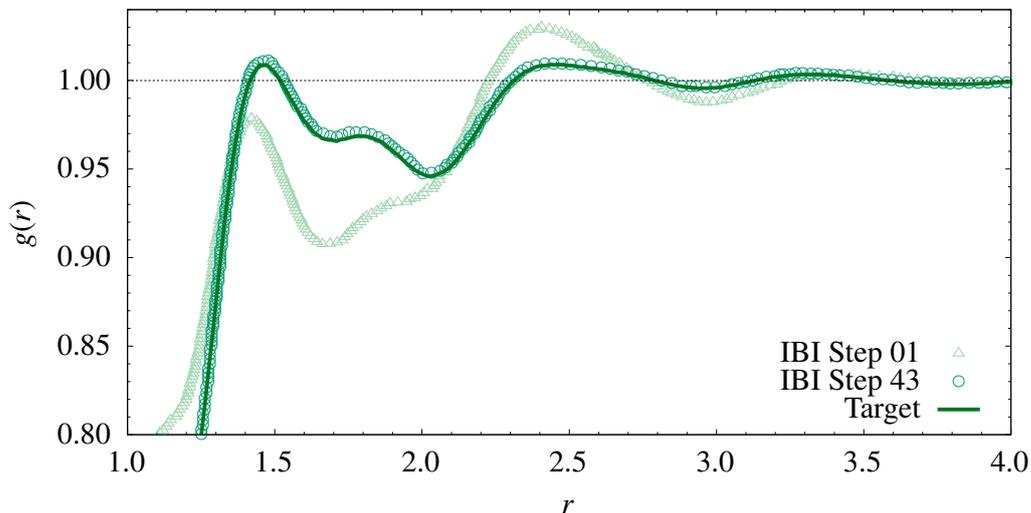


Figure 3.7.: Convergence of the IBI method for the CG2 system at $\rho = 0.9$ and $T = 3.0$. Target RDF (continuous line) is shown, as well as those obtained in step 1 (hollow triangles) of the iterative procedure and step 43 (hollow circles). Region purposely zoomed in to highlight the features of the curves.

3.5.1. Self-Diffusion Coefficient

The basic equation to calculate a self-diffusion coefficient in this work is given by Einstein relation in Equation (2.46), which requires the mean-square displacement (MSD),

$$\text{MSD}(t) = \frac{1}{N} \sum_{i=1}^N |\mathbf{r}_i(t) - \mathbf{r}_i(0)|^2 \quad (3.99)$$

where $\mathbf{r}_i(t)$ is the Cartesian coordinate of particle i at a time t . As seen in the Einstein relation, the self-diffusion coefficient is proportional to the slope of the ensemble-averaged MSD curve in the long time regime. Thus it is practical to take a linear function $f(t) = at + c$ and fit it to the long-time region of the curve $\langle \text{MSD}(t) \rangle$, and since $c \approx 0$ in this long-time region the self-diffusion may be obtained from the slope coefficient of the fitted function, i.e., $D = a/6$, where the denominator is due to Equation (2.46). Errors were calculated using the method from Chitra and Yashonath [139] which is based on computing block averages and their respective variance.

When calculating MSDs one must decide what object exactly should be tracked in order to assess its displacement. For our systems there are two seemingly obvious possibilities, namely individual monomers, or the center of mass of each chain. If only short times are considered then it is expected that the diffusivity will be different for both cases as the individual monomers will jiggle constantly but the chain as whole will not move about very dramatically in a short time period due to its size. However, in the long-time diffusion regime the behavior of both cases should converge to the same diffusion rate. This was exactly what is observed in our calculations as it can be seen in Figure 3.8 where, for all resolutions, the MSD of the monomers and the chains converge for longer times.

The computation of the ensemble average of the MSD, $\langle \text{MSD}(t) \rangle$, was performed using the sliding window method [65] described in Section 2.1.5. In principle one may use the whole trajectory of with duration $\tau_L = 400 \times 10^6$ time steps to calculate $\langle \text{MSD}(t) \rangle$, however the long-time diffusion regime is reached in a much shorter time than $\mathcal{O}(10^8)$ time steps for our systems. As it can be seen in Figure 3.8 a trajectory of $\mathcal{O}(10^6)$ time steps (i.e., $\mathcal{O}(10^4)$ time units) is typically already enough for the FG systems that has comparatively slower dynamics, and for the CG systems with faster dynamics it is enough to use a trajectory of $\mathcal{O}(10^5)$ time steps. Moreover, if the MSD is directly calculated from the whole trajectory then the number of observations contributing to each point will not be the equal for all points, i.e., values of $\text{MSD}(t)$ at the

end of the trajectory will be averaged from a relatively small number of observations since there is not many points left in the trajectory. For these reasons the sliding window approach was employed.

Additionally, the sliding window approach was combined with a block strategy and not applied directly to the complete trajectory of $\tau_L = 400 \times 10^6$ time steps. This trajectory was divided in $N_b = 20$ blocks of equal length, namely $\tau_b = 20 \times 10^6$ time steps. The sliding window method was then applied to each block with a spacing of $\tau_s = 2000$ time steps and a window length of $\tau_w = 10 \times 10^6$ time steps, thus the number of origins per block was $n_0 \approx 5000$ according to Equation (2.40), where the block length, τ_b , was used

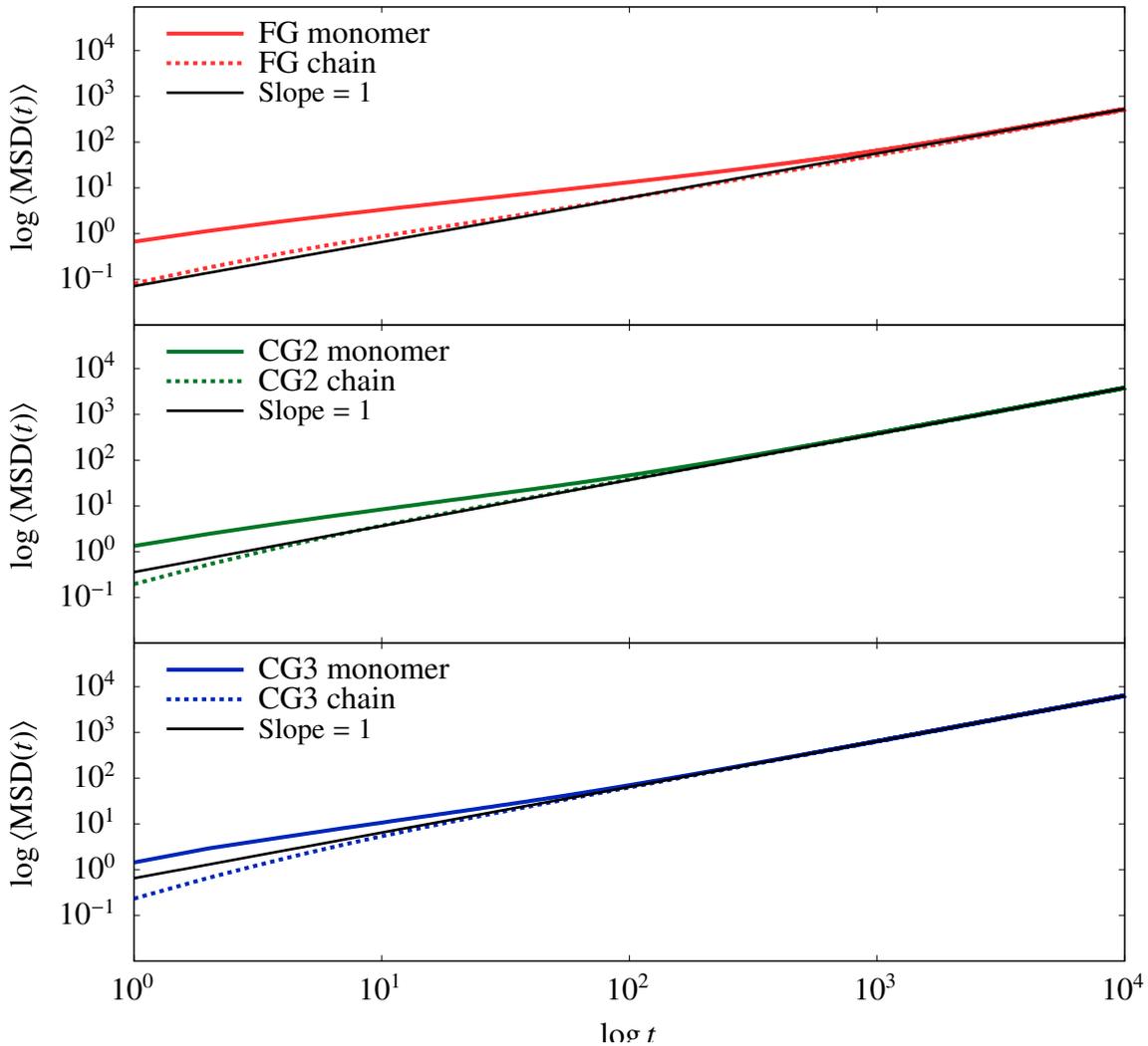


Figure 3.8.: Mean-square displacement of the monomers (solid curves) and center-of-mass of the chains (dotted curves) for the three resolutions FG (top panel), CG2 (middle panel), and CG3 (bottom panel) at $T = 2.4$ and $\rho = 0.8$. Solid black line indicates unity slope in the logarithm scale.

instead of the total trajectory length, τ_L . This procedure resulted in one estimate of $\langle \text{MSD}(t) \rangle_k$, and thus D_k , for each block k . The final self-diffusion coefficient was then simply averaged over the block estimates, $D = \frac{1}{N_b} \sum_{k=1}^{N_b} D_k$, with the error calculated from the standard deviation of the block averages.

3.5.2. Viscosity

The viscosity for all all systems was calculated using the GK approach as given by Equations (2.49) and (2.50). In Figure 3.9 the running integrals are shown for a several independent simulations of the FG system at a particular state point. Even though each calculation employed multiple, well-spaced origins, the running curves are noisy and identifying a finite plateau region where each curve may be regarded as converged is not an easy task. The average of all curves is much smoother due to the good quality of the results and even though it is easier to identify a plateau in this case, it is still difficult to determine the viscosity in an unambiguous and systematic way, e.g., if a cutoff is to be used, what should it be?

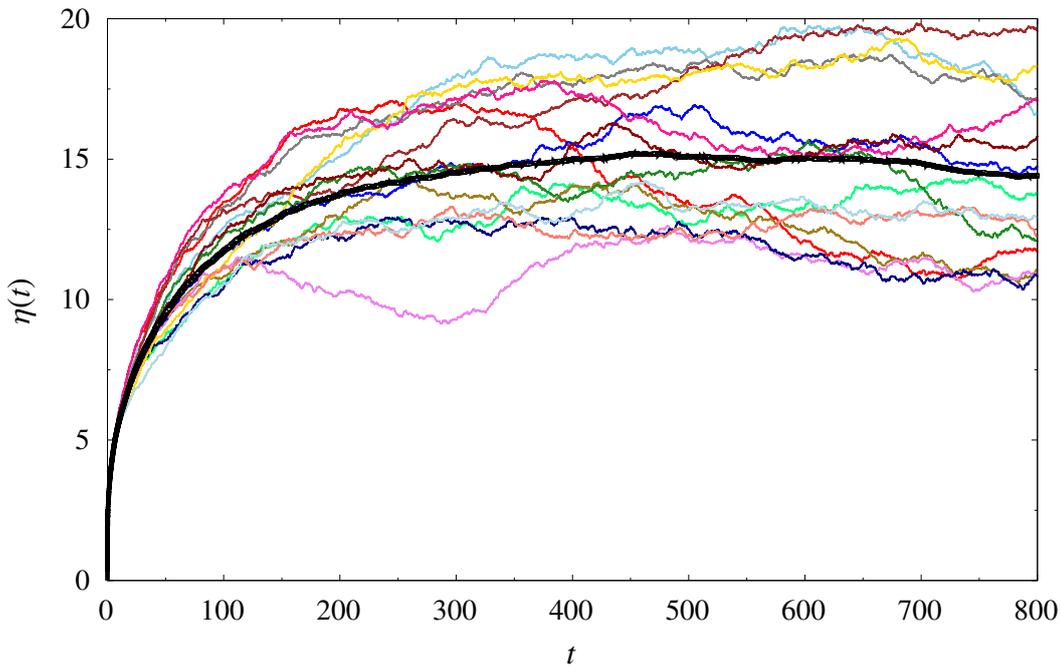


Figure 3.9.: Viscosity calculations using the GK equilibrium method for the FG system at $T = 2.0$ and $\rho = 0.8$. The colorful noisy curves correspond to running GK integrals obtained from 20 MD trajectories with duration 20×10^3 time units each, for which the viscosities were calculated over a sliding window of duration 800 time units with multiple origins. The bold black curve corresponds to the average of all curves.

In order to calculate viscosities in a systematic and reliable manner, we have employed in this work the time decomposition method from Zhang et al. [73], which prescribes a reproducible and unambiguous algorithm that can be easily applied to our MD workflow. The method has been marginally adapted to our data set and the steps are outlined below.

Step 1 For each system at each state point, split the long MD trajectory with total duration τ_c time units in N_b blocks of τ_b time units each.

Step 2 For each block, calculate the viscosity using the GK approach from Equations (2.49) and (2.50) in a sliding window of duration τ_w time units and using multiple origins spaced in τ_s time units from each other. This step yields N_b running integrals $\eta_i(t)$, with $i = 1 \dots N_b$.

Step 3 Calculate the average of the N_b running integrals,

$$\bar{\eta}(t) = \frac{1}{N_b} \sum_{i=1}^{N_b} \eta_i(t), \quad (3.100)$$

and the respective time-dependent standard deviation,

$$\sigma(t) = \sqrt{\frac{1}{1 - N_b} \sum_{i=1}^{N_b} (\eta_i(t) - \bar{\eta}(t))^2}. \quad (3.101)$$

Step 4 Fit the standard deviation to a power-law function,

$$\sigma(t) \approx g(t) = Dt^b. \quad (3.102)$$

where D and b are fit parameters.

Step 5 Perform a weighted fit of the average running integral $\bar{\eta}(t)$ to a double exponential function,

$$\bar{\eta}(t) \approx h(t) = A\alpha\tau_1(1 - \exp -t/\tau_1) + A(1 - \alpha)\tau_2(1 - \exp -t/\tau_2), \quad (3.103)$$

where A , α , τ_1 , and τ_2 are fit parameters, and using a fitting weight of $1/t^b$, where b is the fit parameter from Step 4. The fit to the exponential function should be performed only in the time

range $[0, \tau_{\text{cut}}]$, where the cutoff time τ_{cut} is the time at which the standard deviation $\sigma(t)$ is 20% of $\bar{\eta}(t)$. See Figure 3.10.

Step 6 The viscosity is obtained by taking the long time limit of the fitted double exponential function from Equation (3.103).

Step 7 The error estimate of the final viscosity, ϵ_η , is obtained from the standard deviation of the block average $\bar{\eta}$ at the end of the sliding window for each block, i.e., $\epsilon_\eta = \sigma(\tau_w)/\sqrt{N_b}$, where the standard deviation σ comes from Equation (3.101) in Item Step 3.

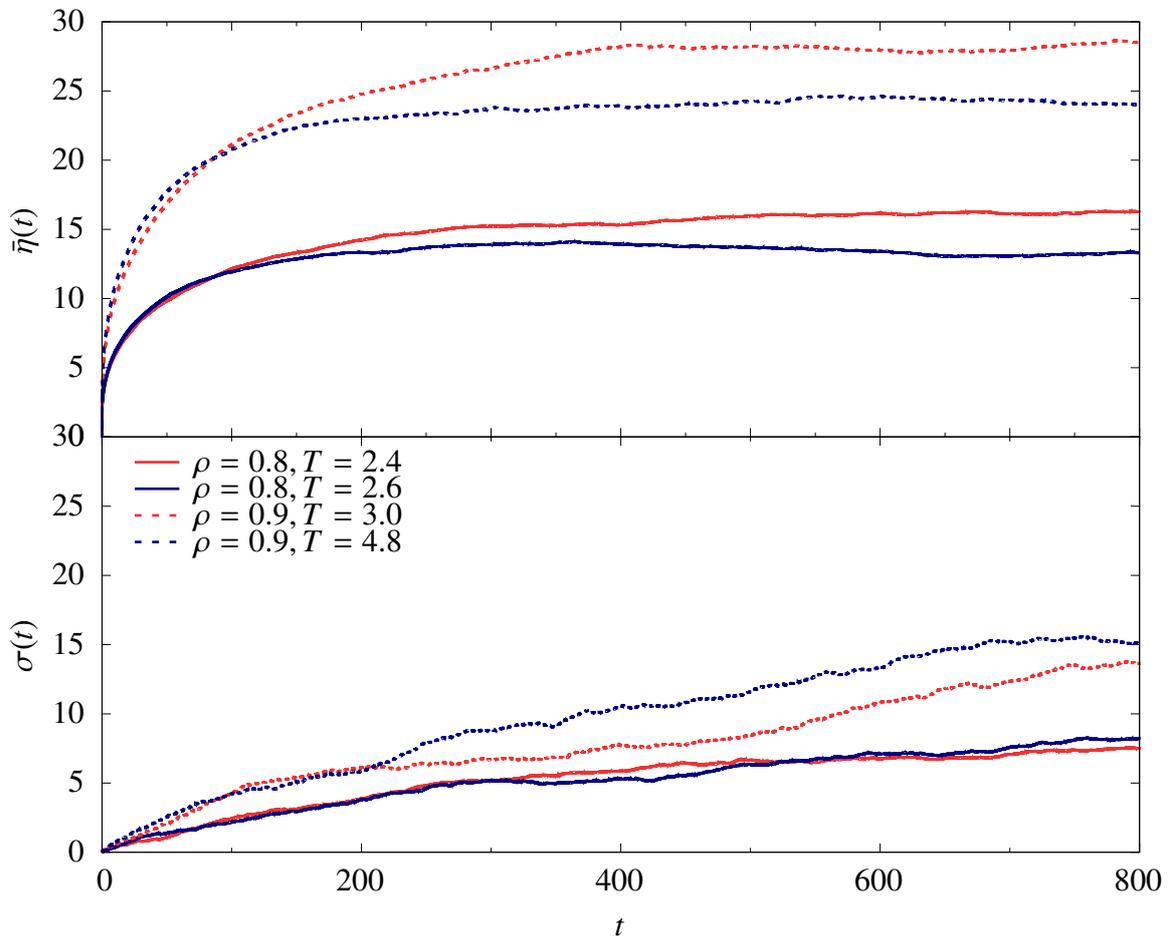


Figure 3.10.: Viscosity for the FG system at selected state points calculated using the time decomposition method. The running average viscosity $\bar{\eta}(t)$ (top panel) were calculated from 20 blocks of duration 20×10^3 time units each within a sliding window with duration 800 time units. The corresponding running standard deviation $\sigma(t)$ is also shown (bottom panel) as a function of time.

The total duration of the complete trajectory noted in Item Step 1 is $\tau_c = 400 \times 10^3$ time units and it was split in $N_b = 20$ blocks, each of length $\tau_b = 20 \times 10^3$ time units. The length of the sliding window over which the viscosity was calculated for each block noted in Item Step 2 is $\tau_w = 800$ time units. The spacing between the origins of the sliding window was determined by performing a statistical inefficiency analysis [48] with respect to the pressure obtained from the MD run. The statistical inefficiency provides an estimate how long one must wait in a MD simulation until uncorrelated measures can be made. In order to calculate it we first break the trajectory up in N_b blocks of τ_b length in a similar fashion to how the viscosity was calculated. A pressure average is then calculated for each block i , i.e.,

$$\bar{p}_i = \frac{1}{n_b} \sum_{j=1+(i-1)n_b}^{in_b} p_j, \quad (3.104)$$

where $n_b = \tau_b/\Delta t$ is the number of steps in a block of length τ_b , and a pressure average is also calculated for the complete trajectory,

$$\bar{p} = \frac{1}{n_c} \sum_1^{n_c} p_j, \quad (3.105)$$

where $n_c = \tau_c/\Delta t$ is the total number of steps in the complete trajectory. The statistical inefficiency for the pressure is then calculated as

$$s = \lim_{\tau_b \rightarrow \infty} \frac{\tau_b \sigma^2(\bar{p}_b)}{\sigma^2(\bar{p})}, \quad (3.106)$$

where $\sigma^2(\bar{p}_b)$ is the variance of the block average from Equation (3.104) and $\sigma^2(\bar{p})$ is the variance of the average from Equation (3.105). The statistical inefficiency is the limiting ratio of the variance of an average and the variance expected when uncorrelated Gaussian statistics is assumed. In practice, the block average is calculated for blocks of increasing lengths and then s is plotted against $\tau_b^{1/2}$ until a plateau can be identified, as shown in Figure 3.11 where the statistical inefficiency of one component of the pressure tensor was calculated for block lengths increasing from 0 to 10^3 time units in steps of 0.2 time units. The plateau was identified at $s = 0.153$, which means that only one measurement of the pressure tensor in about 0.15 time units contribute to new information. This calculation was repeated for all independent components of the pressure tensor and the largest value of s among them was used as the spacing between the time origins, τ_s , in Item Step 2 of the time decomposition algorithm.

In order to validate the viscosity calculation, the results from the GK approach described so far were

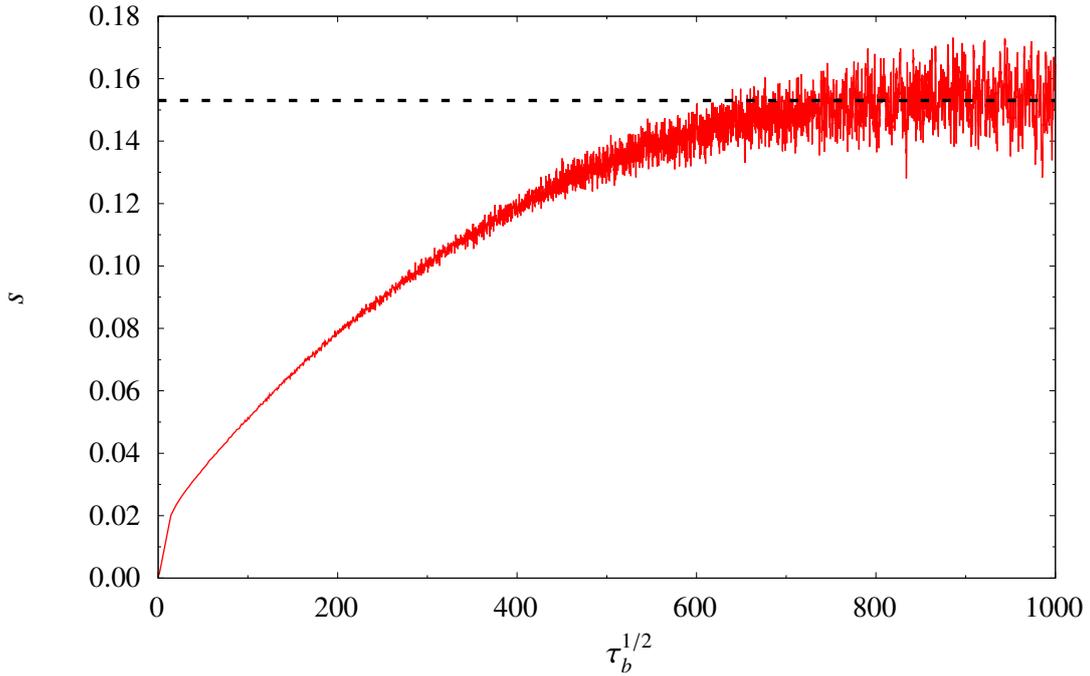


Figure 3.11.: Statistical inefficiency (s) for the FG system at $T = 2.0$ and $\rho = 0.8$ (red continuous curve) as a function of the square root of the block length, $\tau_b^{1/2}$, calculated according to Equation (3.106) from the xy component of the pressure tensor. The dotted black line indicates the plateau reached at $s = 0.153$.

compared with results from the RNEMD method of Müller-Plathe [74, 75] as implemented in LAMMPS [130]. The RNEMD simulation setup is as follows: the simulation box is divided in 20 slabs in the z direction and every 400 steps a pair of particles is selected and the v_x component of their velocities is exchanged. The pair of atoms is picked such that one of them is that with the most positive v_x velocity component in slab 1 and the other is that with the most negative v_x velocity component in slab 11. These velocities swaps induces a momentum flux in the z direction and consequently a shear profile from which the viscosity is calculated. The RNEMD simulations were realized within the canonical ensemble and with a time step of $\Delta t = 0.001$, as usual. The simulations were allowed to equilibrate for 2×10^3 time units until the observed velocity profile was linear and smooth, followed by production runs with duration 10×10^3 time units. The comparison of the viscosities obtained with GK and RNEMD are shown in Figure 3.12, where it can be seen that both techniques yield similar results, with RNEMD having smaller error bars.

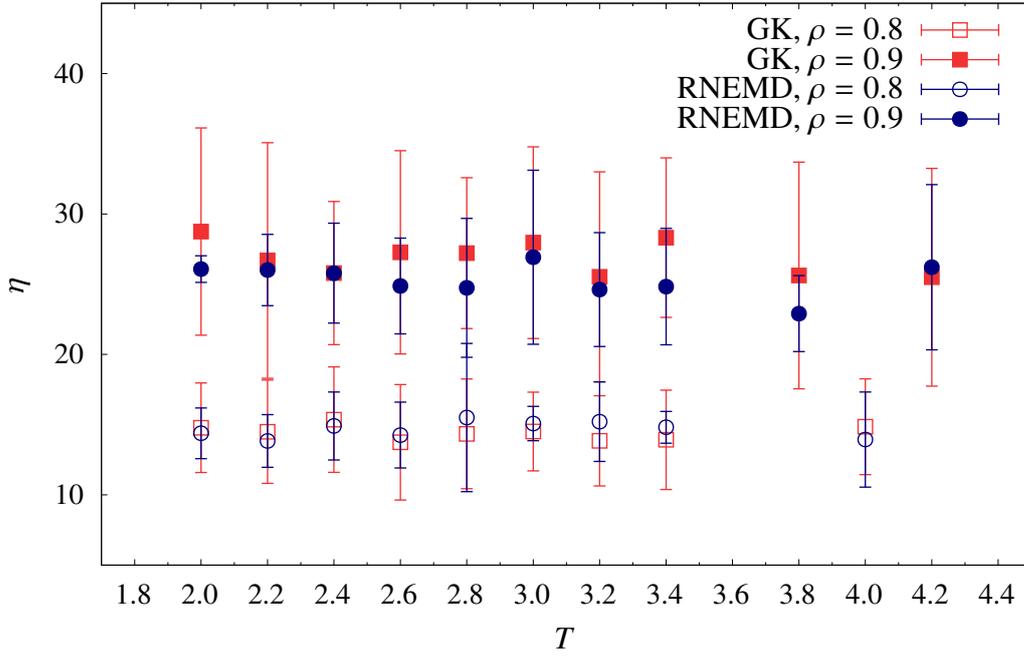


Figure 3.12.: Comparison of the viscosities *versus* temperature obtained with the GK and with the RNEMD for the FG systems at all thermodynamic state points considered.

3.5.3. Bond Reorientation Time

The bond reorientation ACF from Equation (2.51) is calculated based on direct averages of properties of all monomers, and thus differently from, e.g., the pressure tensor ACF, this function decays to zero fairly rapidly (cf. Figure 3.13). Given the rapid decay and the aim to minimize numerical errors, the Gauss-Newton algorithm [137, 140] was used to perform a non-linear least-square fit of the bond reorientation ACF to a modified Kohlrausch-Williams-Watts (mKWW) function [78, 141, 142],

$$\Phi_b(t) \approx f(t) = a \exp(-t/b) + (1 - a) \exp\left(-[t/c]^d\right), \quad (3.107)$$

for which it is straightforward to evaluate the time integral analytically, yielding a direct expression for the bond reorientation time,

$$\vartheta = \int_0^\infty f(t) dt = ab + (1 - a)c \frac{\Gamma(d^{-1})}{d}. \quad (3.108)$$

where a , b , c , and d are coefficients from the fit to the mKWW function in Equation (3.107) and Γ is the gamma function. The fitting of the function to the three resolutions for a particular state point are shown

in Figure 3.13. The estimate of the bond reorientation times obtained from the analytical result above for the mKWW functions was compared with the results from a straightforward numerical integration of Equation (2.53) using the trapezoidal rule, as shown in Figure 3.14 for the FG systems. The results are in very good quantitative agreement with marginally smaller error bars for the mKWW results.

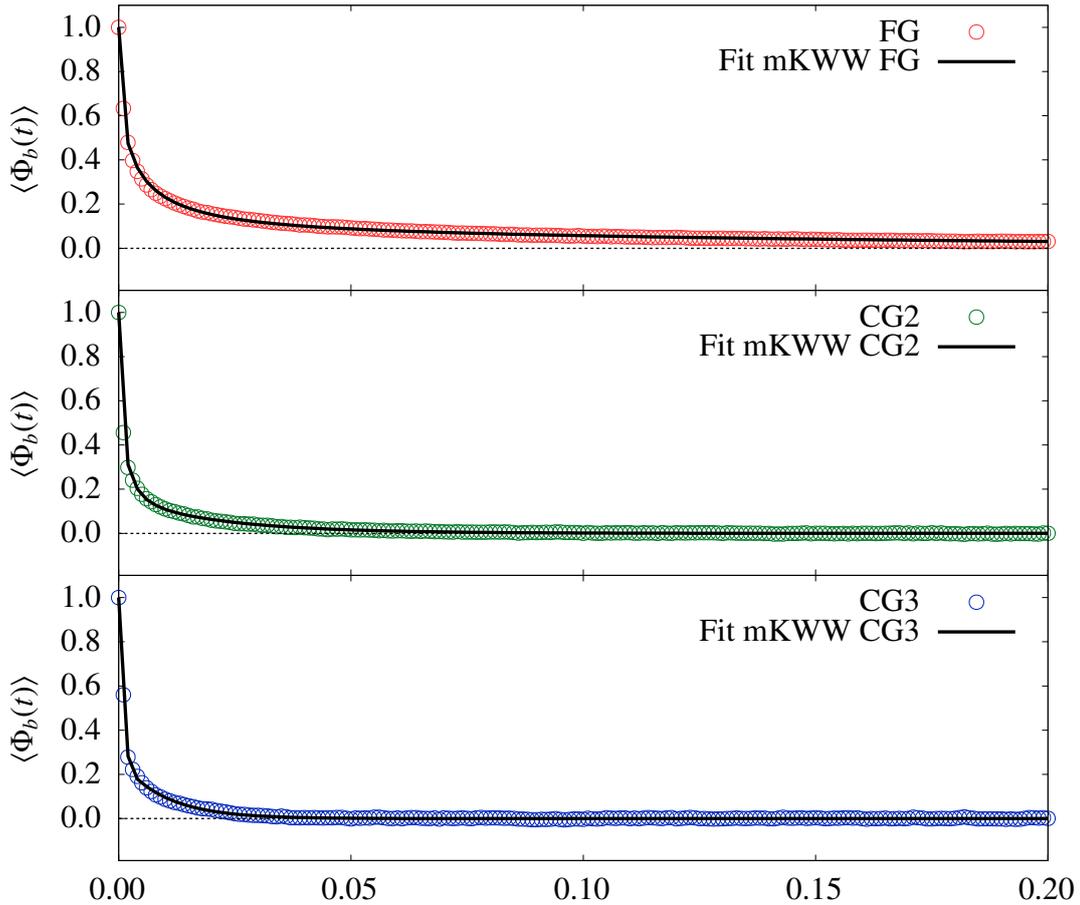


Figure 3.13.: Bond reorientation ACF (hollow circles) versus time for the FG (top panel), CG2 (middle panel), and CG3 (bottom panel) resolutions at $T = 2.4$ and $\rho = 0.8$. Solid black line indicates the modified Kohlrausch-Williams-Watts (mKWW) function fitted to the data points for each resolution.

Similarity to what was done with the MSD calculation for the self-diffusion coefficients, the bond reorientation ACF was calculated by first splitting the trajectory in blocks and then applying the sliding window approach to the blocks. In this case, $N_b = 40$ blocks were also used, with block length $\tau_b = 10 \times 10^6$ time steps each. For the sliding window approach the parameters were $\tau_s = 10^3$ time steps of spacing,

window length of $\tau_w = 2 \times 10^3$ time steps, thus $n_0 \approx 10000$ origins. The result was 40 estimates of $\langle \Phi_b(t) \rangle$ and thus of the bond reorientation time, ϑ , which were averaged and the error calculated from the standard deviation of the average. As it can be seen in Figure 3.13, at $t = 0.2$ time units the bond reorientation ACF is already very close to zero even for the FG resolution, thus a window length of 2×10^3 time steps, i.e., 2 time units, is a sufficient choice.

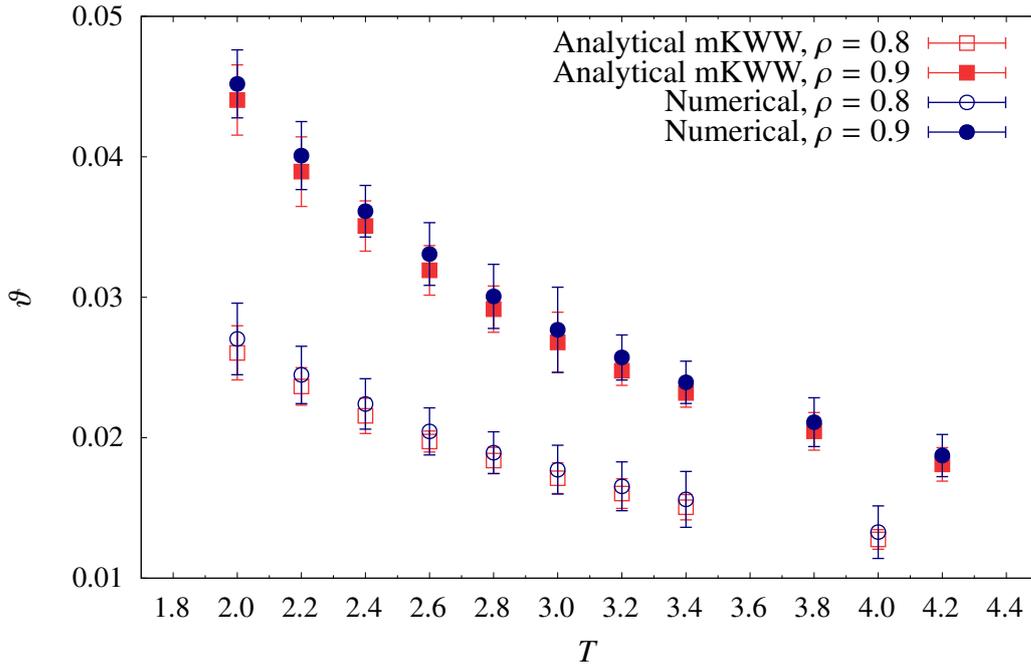


Figure 3.14.: Comparison of the bond reorientation time, ϑ as a function of temperature between results obtained from the analytical integration of the mKWW function and results from the numerical integration of the bond reorientation ACF for the FG systems at all thermodynamic state points considered.

3.6. Excess Entropy Estimation

The sections that follow describe methodological details of how the entropy calculations were performed.

3.6.1. Thermodynamic Integration

The utilization of TI in estimating excess entropies in this work as based on a multi-step methodology that was applied consistently across the three resolutions. As introduced in Section 2.5, the excess entropy of a

system is defined as $S_{\text{exc}} = S_{\text{ig}} - S$, where S is the absolute entropy of the system and S_{ig} is the entropy of an IG at the same thermodynamic conditions as the system. Our TI approach focuses on calculating the absolute entropy for all systems and then subtracting it from the corresponding IG entropy. This simplistic description convey indeed a dichotomy, after all TI is only able to calculate free energy and entropy differences, and not absolute quantities. The resolution to this dilemma is that if the absolute entropy of a similar system is somehow known, then TI can be applied using that system as one of the integration end points and the target system as the other end point. The approach used herein follows this idea and uses as a reference system the freely-jointed chain (FJC) model, in which non-interacting monomers are connected through rigid bonds to form linear chains. There are no interactions in the FJC model and the only correlations are due to the rigid bonds in each chain. The steps for obtaining the absolute entropies of the FG and CG systems can be summarized as follows and is also depicted on Figure 3.15:

- Step 1** Calculation of the absolute entropy of the FJC by the means of an exact analytical framework [143].
- Step 2** Use of TI to calculate the entropy difference between the FJC model system with non-interacting chains and rigid bonds and an FJC model system that is still non-interacting but with flexible harmonic bonds.
- Step 3** Use of TI to calculate the entropy difference between the FJC model system with non-interacting chains and flexible harmonic bonds and the actual models used in this work, i.e., FG, CG2, and CG3.
- Step 4** Determination of the absolute entropies from the entropic quantities of Items Step 1 to Step 3

Following this approach the only uncertainties involved in the calculation of absolute entropies for the models at any level of resolution are statistical uncertainties. They are related to the quality of sampling during the simulations, i.e., there is no systematic error involved. The main drawback of TI is that it demands numerous, often long simulations. However, a method that yields an exact entropy offers an important advantage: it provides a gold standard against which one can compare other entropy estimates obtained with methods or approximations that require less computational efforts.

The absolute entropy for the actual systems at all three resolutions is then decomposed into different contributions:

$$S = S_{\text{FJC}} + \Delta S_{\text{b}} + \Delta S_{\text{nb}}, \quad (3.109)$$

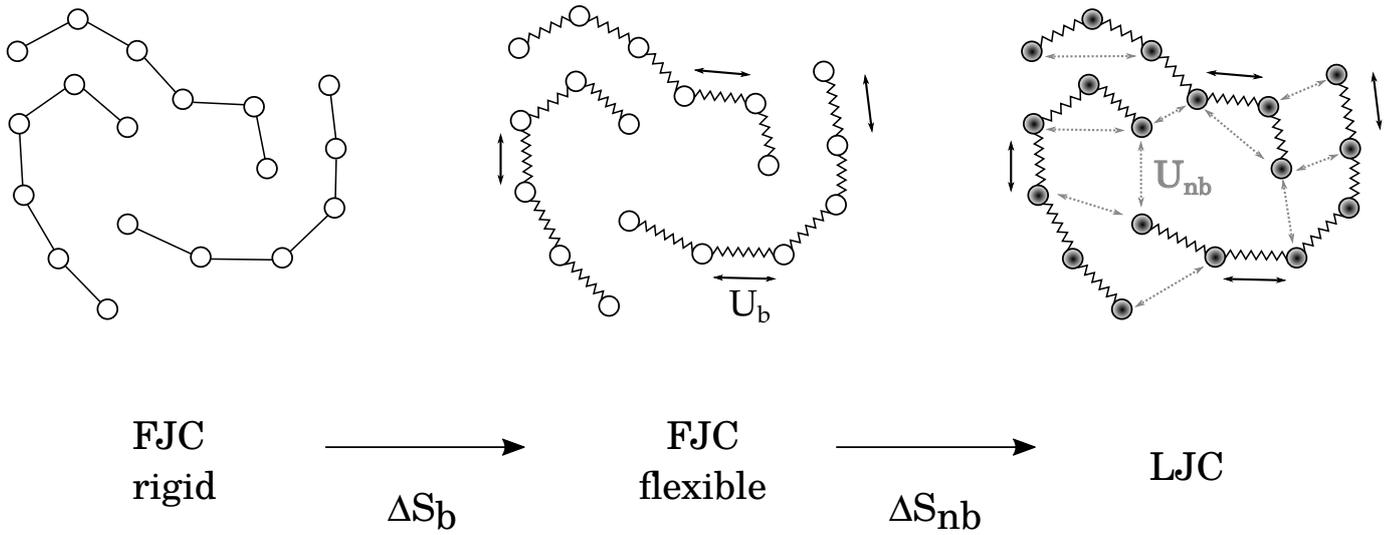


Figure 3.15.: Schematic representation of the TI process. The starting state corresponds to the non-interacting freely-jointed chains with rigid bonds (left), then flexible bonds are introduced (middle), and then the non-bonded interactions are introduced (right).

where S_{FJC} is the absolute entropy of the non-interacting FJC model with rigid bonds and same chain length of the actual systems, ΔS_b is the entropy difference between the absolute entropies of the non-interacting rigid-bond FJC model and of a non-interacting flexible-bond FJC model in which the bonds are modeled by the same bond potentials of the actual systems, and ΔS_{nb} is the entropy difference between the absolute entropies of the non-interacting flexible-bond FJC models and the actual systems, with intra- and inter-molecular non-bonded interactions (cf. Figure 3.15).

Once S is known, the excess entropy is simply obtained from subtracting it from S_{ig} , which is the absolute entropy of an equivalent monoatomic IG system at the same thermodynamic state and with same number of particles, and analytically available through the Sackur-Tetrode equation [96, 144],

$$s_{\text{ig}} = \frac{5}{2} - \ln(\rho\Lambda^3), \quad (3.110)$$

where $s_{\text{ig}} = S_{\text{ig}}/Nk_B$, the gas density is $\rho = N/V$, and Λ is the thermal de Broglie wavelength,

$$\Lambda = \sqrt{\frac{h^2}{2\pi mk_B T}} \quad (3.111)$$

where h is Planck's constant, k_B is Boltzmann constant, m is the mass of each IG particle, and T is the system temperature. Planck's constant in this work was adopted with value $h = 0.18292026$ obtained from the LJ parameters for Argon, however it could as well be set to unity, as long as it is used consistently in all calculations. This rather specific value for h was used as it was also adopted internally for LJ reduced units by the LAMMPS package [130, 131] the versions that were utilized to perform our simulations.

Absolute Entropy of the FJC

The absolute entropy of the non-interacting rigid-bond FJC, S_{FJC} in Equation (3.109), was calculated exactly from the statistical mechanical formalism by Mazars [143], who derived the canonical partition function of a gas of N_c freely-jointed chains in a three-dimensional box, and from it the Sackur-Tetrode equation equivalent for this system,

$$S_{\text{FJC}} = k_B N_c \left(N + \frac{3}{2} \right) + k_B \ln Z_{\text{FJC}} \quad (3.112)$$

where N is the number of monomers per FJC and Z_{FJC} is the canonical partition function of the FJC gas and written as

$$Z_{\text{FJC}} = \frac{1}{N_c!} \left[q Q \left(\frac{T}{T_0} \right)^N \right]^{N_c}, \quad (3.113)$$

where T is the temperature, and the expressions for q , Q and T_0 are shown in Appendix A.2.

Entropy change due to flexible bonds

The entropy difference between the non-interacting rigid-bond FJC and the non-interacting flexible-bond FJC, ΔS_b , is calculated using Equation (2.79), that is,

$$\Delta S_b = \frac{1}{T} [(\langle E \rangle_{\lambda=1} - \langle E \rangle_{\lambda=0}) - \Delta F_b], \quad (3.114)$$

where the ensemble average of the internal energy, $\langle E \rangle_\lambda$, are calculated for the simulations at the end points of the integration, and the associated free energy difference is calculated using TI,

$$\Delta F_b = \int_0^1 \left\langle \frac{\partial U_b(\lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda \quad (3.115)$$

where the thermodynamic end points of the integration $\lambda = 0$ and $\lambda = 1$ correspond to the rigid-bond and flexible-bond states, respectively. This means that the bonded potential U_b must be properly coupled with the λ -parameter in order to comply with this scenario, i.e., an integration path through which the very nature of the system is changed, and from which we can compute the resulting free energy change. Our expression for the λ -dependent bonded potential is

$$U_b(r, \lambda) = [U_{b,0}(r) + 1]^{(1-\lambda)} [U_{b,1}(r) + 1]^\lambda - 1, \quad (3.116)$$

which yields $U_b(r, \lambda = 0) = U_{b,0}(r) + 1$, where $U_{b,0}(r)$ is a harmonic potential with a stiff spring constant $k = 70,000$ modeling the rigid bonds of the FJC,[126, 143] and $U_b(r, \lambda = 1) = U_{b,1}(r) + 1$, where $U_{b,1}(r)$ is the bonded potential of the actual system, i.e., FG, CG2, or CG3. Unity was added to both terms in the square brackets of the right-hand side of Equation (3.116) to prevent either from becoming zero, which would render the logarithm that results from carrying out the partial derivative appearing in Equation (3.115) undefined. Unity was subtracted from the right-hand-side of Equation (3.116) for aesthetic purposes, i.e., so that $U_b(r_{\text{eq}}, \lambda) = 0$, where r_{eq} is defined as the equilibrium bond-length. The dependence on the power of λ used in Equation (3.116) allows a smooth, reversible transition from one state to the other, with well distributed intermediate states, as shown on the left column of Figure 3.16. In the top portion of Figure 3.17 the variation of $\delta U_b / \delta \lambda$ with changing values of λ is shown, where it can be seen that the parameter coupling strategy used here for U_b yields a smooth λ -derivative and thus very convenient for numerical integration.

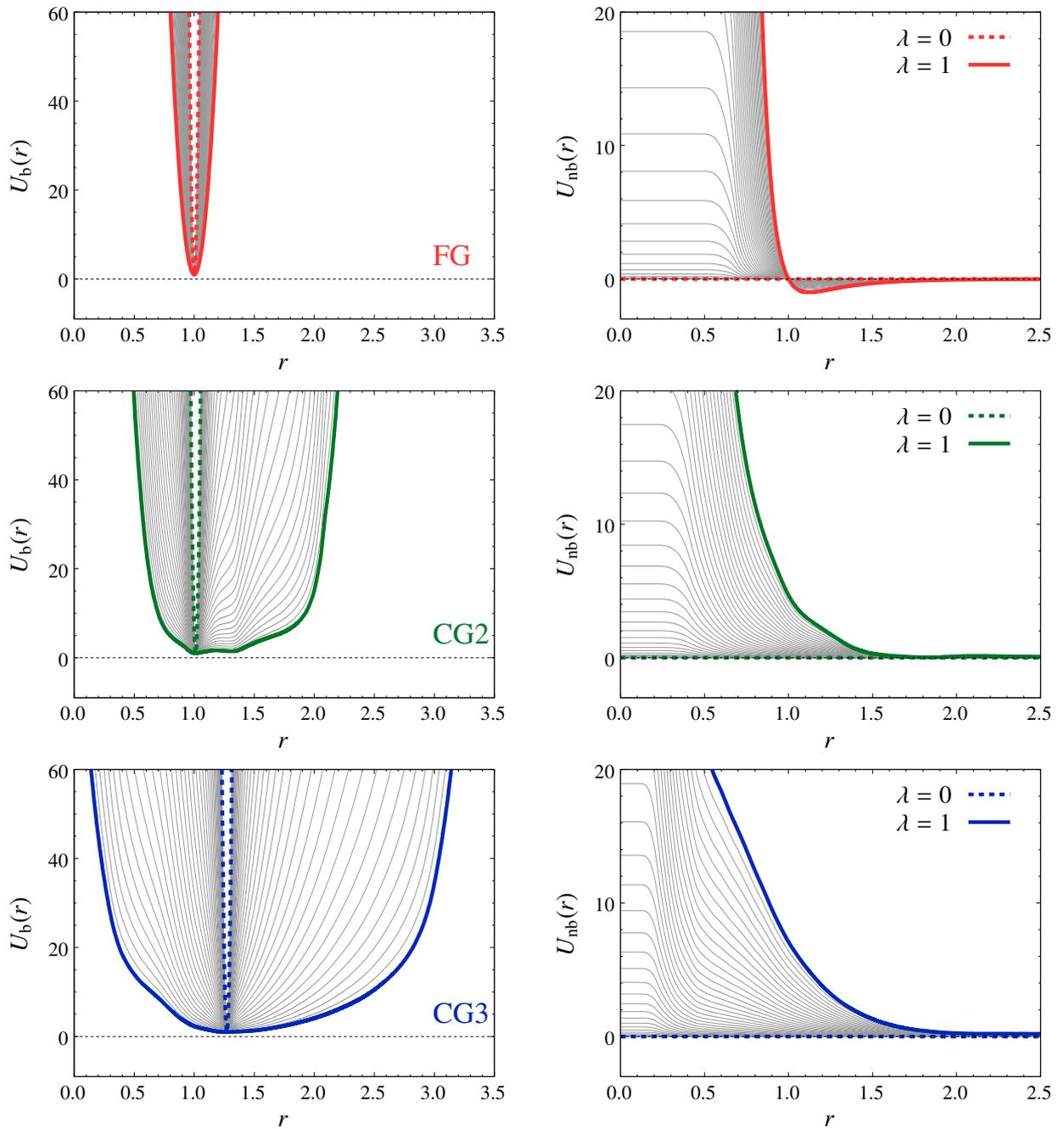


Figure 3.16.: Interaction potentials used in the TI as a function of the separation between the monomers in reduced length units for the three resolutions at $T = 2.8$ and $\rho = 0.8$, FG (top row), CG2 (middle row), and CG3 (bottom row). Dependence on the coupling parameter λ is shown for the bonded (left column) and non-bonded (right column) potentials. The potentials for $\lambda = 0$ and $\lambda = 1$ correspond to the ones used at the end points of the integration, whereas the various grey lines between them are the intermediate states, each corresponding to a distinct value of λ .

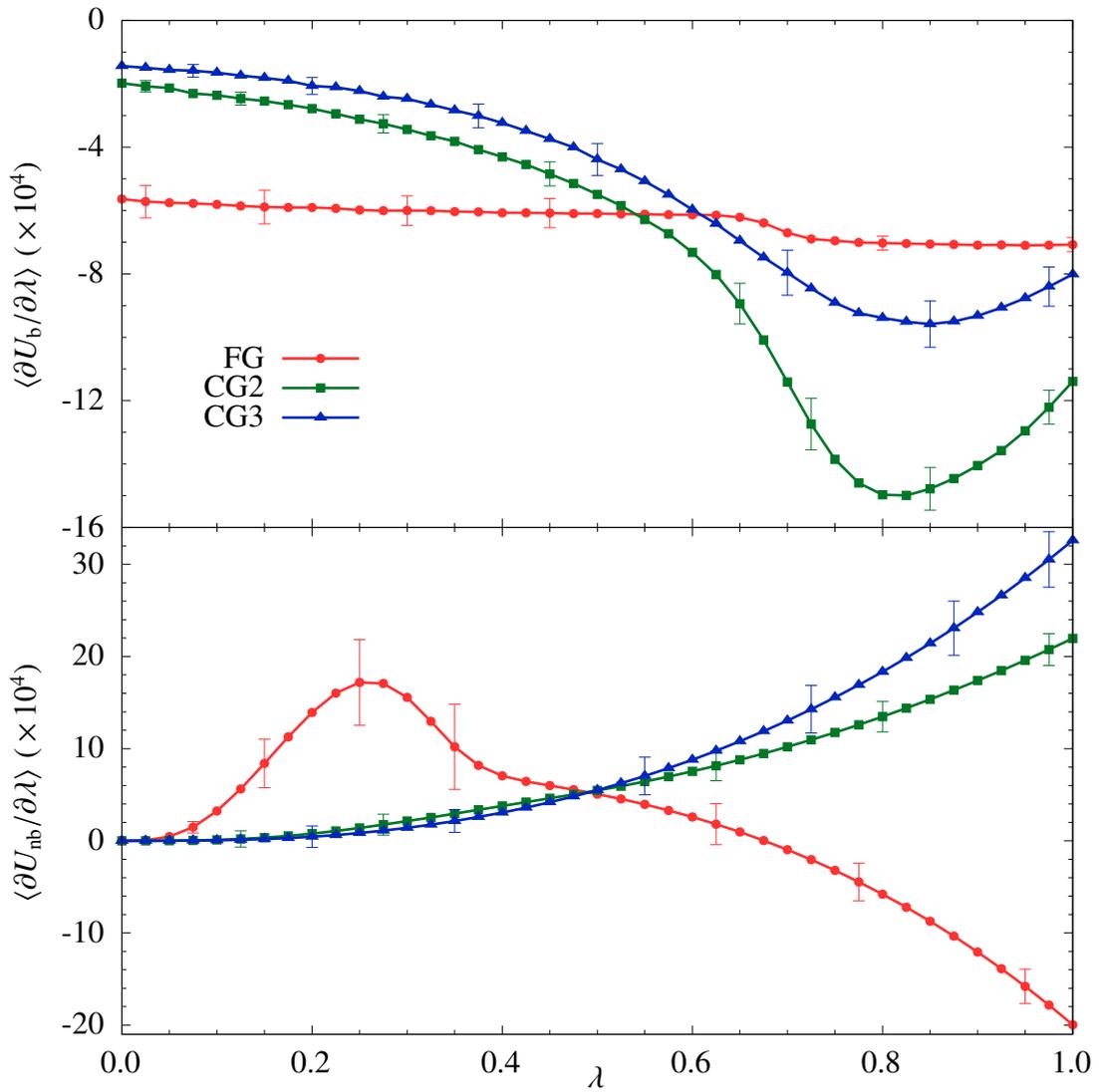


Figure 3.17.: Dependence of the λ -derivatives of the bonded (top) and non-bonded (bottom) potentials to be integrated in the in TI scheme for obtaining ΔS_b and ΔS_{nb} , respectively, via the free energy route as prescribed by Equation (2.78) with respect to the coupling parameter λ for the three CG resolutions FG, CG2, and CG3 at $T = 2.8$ and $\rho = 0.8$. Error bars are shown for selected values of λ to illustrate the characteristic statistical errors.

Entropy change due to non-bonded interactions

The entropy difference between the non-interacting flexible-bond FJC and the fully-interacting LJC, ΔS_{nb} , is calculated using Equation (2.79), as before, that is,

$$\Delta S_{nb} = \frac{1}{T} [(\langle E \rangle_{\lambda=1} - \langle E \rangle_{\lambda=0}) - \Delta F_{nb}] , \quad (3.117)$$

where the associated free energy difference is calculated using TI,

$$\Delta F_{nb} = \int_0^1 \left\langle \frac{\partial U_{nb}(\lambda)}{\partial \lambda} \right\rangle_{\lambda} d\lambda \quad (3.118)$$

where the thermodynamic end points of the integration $\lambda = 0$ and $\lambda = 1$ correspond to the non-interacting and interacting states, respectively. Differently from the case for the bonded potential in which the end point states were both associated with well-defined bonded potentials, in this case the goal is to completely switch off the non-bonded potential at $\lambda = 0$ and then progressively turn it on again as λ increases towards $\lambda = 1$, at which point the non-bonded potential should be completely recovered. One naive approach is simply to adopt a linear interpolation approach, typically of the form $U(\lambda) = \lambda U_{nb,0} + (1 - \lambda)U_{nb,1}$, which in our case would translate to $U(\lambda) = \lambda U_{nb}$, where U_{nb} is simply the full non-bonded potential of the actual FG and CG systems. The problem with this approach is the unbounded nature of the non-bonded interactions at short range, which means that the hard-core nature will not disappear even for $\lambda \equiv 0$, thus singularities are observed. A clear advantage of such linear schemes is that the derivative of $U(\lambda)$ with respect to λ is trivial, i.e., it's simply U , and moreover the internal energy of intermediate states can be easily interpolated.

One popular solution to this problem is the use of soft-core schemes such as the one proposed by Beutler et al. [102] in which the LJ potential is parameterized in terms of λ as

$$U(r, \lambda) = \lambda^n 4\epsilon \left[\frac{1}{[\alpha(1 - \lambda)^2 + (r/\sigma)^6]^2} - \frac{1}{\alpha(1 - \lambda)^2 + (r/\sigma)^6} \right] , \quad (3.119)$$

where ϵ and σ are the usual LJ parameters, r is the scalar distance between the two particles, n is a positive integer constant, and α is a positive constant. The soft-scaling character is conferred by the introduction of $\alpha(1 - \lambda)^2$ in the denominator and disappears as λ increases from 0 to 1. This parametrization leads to

stable simulations for every value of λ , with continuous derivatives of the Hamiltonian and no divergence of its ensemble average as $\lambda \rightarrow 0$. This scheme has been successfully used by Voyiatzis et al. [122] in a investigation of systems similar to the FG model examined here. However, it cannot be used for tabulated CG potentials thus making it unsuitable for this work as one of our goals here is to be as consistent as possible across all resolutions.

In this work we circumvent this problem by using the method of Buelens and Grubmüller [145], namely the Linear-scaling Soft-Core (LSC) scheme, in which the non-bonded potentials are modified such that two main conditions are met. The first is that the ensembles generated by using the potentials at the end points $\lambda = 0$ and $\lambda = 1$ must be indistinguishable from those generated when the native, i.e., unmodified, potentials are used. The second condition is that the regions that are energetically inaccessible in the non-bonded potentials under the thermodynamic conditions of the simulations, and thus regions that have no influence in the resulting ensemble, must be optimized so that linear λ -scaling of the resulting potential is well behaved numerically. This is achieved by dividing the LSC potential U_{LSC} in two regions, one native in which the potential is simply the unmodified U_{nb} potential, and a capped region U_{cap} , and both regions are separated by a threshold distance r_{cap} , i.e.,

$$U_{\text{LSC}}(r) = \begin{cases} U_{\text{nb}}(r) & \text{for } r \geq r_{\text{cap}} \\ U_{\text{cap}}(r) & \text{for } r < r_{\text{cap}}, \end{cases} \quad (3.120)$$

where $U_{\text{cap}}(r_{\text{cap}}) = U_{\text{nb}}(r_{\text{cap}})$ and the region $r < r_{\text{cap}}$ modified to eliminate the possibility of singularities. Determining r_{cap} is achieved by performing equilibrium MD simulations using the unmodified potentials for each system of interest and then recording the maximum value of U_{nb} that is ever evaluated during a simulation for a each system, i.e., U_{max} , which determines a pair-wise non-bonded energy limitation that cannot be exceeded during a simulation at the respective temperature and density. Then r_{cap} is naturally defined such that $U_{\text{nb}}(r_{\text{rcap}}) = U_{\text{max}}$. In practice it is advisable to be conservative with U_{max} and set it about 5% higher than the actually recorded maximum during the MD run, then there is no risk in affecting the end state simulations when altering the non-bonded potential below r_{cap} since that region will never be accessed during a simulation.

The form of U_{cap} must be such that the potential energy of the capped region is less favorable than at r_{cap} , while keeping good properties such as elimination of the singularity at zero separation and continuous

derivative with respect to the separation distance in order to avoid problems with the MD machinery. This is achieved by defining a switch region of width r_{switch} over which the potential is interpolated using a cubic polynomial form,

$$U_{\text{LSC}}(r) = \begin{cases} U_{\text{nb}}(r_{\text{cap}}) + U_{\text{switch}} & \text{for } r \leq (r_{\text{cap}} - r_{\text{switch}}) \\ ar^3 + br^2 + cr + d & \text{for } (r_{\text{cap}} - r_{\text{switch}}) < r < r_{\text{cap}} \\ U_{\text{nb}}(r) & \text{for } r \geq r_{\text{cap}}, \end{cases} \quad (3.121)$$

with the force

$$F_{\text{LSC}}(r) = \begin{cases} 0 & \text{for } r \leq (r_{\text{cap}} - r_{\text{switch}}) \\ -(3ar^2 + 2br + c) & \text{for } (r_{\text{cap}} - r_{\text{switch}}) < r < r_{\text{cap}} \\ F_{\text{nb}}(r) & \text{for } r \geq r_{\text{cap}}, \end{cases} \quad (3.122)$$

where $U_{\text{nb}}(r)$ and $F_{\text{nb}}(r)$ are the unmodified non-bonded potential and force, respectively, and $U_{\text{LSC}}(r)$ and $F_{\text{LSC}}(r)$ are the resulting linear-scaling soft-core potential and force, respectively. The constant U_{switch} is an energy penalty for zero separation, and r_{switch} is the distance over which U_{switch} is reached. The parameters a , b , c , and d are calculated to match the boundary conditions at r_{cap} and r_{switch} . A schematic representation of the scheme is shown in Figure 3.18.

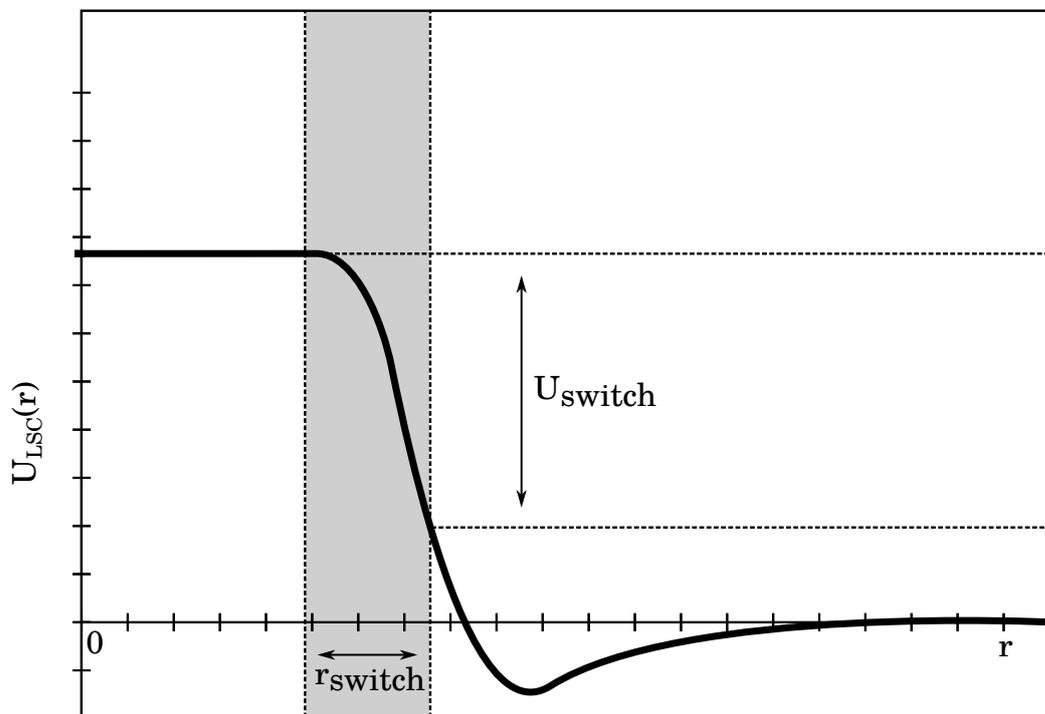
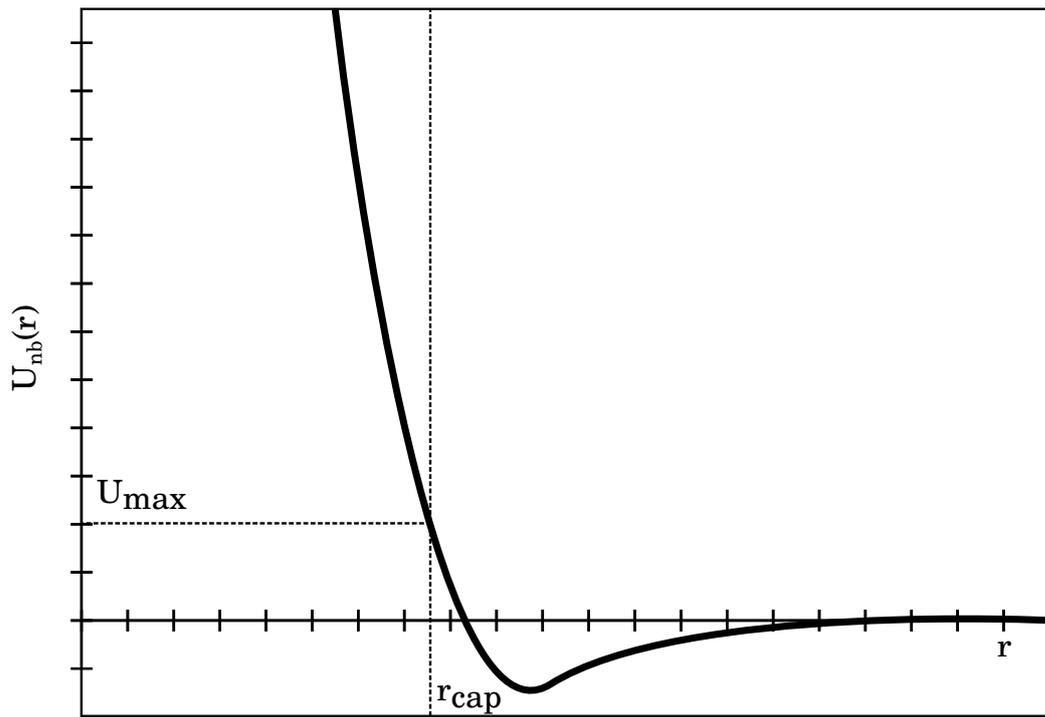


Figure 3.18.: Schematic representation of the Linear-scaling Soft-Core scheme applied to the LJ non-bonded potential.

One advantage of the LSC scheme is that it can be applied to tabulated potentials since the manipulations necessary in the potentials are simple enough to be applied to discrete points. Once the U_{LSC} potential is obtained, the coupling with λ for the TI is simply done using a linear scaling, i.e.,

$$U_{\text{nb}}(r; \lambda) = \lambda U_{\text{LSC}}(r), \quad (3.123)$$

which is now possible since the $U_{\text{LSC}}(r)$ is well behaved at short separations. The resulting potentials for the FG and CG systems at a selected state point can be seen in the right column of Figure 3.16, where it is possible to notice the capping of the potential at short separations and a balanced spread of the intermediate potentials over the λ points. The ensemble averages of the λ -derivatives and their dependency with λ are shown in the bottom panel of Figure 3.17 and it can be seen that they are continuous and well behaved.

Simulation protocol for thermodynamic integration

The calculation of ΔS_{b} and ΔS_{nb} were performed separately and for each integration we used a total of 41 λ -points distributed between 0 and 1 in steps of $\Delta\lambda = 0.025$. We performed an *NVT* simulation of length of 10^7 time steps using an integration step $\Delta t = 0.001$ for each value of λ . Energetic data was collected every 1000 steps, which rendered very smooth curves to be integrated. These were fitted to cubic splines and numerically integrated using the Romberg algorithm [140]. The end points corresponding to $\lambda = 0$ and $\lambda = 1$ were simulated for 10^8 steps in order to generate more accurate estimates of the internal energy difference, ΔE , required in Equations (2.74), (3.114) and (3.117).

3.6.2. Two-Body Approximation

As previously stated in Section 2.4.2, s_2 as given in Equation (2.86) is the translational two-body term of the excess entropy in which the angular contributions have been averaged out, which is a valid approximation in the context of dense and homogeneous amorphous polymer melts such as the LJC considered here as for these systems there is no strong angular dependence of the RDF. For the s_2 results discussed in this thesis Equation (2.86) has been used with the RDF $g(r)$ calculated talking into account the contribution of the first neighbors for the pair correlations. As this is not usually done when evaluating $g(r)$, e.g., for bottom-up coarse-graining, it is worth mentioning here. The inclusion of the immediate bonded monomer

makes sense for this case as s_2 must capture all pair-correlations in the system. In Figure 3.19 it is shown a comparison for an FG system of how $g(r)$ and s_2 differs depending whether the first bonded neighbor is included or excluded when computing $g(r)$. When included, a distinct peak is clear at $r = 1$, which is the equilibrium distance of the harmonic bonds in the chain. The s_2 response to this peak is noticeable decrease, as it is expected given that a strong two-body spatial correlation exists when the first bonded neighbor is included.

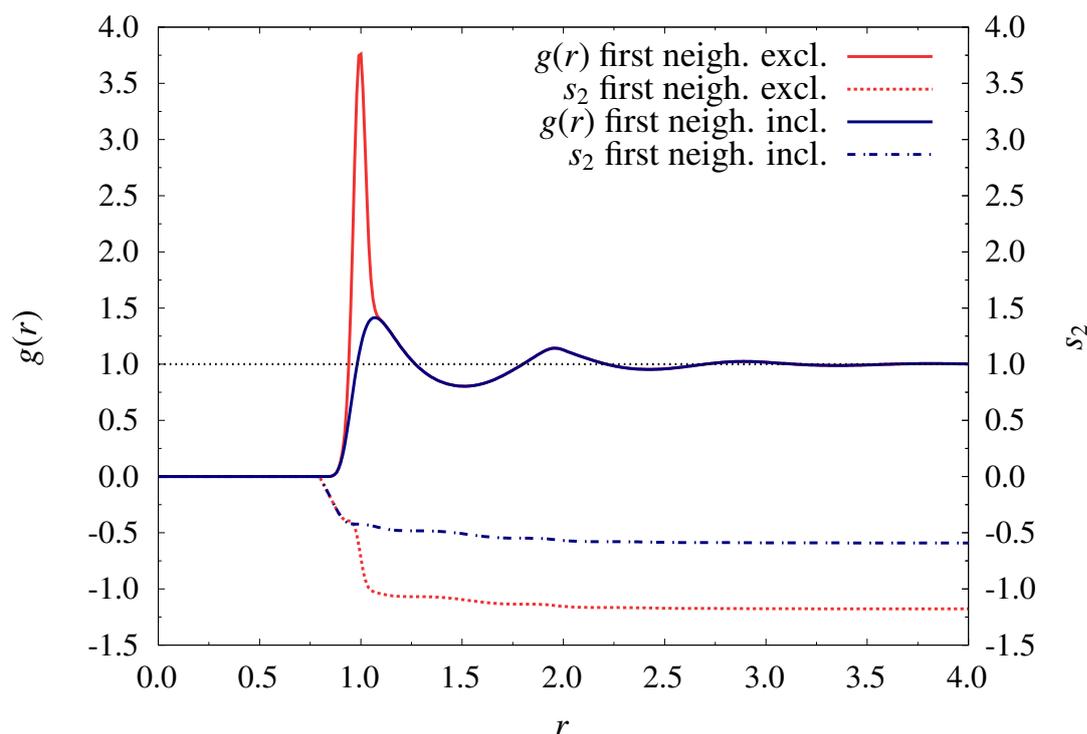


Figure 3.19.: RDF $g(r)$ and respective two-body excess entropy s_2 according to Equation (2.86) including (red) and excluding (blue) the first bonded neighbor contribution to the pair correlations for the FG system at $T = 4.0$ and $\rho = 0.8$.

3.7. Accelerated Dynamics and Excess Entropy Scaling

The aim of this work is to evaluate how the acceleration of the dynamics of CG models can be cast into excess entropy differences. For this reason, the excess entropy scaling relations presented in Section 2.5 take slightly different forms. Instead of directly relating the reduced dynamical properties to the excess

entropy, we rather look at the ratios of these properties across resolutions and relate them to excess entropy differences, i.e.,

$$\ln \frac{X_{\text{CG}}}{X_{\text{FG}}} = \ln A + \alpha \Delta S_{\text{exc}}, \quad (3.124)$$

where X_{CG} and X_{FG} are dynamical properties of the CG and corresponding FG system, respectively, and A and α are fit coefficients. We measure the acceleration of the dynamics of the CG models by taking the appropriate ratio of properties at different resolutions, i.e., $X_{\text{CG}}/X_{\text{FG}}$, where X is a dynamical property (ϑ , D , η). In the case of bond reorientation times, ϑ , the ratio $\vartheta_{\text{CG}}/\vartheta_{\text{FG}}$ corresponds to how much faster a bond vector reorients for the CG model compared with the FG model. The same procedure is used for the viscosity, $\eta_{\text{CG}}/\eta_{\text{FG}}$, since the viscosities of the CG models are considerably lower than for the reference FG model. In the case of the diffusion coefficient, however, the ratio must be considered as $(D_{\text{CG}}/D_{\text{FG}})^{-1}$, since D increases as the dynamics is accelerated, as opposed to what happens with ϑ and η . In this work we refer to such ratios as simply the acceleration or speed-up of a model with respect to another, but they are also known in the literature as scaling factors [6] and dynamical errors associated with coarse-graining [34]. The term ΔS_{exc} in Equation (3.124) is typically the excess entropy difference per monomer for two resolutions, e.g.,

$$\Delta S_{\text{exc}} = \frac{S_{\text{exc}}^{\text{CG}}}{N_{\text{CG}}} - \frac{S_{\text{exc}}^{\text{FG}}}{N_{\text{FG}}}, \quad (3.125)$$

where N_{CG} and N_{FG} are the number of monomers in the simulation box for the CG and FG systems, respectively, and the k_B has been omitted from the denominator since $k_B = 1$ in the unit system adopted in this work (cf. Section 3.1.1). Occasionally the excess entropy difference may also be normalized per molecule, as seen in Chapters 4 and 5.

4. Coarse-Graining Effects on Dynamics and Entropy

In this chapter we investigate the effects that coarse-graining has on the dynamics and entropies of the systems. In particular, we highlight that the dynamics of the CG systems are profoundly accelerated with respect to the FG counterparts. Moreover, the elimination of DOF that comes naturally with coarse-graining results in a significantly lower excess entropy for the CG systems. We also discuss how well the two-body term of the many-body entropy expansion captures the full excess entropy.

4.1. Dynamics in Coarse-Grained Systems

As stated in the introduction of this thesis, the removal of friction upon coarse-graining leads to an artificial acceleration of the CG dynamics in comparison to the FG model. This consequence of coarse-graining is clearly illustrated for a particular state point in Figures 4.1 to 4.3, which shows for the CG and FG systems the MSD of the monomers, the running viscosity integral resulting from the GK relation, and the normalized bond reorientation ACF, respectively. It can be observed that the ordering of the curves persists for the three plots, i.e., acceleration of CG3 is higher than of CG2. This is a good sign that the CG models have been built in a correct way and the results agree with the central argument of friction removal. In Figures 4.1 to 4.3 we can also see the temperature-dependence of the dynamical properties for all resolutions. For FG they are as expected, i.e., higher temperature means higher diffusivity, lower viscosity, faster local relaxations. For the CG systems these dependencies are discussed in what follows.

The $\text{MSD}(t)$ profile and the integrated viscosity are associated to diffusion processes of entire chains and they take place on much longer time scales in comparison to the bond reorientations, therefore they are taken as indicators of global dynamics. The MSD reveals that the chain motions are higher for coarser resolutions, as expected, with the average self-diffusion of the FG system being only about 12% and 7% of the average self-diffusion of CG2 and CG3, respectively. For example, for the density $\rho = 0.8$ the self-diffusion of the FG system over the temperatures considered ranges from 0.0071 ± 0.0003 to 0.0144 ± 0.0005 , while for CG2 it ranges from 0.056 ± 0.008 to 0.085 ± 0.004 , and for CG3 from 0.095 ± 0.003 to 0.144 ± 0.005 . Additionally, the onset of the long-time diffusion regime characterized by $\text{MSD}(t) \propto t$ (i.e., slope = 1 in the log-log plot) happens earlier in CG3 than in CG2, and also in CG2 than in FG, e.g., for the state point shown in Figure 4.1, unity-slope is reached for CG3 at about $t = 50$, while for CG2 it is reached at about $t = 500$, and for FG this regime is reached at about $t = 3000$. As temperature increases the increase of the self-diffusion coefficient is more pronounced for the CG systems, as it can be seen in the bottom panel of Figure 4.1, thus the temperature dependence is stronger for the coarser resolutions, with the slopes of the $D \times T$ curve being 0.00373 ± 0.00007 and 0.00233 ± 0.00004 for FG, 0.0151 ± 0.0009 and 0.0135 ± 0.0004 for CG2, and 0.025 ± 0.001 and 0.019 ± 0.002 for CG3, for densities $\rho = 0.8$ and $\rho = 0.9$, respectively.

As it was to be expected, the viscosity is significantly decreased in the coarser models, also reflecting the faster CG dynamics. Similarly to what is observed in the diffusion, the average viscosities of CG2 and CG3 are about 12% and 7% of that of FG. This shows that the decrease of the viscosity is greater for CG2

with respect to FG than for CG3 with respect to CG2, indicating that, for this LJC system, the lesser degree of coarse-graining, i.e., 2:1 mapping, is already enough to remove most of the friction from the system. The mapping using 3:1 scheme does not remove much extra friction, as $\eta(t)$ is only slightly lower in CG3 than in CG2, e.g., for the density $\rho = 0.8$ and the temperatures considered, it ranges from 11.918 ± 1.67 to 12.33 ± 1.98 for FG, 1.80 ± 0.12 to 2.11 ± 0.14 for CG2, and 0.93 ± 0.12 to 1.26 ± 0.09 for CG3. It can be argued that at CG3 we are already very close to the limit of coarse-graining possible in the LJC model before chains start to cross each other. In the case of viscosity the temperature dependence becomes weaker as the resolution is lowered, with the slope of the $\eta \times T$ curves in the bottom panel of Figure 4.2 being -0.3 ± 0.3 and -0.9 ± 0.4 for FG, -0.23 ± 0.04 and -0.27 ± 0.03 for CG2, and -0.16 ± 0.02 and -0.18 ± 0.02 for CG3, for densities $\rho = 0.8$ and $\rho 0.9$, respectively.

The normalized bond reorientation ACF describes relaxation processes at very short time and length scales and thus is regarded in this work as a local dynamical property. For densities 0.8 and 0.9, the average bond reorientation times for FG are 0.019 ± 0.002 and 0.029 ± 0.003 , respectively, while for CG2 they are 0.0041 ± 0.0005 and 0.0039 ± 0.0006 , and for CG3 they are 0.0031 ± 0.0004 and 0.0027 ± 0.0005 , respectively. Thus, in average, ϑ_{CG2} and ϑ_{CG3} are about 16 % and 12 % of ϑ_{FG} , i.e., the decorrelation of bond directions for the two coarser resolutions happens at time scales that are roughly one order of magnitude less than for FG. Furthermore, as shown in the bottom panel of Figure 4.3, the reduction of the bond reorientation time as temperature increases is much less pronounced in the CG systems than in the FG, showing a weak temperature dependence, e.g., for CG3 and $\rho = 0.9$ the fit slope is only -0.00054 ± 0.00003 .

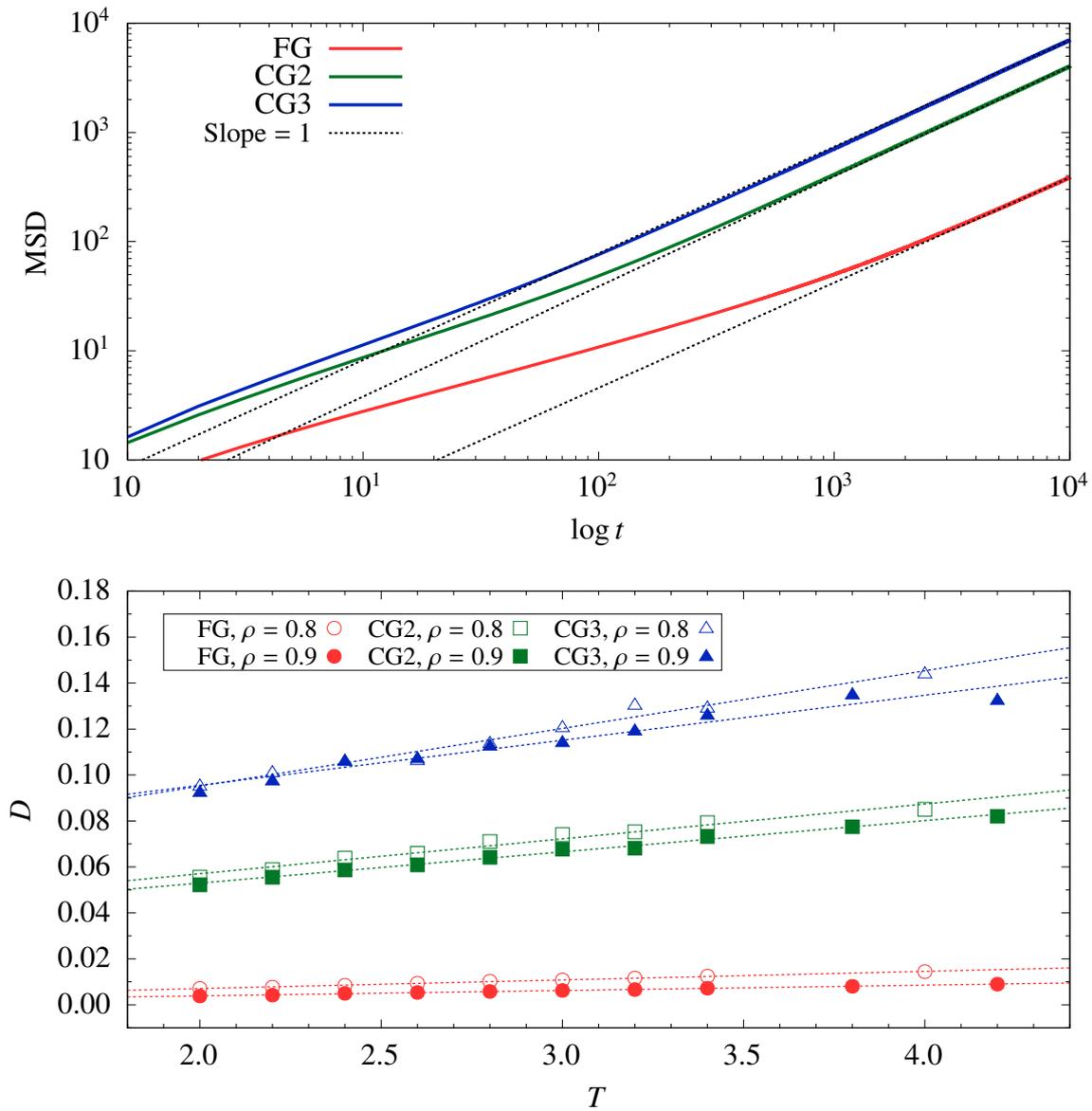


Figure 4.1.: Comparison of the mean-square displacements (top panel) and the temperature dependence of the self-diffusion coefficient (bottom panel) for the FG, CG2, and CG3 systems. The mean-square displacements are shown for the state point $\rho = 0.9$ and $T = 3$ as a function of time. Dotted lines in the temperature-dependence plot are the best linear fits for each resolution and density.

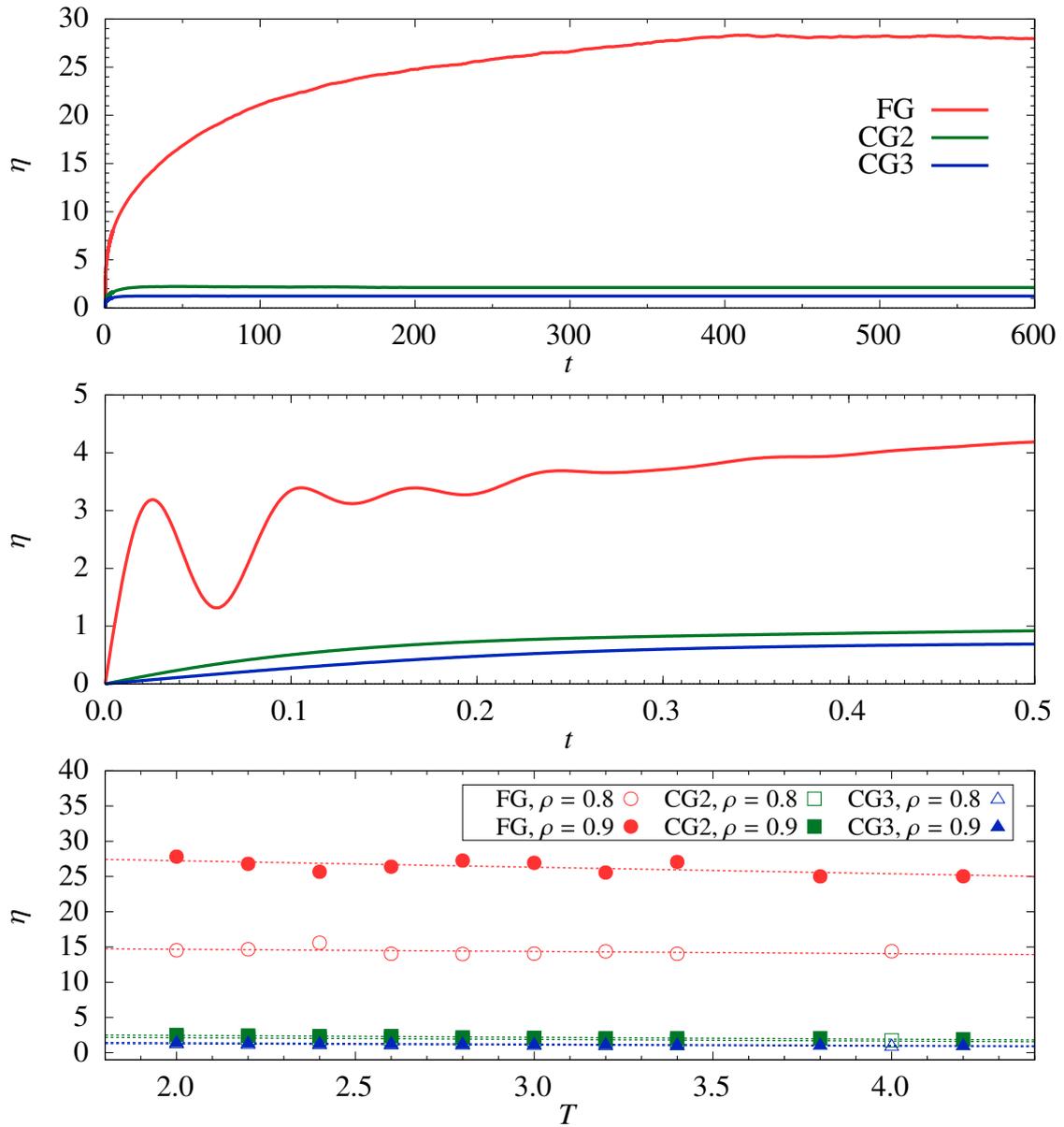


Figure 4.2.: Comparison of the viscosity integrals (top and middle panels) and the temperature dependence of the viscosity (bottom panel) for the FG, CG2, and CG3 systems. The viscosity integrals are shown for the state point $\rho = 0.9$ and $T = 3$ as a function of time. The middle panel plots the same data as the top panel, but zoomed in the short-time region in order to highlight the details. Dotted lines in the temperature-dependence plot are the best linear fits for each resolution and density.

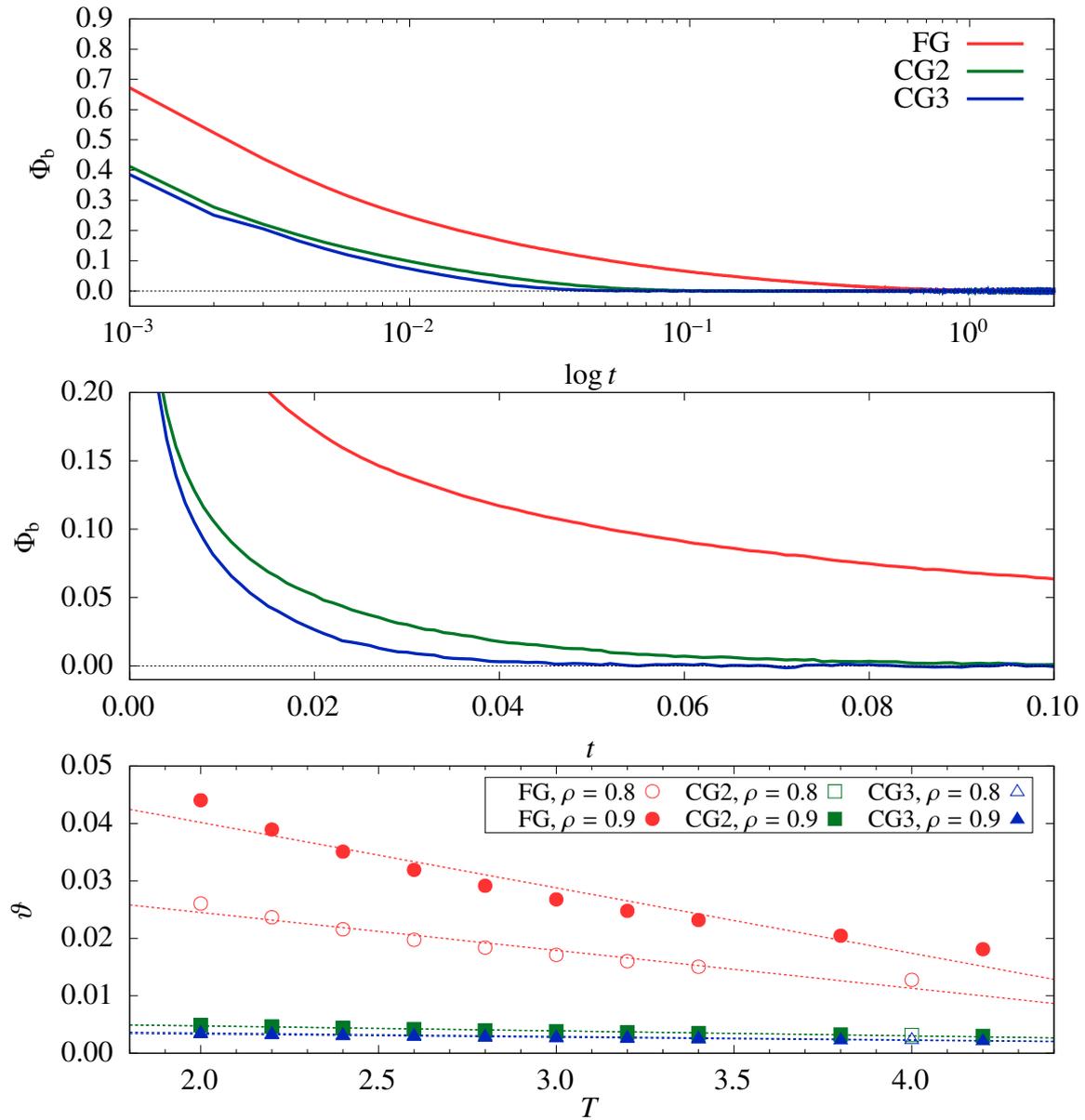


Figure 4.3.: Comparison of the bond reorientation ACF (top and middle panels) and the temperature dependence of the bond reorientation time (bottom panel) for the FG, CG2, and CG3 systems. The ACF are shown for the state point $\rho = 0.9$ and $T = 3$ as a function of time. The middle panel plots the same data as the top panel, but zoomed in the short-time region in order to highlight the details. Dotted lines in the temperature-dependence plot are the best linear fits for each resolution and density.

4.2. Entropy in Coarse-Grained Systems

The main observation that can be made when analyzing the entropies in light of coarse-graining is that the total absolute entropy of the melts, i.e., the absolute entropy calculated by TI according to Section 3.6, S , of a simulation box containing $N_c = 500$ chains, decreases upon coarse-graining, with a greater decrease for coarser resolutions. As the number of chains is unchanged for all three resolutions, the absolute entropy per chain, S/N_c , behaves in the same way, as seen in Table 4.1. What is interesting, however, is that the factor of decrease of the entropy is not equal to the factor of decrease in the number of monomers. The absolute entropies in CG2 and CG3 for a given state point are not equal to half and one third, respectively, to that of the corresponding FG system, even though the latter has two and three times as many monomers per chain as CG2 and CG3, respectively.

Table 4.1.: Absolute and excess entropies for the FG and CG systems obtained with TI. Absolute entropy per chain, S/N_c , absolute entropy per monomer, S/N , and excess entropy per monomer, S_{exc}/N , where N is the total number of monomers in the respective resolution.

ρ	T	S/N_c			S/N			S_{exc}/N		
		FG	CG2	CG3	FG	CG2	CG3	FG	CG2	CG3
0.8	2.0	173.16	134.25	98.61	7.21	11.18	12.32	-4.58	-2.06	-1.70
0.8	2.2	175.96	134.57	98.78	7.33	11.21	12.34	-4.54	-1.96	-1.55
0.8	2.4	178.52	135.13	99.28	7.43	11.26	12.41	-4.52	-1.88	-1.46
0.8	2.6	180.98	135.88	99.28	7.54	11.32	12.41	-4.50	-1.84	-1.40
0.8	2.8	183.22	136.43	99.58	7.63	11.36	12.44	-4.48	-1.79	-1.32
0.8	3.0	185.24	137.00	99.75	7.71	11.41	12.46	-4.46	-1.72	-1.24
0.8	3.2	187.19	137.69	99.99	7.79	11.47	12.49	-4.45	-1.66	-1.12
0.8	3.4	188.91	137.78	100.06	7.87	11.48	12.50	-4.42	-1.58	-1.05
0.8	4.0	193.75	139.52	100.78	8.07	11.62	12.59	-4.40	-1.46	-0.93
0.9	2.0	163.14	132.70	97.80	6.79	11.05	12.22	-4.79	-2.11	-1.70
0.9	2.2	166.27	133.42	98.23	6.92	11.11	12.27	-4.78	-2.07	-1.62
0.9	2.4	169.18	133.80	98.28	7.04	11.15	12.28	-4.76	-1.98	-1.49
0.9	2.6	171.93	134.34	98.68	7.16	11.19	12.33	-4.75	-1.92	-1.43
0.9	2.8	174.32	134.86	99.05	7.26	11.23	12.38	-4.74	-1.86	-1.34
0.9	3.0	176.65	135.38	99.23	7.36	11.28	12.40	-4.73	-1.80	-1.26
0.9	3.2	178.72	135.85	99.28	7.44	11.32	12.41	-4.72	-1.73	-1.20
0.9	3.4	180.68	136.19	99.54	7.52	11.34	12.44	-4.72	-1.66	-1.13
0.9	3.8	184.30	137.15	99.86	7.67	11.42	12.48	-4.71	-1.56	-1.00
0.9	4.2	187.57	138.45	100.39	7.81	11.53	12.54	-4.70	-1.47	-0.92

This scenario is changed when the absolute entropy is normalized by the number of monomers at each resolution, as shown in Figure 4.4 and Table 4.1. The absolute entropy per monomer, S/N , increases as the level of resolution gets coarser. This can be explained by the fact that a fraction of the DOF removed from the FG model upon coarse-graining do not contribute significantly to entropy of the system, in particular the hard DOF such as those associated with the stiff harmonic bonds. Thus the drop in entropy is not proportional to the removal of DOF, but comparatively lower, yielding a higher value per remaining DOF.

A similar result was found by Voyiatzis and Böhm [123], who computed the quasi-harmonic contributions to the entropy of the various DOF and offered quantitative support to the argument presented here. Moreover, the DOF of the CG systems are softer and allow a large number of microstates that are unavailable to the FG system to be realized. The entropy per particle, S/N , is a linear function of the temperature. The slope of the fits at each resolution is greater than zero and equal within the errors for both densities. They are 0.43 ± 0.02 , 0.21 ± 0.01 , and 0.140 ± 0.006 for FG, CG2, and CG3, respectively. Interestingly, they match the factors of decrease in number of monomers, i.e., 0.21 is approximately half of 0.43, and 0.14 is close to one third of the latter. Given that the pressures in all systems are approximately equal by construction of the CG potentials, and based on the fact that $T(\partial S/\partial T)_P$ is the heat capacity at constant pressure, c_P , two arguments can be given to understand the observed decreased temperature dependence for coarser resolutions. First, the decrease of DOF upon coarse-graining, in conjunction with the softness of the CG vibrational potential, leads to a decreased heat capacity for these systems, and thus decreased slope. Second, the coarse-graining flattens out the potential energy surface and makes it smoother, thus energy fluctuations on this surface are smaller, leading to a decrease in c_P for the CG systems [94, 146].

The excess entropy per monomer is also shown in Figure 4.4. The ordering of the resolutions is preserved in comparison with the absolute entropy per monomer, however, the temperature dependence is modified. For all resolutions we confirm the thermodynamic result, $\partial S_{\text{exc}}/\partial T > 0$, and since $S_{\text{exc}} < 0$ by definition, we have $S_{\text{exc}} \rightarrow 0$ as $T \rightarrow \infty$. This is expected [27]: the difference between the real system and the IG system diminishes as temperature increases. Moreover, the excess entropy is closer to zero for $\rho = 0.8$ than for 0.9, which is also expected given that $S_{\text{exc}} = 0$ at zero density. We can also see that the slopes are steeper for coarser resolutions, taking the averaged values 0.065 ± 0.005 , 0.298 ± 0.009 , and 0.37 ± 0.01 , and for FG, CG2, and CG3, respectively, with the averages calculated over the two densities for each resolution. Hence the slopes indicate that as the resolution is lowered and the level of CG is increased, the excess

entropy approaches zero more rapidly with increasing temperature. This complies with the fact that more coarse-graining means less hard-core repulsion and a smoother potential surface. Thus a lower-resolution system, in this context, is closer to the IG behavior than a higher-resolution one, a fact that is also evident in the actual values of the excess entropies, i.e., closer to zero for CG3.

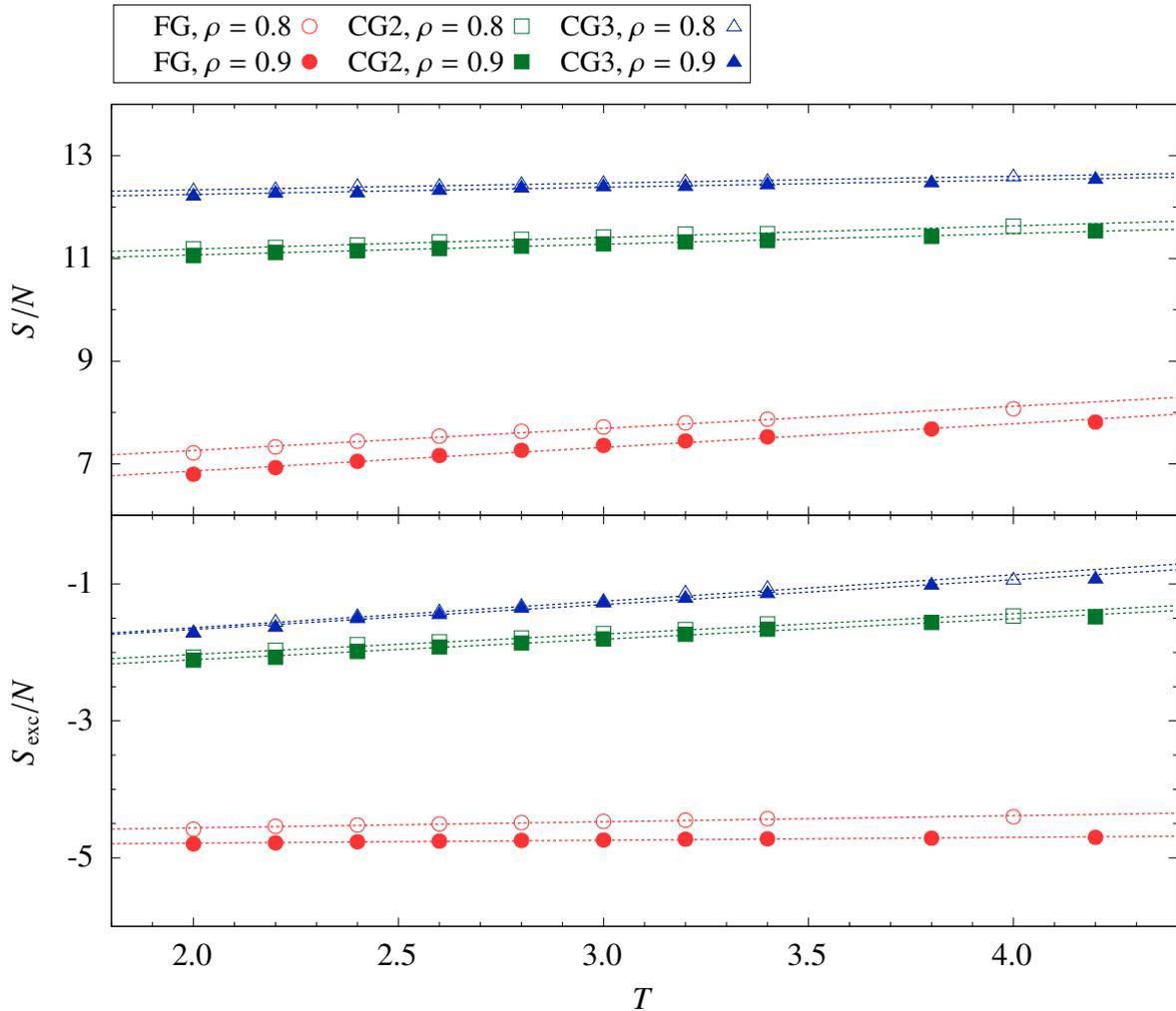


Figure 4.4.: Temperature dependence of the absolute entropy obtained with TI, S , and of the excess entropy, S_{exc} , for the FG and CG systems. Entropies are shown for all thermodynamic state points considered and normalized by the number of monomers of a chain at each resolution, N . Dotted lines are the best linear fits for each resolution and density.

It is also interesting to analyze what happens to the entropies, absolute and excess, at the different phases of the TI as given by Equation (3.109). Borrowing the form of energy-levels diagram often used in quantum chemistry, in Figure 4.5 we show an “entropy-level diagram” for the state point $T = 3$ and $\rho = 0.9$, which proves to be quite useful. As seen in the ordering presented in the diagram, and in agreement with the previous discussions, the absolute entropy per chain (monomer) is larger at higher (lower) resolutions. The largest entropy difference with respect to the IG reference happens in the first phase and corresponds to the entropy loss associated with the polymerization of an IG into a system of non-interacting FJCs with rigid bonds of fixed length. In the second phase, characterized in the TI by the softening of the FJC bonds, a significant entropy gain is observed as the more flexible bonds allow more microstates to be realized, increasing the conformational space of the chains and, consequently, increasing the absolute entropy. We notice that the entropy gain per monomer is relatively larger for the coarser resolutions, as expected given that the bond-length distributions are wider in the CG systems, as seen in Chapter 3. Finally, as the non-bonded LJ or CG interactions are introduced to the system of non-interacting FJC with flexible bonds in the third phase of the TI, the resulting correlations impose constraints that limit the possible conformations, which causes another entropy loss.

4.2.1. Two-body Approximation to the Excess Entropy

The two-body entropy is the second term in the expansion $S = S_{\text{ig}} + S_2 + S_3 + \dots$, and considering that $S_{\text{exc}} = S - S_{\text{ig}}$, then S_2 is a natural approximation to S_{exc} . But how good is it? It has been argued in the literature that for monoatomic LJ fluids it captures up to 90% of the excess entropy [96, 108, 122, 147, 148]. In this work we calculate S_2 using only its translational component and compare it with the exact excess entropy obtained from TI to assess how well they agree, as shown in *Figure 4.6*, where the temperature dependence of the monomer-normalized values of S_2 and S_{exc} are plotted together.

Regarding the temperature dependence, in the case of the FG system it can be seen that S_{exc} and S_2 have a relatively similar behavior, with the fit slope coefficient for S_{exc} being 0.089 ± 0.006 and 0.042 ± 0.003 , and for S_2 equal to 0.11 ± 0.01 and 0.16 ± 0.01 , in each case for $\rho = 0.8$ and $\rho = 0.9$, respectively, thus the value of S_2 is spread across a larger range than S_{exc} for FG. For the CG systems S_{exc} becomes strongly dependent on temperature as the resolution gets coarser, with the fit slope of 0.29 ± 0.01 and 0.30 ± 0.01 for CG2, and 0.38 ± 0.02 and 0.36 ± 0.01 for CG3, for $\rho = 0.8$ and $\rho = 0.9$, respectively. However, the S_2

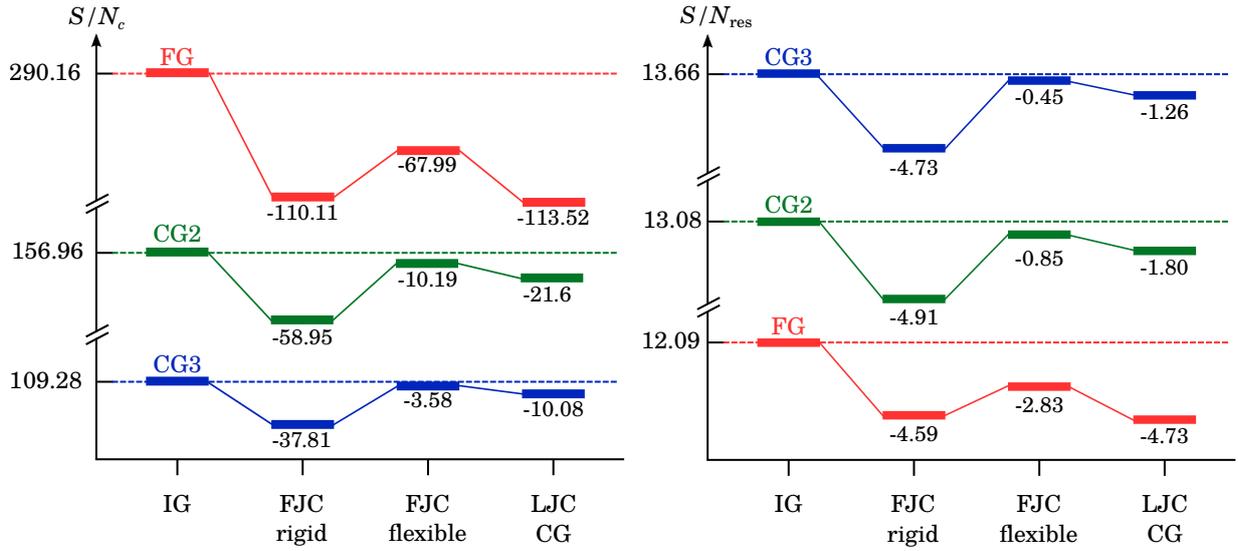


Figure 4.5.: Entropy level diagram for the different phases of the TI for state point $\rho = 0.9$ and $T = 3.0$, normalized by number of chains (top), N_c , and by total number of monomers (bottom) at each resolution, N_{res} , where res is FG, CG2, or CG3. Absolute entropy, S , of the respective IG system at the same thermodynamic conditions is shown on the vertical axis. The values below the bars correspond to excess entropy, i.e., the entropy loss with respect to the IG reference, of the rigid-bond non-interacting freely-jointed chains (FJC rigid), the flexible-bond non-interacting freely-jointed chains (FJC flexible), and the interacting LJC or coarse-grained chains (LJC/CG).

dependence becomes weaker, with the fit slope of 0.0125 ± 0.0003 and 0.0135 ± 0.0004 , and for CG2, and 0.0034 ± 0.0003 and 0.0038 ± 0.0001 for CG3, for $\rho = 0.8$ and $\rho = 0.9$, respectively.

Concerning how well the two-body approximation captures the excess entropy, this also depends on the resolution level. For the FG system and $\rho = 0.8$, the value of S_2 is only 33% of S_{exc} for the lowest temperature considered, and 29% for the highest temperature, with a similar quantitative behavior observed for $\rho = 0.9$. Thus for FG the translational part of S_2 captures roughly one third of the true excess entropy. It may be that adding the orientational portion of S_2 already improves this contribution significantly, but higher terms of the expansion may also play an important role. This is, however, only speculation. For CG2 at $\rho = 0.8$ the contribution of S_2 to S_{exc} is 35% at the lowest temperature, a figure that is similar to that of FG, but it increases up to 49% for the highest temperature. For CG3 at $\rho = 0.8$ this contribution is also ca. 35% at the lowest temperature but 64% for the highest temperature. Thus for increasing temperature and increasing level of coarse-graining the configurational portion of the two-body entropy becomes a better approximation to the excess entropy. However, the values seen here are well below those normally reported in the literature for LJ fluids and hard-sphere models [96, 149]. This discrepancy may be attributed, in part,

to the fact that S_2 as calculated here describes only the translational part of the two-body entropy, ignoring orientational contributions. A more relevant reason is the fact that we are dealing with polymer melts rather than small-molecule liquids. The monomer RDF is very close to that of a monomer fluid, lacking information about the conformational and translational entropies of the polymer chains. Particularly the former is dominant in polymer melts, and thus this simple S_2 estimate does not capture most of the actual excess entropy. However, this cheap entropy estimate is still useful in the context of this work, as shown in Chapter 5.

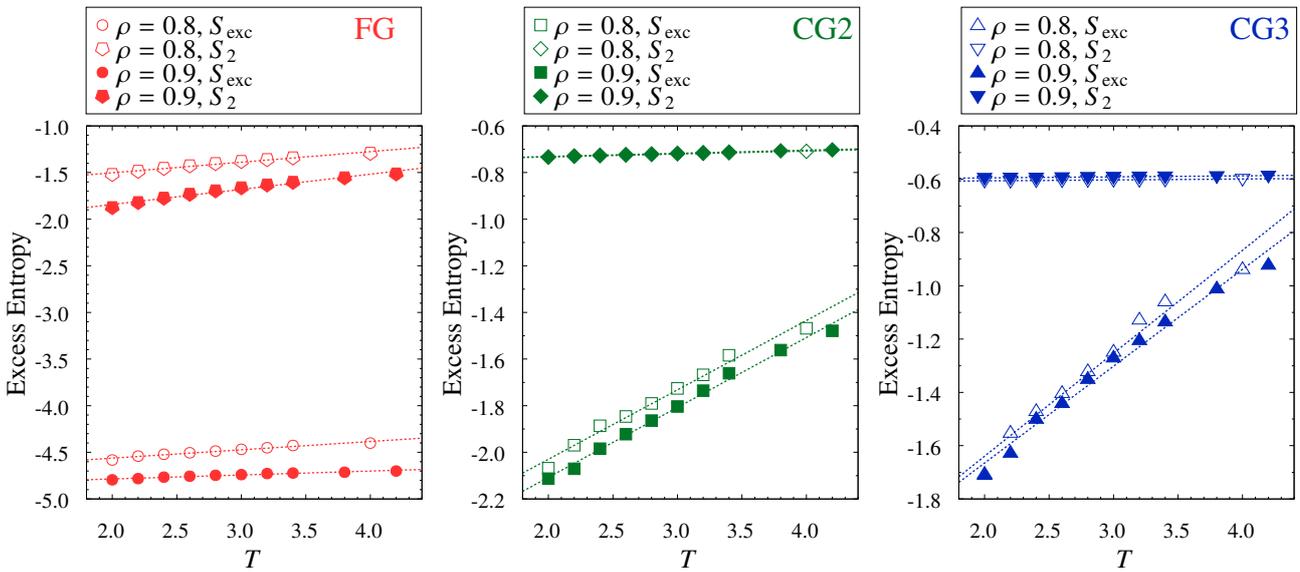


Figure 4.6.: Temperature dependence of the excess entropy obtained with TI, S_{exc} , and the two-body term of the absolute entropy, S_2 , for the FG and CG systems. Values given are normalized by the number of monomers of each resolution. Dotted lines are the best linear fits for each resolution and density.

It is also convenient to analyze how S_2 and S_{exc} correlate, as shown in Figure 4.7. For both densities and all resolutions there are clear linear correlations between the two quantities, although all lines are far from the ideal line $S_2/N_{res} = S_{exc}/N_{res}$. The slopes are greater than zero for every data set, but decrease as the resolution is coarsened. More importantly, the slopes of the fitted lines are < 1 for CG2 and CG3, and > 1 for FG. In the former case, the consequence is that as $S_{exc} \rightarrow 0$, S_2 converges to a negative non-zero value for CG2 and CG3, approximately at -0.61 for both resolutions. This is counter-intuitive and seems to indicate that even though the excess entropy is zero (i.e., the IG limit has been reached), there is still some remaining two-body entropy. This cannot be the case and we attribute this problem to the fact that S_2

as calculated here does not consider angular variations in the RDFs. Thus, this approximation behaves poorly for the coarser resolutions in the low excess entropy region. For FG the behavior is more in line with expectation: $S_2 \rightarrow 0$ well before $S_{\text{exc}} \rightarrow 0$. From the fitted lines in Figure 4.7 it seems that S_2 can assume values > 0 for S_{exc} closer to zero. This is not a physical behavior, and one should not try to extrapolate S_2 to that region. As just mentioned above, S_2 as calculated here disregards the orientational dependence in the RDFs and it completely ignores the conformational contributions to the entropy.

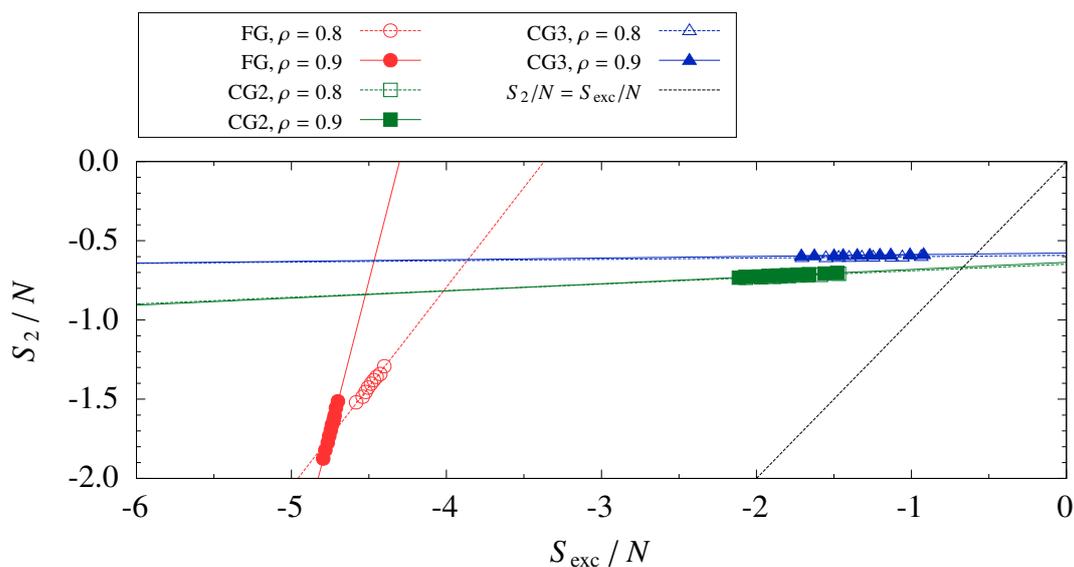


Figure 4.7.: Correlations between the two-body term of the absolute entropy, S_2 , and the excess entropy obtained with TI, S_{exc} , for the FG and CG systems. Values given are normalized by the number of monomers of each resolution. Dotted lines are the best linear fits for each resolution and density.

5. Dynamics and Excess Entropy Differences

In this chapter the observed correlations between the acceleration of the CG dynamics and the differences in excess entropy are presented and discussed. The acceleration of the dynamics of the CG models is measured by taking the appropriate ratios of the dynamical properties at different resolutions, as discussed in Section 3.7.

5.1. Note on Self-Diffusion and Viscosity

Before considering the correlations for the global dynamics by analyzing the acceleration with respect to the ratios of self-diffusion coefficient and viscosity, we note that according to the Rouse theory there is an inverse relationship between diffusion coefficient and viscosity for unentangled polymers [150, 151]. This is manifest in our results, as revealed by the linear correlation between η_{CG}/η_{FG} and $(D_{CG}/D_{FG})^{-1}$ shown in Figure 5.1, where the slope of fit line is unity, as it should due to the Stokes-Einstein relation which is known to work reasonably well for unentangled polymer melts [152]. We, therefore, discuss only one of the ratios in the remainder of this chapter. We chose to focus on the viscosity given that viscosities of polymer melts are easier to find in the literature and of greater relevance in polymer processing than diffusion coefficients. Moreover, recent methods for viscosity prediction based on excess entropy scaling [153, 154] make this property interesting to investigate in this context. Part of the analysis that follows for the viscosity ratios has also been performed for the diffusion coefficients with analogous results (cf. Appendix C).

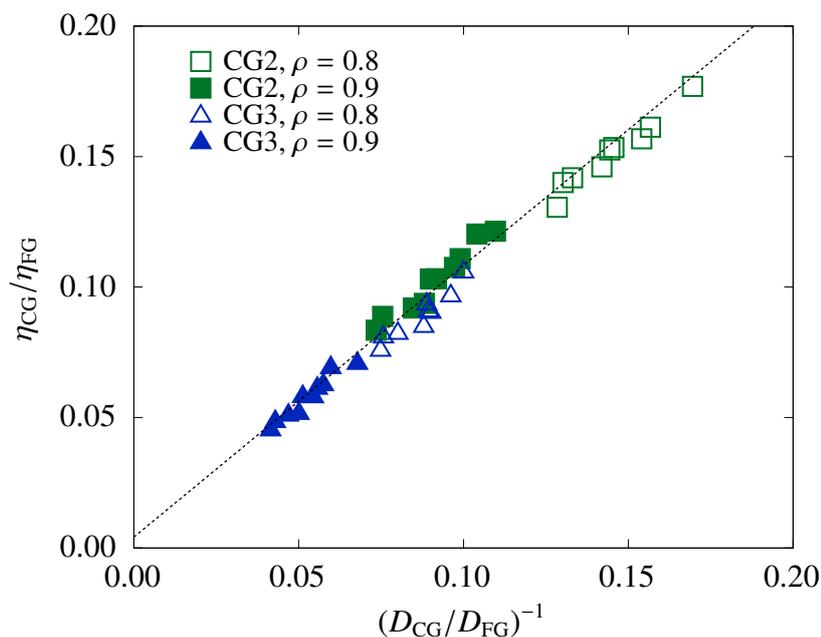


Figure 5.1.: Correlation between the ratio of the viscosity and the inverse ratio of self-diffusion coefficient of the CG models, CG3 and CG2, with respect to the FG model, FG. The dotted line corresponds to the best linear fit considering all the data points and has slope equal to 0.99 ± 0.02 .

5.2. Correlations Between Coarse-Grained and Fine-Grained Systems

The acceleration of the local dynamics in the CG systems with respect to the FG resolution, as given by the ratio of the bond reorientation time, is shown in Figure 5.2 as a function of the difference of the excess entropy per monomer between each corresponding resolution, $\Delta S_{\text{exc}} = S_{\text{exc}}^{\text{CG}}/N_{\text{CG}} - S_{\text{exc}}^{\text{FG}}/N_{\text{FG}}$. There is a distinct linear behavior in the semi-log plot for both resolutions and densities, and this strong correlation indicates that the excess entropy difference is a good estimator to assess by how much the coarse-graining accelerates the local dynamics. The relation given in Equation (3.124) holds for the bond reorientation time, where in this case $X = \vartheta$, with the intercept coefficient A and the slope coefficient α obtained by fitting Equation (3.124) to the data shown in Figure 5.2. In this case of ϑ , there appears to exist a master curve for each density and thus A and α are density dependent. However they do not depend on the level of coarse-graining considered. For $\rho = 0.8$ the values obtained in the fit are $A = 1.09 \pm 0.22$ and $\alpha = -0.60 \pm 0.02$, with an adjusted coefficient of determination $R^2 = 0.993$, while for $\rho = 0.9$ they are $A = 0.94 \pm 0.29$, $\alpha = -0.66 \pm 0.02$, with $R^2 = 0.997$. The difference between the slope coefficients is roughly 10%, i.e., about three times larger than the error of the regression, thus there is an actual dependence of the slope on the density. We highlight the fact that the logarithm of the ratio decreases in absolute value (towards zero) as the excess entropy difference is lowered. As $\Delta S_{\text{exc}} \rightarrow 0$, one would expect $\vartheta_{\text{CG}}/\vartheta_{\text{FG}} \rightarrow 1$, which is in qualitative agreement with the observed behavior.

Moving on to the global dynamics as described by the viscosity, the relationship of its ratios with the excess entropy differences are as shown in Figure 5.3. The dependence of the ratios on ΔS_{exc} is exponential in the same fashion as the ratios of the bond reorientation times. In the case of the viscosities, however, there is not a different curve for each density for the distinct CG resolutions. If lines are fitted through each data set, i.e., for data related to each pair of resolution and density, then their slopes are not that different. These fitted lines are not shown in Figure 5.3 to avoid clutter, but the fit coefficients are shown in Table 5.1. There appear to be offsets for both the resolution and density, as the CG3 system has lower η than CG2 at the same excess entropy difference, and as higher densities have a lower viscosity at the same excess entropy difference. Both offsets are numerically close so there is a coincidental overlap of the data sets for CG2 at $\rho = 0.9$ and CG3 at $\rho = 0.8$. The observed behavior suggests that the local dynamics may capture a “non-polymer behavior” that follows a distinct scaling with excess entropy differences, in analogy to what is observed for CG models of molecular liquids [35]. Nevertheless, excess entropy differences alone are not

sufficient to discern this “polymer behavior”, at least for the chain sizes considered in this work.

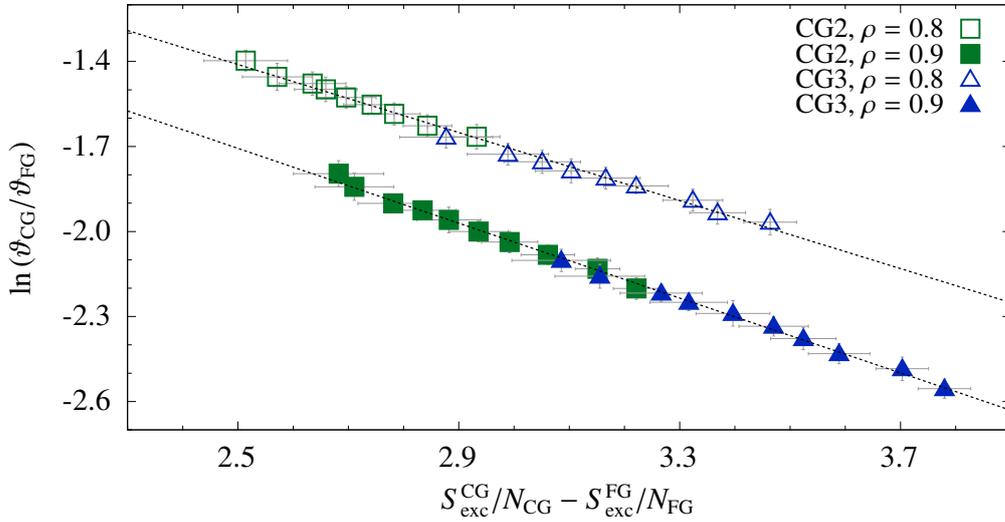


Figure 5.2.: Acceleration of the dynamics measured by the ratio of the bond reorientation time at CG resolutions CG2 and CG3 with respect to the FG model, FG, versus the corresponding excess entropy difference per monomer. Dotted lines correspond to the best linear fit for each density.

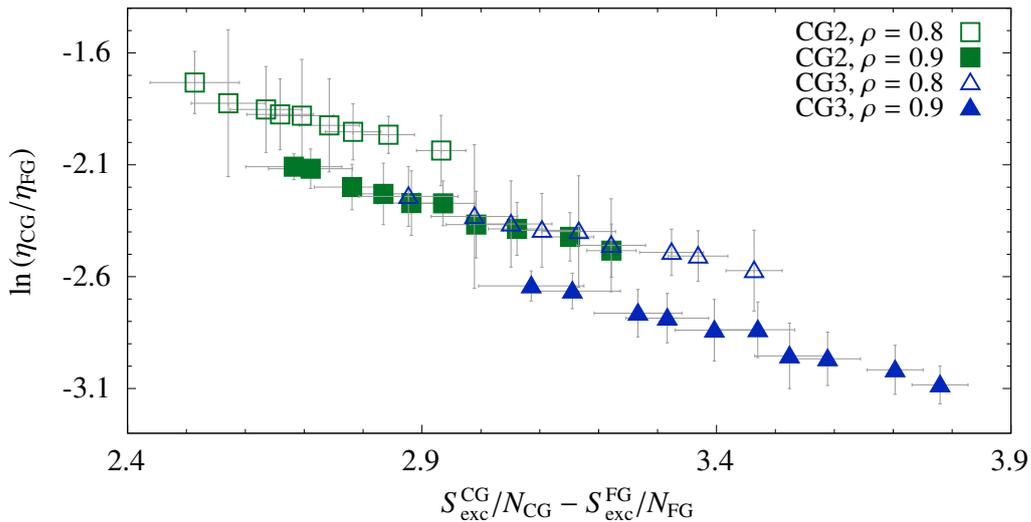


Figure 5.3.: Acceleration of the dynamics measured by the ratio of the viscosity at CG resolutions CG2 and CG3 with respect to the FG model, FG, versus the corresponding excess entropy difference per monomer.

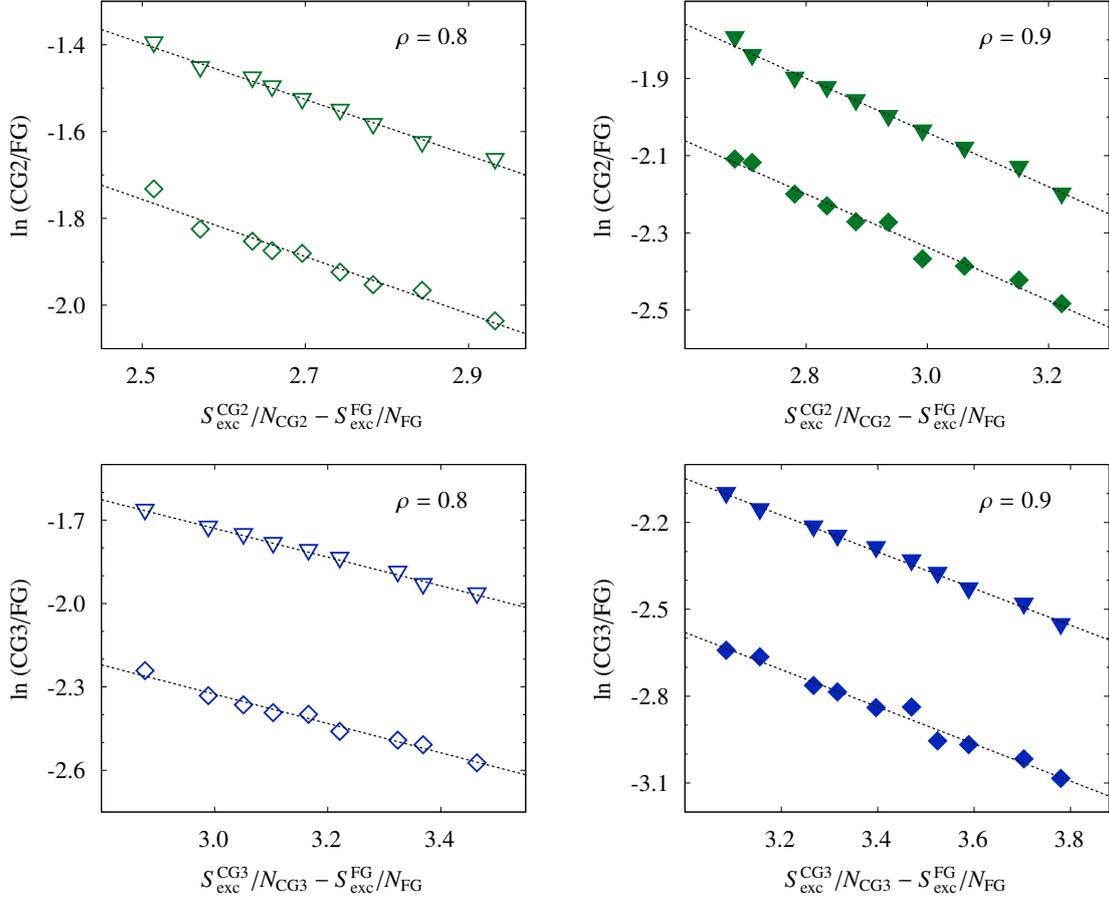


Figure 5.4.: Comparison of the acceleration of the local dynamics, as given by the ratio of bond reorientation times, and the global dynamics, as given by the ratio of viscosities, with respect to excess entropy differences for each density and CG resolution. Ratios for CG2 (CG3) with respect to FG are shown in the two top (bottom) panels. Dotted lines correspond to the best linear fit for each set of points.

Table 5.1.: Fit coefficients for the lines shown in Figure 5.4. Coefficients α and A were introduced in Equation (3.124).

ρ	Ratio	CG2		CG3	
		α	A	α	A
0.8	ϑ	-0.64 ± 0.02	1.23 ± 0.08	-0.51 ± 0.02	0.83 ± 0.09
	η	-0.65 ± 0.04	0.89 ± 0.11	-0.52 ± 0.03	0.47 ± 0.10
0.9	ϑ	-0.70 ± 0.04	1.07 ± 0.10	-0.63 ± 0.03	0.86 ± 0.11
	η	-0.69 ± 0.03	0.76 ± 0.11	-0.64 ± 0.04	0.52 ± 0.12

It is useful to visualize both ratios in the same plot, i.e., the acceleration for ϑ and η for each pair of resolution and density, as shown in Figure 5.4. The remarkable feature in those plots is that, for every system, both global and local dynamics scale similarly. The coefficients of the fits are shown in Table 5.1, following the notation introduced in Equation (3.124). The slope coefficients, α , are the same within the error for both dynamical ratios at each resolution for a particular density. The absolute values of these slopes increase with increasing density but decrease as the resolution is coarsened. Moreover, the intercept coefficients, $\ln A$, also present a systematic behavior: for CG2 the viscosity ratio coefficients are 72 % and 71 % lower than those of the bond reorientation times for densities 0.8 and 0.9, respectively, while for CG3 they are 56 % and 60 % lower for the same densities, respectively. The fact that a local descriptor of the short-time dynamics correlates with the changes in excess entropies in a very similar way as a global descriptor of the long-time dynamics reveals that excess entropy differences may capture a single dynamical behavior over different time scales. It also opens interesting possibilities in practical simulations, e.g., one may take a system at two resolutions, a finer and a coarser one, with the latter derived systematically from the former, and perform very short tandem simulations to probe for the short-time dynamics. From these short simulations, and using a computationally inexpensive method to approximate the excess entropy, one may derive a scaling factor of the global dynamics that allows mapping the CG dynamics back to the reference one, or even derive friction coefficients to be used in stochastic simulations to recover the proper dynamics.

5.3. Correlations Between Coarse-Grained Systems

This project was designed to analyze the acceleration caused by coarse-graining in the context of excess entropy differences when a CG system and the FG system upon which it is based are considered, as discussed in the previous session. It is interesting however to carry out a similar analysis looking at possible correlations when CG3 is considered with respect to CG2, as shown in Figure 5.5. Even though the error bars are significant in this case due to the close proximity of the values, it is clear that the same essential correlations as seen with respect to FG are still manifest. This is rather puzzling because not only the CG potentials for CG3 were not derived systematically from CG2, but the former cannot be structurally mapped to the latter in a straightforward manner. We speculate that such correlations are not mere artifacts of the CG procedure, but rather a fundamental relation between the loss of DOF, and thus of friction, and

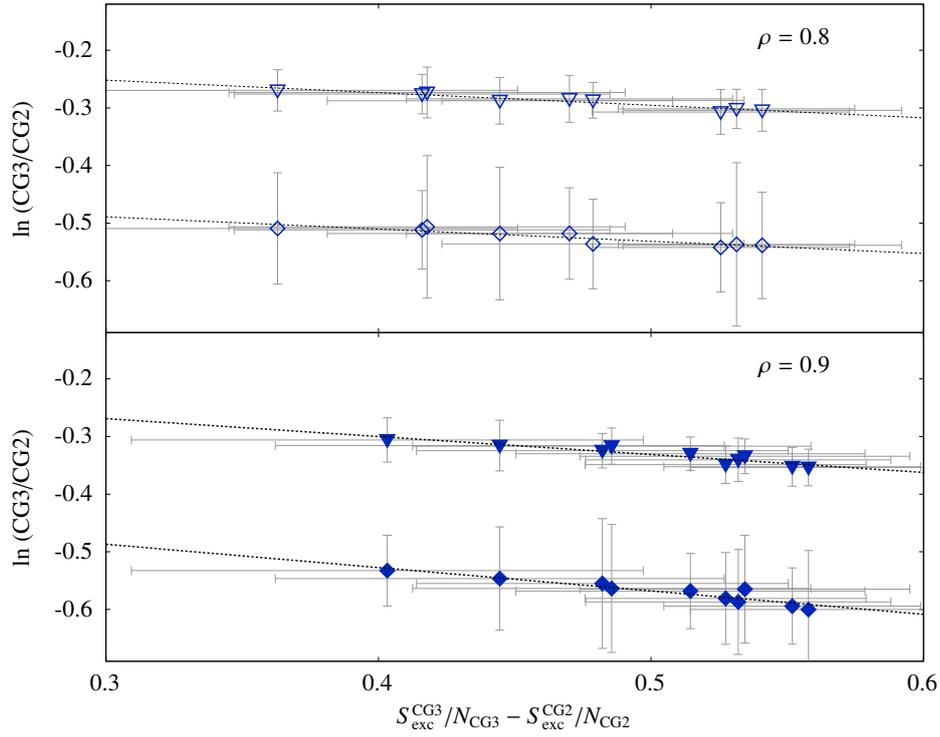


Figure 5.5.: Acceleration of the dynamics measured by the ratio of the bond reorientation times (upside-down triangles) and the viscosities (diamonds) for CG resolution CG3 with respect to the CG resolution CG2 versus the corresponding change in excess entropy per bead.

the corresponding gain in excess entropy per monomer. The slope coefficient, α , for the bond relaxation time and viscosity ratios in the case of density $\rho = 0.8$ are both equal to -0.21 ± 0.03 . For $\rho = 0.9$ they are -0.31 ± 0.04 and -0.40 ± 0.05 , respectively. The values for the intercept coefficient A within each density are different, as it is clear by a visual inspection of the plots, but are comparable within each single dynamic property. For the bond relaxation time ratios they are 0.82 ± 0.05 and 0.65 ± 0.07 for $\rho = 0.8$ and 0.9 , respectively, while for the viscosity ratios they are 0.83 ± 0.03 and 0.69 ± 0.05 for $\rho = 0.8$ and 0.9 , respectively.

5.4. Correlations Based on the Two-Body Entropy

The correlations for the ratios of ϑ and η when the difference in the two-body entropy across resolutions is used instead of exact excess entropy differences are shown in Figure 5.6. The linear behavior in the semi-log plots shows that the ratios are well described by a relation such as given by Equation (3.124),

with the coefficients determined from fitting shown in Table 5.2. The fact that the same mathematical form is found when S_2 is employed follows from the linear relation between S_{exc} and S_2 , and is encouraging given that this is a quantity that can be readily obtained from even short MD simulations. From Figure 5.6 we see that the correlations involving differences in S_2 appear more systematic and consistent across the two properties, not following the exact scaling seen Figure 5.2 and Figure 5.3. The offsets for different CG levels and different densities appear to be slightly smaller for the local dynamics (ϑ) and slightly larger for the global dynamics (η). Interestingly, the ratios A_ϑ/A_η for correlations involving S_2 are very similar to those obtained from the correlations involving S_{exc} , as shown in Table 5.3, supporting the notion that S_2 , approximate as it is, captures well the behavior of the excess entropy for the purposes of predicting the acceleration in dynamics.

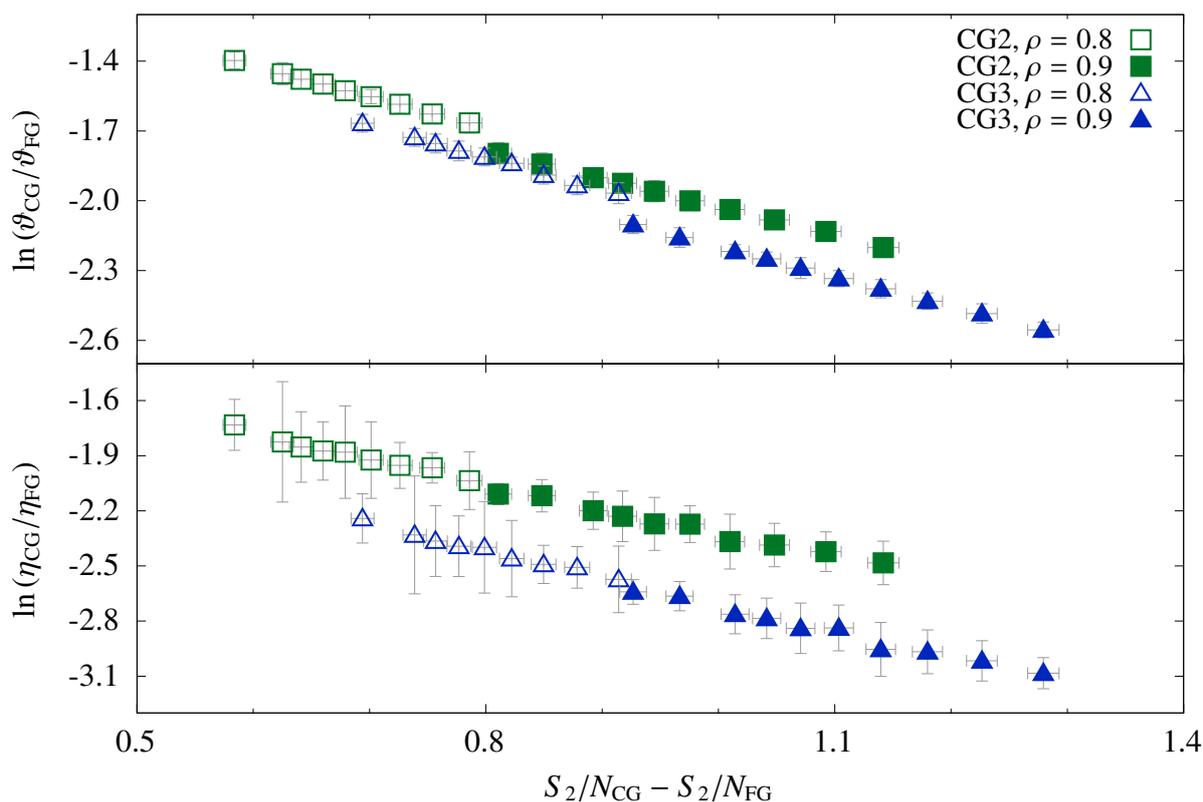


Figure 5.6.: Acceleration of the dynamics measured by the ratio of the bond reorientation time (top) and ratio of viscosity (bottom) at CG resolutions CG2 and CG3 with respect to the FG model versus the corresponding change in the two-body entropy per monomer.

Table 5.2.: Fit coefficients for the data points shown in Figure 5.6. Coefficients α and A were introduced in Equation (3.124).

ρ	Ratio	CG2		CG3	
		α	A	α	A
0.8	ϑ	-1.32 ± 0.01	0.53 ± 0.03	-1.39 ± 0.02	0.49 ± 0.08
	η	-1.35 ± 0.08	0.38 ± 0.12	-1.42 ± 0.05	0.28 ± 0.12
0.9	ϑ	-1.20 ± 0.01	0.44 ± 0.03	-1.27 ± 0.01	0.39 ± 0.03
	η	-1.18 ± 0.05	0.32 ± 0.11	-1.29 ± 0.06	0.23 ± 0.10

Table 5.3.: Comparison of the ratio A_{ϑ}/A_{η} for different densities and using the coefficients obtained from curves fitted to data sets representing the correlation between dynamical acceleration and different excess entropy estimates. Coefficient A was introduced in Equation (3.124). Values of A_{ϑ} and A_{η} employed to calculate the ratios were originally reported in Table 5.1 and Table 5.2.

ρ	CG2		CG3	
	S_{exc}	S_2	S_{exc}	S_2
0.8	1.38 ± 0.13	1.76 ± 0.13	1.39 ± 0.12	1.75 ± 0.14
0.9	1.41 ± 0.14	1.65 ± 0.16	1.38 ± 0.11	1.69 ± 0.10

5.5. Correlations Involving Renormalized Dynamical Properties

The empirical relations between the ratios of dynamical quantities and excess entropy differences discussed in the previous section have not included any sort of reduction of the dynamics. The Dzugutov relations are known to apply not only for simple liquids, but also for polymer melts of entangled and unentangled chains [78, 122]. In order to test how the Dzugutov scheme affects the correlations presented thus far, we have computed the ratios of the $\tilde{\vartheta}$ and $\tilde{\eta}$ for the CG resolutions with respect to the FG and correlated them to the difference in the two-body entropy between the corresponding resolutions, as shown in Figure 5.7. The choice of S_2 is motivated by the fact that Dzugutov's rules originally uses this approximation in place of the full excess entropy. The relation is of the form of Equation (3.124), but $\Delta S_2 \equiv S_2^{\text{CG}}/N_{\text{CG}} - S_2^{\text{FG}}/N_{\text{FG}}$ instead of ΔS_{exc} . An immediate appeal of the Dzugutov reduction is evident: the correlation for each CG resolution

and each property is described by a single density-independent master curve. The density independence is presumably brought about by the inclusion of the number density in the reduction coefficient Γ_E . Moreover, the ordering of the densities is preserved in the four curves, with lower densities corresponding to lower S_2 differences. Such systematic behavior is not seen in any of the non-reduced analyses presented in the previous sections. The slope coefficient, α , for the bond relaxation time ratio as a function of ΔS_2 is -2.08 ± 0.08 and -2.26 ± 0.09 for CG2 and CG3. In the case of the viscosity ratio, the slope coefficients for CG2 and CG3 are -0.87 ± 0.06 and -0.91 ± 0.07 . Therefore, for each particular ratio the slopes of both curves are approximately equal within the error. Interestingly, the ratios of the intercept coefficients, A_{CG3}/A_{CG2} , are 1.21 and 1.36 for $\tilde{\nu}$ and $\tilde{\eta}$, respectively, with errors of ca. 6%. This means that for the same ΔS_2 , both local and global dynamics of the two resolutions differ by approximately the same factor.

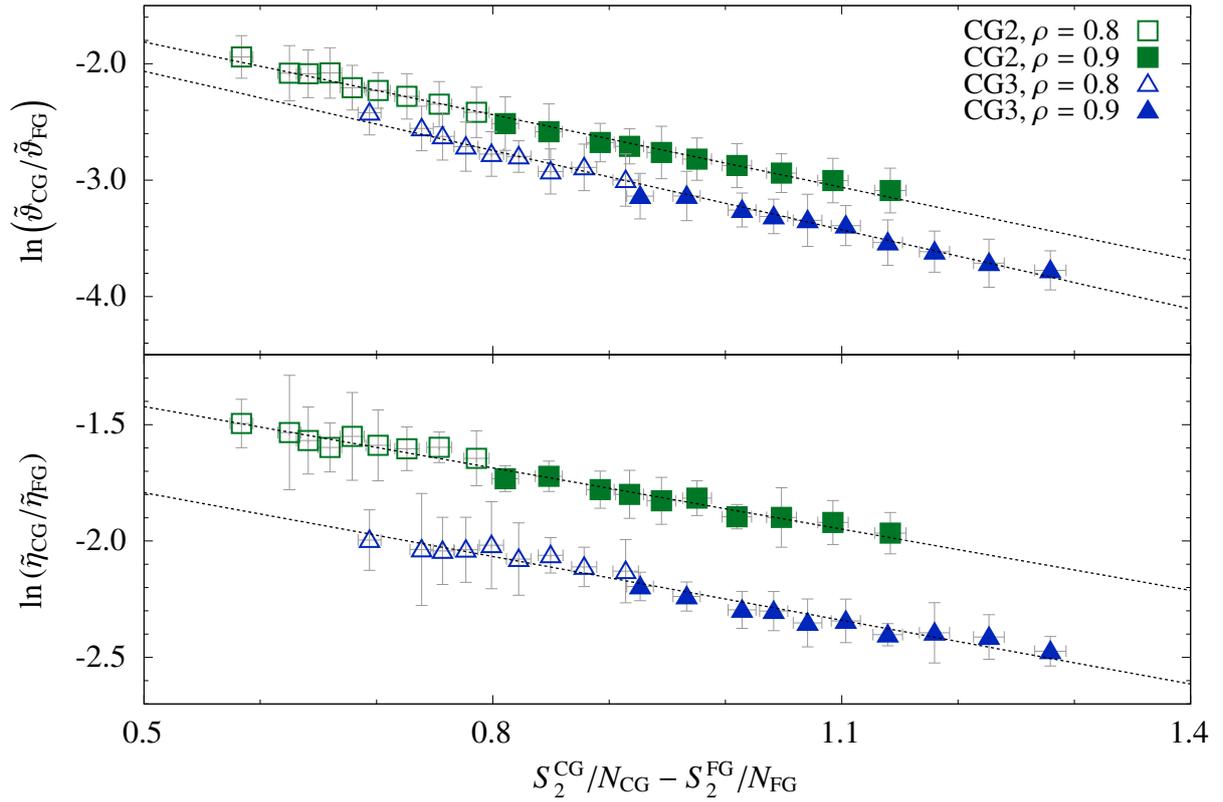


Figure 5.7.: Correlation between the ratios of the Dzugutov-renormalized dynamical properties and the difference in the two-body entropy per monomer for CG systems CG2 and CG3 with respect to the the FG model, FG. Ratio of renormalized bond reorientation times (top) and viscosities (bottom) are shown for densities $\rho = 0.8$ and 0.9 .

Another way of examining excess entropy differences, perhaps more natural even, is to consider its value per chain and not per monomer given that the number of chains is the same for all resolutions, in contrast to the number of monomers. If such normalization is employed within the Dzugutov's reduction of the viscosity and considering now the exact excess entropy differences and not the two-body approximation, then the quite interesting behavior shown in Figure 5.8 is observed. A single master curve captures the correlation for all state points and CG resolutions, with a fitting to the expression $2.34 \exp(-0.045 \Delta S_{\text{exc}})$ yielding a squared correlation coefficient $R^2 = 0.993$, with $\Delta S_{\text{exc}} = (S_{\text{exc}}^{\text{CG}} - S_{\text{exc}}^{\text{FG}}) / M$. It may very well be that this correlation is purely coincidental as it does not hold for the bond reorientation time, and thus it is difficult to provide a microscopic argument to justify this empirical behavior. This behavior deserves further investigation in the future.

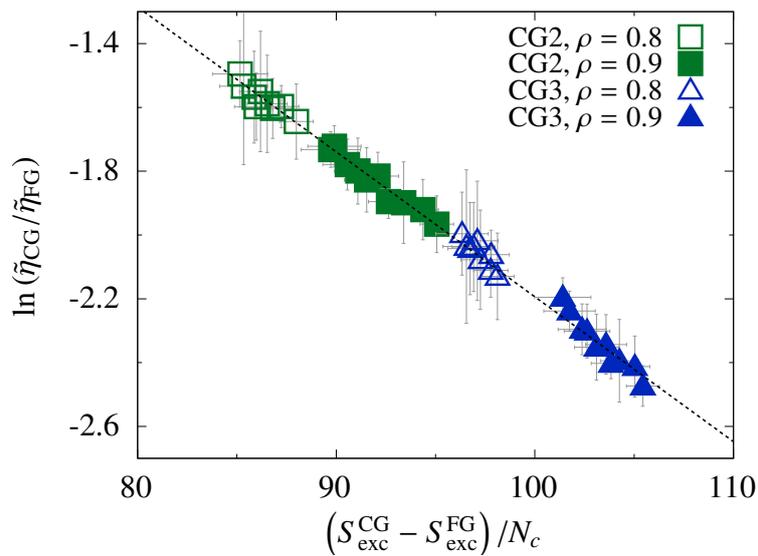


Figure 5.8.: Correlation between the ratios of the Dzugutov-renormalized viscosities and the difference in the exact excess entropy per chain for CG systems CG2 and CG3 with respect to the the FG.

6. Conclusions and Outlook

This work analyzed the connection between the acceleration of the dynamics brought by coarse-graining and the related changes in excess entropy across different CG resolutions. The FG system used as reference, FG, was a melt of unentangled LJC with 24 monomers which were coarse-grained using IBI into two lower-resolution models, CG2 and CG3, containing 12 and 8 monomers per chain, respectively. The bond reorientation times were selected as descriptors of the short-time local dynamics, and the viscosity and self-diffusion coefficients as descriptors of the long-time global dynamics. These properties were calculated from the raw trajectories obtained in extensive canonical MD simulations using well known methods. The acceleration of the CG dynamics was quantified by computing the ratio of the relevant dynamical quantities for a coarser resolution with respect to those of a finer resolution. The exact thermodynamic excess entropy was obtained by carrying out a detailed two-step TI procedure that considered a sufficiently dense integration mesh based on reference states described by exact statistical mechanical formulations. Care was taken in terms of consistency throughout all calculations so that meaningful comparison of the resulting quantities among the different resolutions was possible, particularly for entropic quantities. In particular, the TI procedure followed the same intermediate steps and the exact same methodology was applied to all resolutions in order to exclude artifacts introduced by different approaches.

The observed acceleration of the dynamics in the CG models was as already expected and documented in the literature with coarser models presenting higher mobility, faster dynamics, and overall quicker relaxation. The fact that an ordering of the dynamics was kept following the same ordering of the resolution level adds a level of confidence in the methodology. Perhaps counter-intuitive is the fact that the absolute entropy per monomer increases in coarser resolutions, a consequence of removing hard DOF that contribute little to the entropy in the high-resolution reference system. Relations of the form $X_{CG}/X_{FG} = A \exp(\alpha \Delta S_{exc})$ were observed, for X being either the bond reorientation time, ϑ , or the viscosity, η . The same slope

coefficient, α , is found for both properties at each density and resolution. The bond reorientation times follow a master curve for each density, which is independent of CG resolution, a feature not observed for the viscosity. That relation holds even when comparing one CG resolution with another, suggesting a general behavior upon coarse-graining that connects the change in excess entropy with the friction loss and smoothing of the potential energy surface. Analogous behavior is found using a much simpler approximation to the excess entropy, namely the two-body entropy, S_2 . Using the Dzugutov relations within this approximation is advantageous to renormalize the dynamical properties before considering their ratios. Such renormalization leads to density-independent behavior of the exponential relation for bond reorientation times and also for the viscosities. When using the Dzugutov formalism with the exact excess entropy normalized per chain and not the two-body approximation, a single master curve valid for all densities and resolutions considered here was observed.

The results presented herein for generic bead-spring polymer models at different resolutions confirm that it is possible to cast the effects of coarse-graining in terms of excess entropy differences. A fact that had been verified only for fluids of small molecules [34, 35] is now confirmed for chain molecules. One advantage of the approach proposed in this thesis is the simplicity of the principle behind it. Given a FG model and a CG model derived from it, it may be sufficient to perform a few molecular dynamics simulations to obtain estimates of different dynamical properties and then compare their ratios to these properties to excess entropy differences. These, in turn, could be computed using different levels of approximation. However, it is fair to point out that obtaining the dynamical properties from the FG model in order to relate the ratio of acceleration of the dynamics to ΔS_{exc} may not be possible at all, e.g., in the case of entangled, high-molecular-weight polymer melts. A combined approach may be required in which not only excess entropy is taken into account, but also mechanical properties [12]. Moreover, one possible route of exploration is by employing adaptive resolution simulations [155], in which different regions of a system are simulated at different resolutions. One advantage of such a simulation is that the difference free energy between the different phases of the system is readily available from the method, which could potentially make the assessment of the excess entropy difference significantly easier.

Obtaining systematic and more general relations similar to those presented here requires experimenting with more CG resolutions and a more comprehensive range of temperatures and densities. One fundamental step is to extend the work to realistic entangled models of polymers, which will provide insights whether or

not the significantly slower dynamics of entangled melts still correlate with the excess entropy differences, and if losing entanglements due to coarse-graining would lead to significant time-scale separations. This work, however, offers a possible direction of exploration employing empirical approaches to predict the loss of friction due to coarse-graining by analyzing the behavior of the actual excess entropy or approximations to it. We hope that it helps to pave the way for simple and cost-effective methods to renormalize the transport coefficients of CG systems back to the atomistic realm, thereby making such calculations predictive. Moreover, the computational experiments presented here contribute evidence for theoretical arguments establishing the scaling of coarse-graining errors in dynamic quantities [34], a subject that has not been fully explored.

References

- (1) Noid, W. G. Perspective: Coarse-grained models for biomolecular systems. *Journal of Chemical Physics* **2013**, *139*, 090901.
- (2) Brini, E.; Algaer, E. A.; Ganguly, P.; Li, C.; Rodríguez-Ropero, F.; van der Vegt, N. F. A. Systematic coarse-graining methods for soft matter simulations - A review. *Soft Matter* **2013**, *9*, 2108–2119.
- (3) Salerno, K. M.; Agrawal, A.; Perahia, D.; Grest, G. S. Resolving dynamic properties of polymers through coarse-grained computational studies. *Physical Review Letters* **2016**, *116*, 058302.
- (4) Depa, P.; Chen, C.; Maranas, J. K. Why are coarse-grained force fields too fast? A look at dynamics of four coarse-grained polymers. *Journal of Chemical Physics* **2011**, *134*, 014903.
- (5) Li, Y.; Abberton, B. C.; Kröger, M.; Liu, W. K. Challenges in multiscale modeling of polymer dynamics. *Polymers* **2013**, *5*, 751–832.
- (6) Fritz, D.; Koschke, K.; Harmandaris, V. A.; van der Vegt, N. F. A.; Kremer, K. Multiscale modeling of soft matter: scaling of dynamics. *Physical Chemistry Chemical Physics* **2011**, *13*, 10412–10420.
- (7) Yamamoto, U.; Schweizer, K. S. Theory of entanglements and tube confinement in rod–sphere nanocomposites. *ACS Macro Letters* **2013**, *2*, 955–959.
- (8) Milano, G.; Müller-Plathe, F. Mapping atomistic simulations to mesoscopic models: A systematic coarse-graining procedure for vinyl polymer chains. *Journal of Physical Chemistry B* **2005**, *109*, 18609–18619.
- (9) Harmandaris, V. A.; Kremer, K. Dynamics of polystyrene melts through hierarchical multiscale simulations. *Macromolecules* **2009**, *42*, 791–802.

-
- (10) Harmandaris, V. A.; Kremer, K. Predicting polymer dynamics at multiple length and time scales. *Soft Matter* **2009**, *5*, 3920–3926.
 - (11) Peters, B. L.; Salerno, K. M.; Agrawal, A.; Perahia, D.; Grest, G. S. Coarse-grained modeling of polyethylene melts: Effect on dynamics. *Journal of Chemical Theory and Computation* **2017**, *13*, 2890–2896.
 - (12) Meinel, M. K.; Müller-Plathe, F. Loss of molecular roughness upon coarse-graining predicts the artificially accelerated mobility of coarse-grained molecular simulation models. *Journal of Chemical Theory and Computation* **2020**, *16*, 1411–1419.
 - (13) Lyubimov, I. Y.; McCarty, J.; Clark, A.; Guenza, M. G. Analytical rescaling of polymer dynamics from mesoscale simulations. *Journal of Chemical Physics* **2010**, *132*, 224903.
 - (14) Lyubimov, I.; Guenza, M. G. First-principle approach to rescale the dynamics of simulated coarse-grained macromolecular liquids. *Physical Review* **2011**, *84*, 031801.
 - (15) Lyubimov, I. Y.; Guenza, M. G. Theoretical reconstruction of realistic dynamics of highly coarse-grained cis-1,4-polybutadiene melts. *Journal of Chemical Physics* **2013**, *138*, 12A546.
 - (16) Markutsya, S.; Lamm, M. H. A coarse-graining approach for molecular simulation that retains the dynamics of the all-atom reference system by implementing hydrodynamic interactions. *Journal of Chemical Physics* **2014**, *141*, 174107.
 - (17) Langeloth, M.; Masubuchi, Y.; Böhm, M. C.; Müller-Plathe, F. Recovering the reptation dynamics of polymer melts in dissipative particle dynamics simulations via slip-springs. *Journal of Chemical Physics* **2013**, *138*, 104907.
 - (18) Xia, W.; Hansoge, N. K.; Xu, W.-S.; Phelan, F. R.; Keten, S.; Douglas, J. F. Energy renormalization for coarse-graining polymers having different segmental structures. *Science Advances* **2019**, *5*, eaav4683.
 - (19) Xia, W.; Song, J.; Jeong, C.; Hsu, D. D.; Phelan, F. R.; Douglas, J. F.; Keten, S. Energy-renormalization for achieving temperature transferable coarse-graining of polymer dynamics. *Macromolecules* **2017**, *50*, 8787–8796.
 - (20) Rudzinski, J. F. Recent progress towards chemically-specific coarse-grained simulation models with consistent dynamical properties. *Computation* **2019**, *7*, 42.

-
- (21) Klippenstein, V.; Tripathy, M.; Jung, G.; Schmid, F.; van der Vegt, N. F. A. Introducing memory in coarse-grained molecular simulations. *Journal of Physical Chemistry B* **2021**, *125*, 4931–4954.
- (22) Guenza, M. Thermodynamic consistency and other challenges in coarse-graining models. *The European Physical Journal Special Topics* **2015**, *224*, 2177–2191.
- (23) Riniker, S.; van Gunsteren, W. F. Mixing coarse-grained and fine-grained water in molecular dynamics simulations of a single system. *Journal of Chemical Physics* **2012**, *137*, 044120.
- (24) Riniker, S.; Alison, J. R.; van Gunsteren, W. F. On developing coarse-grained models for biomolecular simulation: A review. *Physical Chemistry Chemical Physics* **2012**, *14*, 12423–12430.
- (25) Rosenfeld, Y. Relation between the transport coefficients and the internal entropy of simple systems. *Physical Review A* **1977**, *15*, 2545–2549.
- (26) Rosenfeld, Y. A quasi-universal scaling law for atomic transport in simple fluids. *Journal of Physics: Condensed Matter*, *11*, 5415.
- (27) Dyre, J. C. Perspective: Excess-entropy scaling. *Journal of Chemical Physics* **2018**, *149*, 210901.
- (28) Abramson, E. H. Viscosity of argon to 5 GPa and 673 K. *High Pressure Research* **2011**, *31*, 544–548.
- (29) Vaz, R. V.; Magalhaes, A. L.; Fernandes, D. L. A.; Silva, C. M. Universal correlation of self-diffusion coefficients of model and real fluids based on residual entropy scaling law. *Chemical Engineering Science* **2012**, *79*, 153–162.
- (30) Ma, X.; Chen, W.; Wang, Z.; Peng, Y.; Han, Y.; Tong, P. Test of the universal scaling law of diffusion in colloidal monolayers. *Physical Review Letters* **2013**, *110*, 078302.
- (31) Bell, I. H. Probing the link between residual entropy and viscosity of molecular fluids and model potentials. *Proceedings of the National Academy of Sciences* **2019**, *116*, 4070–4079.
- (32) Chaimovich, A.; Shell, M. S. Coarse-graining errors and numerical optimization using a relative entropy framework. *Journal of Chemical Physics* **2011**, *134*, 094112.
- (33) Shell, M. S. In *Advances in Chemical Physics*, Rice, S. A., Dinner, A. R., Eds.; John Wiley & Sons, Ltd: 2016; Vol. 161, pp 395–441.
- (34) Shell, M. S. Systematic coarse-graining of potential energy landscapes and dynamics in liquids. *Journal of Chemical Physics* **2012**, *137*, 084503.

-
- (35) Armstrong, J. A.; Chakravarty, C.; Ballone, P. Statistical mechanics of coarse graining: Estimating dynamical speedups from excess entropies. *Journal of Chemical Physics* **2012**, *136*, 124503.
- (36) Rondina, G. G.; Böhm, M. C.; Müller-Plathe, F. Predicting the mobility increase of coarse-grained polymer models from excess entropy differences. *Journal of Chemical Theory and Computation* **2020**, *16*, 1431–1447.
- (37) Rahman, A. Correlations in the motion of atoms in liquid argon. *Physical Review* **1964**, *136*, A405–A411.
- (38) Harp, G. D.; Berne, B. J. Linear- and angular-momentum autocorrelation functions in diatomic liquids. *Journal of Chemical Physics* **1968**, *49*, 1249–1254.
- (39) Harp, G. D.; Berne, B. J. Time-correlation functions, memory functions, and molecular dynamics. *Physical Review A* **1970**, *2*, 975–996.
- (40) Zurek, W. H.; Schieve, W. C. Pattern recognition in molecular dynamics. *Computer Physics Communications* **1977**, *13*, 75–79.
- (41) Bishop, M.; Kalos, M. H.; Frisch, H. L. Molecular dynamics of polymeric systems. *Journal of Chemical Physics* **1979**, *70*, 1299–1304.
- (42) Warshel, A.; Levitt, M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *Journal of Molecular Biology* **1976**, *103*, 227–249.
- (43) Car, R.; Parrinello, M. Unified approach for molecular dynamics and density-functional theory. *Physical Review Letters* **1985**, *55*, 2471–2474.
- (44) Kresse, G.; Hafner, J. Ab initio molecular dynamics for liquid metals. *Physical Review B* **1993**, *47*, 558–561.
- (45) Hafner, J. Ab-initio simulations of materials using VASP: Density-functional theory and beyond. *Journal of Computational Chemistry* **2008**, *29*, 2044–2078.
- (46) Senn, H. M.; Thiel, W. QM/MM methods for biomolecular systems. *Angewandte Chemie International Edition* **2009**, *48*, 1198–1229.

-
- (47) Blum, V.; Gehrke, R.; Hanke, F.; Havu, P.; Havu, V.; Ren, X.; Reuter, K.; Scheffler, M. Ab initio molecular simulations with numeric atom-centered orbitals. *Computer Physics Communications* **2009**, *180*, 2175–2196.
- (48) Allen, M. P.; Tildesley, D. J., *Computer Simulation of Liquids*; Oxford University Press: 2017.
- (49) Haile, J. M., *Molecular Dynamics Simulation: Elementary Methods*; John Wiley & Sons, Inc.: 1992.
- (50) Leach, A., *Molecular Modelling: Principles and Applications*; Pearson: 2001.
- (51) Frenkel, D.; Smit, B., *Understanding Molecular Simulation: From Algorithms to Applications*; Academic Press: 2001.
- (52) Verlet, L. Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Physical Review* **1967**, *159*, 98–103.
- (53) Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *Journal of Chemical Physics* **1982**, *76*, 637–649.
- (54) Braun, E.; Moosavi, S. M.; Smit, B. Anomalous effects of velocity rescaling algorithms: The flying ice cube effect revisited. *Journal of Chemical Theory and Computation* **2018**, *14*, 5262–5272.
- (55) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *Journal of Chemical Physics* **1984**, *81*, 3684–3690.
- (56) Hünenberger, P. H. In *Advanced Computer Simulation: Approaches for Soft Matter Sciences I*, Holm, C., Kremer, K., Eds.; Springer: 2005; Vol. 173, pp 105–149.
- (57) Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics* **1984**, *52*, 255–268.
- (58) Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *Journal of Chemical Physics* **1984**, *81*, 511–519.
- (59) Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A* **1985**, *31*, 1695–1697.
- (60) Hockney, R. W.; Eastwood, J. W., *Computer Simulation Using Particles*; Routledge: 1988.

-
- (61) Sutton, A. P.; Chen, J. Long-range Finnis–Sinclair potentials. *Philosophical Magazine Letters* **1990**, *61*, 139–146.
- (62) Mishin, Y.; Farkas, D.; Mehl, M. J.; Papaconstantopoulos, D. A. Interatomic potentials for monoatomic metals from experimental data and ab initio calculations. *Physical Review B* **1999**, *59*, 3393–3407.
- (63) Tersoff, J. Modeling solid-state chemistry: Interatomic potentials for multicomponent systems. *Physical Review B* **1989**, *39*, 5566–5568.
- (64) Rajasekaran, G.; Kumar, R.; Parashar, A. Tersoff potential with improved accuracy for simulating graphene in molecular dynamics environment. *Materials Research Express* **2016**, *3*, 035011.
- (65) Nevins, D.; Spera, F. J. Accurate computation of shear viscosity from equilibrium molecular dynamics simulations. *Molecular Simulation* **2007**, *33*, 1261–1266.
- (66) Kestin, J. Experimental determination of the viscosity and thermal conductivity of fluids. *Physics and Chemistry of the Earth* **1981**, *13-14*, 295–319.
- (67) Son, Y. Measurement of pressure dependence on the shear viscosity of polymer melts. *Journal of Polymer Research* **2009**, *16*, 667.
- (68) Hess, S.; Evans, D. J. Computation of the viscosity of a liquid from time averages of stress fluctuations. *Physical Review E* **2001**, *64*, 011207.
- (69) Hess, B. Determining the shear viscosity of model liquids from molecular dynamics simulations. *Journal of Chemical Physics* **2002**, *116*, 209–217.
- (70) Davis, P. J.; Evans, D. J. Comparison of constant pressure and constant volume nonequilibrium simulations of sheared model decane. *Journal of Chemical Physics* **1994**, *100*, 541–547.
- (71) Mondello, M.; Grest, G. S. Viscosity calculations of n-alkanes by equilibrium molecular dynamics. *Journal of Chemical Physics* **1997**, *106*, 9327–9336.
- (72) Chen, T.; Smit, B.; Bell, A. T. Are pressure fluctuation-based equilibrium methods really worse than nonequilibrium methods for calculating viscosities? *Journal of Chemical Physics* **2009**, *131*, 246101.

-
- (73) Zhang, Y.; Otani, A.; Maginn, E. J. Reliable viscosity calculation from equilibrium molecular dynamics simulations: A time decomposition method. *Journal of Chemical Theory and Computation* **2015**, *11*, 3537–3546.
- (74) Müller-Plathe, F. Reversing the perturbation in nonequilibrium molecular dynamics: An easy way to calculate the shear viscosity of fluids. *Physical Review E* **1999**, *59*, 4894–4898.
- (75) Müller-Plathe, F.; Reith, D. Cause and effect reversed in non-equilibrium molecular dynamics: an easy route to transport coefficients. *Computational and Theoretical Polymer Science* **1999**, *9*, 203–209.
- (76) Hyde, P. D.; Waldow, D. A.; Ediger, M. D.; Kitano, T.; Ito, K. Local segmental dynamics of polyisoprene in dilute solution: picosecond holographic grating experiments. *Macromolecules* **1986**, *19*, 2533–2538.
- (77) Barbieri, A.; Campani, E.; Capaccioli, S.; Leporini, D. Molecular dynamics study of the thermal and the density effects on the local and the large-scale motion of polymer melts: Scaling properties and dielectric relaxation. *Journal of Chemical Physics* **2004**, *120*, 437–453.
- (78) Voyiatzis, E.; Müller-Plathe, F.; Böhm, M. C. Excess entropy scaling for the segmental and global dynamics of polyethylene melts. *Physical Chemistry Chemical Physics* **2014**, *16*, 24301–24311.
- (79) Tripodo, A.; Puosi, F.; Malvaldi, M.; Capaccioli, S.; Leporini, D. Coincident correlation between vibrational dynamics and primary relaxation of polymers with strong or weak Johari-Goldstein relaxation. *Polymers* **2020**, *12*, 761.
- (80) Harmandaris, V. A.; Adhikari, N. P.; van der Vegt, N. F. A.; Kremer, K. Hierarchical modeling of polystyrene: From atomistic to coarse-grained simulations. *Macromolecules* **2006**, *39*, 6708–6719.
- (81) Müller-Plathe, F. Coarse-graining in polymer simulation: From the atomistic to the mesoscopic scale and back. *ChemPhysChem* **2002**, *3*, 754–769.
- (82) Johnston, K.; Harmandaris, V. Hierarchical multiscale modeling of polymer–solid interfaces: Atomistic to coarse-grained description and structural and conformational properties of polystyrene–gold systems. *Macromolecules* **2013**, *46*, 5741–5750.
- (83) Peter, C.; Kremer, K. Multiscale simulation of soft matter systems – from the atomistic to the coarse-grained level and back. *Soft Matter* **2009**, *5*, 4357–4366.

-
- (84) Mullinax, J. W.; Noid, W. G. Extended ensemble approach for deriving transferable coarse-grained potentials. *Journal of Chemical Physics* **2009**, *131*, 104110.
- (85) Shell, M. S. The relative entropy is fundamental to multiscale and inverse thermodynamic problems. *Journal of Chemical Physics* **2008**, *129*, 144108.
- (86) Noid, W. G.; Chu, J.-W.; Ayton, G. S.; Krishna, V.; Izvekov, S.; Voth, G. A.; Das, A.; Andersen, H. C. The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models. *Journal of Chemical Physics* **2008**, *128*, 244114.
- (87) Reith, D.; Pütz, M.; Müller-Plathe, F. Deriving effective mesoscale potentials from atomistic simulations. *Journal of Computational Chemistry* **2003**, *24*, 1624–1636.
- (88) Rühle, V.; Junghans, C.; Lukyanov, A.; Kremer, K.; Andrienko, D. Versatile object-oriented toolkit for coarse-graining applications. *Journal of Chemical Theory and Computation* **2009**, *5*, 3211–3223.
- (89) Rosenberger, D.; Hanke, M.; van der Vegt, N. F. A. Comparison of iterative inverse coarse-graining methods. *The European Physical Journal Special Topics* **2016**, *225*, 1323–1345.
- (90) Lyubartsev, A. P.; Karttunen, M.; Vattulainen, I.; Laaksonen, A. On coarse-graining by the inverse Monte Carlo method: Dissipative particle dynamics simulations made to a precise tool in soft matter modeling. *Soft Materials* **2002**, *1*, 121–137.
- (91) Lyubartsev, A. P.; Naômé, A.; Vercauteren, D. P.; Laaksonen, A. Systematic hierarchical coarse-graining with the inverse Monte Carlo method. *Journal of Chemical Physics* **2015**, *143*, 243120.
- (92) Ercolessi, F.; Adams, J. B. Interatomic potentials from first-principles calculations: the force-matching method. *Europhysics Letters* **1994**, *26*, 583–588.
- (93) Brini, E.; Marcon, V.; van der Vegt, N. F. A. Conditional reversible work method for molecular coarse graining applications. *Physical Chemistry Chemical Physics* **2011**, *13*, 10468–10474.
- (94) Brini, E.; van der Vegt, N. F. A. Chemically transferable coarse-grained potentials from conditional reversible work calculations. *Journal of Chemical Physics* **2012**, *137*, 154113.
- (95) Pak, A. J.; Voth, G. A. Advances in coarse-grained modeling of macromolecular complexes. *Current Opinion in Structural Biology* **2018**, *52*, 119–126.

-
- (96) Baranyai, A.; Evans, D. J. Direct entropy calculation from computer simulation of liquids. *Physical Review A* **1989**, *40*, 3817–3822.
- (97) Heinz, L. P.; Grubmüller, H. Per|Mut: Spatially resolved hydration entropies from atomistic simulations. *Journal of Chemical Theory and Computation* **2021**, *17*, 2090–2098.
- (98) Ekberg, V.; Ryde, U. On the use of interaction entropy and related methods to estimate binding entropies. *Journal of Chemical Theory and Computation* **2021**, *17*, 5379–5391.
- (99) Peter, C.; Oostenbrink, C.; van Dorp, A.; van Gunsteren, W. F. Estimating entropies from molecular dynamics simulations. *Journal of Chemical Physics* **2004**, *120*, 2652–2661.
- (100) Schäfer, H.; Mark, A. E.; van Gunsteren, W. F. Absolute entropies from molecular dynamics simulation trajectories. *Journal of Chemical Physics* **2000**, *113*, 7809–7817.
- (101) Sun, T.; Xian, J.; Zhang, H.; Zhang, Z.; Zhang, Y. Two-phase thermodynamic model for computing entropies of liquids reanalyzed. *Journal of Chemical Physics* **2017**, *147*, 194505.
- (102) Beutler, T. C.; Mark, A. E.; van Schaik, R. C.; Gerber, P. R.; van Gunsteren, W. F. Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chemical Physics Letters* **1994**, *222*, 529–539.
- (103) Bennett, C. H. Efficient estimation of free energy differences from Monte Carlo data. *Journal of Computational Physics* **1976**, *22*, 245–268.
- (104) Kirkwood, J. G. Statistical mechanics of fluid mixtures. *Journal of Chemical Physics* **1935**, *3*, 300–313.
- (105) Zielkiewicz, J. Two-particle entropy and structural ordering in liquid water. *Journal of Physical Chemistry B* **2008**, *112*, 7810–7815.
- (106) Singh, M.; Dhabal, D.; Nguyen, A. H.; Molinero, V.; Chakravarty, C. Triplet correlations dominate the transition from simple to tetrahedral liquids. *Physical Review Letters* **2014**, *112*, 147801.
- (107) Lazaridis, T.; Karplus, M. Orientational correlations and entropy in liquid water. *Journal of Chemical Physics* **1996**, *105*, 4294–4316.
- (108) Baranyai, A.; Evans, D. J. Three-particle contribution to the configurational entropy of simple fluids. *Physical Review A* **1990**, *42*, 849–857.

-
- (109) Dzugutov, M. A universal scaling law for atomic diffusion in condensed matter. *Nature* **1996**, *381*, 137–139.
- (110) Bretonnet, J. L. Excess entropy scaling for the diffusion coefficient in expanded liquid metals. *Journal of Chemical Physics* **2004**, *120*, 11100–11106.
- (111) Dhabal, D.; Nguyen, A. H.; Singh, M.; Khatua, P.; Molinero, V.; Bandyopadhyay, S.; Chakravarty, C. Excess entropy and crystallization in Stillinger-Weber and Lennard-Jones fluids. *Journal of Chemical Physics* **2015**, *143*, 164512.
- (112) Berthier, L.; Ozawa, M.; Scalliet, C. Configurational entropy of glass-forming liquids. *Journal of Chemical Physics* **2019**, *150*, 160902.
- (113) Rosenfeld, Y. Comments on the transport coefficients of dense hard core systems. *Chemical Physics Letters* **1977**, *48*, 467–468.
- (114) Das, S. P. Fluctuating hydrodynamic models for supercooled liquids and development of long relaxation times. *Physical Review E* **1996**, *54*, 1715–1719.
- (115) Kaur, C.; Harbola, U.; Das, S. P. Nature of the entropy versus self-diffusivity plot for simple liquids. *Journal of Chemical Physics* **2005**, *123*, 034501.
- (116) Malvaldi, M.; Chiappe, C. Excess entropy scaling of diffusion in room-temperature ionic liquids. *Journal of Chemical Physics* **2010**, *132*, 244502.
- (117) Jabes, B. S.; Chakravarty, C. Relating composition, structural order, entropy and transport in multi-component molten salts. *Journal of Chemical Physics* **2012**, *136*, 144507.
- (118) Chopra, R.; Truskett, T. M.; Errington, J. R. On the use of excess entropy scaling to describe single-molecule and collective dynamic properties of hydrocarbon isomer fluids. *Journal of Physical Chemistry B* **2010**, *114*, 16487–16493.
- (119) Galliero, G.; Boned, C.; Fernández, J. Scaling of the viscosity of the Lennard-Jones chain fluid model, argon, and some normal alkanes. *Journal of Chemical Physics* **2011**, *134*, 064505.
- (120) Yokoyama, I. On Dzugutov's scaling law for atomic diffusion in condensed matter. *Physica B: Condensed Matter* **1999**, *269*, 244–248.

-
- (121) Goel, T.; Patra, C. N.; Mukherjee, T.; Chakravarty, C. Excess entropy scaling of transport properties of Lennard-Jones chains. *Journal of Chemical Physics* **2008**, *129*, 164904.
- (122) Voyiatzis, E.; Müller-Plathe, F.; Böhm, M. C. Do transport properties of entangled linear polymers scale with excess entropy? *Macromolecules* **2013**, *46*, 8710–8723.
- (123) Voyiatzis, E.; Böhm, M. C. An open-source implementation of a quasi harmonic approach to compute the intramolecular entropy of particle systems. *Computer Physics Communications* **2019**, *234*, 286–293.
- (124) Chapman, S.; Cowling, T. G., *The Mathematical Theory of Non-Uniform Gases*; University Press: Cambridge: 1970.
- (125) Miyazaki, K.; Srinivas, G.; Bagchi, B. The Enskog theory for transport coefficients of simple fluids with continuous potentials. *Journal of Chemical Physics* **2001**, *114*, 6276–6285.
- (126) Silmore, K. S.; Howard, M. P.; Panagiotopoulos, A. Z. Vapour-liquid phase equilibrium and surface tension of fully flexible Lennard-Jones chains. *Molecular Physics* **2017**, *115*, 320–327.
- (127) Kremer, K.; Grest, G. S. Dynamics of entangled linear polymer melts: A molecular-dynamics simulation. *Journal of Chemical Physics* **1990**, *92*, 5057–5086.
- (128) Hou, J.-X. Note: Determine entanglement length through monomer mean-square displacement. *Journal of Chemical Physics* **2017**, *146*, 026101.
- (129) MacDowell, L. G.; Blas, F. J. Surface tension of fully flexible Lennard-Jones chains: Role of long-range corrections. *Journal of Chemical Physics* **2009**, *131*, 074705.
- (130) Plimpton, S. Fast parallel algorithms for short-range molecular dynamics. *Journal of Computational Physics* **1995**, *117*, 1–19.
- (131) Thompson, A. P.; Aktulga, H. M.; Berger, R.; Bolintineanu, D. S.; Brown, W. M.; Crozier, P. S.; in 't Veld, P. J.; Kohlmeyer, A.; Moore, S. G.; Nguyen, T. D.; Shan, R.; Stevens, M. J.; Tranchida, J.; Trott, C.; Plimpton, S. J. LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Computer Physics Communications* **2022**, *271*, 108171.
- (132) Rossi, G.; Monticelli, L.; Puisto, S. R.; Vattulainen, I.; Ala-Nissila, T. Coarse-graining polymers with the MARTINI force-field: polystyrene as a benchmark case. *Soft Matter* **2011**, *7*, 698–708.

-
- (133) Winger, M.; Trzesniak, D.; Baron, R.; van Gunsteren, W. F. On using a too large integration time step in molecular dynamics simulations of coarse-grained molecular models. *Physical Chemistry Chemical Physics* **2009**, *11*, 1934–1941.
- (134) Jülich Supercomputing Centre JURECA: Modular supercomputer at Jülich Supercomputing Centre. *Journal of large-scale research facilities* **2018**, *4*.
- (135) In 't Veld, P. J.; Rutledge, G. C. Temperature-dependent elasticity of a semicrystalline interphase composed of freely rotating chains. *Macromolecules* **2003**, *36*, 7358–7365.
- (136) Polak, E.; Ribiere, G. Note sur la convergence de méthodes de directions conjuguées. *Mathematical Modelling and Numerical Analysis-Modélisation Mathématique et Analyse Numérique* **1969**, *3*, 35–43.
- (137) Nocedal, J.; Wright, S., *Numerical Optimization*; Springer-Verlag: 2006.
- (138) Tschöp, W.; Kremer, K.; Batoulis, J.; Bürger, T.; Hahn, O. Simulation of polymer melts. I. Coarse-graining procedure for polycarbonates. *Acta Polymerica* **1998**, *49*, 61–74.
- (139) Chitra, R.; Yashonath, S. Estimation of error in the diffusion coefficient from molecular dynamics simulations. *Journal of Physical Chemistry B* **1997**, *101*, 5437–5445.
- (140) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P., *Numerical Recipes 3rd Edition: The Art of Scientific Computing*; Cambridge University Press: 2007.
- (141) Doxastakis, M.; Theodorou, D. N.; Fytas, G.; Kremer, F.; Faller, R.; Müller-Plathe, F.; Hadjichristidis, N. Chain and local dynamics of polyisoprene as probed by experiments and computer simulations. *Journal of Chemical Physics* **2003**, *119*, 6883–6894.
- (142) Logotheti, G. E.; Theodorou, D. N. Segmental and chain dynamics of isotactic polypropylene melts. *Macromolecules* **2007**, *40*, 2235–2245.
- (143) Mazars, M. Statistical physics of the freely jointed chain. *Physical Review* **1996**, *53*, 6297–6319.
- (144) Grimus, W. 100th anniversary of the Sackur–Tetrode equation. *Annalen der Physik* **2013**, *525*, A32–A35.
- (145) Buelens, F. P.; Grubmüller, H. Linear-scaling soft-core scheme for alchemical free energy calculations. *Journal of Computational Chemistry* **2012**, *33*, 25–33.

-
- (146) Izvekov, S.; Voth, G. A. Multiscale coarse graining of liquid-state systems. *Journal of Chemical Physics* **2005**, *123*, 134105.
- (147) Giuffré, E.; Prestipino, S.; Saija, F.; Saitta, A. M.; Giaquinta, P. V. Entropy from correlations in TIP4P water. *Journal of Chemical Theory and Computation* **2010**, *6*, 625–636.
- (148) Gallo, P.; Rovere, M. Relation between the two-body entropy and the relaxation time in supercooled water. *Physical Review* **2015**, *91*, 012107.
- (149) Borzsák, I.; Baranyai, A. On the convergence of Green's entropy expansion. *Chemical Physics* **1992**, *165*, 227–230.
- (150) Doi, M.; Edwards, S. F., *The Theory of Polymer Dynamics*; Clarendon Press: 1988.
- (151) Harmandaris, V. A.; Mavrantzas, V. G.; Theodorou, D. N.; Kröger, M.; Ramírez, J.; Öttinger, H. C.; Vlassopoulos, D. Crossover from the rouse to the entangled polymer melt regime: signals from long, detailed atomistic molecular dynamics simulations, supported by rheological experiments. *Macromolecules* **2003**, *36*, 1376–1387.
- (152) Chremos, A.; Jeong, C.; Douglas, J. F. Influence of polymer architectures on diffusion in unentangled polymer melts. *Soft Matter* **2017**, *13*, 5778–5784.
- (153) Lötgering-Lin, O.; Gross, J. Group contribution method for viscosities based on entropy scaling using the perturbed-chain polar statistical associating fluid theory. *Industrial & Engineering Chemistry Research* **2015**, *54*, 7942–7952.
- (154) Rokni, H. B.; Moore, J. D.; Gupta, A.; McHugh, M. A.; Gavaises, M. Entropy scaling based viscosity predictions for hydrocarbon mixtures and diesel fuels up to extreme conditions. *Fuel* **2019**, *241*, 1203–1213.
- (155) Español, P.; Delgado-Buscalioni, R.; Everaers, R.; Potestio, R.; Donadio, D.; Kremer, K. Statistical mechanics of Hamiltonian adaptive resolution simulations. *Journal of Chemical Physics* **2015**, *142*, 064115.

A. Supplement for Theory and Methodology

In this chapter we have collected selected supplementary information relative to the theory and also to the methodology employed in this project.

A.1. Example of Mapping in Coarse-Graining

As an illustration of this formalism let us consider an FG system consisted of a linear chain of $N_{\text{FG}} = 4$ atoms as shown in Figure A.1. The positions of these FG atoms are

$$\mathbf{r}_1 = 1\hat{x} + 1\hat{y}, \quad (\text{A.126})$$

$$\mathbf{r}_2 = 2\hat{x} + 2\hat{y},$$

$$\mathbf{r}_3 = 3\hat{x} + 1\hat{y},$$

$$\mathbf{r}_4 = 4\hat{x} + 2\hat{y},$$

and the mapping operator is defined as

$$\mathbf{\Omega} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}, \quad (\text{A.127})$$

which is a 2×4 matrix, meaning that it defines $N_{CG} = 2$ CG beads out of the 4 FG atoms. Applying this operator to the FG coordinates results in

$$\begin{aligned}\mathbf{R}_1 &= \Omega_{11}\mathbf{r}_1 + \Omega_{12}\mathbf{r}_2 + \Omega_{13}\mathbf{r}_3 + \Omega_{14}\mathbf{r}_4 = \frac{1}{2}\mathbf{r}_1 + \frac{1}{2}\mathbf{r}_2 = \frac{3}{2}\hat{x} + \frac{3}{2}\hat{y}, \\ \mathbf{R}_2 &= \Omega_{21}\mathbf{r}_2 + \Omega_{22}\mathbf{r}_2 + \Omega_{23}\mathbf{r}_3 + \Omega_{24}\mathbf{r}_4 = \frac{1}{2}\mathbf{r}_3 + \frac{1}{2}\mathbf{r}_4 = \frac{7}{2}\hat{x} + \frac{3}{2}\hat{y},\end{aligned}\tag{A.128}$$

which are the coordinates of the two CG beads shown in Figure A.1. Assuming that each FG atom has unitary mass, the mass of CG bead 1 is

$$m_{1,CG} = \left(\frac{\Omega_{11}^2}{m_{1,FG}} + \frac{\Omega_{12}^2}{m_{2,FG}} \right)^{-1} = \left(\frac{1}{4} + \frac{1}{4} \right)^{-1} = \left(\frac{2}{4} \right)^{-1} = 2,\tag{A.129}$$

which is the same as the mass of CG bead 2.

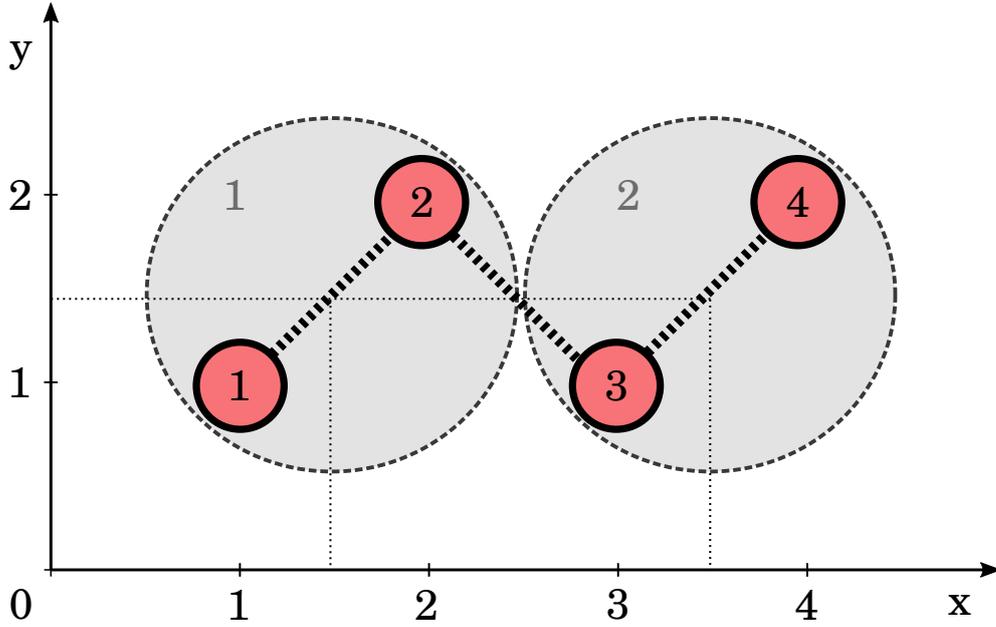


Figure A.1.: Two-dimensional fine-grained system with four atoms (red) and a coarse-grained system with two beads (shaded gray) defined in term of the fine-grained atoms and the respective coordinate system.

A.2. Absolute Entropy of the FJC

The equivalent Sackur-Tetrode equation for a gas of freely-jointed chains was derived by Mazars [143],

$$S_{\text{FJC}} = k_B N_c \left(N + \frac{3}{2} \right) + k_B \ln Z_{\text{FJC}} \quad (\text{A.130})$$

where k_B is Boltzmann's constant, N_c is the number of chains in the system, N is the number of monomers per chain, and Z_{FJC} is the partition function of the FJC system,

$$Z_{\text{FJC}} = \frac{1}{N_c!} \left[qQ \left(\frac{T}{T_0} \right)^N \right]^{N_c}, \quad (\text{A.131})$$

where Q is analogous to the partition function of a particle in a box of volume V ,

$$Q = \left(\frac{mk_B T}{4\pi^2 \hbar^2} \right)^{1/2} V, \quad (\text{A.132})$$

where m is the mass of the monomer and \hbar is $h/2\pi$, where h is Planck's constant. The factor q in Equation (A.131) is

$$q = \left(\frac{\pi}{2} \right)^{1/2} \frac{B}{\Gamma(1)^2}, \quad (\text{A.133})$$

where Γ is the gamma function and $B = 0.944$. The factor T_0 in Equation (A.131) is

$$T_0 = \frac{1}{k_B \beta_0}, \quad (\text{A.134})$$

where β_0 is given by

$$\beta_0 = A \left(\frac{1}{2} \right)^{7/2} \frac{\Gamma(1/2) m a^2}{\Gamma(3/2) \hbar^2}, \quad (\text{A.135})$$

where a is the bond length, i.e., $a = 1$ in this work, and $A = 1.051$. The factors A and B depend on the dimension of the system and the values given above are for a 3D system.

B. Building Lennard-Jones Chains

The package EMC [135] is a very powerful tool to build polymer systems. As the learning curve to it can be steep before fluent use is possible, an EMC input file for the fine-grained LJC system is provided in Listing B.1. EMC uses its own declarative scripting language and supports generating output data files that are compatible with LAMMPS [130, 131], making it very easy to integrate in an MD workflow. This input file was tested with EMC version 9.3.9.

```

1  (* EMC: Script *)
2
3  variables = {
4      nmolecules    ->    500,
5      nrepeat      ->    24,
6      dens         ->    0.8,
7      temp         ->    2.0,
8      ljcutoff     ->    5.0,
9      skinv        ->    1.0,
10     relaxradius  ->    2.0*(ljcutoff+skinv),
11     relaxiter    ->    200,
12     growiter     ->    1000,
13     growbonded   ->    50,
14     growtrials   ->    50,
15     mcsteps      ->    5000000,
16     mcprint      ->    1000
17 };
18
19 sites = {
20     site          ->    {
21         id         ->    LJA,
22         name       ->    "Lennard-Jones",
23         comment    ->    "Lennard-Jones atom",
24         mass       ->    1
25     }
26 };
27
28 groups = {
29     group         ->    {
30         id         ->    monomer,
31         terminator ->    true,
32         chemistry  ->    ">[LJA]<",
33         connects   ->    {
34             {head, {monomer, tail}},
35             {tail, {monomer, head}}
36         }
37     }
38 };
39
40 clusters = {
41     polymer       ->    {
42         id         ->    ljchain,
43         system     ->    main,
44         n          ->    nmolecules,
45         groups     ->    monomer,
46         nrepeat    ->    nrepeat
47     }
48 };
49
50 simulation = {
51     units         ->    {
52         type       ->    reduced,
53         reduced    ->    true,
54         seed       ->    -1
55     },
56     types         ->    {
57         merge      ->    true,
58         virial     ->    false,

```

```

59     neighbor -> sector,
60     stencil  -> standard,
61     skin     -> skinv,
62     standard -> {
63         bond -> {
64             active -> true,
65             data   -> {
66                 {
67                     i0 -> LJA,
68                     i1 -> LJA,
69                     k  -> 3000,
70                     l  -> 1.0
71                 }
72             }
73         },
74     angle -> { active -> false },
75     torsion -> { active -> false },
76     improper -> { active -> false },
77     pair -> {
78         active -> true,
79         nbonded -> 0,
80         mix -> none,
81         shift -> true,
82         cutoff -> ljcutoff,
83         mode -> global,
84         data -> {
85             {
86                 i0 -> LJA,
87                 i1 -> LJA,
88                 epsilon -> 1.0,
89                 sigma -> 1.0
90             }
91         }
92     }
93 }
94 };
95
96
97 build = {
98     system -> {
99         id -> main,
100        density -> dens,
101        temperature -> temp,
102        flag -> {
103            charge -> false,
104            map -> false
105        }
106    },
107    select -> {
108        progress -> list,
109        frequency -> 5,
110        message -> nkt,
111        center -> false,
112        order -> random,
113        check -> false,
114        cluster -> ljchain,
115        relax -> {
116            ncycles -> relaxiter,

```

```

117         radius      ->    relaxradius
118     },
119     name          ->    "error",
120     grow          ->    {
121         method      ->    energetic,
122         check       ->    all,
123         grace       ->    {0.9999, 0.9999},
124         niterations ->    growiter,
125         nbonded     ->    growbonded,
126         ntrials     ->    growtrials
127     }
128 }
129 };
130
131 focus = { clusters -> ljchain };
132
133 force = { style -> none, message -> nkt };
134 force = { style -> list, message -> nkt };
135 force = { style -> init, message -> nkt };
136
137 pdb = {
138     name          ->    "init_after_mc",
139     compress      ->    false,
140     system        ->    0,
141     forcefield    ->    standard,
142     cut           ->    false,
143     detect        ->    false
144 };
145
146 lammps = {
147     name          ->    "init_after_build",
148     compress      ->    false,
149     system        ->    0,
150     mode          ->    put,
151     forcefield    ->    standard,
152     atomistic     ->    false,
153     charges       ->    true,
154     ewald         ->    false,
155     bonds         ->    true,
156     types         ->    true,
157     parameters    ->    false
158 };
159
160 timing = { style -> reset };
161
162 restart = {
163     name          ->    "restart",
164     frequency     ->    1000
165 };
166
167
168 run = {
169     ncycles       ->    mcsteps,
170     nblocks       ->    mcprint,
171     cycle         ->    0
172 };
173
174 focus = { clusters -> ljchain };

```

```
175
176 force = { style -> none, message -> nkt };
177 force = { style -> list, message -> nkt };
178 force = { style -> init, message -> nkt };
179
180 timing = { style -> reset };
181
182 pdb = {
183     name      ->    "init_after_mc",
184     compress  ->    false,
185     system    ->    0,
186     forcefield ->    standard,
187     cut       ->    false,
188     detect    ->    false
189 };
190
191 lammps = {
192     name      ->    "init_after_mc",
193     compress  ->    false,
194     system    ->    0,
195     mode      ->    put,
196     forcefield ->    standard,
197     atomistic ->    false,
198     charges   ->    true,
199     ewald     ->    false,
200     bonds     ->    true,
201     types     ->    true,
202     parameters ->    false
203 };
```

Listing B.1: Input file for the EMC tool to build the fine-grained LJC system with 500 chains at density 0.8 and at temperature 2.0 reduced units.

C. Correlation Involving the Self-Diffusion Coefficients

In Figure C.1 it is shown the correlation between the acceleration as measured by the logarithm of $(D_{CG}/D_{FG})^{-1}$ and excess entropy differences. This is the equivalent of Figure 5.3 of the main text, but using the self-diffusion coefficient instead of the viscosity. The fit coefficients of Equation (3.124) for the self-diffusion ratios are shown in Table C.1, approximately matching those shown for the viscosity ratios in Table 5.1.

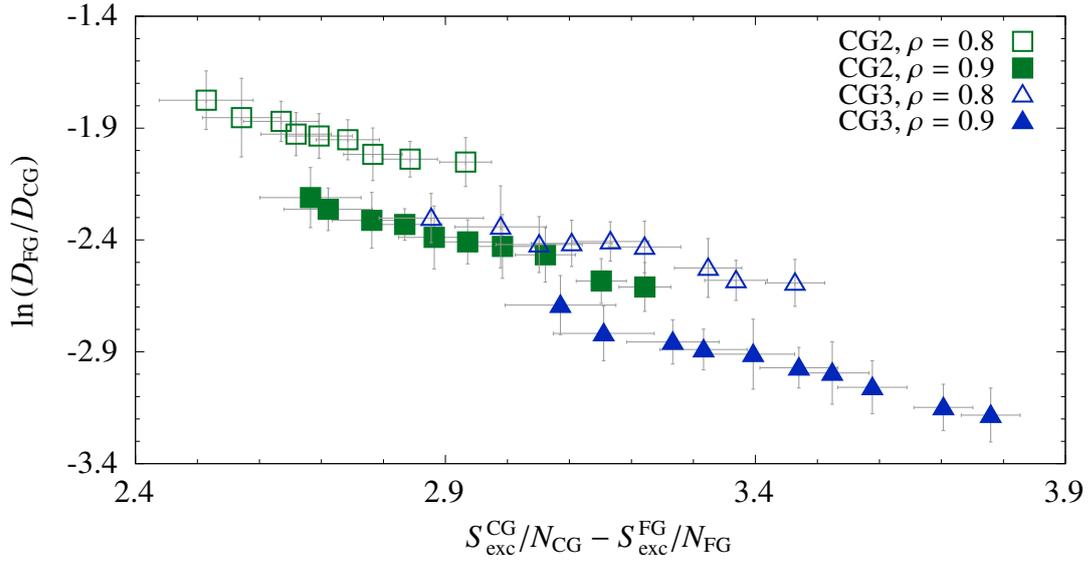


Figure C.1.: Acceleration of the dynamics measured by the ratio of the viscosity at coarse-grained resolutions CG2 and CG3 with respect to the fine-grained model, FG, versus the corresponding excess entropy difference per monomer.

Table C.1.: Fit coefficients for data sets shown in Figure C.1. Coefficients α and A were introduced in Equation (3.124) of the main text.

ρ	CG2		CG3	
	α	A	α	A
0.8	-0.67 ± 0.06	0.89 ± 0.21	-0.51 ± 0.05	0.43 ± 0.15
0.9	-0.71 ± 0.04	0.71 ± 0.18	-0.65 ± 0.03	0.49 ± 0.12

D. List of Publications

The following publication directly relates to the contents of this thesis and features part of the results presented here.

- Rondina, G. G.; Böhm, M. C.; Müller-Plathe, F. Predicting the mobility increase of coarse-grained polymer models from excess entropy differences. *Journal of Chemical Theory and Computation* **2020**, *16*, 1431-1447.

The following publications were co-authored with research collaborators during my PhD studies. My contributions include writing and revising parts of the texts, software development, design and execution of simulation experiments.

- Cezar, H. M.; Rondina, G. G.; Da Silva, J. L. F. Thermodynamic properties of 55-atom Pt-based nanoalloys: Phase changes and structural effects on the electronic properties. *Journal of Chemical Physics* **2019**, *151*, 204301.
- Schätti, J.; Rieser, P.; Sezer, U.; Richter, G.; Geyer, P.; Rondina, G. G.; Häussinger, D.; Mayor, M.; Shayeghi, A.; Köhler, V.; Arndt, M. Pushing the mass limit for intact launch and photoionization of large neutral biopolymers. *Communications Chemistry* **2018**, *1*, 93.
- Cezar, H. M.; Rondina, G. G.; Da Silva, J. L. F. Parallel tempering Monte Carlo combined with clustering Euclidean metric analysis to study the thermodynamic stability of Lennard-Jones nanoclusters. *Journal of Chemical Physics* **2017**, *146*, 064114.
- Souza, D. G. D.; Cezar, H. M.; Rondina, G. G.; Oliveira, M. F. d.; Da Silva, J. L. F. A basin-hopping Monte Carlo investigation of the structural and energetic properties of 55- and 561-atom bimetallic nanoclusters: the examples of the ZrCu, ZrAl, and CuAl systems. *Journal of Physics: Condensed Matter* **2016**, *28*, 175302.