The Journal of Chemical Physics

Coarse-graining of polyisoprene melts using inverse Monte Carlo and local density potentials

Nobahar Shahidi,[1] Antonis Chazirakis,[2] Vagelis A. Harmandaris,[2,3] and Manolis Doxastakis[1,a)]

[1)] *Department of Chemical and Biomolecular Engineering, University of Tennessee, Knoxville, Tennessee 37996, United States*

[2)] *Department of Mathematics and Applied Mathematics, University of Crete, Heraklion, GR-71110, Greece*

[3)] *Institute of Applied and Computational Mathematics, Foundation for Research and Technology - Hellas, GR-71110 Heraklion, Crete, Greece*

Bottom-up coarse-graining of polymers is commonly performed by matching structural order parameters such as distribution of bond lengths, bending and dihedral angles and pair distribution functions. In this study, we introduce the distribution of nearest-neighbors as an additional order parameter in the concept of local density potentials. We describe how the inverse-Monte Carlo method provides a framework for forcefield development that is capable of overcoming challenges associated with the parametrization of interaction terms in polymer systems. The technique is applied on polyisoprene melts as a prototype system. We demonstrate that while different forcefields can be developed that perform equally in terms of matching target distributions, the inclusion of nearest-neighbors provides a facile route to match both thermodynamic and conformational properties. We find that several temperature state points can also be addressed provided that the forcefield is refined accordingly. Finally, we examine both the single-particle as well as the collective dynamics of the coarse-grain models, demonstrating that all forcefields present a similar acceleration relative to the atomistic systems.

Keywords: Polyisoprene, Multi-scale modeling, Nearest-neighbor potentials, Inverse Monte Carlo

[a)]Electronic mail: edoxasta@utk.edu

I. INTRODUCTION

Molecular modeling of polymers relies on efficient lattice or continuum representations to capture material properties over the large length-scale and long time-scales of interest.[1] Methods to build coarse-grained (CG) models representative of chemically-specific systems (including polymer melts) have improved over the last two decades.[2–8] Bottom-up CG polymer models are often assembled by utilizing detailed atomistic trajectories of oligomer chains that exhibit short relaxation times in contrast to commercial grade macromolecules. In structure based coarse-graining methods, such as iterative Boltzmann inversion (IBI) and Inverse Monte Carlo (IMC), construction of the pair-wise model aims for matching the structure proposed by a higher resolution model (often atomistic) by tuning the interaction potential.[9–11] Forcefield development proceeds by an iterative refinement of the CG interaction terms guided by deviations from the target pair distribution function, sampled bond lengths, bending and dihedral angles. Due to coupling of separate potential terms, several optimization cycles are commonly performed in sequence[11] with hope that a pair-wise tabulated CG forcefield that lacks cross terms, e.g. bond-angle, will adequately describe chemically-specific polymeric systems. Polyisoprene (PI) is just one of many examples where such CG models have been successfully employed to study the melt or hybrid systems with solid surfaces.[3,11–15] Note that coarse-grain modeling that focuses on capturing structural details of the atomistic systems can serve as an intermediate step towards creation of well-equilibrated initial structures of macromolecular systems by re-introducing atomistic details. Studying local dynamics or structure in atomistic detail is limited by the long chain relaxation times of polymers and back-mapping procedures can address these challenges by transitioning between models at different resolutions.[15–18]

Apart from structure, it is highly desirable that the CG model captures thermodynamic properties, enabling studies under constant pressure. Deviations in the pressure of the CG models developed with IBI are often addressed by the addition of a linear attractive tail to the pair non-bonded interactions.[2,11] Pressure matching via this process requires readjustment of the remaining potential terms because of their coupled contribution in matching the desired structure as proposed by the atomistic model.[3,11–15,19] Thermodynamic consistency is not a challenge exclusive to IBI;[8,20,21] IMC or force-matching procedures often require the addition of constraints to match the thermodynamic state of the input model.[22–24]

Matching volume (V) fluctuations as expressed by the isothermal compressibility $\kappa_T = (\langle V^2\rangle - \langle V\rangle^2)/(k_B T\langle V\rangle)$ is even more demanding. This has been shown in several studies, i.e. on water[19], polybutadiene[25] and polystyrene,[16] where a choice between matching either pressure or compressibility is presented.[19] It is important to emphasize that thermodynamic consistency improves with higher resolution CG models[26,27] where fewer degrees of freedom are eliminated. This was recently shown for PI with a 2-bead per monomer CG model[28] that is closer to the soft degrees of freedom represented by dihedral angles along the PI chain (3 per monomer).[29,30] However, such models offer moderate computational benefit relative to united atom models (5 interaction sites per monomer) and require the development of more complex forcefields given the increased number of particle types (and consequently distinct interaction terms) in the chosen CG representation.

Beyond modifying pair-potentials or increasing resolution, a different proposition is to supplement the forcefield with a volume-dependent term determined by the pressure profile of the mapped atomistic trajectory.[26] The posterior addition of extra potential terms to a CG forcefield requires further refinement due to the expected coupling between each term.[6,31,32] A different approach based on the integral equation coarse-graining (IECG) method, treats the density as an intrinsic property and active variable of the CG model.[8] In IECG the effective CG interactions are developed analytically using the Ornstein-Zernike equation of the liquid state theory.

A distinct path that motivated this study concerns the development of local density potentials to supplement pair interactions with terms that are multi-body by design. While the application of multi-body potentials has been common practice for many decades (for example in simulations with the embedded atom method[33,34] or in simulations using dissipative particle dynamics[35,36]), the introduction of such interactions has found renewed interest in terms of constructing bottom-up chemically-specific CG models.[37–47] In the field of molecular liquids, multi-body potentials can serve to improve predictions on structure and thermodynamics. CG models are often described by a single bead and the potentials include a term that depends on the local density. Techniques developed for optimization of solely pair potentials can be extended with the incorporation of multi-body terms to improve target properties. In one such recent study, Moore *et al.* optimized multi-body interactions by matching the pressure in simulations of hexahydro-1,3,5-trinitro-s-triazine with pair interactions developed by the force-matching method.[42] In another study, Rosenberg used the

relative-entropy optimization supplemented with local density potentials;[47] in this scheme the pair and density potentials where simultaneously optimized by calculating the relative entropy at each iteration using short trial simulations in the NVT ensemble. Results where compared to CG models with solely pair interactions developed with the Inverse Monte Carlo method. In the field of polymers, Sanyal and Shell used the relative entropy framework to parametrize local density potentials for a bead-spring model in an implicit solvent.[41] The optimization performed in the NVT ensemble targeted the development of the non-bonded interactions while the bond and angle potentials where retained from the explicit water model. A different study by Agrawal *et al.* introduced a density-dependent potential term using the embedded-atom method to model semi-crystalline polyethylene.[48] CG bond and bending angle potentials where developed with the IBI method while the parameters of the embedding energy function were calibrated using least squares minimization to match the macroscopic pressure of the system at different state points.

Finally, we should also briefly mention that many-body CG potentials have also been developed rigorously based on cluster expansion techniques.[49] It was found that the cluster expansion formalism can provide accurate effective pair and three-body CG potentials, mainly at high temperatures and low densities. Such a methodology is computationally efficient since the hierarchy of the effective non-bonded CG potentials terms is developed by the conditional sampling of a few CG particles, and no information from long n-body (bulk) simulations is required. However, for dense liquid systems, as the bulk polymer studied here, this method is not directly applicable since numerical estimation of higher than three-body order terms is required.

In this study we develop a methodology to introduce details of the (many-body) local environment in a bottom-up coarse-graining approach applied to an atomistic model of *cis*-1,4 PI. As described previously, PI has been studied extensively with IBI and a recent report highlighted the challenges in capturing structure, density and compressibility.[28] We implement the Inverse Monte Carlo method, which to the best of our knowledge has not been applied previously on polyisoprene melts and we supplement the forcefields with a local density term. In our study, all potential terms are developed simultaneously while targeting specific order parameters and their covariance is accounted for in the optimization process. Section II presents the methodology together with a systematic evaluation of the first "solvation" shell and the validity of an isotropic approximation of the arrangement

of nearest neighbors. Given the limited applications of IMC in polymer systems, we also provide a short description of the method in Section II C. Section III presents our findings on the optimization process with three different forcefields (one solely with pair interactions and two that include local density terms). We examine the ability of different approaches to capture the local and global structure and evaluate the performance of the CG forcefields with regards to the dynamics of PI oligomers. We finish with a brief discussion of our findings and opportunities that our study reveals.

II. METHODS

A. Atomistic simulations

We performed atomistic simulations with a united atom (UA) model for *cis*-1,4 PI that captures the polymer properties quantitatively as shown in past studies.[28,30,50–53] Molecular dynamics simulations were executed for a system of 100 chains with N=30 repeat units (MW=2043 gr/mol) using the GROMACS 2018.3 software.[54] We employed a leap-frog algorithm with a 2 fs time-step and a cutoff of 1.2 nm for van der Waals interactions. The model does not include any partial charges so electrostatic interactions were not considered. Simulations were conducted in the NPT ensemble with the Parrinello-Rahman pressure coupling[55,56] and Nosé-Hoover thermostat[57,58] with a 0.5 and 2 fs time-step respectively. Properties were calculated over 100 ns, 250 ns and 400 ns at 413 K, 353 K and 298 K and 1 bar following an equilibration period of 120 ns, 200 ns and 1.3 μs respectively. While the average chain relaxation time (slowest mode) at 298 K is estimated as 5.6 μs, the calculated properties show small fluctuations after the equilibration period of 1.3 μs. The model predicts a density of $\rho \approx 0.894$ g/m^3 and an isothermal compressibility of $\kappa_T \approx 8.7\times10^{-5}$ bar^{-1} at 298 K. Experimental data values for high molecular weight PI at room temperature are 0.9 kg/m^3[59] and $\kappa_T \approx$5-5.3$\times10^{-5}$ bar^{-1}[60,61], which are in very good agreement with our data given the lower molecular weight of our model[51] and the presence of isomers in experimental samples.

B. Coarse-grain models

A critical parameter in CG modeling is the choice of CG mapping. Different schemes have been employed for CG models of PI with the most common placing a single bead per repeat unit at the center of the bond between two consecutive monomers.[11,14,15] We found that this mapping leaves chain ends with a very anisotropic environment and introduces additional challenges to the coarse-graining procedure adopted in this work. To simplify the application of the multi-body terms that depend on local densities, we represent each repeat unit with a single CG bead centered on the monomer center of mass.[62]

There are three CG models developed in this study, CG-pair, CG-CN and CG-NN. A summary of these models together with a brief description of their differences is provided in Table I.

Model	Nearest-Neighbor potential	Ensemble
CG-pair	None	NVT
CG-CN	Attractive	NPT
CG-NN	Full	NPT

TABLE I. Nomenclature for the developed CG models.

The CG-pair model contains bonded interactions and only pair-wise non-bonded interactions. The effective interactions are optimized to match specific distributions. We calculated the distributions of bond lengths ($\mathbf{S}_\mathrm{b}$), bending angles ($\mathbf{S}_\mathrm{a}$), dihedral angles ($\mathbf{S}_\mathrm{d}$) and distances between non-bonded pairs ($\mathbf{S}_\mathrm{nb}$, calculated for intra- and inter-molecular neighbors that are separated by more than three CG bonds) as described further in Section II C. At each temperature, the discretized distributions were extracted from the mapped trajectory and concatenated to the vector $\mathbf{S}_\mathrm{target}$ used as the target for coarse-graining. Furthermore, to account for the many-body interactions beyond the pair-wise assumption, we derived two forcefields (CG-NN and CG-CN) by employing an additional multi-body potential ($U_\mathrm{NN}(n)$), where n is the number of nearest-neighbors (NN). The optimization of these forcefields with IMC required the introduction of the distribution of n ($\mathbf{S}_\mathrm{NN}$) to the set of target distributions. It has been reported that pair potentials often fail to quantitatively capture the short-range effective interactions from neighbors in the first coordination shell (first minimum of radial distribution function $g(r)$) and corrections via multi-body contributions are

necessary to capture the local density as well as density fluctuations.[24,39] A computationally inexpensive way to calculate the number of nearest-neighbors (n) of a bead on-the-fly is through a continuous isotropic indicator function ($\Theta(r)$) that reduces to zero at a fixed cut-off. A common choice for $\Theta(r)$ is the Lucy function.[42,63–65] Sanyal et al.[41,46,47] optimized the cut-off radius as a system parameter using the relative entropy coarse-graining framework. Faure derived the local density using Voronoi tessellation.[40,66–68] Here, we explore the use of the solid-angle nearest-neighbor (SANN) algorithm developed by Meel et al.[69] to extract n. The SANN algorithm defines n neighbors on a particle i based on the separation between i and all remaining particles as well as their individual contributions towards a total solid angle 4π which leads to the following condition for the maximum neighbor radius R_i^n

$$R_i^n = \frac{\sum_{j=1}^{n} r_{i,j}}{n-2} < r_{i,n+1}{}^{69} \tag{1}$$

The resulting distribution of n is shown in Fig. 1a at three different temperatures. Note that the distribution of nearest-neighbors defined by SANN presents minimal changes over a broad range of temperatures. The increase of the density at lower temperatures translates to a lower volume occupied by the first neighbors (defined by R_i^n). An illustration of nearest neighbors is shown in Figure 1b. We observed that chain ends exhibit a slightly increased n (positioned at longer distances) due to their increased free volume.[51] While in principle this could translate to a need for a separate density potential this effect is minimized when introducing a uniform indicator function (described below) since for a fixed range similar distributions were observed irrespective of the position of the CG bead along the chain.

The SANN analysis provides a systematic method to define the indicator function from the contribution of nearest-neighbors to the local density surrounding the CG bead and allowed us to quantitatively assess the accuracy of the isotropic approximation of local density. The indicator function Θ is defined as the fraction of the total pair distribution function from the nearest-neighbors $\Theta(r) = g_{\mathrm{NN}}(r)/g_{\mathrm{tot}}(r)$ and extrapolated to 1 for short separations before the first peak (Fig. 1c, ≈ 0.27 nm). The resulting indicator function calculated at three temperatures are depicted in Fig. 1d. It is evident that the aforementioned definition of a "local volume" decreases with temperature as the size of the first coordination shell becomes progressively smaller. The number of nearest-neighbors (n) around a particle i can be calculated using $\Theta(r)$ as

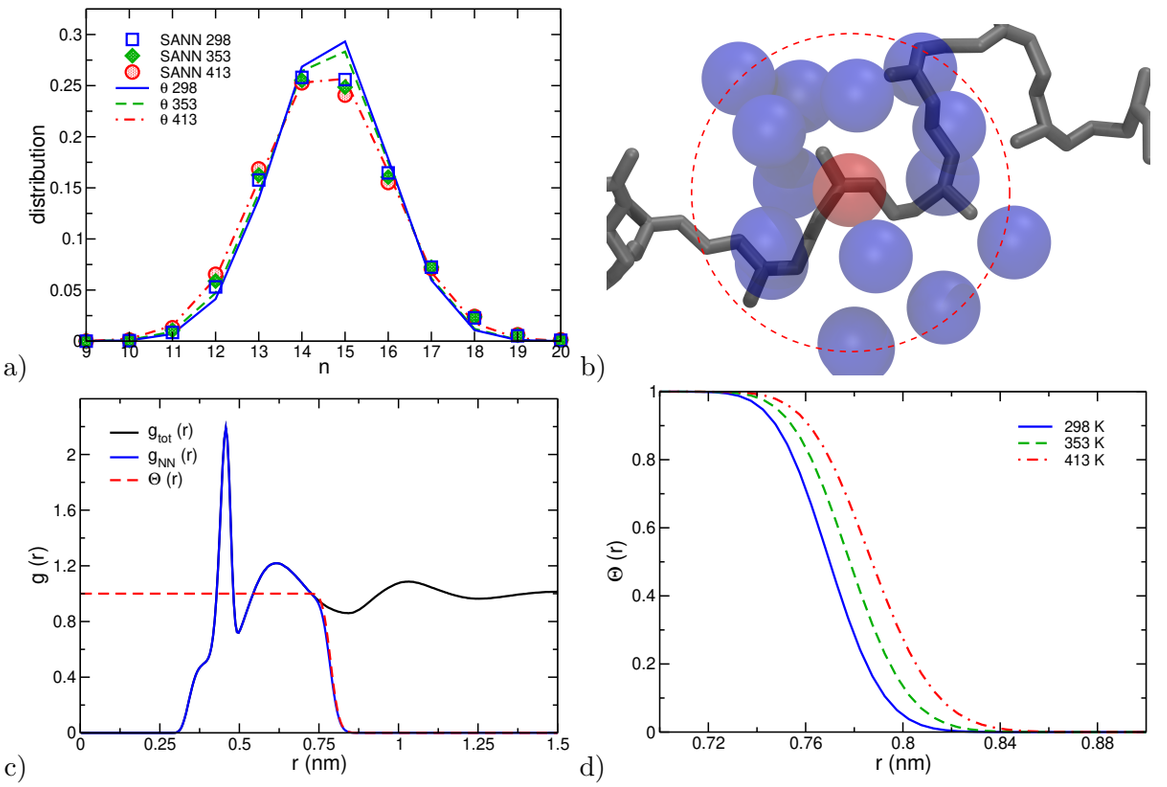

FIG. 1. a) Distribution of number of nearest-neighbors (n) extracted with SANN[69] (symbols) and by the use of indicator functions (lines) at 298 K (blue), 353 K (green) and 413 K (red). b) Nearest-neighbors (blue) of a CG bead (red) as extracted from SANN analysis on a mapped atomistic configuration. The dashed red lines represent a circle centered on the examined bead. c) Total pair-distribution function $g(r)$ and contribution from nearest-neighbors ($g_{\mathrm{NN}}(r)$) as determined by SANN at 413 K. The dashed line depicts the extracted indicator function $\Theta(r)$. d) Indicator functions extracted at different temperatures.

$$n_i = \sum_{j \neq i}^{N} \Theta(r_{ij}) \tag{2}$$

Note that for consistency with pair-distribution functions in defining local density, the center atom does not contribute to the calculation of n.[70] Intra-molecular neighbors that participate in bond stretching, angle bending and torsional potentials were included in the list of potential neighbors which underlines a complexity present when working with polymers, i.e. the unavoidable direct or indirect correlation between the calculated distributions of the corresponding order parameters. Comparing the direct results of SANN and the

analysis performed using an isotropic $\Theta(r)$, we observe that deviations between the two calculations increase as temperature decreases. The isotropic assumption depends on the mapping scheme employed as well as the specific system and conditions studied. SANN was introduced originally as a rapid algorithm to perform NN calculations on-the-fly as an alternative to simulations with Voronoi tessellation of space which are computationally demanding. Although we opted for the simpler $\Theta(r)$ route for calculating n in this study, we note that SANN could be explored for on-the-fly calculations in Monte Carlo simulations, particularly at lower temperatures where the application of the spherical approximation presents significant deviations. Using n calculated from Eq. 2, forcefields for the CG model were assembled in tabulated forms and the potential of the CG-CN and CG-NN models assumed the general form of

$$U_{\text{total}} = U_{\text{bond}}(l) + U_{\text{angle}}(\theta) + U_{\text{dihedral}}(\phi) + U_{\text{nonbond}}(r) + U_{\text{NN}}(n) \tag{3}$$

where U_{bond}, U_{angle} and U_{dihedral} are the bonded potentials that depend on bond length (l), angle (θ) and dihedral (ϕ) of consecutive beads along the chain. U_{nonbond} is the pair interaction between non-bonded beads and is a function of their separation (r). U_{NN} depends on the number of nearest-neighbors surrounding the bead (n) which includes both bonded and non-bonded neighbors. The force on a particle i from a surrounding neighbor j can be calculated as[42,71,72]

$$f_{i,j} = -\left[\frac{dU_{NN}}{dn}(n_i) + \frac{dU_{NN}}{dn}(n_j)\right]\frac{d\Theta(r_{ij})}{dr}\frac{r_i - r_j}{r_{ij}} \tag{4}$$

The above equation underlines the advantage of choosing an indicator function that varies continuously with r. As described in the next section, we derived three different forcefields in this study, one with only pair-wise interactions and two with local density potentials .

C. Inverse Monte Carlo

We constructed the CG force fields using the inverse Monte Carlo (IMC) method (also known as the inverse Newton method).[73,74] IMC provides a numerical framework to derive a set of potentials that reproduce the target distribution vector ($\mathbf{S}_{\text{target}}$) derived from the atomistic model. The tabulated potentials (as constant grids) are updated at each iteration

of the IMC procedure. The potential at each bin, E_α, contributes to the bonded, non-bonded or NN energy with the corresponding occurrence S_α. The total potential energy of the system is calculated by

$$U_{\text{total}} = \sum_\alpha S_\alpha E_\alpha \tag{5}$$

The initial estimate for potentials at 413 K was derived from the Boltzmann inversion of target distributions,[11] e.g, $E^0_{\text{nb}}(r) = -\text{k}_\text{B}T \ln g_{\text{nb}}(r)$ where $g_{\text{nb}}(r) = S_{\text{nb}}(r)/(4\pi r^2 N)$. Likewise, we estimated the nearest-neighbor potential from the distribution $S_{\text{NN}}(n)$ as $E^0_{\text{NN}}(n) = -\text{k}_\text{B}T \ln S_{\text{NN}}(n)$ (see Figure 1a). We initiated the optimization at lower temperatures using potentials from the immediate higher temperature as the initial guess. The deviation of the distribution vector from the target $\Delta\langle\mathbf{S}\rangle = \langle\mathbf{S}_{\text{obs}}\rangle - \langle\mathbf{S}_{\text{target}}\rangle$ was sampled at each iteration and the potential was updated by $\Delta\mathbf{E}$ after solving the inverse problem:

$$\Delta\langle S_\alpha\rangle = \sum_\beta \frac{\partial\langle S_\alpha\rangle}{\partial E_\beta}\Delta E_\beta + \mathcal{O}(\Delta\mathbf{E}^2) \tag{6}$$

Note again that the vector $\mathbf{S}$ is a concatenated vector with populations sampled for each bin of the tabular potentials (including bonded, angular, dihedral, pair and NN terms) and the vector $\Delta\mathbf{E}$ consists of an update to all tabulated terms. The Jacobian matrix ($\mathbf{A}$) employed in the equation above is calculated using the relationship[73,74]

$$A_{\alpha\beta} = \frac{\partial\langle S_\alpha\rangle}{\partial E_\beta} = -\frac{\langle S_\alpha S_\beta\rangle - \langle S_\alpha\rangle\langle S_\beta\rangle}{k_B T} \tag{7}$$

The linear problem in Eq. 6 is ill-defined as can be further shown by singular value decomposition[23,75], e.g., the distribution of sampled bond lengths depend on differences in the bonded potential and not the value at the minimum. Therefore we minimized the following objective function instead[23]

$$\text{objective function} = ||\mathbf{A}\Delta\mathbf{E} - \Delta\mathbf{S}||_2 + \gamma||\Delta\mathbf{E}||_2 \tag{8}$$

that includes a regularization term to penalize large changes in potentials (expressed as the Euclidean norm of the vector $\Delta\mathbf{E}$) and reduce the impact of statistical noise. We employed the Sequential Quadratic Programming algorithm as implemented in standard Python distributions. The regularization parameter γ controls the stability of the scheme at the expense of sensitivity; we decreased γ as $\mathbf{S}$ approaches the target up to the point where

the method becomes unstable given the presence of statistical noise in our data; for our particular system and target vector we employed λ values in the range of 20 to 0.3. At each iteration the minimum of the bonded potentials was shifted to zero while for the non-bond potentials the value at $R_{\text{cut}} = 1.5$ nm was constrained to zero. The bonded and non-bond tables consisted of $100 - 150$ bins each and the NN potential had about 15; we found that the choice of grid affects the sensitivity of the optimization scheme as well as the strength of the relative contribution of each distribution to $\Delta\mathbf{S}$.

Several past studies restricted multibody non-bonded potentials to solely providing cohesive energy to the system.[41,76] Within the scheme proposed herein, we introduced an additional constraint in the minimization that $\Delta\mathbf{E}$ should result to a monotonic NN potential decreasing as n increases. Effectively all repulsion in this new set of interactions (termed CG-CN model) will be provided by pair interactions testing directly whether such a simpler model provides an equally satisfactory description of the system under study.

The IMC procedure was performed employing our in-house Monte Carlo algorithms[15,77] and included the execution of single particle displacements (89.95%) and reptate moves performed in a configurational-bias scheme.[78] Simulations for the model without NN interactions (CG-pair) were performed in the NVT ensemble to match the density of the UA model. The local density models were simulated in the NPT ensemble at 1 bar by including 0.05% attempts for volume changes. As documented in the literature, IMC requires exhaustive sampling of both distributions and correlations for the scheme to converge to a satisfactory solution.[23,79,80] The distributions herein were calculated over at least 400 million MC iterations (after an equilibration period of 20,000 iterations) to reduce the statistical noise with more iterations performed at lower values of λ. A technical but important detail of the scheme is that the sampling for the Jacobian matrix (Eq. 7) was performed on-the-fly every 2,000 steps.

D. CG Molecular dynamics

Molecular dynamics simulations of CG models were performed by a custom in-house parallel MD code, using the CG potentials derived from the IMC procedure to study dynamics. The (additional) computational cost of including many-body terms, based on the distribution of nearest-neighbors, is about 2-3 times that of a typical simulation with a pair potential.

This is mainly due to the extra neighbor lists which must be separately constructed for the calculation of local density for each atom; the latter has a computational complexity similar to the calculation of a pair potential. The computation of the actual force requires nearly as much time as the forces of a pair potential. Finally, some extra communications are needed for the exchange of the local densities across MPI ranks.

Simulations of the CG-NN and CG-CN models were performed in the NPT ensemble at 1 bar and different temperatures. The Parrinello-Rahman barostat[55,56] and the Nosé-Hoover thermostat[57,58] were used to maintain constant pressure and temperature. Simulations of the CG-pair model were performed in the NVT ensemble to maintain the density obtained from the corresponding UA simulations. The time step was 1 fs. The NVT runs where 10 ns long and the NPT runs 4 ns.

III. RESULTS

In the following sections we present our findings on the structure, conformational properties, thermodynamic properties and dynamics of the CG models compared to the underlying UA model as well as the performance of the method employed. We derived three models: (a) The first, "CG-pair", assumes solely pair-wise interactions between particles that do not interact via bond, angle or dihedral potentials. The other two CG potentials include the multi-body potential acting on particles due to the number of NN in their local environment; (b) the "CG-NN" includes both attractive and repulsive multi-body terms, whereas (c) the "CG-CN" includes only attractive terms due to the constraint implemented during the forcefield development process.

A. Optimization

A clear advantage of the IMC procedure is the simultaneous refinement of separate interaction potentials based on distributions accounting to an extent for their coupled effect on the sampled distributions. It is therefore instructive to consider the performance of our scheme. Fig. 2a presents the norm of the deviation vector from the target distribution for the CG-CN forcefield at 413 K as a function of IMC iterations. Within 5 iterations, the initial $||\Delta\mathbf{S}||_2$ decreases by an order of magnitude while subsequent optimization is slower

as the distribution vector approaches the target. The convergence route of individual terms depends on the choice of distribution grid (which determines the weight in $\mathbf{S}$) and the choice of the initial estimate. The initial potential results in high deviations in the non-bonded pair distribution which is the main contributor to the objective function. The inversion of the pair distribution function results in a repulsive potential (Fig. 2b, iteration 0) with little cohesion. Such an estimate could jeopardize the stability of CG NPT simulations during forcefield development. In our scheme, the inverted nearest-neighbor potential (Fig. 2c for the CG-CN model) provides the necessary attraction preventing the sampling of configurations with reduced local density. As the optimization progresses, the pair potential develops a first minimum to match the distribution of pairs (Iteration 10). The density of the CG-CN model is rapidly converging to the targeted value at a faster pace than the matching of intramolecular distributions.

The initial intra-molecular potentials provide a reasonable estimate with small deviations. Therefore, the bonded terms undergo significant improvement only at later stages of the algorithm when sensitivity is increased by lowering the γ parameter. We reduced γ every 5 iterations to increase the allowed size of the update $\Delta\mathbf{E}$ in the forcefield at each step. After 10 iterations the non-bond distribution resembles the target and the algorithm switches focus towards capturing intramolecular distributions; remarkably this coincides with the algorithm attempting to match the tails of the nearest-neighbor distribution (see Iteration 10 and Iteration 22 in Fig. 2c) with a parallel rendering of the pair-potential at short distances positive and a second clear minima over the second "solvation shell". Such observations highlight that if the fidelity in matching distributions is overlooked very different forcefields could be derived. We also emphasize (and will demonstrate further) that intramolecular neighbors and their placement is coupled to capturing the local density accurately.

The iterative minimization is limited by a system-dependent minimum value of γ as further reduction results in instability. We found that this limit is coupled to the quality of distributions and cross-correlation terms; noise in the Jacobian is detrimental to the method therefore we opted to sample the distributions on-the-fly circumventing challenges limitations imposed by storage of large trajectories. All three approaches (CG-pair in NVT, CG-NN and CG-CN) converged to forcefields that were capable to capture the target distributions satisfactory with small deviations from the target distributions. The potentials at 413 K provided a good initial estimate for rapid convergence at 353 K however at 298 K the

slow structural evolution made the optimization process challenging even while the needed changes from the potentials refined at 353 K were minimal.

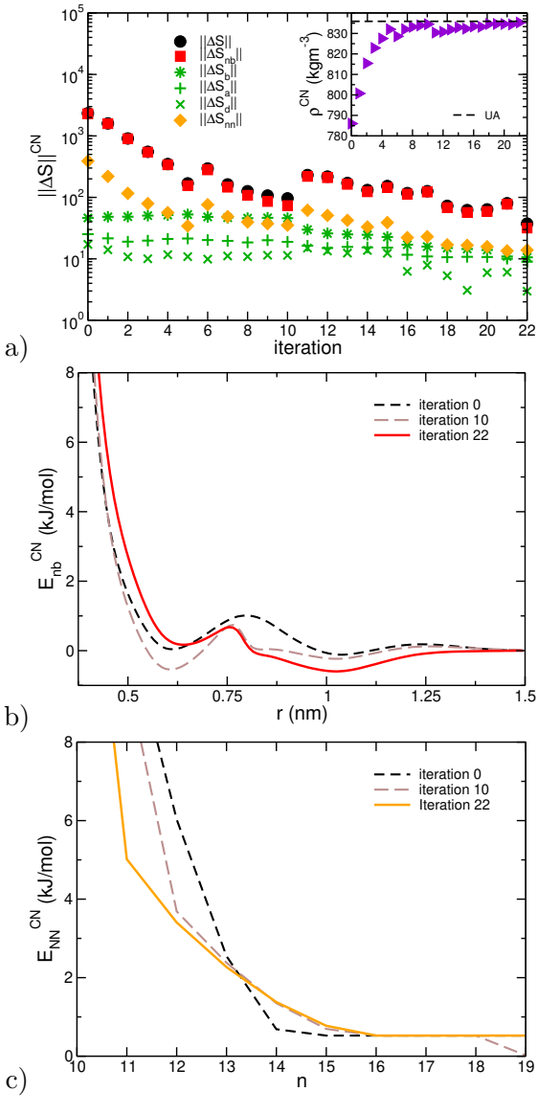

FIG. 2. a) Convergence of distributions to target values during the IMC process (inset shows the convergence of density to the UA value at 1 bar) and optimization of the b) pair non-bonded potentials and c) nearest-neighbor potentials at 413 K for the CG-CN in the NPT ensemble.

B. Forcefields

The forcefields of the three CG models optimized at 353 K are presented in Fig. 3 to demonstrate their significant differences. The CG-pair model is able to capture the target

distributions quantitatively. However it lacks necessary cohesion and is incapable of capturing the correct pressure in the NVT ensemble (and accordingly the correct density of the melt in an NPT simulation). In contrast both CG-NN and CG-CN PI models include attractive non-bonded interactions with the latter limited to solely monotonic local density potentials as often considered in coarse-grain models with multi-body terms.[31,41,46,81] Note that constraining the local density potentials to cohesive forces results in a increased pair-potential within the range of the chosen indicator function (Fig. 3e,1d). Similarly, the repulsion provided by the local density potential in CG-NN at higher values of n (Fig. 3e) is counteracted by a more attractive pair potential in the range of the chosen indicator function (Figs 3e,1d) to match $\mathbf{S}_{\mathrm{nb}}$ in the first coordination shell. The above findings further underline: a) the strong coupling between pair non-bonded potential and local density potentials and b) the significance of the choice of "local volume" (or range of the multi-body term) in such forcefields. Clearly the shape of the indicator function holds also a prominent role in guiding the potentials extracted; herein we followed a systematic approach based on the contribution of the first-neighbor shell to the total density profile. It is important to emphasize the unique challenges present in bottom-up coarse-graining of polymers by direct contributions from proximal intramolecular neighbors along the chain to the local density profile of a segment. This is evident by differences in the bond, angle and dihedral potentials extracted as shown in Fig. 3 at 353 K with similar results at all temperatures studied. While all three models are capable of capturing quantitatively the corresponding distributions significant variations can be observed by examining individual terms between the forcefields.

All CG forcefields (CG pair, CG-NN and CG-CN) were able to reproduce the respective target distributions at different temperatures with minimal deviations; however none was found to be temperature transferable. A refinement was introduced as described previously with additional optimizations performed to quantitatively capture the distributions at a different state point. Several strategies could be explored to optimize the potentials on multiple state-points and improve transferability.[43,82,83] However in our method two additional features need to be considered: a) the definition of "local volume" changes across temperatures and correspondingly the employed indicator function and b) the assumption of a spherical local volume at low temperatures needs to be examined. Fig. 1c presents a systematic approach to extract the temperature dependence of the local volume; using

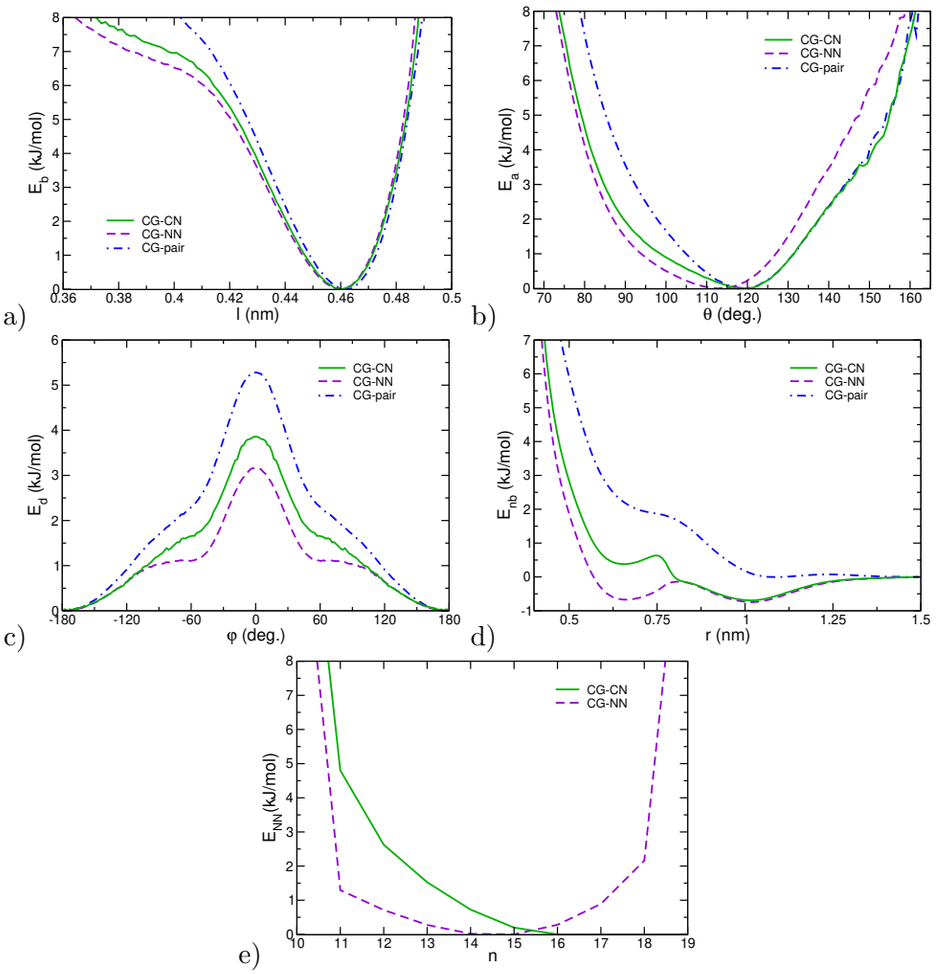

FIG. 3. a) Bond-stretching, b) Angle-bending, c) Dihedral, d) Pair non-bonded and e) Nearest-neighbor potentials for CG-CN (−), CG-NN (- -) and CG-pair (· - ·) at 353 K.

these data the extracted non-bonded potentials after refinement are presented in Fig. 4. It is evident that after accounting for the decrease of the local volume (together with remarkably similar distributions of first neighbors, Fig. 1a) the non-bonded potentials derived independently at different temperatures are similar but not identical as expected. While a temperature-transferable CG model would be highly desirable we note that the reduced resolution offered by the model (together with the lost entropy of the eliminated degrees of freedom) does come at a cost; as shown in Fig. 1a, the spherical approximation of the local volume becomes progressively less accurate therefore additional constraints to capture multiple-state points with a single forcefield could result to reduced quality in matching the target distributions at individual temperatures.

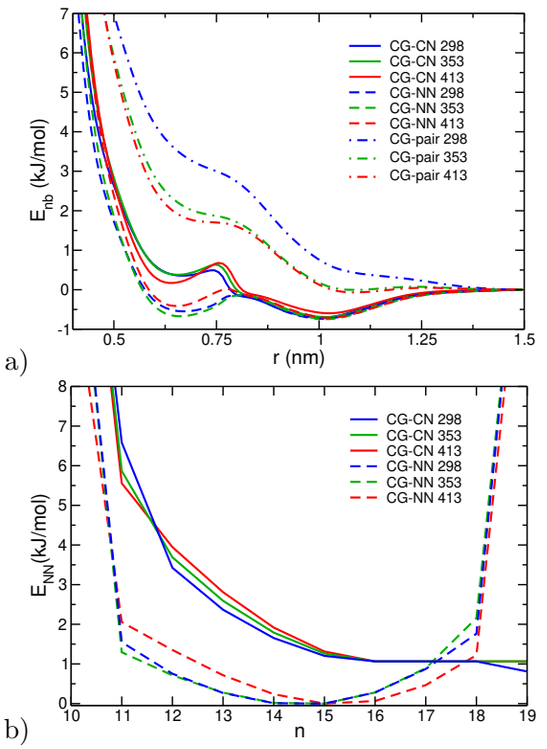

FIG. 4. a) Pair non-bonded and b) nearest-neighbor potentials for CG-NN (−) and CG-CN (- -) at 298 K (blue), 353 K (green) and 413 K (red).

C. Conformational and thermodynamic properties

We first evaluated the ability of the forcefields to drive CG simulations that exhibit distributions matching those calculated with the atomistic model. In all cases the method converged to forcefields that were able to reproduce the target distributions with minimal deviations. To demonstrate the successful implementation we present in Fig. 5 the distributions using all three models at 298 K compared to the UA target. Note that this is the state point where the largest deviations persisted; in all other cases the CG model provided an even finer description of target distributions. As it is evident, the CG forcefields are able to capture all essential features recorded with the atomistic model based on the mapping performed. Note however that the shown distribution of nearest neighbors is an approximation based on spherical sampling as discussed previously (and not the one from SANN analysis).

Coarse-graining of polymers aims to facilitate the modeling of long macromolecules. To this extend, it is informative to examine how successive local degrees of freedom vary using common measures as the distribution of internal separations as well as the chain end-to-end

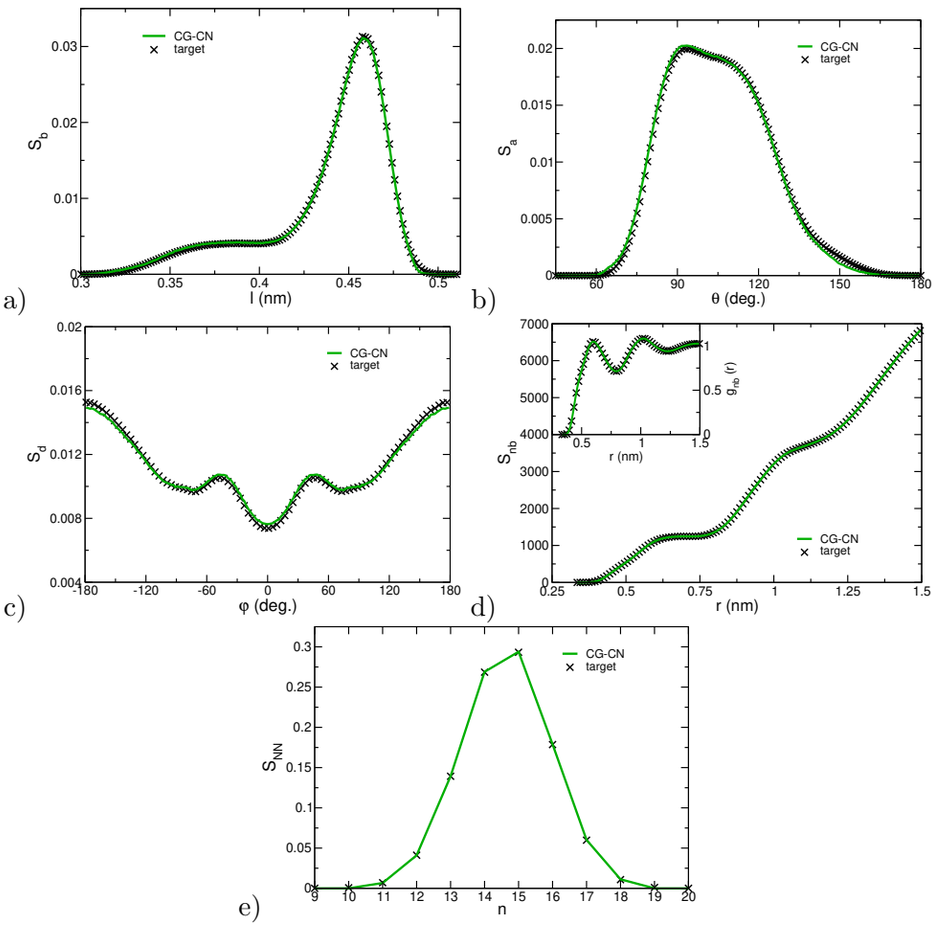

FIG. 5. Distribution of a) bond lengths, b) bending angles, c) dihedral angles, d) distance between nonbonded pairs and c) number of nearest-neighbors as produced by the CG-CN forcefield (lines) and mapped UA trajectory (target, symbols) at 298 K. At higher temperatures the agreement is more favorable.

distance. The average separation of internal segments casted in the form of characteristic ratio is presented in Figure 6a. As seen all models capture well the average distances in close agreement to polyisoprene's conformational characteristics.[30] Beyond a small number of monomers the chains resemble gaussian coils; the whole distribution of chain end-to-end distances is presented in Figure 6b where again a quantitative match between all models is observed. However, if we consider the separation between a small number of segments (1-4, interacting via dihedral and density potential terms) we find that while the coarse-grain models capture the average separation (see Figure 6a) they fail to quantitatively describe the distribution of distances. Obviously capturing the dihedral distribution (internal coordi-

nates) does not guarantee a similar result in Cartesian coordinates. Atomistic models retain excluded volume between repeat units separated by three monomers while we opted to exclude such interactions between corresponding coarse-grain beads. Another option would be to include direct non-bonded interactions however we found that the general non-bonded potential is not necessarily applicable between such pairs. In fact, even in atomistic forcefields the inclusion of a portion of non-bonded interactions between atoms interacting via the dihedral potential is a design variable. Given the satisfactory agreement for average separations as well as their distributions when examining segments separated by more CG bonds we chose the scheme with no 1-4 non-bonded interactions. It is also important to emphasize another feature present in our study with regards to 1-4 pairs: based on the data from mapped atomistic trajectory such segments switch between the first and second solvation shell of each other (symbols in Fig. 6c). This further cautions that development of the CG forcefield is subject to coupling between individual terms and supports the selection of an indicator function that clearly distinguishes between the two populations.

T (K)	413			353			298		
Model	UA	CG-CN	CG-NN	UA	CG-CN	CG-NN	UA	CG-CN	CG-NN
ρ (kg/m^3)	836	835	836	866	866	866	894	894	894
$\kappa_T \times 10^5$ (bar^{-1})	10.0	9.9	9.6	9.3	8.2	10.0	8.7	11.7	10.3

TABLE II. Comparison of density (ρ) and isothermal compressiblity (κ_T) of CG models with UA for various temperatures (T) at 1 bar.

Finally, we compared the predictions of all three CG models on PI's thermodynamic properties. The CG-pair model performs satisfactory with regards to structure and conformational properties at the target density, i.e, $R_g \approx 1.35$ at 413 K. However, it overestimates the pressure by three orders of magnitude, $P \approx 5{\times}10^3$ bar. Introducing the nearest-neighbor potentials does not only help to drive the system to the target density (inset in Fig. 2a) under an optimization process within the NPT ensemble but also assists in capturing density fluctuations. The isothermal compressibility was calculated using:[84]

$$\kappa_T = \frac{\langle V^2 \rangle - \langle V \rangle^2}{k_B T \langle V \rangle} \tag{9}$$

The CG-NN and CG-CN models reproduce both the atomistic density and volume fluc-

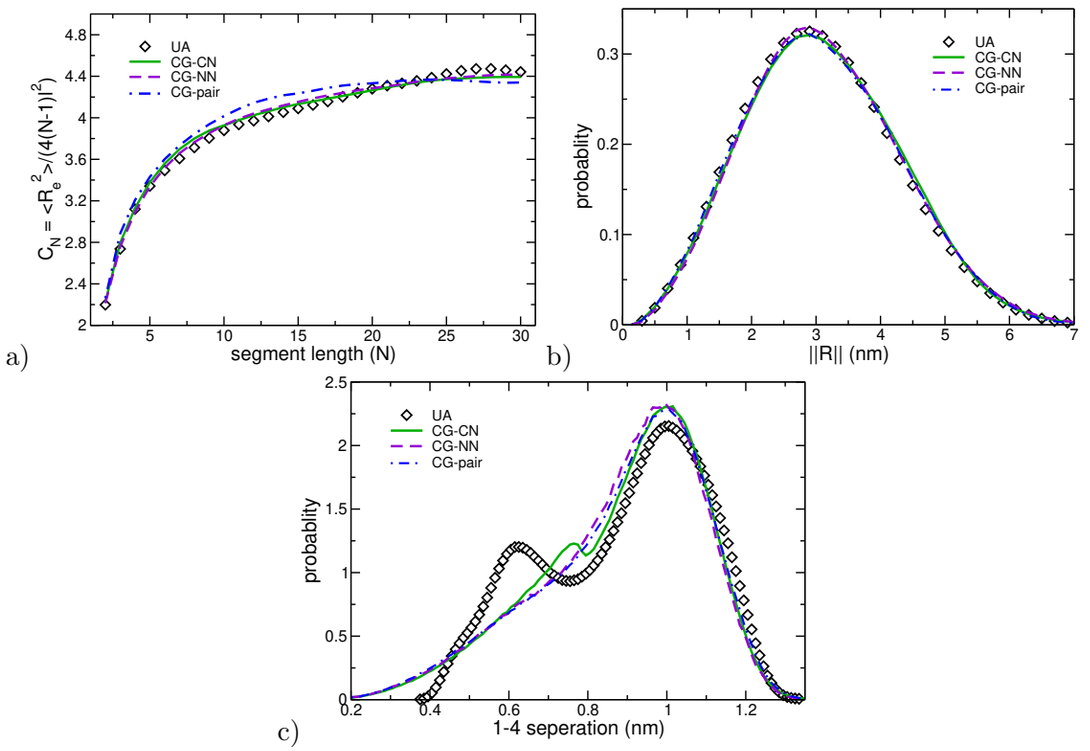

FIG. 6. a) Characteristic ratio (C_N) at as a function of number of monomers (N) in the 30mer chain, b) distribution of chain end-to-end distances and c) distribution of distances for CG beads separated by three bonds for mapped UA and CG models (CG-NN, CG-CN and CG-pair) at 353 K.

tuations with excellent agreement, i.e. $\rho \approx 835$ kg/m^3 and isothermal compressibility $\kappa_T \approx 9.6\times10^{-5}$ bar^{-1} for the CG-NN model in the NPT ensemble at 1 bar and 413 K. Table II presents the thermodynamic and conformations properties predicted by the CG-NN and CG-CN models at three temperatures compared to the UA model. It is important to emphasize that in contrast to techniques that add potential terms to the extensive volume of the system (or density) herein we aimed at capturing the fluctuations at local length-scales; therefore it is not guaranteed that volume fluctuations would be matched. In fact, upon close inspection we find that the atomistic model predicts a decrease of compressibility as we move to lower temperatures; this trend does not appear in the CG model. We attribute this disagreement to the non-spherical packing of neighbors (or the limitation of a spherical description of CG beads) as well as to potential coupling between neighboring solvation shells. At lower temperatures we anticipate that the correlation length increases while the solvation shells over which the many-body term applies become progressively smaller

(Fig. 2d). Despite these small deviations, the agreement between the models is remarkable given challenges present to capture both density and fluctuations with simpler CG models.[28]

D. Dynamics

In this section we examined the ability of the different CG models to qualitatively predict the dynamics of the atomistic PI system. The main advantage of coarse graining in polymers is the increase in length and time scales accessible by simulations. However, as discussed in the Introduction, the direct use of the raw CG data for the quantitative predictions of polymer dynamics is not straightforward due to the use of softer effective CG potentials. Indeed, as a consequence of the softening of the energy landscape at the mesoscopic description, and the fact that we neglect the friction and random forces,[7,17] the time in the dynamic CG simulations does not correspond to the real time of the underlying polymeric system and has to be properly scaled. For systems characterized by a scalar friction, such as the homopolymer melts studied here, the scaling parameter (τ) is expected to be a scalar variable, and corresponds, in the long time regime, to the ratio between the atomistic friction coefficient and the one in the CG description.[85,86]

$$\tau = \frac{\zeta_{AT}}{\zeta_{CG}} \tag{10}$$

Due to complex and strongly fluctuating potential energy surface (PES), it is not possible to derive an analytical prediction of τ. Thus, typically τ is calculated either from experimental data or from detailed atomistic simulations.[85,86] Here, we used data from the UA MD simulations and matched the mean square displacement (MSD) of the CG particles over a considerable time, where both amplitude and shape coincide. Thus, the derived scaling factor determines the real time unit to which the CG time corresponds. In addition, such a methodology provides direct insight into the spatio-temporal scales the particular CG simulation can be used for. The MD simulations for the CG-pair model were performed in the NVT ensemble at the corresponding UA density while all other simulations were performed in the NPT ensemble at 1 bar.

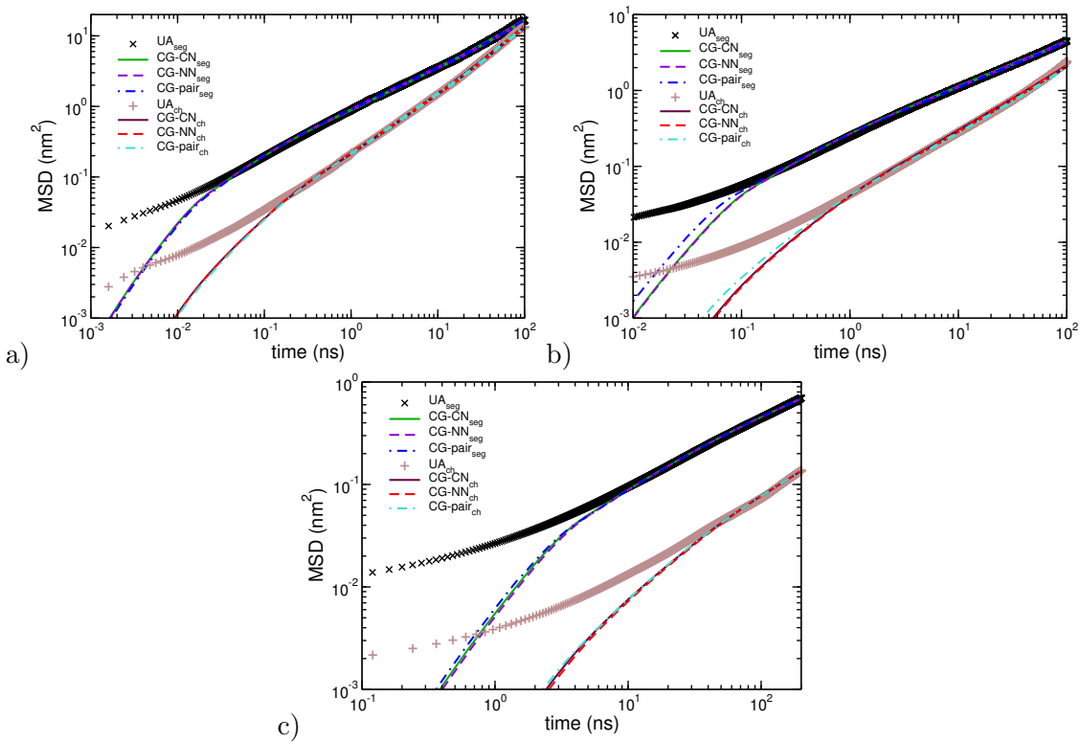

FIG. 7. Mean square displacement (MSD) for segment (monomer) and chain center-of-mass from UA and CG simulations (CG-CN, CG-NN and CG-pair) at different temperatures: a) 413 K, b) 353 K and c) 298 K.

1. Mean Square Displacement

We compared the segmental (monomer) and chain center-of-mass MSD of PI from the UA and the different CG MD simulations at different temperatures in Fig. 7. In all cases the CG data are scaled with a scaling factor, τ, obtained by overlapping the CG and the atomistic curves in the long time (Fickian) regime. Note that although the simulation lengths are shorter than the time scale of the Fickian regime, the scaling factors for the segmental and chain MSD have similar values at the observed time. In Fig. 7a, for the two density dependent CG models, CG-NN and CG-CN, we see that segmental $\tau = 20 \pm 3$; i.e. for this specific system and state point, the CG model exhibits segmental dynamics that is about 20 times faster than the more accurate atomistic one.

It is clear from the data shown in Fig. 7 that atomistic and CG dynamics overlap for length scales around 0.35 nm (MSDs of 0.1 nm^2) for both chain and segmental MSD at all temperatures. This length scale corresponds to the size of the PI monomer (similar to CG

bead in the current 1:1 mapping). This is not surprising if we consider that a CG model is not expected to accurately capture the displacement of the COM of the CG bead at distances shorter than its size therefore it will underestimate such short-length displacement due to the lack of necessary degrees of freedom. The time scale where segmental MSD curves overlap is expected to be of the same order of magnitude as the characteristic time of the CG bead length scale and increases as temperature decreases from around 0.04 ns at 413 K to about 10 ns at 298 K. Such time scale for segmental dynamics is similar to the $\alpha-$relaxation time in PI which is about 0.04 ns for the UA model at 413 K. The chain MSDs overlap at times above 0.1 ns at 413 K. The CG time scaling does not need to be performed in the diffusive regime, since atomistic and CG MSD curves follow each other even for shorter times. This is due to the fact that our CG model is rather close to the underlying chemistry, i.e. a CG bead corresponds to a PI monomer. Such, a behavior has been observed in the past for other polymers as well.[17,85] Interestingly the CG-pair potential gives rather similar dynamics. However note that the latter is performed in the NVT ensemble; since as previously mentioned the NPT simulations with this model do not yield the correct density. Thus, for the above spatio-temporal scales the proposed density dependent CG model describes correctly, after scaling, both the segmental and the chain translation dynamics of the PI melts as predicted by the atomistic simulations.

The scaling factor is the same for chain and segmental MSDs (within the error) but varies greatly with temperature; from around 20 at 413 K, becomes about 100 at 353 K, up to about 4000 at 298 K. Such a strong temperature dependence of the parameter τ has been observed before for other CG polymer models[87] and since the friction coefficient is inversely proportional to diffusivity the decrease of temperature causes a more pronounced deceleration of the Brownian dynamics of atomistic chains as compared to the CG ones. Finally, we should note that as temperature decreases, approaching T_g, it is clear that the atomistic short time (below 1 ps) dynamics become more heterogeneous; these motions clearly cannot be captured by the CG model since they correspond to length scales below the size of the CG bead (PI monomer). Note also that the time mapping is very similar if a dynamical quantity describing the orientational dynamics, such as the chain end-to-end vector auto-correlation function, is used.

Overall, the possibility to describe accurately the Brownian motion of a PI chain at such small length and short time scales is one of the advantages of the present CG model and is

directly related to the chosen mapping (1:1) scheme, that allows the model to be close to the atomistic structure. However, we should note that as temperature decreases, approaching T_g, it is clear that the atomistic local (segmental) dynamics of PI become more heterogeneous for very short length scales; see atomistic MSD data of about 0.1 nm^2 in Fig. 7c. Such dynamical heterogeneities are related to local motions, such as bond vibrations and angle librations, which clearly cannot be captured by the CG model, since they correspond to length scales below the size of the CG bead. This observation supports the importance of the atomistic detail in order to describe local dynamical modes of PI, particularly at low temperatures; note though that the CG models can still be practical by providing initial starting points for backmapping procedures.

2. Dynamic Structure Factor

The previous analysis examined the ability of the CG models considered in this work, to predict the translational dynamics of PI chains (quantified via the MSDs) at the segmental and chain-center-of-mass level, using as a reference the data from the detailed atomistic simulations. Another quantity that typically probes molecular motions at multiple length scales is the coherent dynamic structure factor $S(q,t)$ (measured directly by neutron spin echo, NSE experiments), defined as:

$$S(q,t) = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{\sin(qr_{ij}(t))}{qr_{ij}(t)}/S(q,0) \tag{11}$$

The above equation can be used to extract $S(q,t)$ directly, using the configurations that are stored during the MD simulations, by recording the time correlation function of segments i and j belonging to the same chain.[88–90] The advantage of $S(q,t)$ is that by varying the value of the scattering vector q the dynamics over distinct length scales can be studied and compared between different models.

Fig. 8 shows the normalized single-chain dynamic structure factor $S(q,t)$ for different values of the scattering vector (q = 3, 6, 9, and 12 nm^{-1}) as extracted by the mapped atomistic trajectory and the CG density dependent NPT simulations, at 413 K. All three CG models have very similar dynamics as shown in MSD data, so in the following we report only the data from the CG-NN simulations. Note, that the first peak of the static structure

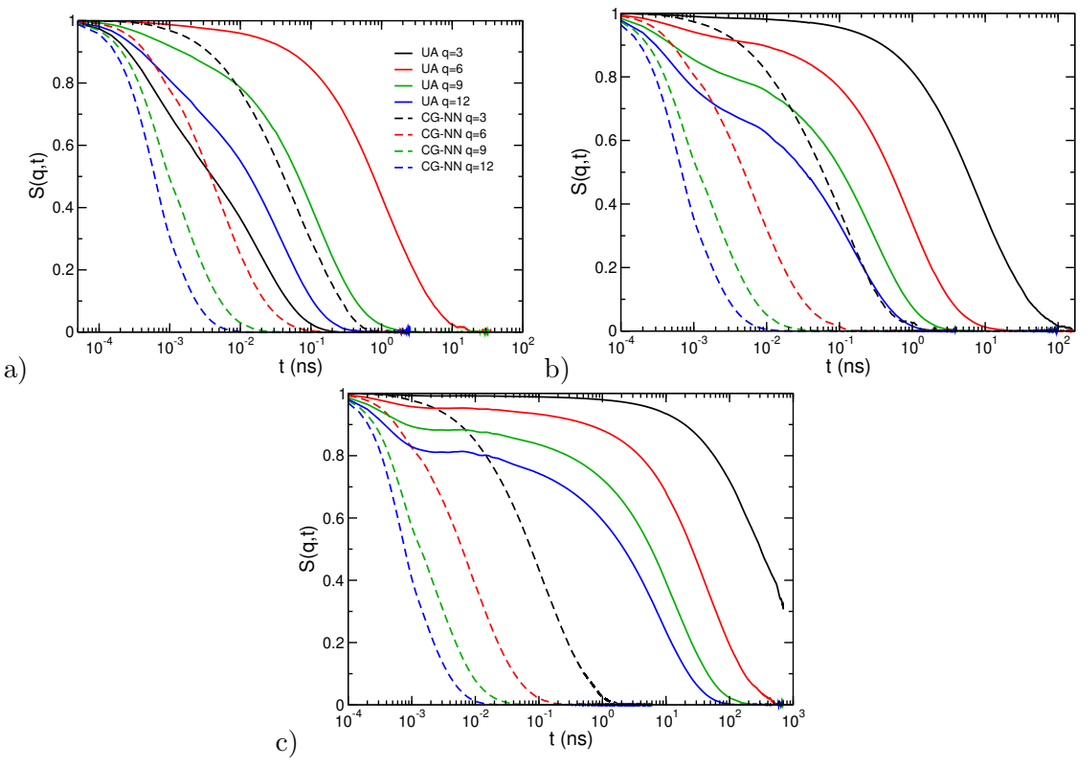

FIG. 8. Single chain dynamic structure factor S(q,t) from UA and CG-NN models at $q = 3, 6, 9,$ and 12 nm^{-1}, at different temperatures: a) 413 K, b) 353 K and c) 298 K.

factor $S(q)$ occurs at $q = 18$ nm^{-1}. Moreover, as it was shown by Moe and Ediger[88] for high q values (above 12 nm^{-1}) the total coherent $S(q,t)$ is dominated by intramolecular contributions. The data shown are presented as calculated from the original CG simulations before any time re-scaling. Thus, as expected, $S(q,t)$ from the CG model decorrelate much faster than the atomistic ones.

A convenient dynamical quantity, used typically in NSE experiments to describe the relaxation of over associated length scales of the polymer chain, is a characteristic time τ_{NSE} defined as the area under a curve of $S(q,t)$. To compute τ_{NSE} values from the atomistic and the CG-NN dynamic simulations, we modeled the curves shown in Fig. 8 using two terms, a fast exponential and a slower stretched exponential one:

$$S(q,t) = \alpha \exp(-t/\tau_0) + (1-\alpha)\exp(-(t/\tau_1)^{\beta}) \tag{12}$$

Fig. 9 shows the relaxation time τ_{NSE}, for various q values, derived via the above procedure for both the UA, τ_{NSE}^{UA}, and the CG, τ_{NSE}^{CG}, PI models. As q increases, $S(q,t)$ describes the dynamics at smaller length scales of a polymer chain and as expected the dynamics

become faster, i.e. the characteristic time τ_{NSE} decreases. For example, τ_{NSE} from the atomistic simulations at 413 K ranges from about 2 ns for $q = 3$ nm^{-1} to about 20 ps for $q = 12$ nm^{-1}. The CG models follow the same trend for different q values supporting further that similar dynamic characteristics are probed by the two models.

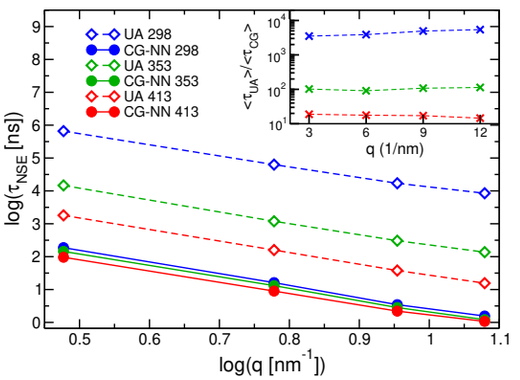

FIG. 9. Average relaxation time of the single chain structure factor (τ_{NSE}) as a function of the scattering vector (q) at 413 K (red), 353 K (green) and 298 K (blue).

As a final check of the consistency of the CG density dependent PI models to reproduce qualitatively, the PI dynamics at multiple length scales, we report in Fig. 9 the $\tau^{UA}_{NSE}/\tau^{CG}_{NSE}$ ratios for all systems studied here. For each temperature the ratio $\tau^{UA}_{NSE}/\tau^{CG}_{NSE}$ is very similar to the time mapping factor, τ, reported above (see Fig. 7). This is not surprising since the relaxation of $S(q,t)$ is a measure of the Brownian motion of the PI chains, and as mentioned above, τ is a measure of the ratio between the atomistic and the CG PI friction coefficient (Eq. 10). Nevertheless, it is interesting that the ratio $\tau^{UA}_{NSE}/\tau^{CG}_{NSE}$ does not depend on the value of the q vector. This is a clear indication that the CG models predict consistently the long-time dynamics of bulk PI, at multiple length scales, from the chain dimension (radius of gyration) down to the size of a CG bead. However, naturally there are differences between S(q,t) of atomistic and CG PI in the short time regime, as shown in Fig. 8. This is clear in the first local decay of the atomistic model, which develops with increasing q and decreasing T, while it is absent in the CG one. The above is in agreement with the short time heterogeneous segmental dynamics of the atomistic PI at lower temperatures (see Fig. 7), which, as mentioned above, cannot be described by the CG model.

Finally, note the strong dependence of $\tau^{UA}_{NSE}/\tau^{CG}_{NSE}$ on temperature. Similar to the scaling factor τ, it varies from around 20 for the higher temperature (413 K) up to several thousands

for the lower one (298 K). Such a strong temperature dependence further emphasizes the different temperature dependence of the friction factor of the atomistic and the CG models.

IV. CONCLUSIONS

We developed three CG models for polyisoprene melts and demonstrated the application of multi-body neighbor potentials based on the local environment proposed by a more detailed united atom model. We introduced such interaction terms through an order parameter with a well-defined free energy minimum ($\approx 14 - 15$ first neighbors) based on a geometric analysis. Nearest-neighbor potentials facilitate CG forcefield development by driving the system to the target density under constant pressure simulations in addition to offering another set of parameters to improve matching the target distributions.

The implementation of many-body potentials is a developing area with several options to be explored such as the applicability of isotropic sampling of nearest-neighbors, the range and shape of the introduced indicator function, the shape of the multi-body potential and most importantly the algorithm to be employed to develop such terms. All these options play a prominent role in the success of this approach. For the system studied herein, at all state-points examined, we found that solely attractive multi-body interactions and isotropic sampling suffice to provide an accurate description of the structure at all but the lowest temperature point. We observed small deviations at 298 K (both in terms of spherical packing as well as in the compressibility of the CG system), indicating that there could be cases were improvements might be sought with on-the-fly calculations of the nearest-neighbors. This was also the motivation for the original design of the SANN algorithm. We found that a purely attractive multi-body term performs equally well to one that presents a clear minimum; the pair nonbonded interactions adjust accordingly to match the target distributions. With regards to the algorithm, it is important to emphasize that the application of the Inverse Monte Carlo method, despite the numerical challenges it imposes, offers clear advantages over sequential refinement of each potential term. Beyond the clear coupling between pair and multi-body non-bonded terms we have demonstrated that for polymers additional complications arise; i.e. the dihedral potential can be coupled to the nearest-neighbor distributions.

Furthermore, we've characterized the dynamics of the developed CG models at different

temporal and length-scales. We found that in all cases, the dynamics were accelerated with a ratio that is highly sensitive on the temperature studied. Most importantly, we provided evidence that the introduction of the many-body term did not alter the character of the relaxation modes probed at different scattering vectors.

The bottom-up coarse graining methodology for PI presented here, is in general applicable to other polymeric systems. From the methodology point of view we would like to relate the hybrid particle-density field CG polymer models, as the one presented here, with many-body CG potentials based on rigorous renormalization group theory approaches, such as the cluster expansion techniques.[49] In addition, future studies will focus both on improving temperature transferability and exploring different systems, *i.e.* polymer mixtures, copolymers or heterogeneous nanostructured systems in the presence of polymer/solid or polymer/liquid interfaces.[3,4,91,92] The development and implementation of multi-body interaction terms for such systems presents significant potential as well as challenges.

REFERENCES

[1]K. Binder, *Monte Carlo and molecular dynamics simulations in polymer science* (Oxford University Press, 1995).

[2]F. Müller-Plathe, Chem. Phys. Chem. **3**, 754 (2002).

[3]Y. N. Pandey, G. J. Papakonstantopoulos, and M. Doxastakis, Macromolecules **46**, 5097 (2013).

[4]K. Johnston and V. Harmandaris, Soft Matter **9**, 6696 (2013).

[5]R. Potestio, C. Peter, and K. Kremer, Entropy **16**, 4199 (2014).

[6]N. J. H. Dunn, T. T. Foley, and W. G. Noid, Acc. Chem. Res. **49**, 2832 (2016).

[7]E. Kalligiannaki, A. Chazirakis, A. Tsourtis, M. Katsoulakis, P. Plecháč, and V. Harmandaris, Eur. Phys. J. Spec. Top. **225**, 1347 (2016).

[8]M. Dinpajooh and M. G. Guenza, J. Phys. Chem. B **122**, 3426 (2018).

[9]W. Schommers, Phys. Rev. A **28**, 3599 (1983).

[10]A. Soper, Chem. Phys. **202**, 295 (1996).

[11]D. Reith, M. Pütz, and F. Müller-Plathe, J. Comput. Chem. **24**, 1624 (2003).

[12]R. Faller, Polymer **45**, 3869 (2004).

[13]Q. Sun and R. Faller, J. Chem. Theory Comput. **2**, 607 (2006).

[14]Q. Sun and R. Faller, J. Chem. Phys. **126**, 144908 (2007).

[15]Y. N. Pandey, A. Brayton, C. Burkhart, G. J. Papakonstantopoulos, and M. Doxastakis, J. Chem. Phys. **140**, 054908 (2014).

[16]T. Spyriouni, C. Tzoumanekas, D. Theodorou, F. Müller-Plathe, and G. Milano, Macromolecules **40**, 3876 (2007).

[17]V. Harmandaris, N. Adhikari, N. van der Vegt, and K. Kremer, Macromolecules **39**, 6708 (2006).

[18]G. G. Zhang, A. Chazirakis, V. Harmandaris, T. Stuehn, K. Daoulas, and K. Kremer, Soft Matter **15**, 289 (2019).

[19]H. Wang, C. Junghans, and K. Kremer, Eur. Phys. J. E **28**, 221 (2009).

[20]M. Guenza, Eur. Phys. J. Spec. Top. **224**, 2177 (2015).

[21]M. Dinpajooh and M. G. Guenza, Soft Matter **14**, 7126 (2018).

[22]S. Izvekov and G. A. Voth, J. Chem. Phys. **123**, 134105 (2005).

[23]T. Murtola, E. Falck, M. Karttunen, and I. Vattulainen, J. Chem. Phys. **126**, 075101

(2007).

[24]S. Izvekov, P. W. Chung, and B. M. Rice, J. Chem. Phys. **133**, 064109 (2010).

[25]G. Maurel, B. Schnell, F. Goujon, M. Couty, and P. Malfreyt, J. Chem. Theory Comput. **8**, 4570 (2012).

[26]A. Das and H. C. Andersen, J. Chem. Phys. **132**, 164106 (2010).

[27]C.-C. Fu, P. M. Kulkarni, M. Scott Shell, and L. Gary Leal, J. Chem. Phys. **137**, 164106 (2012).

[28]T. Ohkuma and K. Kremer, Polymer **130**, 88 (2017).

[29]J. E. Mark, J. Am. Chem. Soc. **88**, 4354 (1966).

[30]M. Doxastakis, V. G. Mavrantzas, and D. N. Theodorou, J. Chem. Phys. **115**, 11339 (2001).

[31]N. J. H. Dunn and W. G. Noid, J. Chem. Phys. **143**, 243148 (2015).

[32]D. Rosenberger and N. F. A. van der Vegt, Phys. Chem. Chem. Phys. **20**, 6617 (2018).

[33]M. S. Daw and M. I. Baskes, Phys. Rev. Lett. **50**, 1285 (1983).

[34]M. S. Daw and M. I. Baskes, Phys. Rev. B **29**, 6443 (1984).

[35]I. Pagonabarraga and D. Frenkel, Mol. Simul. **25**, 167 (2000).

[36]S. Merabia and I. Pagonabarraga, J. Chem. Phys. **127**, 054903 (2007).

[37]E. C. Allen and G. C. Rutledge, J. Chem. Phys. **128**, 154115 (2008).

[38]E. C. Allen and G. C. Rutledge, J. Chem. Phys. **130**, 034904 (2009).

[39]S. Izvekov, P. W. Chung, and B. M. Rice, J. Chem. Phys. **135**, 044112 (2011).

[40]G. Faure, J.-B. Maillet, and G. Stoltz, J. Chem. Phys. **140**, 114105 (2014).

[41]T. Sanyal and M. S. Shell, J. Chem. Phys. **145**, 034109 (2016).

[42]J. D. Moore, B. C. Barnes, S. Izvekov, M. Lsal, M. S. Sellers, D. E. Taylor, and J. K. Brennan, J. Chem. Phys. **144**, 104501 (2016).

[43]V. Agrawal, G. Arya, and J. Oswald, Macromolecules **47**, 3378 (2014).

[44]J. W. Wagner, T. Dannenhoffer-Lafage, J. Jin, and G. A. Voth, J. Chem. Phys. **147**, 044113 (2017).

[45]J. P. Larentzos, J. M. Mansell, M. Lsal, and J. K. Brennan, Mol. Phys. **0**, 1 (2018).

[46]T. Sanyal and M. S. Shell, J. Phys. Chem. B **122**, 5678 (2018).

[47]D. Rosenberger, T. Sanyal, M. S. Shell, and N. F. van der Vegt, J. Chem. Theory Comput. , 2881 (2019).

[48]V. Agrawal, P. Peralta, Y. Li, and J. Oswald, J. Chem. Phys. **145**, 104903 (2016).

[49]A. Tsourtis, V. Harmandaris, and D. Tsagkarogiannis, Entropy **19**, 395 (2017).

[50]M. Doxastakis, V. G. Mavrantzas, and D. N. Theodorou, J. Chem. Phys. **115**, 11352 (2001).

[51]V. A. Harmandaris, M. Doxastakis, V. G. Mavrantzas, and D. N. Theodorou, J. Chem. Phys. **116**, 436 (2002).

[52]M. Doxastakis, D. N. Theodorou, G. Fytas, F. Kremer, R. Faller, F. Müller-Plathe, and N. Hadjichristidis, J. Chem. Phys. **119**, 6883 (2003).

[53]P. Sharma, S. Roy, and H. A. Karimi-Varzaneh, J. Phys. Chem. B **120**, 1367 (2016).

[54]M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, SoftwareX **1**, 19 (2015).

[55]M. Parrinello and A. Rahman, J. Appl. Phys. **52**, 7182 (1981).

[56]S. Nosé and M. Klein, Mol. Phys. **50**, 1055 (1983).

[57]S. Nosé, Mol. Phys. **52**, 255 (1984).

[58]W. G. Hoover, Phys. Rev. A **31**, 1695 (1985).

[59]W. Graessley and S. Edwards, Polymer **22**, 1329 (1981).

[60]R. A. Orwoll, "Densities, coefficients of thermal expansion, and compressibilities of amorphous polymers," in *Physical Properties of Polymers Handbook*, edited by J. E. Mark (Springer New York, New York, NY, 2007) pp. 93–101.

[61]Y. Yi and P. Zoller, J. Polym. Sci., Part B: Polym. Phys. **31**, 779 (1993).

[62]M. S. Uddin and J. Ju, Polymer **101**, 34 (2016).

[63]L. B. Lucy, Astron. J. **82**, 1013 (1977).

[64]R. A. Gingold and J. J. Monaghan, Mon. Not. R. Astron. Soc **181**, 375 (1977).

[65]P. Espanol and M. Revenga, Phys. Rev. E **67**, 026705 (2003).

[66]A. Okabe, B. Boots, K. Sugihara, and S. N. Chiu, *Spatial tessellations: concepts and applications of Voronoi diagrams*, Vol. 501 (John Wiley & Sons, 2009).

[67]D. E. Knuth, *Art of computer programming, volume 2: Seminumerical algorithms* (Addison-Wesley Professional, 2014).

[68]P. W. Jones, A. Osipov, and V. Rokhlin, Proc. Natl. Acad. Sci. U.S.A **108**, 15679 (2011).

[69]J. A. van Meel, L. Filion, C. Valeriani, and D. Frenkel, J. Chem. Phys. **136**, 234107 (2012).

[70]S. Trofimov, E. Nies, and M. Michels, J. Chem. Phys. **117**, 9383 (2002).

[71]P. B. Warren, Phys. Rev. E **68**, 066702 (2003).

[72]P. B. Warren, Phys. Rev. E **87**, 045303 (2013).

[73]A. P. Lyubartsev and A. Laaksonen, Phys. Rev. E **52**, 3730 (1995).

[74]A. P. Lyubartsev and A. Laaksonen, "On the reduction of molecular degrees of freedom in computer simulations," in *Novel Methods in Soft Matter Simulations*, edited by M. Karttunen, A. Lukkarinen, and I. Vattulainen (Springer Berlin Heidelberg, Berlin, Heidelberg, 2004) pp. 219–244.

[75]A. Savelyev and G. A. Papoian, Biophys. J. **96**, 4044 (2009).

[76]I. Pagonabarraga and D. Frenkel, J. Chem. Phys. **115**, 5015 (2001).

[77]Y. N. Pandey and M. Doxastakis, J. Chem. Phys. **136**, 094901 (2012).

[78]D. Frenkel and B. Smit, *Understanding Molecular Simulation From Algorithms to Applications* (Academic Press, California, 2002).

[79]A. Lyubartsev, A. Mirzoev, L. Chen, and A. Laaksonen, Faraday Discuss. **144**, 43 (2010).

[80]D. Rosenberger, M. Hanke, and N. F. van der Vegt, Eur. Phys. J. Spec. Top. **225**, 1323 (2016).

[81]M. R. DeLyser and W. G. Noid, J. Chem. Phys. **147**, 134111 (2017).

[82]T. C. Moore, C. R. Iacovella, and C. McCabe, J. Chem. Phys. **140**, 06B606_1 (2014).

[83]T. C. Moore, C. R. Iacovella, R. Hartkamp, A. L. Bunge, and C. McCabe, J. Phys. Chem. B **120**, 9944 (2016).

[84]M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, New York, 1987).

[85]V. Harmandaris and K. Kremer, Macromolecules **42**, 791 (2009).

[86]V. Harmandaris and K. Kremer, Soft Matter **5**, 3920 (2009).

[87]V. Harmandaris, N. Adhikari, N. van der Vegt, K. Kremer, B. Mann, R. Voelkel, C. Liew, and H. Weiss, Macromolecules **40**, 7026 (2007).

[88]N. E. Moe and M. Ediger, Phys. Rev. E **59**, 623 (1999).

[89]J. Salacuse, A. Denton, and P. Egelstaff, Phys. Rev. E **53**, 2382 (1996).

[90]V. Harmandaris, V. Mavrantzas, D. Theodorou, M. Kröger, J. Ramirez, H. C. Öttinger, and D. Vlassopoulos, Macromolecules **36**, 1376 (2003).

[91]M. Doxastakis, K. Chrissopoulou, A. Aouadi, B. Frick, T. P. Lodge, and G. Fytas, J. Chem. Phys. **116**, 4707 (2002).

[92]V. Harmandaris and M. Doxastakis, J. Chem. Phys. **139**, 034904 (2013).