

## Structure-based coarse-graining in liquid slabs

Mara Jochum, Denis Andrienko, Kurt Kremer, and Christine Peter  
*Max Planck Institute for Polymer Research, Ackermannweg 10, 55128 Mainz, Germany*

(Received 25 April 2012; accepted 19 July 2012; published online 8 August 2012)

Structure-based coarse-graining relies on matching the pair correlation functions of a reference (atomistic) and a coarse-grained system. As such, it is designed for systems with uniform density distributions. Here, we demonstrate how it can be generalized for inhomogeneous systems by coarse-graining slabs of liquid water and methanol in vacuum, as well as a single benzene molecule at the water-vacuum interface. Our conclusion is that coarse-graining performed in inhomogeneous systems improves thermodynamic properties and the structure of interfaces without significant alterations to the local structure of the bulk liquid. © 2012 American Institute of Physics. [<http://dx.doi.org/10.1063/1.4742067>]

### I. INTRODUCTION

Coarse-graining (CG) is a procedure which reduces the number of degrees of freedom used to model a physical system. By introducing softer interaction potentials, it accelerates diffusion processes,<sup>1</sup> increases intrinsic length scales, and consequently provides access to longer simulation time- and length-scales. CG models have been developed for various types of systems, such as polymer melts,<sup>2</sup> organic solvents,<sup>3–5</sup> lipid membranes,<sup>6,7</sup> conjugated polymers,<sup>8–10</sup> peptides,<sup>11</sup> surfactants,<sup>12</sup> and proteins.<sup>13</sup>

To derive CG potentials, one needs to project a many-body potential of mean force onto a CG force field.<sup>14</sup> The projection operator is not unique and depends on the set of thermodynamic properties of the system which should be preserved during coarse-graining. Force matching, for example, aims to reproduce the entire distribution of states in a canonical ensemble,<sup>14–17</sup> while structure-based coarse-graining such as iterative Boltzmann inversion<sup>18</sup> (IBI) or inverse Monte Carlo<sup>19</sup> try to match the radial distribution functions (RDFs) of the CG and reference systems.

In this paper, we will focus on iterative Boltzmann inversion, a numerical scheme which refines a CG potential,  $U^{cg}$ , while attempting to match the reference (atomistic) and coarse-grained radial distribution functions,  $g^{at}(r)$  and  $g^{cg}(r)$ , respectively, where  $r$  is a set of coarse-grained degrees of freedom. As an initial guess, it often uses the potential of mean force,  $U^{(0)}(r) = -k_B T \ln g^{at}(r)$ . At each subsequent iteration  $i$ , a correction is added to the interaction potential based on the differences in  $\ln g(r)$

$$U^{(i+1)}(r) = U^{(i)}(r) + \alpha \Delta U^{(i)}(r), \quad (1)$$

$$\Delta U^{(i)}(r) = k_B T \ln \frac{g^{(i)}(r)}{g^{at}(r)}, \quad (2)$$

where  $g^{(i)}(r)$  is the radial distribution function of the CG system at iteration  $i$ , and  $\alpha \in (0, 1]$  is a scaling factor used to control the stability of the scheme. Convergence is monitored

by evaluating the difference in RDFs

$$\Delta g^{(i)} = \int_0^{r_c} [g^{(i)}(r) - g^{at}(r)]^2 dr, \quad (3)$$

where  $r_c$  is the cutoff distance for the CG interaction potential. Once a given convergence criterion has been met (in this case  $\Delta g < 10^{-3}$  nm), the update is terminated and one has obtained a CG potential which reproduces the RDF of the reference system.

In many cases, it is desirable to adjust additional thermodynamic properties, such as the pressure of the CG system.<sup>18,20</sup> This can only be achieved by relaxing the convergence criterion<sup>21</sup> in Eq. (3) or, in other words, by seeking potential updates which offer a compromise between structure and thermodynamic properties.

By construction, IBI relies on the radial distribution function which is only defined for homogeneous systems. Hence, in order to treat any phenomena which take place at interfaces (i.e., the aqueous/organic interface in biological cells) an extension of IBI to inhomogeneous systems is required. In this paper, we will show how IBI can be performed on a liquid slab as well as on a solute molecule solvated within this slab.

### II. METHODOLOGY

In a homogeneous system, the radial distribution function,  $g(r)$ , is calculated by counting the number of particles,  $N_{\text{shell}}(r)$ , in a spherical shell of a radius  $r$  and thickness  $\Delta r$ . This number is normalized by the shell volume,  $V_{\text{shell}}(r) = 4\pi r^2 \Delta r$ , and the number density,  $\rho = N/V$ , to ensure that  $g(r) = 1$  at large  $r$ . To improve accuracy,  $g(r)$  is then averaged over all particles in the system

$$g(r) = \frac{\langle N_{\text{shell}}(r) \rangle}{\rho V_{\text{shell}}(r)}, \quad (4)$$

where  $\langle \dots \rangle$  denotes an ensemble average.

If the same protocol is applied to a system consisting of a slab of thickness  $w$ , sandwiched between two vacuum layers in a box of length  $l_s \geq 2w$  (see Fig. 1(a)), two things change. First, the number density used for normalization,  $\rho_s = N/V_s$ , becomes smaller due to a larger size of the simulation box in

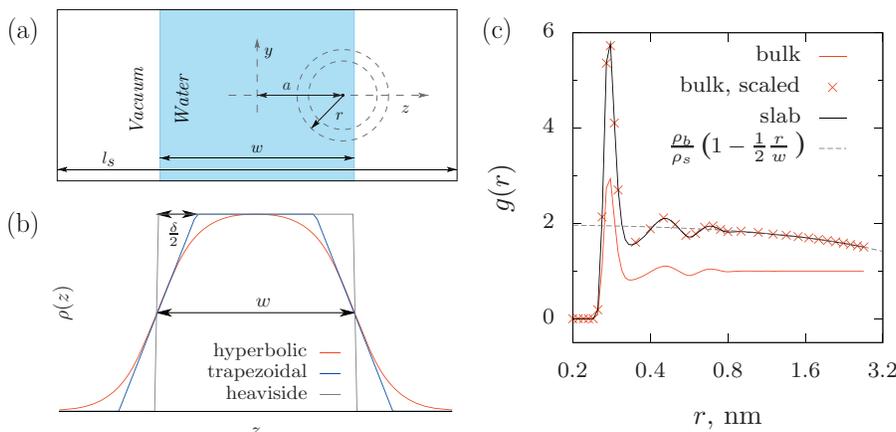


FIG. 1. (a) Schematic diagram of a liquid slab of width  $w$  in a simulation box of length  $l_s$ . A partially filled shell used for RDF calculations is shown. (b) Density profiles,  $\rho(z)$ . The hyperbolic function, Eq. (A6), is used to fit simulation data to determine  $w$ ,  $\delta$ , and  $\rho_0$ . Trapezoidal and Heaviside functions are used to obtain the ratio of RDFs calculated in a slab and in bulk. (c) Radial distribution functions of atomistic water calculated in the bulk and slab systems, as well as a slab RDF obtained using Eq. (7). The ratio between the RDFs is also shown for a density profile with sharp interfaces (Eq. (7)).

the  $z$ -direction,  $l_s$ , and second, the number of particles in the shell is no longer uniform and becomes a function of the position of the shell center,  $a$ . This number can be estimated by assuming that only the density, but not the local structure of the liquid changes within the slab (see Appendix C for justification)

$$N_{\text{shell}}(r, a) = 2\pi r \Delta r g_b(r) \int_{a-r}^{a+r} dz \rho(z), \quad (5)$$

where  $g_b(r)$  is the RDF of a pure bulk system (without interfaces),  $\rho(z)$  is the number density which depends on the  $z$  coordinate, and  $a$  is the distance from the shell center to the symmetry plane of the slab (as shown in Fig. 1(b)). Averaging over all particles, which is equivalent to integrating  $N_{\text{shell}}(r, a)$  over  $\rho(a)da$  and appropriately normalizing, one obtains

$$g_s(r) = g_b(r) \frac{\int_{-l_s/2}^{l_s/2} da \rho(a) \int_{a-r}^{a+r} dz \rho(z)}{2r \rho_s \int_{-l_s/2}^{l_s/2} da \rho(a)}. \quad (6)$$

Here,  $g_s(r)$  is an object which one obtains by using a standard procedure of calculating an RDF for a system with a slab. In what follows, we will refer to it as a *slab* RDF.

To illustrate how Eq. (6) works in practice, let us first consider a slab with two *sharp* interfaces. In this case,  $\rho(z)$  can be written as a sum of two Heaviside step functions (see Fig. 1(b)) and, for  $r \leq w$ , Eq. (6) simplifies to

$$g_s(r) = g_b(r) \frac{\rho_b}{\rho_s} \left(1 - \frac{r}{2w}\right), \quad (7)$$

where  $r < w < l_s$ ,  $w$  is the width of the slab, and  $\rho_b = N/V_b$  is the number density of the bulk system. One can see that apart from the local structure of the liquid,  $g_s(r)$  also contains information about the slab width and total number density. The transformation between the bulk and slab RDFs (Eq. (7)) is illustrated for atomistic SPC/E water in Fig. 1(c). Here,  $g_b(r)$  obtained from simulations of bulk water, when scaled, lies exactly on top of  $g_s(r)$  obtained from simulations of a slab of water. Note, that the slab system was prepared from the smaller, equilibrated using periodic boundary condi-

tions, bulk system by increasing the box size in the  $z$  direction by a factor of two, such that  $\rho_b/\rho_s = 2$ .

It is not surprising that Eq. (7) works perfectly for a water slab, since its interfaces are very sharp. For more diffuse interfaces it is still possible to obtain an analytical relation between the RDFs by approximating the slab density profile by either a trapezoidal (see Fig. 1(b)) or a more complicated shape, as discussed in Appendix A. Moreover, a numerical relation, which also accounts for periodic boundaries and was used in all examples presented below, can be established using Fourier series, as discussed in Appendix B.

The relation between the slab and bulk RDFs enables us to rewrite the potential update, Eq. (2), for a slab system. Substituting Eq. (7) into Eq. (2) one obtains

$$\frac{\Delta U^{(i)}}{k_B T} = \ln \frac{g_b^{(i)}(r)}{g_b^{\text{at}}(r)} = \ln \frac{g_s^{(i)}(r)}{g_s^{\text{at}}(r)} - \ln \frac{2w^{(i)} - r}{2w^{\text{at}} - r}, \quad (8)$$

where  $w^{\text{at}}$  and  $w^{(i)}$  are, respectively, slab widths of the reference and coarse-grained (at iteration  $i$ ) systems. Note that we neglected all constant offsets to the potential. One can see that  $\Delta U^{(i)}$  splits into two contributions. The first depends on the ratio of the coarse-grained to atomistic *slab* RDFs, and the second is a function of the widths of the atomistic and coarse-grained slabs.

The derived update (Eq. (8)) might give the impression that it is sufficient to match *bulk* RDFs in order to satisfy  $w^{\text{cg}} = w^{\text{at}}$ . This is, however, incorrect, since structural coarse-graining does not preserve all thermodynamic properties of a system,<sup>20,22,23</sup> among them the interfacial width of the atomistic reference. To illustrate this, we performed conventional IBI with pressure correction on a bulk system to obtain a one-site CG water model, the tabulated pair interaction potential of which is shown in Fig. 3(a) (blue dashed line). As expected, simulations of a water slab using this model produced very wide and diffuse interfaces (see Fig. 3(c)) which, even though *bulk* RDFs were perfectly matched, resulted in different *slab* RDFs. In other words,  $w^{\text{cg}} \neq w^{\text{at}}$  and  $g_s^{\text{cg}}(r) \neq g_s^{\text{at}}(r)$ . Note that this effect is especially severe in the case of a one-site coarse-grained *water* model, as it lacks

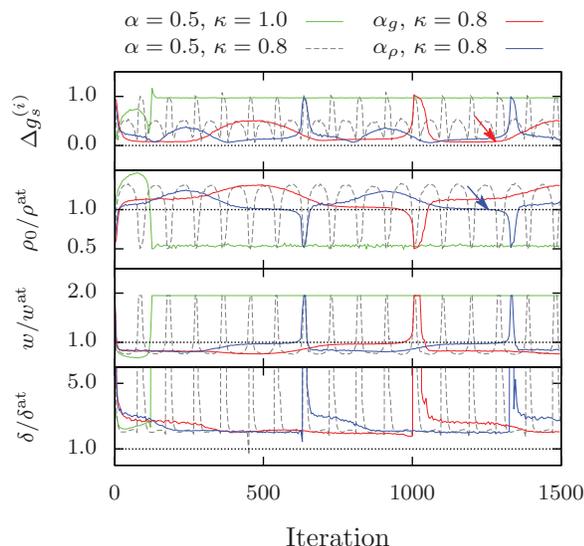


FIG. 2. Convergence for slab RDF,  $\Delta g_s$ , maximum of the density in a slab,  $\rho_0$ , width of the slab,  $w$ , and width of the interface,  $\delta$ , all normalized by the reference values taken from atomistic simulations of SPC/E water,  $\rho^{\text{at}} = 994.858 \text{ kg/m}^3$ ,  $w^{\text{at}} = 5.4 \text{ nm}$ , and  $\delta^{\text{at}} = 0.43 \text{ nm}$ .

three-body contributions to the potential of mean force.<sup>23</sup> A consistent approach to this problem is to either include a three-body interaction potential in the one-site water model<sup>23</sup> or to switch to a different representation which can reduce the role of such contributions, e.g., by mapping more than one water molecule onto a CG bead.<sup>7,12,24</sup>

If, however, a one-site representation with a pair interaction potential is still desired, one should try to find a balance between matching the pair distribution function and the thermodynamic properties of the system, in our case the interfacial tension of the interface. Equation (7) offers a practical solution as one can perform all simulations in a system with a slab, determine interfacial and slab widths,  $\delta$  and  $w$ , calculate slab RDFs,  $g_s(r)$ , and subsequently change the interaction potential according to Eq. (8).

This idea, however, has both technical and conceptual issues. On the technical side, Eq. (8) neither takes into account periodic boundary conditions present in MD simulations nor is it exact due to diffuse interfaces. These issues can, however,

be improved by employing more realistic fitting functions for the density profile, or, ultimately, the Fourier series, as described in Appendices A and B.

Conceptually, the Henderson theorem states that there exists a unique (up to an additive constant) interaction potential which reproduces a given RDF of a homogeneous system. Our experience also teaches us that this is the stationary point of the IBI method. Since this statement does not depend on the density of the system, same RDFs can correspond to different interaction potentials if the respective densities are different. The immediate implication is that if we now run IBI on an inhomogeneous system using the exact form of Eq. (8), the system will eventually become homogeneous with the RDF reproducing that of the atomistic reference but an interaction potential which corresponds to a lower density (determined by the size of the simulation box). This is illustrated in Fig. 2 (green solid line), where we show the convergence of the slab RDF, density, interfacial and slab widths as a function of iteration. Independently from the initial density profile, the system quickly becomes homogeneous and the reference RDF is perfectly matched ( $\Delta g = 0 \text{ nm}$ ).

However, a spatially homogeneous density distribution is not what we would like to achieve here. Equation (8) also offers a solution in this case. Since its second term depends only on the density profile, one can introduce a bias which will make the homogeneous solution unstable. The simplest way of doing this is via a scaling factor for the density-dependent term, e.g., in the case of sharp interfaces, Eq. (8), the potential update becomes

$$\frac{\Delta U^{(i)}}{k_B T} = \ln \frac{g_s^{(i)}(r)}{g_s^{\text{at}}(r)} - \kappa \ln \frac{2w^{(i)} - r}{2w^{\text{at}} - r}, \quad \kappa < 1. \quad (9)$$

The scaling  $\kappa$  effectively adds a long-range attractive term to the potential once the interface dissociates and hence destabilizes the homogeneous solution, leading to oscillations between the homogeneous and inhomogeneous density distributions. Such oscillations are shown in Fig. 2 for  $\kappa = 0.8$  (gray dashed line) and tell us that the homogeneous density distribution is no longer a stationary point of the iterative scheme, since the structure- and density-dependent components of the update are unbalanced.

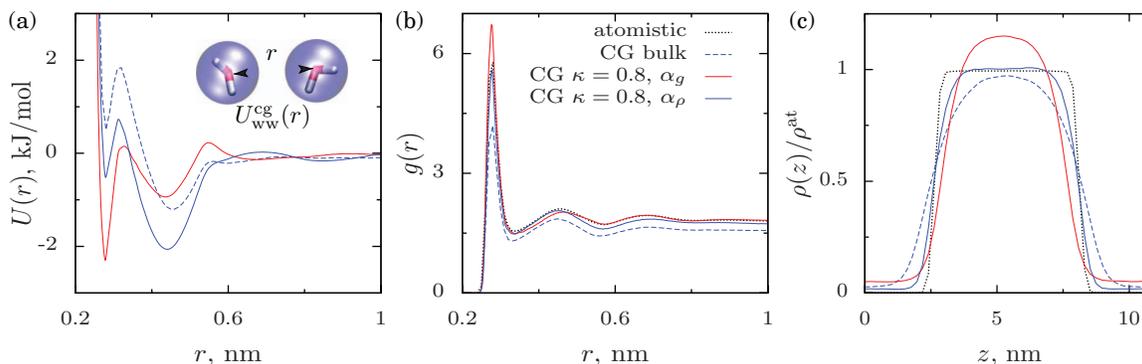


FIG. 3. (a) Water-water pair interaction potentials,  $U(r)$ , (b) radial distribution functions,  $g(r)$ , and (c) slab density profiles,  $\rho(z)$ , for systems coarse-grained in the bulk (CG bulk), in a slab using the update from Eq. (8) using slab RDFs (CG  $\kappa = 0.8$ ,  $\alpha_g$ ), or densities (CG  $\kappa = 0.8$ ,  $\alpha_\rho$ ) indicated by red and blue arrows in Fig. 2). For RDFs and density profiles the atomistic reference is also shown (atomistic).

To find a compromise between liquid structure and density profile, one can slow down the oscillation dynamics by making a global scaling factor  $\alpha$  in Eq. (2) dependent on the convergence of the tail of the slab RDF,  $\alpha_g = |1 - g(r_c)/g^{\text{at}}(r_c)|$ , where  $r_c$  is the cutoff distance. Similarly, by introducing  $\alpha_\rho = |1 - \rho_0/\rho_0^{\text{at}}|$ , one could ensure that the bulk densities,  $\rho_0$  and  $\rho_0^{\text{at}}$ , defined as the maximum of the density profile, match. The respective system behavior is shown in Fig. 2 ( $\kappa = 0.8$ ,  $\alpha_g$  and  $\alpha_\rho$ , red and blue solid lines). The “pseudo-stationary” points (depicted by the arrows) can be clearly identified and the corresponding potentials which, as we will see in the following examples, provide a good compromise between reproducing the thermodynamic and structural properties of a liquid.

### III. EXAMPLES

#### A. Slab of SPC/E water

Putting this into practice, IBI was performed on a slab of SPC/E water using  $g_s^{\text{at}}(r)$  as a target and the pressure-corrected potential for bulk water as an initial guess. All simulations were performed with the GROMACS simulation package,<sup>25</sup> while the coarse-graining tools (mapping, IBI) stem from the VOTCA package.<sup>26</sup> Atomistic simulations were performed in a box of approximately  $5.5 \text{ nm} \times 5.5 \text{ nm} \times 11 \text{ nm}$ , containing 5439 type SPC/E water molecules. The GROMOS53a6 force field was used to describe inter- and intramolecular interactions. For electrostatic interactions, the smooth particle mesh Ewald<sup>27</sup> method was employed using a real space cutoff of 1 nm, while the van der Waals cutoff was set to 1.4 nm. The temperature was coupled to 300 K via the Berendsen thermostat<sup>28</sup> and a coupling constant of 0.1 ps. Atomistic reference simulations were run for 20 ns with a time step of 2 fs, while CG simulations were performed for 100 ps (per iteration) with the same time step. Fourier-series and scaling factor  $\kappa = 0.8$  were used to construct the potential update (see Appendix B).

To analyze simulations, the density profile,  $\rho(z)$ , was fitted to a sum of two hyperbolic tangents (see Fig. 1(b) and Appendix A) at each iteration step. Fitted parameters, namely, the height of the slab density profile,  $\rho_0$ , the width of the slab,  $w$ , and the width of the interface,  $\delta$ , were monitored over the course of the simulation and are shown in Fig. 2, together with the convergence for the slab RDF,  $\Delta g_s$ . The points with matching at  $r_c$  slab RDFs or bulk densities were identified (these are shown with red and blue arrows) and the corresponding interaction potentials, RDFs, and density profiles are summarized in Fig. 3.

One can see that, depending on the potential update ( $\alpha_g$  or  $\alpha_\rho$  scaling), either the slab RDF or the density profile is better reproduced. This implies that a certain compromise has been achieved between the local structure, characterized by  $g_s(r)$ , and the geometry of the interface, represented by  $w$  and  $\delta$ . In other words, the two terms in the potential update (Eq. (8)) compensate each other within some numerical accuracy. Remarkably, in both cases the density profile and the slab RDF are significantly better than those obtained by using

interaction potentials based on the IBI procedure in the bulk system.

We have also calculated the surface tension from the pressure tensor  $P_{\alpha\beta}$

$$\gamma = \frac{1}{2} \int [2P_{zz}(z) - P_{xx}(z) - P_{yy}(z)] dz, \quad (10)$$

where  $P_{zz}(z)$  is the perpendicular component and  $P_{xx}(z)$  and  $P_{yy}(z)$  are the transverse components of the pressure tensor with respect to the plane of the interface. The atomistic reference system has a surface tension of  $\gamma^{\text{at}} = 59.4 \pm 0.3 \text{ mN/m}$ . All CG models reproduce this value reasonably close,  $\gamma_{\text{bulk}}^{\text{CG}} = 57.5 \pm 0.3 \text{ mN/m}$ ,  $\gamma_{\alpha_g, \kappa=0.8}^{\text{CG}} = 45.6 \pm 0.9 \text{ mN/m}$ , and  $\gamma_{\alpha_\rho, \kappa=0.8}^{\text{CG}} = 31.7 \pm 1.3 \text{ mN/m}$ . Surface tension  $\gamma$ , however, is a scalar which can be obtained by integrating an expression which includes the radial distribution function, the density profile at the interface, and the derivative of the interaction potential,<sup>29</sup> which makes it possible to have the same value of  $\gamma$  for very different  $\rho(z)$  and  $g(r)$ . As such, one should not validate the CG model on surface tension alone.

#### B. Slab of liquid methanol

We now investigate a slab of methanol, which although is in a liquid state, exhibits significant differences in behavior as compared to water. First, its liquid-vacuum interface is much wider ( $\delta^{\text{at}} \sim 0.6 \text{ nm}$ ) than that of water due to the substitution of one hydrogen by a methyl group. Second, a CG model with one site per molecule can reproduce both structural and thermodynamic properties of the bulk system better than the corresponding CG model of water, since similar coarse-grained potentials can be obtained from force-matching or conventional iterative Boltzmann inversion.<sup>26</sup>

In spite of these differences, the dynamics of convergence is qualitatively very similar: stationary homogeneous density distribution for  $\kappa = 1.0$ , oscillations between homogeneous and heterogeneous (with a slab) states for  $\kappa = 0.8$ , and slow oscillations for both  $\alpha_g$  and  $\alpha_\rho$ . The respective interaction potentials, radial distribution functions, and density profiles are shown in Fig. 4. Again, by coarse-graining in a slab, one can achieve much steeper density profiles, matching slab widths, while the reference and atomistic RDFs still reproduce each other. Note, that if a “naive” substitution of the IBI update is used ( $\kappa = 0$ ) neither the RDFs, nor the density profiles are well reproduced.

The surface tension of the atomistic reference, calculated using Eq. (10), was  $\gamma^{\text{at}} = 21.3 \pm 0.4 \text{ mN/m}$ , while the CG models yielded values of  $\gamma_{\text{bulk}}^{\text{CG}} = 20.3 \pm 0.2 \text{ mN/m}$ ,  $\gamma_{\alpha_g, \kappa=0.8}^{\text{CG}} = 10.1 \pm 0.3 \text{ mN/m}$ , and  $\gamma_{\alpha_\rho, \kappa=0.8}^{\text{CG}} = 6.4 \pm 0.3 \text{ mN/m}$ . We have also checked that the potentials obtained from the update using the Fourier series are transferable to systems of different slab widths ( $w$ ) as long as the interaction range is larger than  $w$  and there is a well defined plateau in the middle of the density profile.

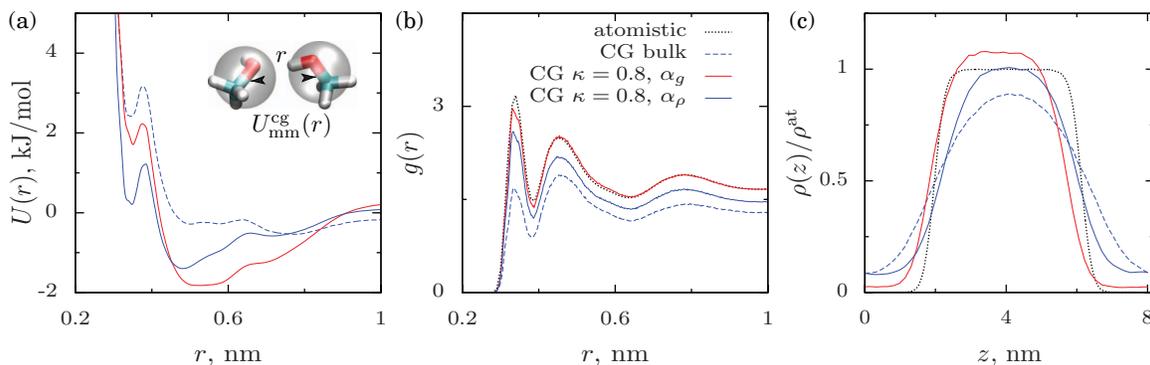


FIG. 4. (a) Methanol-methanol (centers of mass) pair interaction potentials,  $U(r)$ , (b) radial distribution functions,  $g(r)$ , and (c) slab density profiles,  $\rho(z)$ , for systems coarse-grained in the bulk (CG bulk) and in a slab using the Fourier update (CG  $\alpha = 0.5, \kappa = 1$ ). For RDFs and density profiles the atomistic reference is also shown (atomistic).

### C. Benzene in water

We will now describe a more complicated situation, that is, determining solute-solvent coarse-grained potentials. As an example, we consider a benzene molecule in water. Benzene is a hydrophobic molecule and hence predominantly occupies the vacuum-water interface region. This has been confirmed by analyzing the distribution of its position,  $p^{at}(z)$ , along the  $z$  coordinate from atomistic simulations, as shown in Fig. 5(c). One can see that this distribution is peaked around the positions of the two water-vacuum interfaces of the slab. Simulation details were the same as for the water slab, but with smaller simulation box sizes of  $4 \text{ nm} \times 4 \text{ nm} \times 8 \text{ nm}$ . In addition, due to the low benzene concentration (1 benzene in 2175 water molecules), simulation times for both the atomistic references as well as the CG simulations have been increased to 100 ns and 10 ns (per iteration), respectively, to obtain better sampling. The benzene-water interaction was assigned a Van der Waals cutoff of 1.4 nm.

If one now employs a usual IBI coarse-graining procedure (in bulk) and reproduces the bulk RDF between the benzene bead (which is treated as a single interaction site) and the surrounding water molecules, simulations in a slab with this potential result in benzene distributions which are peaked around the slab center (shown in Fig. 5(c)). This means that

the benzene bead mostly occupies the bulk water region, a circumstance which is directly reflected in differences in *slab* RDFs of the coarse-grained and atomistic systems, as shown in Fig. 5(b). It arises due to the different probability distributions,  $p^{cg}(z)$  and  $p^{at}(z)$ , of the benzene positions along the slab, since now, in order to average over the ensemble of benzene positions, one has to integrate  $N_{\text{shell}}$  over  $p(a)da$  instead of  $\rho(a)da$  in order to obtain the equivalent of Eq. (6), as described in Appendix B.

In other words, to perform an update of the interaction potential analogous to Eq. (8), we sampled the benzene position along the slab and decomposed it into Fourier series in order to obtain the relation between  $g_s$  and  $g_b$ . Starting from an initial guess of a pressure-corrected potential for bulk methanol and applying this procedure we eventually recovered the shape of the atomistic slab RDF, as shown in Fig. 5(b).

By matching the slab RDFs, one automatically matches the distributions of the benzene position along the slab, as can be seen in Fig. 5(c). The benzene bead is now primarily located at the interfaces of the slab which coincides with its behavior in atomistic simulations. It is notable that the two coarse-grained potentials, shown in Fig. 5(a), are remarkably similar: the one obtained by coarse-graining in a slab has a slightly more long-range attractive tail and smaller repulsive

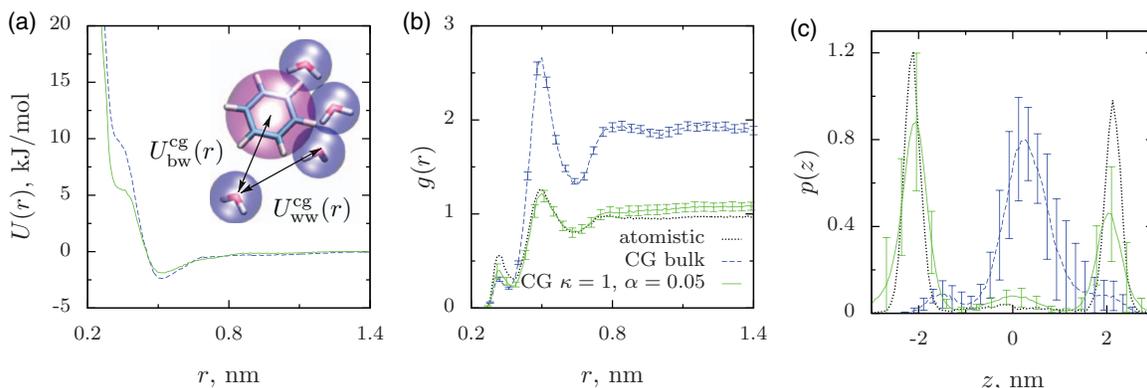


FIG. 5. (a) Benzene-water pair interaction potentials,  $U(r)$ , (b) benzene-water radial distribution functions,  $g(r)$ , and (c) probability distribution to find a benzene molecule at a position  $z$ ,  $p(z)$ . All plots show results of coarse-graining in the bulk (CG bulk) and in a slab using the Fourier-series update ( $\kappa = 1, \alpha = 0.05$ ). For RDFs and probability distributions the atomistic reference is also shown (atomistic).

bump at the beginning than the one obtained from a bulk system. This small difference, however, is what leads to the distinct hydrophobic behavior of benzene when simulated in a slab.

#### IV. CONCLUSIONS

In conclusion, we have proposed a scheme for obtaining coarse-grained potentials for inhomogeneous systems, analogous to the iterative Boltzmann inversion method. The main idea is to construct an update for the interaction potential based on the radial distribution function calculated in a slab geometry. Apart from the local liquid structure, this update also carries information about the geometry of the system, namely, the slab and interfacial widths. Since these geometric features are very sensitive to the thermodynamic properties of the system (surface tension, pressure tensor), a partially controllable balance between the local structure and thermodynamic properties can be achieved.

There are also drawbacks to the proposed method. First, simulation times for each iteration need to be increased to attain adequate accuracy when measuring thermodynamic properties of the system, such as a density profile or a distribution of positions of a solute within a solvent. Second, one needs to introduce a small scaling factor for the potential update in order to improve convergence and to avoid either evaporation or solidification of the system. Finally, it is not easy to achieve a controlled balance between the thermodynamic properties and the local structure.

In spite of these limitations, the method offers a practical way of finding a compromise between the local structural

and intensive thermodynamic properties of the system, and as such can be useful for building simple coarse-grained models for inhomogeneous and open systems.

#### APPENDIX A: DENSITY PROFILES AND UPDATE FUNCTIONS

In this section, we simplify the derived integral (Eq. (6)) which relates the bulk and slab RDFs and evaluate it for several functional forms of the density profile. The obtained analytical expressions may be used to analyze density profiles and construct updates for the interaction potentials.

In order to evaluate the integral

$$I(r) = \int_{-l_s/2}^{l_s/2} da \rho(a) \int_{a-r}^{a+r} dz \rho(z), \quad (\text{A1})$$

we assume that sufficiently far from the slab interfaces, the particle density vanishes such that all particles are (on average) located within the slab. Under this condition, the integral over  $da$  can be extended to cover all space  $[-\infty, \infty]$ . Furthermore, we introduce a new variable,  $t = z - a$ , and change the order of integration

$$I(r) = \int_{-r}^r dt \int_{-\infty}^{\infty} da \rho(a) \rho(t+a) = \int_{-r}^r dt h(t), \quad (\text{A2})$$

where  $h(t)$  is the correlation of  $\rho(a)$ . Using the correlation (convolution) theorem, we can write

$$\hat{h}(\xi) = \hat{\rho}(\xi)^2, \quad (\text{A3})$$

TABLE I. Density profiles, their Fourier transforms, and relations between RDFs calculated in a slab and in bulk. For the exponential profile the analytical expression for the relation between RDFs exists but is too lengthy to be shown here. From the practical point of view, fitting with a hyperbolic profile and using the relation between the RDFs provided by the trapezoidal profile suffices. All profiles, their Fourier transforms, and scaling relations are also shown in Fig. 6. Here  $\rho_0 = \rho(z=0)$ .

Type	$\rho(z)/\rho_0$	$\hat{\rho}(\xi)/\rho_0$	$\rho_s g_s(r)/\rho_b g_b(r)$
Heaviside	$\begin{cases} 0 & -\infty \leq z < -\frac{1}{2}w \\ 1 & -\frac{1}{2}w \leq z < \frac{1}{2}w \\ 0 & \frac{1}{2}w \leq z \leq \infty \end{cases}$	$\frac{2}{\xi} \sin \frac{w\xi}{2}$	$\begin{cases} 1 - \frac{1}{2} \frac{r}{w} & r \leq w \\ \frac{1}{2} \frac{w}{r} & r > w \end{cases}$
Trapezoidal	$\begin{cases} 0 & -\infty \leq z < -\frac{1}{2}(w+\delta) \\ \frac{1}{2} \left(1 + \frac{w+2z}{\delta}\right) & -\frac{1}{2}(w+\delta) \leq z < -\frac{1}{2}(w-\delta) \\ 1 & -\frac{1}{2}(w-\delta) \leq z < \frac{1}{2}(w-\delta) \\ \frac{1}{2} \left(1 + \frac{w-2z}{\delta}\right) & \frac{1}{2}(w-\delta) \leq z < \frac{1}{2}(w+\delta) \\ 0 & \frac{1}{2}(w+\delta) \leq z \leq \infty \end{cases}$	$\frac{4}{\delta\xi^2} \sin \frac{\delta\xi}{2} \sin \frac{w\xi}{2}$	$\begin{cases} 1 + \frac{r^3}{12w\delta^2} - \frac{r^2}{3w\delta} - \frac{\delta}{3w} & r < \delta \\ 1 - \frac{r}{2w} - \frac{\delta^2}{12rw} & \delta \leq r < w \end{cases}$
Exponential <sup>29</sup>	$\frac{e^{\frac{w}{\delta}}}{2e^{\frac{w}{\delta}} - 1} \begin{cases} \exp \frac{w+2z}{\delta} & -\infty \leq z < -\frac{1}{2}w \\ 2 - \exp \frac{-w-2z}{\delta} & -\frac{1}{2}w \leq z < 0 \\ 2 - \exp \frac{-w+2z}{\delta} & 0 \leq z < \frac{1}{2}w \\ \exp \frac{w-2z}{\delta} & \frac{1}{2}w \leq z \leq \infty \end{cases}$	$\frac{4 \left(4e^{\frac{w}{\delta}} \sin \frac{\xi w}{2} + \xi \delta\right)}{\left(2e^{\frac{w}{\delta}} - 1\right) \xi (\delta^2 \xi^2 + 4)}$	
Hyperbolic	$\frac{1 - \tanh \frac{2z+w}{\delta} \tanh \frac{2z-w}{\delta}}{1 + \tanh^2 \frac{w}{\delta}}$	$\frac{\delta}{\tanh \frac{w}{\delta}} \frac{\sin \left(\frac{\xi w}{2}\right)}{\sinh \left(\frac{\xi \delta}{2}\right)}$ (see Ref. 30)	

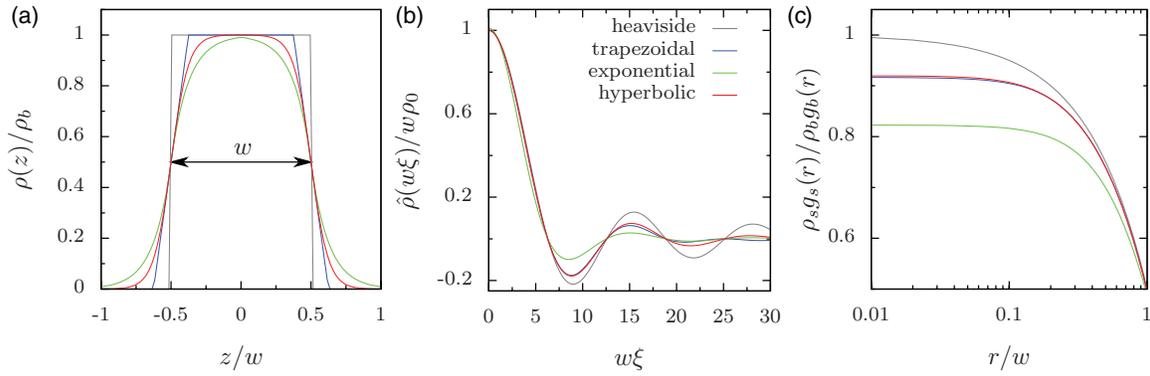


FIG. 6. (a) Density profiles, (b) their Fourier transforms, and (c) relations between the slab and bulk RDFs for all functions listed in Table I.  $\delta/w = 0.25$ .

where the hat notation implies the Fourier transform of the function

$$\hat{\rho}(\xi) = \int_{-\infty}^{\infty} da \rho(a) e^{-i\xi a}. \quad (\text{A4})$$

Back-transforming  $\hat{h}(\xi)$  and integrating over  $t$ , one obtains

$$I(r) = \frac{1}{\pi} \int_{-\infty}^{\infty} d\xi \hat{\rho}^2(\xi) \frac{\sin(r\xi)}{\xi}. \quad (\text{A5})$$

Hence, the problem has now been reduced to simply calculating the Fourier transform of the density profile and evaluating the integral (Eq. (A5)). We have calculated these for several common functional forms of the density profile, which are summarized in Table I together with their Fourier transforms and the resulting relations between the bulk and slab RDFs.

For the systems considered here, we have found that a sum of two hyperbolic tangent functions, adjusted to account for the slab drift,  $z_0$ , and a finite concentration of particles outside the slab,  $\rho_v$ , is the most convenient way to analyze (fit) the simulated density profiles

$$\rho(z) = \rho_v + (\rho_0 - \rho_v) \frac{1 - \tanh \frac{2z-2z_0+w}{\delta} \tanh \frac{2z-2z_0-w}{\delta}}{1 + \tanh^2 \frac{w}{\delta}}. \quad (\text{A6})$$

On the other hand, it also suffices to use much simpler expressions based on the trapezoidal-shaped density profile (Table I and Fig. 6)

$$g_s(r) = g_b(r) \frac{\rho_b}{\rho_s} \begin{cases} 1 + \frac{r^3}{12w\delta^2} - \frac{r^2}{3w\delta} - \frac{\delta}{3w} & r < \delta \\ 1 - \frac{r}{2w} - \frac{\delta^2}{12rw} & \delta \leq r < w \end{cases}, \quad (\text{A7})$$

where  $\delta$  is the width of the interface. Note that Eq. (A7) generalizes Eq. (7) to a slab with diffuse interfaces.

## APPENDIX B: PERIODIC BOUNDARY CONDITIONS

During a single iteration step, interfaces may become so diffuse that the concentration of particles at the box bound-

aries ( $z = -l_s/2$  and  $z = l_s/2$ ) is no longer zero. In this case, the link between  $g_b(r)$  and  $g_s(r)$  must account for periodic boundary conditions (PBC). In other words, one can no longer change the range of integration from  $[-l_s/2, l_s/2]$  to  $[-\infty, \infty]$  in Eq. (6) and Fourier *series* must be used to account for PBC. Thus, expanding the density profile

$$\rho(z) = \sum_{n=-\infty}^{\infty} a_n \exp\left(2\pi i n \frac{z}{l_s}\right), \quad (\text{B1})$$

$$a_n = \frac{1}{l_s} \int_{-l_s/2}^{l_s/2} dz \rho(z) \exp\left(-2\pi i n \frac{z}{l_s}\right), \quad (\text{B2})$$

and performing similar steps to those in Appendix A and taking into account that  $\rho(-z) = \rho(z)$ , one obtains

$$I(r) = l_s^2 \sum_{n=-\infty}^{\infty} \frac{a_n^2}{\pi n} \sin \frac{2\pi n r}{l_s}. \quad (\text{B3})$$

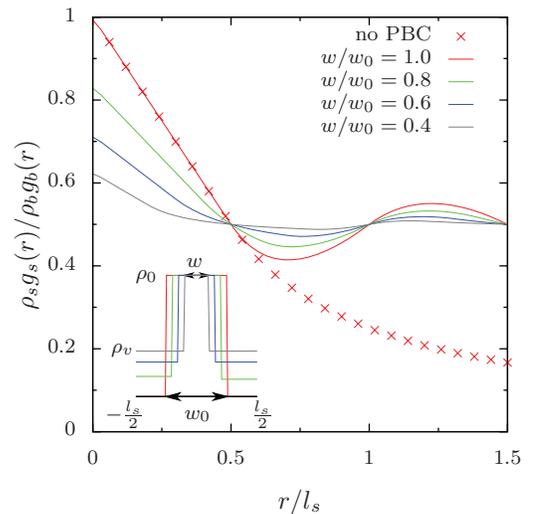


FIG. 7. Ratio between the slab and bulk RDFs for a slab with two sharp interfaces, Eq. (B4). Two cases are shown: without taking into account periodic boundary conditions (no PBC), Eq. (7), and with PBC for different concentrations of particles outside the slab,  $\rho_v$  (or different slab widths,  $w$ ). The difference for  $r > l_s/2$  is due to the slab periodic image and exists even if there are no particles outside the slab ( $w = w_0$ ).

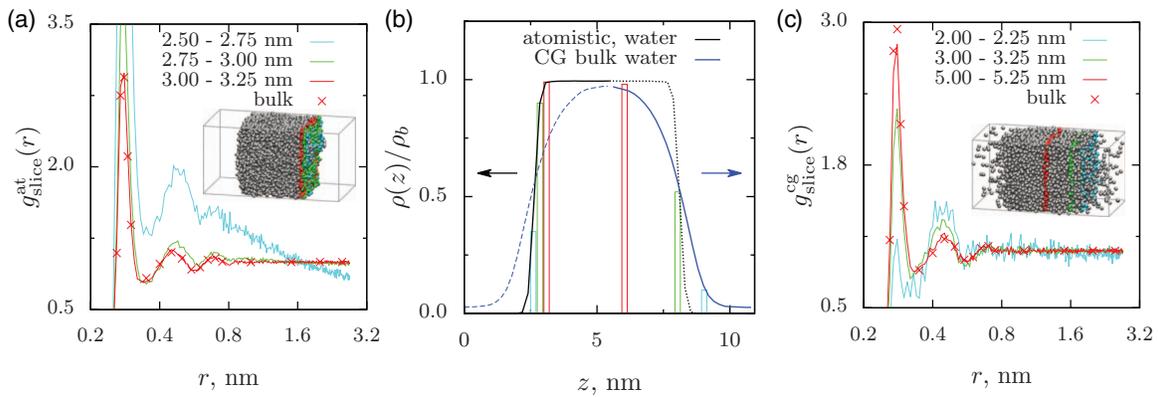


FIG. 8. Slice-resolved and appropriately scaled radial distribution functions,  $\kappa(r)g_{\text{slice}}(r)$  for (a) a slab of atomistic water (with sharp interfaces) and (c) a slab of coarse-grained, using IBI in bulk, water (wide interfaces). The corresponding density profiles are shown in (b), together with the slices, which are also depicted as insets of (a) and (c). Slice width of  $\Delta z = 0.25$  nm was used for RDF calculations.

To illustrate the effect of having a finite concentration of particles in vacuum, we will use the functional form for the slab with two sharp interfaces

$$\rho(z) = \begin{cases} \rho_v & -l_s/2 \leq z < -\frac{1}{2}w \\ \rho_0 - \rho_v \left( \frac{w_0 - w}{l_s - w} \right) & -\frac{1}{2}w \leq z < \frac{1}{2}w \\ \rho_v & \frac{1}{2}w \leq z \leq l_s/2 \end{cases}, \quad (\text{B4})$$

where  $\rho_v$  is the particle concentration in vacuum. Note that the total number of particles, that is, the integral over  $z \in [-l_s, l_s]$  is  $\rho_0 w_0$ , where  $w_0$  denotes the width of the slab when  $\rho_v = 0$  (no particles in the vacuum).

Evaluating Eq. (B2), we obtain the corresponding Fourier coefficients

$$a_n = \rho_0 \frac{l_s - w_0}{l_s - w} \frac{\sin(\pi n \frac{w}{l_s})}{\pi n}. \quad (\text{B5})$$

The relation between slab and bulk RDFs,  $\rho_s g_s(r)/\rho_b g_b(r) = I(r)/(\rho_0^2 w_0 2r)$ , is shown in Fig. 7 for several widths of the slab,  $w/w_0$ , which correspond to different densities of particles in the vacuum. One can see that even when  $w = w_0$  (all particles are part of the slab), the relation for a system with PBC starts to deviate from that with two infinite vacuum layers, starting at  $r > l_s/2$ . This is expected since at this point, the spherical shell of radius  $r$  starts to touch the periodic image of the slab. However, this is of no relevance for constructing the IBI update, as long as  $g(r)$  is calculated only until the cutoff distance,  $r_c < l_s/2$ .

If  $w \neq w_0$ , the deviation can already be seen at small  $r$  and, strictly speaking, one should calculate the update using the expansion in Eq. (B2), which is the simplest way of constructing the potential update as it only relies on the Fourier expansion coefficients of the density profile,  $\rho_n$ , does not require fitting of the simulated density profile, and avoids assumptions about the shape of the interface when constructing the update. This is especially useful in case of solute-solvent systems, where the distribution of the positions of the solute,  $p(z)$ , is a tabulated function of the  $z$  coordinate. In this case

the update can be calculated as

$$\begin{aligned} \frac{g_s(r)}{g_b(r)} &= \frac{l_s}{2r\rho_0\rho_s} \sum_{-\infty}^{\infty} \frac{\rho_n p_{-n}}{\pi n} \sin \frac{2\pi nr}{l_s} \\ &= \frac{\rho_0}{\rho_s} + \frac{l_s}{\rho_s \rho_0} \sum_{n=1}^{\infty} \frac{\rho_n p_n^* + \rho_n^* p_n}{\pi n} \sin \frac{2\pi nr}{l_s}, \end{aligned} \quad (\text{B6})$$

where  $p_n$  are the Fourier expansion coefficients of  $p(z)$ .

## APPENDIX C: SLICE-RESOLVED RADIAL DISTRIBUTION FUNCTION

The main assumption used when deriving Eq. (6) was that the local structure of the liquid does not change in the interfacial regions, i.e., only the density,  $\rho(z)$ , but not  $g_b(r)$  is a function of the  $z$  coordinate. To verify this assumption, we have calculated RDFs,  $g_{\text{slice}}(r)$ , of thin slices of width  $\Delta z$  in different regions of the slab. Since each slice is simply a slab with two sharp interfaces, we rescaled the  $g_{\text{slice}}(r)$  by a factor

$$\kappa(r) = \frac{l_s}{\Delta z} \begin{cases} 1 - \frac{1}{2} \frac{r}{\Delta z} & r \leq \Delta z \\ \frac{1}{2} \frac{\Delta z}{r} & r > \Delta z \end{cases}, \quad (\text{C1})$$

which stems from the relation between the slab and bulk RDFs, calculated for a slab with two sharp interfaces (Table I). These RDFs are shown in Fig. 8 for a set of selected slices. One can see that, in both the bulk and the slab cases, the deviation from the bulk RDF is only noticeable for the outermost interfacial layer (ca. one or two molecules thick). After this, it perfectly reproduces the bulk RDF, which justifies the assumption made for deriving Eq. (6).

## ACKNOWLEDGMENTS

This work was partially supported by the Deutsche Forschungsgemeinschaft (DFG) Emmy Noether program, DFG program IRTG 1404, DFG grant SPP 1355, and BMBF grant MESOMERIE. We are grateful to Sebastian Fritsch for critical reading of the manuscript. Siola Garbelino is thanked for kindly providing the atomistic reference.

- <sup>1</sup>D. Fritz, K. Koschke, V. A. Harmandaris, N. F. A. van der Vegt, and K. Kremer, *Phys. Chem. Chem. Phys.* **13**, 10412 (2011).
- <sup>2</sup>W. Tschöp, K. Kremer, J. Batoulis, T. Burger, and O. Hahn, *Acta Polym.* **49**, 61 (1998).
- <sup>3</sup>N. Zacharopoulos, N. Vergadou, and D. N. Theodorou, *J. Chem. Phys.* **122**, 244111 (2005).
- <sup>4</sup>J. W. Mullinax and W. G. Noid, *J. Phys. Chem. C* **114**, 5661 (2009).
- <sup>5</sup>J. W. Mullinax and W. G. Noid, *J. Chem. Phys.* **133**, 124107 (2010).
- <sup>6</sup>T. Murtola, A. Bunker, I. Vattulainen, M. Deserno, and M. Karttunen, *Phys. Chem. Chem. Phys.* **11**, 1869 (2009).
- <sup>7</sup>S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. de Vries, *J. Phys. Chem. B* **111**, 7812 (2007).
- <sup>8</sup>D. M. Huang, R. Faller, K. Do, and A. J. Moule, *J. Chem. Theory Comput.* **6**, 526 (2010).
- <sup>9</sup>A. Lukyanov, A. Malafeev, V. Ivanov, H. Chen, K. Kremer, and D. Andrienko, *J. Mater. Chem.* **20**, 10475 (2010).
- <sup>10</sup>V. Rühle, J. Kirkpatrick, K. Kremer, and D. Andrienko, *Phys. Stat. Solidi B* **245**, 844 (2008).
- <sup>11</sup>A. Villa, C. Peter, and N. F. A. van der Vegt, *Phys. Chem. Chem. Phys.* **11**, 2077 (2009).
- <sup>12</sup>W. Shinoda, R. DeVane, and M. L. Klein, *Soft Matter* **4**, 2454 (2008).
- <sup>13</sup>L. Monticelli, S. Kandasamy, X. Periole, R. Larson, P. Tieleman, and S. Marrink, *J. Chem. Theory Comput.* **4**, 819 (2008).
- <sup>14</sup>W. G. Noid, J. Chu, G. S. Ayton, V. Krishna, S. Izvekov, G. A. Voth, A. Das, and H. C. Andersen, *J. Chem. Phys.* **128**, 244114 (2008).
- <sup>15</sup>F. Ercolessi and J. B. Adams, *Europhys. Lett.* **26**, 583 (1994).
- <sup>16</sup>S. Izvekov and G. A. Voth, *J. Phys. Chem. B* **109**, 2469 (2005).
- <sup>17</sup>G. S. Ayton, W. G. Noid, and G. A. Voth, *MRS Bull.* **32**, 929 (2007).
- <sup>18</sup>D. Reith, M. Pütz, and F. Müller-Plathe, *J. Comput. Chem.* **24**, 1624 (2003).
- <sup>19</sup>A. P. Lyubartsev and A. Laaksonen, *Phys. Rev. E* **52**, 3730 (1995).
- <sup>20</sup>H. Wang, C. Junghans, and K. Kremer, *Eur. Phys. J. E* **28**, 221 (2009).
- <sup>21</sup>R. L. Henderson, *Phys. Lett. A* **49**, 197 (1974).
- <sup>22</sup>M. E. Johnson, T. Head-Gordon, and A. A. Louis, *J. Chem. Phys.* **126**, 144509 (2007).
- <sup>23</sup>L. Larini, L. Lu, and G. A. Voth, *J. Chem. Phys.* **132**, 164107 (2010).
- <sup>24</sup>X. He, W. Shinoda, R. DeVane, and M. L. Klein, *Mol. Phys.* **108**, 2007 (2010).
- <sup>25</sup>D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. Mark, and H. Berendsen, *J. Comput. Chem.* **26**, 1701 (2005).
- <sup>26</sup>V. Rühle, C. Junghans, A. Lukyanov, K. Kremer, and D. Andrienko, *J. Chem. Theory Comput.* **5**, 3211 (2009).
- <sup>27</sup>U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, *J. Chem. Phys.* **103**, 8577 (1995).
- <sup>28</sup>H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, *J. Chem. Phys.* **81**, 3684 (1984).
- <sup>29</sup>M. V. Berry, R. F. Durrans, and R. Evans, *J. Phys. A* **5**, 166 (1972).
- <sup>30</sup>H. K. Johansen and K. Sorensen, *Geophys. Prosp.* **27**, 876 (1979).