Eur. Phys. J. E **28**, 221–229 (2009)
DOI 10.1140/epje/i2008-10413-5

THE EUROPEAN PHYSICAL JOURNAL E

Regular Article

Comparative atomistic and coarse-grained study of water: What do we lose by coarse-graining?

Han Wang[1,2], Christoph Junghans[1,a], and Kurt Kremer[1]

[1] Max Planck Institute for Polymer Research, Ackermannweg 10, D-55128 Mainz, Germany
[2] LMAM and School of Mathematical Sciences, Peking University, Beijing 100871, PRC

Received 5 October 2008 and Received in final form 5 December 2008
Published online: 14 January 2009 –

Abstract. We employ the inverse Boltzmann method to coarse-grain three commonly used three-site water models (TIP3P, SPC and SPC/E) where one molecule is replaced with one coarse-grained particle with isotropic two-body interactions only. The shape of the coarse-grained potentials is dominated by the ratio of two lengths, which can be rationalized by the geometric constraints of the water clusters. It is shown that for simple two-body potentials either the radial distribution function or the geometrical packing can be optimized. In a similar way, as needed for multiscale methods, either the pressure or the compressibility can be fitted to the all atom liquid. In total, a speed-up by a factor of about 50 in computational time can be reached by this coarse-graining procedure.

PACS. 05.10.-a Computational methods in statistical physics and nonlinear dynamics – 61.20.Ja Computer simulation of liquid structure – 65.20.Jk Studies of thermodynamic properties of specific liquids

1 Introduction

Due to its common appearance and general importance water is one of the most studied liquids and despite its simplicity it is still not completely understood. Over the years, simulations have begun to play an increasing role in the attempt to provide a better understanding of water as a liquid but also as a solvent [1,2]. Many different classical simulation models have been developed which allow to capture different aspects of water. Density functional theory can give water structures in bulk to some extent and close to certain metal surfaces [3]. Molecular-dynamics (MD) simulations can give insight *i.e.* into the role of water in protein folding. On a much larger, solely particle-based scale, approaches like the lattice Boltzmann method [4] or dissipative particle dynamics (DPD) simulations [5] are employed to describe hydrodynamic effects. DPD simulations do a reasonable job in giving a qualitative insight into the behavior of clusters of particles, so-called DPD particles[1]. However, it is not possible to reproduce the structural properties of water with these rather coarse methods. This is still possible with the so-called coarse-grained (CG) MD [7], where each water molecule is represented by a coarse-grained bead, and can reproduce parts of the structure. Of course, some properties of water are lost, but a significant speed-up can be obtained due to a smaller number of degrees of freedom, simpler potentials and larger time steps. These models are also applicable for methods in which all atom simulations of water are coupled to coarse-grained simulations or directly coupled to a continuum model [8]. Other approaches for speeding up the simulation, such as coarse-graining in interaction space can be found in [9], however, in this paper we study, compare and extend different coarse-grained water models, which conserve first, the structural properties, namely the radial distribution function $g(r)$, and second, other properties such as pressure or compressibility.

2 Model and methods

2.1 Atomistic water models

The development of all-atom water models has a long history. The first ideas go back to Bernal and Fowler in 1933 [10]. The modern development of water modeling in computer simulation started in the early 1970s and several of the currently used models were developed in the 1980s. The SPC model was introduced by Berendsen and co-workers in 1981 [11] and TIP3P and TIP4P were published in 1983 [12]. The most widely studied extension of the SPC model, the SPC/E model (1987) [13], takes into

[a] e-mail: junghans@mpip-mainz.mpg.de

[1] The stochastic part of the DPD approach can also be used as a hydrodynamics conserving thermostat in standard MD. There the background friction can be varied in order to adjust the diffusion constant and the viscosity [6].

account the averaged polarization effects. A more recent development is the TIP5P (2000) [14] model that is able to reproduce the density anomaly near 277 K as well as to maintain high-quality structural and thermal properties. Over the last 30 years a huge amount of work has been devoted to develop improved water models [15] and this will stay an active area in the future.

All-atom models are designed and parameterized to fit one or more physical properties, such as the radial distribution function, density anomaly, heat of vaporization, dipole moment, etc. Alternatively, they can be based on *ab initio* calculations of water dimers or higher clusters. None of the classical all-atom models is able to simultaneously reproduce all physical properties of water, as is pointed out and analyzed in the literature [15–18]. Moreover, all present models take only two-body interactions into account, while three-body interactions, which are shown to contribute less than 14.5% to the total internal energy [19], are neglected. This also gives an impression about the typical accuracy achieved by classical all-atom models used so far. Although the water molecule is small and one of the most basic molecules in nature, it still poses difficult problems. Nevertheless MD simulations employing classical atomistic water models play an increasingly important role in many areas of computational physics and chemistry.

In the present work three simple rigid and nonpolarizable three-site models (TIP3P [12], SPC [11] and SPC/E [13]) are studied. All more complex models are extensions thereof. The methods described in this paper can be extended to any all-atom water model without substantial problems. Therefore, we choose the simplest water models.

The oxygen atoms interact through a Lennard-Jones potential which defines the overall size of the molecule, while the Coulomb interaction is assigned to charges on the hydrogen and oxygen nuclei to model hydrogen bonds. The intermolecular potential of these models can be expressed as

$$V_{\alpha\beta} = \sum_{i\in\alpha}\sum_{j\in\beta} \frac{q_i q_j e^2}{4\pi\epsilon_0 r_{ij}} + \frac{C_{12}}{r_{\mathrm{OO}}^{12}} - \frac{C_6}{r_{\mathrm{OO}}^{6}}\,, \qquad (1)$$

where α and β stand for two different molecules and r_{ij} denotes the distance between atom i on molecule α and atom j on molecule β. r_{OO} denotes the distance between oxygen atoms. The intramolecular interactions are not considered because the geometry of the molecule is kept rigid. The parameters of the interactions and geometrical constraints are listed in Table 1.

Various water models with internal degrees of freedom have been studied as well. These internal degrees are bond length and angle vibrations [20–23] or even polarization effects [24–28]. Such models are not considered here.

Table 1. Parameters of all-atom water models.

	TIP3P	SPC	SPC/E
r_{OH} (Å)	0.9572	1.0	1.0
$\angle$HOH (°)	104.52	109.47	109.47
$10^3 C_6$ (kJ/mol nm^6)	2.4889	2.6171	2.6171
$10^6 C_{12}$ (kJ/mol nm^{12})	2.4352	2.6331	2.6331
q_{O} (e)	−0.834	−0.820	−0.8476
q_{H} (e)	+0.417	+0.410	+0.4238

2.2 Coarse-grained water models

Generally speaking, coarse-graining methods aim at finding a way to map an all-atom model onto a less structured and simpler model, which is computationally much more efficient. A coarse-grained system should preserve as many properties of the underlying all-atom system as possible or as required by the physical question under consideration. In this work, we map the entire water molecule onto one coarse-grained bead at the center of the oxygen. Recently, this procedure was adopted to study TIP3P water in an adaptive resolution simulation (AdResS method) [29]. In that study the parameterization of the coarse-grained model was based on TIP3P water with a bond angle of 112.19° instead of 104.52°. This led to a shift of the relative depths of the minima in the coarse-grained potential [29]. This difference in coarse-grained models motivates, to some extend the present study, in which different models are compared in a systematic way.

There are different options to parameterize the effective coarse-grained particle-particle interactions in order to resemble the overall water structure most closely. Typical examples are reverse Monte Carlo simulations [30] or iterative Boltzmann inversion [31]. As discussed above coarse-grained models cannot reproduce every property of the corresponding all-atom model [32], therefore one has to adjust the coarse-grained models based on which question it is supposed to deal with. Here we are mostly concerned with the structural properties of liquid water at a given temperature and pressure. The construction of an effective interaction, consists of two steps:

1) Iterative Boltzmann inversion. This scheme constructs an effective potential that reproduces the center of mass or O-O radial distribution function of the all-atom model. So if a property of interest is mostly relying on the RDF, then it is naturally reproduced by the resulting coarse-grained model. For the three water models considered in this paper, the differences in the peaks and minima between the center of mass and O-O radial distribution functions are no more than 2%.
2) Adjust the obtained potential produced by the first step to reproduce further properties we are interested in. Of course the correction may sacrifice the accuracy of other properties that are not so important to the problem. This step can be iterated with every step of iterative Boltzmann inversion or after the whole inversion procedure.

2.2.1 Iterative Boltzmann inversion

In this paper we derive the effective interactions between the coarse-grained water particles by the iterative

Boltzmann inversion as introduced by Reith *et al.* [31]. The theoretical basis of this method is a system in which particles interact through a series of potentials $V^{(1)}(\boldsymbol{r})$, $V^{(2)}(\boldsymbol{r}_1, \boldsymbol{r}_2)$, $\cdots$, $V^{(n)}(\boldsymbol{r}_1, \cdots, \boldsymbol{r}_n)$, where $V^{(i)}$ denotes the i-body interaction. All thermodynamic properties can be determined by multi-body distributions $\rho^{(1)}(\boldsymbol{r})$, $\rho^{(2)}(\boldsymbol{r}_1, \boldsymbol{r}_2)$, $\cdots$, $\rho^{(n)}(\boldsymbol{r}_1, \cdots, \boldsymbol{r}_n)$ [33]. In practice, the calculation of i-body ($i \geq 3$) interactions is extremely time consuming, therefore only two-body interactions are taken into account. Thus the effective potential is optimized to fit the pair distribution $\rho_2(\boldsymbol{r}_1, \boldsymbol{r}_2)$. Furthermore, since the beads are spatially isotropic, only the radial distribution function (RDF) $g(r)$ is available.

After starting with an initial guess of the pair interaction $V_0(r)$, the interaction of the $(i+1)$-th step is given by

$$V_{i+1}(r) = V_i(r) + k_B T \ln\left[\frac{g_i(r)}{g^{\text{target}}(r)}\right], \tag{2}$$

where $g^{\text{target}}(r)$ is the target RDF we want to fit, and $g_i(r)$ the RDF of the i-th recursion step. A good initial guess for the iteration is the potential of mean force

$$V_0(r) = -k_B T \ln g^{\text{target}}(r), \tag{3}$$

which is exact for a system interacting through two-body interactions at the density limit of $\rho \to 0$.

2.2.2 Pressure correction

The pressure of the coarse-grained system, parameterized by the method described above, is not the same as the original system. By substituting a water molecule with a bead, the internal structure of the water molecules is lost, *i.e.* the information carried by the RDF of H-O and H-H, and thus the related interactions contributing to the pressure. A perfect match of the distribution function would perfectly reproduce the compressibility, which determines density fluctuations, would be properly reproduced (as long as two-body interactions are dominant). So in general one cannot expect the coarse-grained models to describe the thermodynamic properties correctly without any further adjustment.

In the original paper of the presented iterative Boltzmann inversion [31], a linear correction method is introduced to get the right pressure at the cost of losing some of the accuracy in the RDF (and thus of the compressibility). The correction to the potential is

$$\Delta V(r) = A\left(1 - \frac{r}{r_{cut}}\right). \tag{4}$$

In [31], the constant A is set to $-0.1 k_B T$. We estimate A in a different way. The virial expression of the pressure P_i is given by

$$P_i V = N k_B T - \frac{2}{3}\pi N \rho \int_0^{+\infty} r^3 \frac{\mathrm{d}V_i(r)}{\mathrm{d}r} g_i(r)\,\mathrm{d}r, \tag{5}$$

where $V_i(r)$ and $g_i(r)$ are the potential and RDF of the i-th step the iteration, respectively. The corrected potential is ought to match the correct pressure P^{target}, that is

$$P^{\text{target}} V \approx N k_B T - \frac{2}{3}\pi N \rho \int_0^{+\infty} r^3 \frac{\mathrm{d}}{\mathrm{d}r}\left[V_i(r) + \Delta V_i(r)\right] g_i(r)\,\mathrm{d}r. \tag{6}$$

The approximation appears due to the fact that $g_i(r)$ is the RDF of the uncorrected system. Therefore, the unknown A_i satisfies

$$-\left[\frac{2\pi N\rho}{3\, r_{cut}} \int_0^{r_{cut}} r^3 g_i(r)\,\mathrm{d}r\right] A_i \approx (P - P_{\text{target}})V. \tag{7}$$

Usually one cannot reach the target pressure by one step of the pressure-correction. Therefore, we first apply the pressure-correction at each step of the iterative Boltzmann inversion. When the RDF is approximated well enough, we apply an iteration of pressure-correction to obtain the correct pressure. The additional iteration of pressure-correction will not change the RDF a lot (less than 0.5%). It should however be mentioned at this point, that the pressure-correction leads to a potential, where the isothermal compressibility deviates significantly from that of the all atom model.

2.3 Liquid structure —tetrahedral packing

A special property of liquid water is the tetrahedral packing due to hydrogen bonds. For the all-atom models, this local structure is the result of the nearest-neighbor Coulomb interaction of the partial charges on the oxygens and hydrogens. In contrast to the all-atom models, the effective potential is spatially isotropic. For that reason the structure is a result of two length scales in the effective potential. In this paper, we will refer to the molecule in the center of four nearest neighbors as the reference molecule and analyze whether such a cluster resembles the tetrahedral packing of the underlying all-atom model. The tetrahedral packing is measured by a parameter q_4 which is computed by a sum over the deviation between each "bond angle" (there are no actual bonds between the coarse-grained beads) and the perfect tetrahedral angle $\arccos(-\frac{1}{3}) \approx 109.471°$:

$$q_4 = 1 - C_4 \sqrt{\sum_{i<j}\left[\theta_{ij} - \arccos\left(-\frac{1}{3}\right)\right]^2}, \tag{8}$$

where θ_{ij} is the bond angle between particle i, the reference particle and particle j, where i and j go over the first four nearest neighbors of the reference particle. The bond angle are measured between the oxygen atoms in the all-atom description and between the beads in the coarse-grained model. $C_4 = 1.8165$ is a constant chosen such that $q = 0$ when "bond angles" are randomly distributed and $q = 1$ when the particles exhibit perfect tetrahedral packing. As the tetrahedral packing becomes weaker, the value

Table 2. Structural properties of different water models: All-atom (AA), coarse-grained (CG), pressure-corrected coarse-grained (PC CG) and potential of interpolation (POE) results are shown. The properties presented in the table are from left to right: ϕ (the ratio between peak position and first-well position), γ (the ratio between second-well position and first-well position), tetrahedral packing parameter and the probability that q_4 is a large value (not less than 0.7). The statistical error is given in parentheses.

	ϕ (1.155)	γ (1.633)	$\bar{q}_4$	$P(q_4 \geq 0.7)$
TIP3P AA			0.3512 (0.0009)	0.0251
TIP3P CG	1.120	1.617	0.2530 (0.0003)	0.0049
TIP3P PC CG	1.127	1.531	0.2525 (0.0003)	0.0048
SPC AA			0.3823 (0.0009)	0.0359
SPC CG	1.120	1.636	0.2592 (0.0003)	0.0059
SPC PC CG	1.128	1.575	0.2587 (0.0003)	0.0060
SPC/E AA			0.4109 (0.0009)	0.0479
SPC/E CG	1.121	1.643	0.2670 (0.0003)	0.0078
SPC/E PC CG	1.125	1.587	0.2662 (0.0003)	0.0071
POE	1.131	1.632	0.3142 (0.0004)	0.0251
POE PC	1.131	1.611	0.3238 (0.0004)	0.0297

of q becomes smaller. This definition is very similar to the parameter introduced by Errington and Debenedetti [34]. The difference is that our parameter measures the deviation of the "bond angles" from the perfect tetrahedral angle rather than of the cosine values.

Tetrahedral clusters can be characterized by two typical distances. The first is the distance between the reference molecule and the nearest neighbors, and the second is the distance between pairs of nearest neighbors. The ratio between the second typical distance and the first distance is $2\sqrt{2} : \sqrt{3}$. Not surprisingly, the ratio between the positions of the minima of the two wells of the coarse-grained effective potential is roughly identical. We denote this ratio by γ and list its value for different coarse-grained models in Table 2. For a coarse-grained model that is required to reproduce the tetrahedral packing in liquid state, γ should not deviate from the perfect ratio $(2\sqrt{2} : \sqrt{3})$ too much. We denote the ratio between the position of the peak between two wells and the first potential minimum by ϕ. The value of this ratio is roughly $2 : \sqrt{3}$, which is the position preventing other molecules from entering the tetrahedral cluster.

3 Results

3.1 Effective coarse-grained potentials

NVT simulations for the all-atom water models TIP3P, SPC and SPC/E were performed by using GROMACS 3.3 [38–40]. $24^3 = 13834$ molecules are studied inside a cubic box with periodic boundary conditions. The temperature is adjusted to 300 K by a Berendsen thermostat [41]. The densities (see Tab. 3) of the systems are the average densities obtained from NPT simulations of the same systems, where the pressures were kept at one bar by the Berendsen barostat [41]. The long-range electrostatic interaction are calculated by the particle mesh Ewald (PME) method [42]. Parts of the coarse-grained and pressure-corrected coarse-grained simulations are carried using the Espresso package [43]. Our coarse-grained simulations contain 10^4 particles and the size of the periodic simulation boxes is adjusted and fixed so that the densities are the same as for the corresponding all-atom simulations. The cut-off was set 0.7 nm, which contains most of the structure (see Fig. 1). Temperature is kept constant at 300 K by a Langevin thermostat [44] with a friction constant of $5\,\mathrm{ps}^{-1}$. The systems are integrated at a time step of 0.004 ps. Thus the results are targeted towards static properties. The diffusion constants of the coarse-grained models are measured by simulations employing GROMACS 3.3 with the same system settings as for all-atom simulations.

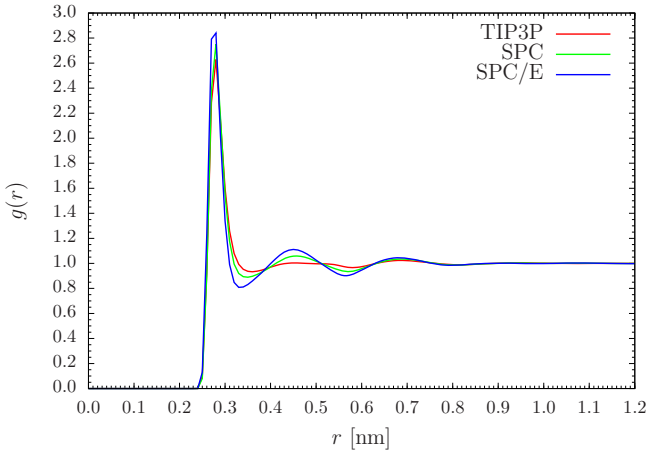

Fig. 1. The radial distribution function of the all-atom TIP3P, all-atom SPC and all-atom SPC/E model.

The RDFs of all employed the all-atom models are shown in Figure 1. One can see that the RDF of the SPC/E model has the strongest peaks and wells while the RDF of the TIP3P model has the weakest structure. A

Table 3. Thermodynamic properties of different water models: All-atom (AA), coarse-grained (CG), pressure-corrected coarse-grained (PC CG) and potential of extrapolation (POE) results are shown. The properties presented in the table are from left to right: density, virial pressure, isothermal compressibility and diffusion constant. In parentheses are the statistical errors. The last line shows the experimental data.

	ρ (g/cm^3)	P (bar)	κ_T (10^{-10} m^2/N)	D (10^{-9} m^2/s)
TIP3P AA	0.9846	1.00 (0.65)	5.76 (0.02)	5.9319 (0.0737)
TIP3P CG		8536 (0.67)	4.79 (0.02)	19.3899 (0.0556)
TIP3P PC CG		0.70 (0.68)	27.12 (0.51)	19.4417 (0.2585)
SPC AA	0.9769	0.82 (0.67)	5.28 (0.02)	4.4374 (0.0643)
SPC CG		8994 (0.72)	4.66 (0.02)	17.9753 (0.0355)
SPC PC CG		1.67 (0.72)	28.62 (0.61)	17.8028 (0.2597)
SPC/E AA	0.9984	0.76 (0.73)	4.56 (0.02)	2.7866 (0.0310)
SPC/E CG		9886 (0.81)	4.38 (0.02)	15.7114 (0.0351)
SPC/E PC CG		0.45 (0.82)	29.71 (0.74)	15.6021 (0.2622)
POE	0.9984	8982 (1.10)	4.78 (0.03)	5.8013 (0.1331)
POE PC		3.61 (1.12)	28.15 (0.92)	4.7760 (0.0053)
exp. (298 K)	0.99705 [35]	-	4.599 [36]	2.272 [37]

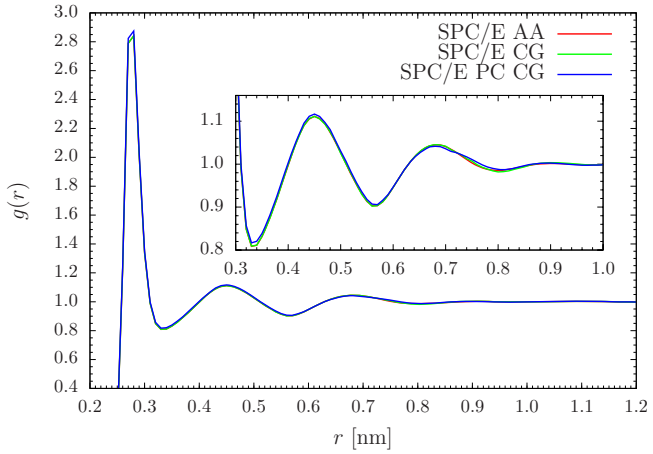

Fig. 2. Comparison between the RDF of all-atom SPC/E model (SPC/E AA), coarse-grained SPC/E model (SPC/E CG) and pressure-corrected coarse-grained SPC/E model (SPC/E PC CG). The all-atom and the coarse-grained RDFs coincide with each other quite well, while the pressure-corrected RDF deviates slightly at the minima and maxima as the insertion shows.

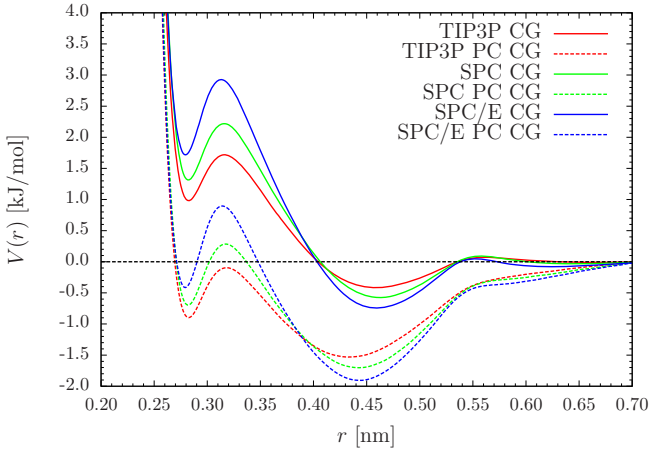

Fig. 3. Effective coarse-grained (CG) potentials generated by iterative Boltzmann inversion. Solid lines are those without pressure-correction (PC) and the dashed lines are those with pressure-correction.

similar conclusion is reached for q_4 by looking at the plot of the distribution of tetrahedral parameter q_4, see the dashed lines in Figure 4.

The derived coarse-grained models match the all-atom RDFs (we only plot the RDFs of SPC/E model, see Fig. 2) extremely well. The corresponding effective potentials are plotted as solid lines in Figure 3, from which we see that larger magnitude of the peaks and wells in RDFs results in stronger peaks and wells in effective potentials. However, all coarse-grained models produce a significantly weaker tetrahedral order than the corresponding all-atom models as shown in Figure 4. SPC/E displays the most pronounced tetrahedral packing and TIP3P presents the least, again in agreement with the shape of the coarse-grained potential. The reason is that the hydrogen bonds originating from the Coulomb interactions of the partial charges of the molecule are replaced by isotropic potential wells with a single origin. Our data shows that the tetrahedral packing for electric interactions is stronger than for effective potentials.

We observe only a small discrepancy between the all-atom RDFs and the pressure-corrected coarse-grained RDFs (see Fig. 2) even though the effective potentials look quite different from those of the coarse-grained models without pressure-correction, see Figure 3. Henderson [45] has shown that an effective potential designed to reproduce a given RDF is unique up to a constant. While this is a rigorous result, the coarse-graining procedure shows that rather small variations in the RDFs

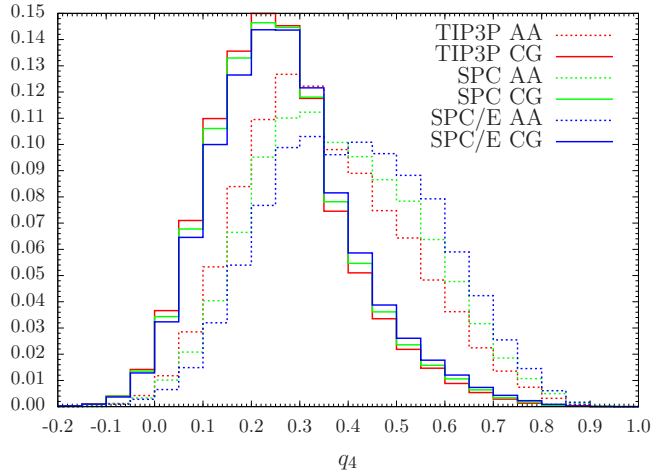

Fig. 4. The distribution function of the tetrahedral packing parameter q_4 of all-atom (AA) water models and coarse-grained (CG) water models. The results of pressure-corrected potentials are not plotted because they are almost indistinguishable from those without pressure-correction.

Table 4. Life time analysis of structural properties of different water models: all-atom (AA), coarse-grained (CG), pressure-corrected coarse-grained (PC CG) and potential of extrapolation (POE). Mean values are measured and standard deviations (not the statistical error) are given in parenthesis. t_T is the lifetime of a tetrahedral cluster, while t_H is the lifetime of a hydrogen bond.

Model	t_T (ps)	t_H (ps)
TIP3P	0.057 (0.059)	0.868 (0.928)
SPC	0.067 (0.072)	0.994 (1.120)
SPC/E	0.070 (0.073)	1.268 (1.544)
TIP3P CG	0.044 (0.042)	0.546 (0.456)
SPC CG	0.043 (0.040)	0.615 (0.527)
SPC/E CG	0.044 (0.045)	0.735 (0.654)
TIP3P PC CG	0.045 (0.046)	0.550 (0.459)
SPC PC CG	0.044 (0.043)	0.610 (0.530)
SPC/E PC CG	0.047 (0.043)	0.741 (0.664)
POE	0.061 (0.069)	1.306 (1.291)
POE PC	0.061 (0.072)	1.362 (1.380)

can result in significant changes of the potential. Despite the different effective potentials and a discrepancy in γ between coarse-grained and pressure-corrected coarse-grained models (Tab. 2), the distributions of q_4 is nearly the same. Thus we will not plot them in Figure 4 and refer the reader to the mean values and the probability of large q_4 in Table 2.

As shown in Table 3, the virial pressure of the coarse-grained models does not agree with the underlying all-atom model. This significant deviation reflects the fact that the coarse-grained models approximate a system with significantly more degrees of freedom, more spatially varying and effectively anisotropic interactions. This is well known for coarse-grained simulations [32,46]. All pressure-corrected coarse-grained models reproduce the pressures of the all atom models within the error bars, supporting our pressure-correction strategy. In order to deal with a mixed all-atom and coarse-grained simulation (AdResS scheme) [47,29] it is more appropriate to adjust the compressibility rather than the pressure. One can determine the isothermal compressibility of the system by the following finite-difference method:

$$\kappa_T = -\frac{1}{V}\left(\frac{\partial V}{\partial p}\right)_T \approx -\frac{1}{V}\frac{V^+ - V^-}{p^+ - p^-}\,. \tag{9}$$

Here, we want to note that the resulting effective potential depends on the temperature T and density ρ [32] for which the coarse-graining procedure is carried out [48]. However, since the compressibility is quite small and the system reasonably big, it is sufficient for this estimate not to account for any dependence of the coarse-grained potential on the states $\{N, V^+, T\}$ and $\{N, V^-, T\}$. As shown in Table 3, the side effect of the pressure-correction is a strong deviation in the isothermal compressibilities. The agreement between the isothermal compressibilities of the all-atom system and the coarse-grained system without pressure-correction is much better. The isothermal compressibility can also be determined by the following formula [49]:

$$\rho k_B T \kappa_T = 1 + 4\pi\rho \int r^2 \left[g(r) - 1\right] \mathrm{d}r. \tag{10}$$

The large deviations in the compressibility between the coarse-grained and the pressure-corrected coarse-grained results show that the compressibility is a very sensitive measure of the overall agreement between different RDFs.

Table 2 also shows that the intrinsic timescale of the coarse-grained water models is approximately 4 times faster than that of the all-atom simulations. This is due to the softer interaction potentials and the resulting reduced friction between molecules, as has already been observed in a different context [46].

The lifetime (see Tab. 4) of tetrahedral clusters (defined as the length of the time period when a cluster has a q_4 larger than 0.7) is about 0.04–0.07 ps and the standard deviation is roughly the same as the mean value. The lifetime of individual hydrogen bonds is significantly longer, roughly 0.55–1.36 ps. Of course the definition of a "hydrogen bond" is somewhat arbitrary within these models. Taking the RDFs, we count a pair of molecules hydrogen bonded when the distance of two oxygens is less than 0.31 nm. This definition is not a problem for coarse-grained models, but obviously not truly adequate for all-atom models, however well suited for a comparison of the different models.

3.2 Optimizing the tetrahedral packing

3.2.1 The shape of effective potential and tetrahedral packing

Figure 1 shows that the neighbor distribution of SPC/E displays a weaker structure than that of the SPC model, while the TIP3P model has a nearly flat distribution of

neighbors beyond the first peak of the RDF. Along with this the strength of the tetrahedral packing also increases from TIP3P to SPC and even more to SPC/E. (see Fig. 4). The coarse-grained potentials (see Fig. 3) suggests a more detailed check on how the tetrahedral packing and the shape of the potentials are correlated.

From the analysis in Section 2.3, one knows how to obtain a modified effective potential that can produce better tetrahedral packing, namely to keep the position of the wells and the peaks, but varying their height/depth. We find that the distribution of q_4 is insensitive to the change of the depth of the first well as long as the barrier is lower than some critical value (roughly $kT \approx 2.5\,\mathrm{kJ/mol}$). When we increase the height of the peak or the depth of the second well, the tetrahedral packing also becomes more pronounced. This effect may be understood by the fact that the difference between the peak and the second well acts as a barrier which prevents more particles from entering the first-neighbor shell. This result suggests that in order to reach better tetrahedral packing, we have to increase the depth of the second well and the height of the peak of the potential while keeping them at the right position.

3.2.2 Extrapolation of effective coarse-grained potentials

To check the ideas raised in the previous section, we manipulate the effective coarse-grained potentials by linearly interpolating two coarse-grained potentials under the constraint of keeping the region where they are nearly equal unchanged. Since it does not matter which potential is used as base potential in the extrapolation, we have chosen a combination of the coarse-grained SPC and coarse-grained SPC/E potential. The extrapolation can be described by

$$V_{\mathrm{POE}}(r) = V_{\mathrm{SPC}}(r) + \lambda(r)(V_{\mathrm{SPC/E}}(r) - V_{\mathrm{SPC}}(r)), \quad (11)$$

where $\lambda(r)$ is a first-order continuously differentiable function, which was generated by cubic splines. The requirement of smoothness of $\lambda(r)$ is the necessary and sufficient condition for a continuous effective force resulting from the potential $V_{\mathrm{POE}}(r)$. A value $0 < \lambda < 1$ allows us to locally change the potential between the two original coarse-grained potentials considered. A value $\lambda > 1$ corresponds to adding even more "SPC/E structure". Outside this region we use the potential as for SPC/E ($\lambda = 1$). Negative values of λ point more towards the coarse-grained version of the TIP3P model.

3.2.3 Simulation results

The weight function $\lambda(r)$ used in this study is plotted in Figure 5, where the values of λ at the position of the first well, the peak and the second well are 3, 3 and 8, respectively.

The simulation results of the extrapolation potentials are shown in Figures 5, 6 and 7 and Tables 2 and 3. From Figure 7 we deduce that the extrapolation potential can

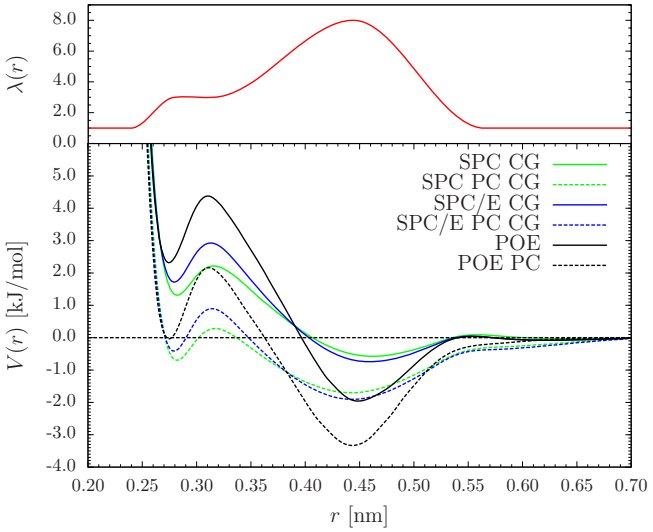

Fig. 5. Weight function $\lambda(r)$, the potential of extrapolation (POE) and its pressure-corrected version (POE PC). These potentials are plotted in comparison with coarse-grained (CG) and pressure-corrected coarse-grained potentials (PC CG) of SPC and SPC/E.

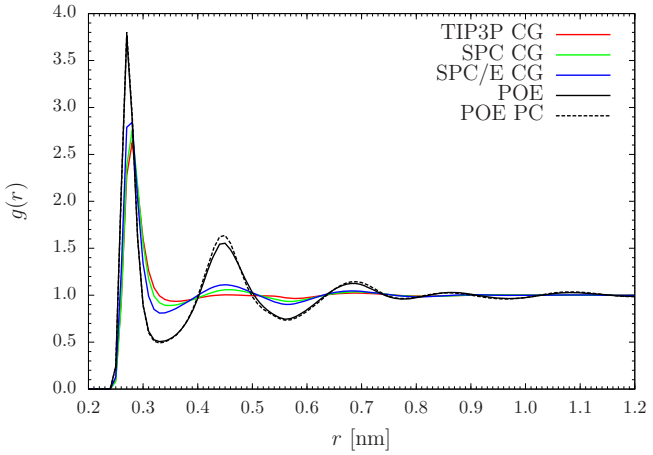

Fig. 6. The RDF of coarse-grained (CG) models and the extrapolated potential (POE).

only fit the all-atom TIP3P result when $q_4 \geq 0.7$. There is still an obvious discrepancy for smaller q_4 values. We could improve the fit of the distribution of smaller q_4 by increasing the weight function λ, but then the agreement at large q_4 values would be destroyed. In practice, the region of large q_4 (where clusters are well tetrahedrally packed) is more important than the region of small q_4. Thus we focus on changes of the coarse-grained potentials, which better fit q_4 in the $[0.7, 1]$ interval. For this new potential, Figure 6 shows that the RDFs deviate strongly. With deeper wells and a higher peak we fit the higher-order structural aspects of the liquid better, but introduce more structure in the RDFs. It seems impossible to fit both the RDF and the q_4 distribution at the same time with isotropic central potentials.

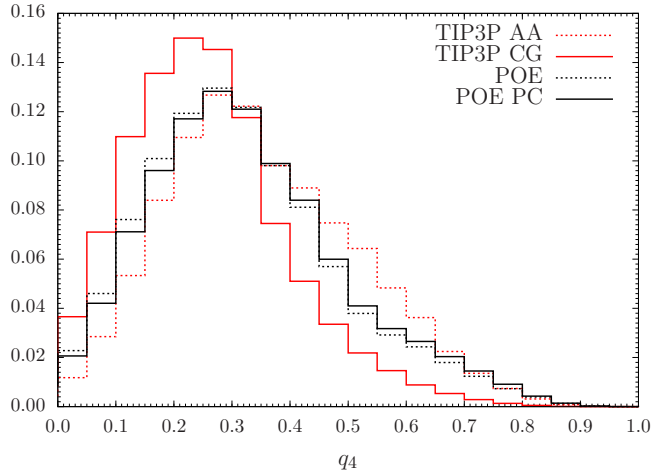

Fig. 7. The distribution of tetrahedral packing parameter q_4 of the extrapolated potential (POE) and its pressure-corrected version (POE PC) comparing with all-atom TIP3P model (TIP3P AA) and coarse-grained TIP3P model (TIP3P CG).

Even considering even deeper potential wells and higher peaks, we cannot reach the q_4 distribution of the all-atom SPC and SPC/E models. Since the strength of the potential wells is then comparable to $kT \approx 2.5\,\mathrm{kJ/mol}$, the coarse-grained water is no longer a homogeneous liquid.

Surprisingly, the isothermal compressibility of the extrapolation potential is preserved very well though the fit of the RDF becomes poor. For the coarse-grained SPC and SPC/E model the virial pressure is too large, which has been shown above for all non-pressure-corrected coarse-grained potentials. To obtain a coarse-grained model that can reproduce the all-atom pressure, one can also extrapolate between pressure-corrected coarse-grained SPC and SPC/E models instead of between the original coarse-grained potentials. But then the resulting pressure is too low due to the stronger structure in the potential. To solve this problem, we apply a tail correction to the resulting potential and vary the size of the support of the correction to arrive at the correct pressure. In Table 3 we show that the pressure is correct now, but the compressibility is wrong, which is the same as for all pressure-corrected models. From the dashed lines in Figure 6 and 7, we find no obvious difference in the structural properties between the pressure-corrected extrapolated potential and the non–pressure-corrected potential.

4 Summary

In this paper it was shown how simple rigid water models can be coarse-grained in a straight forward way by replacing one molecule by a coarse-grained bead. An efficient coarse-grained interaction between the oxygen atoms was derived by the inverse Boltzmann method, which naturally gives a coarse-grained model a very similar structure and nearly the same compressibility as the all-atom model. Under the condition of losing the agreement of compressibility, we were able to adjust the pressure to the same value as in the all-atom case by a simple linear correction. It seems to be, however, impossible to simultaneously adjust the pressure and the compressibility by simple isotropic two-body potentials.

The lifetime analysis of the tetrahedral clusters shows that for all the models a cluster lives 50 fs on average, which makes our approach of coarse-graining much more physical than the idea of replacing five water molecules (one tetrahedral cluster) by a coarse-grained bead. The lifetime gives a rough estimate of the maximal physical timescale of these kinds of configurations.

We also showed how to improve the tetrahedral packing of coarse-grained models by introducing a higher barrier in the coarse-grained potentials. On the other hand, this also leads to more structure in the RDF. This is why we think good tetrahedral packing and conserved structure is not possible at the same time with such a simple spherical interaction. In general for all these coarse-grained potentials there exist two characteristic length ratios. First, the ratio of the first well to the second well, which is also the ratio between the nearest- and the second-nearest–neighbor distances in the tetrahedral cluster. Second, the ratio of the first well to the first peak, which prevents molecules from entering the first neighbor shell.

Altogether such a coarse-graining procedure leads to a factor of 10 (9 Coulomb + 1 Lennard Jones *vs.* 1 tabulated interaction) in computation speed plus a gain due to the lack of electrostatic interaction and the four times larger intrinsic timescale. This accumulates to a speed-up of the order of 50 in computer time. While for small systems this is not decisive, for huge simulations of *i.e.* biomolecular systems with surrounding water this will be crucial in many cases.

Of course models with just isotropic interactions cannot capture all properties compared to more complicated ones. But this is the same for all coarse-grained or simplified models and in order to overcome such shortcomings links to other approaches have to be employed. One possibility is given by the AdResS method [29], where all atom simulations and coarse-grained simulations are coupled in a way that there is a free exchange of molecules and their representation. For the effect of long range electrostatic interactions the model can be linked to, *i.e.* a fast Poisson-Boltzmann solver [50]. However all these extensions require the detailed understanding of the possibilities and limitations of both the underlying atomistic and coarse-grained models.

We thank M. Praprotnik for providing comparison data for the TIP3P water model. In addition we thank A. Falk, T. Vehoff and K. Johnston for critical reading of the manuscript. H.W. thanks the China Scholarship Council for financial support.

References

1. S. Koneshan, J.C. Rasaiah, R.M. Lynden-Bell, S.H. Lee, J. Phys. Chem. B **102**, 4193 (1998).
2. M.R. Shirts, V.S. Pande, J. Chem. Phys. **122**, 134508 (2005).
3. D. Sebastiani, L. Delle Site, J. Chem. Theor. & Computat. **1**, 78 (2005).
4. B. Duenweg, A.J.C. Ladd, Adv. Polym. Sci. **291**, 89 (2009).
5. P.J. Hoogerbrugge, J.M.V.A. Koelman, Europhys. Lett. **19**, 155 (1992).
6. C. Junghans, M. Praprotnik, K. Kremer, Soft Matter **4**, 156 (2008).
7. A.P. Lyubartsev, M. Karttunen, I. Vattulainen, A. Laaksonen, Soft Materials **1**, 121 (2002).
8. R. Delgado-Buscalioni, K. Kremer, M. Praprotnik, J. Chem. Phys. **128**, 114110 (2008).
9. S. Izvekov, J.M.J. Swanson, G.A. Voth, J. Phys. Chem. B **112**, 4711 (2008).
10. J.D. Bernal, R.H. Fowler, J. Chem. Phys. **1**, 515 (1933).
11. H.J.C. Berendsen, J.P. Postma, W.F. van Gunsteren, J. Hermans, in *Intermolecular Forces*, edited by B. Pullman (D. Reidel Publishing Company, Dordrecht, 1981), pp. 331–342.
12. W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, J. Chem. Phys. **79**, 926 (1983).
13. H.J.C. Berendsen, J.R. Grigera, T.P. Straatsma, J. Phys. Chem. **91**, 6269 (1987).
14. M.W. Mahoney, W.L. Jorgensen, J. Chem. Phys. **112**, 8910 (2000).
15. B. Guillot, J. Mol. Liquids **101**, 219 (2002).
16. A. Baranyai, A. Bartók, A.A. Chialvo, J. Chem. Phys. **124**, 074507 (2006).
17. A. Brodsky, Chem. Phys. Lett. **261**, 563 (1996).
18. J.L. Finney, J. Mol. Liquids **90**, 303 (2001).
19. V.E. Petrenko, M.L. Dubova, Y.M. Kessler, M.Y. Perova, Russ. J. Phys. Chem. **74**, 1777 (2001).
20. S. Amira, D. Spångberg, K. Hermansson, J. Chem. Phys. **303**, 327 (2004).
21. L.X. Dang, B.M. Pettitt, J. Phys. Chem. **91**, 3349 (1987).
22. K. Toukan, A. Rahman, Phys. Rev. B **31**, 2643 (1985).
23. S.B. Zhu, C.F. Wong, J. Chem. Phys. **98**, 8892 (1993).
24. L.X. Dang, J. Chem. Phys. **97**, 2659 (1992).
25. G. Lamoureux, A.D. MacKerell jr., B. Roux, J. Chem. Phys. **119**, 5185 (2003).
26. S.W. Rick, S.J. Stuart, B.J. Berne, J. Chem. Phys. **101**, 6141 (1994).
27. H.A. Stern, F. Rittner, B.J. Berne, R.A. Friesner, J. Chem. Phys. **115**, 2237 (2001).
28. H. Yu, T. Hansson, W.F. van Gunsteren, J. Chem. Phys. **118**, 221 (2002).
29. M. Praprotnik, S. Matysiak, L. Delle Site, K. Kremer, C. Clementi, J. Phys.: Condens. Matter **19**, 292201 (2007).
30. A.P. Lyubartsev, A. Laaksonen, Phys. Rev. E **52**, 3730 (1995).
31. D. Reith, M. Puetz, F. Mueller-Plathe, J. Comput. Chem. **24**, 1624 (2003).
32. M.E. Johnson, T. Head-Gordon, A.A. Louis, J. Chem. Phys. **126**, 144509 (2007).
33. J. Zwicker, R. Lovett, J. Chem. Phys. **93**, 6752 (1990).
34. J.R. Errington, P.G. Debenedetti, Nature **409**, 318 (2001).
35. F. Franks, *Water: A Matrix of Life* (Royal Society of Chemistry, 2000).
36. M.N. Rodnikova, J. Molec. Liquids **136**, 211 (2007).
37. D. Eisenberg, W. Kauzmann, *The Structure and Properties of Water* (Clarendon Press Oxford, 1969).
38. H.J.C. Berendsen, D. van der Spoel, R. van Drunen, Comput. Phys. Commun. **91**, 43 (1995).
39. E. Lindahl, B. Hess, D. van der Spoel, J. Molec. Model. **7**, 306 (2001).
40. D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, H.J.C. Berendsen, J. Comput. Chem. **26**, 1701 (2005).
41. H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. DiNola, J.R. Haak, J. Chem. Phys. **81**, 3684 (1984).
42. T. Darden, D. York, L. Pedersen, J. Chem. Phys. **98**, 10089 (1993).
43. H.-J. Limbach, A. Arnold, B.A. Mann, C. Holm, Comput. Phys. Commun. **174**, 704 (2006).
44. M.P. Allen, D.J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, USA, 1989).
45. R.L. Henderson, Phys. Lett. A **49**, 197 (1974).
46. V.A. Harmandaris, N.P. Adhikari, N.F.A. van der Vegt, K. Kremer, Macromolecules **39**, 6708 (2006).
47. M. Praprotnik, L. Delle Site, K. Kremer, J. Chem. Phys. **123**, 224106 (2005).
48. W. Tschp, K. Kremer, J. Batoulis, T. Brger, O. Hahn, Acta Polymer **49**, 61 (1998).
49. J.P. Hansen, I.R. McDonald, *Theory of Simple Liquids* (Academic Press, 2006).
50. M. Baptista, R. Schmitz, B. Dünweg, *A simple and robust solver for the Poisson-Boltzmann equation*, submitted to Phys. Rev. E (2008).