

From Atomistic to Coarse Grain Systems – Procedures & Methods*

Frank Römer[†]

Theoretical Soft Matter and Biophysics (ICS-2/IAS-2), Forschungszentrum Jülich GmbH, Germany

(Dated: December 4, 2014)

The physical and mathematical basis as well as the historical background of the most popular coarse graining methods (*Reverse/Inverse Monte-Carlo*, *Iterative Boltzmann Inversion* and *Force Matching* method) in the field of fluids and soft matter are presented here. In terms of lengths and time scale, I refer here to the classical coarse grain systems, which are in between the atomistic and mesoscale systems. The focus is on the path to derive the coarse grain force fields from reference data obtained from atomistic simulations.

I. INTRODUCTION

To coarse grain a system means reducing the granularity of a system. The Wikipedia tells us [1]: “*Granularity is the extent to which a system is broken down into small parts, either the system itself or its description or observation. It is the extent to which a larger entity is subdivided. [...] Coarse-grained systems consist of fewer, larger components than fine-grained systems; a coarse-grained description of a system regards large subcomponents while a fine-grained description regards smaller components of which the larger ones are composed.*”

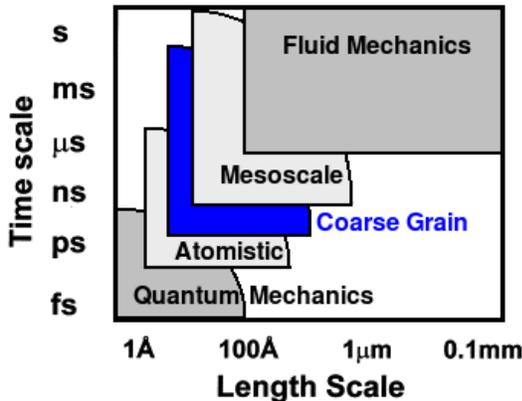


FIG. 1. Time and length scales covered by different computational methods.

In terms of computer simulations it means to reduce the degrees of freedom of a given system. Coming from *ab initio*, were strictly speaking already simplifications like Born-Oppenheimer approximation are introduced, to atomistic models we neglect at least partially ¹ the degree of freedom of the electrons. So, technically speaking every step upwards in Fig. 1 can be interpreted as coarse

graining. However, in this work I will focus on systems between atomistic and mesoscale, which are usually referred as coarse grain systems.

The coarse graining of a system is mostly motivated by the time or length scale of the phenomenon one like to study. Actually, the limiting factor which determines the boundaries of the domains for the different computational methods, as sketched in Fig. 1, is the computational power. Hence, one resign on including certain details of a system in order to capture the phenomenon of interest.

In 1990 B. Smit *et al.* published a short Nature article about computer simulations of a ternary system containing oil, water and amphiphilic molecules [2], which is often cited as one of the first coarse graining attempt. From our present point of view the system is very simple. It contains only two types of particle *o* and *w*. A water molecule is represented by a single *w* particle, an oil molecule by a single *o* particle. The amphiphilic molecule is a chain of two *w* particles, representing the polar head group, followed by five *o* particles, representing hydrophobic tail. All particles interact by a Lennard-Jones potential [3], while the *o*-*o* and *w*-*w* interactions are truncated at 2.5σ and the *o*-*w* interactions at $2^{1/6}\sigma$, which makes this interaction completely repulsive [4]. Even this phenomenological model, as you can see in Fig. 2, shows the formation of a layer of amphiphilic molecules at the oil-water interface, as well as the formation of micelles. Certainly, this model has no analog in the real world, but at those times only more simple Monte-Carlo simulations on a grid were performed before [5] to study the structural diversity of that kind of ternary systems. Nowadays we are able to model systems which are not only able to capture the general qualitative behavior, but also quantitative properties, reflecting properly the chemical nature of the system. Therefore methods and procedures are needed to derive from experimental data or atomistic simulations proper coarse grain models, like atomistic models are derived from *ab initio* calculations [6].

The intention of this work is to introduce different methods and procedures to gain from a atomistic description of a system a competitive coarse grain description. The mathematical and physical basis as well as

* Extended abstract for a talk given at the Mulliken Center in Bonn on November 27th 2014.

[†] f.roemer@fz-juelich.de; personal: frank.roemer@wgsolve.de

¹ There are atomistic force fields, like e.g. bond-order potentials, which reintroduce partially the degree of freedom of the electrons.

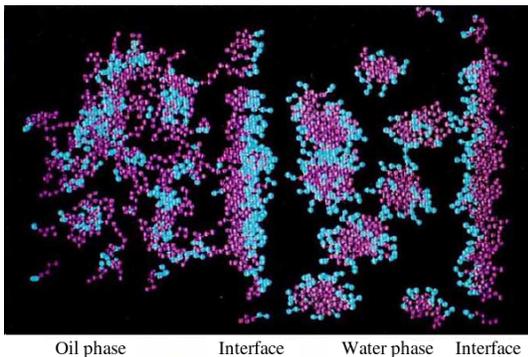


FIG. 2. A typical example of a configuration of surfactants in an oil-water system. The hydrophobic segments are blue and the hydrophilic segments are purple. The positions of the oil and water particles are not shown. [Fig. 2 from Ref. [2]]

the historical background of these methods and procedures will be elucidated, and examples for practical applications provided. The focus here is on fluid and soft matter systems, and to *present* the coarse grain methods and procedures and *not to evaluate or compare* those in detail.

II. FROM ATOMISTIC TO CG FORCE FIELDS

In general, when started from a detailed atomic description, any coarse-graining procedure defines new objects, like building blocks. These objects are intended to mimic, at least partially, the behavior of a group of atoms. This assignment has not to be mandatory bijective. One of these new objects, let's call it CG group, can be used to represent very different groups of atoms as long as the physical and chemical nature, with respect to the subject of study, is equivalent or at least negligible different. The bottom-up approach, deconstruct a molecule in groups of atoms and then find a proper description for each of these groups, is the most prominent. But one can also follow the top-down approach, by defining several CG groups with specific properties, like for example the MARTINI force field [7], and rebuild the target molecule using these CG groups.

To create a model apparently one needs a reference system. Atomistic force fields are parametrized using experimental data, e.g. OPLS force field [8], or *ab initio* respectively DFT calculations, like CFF [6] or TraPPE [9]. Analogous to these, simulations with atomistic force fields are mostly used as reference systems to parametrize CG force fields.

Like in most atomistic force fields, CG force fields contain two types of forces, bonded and non-bonded. In general a total force field can be written as follows.

$$V_{\text{tot}} = \underbrace{(V_{\text{str}} + V_{\text{bend}} + V_{\text{tors}})}_{V_{\text{bonded}}} + \underbrace{(V_{\text{vdw}} + V_{\text{es}})}_{V_{\text{non-bonded}}} \quad (1)$$

The different potential terms for bonded forces, take into account bond stretching V_{str} and bending V_{bend} , as well as the dihedral rotation barriers V_{tors} . The non-bonded forces are formed of the van-der-Waals potential V_{vdw} , which accounts for the excluded volume repulsion as well as the London attractive forces, and the V_{es} electrostatic interactions due to (partial)charges. The usual way is to adjust the terms successively, in the order of their relative contribution to the total force field [10]:

$$V_{\text{str}} \rightarrow V_{\text{bend}} \rightarrow V_{\text{non-bonded}} \rightarrow V_{\text{tors}}$$

I do like to note, that these potentials do not necessarily have to follow a determined functional form, like for example a Lennard-Jones potential, they can be compiled in tables. Often, owing to the specific coarse graining procedure, one obtains discrete potential/force values with respect to the distance, angle and so forth. In some applications determined functions are later fitted to these discrete values.

In general one can divide the coarse graining methods, which use atomistic simulations as reference systems, in two categories: one using structure information from the trajectory and one using forces or energies as well. Here I will focus on the first category, and will present

- the *Reverse Monte-Carlo* (Sec. III),
- the *Iterative Boltzmann Inversion* (Sec. IV), and
- the *Inverse Monte-Carlo method* (Sec. V).

In the last section I will show at least one variant of

- the *Force Matching* method (Sec. VI)

in order to cover the most relevant methods used in applications. Actually, most of the applications one finds in literature did not stick consistently to one of these methods. Often you will find a useful mixture of them owing to the constraints of the practical use cases.

A. Bonded Forces

Starting from an atomistic simulation as reference system, a straightforward approach would be to fit the intramolecular potentials of the coarser system to the energy functions of the fully atomistic system. This approach will be presented later in the context of the *Force Matching* method in Sec. VI where bonded and non-bonded forces are processed together. There are also other coarse graining methods on the market which come up with their own intrinsic methods to derive the bonded and non-bonded forces in “one step”. In this chapter I do like to give just as an example two widely used methods to derive coarse grain potentials for bonded interactions.

1. Harmonic approximation

When introducing a harmonic bond stretching

$$V_{\text{str}} \equiv V_{\alpha\beta}(r) = \frac{k_{\alpha\beta}}{2} (r - r_{\alpha\beta})^2 \quad (2)$$

and bending potential

$$V_{\text{bend}} \equiv V_{\alpha\beta\gamma}(\theta) = \frac{k_{\alpha\beta\gamma}}{2} (\theta - \theta_{\alpha\beta\gamma})^2, \quad (3)$$

for the CG model, one can obtain the equilibrium bond length $r_{\alpha\beta}$ between two CG groups α and β respectively the equilibrium angle $\theta_{\alpha\beta\gamma}$ between three CG groups α , β and γ , directly by calculating the referring average values from a trajectory of a atomistic simulation. Here a *bond* between two CG groups is usually defined between the center of mass of the groups of atoms representing the respective CG groups in the atomistic simulation. The force constants can than derived from the standard deviation following the simple harmonic approximation

$$\begin{aligned} k_{\alpha\beta} &= k_{\text{B}}T / \langle (r - r_{\alpha\beta})^2 \rangle, \text{ and} \\ k_{\alpha\beta\gamma} &= k_{\text{B}}T / \langle (\theta - \theta_{\alpha\beta\gamma})^2 \rangle, \end{aligned} \quad (4)$$

where k_{B} is the Boltzmann constant and T the temperature. Of course, choosing a different functional form for the potential terms, will lead to a different derivation of the force constants and may be of other parameters.

2. Boltzmann inversion method

Neither the straight forward fitting procedure nor the harmonic approximation includes the conformational entropy a priori properly. W. Tschöp *et al.* proposed a method, later referred to as Boltzmann inversion (BI) method, to overcome this problem [11]. The idea of the BI method based on the fact that in a canonical ensemble independent degrees of freedom q obey the Boltzmann distribution:

$$P(q) = Z^{-1} \exp[-U(q)/k_{\text{B}}T] \quad (5)$$

where Z is a partition function, which acts as a normalization prefactor. If $P(q)$ is known, one can invert Eq. 5 and obtain, by neglecting the normalization prefactor which would only appear as an additive constant, the potential:

$$U(q) = -k_{\text{B}}T \ln P(q) \quad (6)$$

To apply this idea to gain intra-molecular potentials for a CG model, one follow this procedure:

1. Like in the previous subsection, a full atomistic NVT simulation is performed. Here the positions of the geometrical center of gravity² of the CG groups are determined.

² Geometrical center of gravity means that here the masses of all atoms are counted the same.

2. From this positions, histograms for bond lengths $H_r(r_{\alpha\beta})$, bond angles $H_\theta(\theta_{\alpha\beta\gamma})$ and torsion angles $H_\varphi(\varphi_{\alpha\beta\gamma\omega})$ are calculated.

3. These histograms has to be rescaled in order to obtain the the volume normalized distribution functions [12]:

$$P_r(r) = \frac{H_r(r)}{4\pi r^2}, P_\theta(\theta) = \frac{H_\theta(\theta)}{\sin \theta}, P_\varphi(\varphi) = H_\varphi(\varphi). \quad (7)$$

4. Assuming a canonical distribution and statistically independent CG degrees of freedom

$$P(r, \theta, \varphi) = \exp[-U(r, \theta, \varphi)/k_{\text{B}}T] \quad (8)$$

$$P(r, \theta, \varphi) = P_r(r) \cdot P_\theta(\theta) \cdot P_\varphi(\varphi) \quad (9)$$

they are equivalent to to the Boltzmann factors, and the interaction potentials for the CG model are given by:

$$\begin{aligned} U(r, \theta, \varphi) &= U_r(r) + U_\theta(\theta) + U_\varphi(\varphi) \\ U_q(q) &= -k_{\text{B}}T \ln P_q(q) \text{ for } q = r, \theta, \varphi \end{aligned} \quad (10)$$

There are four important issues to mention: (i) The resulting CG potentials are obviously temperature dependent, and refers to the temperature used in the atomistic NVT simulation. This means, the validity of the CG potential is only given in a certain temperature range, whose scope depends on the type of the system. It should be noted that this is actually true also for the harmonic approach (Sec. II A 1) and for nearly every CG potentials which are derived from atomistic simulation data. (ii) The independence of the CG degrees of freedom is a crucial point, because it is the main assumption that allows to factorize the probability distributions Eq. 9 and the potential Eq. 10. Hence, one should check carefully for cross-correlations the ensure this assumption is hold in practice. (iii) The potentials, which due to the technique are defined at discrete points, needs to be smoothed in order to provide a continuous force, for example by splines. (iv) Regions of high potential energy, which are poorly or not sampled shall be extrapolated.

B. Non-Bonded Forces

The non-bonded interaction potential are much more complex and costly to obtain than the bonded. To present current methods to derive these potentials, including the historical, mathematical and physical background, is the main focus of this work and will be reviewed in separate chapters. Because the focus lies here on CG methods using structural information from reference system, I do like to illuminate previously some important basics underlying these methods.

1. Radial distribution function $g(r)$ Potential of Mean Force

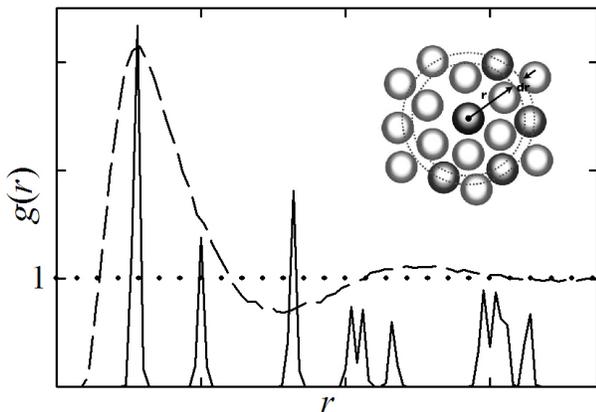


FIG. 3. Schematic sketch of a normalized radial distribution function for an ideal gas (dotted line), a liquid (dashed line) and a crystal (solid line). Inlet: 2D-sketch how the radial distribution function is derived from particle coordinates.

The radial distribution function (RDF), (or *pair correlation function*) $g(r)$ in a system of particles (atoms, molecules, colloids, etc.), describes how density varies as a function of distance from a reference particle. Hence one can gain structural information out of it, as illustrated in Fig. 3 the RDF of a crystal shows a significant different from a liquid. It can be derived from neutron scattering or x-ray diffraction experiments via Fourier transformation of the structure factors [13], and can be calculated from particle coordinates yield from simulations by [14]

$$g(r) = \frac{1}{4\pi r^2} \frac{1}{N\rho} \sum_{i=1}^N \sum_{j \neq i}^N \langle \delta(|r_{ij}| - r) \rangle, \quad (11)$$

where the normalization factor $1/4\pi r^2$ considers the volume of the spherical shell and $1/N\rho$ the average density. In case of an infinite diluted gas, one can obtain the real pair interaction potential by using the the potential of mean force (PMF) [15]

$$\psi_{\alpha,\beta}(r) = -kT \ln [g_{\alpha,\beta}(r)] \quad (12)$$

Here $g_{\alpha,\beta}(r)$ is the RDF between atoms of type α and β , k the Boltzmann constant, and T the temperature. In real systems the RDF always includes effects from the many-body interactions. Therefore we refer to those potentials as *effective pair potentials*.

Now the question arises: *Is the effective pair potential derived from a RDF unique?* Fortunately, R. L. Henderson shows very elegant in his 1974 paper [16], that under rather weak conditions two pair potentials which give rise to the same $g(r)$ cannot differ by more than a constant. This constant can easily be eliminated by adding

the condition $\psi(r \rightarrow 0) \rightarrow 0$. Ten years later, Chayes *et al.* provided the proof that such a pair potential always exists, again under rather weak conditions [17, 18]: If the given RDF is a two-particle reduction of any admissible N -particle probability distribution, there always exists a pair potential that reproduces it.

2. Henderson Theorem

The Henderson theorem based on the Gibbs-Bogoliubov inequation (aka Feynman-Kleinert variational principle [19]). For the free energy of two systems with Hamiltonians H_1 and H_2 the inequation

$$F_1 \leq F_2 + \langle H_2 - H_1 \rangle_1 \quad (13)$$

is always valid. Here $\langle \dots \rangle_1$ indicates the (canonical) average appropriate for H_1 . Equality holds only if $H_2 - H_1$ is independent of all degrees of freedom, which implies that the pair potentials can differ only by a constant. Lets consider now two identical systems ($g_1 \equiv g_2$) which differ only in the pair potential $u_1 \neq u_2$. The equality of Eq. 13 is only hold if $u_1 - u_2 = c$ is a constant. If we now consider u_1 and u_2 differ by more than a constant, and therefore also $H_2 - H_1$ we get $F_1 < F_2 + \langle H_2 - H_1 \rangle_1$ or

$$f_1 < f_2 + \frac{1}{2}n \int d^3r [u_2(r) - u_1(r)] g_1(r) \quad (14)$$

and also for the interchange of the systems

$$f_2 < f_1 + \frac{1}{2}n \int d^3r [u_1(r) - u_2(r)] g_2(r) \quad (15)$$

where n refers to the average particle density and f to the free energies per particle. If we now combine Eq. 14 and Eq. 15 and take into account $g_1 \equiv g_2$ we get the contradiction $0 < 0$. Hence, the initial assumption that u_1 and u_2 differs by more than a constant while the RDFs are equal is wrong.

III. REVERSE MONTE-CARLO (RMC) METHOD

The reverse Monte-Carlo scheme was already used in the late 1960's [20] and later [21, 22] to derive a variety of physical quantities. Confusingly, in these and other papers they refer to it inconsistently sometimes as *reverse* and sometimes as *inverse* Monte-Carlo method. R. L. McGreevy and L. Pusztai [23] first came up with an application of the RMC to determine structures in disordered systems from diffraction data. They also came up with the initial idea, that one could use this method to determine atomic potentials. Their approach is shown in the following section. One month after the RMC paper [23], M. Ostheimer and H. Bertagnolli submitted their manuscript [24] titled "*Test of the Inverse Monte Carlo*

Method for the Calculation of Interatomic Potentials in Atomic Liquids". The procedure they proposed based on the work of V. Gerold and J. Kern [22], who introduced a method to determine interaction energies in solid solutions. The methods are slightly different but the results are equivalent. The latter publication was more or less ignored (13 citations) while the previous became prominent (1008 citations). Based on the RMC application by McGreevy *et al.*, A. K. Soper developed his Empirical Potential Monte-Carlo (EPMC) method [25] to derive atomistic potentials. One has to be careful, some authors refer to the work of McGreevy *et al.*, some other to the work of Soper and again some other to both as the RMC method, depending on the field of research.

In the following subsections I will present the application of the RMC method by McGreevy *et al.* (Sec. III A), the EPMC method (Sec. III B) and an example (Sec. III C) how latter can be used in context of coarse graining potentials.

A. Structures in Disordered Materials

In the late eighties of the last century R. L. McGreevy and L. Pusztai came up with a new technique to determine structures in disordered materials [23]. They utilized the Reverse Monte-Carlo method, which has been used already since twenty years to model a variety of physical data [20–22]. The main idea is to match the radial distribution function (RDF) gain from X-ray diffraction experiments by Fourier transformation of the structure factor with the RDF computed from an Monte-Carlo ensemble.

The algorithm for this method can be summarised as follows:

1. An initial configuration of N particles in a three dimensional cubic system is generated by random. For the cubic system of side length L , periodic boundary conditions are applied and the RDF $g_S(r)$ is calculated.
2. By a random motion of one particle a new configuration is generated and the new RDF $g'_S(r)$ is calculated.
3. Both RDFs, $g_S(r)$ and $g'_S(r)$ are now compared with the experimental RDF $g_E(r)$ by using a χ^2 -test:

$$\chi^2 = \sum_{i=1}^{n_r} (g_E(r_i) - g_S(r_i))^2 / \sigma_E^2(r_i) \quad (16)$$

$$\chi'^2 = \sum_{i=1}^{n_r} (g_E(r_i) - g'_S(r_i))^2 / \sigma_E^2(r_i) \quad (17)$$

where n_r is the number of r points and σ_E is the experimental error.

4. If $\chi'^2 < \chi^2$ is accepted. If $\chi'^2 > \chi^2$ it is accepted with a probability that follows a normal distribution, given by

$$P = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\Delta\chi^2}{2\sigma^2}\right), \quad (18)$$

where the with σ is a tuneable parameter.

5. If the new configuration is accepted it becomes the new starting configuration $g_S(r) \leftarrow g'_S(r)$, otherwise the old starting configuration retained. We then repeat from step 2 until χ^2 is converged.
6. Then we continue and saving the accepted configuration. The average $g_S(r)$ calculated from this configurations agrees with the experiment within the experimental error.

From these last ensembles one can now extract further information which are not directly available through the X-ray diffraction data. As a proof of concept, they show excellent agreement³ ($\chi^2/n_r = 0.075$) in the RDF in comparison to experimental results on Argon [26]. For a MC ensemble containing $N = 512$ particles, it takes takes 10697 moves to convergence, where 2070 were accepted.

R. L. McGreevy and L. Pusztai already suggested in this paper, that *there may even be methods whereby the derived structures can actually be used to generate pair potentials*. Due to a comment from R. Evans on this paper [27], one can gain at best an effective pair potential, if $g_E(r)$ correspond to data for real liquids.

For more informataion about the application of the RMC modelling, I do like to refer to a review published in 2001 by McGreevy [28].

B. Empirical Potential Monte-Carlo method

A few years later in the mid-nineties, inspired by the work of R. L. McGreevy and L. Pusztai, A. K Soper shows a simple bootstrap procedure to derive an empirical potential energy function, he called Empirical Potential Monte-Carlo (EPMC) method [25]. He patched in this method two inherent shortcomings of the original RMC method due to the usage of χ^2 (Eq. 16 and 17) as a constraint. Firstly, χ^2 can not distinguish between one configuration with a large statistical uncertainty but matches well the target RDF and a configuration with lower statistical uncertainty but misfits the peaks. Secondly, because of constraints in the number of particles in MC ensemble and numerical accuracy the relative uncertainty of $g_S(r)$ can become one order of magnitude larger than of diffraction data. Therefore, trying to fit $g(r)$ better than this intrinsic uncertainty by reducing σ (in Eq.

³ Due to the mean quality of the figures I resign to re-print them here. See Fig. 1 in Ref. [23].

18) may lock the MC ensemble in a local minimum and prevent a true random walk.

The object of A. K Soper was to extend the RMC method so that it allows the system to sample a wide range of phase space and therefore correctly mimicking the experiment and so that it derives an effective pair potential which, when used in simulations, reproduces quantitatively the experimentally obtained RDF.

We assume to have a set of RDFs $g_{\alpha,\beta}^D(r)$, derived from previous computer simulations or measured partial structure factors, and a corresponding set of reference pair potentials $U_{\alpha,\beta}^{ref}(r)$, which only needs to incorporate known hardcore limitations on atomic overlap (see. Fig. 4) and in case of molecules configurational constraints. The starting point of the idea is the potential of mean force (Eq. 12). The PMF $\psi_{\alpha,\beta}(r)$ calculated for the the input data, which would be per definition a purely pairwise potential, can not be used as a potential in simulation because it already contains the many-body cooperative effects. Nevertheless the PMF can be used to determine where a given pair potential needs to be modified to reproduce a target RDF.

The iterative procedure works as follows:

1. Set up a model fluid with the correct density and temperature. The initial potentials $U_{\alpha,\beta}^0(r)$ are given by $U_{\alpha,\beta}^{ref}(r)$.
2. A Markov-Chain-Monte-Carlo (MCMC) simulation [14] is performed ⁴ for several steps and the $g_{\alpha,\beta}(r)$ calculated.
3. The PMF is now used to generate a new potential energy function $U_{\alpha,\beta}^N(r)$, as a perturbation of the initial/previous:

$$U_{\alpha,\beta}^N(r) = U_{\alpha,\beta}^0(r) + \left[\psi_{\alpha,\beta}^D(r) - \psi_{\alpha,\beta}(r) \right] \quad (19a)$$

$$= U_{\alpha,\beta}^0(r) + kT \ln \left[g_{\alpha,\beta}(r) / g_{\alpha,\beta}^D(r) \right] \quad (19b)$$

4. Now $U_{\alpha,\beta}^N(r)$ replaces $U_{\alpha,\beta}^0(r)$ and we continue going back to step 2, until the process is converged:

$$U_{\alpha,\beta}^0(r) \approx U_{\alpha,\beta}^N(r) = U_{\alpha,\beta}(r) \quad (20)$$

and therefore:

$$g_{\alpha,\beta}(r) \approx g_{\alpha,\beta}^D(r) \quad (21)$$

for all r and all atom pairs.

⁴ Because here, in contrast to McGreevy and Pusztai method, an interaction potential is introduced, one can also perform MD instead of MC simulations.

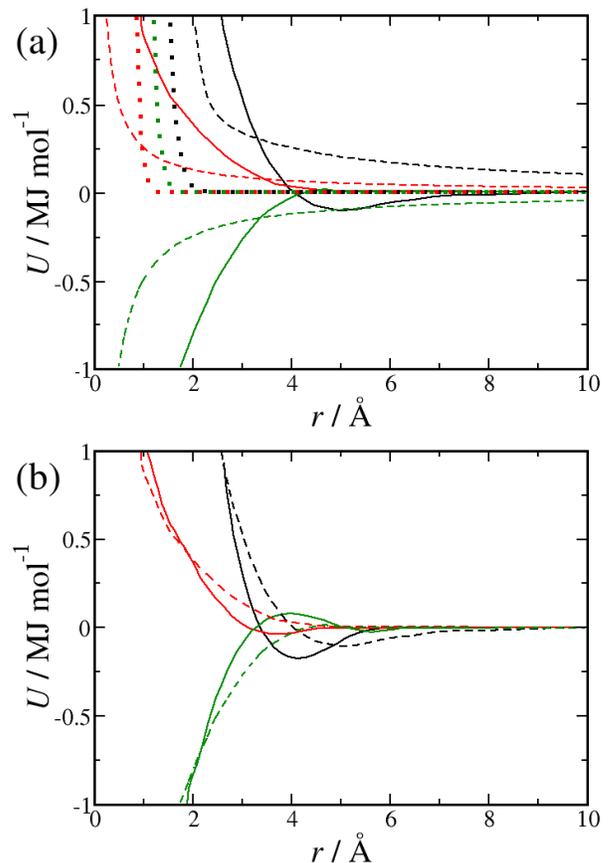


FIG. 4. Potentials for the OO- (black), HH- (red) and OH-pair (green). (a) Comparison of the empirical potential (solid line) derived from RDFs gain from SPC/E simulations, with the true SPC/E [29] (dashed line) and the initial guess $U_{\alpha,\beta}^{ref}(r)$ (dotted line). (b) Comparison of the empirical potential derived from experimental data for water [30] (solid line) with those derived from RDFs gain from SPC/E simulations (dashed line). (The data for the empirical potentials are extracted from Fig. 2&8 from Ref. [25].)

For more details about the implementation of the procedure and some important practical considerations, please consult section 3 of Ref. 25.

A. K Soper tested the proposed method on liquid water. The reference RDF data were taken from MD simulations using the SPC/E potential [29] and from partial structure factors, yield from neutron scattering experiments [30]. As shown in in Fig. 4, the resulting empirical pair potentials are rather different from the true SPC/E potential but showing a similar *qualitative* trends. The agreement with both, the RDFs taken from SPC/E simulations, as well the RDFs calculated from partial structure factors from water, as you can see in Fig. 5, are excellent.

In a more recent publications Soper [31] refers to this method as *empirical potential structure refinement* (EPSR) method. The previous mentioned review about RMC modeling [28] also covers the EPMC/EPSR

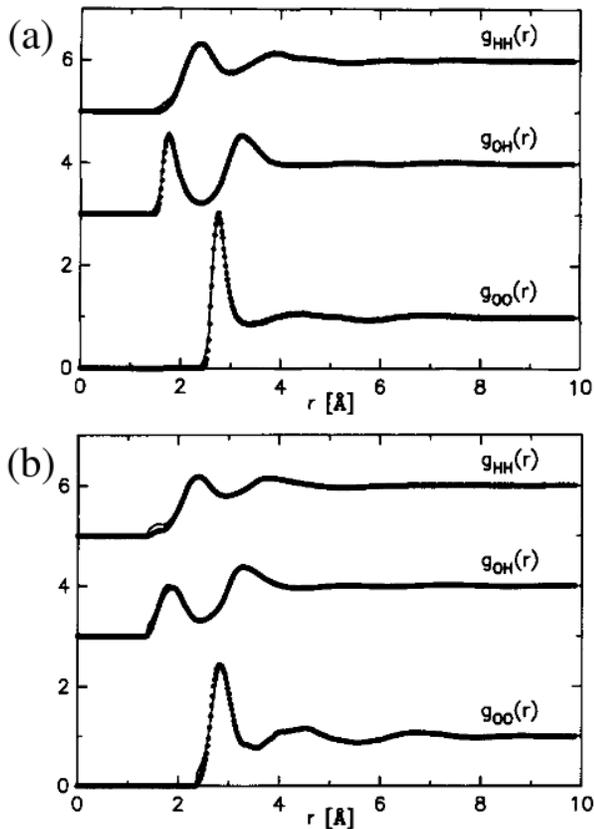


FIG. 5. The RDFs from the empirical potential (solid lines) compared with (a) those obtained from true SPC/E potential and (b) derived from partial structure factors. (Figures are taken from Ref. [25].)

method. In some publications Soper’s method is by mistake subsumed to the *inverse Monte-Carlo* method (Sec. ??) and will explained later.

C. A CG Force Field derived by the RMC/EPMC method

As an example for the application of the RMC/EPMC method [23, 25] I will present the work from J. Elezgaray and M. Laguerre who proposed a systematic method to derive CG force fields for phospholipids [32]. More precisely, it is about dimyristoylphosphatidylcholine (DMPC) and how the CG force field preform in order to model DMPC bilayers.

For the DMPC molecule (see Fig. 6) they consider four different CG groups: choline (CHOL), phosphate (PHOS), glycerol backbone (GLYC) and alkane groups (CH23). There are two charges placed, $q = -1e$ attached to the phosphate group and $q = +1e$ to the choline group. The solvent is modeled by an assembly of three water molecules (3TIP).

The bonded interactions, bond stretching and bending, are modeled by a harmonic potential and the parameters

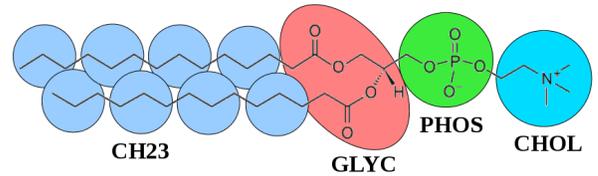


FIG. 6. Structural formula of DMPC underlying the CG model with four different CG group types, as proposed by J. Elezgaray and M. Laguerre [32].

are derived as shown in Sec. II A from atomistic simulations. To calculate the non-bonded interaction potentials, they follow in general the same iterative procedure as shown before in Sec. III B. Instead of atoms, α and β now representing the groups of atoms as defined above. A feasible implementation of Eq. 19 is used as follows

$$U_{\alpha,\beta}^{n+1}(r) = U_{\alpha,\beta}^n(r) + \eta kT \ln \left[\frac{g_{\alpha,\beta}^n(r) + \delta}{g_{\alpha,\beta}^{\text{target}}(r) + \delta} \right], \quad (22)$$

here $\eta = 0.1$ is a tunable parameter to improve convergence of the iteration. The parameter $\delta = 10^{-3}$ prevents divergence of Eq. 22 when $r \rightarrow 0$. The deviation of the RDFs is judged by

$$\epsilon_n = \frac{1}{N_{\text{pair}}} \sum_{\alpha,\beta,\{r < r_{\text{cut}}\}} \left| g_{\alpha,\beta}^n(r) - g_{\alpha,\beta}^{\text{target}}(r) \right|^2, \quad (23)$$

and convergence is assumed when $\epsilon_n < \epsilon_{\text{max}}$ is satisfied. The total pressure tends to increase steadily during the RMC iteration Eq. 22. Therefore, as in Ref. [33], a simple rescaling of the interaction potentials $U_{\alpha,\beta}^n(r)$ of the form $1 + (P_n - P_{\text{target}})$ is applied each time when $P_n > 100$ bar or $\epsilon_n < \delta$. Here P_{target} is set to 1 bar.

1. The initial pair potentials $U_{\alpha,\beta}^0(r)$

Because the solution of the RMC process is not unique and can be strongly biased by the guess for the initial potentials [34], an appropriated choice here is crucial. To use the PMF as a starting point [25, 35] is only useful when dealing with unstructured liquids. In case of structured systems, like lipid bilayers, at may lead to undesirable results because the MC system is hindered to explore a sufficient part of the phase space. To obtain proper initial pair potentials, they preform for each couple (α, β) of CG group types a preliminary all-atom *NVT* simulations. In case of the solute-solute couples, ≈ 10 of each CG groups, as its all-atom representation, were simulated in water and if necessary with counter ions. The broken bonds, that connects them to neighboring groups, were patched with hydrogen atoms. The water-water pair was treated similar to the previous. A meaningful iterative procedure was implemented to gather the water molecules in groups of three, to calculate the respective RDFs. For

the solvent-water couples, a single solute CG group is placed in a 19 Å cubic water box. The RDFs generated for the 15 couples are then used as target functions to obtain the initial pair potentials $U_{\alpha,\beta}^0(r)$ using the RMC with the PMF as the first guess.

2. CG model for a DMPC molecule

The target RDFs were derived from an atomistic *NPT* simulation of a 2×32 DMPC in a $40 \times 40 \times 70$ Å box filled with water. With the initial pair potentials computed before, the RMC reaches convergence ($\epsilon_{\max} = 10^{-2}$) after 20 iterations. The presented examples (CHOL-CHOL, CHOL-3TIB) for comparison of the RDF in Fig. 7 between atomistic and CG simulation shows a good agreement. The density profiles through the bilayer, shown in Fig. 8, are also in a fair good agreement with atomistic simulation results.

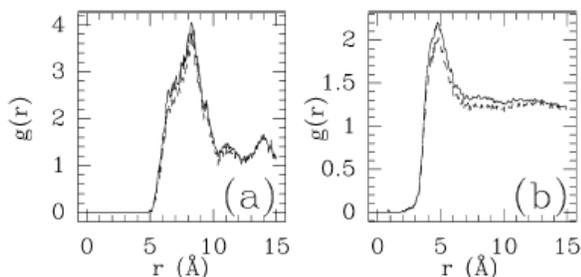


FIG. 7. Comparison of the radial distribution functions $g(r)$ for (a) CHOLCHOL and (b) CHOL3TIP. Continuous line: data obtained with the optimized potentials. Dashed-line data obtained from a coarse-grained version of the reference (full-atom) simulation. [Fig. 2 from Ref. [32]]

It should be mentioned, that this work from J. Elezgaray and M. Laguerre aimed to describe a general framework to derive CG force fields. The transferability of this specific model to other DMPC structures, like micelles, have to be checked as the authors also noted.

IV. ITERATIVE BOLTZMANN INVERSION (IBI) METHOD

The iterative Boltzmann inversion (IBI) is a natural extension of the Boltzmann inversion method [11], as described in Sec. II A 2, proposed for coarse graining of intra-molecular potentials. The IBI was introduced by D. Reith *et al.* [33] for deriving CG potentials, and can be used to refine those intra-molecular potentials, as well as to derive non-bonded CG interaction potentials from atomistic simulations. The histogram, respectively the probability distribution function accounting for the internal degrees of freedom, like for example a bond length $P_r(r) = H_r(r)/4\pi r^2$, are here replaced by the RDF. Because $g(r)$ is already normalized (Eq. 11), we can simply

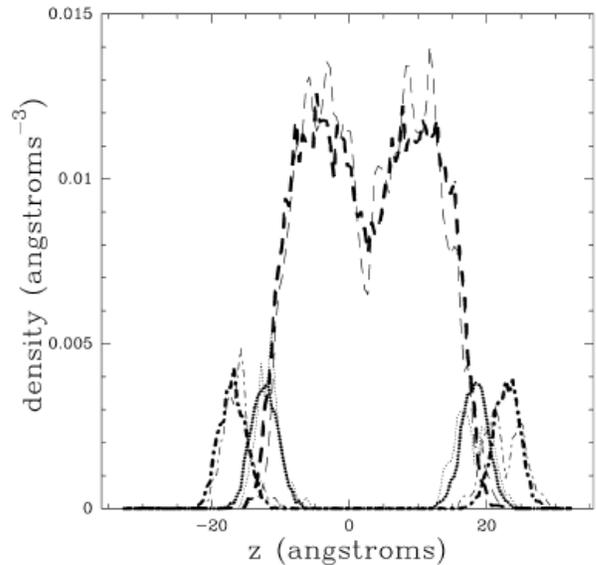


FIG. 8. Comparison of the atomistic (thin) and coarse-grained (thick) density profiles perpendicular to the bilayer for the GLYC (short dashed), PHOS (short-long dashed) and CH23 (long dashed). [Fig. 3 from Ref. [32]]

replace the probability distribution function (Eq. 10) in the IB method by the RDF.

In the IBI method a CG potential is refined in each iteration step n according to the following scheme:

$$U^{n+1} = U^n + \Delta U^n$$

$$\Delta U^n(r) = k_B T \ln \left(\frac{g^n(r)}{g_{\text{ref}}(r)} \right) \quad (24)$$

The correction term for the potential (Eq. 24) here is equivalent to the term (Eq. 19) used in the iteration procedure of the EPMC method from Soper [25], as shown in Sec. III B. Like in the EPMC method, the PMF (Eq. 12) is used to gain initial potential $U^0(r)$ values.

In practice I can not see any difference between the IBI and the EPMC method. Reith *et al.* [33] even cited the Soper’s paper [25] as one of the RMC technique approaches, but claim further: “[...] *The lack of a potential, though, limits the usefulness of the approach. [...]*”, which is actually not true. As discussed before, this was exactly one of the drawbacks of the RMC proposed by McGreevy and Pusztai (Sec. III A) which Soper overcomes with his EPMC method. To conclude, the IBI and the EPMC method are equivalent to each other, the only difference is, that Soper proposed his EPMC method originally to derive atomistic potentials and Reith *et al.* proposed their IBI as a method to derive CG potentials.

V. INVERSE MONTE-CARLO (IMC) METHOD

Nearly at the same time when Soper published his EPMC paper, Lyubartsev and Laaksonen [35] proposed

a method to calculate effective interaction potentials from the RDFs. They first called it “*A reverse Monte-Carlo Approach*”, but later they [36] such as others (e.g. [12, 37]) will refer to it as *inverse Monte-Carlo (IMC)* method.

Inspired by the renormalization group Monte-Carlo method for phase transition studies in the Ising model by R. H. Swendsen [38], they observe the Hamiltonian of the system:

$$H = \sum_{ij} U(r_{ij}), \quad (25)$$

where $U(r_{ij})$ is the pair potential between particle i and j . We assume the potential is short-ranged and introduce a cut-off radius, hence $U(r_{ij}) = 0$ if $r_{ij} \geq r_{\text{cut}}$. The potential is tabulated on a grid of M points, $r_\alpha = \alpha\Delta r$, where $\alpha = [0, M]$, and $\Delta r = r_{\text{cut}}/M$ is the spacing of the grid. Now we can rewrite the Hamiltonian, Eq. 25, as:

$$H = \sum_{\alpha} U_{\alpha} S_{\alpha} \quad (26)$$

where S_{α} is the number of particle pairs where $r_{ij} = r_{\alpha}$ which correspond to the tabulated value of the potential U_{α} . Obviously, the S_{α} can be connected to the RDF:

$$\langle S_{\alpha} \rangle = \frac{N(N-1)}{2} \frac{4\pi r_{\alpha}^2 \Delta r}{V} g(r_{\alpha}) \quad (27)$$

where N is the number of atoms, $N(N-1)/2$ the number of pairs and V the total volume of the system. Because $\langle S_{\alpha} \rangle$ is a function of the potential U_{α} , and can be written as a Taylor series

$$\Delta \langle S_{\alpha} \rangle = \sum_{\gamma} \frac{\partial \langle S_{\alpha} \rangle}{\partial U_{\gamma}} \Delta U_{\gamma} + \mathcal{O}(\Delta U^2). \quad (28)$$

The derivatives can be obtained by using the chain rule:

$$\begin{aligned} A &= \frac{\partial \langle S_{\alpha} \rangle}{\partial U_{\gamma}} \\ &= \frac{\partial}{\partial U_{\gamma}} \frac{\int dq S_{\alpha}(q) \exp \left[-\beta \sum_{\gamma} U_{\gamma} S_{\gamma}(q) \right]}{\int dq \exp \left[-\beta \sum_{\gamma} U_{\gamma} S_{\gamma}(q) \right]} \\ &= \beta (\langle S_{\alpha} \rangle \langle S_{\gamma} \rangle - \langle S_{\alpha} S_{\gamma} \rangle). \end{aligned} \quad (29)$$

with $\beta = 1/k_{\text{B}}T$ and q number of degrees of freedom of the system. Now we can obtain from RDF(s), using relation Eq. 27, the correction terms for the potential(s) by combining Eq. 28+29 and solving a set of linear equations:

$$\langle S_{\alpha} \rangle - S^{\text{ref}} = \sum_{\gamma} A_{\alpha\gamma} \Delta U_{\gamma} \quad (30)$$

here S^{ref} refers to the target RDF. This can now be used in the same iteration environment than in Sec. III B described for the EPMC method. As evaluation function for convergence $\Delta \langle S_{\alpha} \rangle$ is observed.

An advantage of the IMC method in comparison to the RMC method is, in the case of a multicomponent system, that it takes into account cross correlations of the RDFs and therefore the updates for U_{AA} , U_{AB} and U_{BB} are interdependent. Hence, the iteration shall converge faster.

VI. FORCE MATCHING

A. The FM method

First F. Ercolessi and J. B. Adams proposed a method to gain atomistic interaction potentials directly from atomic coordinates and forces obtained from *ab initio* calculations [39], which they call Force-Matching (FM) method. Based on this work, S. Izvekov and G. A. Voth introduced a procedure to derive CG potentials from atomistic simulation data [40].

1. Theoretical background

Let us consider we have a system containing $i = 1, \dots, N$ atoms or CG sites and we perform *ab initio* calculations or atomistic MD simulations to gain $l = 1, \dots, L$ configurations and $\mathbf{F}_{il}^{\text{ref}}$ forces. In the original FM method by Ercolessi and Adams, the classical forces $\mathbf{F}_{il}^{\text{p}}$ are defined in a preselected analytical form depending on a set of M parameters g_1, \dots, g_M , are optimized in a least squares sense by minimizing directly the the objective function

$$\chi^2 = \frac{1}{3LN} \sum_{l=1}^L \sum_{i=1}^N |\mathbf{F}_{il}^{\text{ref}} - \mathbf{F}_{il}^{\text{p}}(g_1, \dots, g_M)|^2. \quad (31)$$

In such a form the FM procedures encounters difficulties as the number of fitting parameter grows, for example due to a large number of species. Also, fitting the potential parameter simultaneously to the whole database of coordinates and forces will limit the accuracy.

Izvekov and Voth replaced in their FM procedure the minimization of Eq. 31 with an *explicit configurational averaging* scheme, which makes the method more accurate and efficient. Since the method can be used for CG as well as atomistic potentials, instead of CG group or atom, the term particle will be used during the further description. They ensure a linear dependency of the force fields $\mathbf{F}_{il}^{\text{p}}$ on its parameters $\{g_j\} = (g_1, \dots, g_M)$ by using cubic splines [41]. The minimization of χ^2 in Eq. 31 with respect to the parameter vector $\{g_j\}$ can be written in a matrix notation [42] as

$$\left\| (\mathbf{F}_{il}^{\text{p}})'_{g_j} \right\|^T \left\| (\mathbf{F}_{il}^{\text{p}})'_{g_j} \right\| \{g_j\} = \left\| (\mathbf{F}_{il}^{\text{p}})'_{g_j} \right\|^T \mathbf{F}_{il}^{\text{ref}}, \quad (32)$$

$$i = [1, N], l = [1, L],$$

where the matrix indices are j and the index numbering components of the vector \mathbf{F} . Equation 32 actually represents a set of linear equations with respect to the set of

parameters $\{g_j\}$ and can be reduced to

$$\mathbf{F}_{il}^p(g_1, \dots, g_M) = \mathbf{F}_{il}^{\text{ref}}, \quad i = [1, N], \quad l = [1, L], \quad (33)$$

which is an overdetermined system of linear equations if $M < N \times L$. A unique solution of an overdetermined system in the least-squares sense can often be found using algorithms such as QR [43] or singular value decomposition [44].

2. Implementation

To fit pairwise central force field with this method, the force $\mathbf{f}_i^p(r_{ij})$ acting between particle i and particle j is partitioned into a short-ranged part and a long-ranged

$$\mathbf{F}_{\alpha il}^{\text{ref}} = - \sum_{\gamma=\text{nb,b}} \sum_{\beta=1}^K \sum_{j=1}^{N_\beta} \left(f(r_{\alpha il, \beta jl}, \{r_{\alpha\beta, \gamma, k}\}, \{f_{\alpha\beta, \gamma, k}\}, \{f''_{\alpha\beta, \gamma, k}\}) + \frac{q_{\alpha\beta}}{r_{\alpha il, \beta jl}^2} \delta_{\gamma, \text{nb}} \right) \mathbf{n}_{\alpha il, \beta jl} \quad (36)$$

In Eq. 36, αil labels the i th particle of species α in the l th configuration from the trajectory; $r_{\alpha il, \beta jl}$ is the distance between the i th particle of type α and the j th particle of type β in the l th configuration; K is the number of species and N_β number of particles of type β . The parameters $f_{\alpha\beta, \gamma, k}$, $f''_{\alpha\beta, \gamma, k}$ and $q_{\alpha\beta}$ are subjected to the fit. The additional index $\gamma = \text{nb, b}$ permits a systematic separation of bonded and non-bonded forces. If the total number of configurations L is large enough, thus the equations overdetermine the force parameter, one can use the QR algorithm or the single value decomposition method to obtain all force parameter in a least square sense. The site partial charges q_α can be readily recovered by solving the system of nonlinear equations

$$q_\alpha q_\beta = \langle q_{\alpha\beta} \rangle \quad (37)$$

where $\langle q_{\alpha\beta} \rangle$ are parameters obtained from Eq. 36 and then averaged over trajectories.

3. Correction

Often CG force fields fail to maintain the proper internal pressure and as a result also predict wrong densities. The reason of this systematic fail can be easily under-

Coulomb part as

$$\mathbf{f}_i^p(r_{ij}) = - \left(f(r_{ij}) + \frac{q_i q_j}{r_{ij}^2} \right) \mathbf{n}_{ij}, \quad (34)$$

where $r_{ij} = \|\mathbf{r}_{ij}\| = \|\mathbf{r}_j - \mathbf{r}_i\|$ is distance between the particles, q is the partial charge of a particle, and $\mathbf{n}_{ij} = \mathbf{r}_{ij}/r_{ij}$. The short ranged term $f(r)$ is expressed by cubic splines connecting a set of point $\{r_k\}$ and the respective tabulated values $\{f_k\}$, $\{f''_k\}$, thus preserving continuity of its functions and their first two derivatives, such that

$$\begin{aligned} f(r, \{r_k\}, \{f_k\}, \{f''_k\}) = \\ A(r, \{r_k\})f_i + B(r, \{r_k\})f_{i+1} \\ + C(r, \{r_k\})f''_i + D(r, \{r_k\})f''_{i+1} \\ \text{with } r \in [r_i, r_{i+1}], \end{aligned} \quad (35)$$

where A, B, C and D are known functions of r respectively $\{r_k\}$. The parameters $\{f_k\}$, $\{f''_k\}$ and $q_{ij} = q_i q_j$ are to be obtained from the fit. By combining Eq.34 and 35 we can now express the known reference forces $\mathbf{F}_{\alpha il}^{\text{ref}}$ for particles of species $\alpha = [1, K]$ and for a given configuration $l = [1, L]$ in the following linear equations:

for each particle of species α : $i = [1, N_\alpha]$.

stood by examining the virial equations used to evaluate the pressure in MD simulation:

$$P = \left(\frac{2}{3} \langle E^{\text{kin}} \rangle + \langle W \rangle \right) / V. \quad (38)$$

where $\langle E^{\text{kin}} \rangle$ is average kinetic energy and $\langle W \rangle$ the virial of the system. The first term in Eq. 38, the average kinetic energy

$$\langle E^{\text{kin}} \rangle = N k_B T / 2 \quad (39)$$

is not conserved in a CG system because of the reduction of the number of system degrees of freedom N . The second term, the virial of the system

$$\langle W \rangle = \left\langle \frac{1}{3} \sum_{i < j} \mathbf{f}_{ij} \cdot \mathbf{r}_{ij} \right\rangle \quad (40)$$

is also not conserved due to contributions which are *intra* with respect to the atomic groups subject to coarse graining. Fortunately, because the virial $\langle W \rangle$ depends linearly on the atomic forces and E_{kin} does not rely on forces at all, the FM force eld can also be constrained to produce the correct pressure. This task can be accomplished by adding to Eq. 36 the constraint

$$3W_l^{\text{atom}} + 2\Delta E_l^{\text{kin}} = \sum_{\gamma=\text{nb,b}} \sum_{\alpha\beta} \sum_{ij} \left(f(r_{\alpha il, \beta jl}, \{r_{\alpha\beta, \gamma, k}\}, \{f_{\alpha\beta, \gamma, k}\}, \{f''_{\alpha\beta, \gamma, k}\}) r_{\alpha il, \beta jl} + \frac{q_{\alpha\beta}}{r_{\alpha il, \beta jl}} \delta_{\gamma, \text{nb}} \right) \quad (41)$$

where W_l^{atom} is the *instantaneous* virial or the l th configuration along the atomistic trajectory and

$$\Delta E_l^{\text{kin}} = E_l^{\text{kin,atom}} - E_l^{\text{kin,CG}} \quad (42)$$

which is the difference in *instantaneous* kinetic energy between the atomistic and CG representation of the system. Latter can be approximated by

$$\Delta E_l^{\text{kin}} = E_l^{\text{kin,atom}} (1 - N^{\text{CG}}/N^{\text{atom}}) \quad (43)$$

where $N^{\text{atom}}, N^{\text{CG}}$ are the number of degrees of freedom of the atomistic and the CG system. Because Eq. 41 depends explicitly on the temperature in the reference atomistic system, it may further reduce the transferability of the CG model to other temperatures and also thermodynamic properties which depend on the derivation of the temperature may be less accurate.

VII. VOTCA

Rühle *et al.* [12, 45] published the *Versatile Object-Oriented Toolkit for Coarse-Graining Applications* (VOTCA) which provides a framework which implements

- the BI for bonded potentials,

- the Iterative Boltzmann Inversion,
- the Inverse Monte Carlo, and
- the Force Matching method,

as well as a flexible modular platform for the further development of coarse-graining techniques. It supported a variety of file formats and therefore also simulation programs, like:

- xtc, trr, tpr
(all formats supported by GROMACS [46])
- DL_POLY [47] FIELD and HISTORY
- LAMMPS [48] dump files
- pdb, xyz
(to use with ESPResSo [49] and ESPResSo++ [50])

ACKNOWLEDGMENTS

Beside the original references, I took some nice formulations from the review article from Mortula *et al.* [37] and from the VOTCA paper from Rühle *et al.* [12].

[1] Wikipedia, “Granularity,” (6.11.2014), <http://en.wikipedia.org/wiki/Granularity>.

[2] B. Smit, P. A. J. Hilbers, K. Esselink, L. A. M. Rupert, N. M. van Os, and A. G. Schlijper, *Nature* **348**, 624 (1990).

[3] J. E. Lennard-Jones, *Proc. Phys. Soc. London* **43**, 461 (1931).

[4] J. D. Weeks, D. Chandler, and H. C. Andersen, *J. Chem. Phys.* **54**, 5237 (1971).

[5] R. G. Larson, *The Journal of Chemical Physics* **89**, 1642 (1988).

[6] J. R. Maple, M.-J. Hwang, T. P. Stockfisch, U. Dinur, M. Waldman, C. S. Ewig, and A. T. Hagler, *J. Comput. Chem.* **15**, 162 (1994).

[7] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. de Vries, *J. Phys. Chem. B* **111**, 7812 (2007).

[8] W. J. Jorgensen and J. Tirado-Rives, *J. Am. Chem. Soc.* **110**, 1657 (1988).

[9] M. G. Martin and J. I. Siepmann, *J. Phys. Chem. B* **102**, 2569 (1998).

[10] D. Reith, H. Meyer, and F. Müller-Plathe, *Macromolecules* **34**, 2335 (2001).

[11] W. Tschöp, K. Kremer, J. Batoulis, T. Bürger, and O. Hahn, *Acta Polymerica* **49**, 61 (1998).

[12] V. Rühle, C. Junghans, A. Lukyanov, K. Kremer, and D. Andrienko, *J. Chem. Theory Comput.* **5**, 3211 (2009).

[13] H. E. Fischer, A. C. Barnes, and P. S. Salmon, *Rep. Prog. Phys.* **69**, 233 (2006).

[14] M. P. Allen and D. J. Tildesley, *Computer simulation of liquids* (Clarendon Press, New York, 1991).

[15] J. Hansen and I. McDonald, *Theory of Simple Liquids*, 2nd ed. (Academic Press, London, 1986).

[16] R. Henderson, *Physics Letters A* **49**, 197 (1974).

[17] J. Chayes and L. Chayes, *Journal of Statistical Physics* **36**, 471 (1984).

[18] J. Chayes, L. Chayes, and E. Lieb, *Communications in Mathematical Physics* **93**, 57 (1984).

[19] R. P. Feynman and H. Kleinert, *Phys. Rev. A* **34**, 5080 (1986).

[20] R. Kaplow, T. A. Rowe, and B. L. Averbach, *Phys. Rev.* **168**, 1068 (1968).

[21] W. Dunn, V. Boffi, and F. O’Foghludha, *Nucl. Inst. and Meth.* **255**, 147 (1987).

[22] V. Gerold and J. Kern, *Acta Met.* **35**, 393 (1987).

- [23] R. L. McGreevy and L. Pusztai, *Mol. Simul.* **1**, 359 (1988).
- [24] M. Ostheimer and H. Bertagnolli, *Mol. Simul.* **3**, 227 (1989).
- [25] A. K. Soper, *Chem. Phys.* **202**, 295 (1996).
- [26] J. L. Yarnell, M. J. Katz, R. G. Wenzel, and S. H. Koenig, *Phys. Rev. A* **7**, 2130 (1973).
- [27] R. Evans, *Mol. Simul.* **4**, 409 (1990).
- [28] R. L. McGreevy, *J. Phys.: Condens. Matter* **13**, R877 (2001).
- [29] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, *J. Phys. Chem.* **91**, 6269 (1987).
- [30] A. K. Soper, *J. Chem. Phys.* **101**, 6888 (1994).
- [31] A. Soper, *Mol. Phys.* **99**, 1503 (2001).
- [32] J. Elezgaray and M. Laguerre, *Comput. Phys. Commun.* **175**, 264 (2006).
- [33] D. Reith, M. Pütz, and F. Müller-Plathe, *J. Comput. Chem.* **24**, 1624 (2003).
- [34] P. Jedlovsky, I. Bakó, G. Pílinkás, T. Radnai, and A. K. Soper, *J. Chem. Phys.* **105**, 245 (1996).
- [35] A. P. Lyubartsev and A. Laaksonen, *Phys. Rev. E* **52**, 3730 (1995).
- [36] A. P. Lyubartsev, M. Karttunen, I. Vattulainen, and A. Laaksonen, *Soft Materials* **1**, 121 (2002).
- [37] T. Murtola, A. Bunker, I. Vattulainen, M. Deserno, and M. Karttunen, *Phys. Chem. Chem. Phys.* **11**, 1869 (2009).
- [38] R. H. Swendsen, *Phys. Rev. Lett.* **42**, 859 (1979).
- [39] F. Ercolessi and J. B. Adams, *Europhysics Letters* **26**, 583 (1994).
- [40] S. Izvekov and G. A. Voth, *J. Chem. Phys.* **123**, 134105 (2005).
- [41] C. De Boor, *A Practical Guide to Splines* (Springer, New York, 1978).
- [42] Wikipedia, “Linear least squares,” (6.11.2014), [http://en.wikipedia.org/wiki/Linear_least_squares_\(mathematics\)](http://en.wikipedia.org/wiki/Linear_least_squares_(mathematics)).
- [43] Wikipedia, “QR decomposition,” (6.11.2014), http://en.wikipedia.org/wiki/QR_decomposition.
- [44] C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems* (Society for Industrial and Applied Mathematics, 1995).
- [45] V. Rühle, C. Junghans, A. Lukyanov, K. Kremer, and D. Andrienko, “Votka – versatile object-oriented toolkit for coarse-graining applications,” <http://www.votka.org>.
- [46] H. Berendsen, D. van der Spoel, and R. van Drunen, *Computer Physics Communications* **91**, 43 (1995).
- [47] I. T. Todorov, W. Smith, K. Trachenko, and M. T. Dove, *J. Mater. Chem.* **16**, 1911 (2006).
- [48] S. Plimpton, *J. Comput. Phys.* **117**, 1 (1995).
- [49] H. Limbach, A. Arnold, B. Mann, and C. Holm, *Computer Physics Communications* **174**, 704 (2006).
- [50] J. D. Halverson, T. Brandes, O. Lenz, A. Arnold, S. Bevc, V. Starchenko, K. Kremer, T. Stuehn, and D. Reith, *Computer Physics Communications* **184**, 1129 (2013).