Multiscale coarse graining of liquid-state systems

Cite as: J. Chem. Phys. **123**, 134105 (2005); https://doi.org/10.1063/1.2038787
Submitted: 22 June 2005 . Accepted: 25 July 2005 . Published Online: 03 October 2005

Sergei Izvekov and Gregory A. Voth

Multiscale coarse graining of liquid-state systems

Sergei Izvekov and Gregory A. Voth[a)]
Center for Biophysical Modeling and Simulation and Department of Chemistry, University of Utah, Salt Lake City, Utah 84112-0850

(Received 22 June 2005; accepted 25 July 2005; published online 3 October 2005)

A methodology is described to systematically derive coarse-grained (CG) force fields for molecular liquids from the underlying atomistic-scale forces. The coarse graining of an interparticle force field is accomplished by the application of a force-matching method to the trajectories and forces obtained from the atomistic trajectory and force data for the CG sites of the targeted system. The CG sites can be associated with the centers of mass of atomic groups because of the simplicity in the evaluation of forces acting on these sites from the atomistic data. The resulting system is called a multiscale coarse-grained (MS-CG) representation. The MS-CG method for liquids is applied here to water and methanol. For both liquids one-site and two-site CG representations without an explicit treatment of the long-ranged electrostatics have been derived. In addition, for water a two-site model having the explicit long-ranged electrostatics has been developed. To improve the thermodynamic properties (e.g., pressure and density) for the MS-CG models, the constraint for the instantaneous virial was included into the force-match procedure. The performance of the resulting models was evaluated against the underlying atomistic simulations and experiment. In contrast with existing approaches for coarse graining of liquid systems, the MS-CG approach is general, relies only on the interatomic interactions in the reference atomistic system. © *2005 American Institute of Physics*. [DOI: 10.1063/1.2038787]

I. INTRODUCTION

The development of accurate, reliable, and computationally efficient force fields is a central focus of molecular modeling. This task becomes more difficult if the force field is intended for the simulation of condensed-phase systems because of many-body effects which can reduce the transferability of the force field. However, it is generally believed that the essential character of the interactions in condensed-phase systems, including many polarization effects, can be effectively included in effective pair potentials. Because computational efficiency is one of the dominant factors in choosing particular models for interactions in simulation studies of complex (e.g., biomolecular) systems, the models based on the effective pairwise potentials have gained popularity for being computationally inexpensive. Unfortunately, even simple pairwise all-atom potentials are computationally still too costly for many applications of practical interest and often place serious limitations on the scale of molecular-dynamics (MD) simulations. To access larger simulated time scales and system sizes a further simplification of the molecular models is required, even those based on empirical potentials.

Several so-called coarse-graining (CG) approaches for the simulation of condensed-phase systems have recently been developed to overcome this problem. These approaches can be viewed as bridges between microscopic and mesoscopic lengths and time scales. The goals of the various CG approaches are the same; a simpler description of interactions in the system while minimizing the losses in the ability of the resulting CG models to predict properties of interest. The coarse graining can be accomplished in space and time through various techniques which on first inspection look very different. Possible ways of achieving this goal include mapping a system onto a structurally less detailed level by, for example, grouping atoms into single interaction sites,[1–13] eliminating fast degrees of freedom by assuming rigidity of certain parts of a system, and time averaging of effective potentials over a finite time span.[14] The level of coarse graining is obviously dictated by the scale of the desired computed properties.

In many cases the technical implementation of CG approaches can be divided into two major phases. The first is mapping a system into a reduced system having a different and less complex structure; while the second is the construction of an effective force field to describe the interaction between the reduced structural units in the new system. The second stage, the prime focus of the present work, represents a major difficulty in the applications of CG approaches. Generally, the CG effective potentials are expected to have much less transferability compared to atomistic ones. This occurs, for example, because an effective interaction between the structural units intended for coarse graining is defined by the average structure (e.g., average orientation and distances) within the complexes formed by these units in a particular thermodynamic state. The structural properties and, thus, the CG potentials even for the same phase can be very sensitive to variations in temperature and other thermodynamic conditions. Moreover, the CG potentials are ideally fitted to the all-atom systems under the same thermodynamic conditions as those for which the new CG system is designed to simu-

[a)]Electronic mail: voth@chem.utah.edu

late. This situation makes the availability of computationally efficient methods for the development of CG effective potentials critical.

Development of any (not necessarily coarse-grained) force field can be further divided into two principal stages. First is the selection of the right empirical form of the interaction between structural units. Once the empirical form is adopted the next step is the fitting of the parameters. Even assuming that the available data are of adequate quality, one still has the formidable problem of model parametrization. In fact, there is no reliable and consistent strategy even for all-atom potentials. Several approaches have been used with varied success to construct CG potentials. Typically, CG potentials of a preselected analytical form are parametrized either to reproduce average structural properties such as the density and radial distribution functions seen in atomistic simulations; for example, using an iterative adjustment of potential parameters starting from an approximation based on the potentials of mean force[6,8] through solving of the Ornstein-Zernicke equation,[2] or using the inverse Monte Carlo technique.[7,12] Alternatively, they can be parametrized to match thermodynamic properties.[11] It is worth noting that the existing approaches do not generally use directly the underlying all-atom interactions as input to derive the CG force fields. For example, this is the case when a CG potential is developed targeting the averaged properties seen for the atomistic MD simulation. Such approaches are independent parametrization procedures, despite the fact that the underlying atomistic interactions might be readily available.

A recently proposed method[15–17] for force matching[18] (FM) has proven to be able to build realistic effective force fields for molecular condensed-phase systems from *ab initio* MD simulations[19] at a low computational cost. This method is a nontrivial extension of the FM approach originally suggested by Ercolessi and Adams.[18] The first application of the new FM method resulted in a family of pairwise nonpolarizable (simple point-charge-like) water models which showed far better agreement with experiment than many other conventional effective water potentials.[15] Other recent applications of this method which seem to be very successful are in the development of a simple two-site model for liquid hydrogen fluoride[16] and to define effective quantum centroid potentials for centroid molecular-dynamics (CMD) simulations.[17]

In principle, the FM method can also be extended to derive realistic CG force fields directly from the underlying atomistic forces. For this purpose, the FM procedure needs to be applied to the trajectory data for the CG sites, i.e., the total forces acting on these sites. To obtain trajectory and force data for a CG image of the system, the trajectories and forces output from an atomistic simulation (not necessarily an *ab initio* simulation but a simulation employing an empirical potential) have to be transformed accordingly. It is convenient to associate CG sites with the centers of mass (COMs) of atomic groups which are subject to coarse graining, because the force acting on a center of mass of an atomic group can be straightforwardly evaluated from the atomistic data. (However, other choices are clearly possible, such as the center of geometry.) The FM procedure then applied to the CG trajectories and forces will in principle yield the effective interaction between CG sites [essentially an approximate many-dimensional potential of mean force (PMF)] as is present in the underlying atomistic simulation. It should be noted that it is not necessarily obvious that the FM approach as originally developed for fitting *ab initio* force fields should work for defining effective CG force fields given the possibly significant entropic component of the latter. As a preliminary step, however, this scheme was successfully employed to develop a CG model of a solvated lipid bilayer.[13] In the present work, we more systematically explore the FM coarse-graining approach by focusing on its application to "simple" liquids such as water or methanol. The feature that makes our approach markedly different from all previous methods for the development of CG force fields is that it derives the effective force field directly from the underlying atomistic potentials. Hence we have called this methodology the "multiscale coarse-grained" (MS-CG) approach.

The main objective of the present work is thus the demonstration of the abilities and universality of the new FM method in the development of CG models. Additionally, as a practical matter the availability of good CG models for certain liquids such as water is especially desirable. For example, any effort to improve the computational efficiency in the simulation of bimolecular systems must be initiated by an attempt to get a better CG model for water, since the dominant computational cost in such simulations generally comes from the large number of force evaluations for explicit water solvation. A CG water model should also reproduce basic structural, dynamic, and thermodynamic properties of pure water in reasonable agreement to experiment and at a low computational cost. These requirements have therefore dictated water as being our primary focus in this paper, so we have applied the MS-CG method to develop one- and two-site CG water models. The coarse graining into two-site representation was accomplished with and without assigning of partial charges to the CG sites. The second application described later is to CG methanol as an example of a more complex molecular fluid. As such, one- and two-site MS-CG methanol models are constructed and evaluated.

This article is organized as follows: In Sec. II we describe our approach in detail, then Sec. III reports applications of the method to derive MS-CG models for water and methanol. Section III starts with a presentation of general details of the FM fitting procedure from the underlying MD simulations. The various properties from the MD simulations employing the new MS-CG models are also analyzed in the respective subsections of Sec. III. In Sec. IV, conclusions from these results are presented.

II. METHOD

An original version of the FM method was proposed by Ercolessi and Adams[18] to fit potential parameters directly to force and atomic coordinate data obtained from *ab initio* simulations. In their scheme the classical forces of a preselected analytical form, which are dependent on the set of M parameters $g_1,\ldots,g_M$, are optimized by trying to match the

forces supplied by first-principles calculations for a large set of different configurations. The match is achieved in a least-squares sense by minimizing directly the objective function equivalent to

$$\chi^2 = \frac{1}{3LN}\sum_{l=1}^{L}\sum_{i=1}^{N}\left|\mathbf{F}_{il}^{\mathrm{ref}} - \mathbf{F}_{il}^{\mathrm{p}}(g_1,\ldots,g_M)\right|^2, \tag{1}$$

where the $\mathbf{F}_{il}^{\mathrm{ref}}, \mathbf{F}_{il}^{\mathrm{p}}$ are, respectively, a reference *ab initio* force and a force predicted using the analytical form which act on the ith atom in the lth atomic configuration; N is the number of atoms in the atomic configuration (we assume the same number of atoms in all configurations); and L is the total number of atomic configurations used in the fit. This method has been successfully applied to derive potentials for certain compounds.[20] However, in such a form, the FM procedure encounters difficulties as the number of fitting parameters grows, especially as a result of the variety of interactions involved, for example, due to a large number of species. Fitting of the potential parameters to the whole coordinate and force database simultaneously limits its accuracy and makes it difficult to use large sets of atomic configurations. The resulting potentials (supplied by *ab initio* calculations) do not often show satisfactory performance in the reproduction of macroscopic properties. As a possible remedy, a number of constraints aimed at fixing physical quantities to desired values are often added to Eq. (1). An important observation is that a least-squares approximation to a sequence of data is simply their averaged value (mathematical expectation), so it would be much more accurate and computationally efficient to replace (or combine) the minimization of the functional equation (1) with the averaging scheme. Furthermore, the averaging for the purpose of coarse graining helps us to pick up the configurational entropy that should have a contribution to the PMF between CG sites.

Our FM method described below includes *explicit configurational averaging* that makes the methodology more accurate and consistent with the philosophy of reconstruction of an effective (or mean) CG force field. The most efficient way to implement configurational averaging is to assume a linear dependence of the force field on its parameters. Such a property can be achieved by using proper (e.g., spline) interpolation. Furthermore, if a force field which is subject to fit depends linearly on the fitting parameters, the least-squares fit in Eq. (1) can be reduced to the solution of *an overdetermined system of linear equations*. Indeed, it can be shown that if the $\mathbf{F}_{il}^{\mathrm{p}}(g_1,\ldots,g_M)$ force is a linear function of its M parameters $\{g_j\}$, then the minimization of the χ^2 in Eq. (1) with respect to a vector of parameters $\{g_j\}$ leads to a system of equations with respect to the parameters g_j. This can in turn be written in a matrix notation as

$$\left\|(\mathbf{F}_{il}^{p})'_{g_j}\right\|^T\left\|(\mathbf{F}_{il}^{p})'_{g_j}\right\|\{g_j\} = \left\|(\mathbf{F}_{il}^{p})'_{g_j}\right\|^T\mathbf{F}_{il}^{\mathrm{ref}},$$
$$i = 1,\ldots,N,\ l = 1,\ldots,L, \tag{2}$$

where the matrix indices are j and the index numbering components of the vector $\mathbf{F}$. Equations (2) are linear equations with respect to the set of parameters $\{g_j\}$ and called the normal equations of the least-squares problem in Eq. (1). These equations can be reduced to

$$\mathbf{F}_{il}^{p}(g_1,\ldots,g_M) = \mathbf{F}_{il}^{\mathrm{ref}}, \quad i = 1,\ldots,N, \quad l = 1,\ldots,L, \tag{3}$$

which is an overdetermined system of linear equations if $M < N \times L$ (N is the total number of atoms in the atomic configuration and L is the total number of atomic configurations used in the fit). Such a system can be built up by equalizing the reference forces and the predicted forces acting on the same atoms in the configurations sampled along the trajectories obtained from a reference MD simulation. By sampling a large enough number L of the atomic configurations it is clearly possible to make the linear system overdetermined. A unique solution of an overdetermined system in the least-squares sense can often be found using algorithms such as QR or singular value decomposition.[21]

Next, the adaptation of the method used to fit a pairwise central (i.e., dependent only on the atom-atom separation) force field to trajectory and force data from a reference MD simulation is outlined. The atom-atom force $\mathbf{f}_i^p(r_{ij})$ (acting from particle j on particle i) can be partitioned into a short-ranged part and a long-ranged Coulomb part as

$$\mathbf{f}_i^p(r_{ij}) = -\left(f(r_{ij}) + \frac{q_i q_j}{r_{ij}^2}\right)\mathbf{n}_{ij}, \tag{4}$$

where r_{ij} is the modulus of the vector $\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i$ connecting two atoms, q_i is the partial atomic charge (subject to fit), and $\mathbf{n}_{ij} = \mathbf{r}_{ij}/r_{ij}$. The short-ranged term $f(r)$ is represented by third-order polynomials (cubic splines)[22] connecting a set of points $\{r_k\}$ (which meshes the interatomic separation up to cutoff radius $r_{k_{\max}}$), thus preserving continuity of its functions and their first two derivatives across the junction, such that

$$f(r,\{r_k\},\{f_k\},\{f_k''\}) = A(r,\{r_k\})f_i + B(r,\{r_k\})f_{i+1} + C(r,\{r_k\})f_i'' + D(r,\{r_k\})f_{i+1}'',$$
$$r \in [r_i, r_{i+1}], \tag{5}$$

where A, B, C and D are known functions of $r, \{r_k\}$, and $\{f_k\}, \{f_k''\}$ are tabulations of $f(r)$ and its second derivative on a radial mesh $\{r_k\}$. A key property for the success of the method is that a spline representation depends linearly on its parameters $\{f_k, f_k''\}$. The parameters $\{f_k, f_k''\}$ and $q_{ij} = q_i q_j$ are to thus be obtained from the fit.

Equalization of known reference forces $\mathbf{F}_{il}^{\mathrm{ref}}$ to forces predicted using the representation in Eqs. (4) and (5), which act on the same atoms in the l th atomic configuration sampled along the reference trajectories, results in the following linear equations:

$$-\sum_{\gamma=nb,b}\sum_{\beta=1,K}\sum_{j=1,N_\beta}\left(f(r_{\alpha il,\beta jl},\{r_{\alpha\beta,\gamma,k}\},\{f_{\alpha\beta,\gamma,k}\},\{f''_{\alpha\beta,\gamma,k}\})+\frac{q_{\alpha\beta}}{r^2_{\alpha il,\beta jl}}\delta_{\gamma,nb}\right)\mathbf{n}_{\alpha il,\beta jl}=\mathbf{F}^{\mathrm{ref}}_{\alpha il},\quad \alpha=1,\ldots,K,\quad i=1,\ldots,N_\alpha \tag{6}$$

with respect to $\{f_{\alpha\beta,\gamma,k}, f''_{\alpha\beta,\gamma,k}, q_{\alpha\beta}\}$, which are force parameters subject to a fit. In Eqs. (6), $\{\alpha il\}$ labels the ith atom of kind α in the lth atomic configuration; $r_{\alpha il,\beta jl}$ is the distance between atoms $\{\alpha i\}$ and $\{\beta j\}$ in the lth atomic configuration; $q_{\alpha\beta}$ is a product of partial charges q_α, q_β of atoms of kinds α and β; and N_β and K are, respectively, the number of atoms of kind β and the total number of atomic kinds in the system.

The method also permits systematic separation of bonded and nonbonded forces. This is important for the sites which have an overlap in regions of intra- and intermolecular motions, causing the FM force field in these regions to be a mixture of nonbonded and bonded components, which is otherwise impossible to separate. To fit explicitly the bonded forces the additional index $\gamma=\{nb,b\}$, which indicates whether the site pair $\alpha il,\beta jl$ is bonded (i.e., $\gamma=b$), is introduced to distinguish the f, f'' parameters in Eqs. (6) for bonded and nonbonded interactions. For bonded pairs the Coulomb term is absent, which is enforced in Eqs. (6) by the $\delta_{\gamma,nb}$ which is unity if $\gamma=nb$ and zero otherwise. It is permissible to choose the spline mesh $\{r_{\alpha\beta,\gamma,k}\}$ to be different for different pairs $\{\alpha\beta\}$ of atomic kinds. Equations (6) can be built up equally well by equalizing only projections of the forces onto preselected direction(s), and these equations for L different atomic configurations (i.e., $l=1,\ldots,L$) can be further combined. If L is large enough these equations can form an overdetermined system (of linear equations) with respect to the fitting parameters, so we will assume that in Eqs. (6) the index l runs over a sufficiently large number L of the atomic configurations such that the equations overdetermine the force parameters. Standard equations which are linear with respect to $\{f_{\alpha\beta,\gamma,k}, f''_{\alpha\beta,\gamma,k}\}$ must also be included into Eqs. (6) to ensure that the $f(r)$'s first derivative $f'(r)$ is continuous across the boundary between two intervals.[22]

The solution of an overdetermined system in a least-squares sense, which is defined as a vector of unknowns minimizing the Euclidean norm of vectors of residuals, is usually found using a QR algorithm or a singular value decomposition.[21] A least-squares solution is unique if a matrix of the system has a full rank or, equivalently, the number of columns in Eqs. (6) is equal to the number of unknowns. The system in Eqs. (6) has a full rank if the sets of atomic configurations used to build the system sample all grid bins in the meshes of the interatomic distances and, therefore, the term corresponding to any grid point index is explicitly present in Eqs. (6). Consequently, having the system in Eqs. (6) to be of full rank can be achieved by choosing more grain meshing or by selecting the atomic configurations which represent a more extensive sampling of the interatomic separations or by both. A finer meshing should be performed in those regions of the interatomic distances in which intramolecular motions occur. The site partial charges q_α can be readily recovered by solving the system of (nonlinear) equations

$$q_\alpha q_\beta=\langle q_{\alpha\beta}\rangle, \tag{7}$$

where $\langle q_{\alpha\beta}\rangle$ are parameters obtained from Eqs. (6) and then averaged over trajectories.

Some additional constraints on the force parameters may be added to Eqs. (6). For example, the atomic partial charges found from Eqs. (6) and (7) normally do not satisfy charge conservation conditions such as electrostatic neutrality if the species/molecules in the system are believed to have a zero total charge. This happens due to an inherent mixing between terms representing the short-ranged and Coulomb interactions. For smaller systems, such a mixing manifests itself more strongly. To avoid phantom charge distributions on the atoms, additional constraints aimed at fixing charges on the species to desired values can be added. If the total charge on a given species is zero, a corresponding constraint can be reduced to set of linear equations with respect to $q_{\alpha\beta}$ as

$$\sum_{\beta\in\Gamma} n_\beta q_{\alpha\beta}=0,\quad \alpha\in\Gamma, \tag{8}$$

where $\Gamma=(\gamma_1,\gamma_2,\ldots,\gamma_n)$ is the set of labels for all kinds of atoms found in the species, and n_β is the number of atoms of kind β in that species. The short-ranged term and optionally its first two derivatives can be constrained to zero at the cutoff radius (that is, the most remote point $r_{\alpha\beta,\max[k]}$ in a respective spline mesh). The respective constraints are linear functions of the force parameters and can be added separately to Eqs. (6). However, normally if the grid extends to sufficiently large distances, the short-ranged term goes to zero smoothly at the cutoff radius without the use of constraint.

The method described above is general and does not rely on the source of the trajectory and force data. Importantly, because of the averaging inherent in our FM methodology, the method can be straightforwardly applied to derive an effective force field for coarse-grained models. For this purpose the atomistic trajectories and forces have to be reduced to the trajectories and forces for the structural units subject to coarse graining. The most natural way to coarse grain a set of atomic groups is to place the CG site at their COM as the force acting on the CG site can be straightforwardly evaluated from the atomistic data. Consequently, the FM algorithm can be applied to the underlying atomistic trajectories as to any other trajectories from fully atomistic simulations. The force field matched to the CG trajectories and forces is then an approximation to the effective interaction, or potential of mean force, between the CG interaction sites. As an alternative the CG site can coincide with the geometrical center of the structure (i.e., the COM of the system assuming that all atoms are of the same mass). Because the FM algorithm does not have masses of interaction sites as an input and the instantaneous interatomic forces are mass independent, the CG force from atomistic data can be evaluated as if the geometrical center is the COM of the structure. Of course, the FM force field will be dependent on where the CG sites are placed.

As shown below such CG force fields perform well in the reproduction of the bulk phase structural properties; however, it sometimes fails to maintain the proper internal pressure in the system and as a result the density is also wrong (too low) in constant pressure CG MD simulations. This behavior is attributed to two primary sources which can be understood from the virial equations used to evaluate the pressure in MD simulation, i.e.,

$$P = \left(\tfrac{2}{3}\langle E^{\mathrm{kin}}\rangle + \langle W\rangle\right)/V, \tag{9}$$

where

$$\langle E^{\mathrm{kin}}\rangle = Nk_BT/2 \tag{10}$$

is the average kinetic energy and

$$\langle W\rangle = \left\langle \frac{1}{3}\sum_{i<j} \mathbf{f}_{ij}\cdot\mathbf{r}_{ij} \right\rangle \tag{11}$$

is the virial of the system. In Eq. (10) N is the number of system degrees of freedom and T is the system temperature. Because the coarse graining of the system eliminates some degrees of freedom as seen from Eq. (10), the first term in Eq. (9) is lower compared to atomistic simulation, which in turn lowers the pressure P. The second reason for an inaccurate density from CG models is related to the second (virial) term in Eq. (9). The coarse graining of the system does not preserve the virial term. In particular, it contracts the system virial of contributions from forces 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 field can also be constrained to produce the correct pressure.* This task can be accomplished by adding to Eqs. (6) the constraint

$$\sum_{\gamma=nb,b}\sum_{\substack{\alpha=1,K\\ \beta=1,K}}\sum_{\substack{i=1,N_\alpha\\ j=1,N_\beta}}\left(f(r_{\alpha il,\beta jl},\{r_{\alpha\beta,\gamma,k}\},\{f_{\alpha\beta,\gamma,k}\},\{f''_{\alpha\beta,\gamma,k}\})r_{\alpha il,\beta jl}+\frac{q_{\alpha\beta}}{r_{\alpha il,\beta jl}}\delta_{\gamma,nb}\right)=3W_l^{\mathrm{atm}}+2\Delta E_l^{\mathrm{kin}}, \tag{12}$$

where the W_l^{atm} is the "instantaneous" virial for the lth configuration along atomistic trajectories and

$$\Delta E_l^{\mathrm{kin}} = E_l^{\mathrm{kin,atm}} - E_l^{\mathrm{kin,CG}}. \tag{13}$$

In Eq. (13), $E_l^{\mathrm{kin,atm}}$ and $E^{\mathrm{kin,CG}}$ are instantaneous kinetic energies for the lth configuration along atomistic trajectories and its CG image, respectively. The difference ΔE_{kin} which accounts for the change in pressure due to the reduction in system degrees of freedom can be calculated, somewhat less accurately, as

$$\Delta E^{\mathrm{kin}} = E_l^{\mathrm{kin,atm}}(1 - N^{\mathrm{CG}}/N^{\mathrm{atm}}), \tag{14}$$

where $N^{\mathrm{atm}}, N^{\mathrm{CG}}$ are the numbers of degrees of freedom of atomistic and CG systems, respectively.

The fitted data in Eq. (12) depend explicitly on the instantaneous kinetic energy and therefore on the temperature in the reference atomistic simulation. This may further reduce the transferability of the present MS-CG models to other temperatures. In particular, thermodynamic properties which rely on the derivatives of the temperature (e.g., thermal-expansion coefficient) may be less accurate. However, as will be shown below from a comparison of FM models for water and methanol, the transferability also depends on the compound and reference atomistic simulation data. Furthermore because the method is general and efficient, new MS-CG force fields can readily be obtained for different thermodynamic conditions.

As was mentioned earlier, in the approach by Ercolessi and Adams the potential parameters are fitted to configuration and force data in a least-squares sense at once. In our approach the linear dependence of the force field on its parameters allows one to carry out the fit repeatedly in an average manner which is more consistent with the philosophy of finding an effective CG force field. In this case Eqs. (6) have to be solved for different smaller sets of atomic configurations well sampled along the reference trajectories, and then the solutions so obtained *should be averaged over a large number of such sets.* If a certain spline parameter is not explicitly present in Eqs. (6) because the respective grid space is not sampled by a particular set of atomic configurations and, therefore, this parameter is not possible to determine, it can be safely assumed to be zero. A subsequent averaging of this spline parameter has to be performed over only those sets of atomic configurations for which the parameter can be evaluated [by solving Eqs. (6)].

It seems evident that the approach outlined above is indeed a way to get a pairwise mean CG force field. It is also instructive to give the following interpretation to this method. A reference (total) force acting on each particle in each configuration along the reference trajectories can always be decomposed into a sum of pairwise forces acting on that particle. Obviously, such a decomposition is not unique for a single configuration because the number of atoms in a single set is less than a number of distinct atomic pairs. By adding different configurations picked up along the reference tajectories, it is possible to determine such a decompostion, $F_{ij}(r_{ij})$ uniquely. This part of the task is accomplished by solving Eqs. (6). Subsequent averaging, $\langle\cdots\rangle$, of these so-

obtained decompositions over the reference trajectories results in a mean (or effective) force field, $\langle F_{ij}(r_{ij})\rangle$. As discussed above this force field approximates all reference forces utilized in the fit in a least-squares sense.

If many-body effects (e.g., due to polarizability of interaction sites) or effects arising from noncentrality of intersite interactions are important in the underlying simulated atomistic system, the atomic configurations used to build each set of Eqs. (6) should not be separated too far in time in order to minimize the difference in the effective pairwise interactions among these sets. This is because the solution of Eqs. (6) approximates the interactions in the utilized sets equally well (in a least-squares sense).

The FM method is also well suited for designing all-atom models having a much-simplified description of interactions (e.g., by a simplified treatment of effective Coulomb interaction using a cutoff). This can also be considered a kind of coarse graining as, for example, the use of cutoff to long-ranged electrostatics in the simulation of polar liquid is equivalent to a treatment of remote layers of molecules as a continuous neutral media. An application of the FM procedure using the representation for the force field in Eq. (4), which has the Coulomb term removed, results in a short-ranged parametrization which is automatically optimized to compensate for the lack of explicit long-ranged electrostatics.

Numerous previously developed CG approaches have targeted the CG force parametrization by fitting a few preselected macroscopic properties. Our method is qualitatively different by mapping the underlying atom-atom interactions into effective interactions between molecular fragments designated as the CG interaction sites. The CG force field in our method therefore relies only on the underlying atomistic interaction for its development, and hence this is why it has been called "multiscale" coarse graining.

A coarse-grained force field is in principle a representation of the PMF between the CG sites, which formally arises after integrating out the thermal fluctuations of the degrees of freedom which are eliminated in the coarse graining of the condensed-phase molecular system. For example, in the MS-CG approach the force-matching procedure yields directly a site-site pairwise decomposable representation of the derivatives of the PMF between the CG sites. The integral of these forces in turn yields an approximate site-site pairwise PMF so, to the extent this is an accurate representation of the real underlying CG PMF, running MD trajectories on the MS-CG potential will yield good equilibrium (structural and thermodynamic) properties. Indeed if the MS-CG (or any CG) methodology reproduces the exact PMF for the CG sites, then the resulting thermodynamic properties calculated from the CG MD simulation, if such properties can be expressed formally as an average over the CG site configurations, must also be exact or nearly so to within the statistical error. The same *is not true* for the dynamics, however. The running of Newtonian (or thermostated Newtonian) MD trajectories on a CG potential is equivalent to the running trajectories on the PMF for the CG sites. This "dynamics" is no longer a real dynamics, meaning that the variable t in the CG dynamics is not the real time. For example, it is common (see examples later) for the self-diffusion of the CG molecules to be significantly faster than the underlying atomistic-level MD. This behavior arises from the fact that degrees of freedom have been removed from the atomistic system upon coarse graining, effectively reducing its phase-space volume, which is in turn usually explored in a more rapid fashion.

One simple concept to deal with the issue of the incorrect CG dynamics is to rescale the time variable to create a "CG time" which equates it to an equivalent atomistic MD time. This rescaling is done based on the difference between the self-diffusion rate in the CG MD and atomistic MD simulations. Such an approach makes some sense for a homogeneous system (e.g., a fluid composed of molecules of one kind), but in more mixed or heterogeneous systems it begins to be conceptually problematical. The real solution to the CG dynamics problems lies in the foundations of nonequilibrium statistical mechanics and, in particular, the second fluctuation-dissipation theorem. Within that framework, it is known that the dynamics of a reduced system, after integrating out the other degrees of freedom, should be described by a set of (possibly nonlinear) generalized Langevin-type equations. Such equations describe dynamics on the PMF for the reduced degrees of freedom, but they *also* contain frictional and random force terms which build back in the missing dynamics from the degrees of freedom that were eliminated from the problem. Thus, for the CG dynamics to be correct (i.e., to give the real dynamics of the CG system), the CG sites should evolve under non-Newtonian generalized Langevin-type equations. Unfortunately, the determination of the proper friction kernels and random force terms in such equations for the CG sites from the underlying atomistic MD data would be challenging to say the least. However, the force-matching approach inherent in the present MS-CG methodology may offer an advantage in thet regard and will be the subject of a future publication.

III. RESULTS AND DISCUSSIONS

A. General details of FM procedures and MD simulations

To derive effective CG force fields, the trajectories and forces yielded by the simulations using empirical potentials for water and methanol were coarse grained, as discussed in Sec. II (the COM was used for the CG sites). The virial constraint was also employed to ensure correct pressure and related thermodynamic properties. The resulting spline representations were then refitted in a least-squares sense using an analytical expansion

$$f(r) = \sum_{n=2}^{n_{\max}} A_{ij}^{n}/\mathrm{r}^{n}, \tag{15}$$

where $n_{\max}$ was separately selected for each model to adequately represent an original spline fit using the simulated properties as criteria. The analytical fits for the models studied here are presented and can easily be used to program the interactions.

All compounds considered here consist of polar molecules. For systems having charged atoms the FM method permits fitting partial atomic charges [see Eqs. (6) and (7)]. On the other hand, a fit can be equally well carried out with the Coulomb term absent in Eq. (4). In this case the short-ranged term $f(r)$ yielded by the FM procedure will contain a portion of the electrostatic interaction within the applied r_{cut}. Such a model is not fully equivalent to a model generated with the Coulomb term explicitly present, which is then truncated at the r_{cut}. This is because in a model parametrized without a Coulomb term, the Coulomb interaction, which is still implicitly present in the $f(r)$, is enforced to go to zero smoothly at r_{cut} as a result of the boundary condition of zero imposed on the $f(r)$ at the r_{cut}. In the present work, the relative performance of two parametrizations, which differ by whether the partial charges were explicitly fitted, was evaluated for two-site water model.

In the MD simulations with force-match models, the short-ranged potentials which are necessary to determine the simulated energetics and thermodynamics were calculated by integrating out the respective terms in the force fields and then shifting them to zero at the r_{cut}. These manipulations can be viewed as a source of error in the simulated energetics. However, the way in which the potentials are calculated should not influence the properties of the simulated system because in the present framework the short-ranged forces are not derived from the potentials.

The specific details of the atomistic simulations and FM procedure for each compound are given in the respective sections below. The classical MD simulations with the new MS-CG models were performed using the DL_POLY 2.9 computer code.[23] The tabulated force fields and potentials for all models are available in a DL_POLY code format upon request. Alternatively, the polynomial expansions given for each model in the tables below can be used conveniently to program the models. The temperature was weakly coupled to a bath with a relaxation time of 0.1 ps (a Nosé-Hoover thermostat was used), and the pressure in the simulations with the constant *NPT* ensemble was weakly coupled to a barostat with a relaxation time of 0.5 ps. The systems were equilibrated typically for 1 ns. A simple velocity scaling method was used to control the system temperature during the equilibration.[24] Production runs were then 2–3 ns long, with an integration time step of 0.5 fs. The isochoric and isobaric heat capacities C_P and C_V, thermal-expansion coefficient α_P, and thermal compressibility κ_T were determined from the MD simulations using both the standard fluctuation formulas[25] and numerical differentiation.[26]

B. Liquid water

The MS-CG models of water were developed from Car-Parrinello molecular-dynamics (CPMD) simulations using FM procedure. The details of the simulations are given in Ref. 15. It should be noted that the CPMD simulations with the Becke-Lee-Yang-Parr (BLYP) functional[27–29] and the Troullier-Martins pseudopotential[30] were carried out with a fictitious electronic mass which does not give converged CPMD and hence yields accidentally good agreement with experimental data.[31–37] However, for the purpose of developing and testing MS-CG models these simulations are adequate.

The reference CPMD simulation used to fit the CG water models was carried out for 64 H_2O molecules in a supercell using the same approximations and the geometries as in Refs. 15 and 32. To fit the force field parameters, the overdetermined system of linear equations (6) was solved repeatedly for approximately 290 different sets of configurations, with each set consisting of two configurations [$l=1, 2$ in Eqs. (6)] which were equidistant in time and separated by approximately 0.034 ps. This corresponds to about 20.0 ps of the CPMD simulation. Then the resulting solutions were averaged over all sets. More specific details are given in Ref. 15.

For all MS-CG water models the short-ranged force component was represented by a spline over a mesh with a grid space of approximately 0.0025 nm in the region of the intramolecular degrees of freedom and 0.005 nm in the region of the intermolecular motions. The r_{cut} was varied depending on the model. Within a particular model, the same r_{cut} to short-ranged interaction for all pairs of CG sites was used. In the simulations using the two-site CG model the molecular geometry was rigid and taken to be that observed on average in the reference CPMD 64-H_2O simulation.[15,32]

The simulations using the FM models were carried out for a system of 512 molecules at T=300 K. The structural and dynamical properties were reported from simulations under constant *NVT* conditions at a density of 1000 $\mathrm{kg/m^3}$. The corresponding box side length of the supercell was 2.48 344 nm. However, the structure and self-diffusion in constant *NVT* and *NPT* simulations were virtually identical.

1. One-site CG water model

In this very "aggressively" coarse-grained model, the water molecule was represented by one spherically symmetrical CG site having a mass of the H_2O molecule which is labeled by "O." The effective force field between CG sites was obtained using trajectories of molecular COMs furnished by the reference all-atom 64-H_2O MD simulation. The cutoff of 0.59 nm was applied to the force field represented by only the short-ranged term in Eq. (4). The CG force field was obtained as a solution of Eqs. (6) and virial constraints equations (12). The resulting parametrization, which is referred to as the $\mathrm{FF_{one\text{-}site}}(64,0.59)$ model (we use a notation similar to that in Ref. 15), is given in Table I. The force and corresponding potential are plotted in Fig. 1. Qualitatively, but not quantitatively, the potential is similar to that obtained using an inverse MC method to match the radial distribution function (RDF) and density in Ref. 7.

In Fig. 2 we show a comparison of the radial distribution function g_{OO} obtained from the MD simulation using the $\mathrm{FF_{one\text{-}site}}(64,0.59)$ model with the g_{OO} (which should be very close to RDF of molecular COM) from the all-atom 64-H_2O MD simulation[15] and neutron-diffraction scattering experiments by Soper.[38] A comparison with the all-atom MD simulation and experiment is far better than one might expect.[8,39] Surprisingly, the absence of a tetrahedral molecular structure

TABLE I. Coefficients A^n_{OO} of the least-squares fit of the $FF_{one\text{-}site}(64,0.59)$ CG force field using the expansion in Eq. (15) with $n_{max}=11$. a.u. were used. The cutoff of 0.59 nm must be applied to this expansion.

n	A^n_{OO}
2	−4797.380 894 912
3	321 106.944 415 7
4	−9 435 745.917 272
5	159 805 932.3305
6	−1 719 523 144.815
7	12 193 219 799.32
8	−56 992 670 350.61
9	169 359 294 902.4
10	−290 395 612 488.0
11	218 964 657 251.3

within the framework of the one-site model does not adversely affect the structural properties for the RDF. These properties are *effectively* built into the one-site MS-CG potential through the FM procedure. Indeed, the positions and shapes of the first and second maxima in the RDFs from the $FF_{one\text{-}site}(64,0.59)$ and all-atom MD simulations are virtually identical. Also, the simulated first peak coordination number of 4.0 is the same.

Some liquid properties of the one-site MS-CG water model are given in Table II. The average configuration energy U^{pot} is about two times higher compared to the all-atom FM water models. The heat capacity C_V is substantially underestimated compared to the all-atom force field $FF^{cl}_{fl}(64,0.78)$ from Ref. 15, which reproduces the properties of water most accurately, and to experiment. However, this is not a surprise based on the expected small magnitude of the averaged fluctuations in U^{pot} which define the value of C_V. The thermal-expansion coefficient α_P is significantly higher compared to the all-atom model. This is likely due to the dependence of reference virial data on the temperature as discussed above [see Eq. (10)]. The self-diffusion coefficient is about four times larger compared to the all-atom model and the experiment. The faster dynamics in the CG model is due to the absence of molecular degrees of freedom (as discussed at the end of Sec. II) and a smoother force field compared to all-atom models. The isothermal compressibility κ_T of the one-site CG water is about three times larger compared to the all-atom model (and experiment). This behavior will be discussed later.

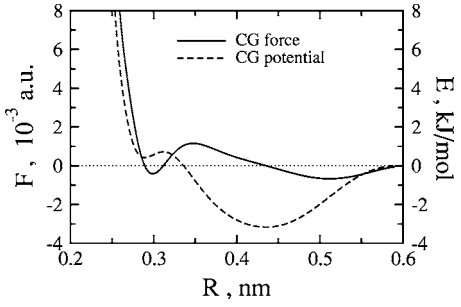

FIG. 1. Effective pairwise force (solid line) and potential (dashed line) between molecular centers of mass in liquid water at ambient conditions as functions of interatomic separation calculated by the force-matching method.

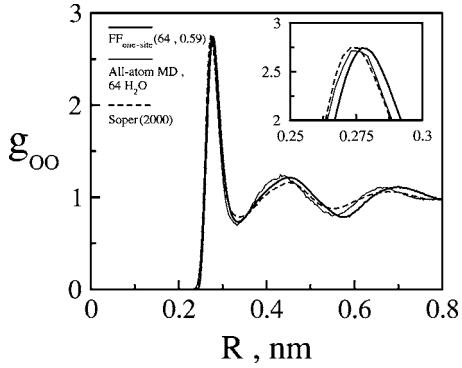

FIG. 2. Site-site RDF from MD simulation using the $FF_{one\text{-}site}(64,0.59)$ model (thick solid line) compared to an oxygen-oxygen RDF from a CPMD simulation of 64-H_2O molecules (thin solid line) and to neutron-diffraction scattering data by Soper (Ref. 38) (dashed line). The inset shows an enlarged region of the first peak tip.

2. Two-site CG water model

The dielectric permittivity of water is a key factor for the simulations of systems involving polar solutes in water or solvated biomolecules which have charged groups. For these applications an absence of the molecular dipole moment in the one-site CG water model may represent a serious drawback. A molecular dipole moment can be modeled by CG models having at least two interaction sites to which partial charges are assigned. In the two-site CG water model presented here, the CG sites were associated with the oxygen and midpoint of HH bond (i.e., center of mass of the HH group). The latter site is denoted as "H_{CG}."

We have tested two model parametrizations. Both models were parametrized using the same $r_{cut}=0.78$ nm to the short-ranged term. The first, which is referred to as the $FF^{cl}_{two\text{-}site}(64,0.78)$ model, was obtained with the explicit treatment of the Coulomb term in Eq. (4) and, thus, has partial charges on its interaction sites. The partial charges are −0.70081 a.u. on the oxygen site and of equal magnitude and opposite sign on the H_{CG} site. The second parametrization, which does not have a long-ranged Coulomb term, is referred to as the $FF^{sr}_{two\text{-}site}(64,0.78)$ model. Surprisingly, the $FF^{cl}_{two\text{-}site}(64,0.78)$ and $FF^{sr}_{two\text{-}site}(64,0.78)$ models show virtually identical qualities in the liquid water simulations. In Table III we present the polynomial expansion coefficients in Eq. (15) for the $FF^{sr}_{two\text{-}site}(64,0.78)$ force field which is plotted in Fig. 3. The bonded part of the $FF^{cl/sr}_{two\text{-}site}(64,0.78)$ force

TABLE II. Properties of different FM water models. Shown are ρ density, U^{pot} average configuration energy, C_V isochoric heat capacity, α_P thermal expansion coefficient, κ_T isothermal compressibility, and D_{diff} self-diffusion coefficient.

Property	$FF_{one\text{-}site}(64,0.78)$	$FF^{cl/sr}_{two\text{-}site}(64,0.78)$	All-atom[a]	Expt.[b]
ρ (kg/m^3)	997	999	1003	997.05
U^{pot} (kJ/mol)	−17.9	−23.3	−32.30	−41.5
C_V (J/[mol K])	21.0	34.0	71.0	74.54
α_P (10^{-5} 1/K)	25.0	17.6	5.7	2.53
κ_T (10^{-5} 1/bar)	14.5	7.8	4.6	4.52
D_{diff} (10^{-9} m^2/s)	9.7	12.0	2.3	2.3

[a]Reference 15, the $FF^{cl}_{fl}(64,0.78)$ potential.
[b]Reference 42 (25 °C).

TABLE III. Coefficients A_{ij}^n of the least-squares fit of the forces of the $FF_{\text{two-site}}^{\text{sr}}(64,0.78)$ model using the expansion in Eq. (15) with $n_{\max}=12$. a.u. were used. At small separations $r<r_{\text{core}}$, the $f_{ij}(r)$ was extrapolated as $f(r<r_{\text{core}})=f(r_{\text{core}})$. The following core radii were used: $r_{\text{core}}^{\text{OO}}=4.8$ a.u., $r_{\text{core}}^{\text{OH}_{\text{CG}}}=4.1$ a.u., and $r_{\text{core}}^{\text{H}_{\text{CG}}\text{H}_{\text{CG}}}=4.7$ a.u. The cutoff of 0.78 nm must be applied to this expansion.

n	A_{OO}^n	A_{OHCG}^n	$A_{\text{H}_{\text{CG}}\text{H}_{\text{CG}}}^n$
2	879.529 357 459 860	30.988 161 797 233 0	60.643 580 780 483 6
3	−70 806.737 842 098 5	−2844.184 015 972 52	−5342.030 213 688 33
4	2 514 598.281 334 69	117 963.347 522 358	205 697.385 493 863
5	−51 915 523.005 594 9	−2 843 111.248 507 82	−4 611 565.846 578 44
6	690 522 038.011 901	43 593 919.860 910 6	67 157 568.707 773 4
7	−6 186 679 895.161 77	−442 751 104.654 784	−666 331 957.688 614
8	37 836 639 738.0748	3 014 505 371.329 83	4 569 034 525.743 29
9	−156 076 292 105.178	−13 595 219 239.1346	−21 377 796 619.6084
10	415 860 302 043.463	38 920 145 856.5204	65 231 595 617.9045
11	−646 720 676 419.486	−63 975 647 948.4449	−116 984 876 345.109
12	446 045 250 387.934	45 941 216 124.4869	93 434 861 444.8085

fields (see Fig. 3) can be used straightforwardly to carry out the simulation with flexible molecules. However, for simplicity, the models used here had a rigid molecular geometry. The O–H_{CG} bond length was constrained to 0.06 nm as observed on average in the reference all-atom MD 64-H_2O simulation. For the $FF_{\text{two-site}}^{\text{cl}}(64,0.78)$ model this geometry gives a molecular dipole moment in the gas phase of 2.02 D. A rigid molecular geometry may give a more structured liquid water in the simulations with FM models.[15]

The RDFs simulated using the (rigid) $FF_{\text{two-site}}^{\text{cl/sr}}(64,0.78)$ models are shown in Fig. 4. The $g_{\text{OH}_{\text{CG}}}$ and $g_{\text{HH}_{\text{CG}}}$ agree closely with those obtained from the reference atomistic simulation up to interatomic separations of about 0.4 nm. The discrepancies, especially those seen in g_{OO}, might be attributed to the diatomic molecular symmetry of the two-site MS-CG model. For such a model, the hydrogen bonding in the liquid phase is expected to differ from tetrahedral coordination because all molecular positive charge is associated with a single interaction site. An increased directionality of interaction in the two-site MS-CG models results in a different hydrogen-bonded network from the fully atomistic MD simulation. Because the two-site MS-CG water "molecule" resembles hydrogen fluoride, it is instructive to compare the liquid H_2O structure arising from the two-site CG approximation with that of the liquid HF. It is known that prevailing aggregates in liquid HF are zigzag chains of hydrogen bonded molecules.[40] This chain formation significantly influences the structure and dynamic properties of liquid HF. A similar type of aggregation may therefore be expected in the liquid phase of the $FF_{\text{two-site}}^{\text{cl/sr}}(64,0.78)$ models. This conjecture is supported if RDFs from the $FF_{\text{two-site}}^{\text{cl/sr}}(64,0.78)$ models and the HF FM model[16] are compared at similar liquid-state conditions. (The development and accuracy of the FM model for liquid HF are discussed in

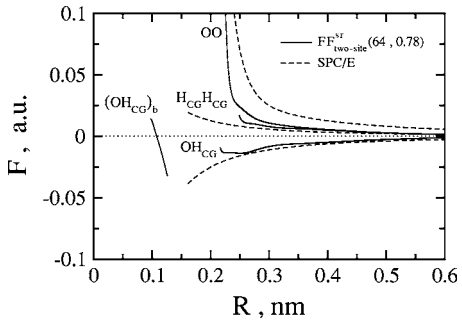

FIG. 3. Effective pairwise forces from the $FF_{\text{two-site}}^{\text{sr}}(64,0.78)$ water model as functions of interatomic separation (solid lines). Index b labels a bonded part of the OH_{CG} interaction. Atom-atom force from the Simple point charge/extended (SPC/E) model (dashed lines) shown for comparison.

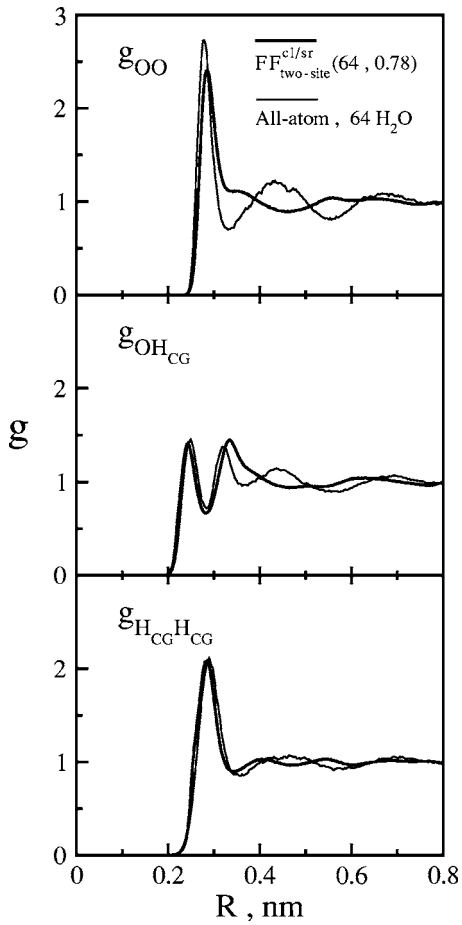

FIG. 4. Intermolecular site-site RDFs from MD simulation using the $FF_{\text{two-site}}^{\text{cl/sr}}(64,0.78)$ water model (thick lines) compared to the reference CPMD simulation (thin lines).

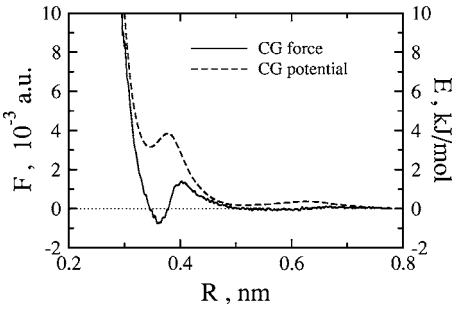

FIG. 5. Effective pairwise force (solid line) and potential (dashed line) between molecular centers of mass in liquid methanol calculated by the force-matching method. Original spline representation was used.

detail in Ref. 16.) A comparison[16] of the g_{FF} to the g_{OO} from the $\mathrm{FF}^{\mathrm{cl/sr}}_{\mathrm{two\text{-}site}}(64,0.78)$ models indeed suggests a similar tendency toward chainlike aggregation. The water self-diffusion coefficient of $12\times10^{-9}\ \mathrm{m}^2/\mathrm{s}$ predicted by the $\mathrm{FF}^{\mathrm{cl/sr}}_{\mathrm{two\text{-}site}}(64,0.78)$ models at ambient conditions is also close to that in liquid HF at similar thermodynamic conditions,[16] further indicating a similarity in the dynamics of both systems. Most importantly, these structural results reflect a key issue in coarse graining, i.e., that *both* the choice of the CG sites *and* their resulting effective interactions are important. The latter can only "fix" so much of a bad choice in the former. In the case of the two site MS-CG water model, the FM procedure, as powerful as it may be, cannot build effective interactions into a model that has the fundamentally wrong symmetry to reproduce the structural features of the tetrahedral hydrogen-bonded network in water. In fact, in some ways the one-site MS-CG model was better at this (at least for the O–O RDF) because its spherical symmetry more accurately reflects, in an approximate sense, the tetrahedral symmetry of the underlying all-atom model.

Some of the liquid properties of the $\mathrm{FF}^{\mathrm{cl/sr}}_{\mathrm{two\text{-}site}}(64,0.78)$ models are presented in Table II. The correct density is a consequence of the use of a virial constraint in the FM procedure. Similar to the one-site model, the coefficient of thermal expansion α_P is much higher than the atomistic counterpart owing in part to the explicit dependence of the virial constraint on the temperature. The isothermal compressibility shows better agreement with the all-atom MD value. The two-site models have an improved (compared to one-site model) but still rather inadequate energetics. Again this is because degrees of freedom that contribute to the energy and its fluctuations have been removed from the system, so this is not unexpected. The large discrepancy in the CG values of the coefficients of thermal expansion and isothermal compressibility compared to the experimental values points to the fact that for new thermodynamic conditions (e.g., temperature and pressure) it may well be best to apply the FM procedure to develop a new MS-CG force field for those new conditions. This is of course easily done and, in fact, should prove quite interesting in revealing how new thermodynamic conditions are manifested in the CG force field.

C. Methanol

In this section methanol (CH_3OH) is chosen as a test case for a more complex molecular fluid composed of anisotropic molecules. The MS-CG models for methanol were developed using trajectory and force output from an all-atom MD simulation with the optimized potentials for liquid simulation with all-atom (OPLS-AA) force field.[41] The OPLS-AA methanol model is six-site and flexible. The MD simulation was for a system of 125 molecules at T=300 K and 2 ns long.

TABLE IV. Coefficients A^n of the least-squares fit of the one-site methanol force field using the expansion in Eq. (15) with $n_{\max}=12$. a.u. were used. At small separations $r<r_{\mathrm{core}}$, the $f(r)$ should be extrapolated as $f(r<r_{\mathrm{core}})=f(r_{\mathrm{core}})$. The $r_{\mathrm{core}}=5.2$ a.u. and the cutoff of 0.788 nm must be applied to this expansion.

n	A^n
2	−5053.676 121 11
3	445 020.339 092
4	−17 358 574.4369
5	394 873 773.657
6	−5 800 439 290.45
7	57 486 913 399.8
8	−389 309 601 374
9	1 779 143 276 396
10	−5 252 356 385 275
11	9 048 213 789 256
12	−6 909 941 376 590

The MS-CG models with one- and two-site CG molecules were obtained. The FM procedure followed the procedure for the CG models of water described earlier. In particular, the virial constraint equation (12) was used to fix the pressure and therefore to ensure the correct density in constant *NPT* simulations. Because the methanol force field used in the atomistic simulation accounts for polarizability only through the flexible molecular geometry (it does not include instantaneous electronic polarizability), one might expect that many-body effects are not so important as in the fully *ab initio* simulations used to match the water force fields. This may or may not enhance the quality of the FM fit and remains to be explored in future work.

1. One-site CG methanol model

To develop a one-site MS-CG model, the site-site interaction was represented by only a short-ranged term up to $r_{\mathrm{cut}}=0.788$ nm. The effective force field and corresponding potential yielded by the FM procedure, which was applied to the trajectories and forces for molecular centers of mass, are shown in Fig. 5. Table IV summarizes the polynomial expansion coefficients in Eq. (15), which is the best least-square fit of the spline data. Because a methanol molecule is highly nonspherical one might argue that the one-site model cannot work. Surprisingly, it does, at least for the CG structural data. In Fig. 6 the RDFs of molecular COM from the reference all-atom simulation are compared to the corresponding RDFs from the MS-CG MD simulation. Both functions are hardly distinguishable on the scale of the graph. Some of the properties of the one-site CG model are compiled in Table VI. As one might expect the self-diffusion of the CG molecules is enhanced, about nine times larger compared to the atomistic simulation, and the average configurational energy and con-

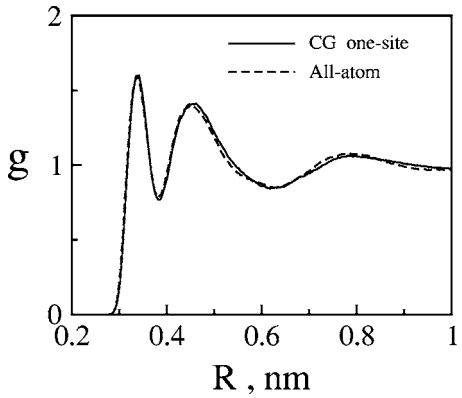

FIG. 6. Site-site RDF from MD simulation at $T=300$ K using the one-site methanol model compared to the RDF of center of mass from the reference fully atomistic simulation.

stant volume heat capacity are underestimated compared to atomistic data for the same reasons as in the case of water. The coefficient of thermal expansion is seen to match more closely to the atomistic MD reference data and experiment. However, similar to water the one-site model exhibits a too high isothermal compressibility.

2. Two-site CG methanol model

The interaction sites in the two-site model were associated with the center of mass of the CH_3 group (CH_3 site) and the center of mass of the OH group (OH site). The CG groups bear charge; however, the force match carried out with the Coulomb term explicitly present in Eq. (4) yielded small values for partial charges on the CG sites. This fact might be explained by a mixing of the short-ranged and Coulomb interactions in the FM procedure due to large dielectric screening effects. For whatever reason the partial charges on the CG sites were small, which indicate that a model with only a short-ranged interaction term can provide a good approximation to the two-site MS-CG force field. The forces for such a model are depicted in Fig. 7, and the polynomial fit [Eq. (15)] is presented in Table V. The FM procedure was conducted with an explicit separation of bonded and nonbonded forces as discussed in Sec. II. However, as seen in Fig. 7 this is not necessary as there is no overlap between

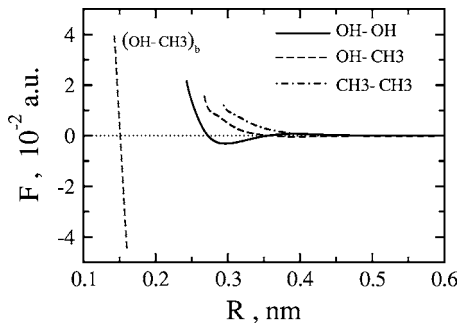

FIG. 7. Effective pairwise forces between interaction OH–OH (solid line), HO–CH_3 (dashed line), and CH_3–CH_3 (dot-dashed line) sites of the two-site methanol model as calculated by the force-matching method.

regions of intra- and intermolecular motions, and the intramolecular interaction can be easily refined on the basis of distance.

The CG site-site RDFs compared with the respective functions from the reference all-atom MD simulation are shown in Fig. 8. The $g_{\mathrm{OH\text{-}OH}}$ and $g_{\mathrm{OH\text{-}CH_3}}$ are in less agreement with the all-atom simulation than the $g_{\mathrm{CH_3\text{-}CH_3}}$. The slight mismatch in $g_{\mathrm{OH\text{-}OH}}$ is likely due to the deviation in cross-site ($g_{\mathrm{OH–CH_3}}$) structure between the CG and all-atom simulations. Recalling that the ability of the two-site CG water model to reproduce structure is less than that of the one-site and all-atom models, one might have expected a tendency for relatively poorer performance of two-site CG models due to an enhanced directionality of interaction in such models. However, the agreement between the structure predicted by the two-site CG methanol model and its all-atom reference structure is very good. The self-diffusion of the two-site CG methanol molecules is about four times larger compared to all-atom simulation, which is not surprising as stated before, given the reduced number of degrees of freedom.

In Fig. 9 the power spectra for the OH and CH_3 sites from all-atom and CG MD simulations are compared. In the atomistic spectrum, in addition to the OH–CH_3 stretching band which is captured by the CG model, another band centered at a higher frequency is manifested. Existence of this band is a consequence of the flexible geometry of the OH and CH_3 groups, which brings additional vibrational modes into the power spectrum of their COMs (i.e., CG sites). The positions of the other two bands agree well between atomistic and CG simulations. The higher intensity of the low-frequency band which corresponds to translational and libration modes in the CG spectrum is obviously due to the faster molecular self-diffusion in MS-CG MD simulation.

Some of the thermodynamic properties of the two-site model are given in Table VI. The properties which rely on the use of virial constraint equation (12), i.e., density, isothermal compressibility, and isothermal expansion, are in satisfactory agreement with their all-atom counterparts. The average configurational energy and C_V, which is derivative from the fluctuations in the configuration energy, are not surprisingly below that of all-atom simulation and experiment. The two-site model performs clearly better in reproducing key thermodynamic properties compared to the one-site model.

Interestingly, the CG methanol models outperform significantly those of water in reproducing the pressure-related properties which are constrained by virial equations in the FM procedure. These results seem to suggest that water is a "special" and challenging (and very interesting) target for coarse graining.

For methanol we tried a CG model having the interaction sites associated with geometrical center of the atomic groups. With such a choice there is overlap between intra- and intermolecular regions that slightly complicates the fit (so the FM procedure with an explicit separation of bonded forces should be used). The resulting model has the same degree of accuracy as for that from COM CG site.

TABLE V. Coefficients A_{ij}^n of the least-squares fit of the forces of the two-site model of methanol using the expansion in Eq. (15) with $n_{\max}=12$ a.u. were used. At small separations $r<r_{\text{core}}$, the $f_{ij}(r)$ should be extrapolated as $f(r<r_{\text{core}})=f(r_{\text{core}})$. The following core radii were used: $r_{\text{core}}^{\text{OH,OH}}=4.5$ a.u., $r_{\text{core}}^{\text{OH,OH}_3}=5.0$ a.u., and $r_{\text{core}}^{\text{OH}_3,\text{OH}_3}=5.6$ a.u. The cutoff of 0.788 nm must be applied to this expansion.

n	$A_{\text{OH,OH}}^n$	$A_{\text{OH,OH3}}^n$	$A_{\text{OH3,OH3}}^n$
2	−940.241 309 480	1029.318 123 17	865.124 021 271
3	77 550.208 251 9	−89 030.702 6047	−71 261.524 180 9
4	−2 823 594.141 20	3 409 298.044 48	2 565 594.186 15
5	59 763 557.4639	−76 125 133.8684	−52 916 296.9269
6	−814 392 492.384	1 097 904 341.97	687 569 721.647
7	7 467 032 182.11	−10 691 276 102.3	−5 813 495 095.62
8	−46 666 967 873.74	71 227 669 400.1	31 714 389 423.5
9	196 393 069 807	−320 772 036 450	−105 268 210 268
10	−532 956 581 491	935 169 258 609	177 490 075 242
11	842 782 132 096	−1 594 832 798 559	−43 481 079 218.9
12	−590 211 718 385	1 208 986 208 065	−196 365 799 673

IV. CONCLUSIONS

In this paper a general and computationally efficient methodology has been presented for the development of effective CG force fields for liquids using the trajectories and forces from all-atom MD simulations. The new force-matching method is markedly different from existing coarse-graining approaches which rely on an inversion of simulated properties rather than the underlying atomistic interactions. The parametrization procedure is systematic and can be applied to a wide range of systems once their atomistic MD simulation data are available. The MS-CG procedure is applied to trajectory and force data after they are appropriately transformed into trajectories and forces for the CG sites. Because the force acting on the center of mass of an atomic group is just the sum of the forces acting on each atom within the groups, it is convenient to associate the CG sites with centers of mass of atomic groups designated for coarse graining. However, other locations (e.g., the geometrical center) for CG sites can instead be used.

The resulting MS-CG method was applied to two examples of polar fluids, water and methanol, which are challenging molecular modeling objects, in order to demonstrate how the method works. The one-site water and one- and two-site MS-CG methanol models are able to capture accurately the liquid structure seen in the underlying all-atom simulations. The virial constraint on the FM procedure also permits the CG models to reproduce pressure-related properties.

The two-site CG water models parametrized assuming either a short-ranged interaction or full electrostatics exhibited less satisfactory agreement to reference all-atom data. This behavior can be explained by an increased directionality of the interactions, leading to directional aggregation behav-

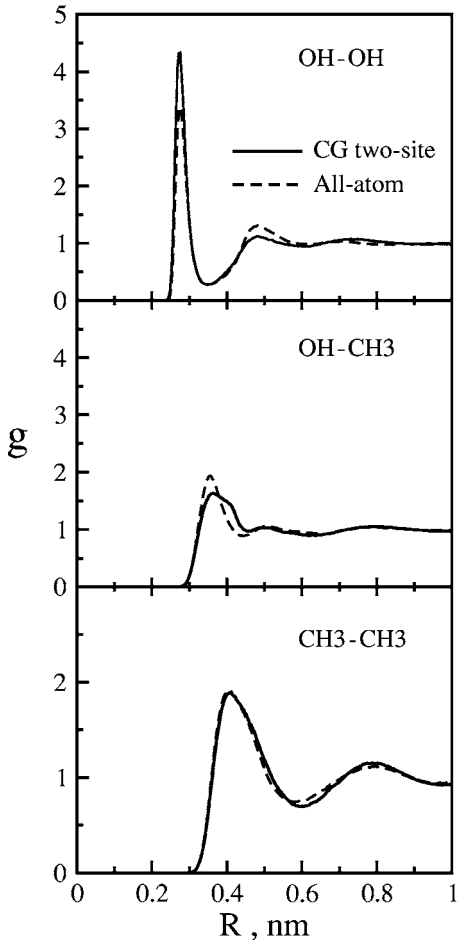

FIG. 8. Intermolecular site-site RDFs from MD simulation at $T=300$ K using the two-site methanol model (solid line) compared to similar RDFs from the reference fully atomistic simulation (dashed line).

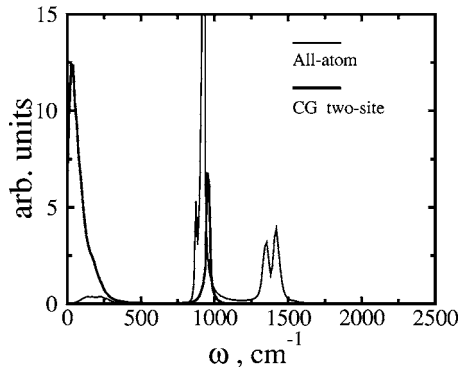

FIG. 9. Power spectra of methanol for CG sites from fully atomistic MD simulation (thin line) and CG MD simulation (thick line).

TABLE VI. Properties of one- and two-site methanol models. Shown are ρ density, U^{pot}, average configuration energy, C_V isochoric heat capacity, α_P thermal-expansion coefficient, κ_T isothermal compressibility, and D_{diff} self-diffusion coefficient.

Property	One-site	Two-site	All-atom	Expt.[a]
ρ (kg/m^3)	766	783	767	785
U^{pot} (kJ/mol)	−8.8	−18.1	⋯	−35.6
C_V [j/(mol K)]	18.6	40.7	65.1	66.41
α_P (10^{-3} 1/K)	2.7	1.7	1.7	1.49
κ_T (10^{-5} 1/bar)	54.7	11.2	12.2	12.14
D_{diff} (10^{-9} m^2/s)	17.2	4.9	2.3	2.4

[a]References 43 and 44.

ior different from the atomistic simulations. These results reflect the importance of the choice of CG sites, which will be the topic of a future publication. However, even here the match in structure from MS-CG and atomistic simulations was still reasonable.

Interestingly, the presence of long-ranged electrostatics is not a feature which makes the properties different for MS-CG water models. The methanol two-site model, which is without explicit treatment of the electrostatics, exhibited only slight differences, which is likely caused by a deviation in the first peak of cross-site CG RDFs from the all-atom one. The nature of the effective electrostatic interactions in liquids will be the topic of a future publication.

The pressure-related thermodynamic properties of MS-CG models of the methanol are in better agreement with the atomistic reference properties compared to the MS-CG water models. This behavior may reflect certain subtle properties of the tetrahedral hydrogen-bonding network in water which will also be explored in the future.

The present applications of multiscale coarse-graining methodology are encouraging and point toward the MS-CG method as a powerful, general, and computationally inexpensive technique for the development of reliable CG models for condensed-phase simulations.

ACKNOWLEDGMENTS

This research was supported by the United States Office of Naval Research and the National Science Foundation. The computations were also supported in part by National Science Foundation (NSF) Cooperative Agreement No. ACI-9619020 for computing resources provided by the National Partnership for an Advanced Computational Infrastructure at the San Diego Supercomputer Center.

[1] B. Smit, K. Esselink, P. A. J. Hilbers, N. M. van Os, L. A. M. Rupert, and I. Szleifer, Langmuir **9**, 9 (1993).
[2] T. Head-Gordon and F. H. Stillinger, J. Chem. Phys. **98**, 3313 (1993).
[3] B. J. Palmer and J. Liu, Langmuir **12**, 746 (1996).
[4] R. Goetz and R. J. Lipowsky, J. Chem. Phys. **108**, 7397 (1998).
[5] J. C. Shelley and M. Y. Shelley, Curr. Opin. Colloid Interface Sci. **5**, 101 (2000).
[6] H. Meyer, O. Biermann, R. Faller, D. Reith, and F. Müller-Plathe, J. Chem. Phys. **113**, 6264 (2000).
[7] S. Garde and H. S. Ashbaugh, J. Chem. Phys. **115**, 977 (2001).
[8] J. C. Shelley, M. Y. Shelley, R. C. Reeder, S. Bandyopadhyay, and M. L. Klein, J. Phys. Chem. B **105**, 4464 (2001).
[9] S. Marrink and A. Mark, J. Am. Chem. Soc. **125**, 15233 (2003).
[10] M. J. Stevens, J. H. Hoh, and T. B. Woolf, Phys. Rev. Lett. **91**, 188102 (2003).
[11] S. J. Marrink, A. H. de Vries, and A. E. Mark, J. Phys. Chem. B **108**, 750 (2004).
[12] T. Murtola, E. Falck, M. Patra, M. Karttunen, and I. Vattulainena, J. Chem. Phys. **121**, 9156 (2004).
[13] S. Izvekov and G. A. Voth, J. Phys. Chem. B **109**, 2469 (2005).
[14] B. Forrest and U. Suter, J. Chem. Phys. **102**, 7256 (1995).
[15] S. Izvekov, M. Parrinello, C. J. Burnham, and G. A. Voth, J. Chem. Phys. **120**, 10896 (2004).
[16] S. Izvekov and G. A. Voth, J. Phys. Chem. B **109**, 6573 (2005).
[17] T. D. Hone, S. Izvekov, and G. A. Voth, J. Chem. Phys. **122**, 54105 (2005).
[18] F. Ercolessi and J. B. Adams, Europhys. Lett. **26**, 583 (1994).
[19] R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).
[20] P. Tangney and S. Scandolo, J. Chem. Phys. **117**, 8898 (2002).
[21] C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems* (Prentice-Hall, Englewood Cliffs, NJ, 1974).
[22] C. De Boor, *Practical Guide to Splines* (Springer, New York, 1978).
[23] T. R. Forester and W. Smith, DL_POLY user manual, CCLRC, Daresbury Laboratory, Daresbury, Warrington, UK, 1995.
[24] D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications* (Academic, San Diego, 1996).
[25] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, Oxford, 1987).
[26] R. Walser, A. E. Mark, W. F. van Gunsteren, M. Lauterbach, and G. Wipff, J. Chem. Phys. **112**, 10450 (2000).
[27] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).
[28] A. D. Becke, Phys. Rev. A **38**, 3098 (1988).
[29] C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).
[30] N. Troullier and J. L. Martins, Phys. Rev. B **43**, 1993 (1991).
[31] P. L. Silvestrelli and M. Parrinello, J. Chem. Phys. **111**, 3572 (1999).
[32] S. Izvekov and G. Voth, J. Chem. Phys. **116**, 10372 (2002).
[33] J. C. Grossman, E. Schwegler, E. W. Draeger, F. Gygi, and G. Galli, J. Chem. Phys. **120**, 300 (2004).
[34] D. Asthagiri, L. R. Pratt, and J. D. Kress, Phys. Rev. E **68**, 041505 (2003).
[35] I.-F. W. Kuo, C. J. Mundy, M. J. McGrath *et al.*, J. Phys. Chem. B **108**, 12990 (2004).
[36] J. VandeVondele, F. Mohamed, M. Krack, J. Hütter, M. Sprik, and M. Parrinello, J. Chem. Phys. **122**, 014515 (2005).
[37] P. H.-L. Sit and N. Marzari, J. Chem. Phys. **122**, 204510 (2005).
[38] A. K. Soper, Chem. Phys. **258**, 121 (2000).
[39] J. L. Finney, J. Mol. Liq. **90**, 303 (2001).
[40] M. W. Johnson, E. Sándor, and E. Arzi, Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem. **B31**, 1998 (1975).
[41] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, J. Am. Chem. Soc. **118**, 11225 (1996).
[42] NIST Chemistry WebBook, NIST Standard Reference Database Number 69, 20899, edited by P. J. Linstrom and W. G. Mallard, NIST, Gaithersburg, MD, 2003 http://webbook.nist.gov
[43] R. D. Goodwin, J. Phys. Chem. Ref. Data Monogr. **16**, 799 (1967).
[44] *CRC Handbook of Chemistry and Physics*, 85th ed., edited by D. R. Lide (CRC, Boca Raton, FL, 2004).