This paper is published as part of a PCCP Themed Issue on:
Coarse-grained modeling of soft condensed matter

Guest Editor: Roland Faller (UC Davis)

Editorial

Coarse-grained modeling of soft condensed matter
Phys. Chem. Chem. Phys., 2009 **DOI:** 10.1039/b903229c

Perspective

Multiscale modeling of emergent materials: biological and soft matter
Teemu Murtola, Alex Bunker, Ilpo Vattulainen, Markus Deserno and Mikko Karttunen, *Phys. Chem. Chem. Phys.*, 2009 **DOI:** 10.1039/b818051b

Communication

Dissipative particle dynamics simulation of quaternary bolaamphiphiles: multi-colour tiling in hexagonal columnar phases
Martin A. Bates and Martin Walker, *Phys. Chem. Chem. Phys.*, 2009 **DOI:** 10.1039/b818926a

Papers

Effective control of the transport coefficients of a coarse-grained liquid and polymer models using the dissipative particle dynamics and Lowe–Andersen equations of motion
Hu-Jun Qian, Chee Chin Liew and Florian Müller-Plathe, *Phys. Chem. Chem. Phys.*, 2009
DOI: 10.1039/b817584e

Adsorption of peptides (A3, Flg, Pd2, Pd4) on gold and palladium surfaces by a coarse-grained Monte Carlo simulation
R. B. Pandey, Hendrik Heinz, Jie Feng, Barry L. Farmer, Joseph M. Slocik, Lawrence F. Drummy and Rajesh R. Naik, *Phys. Chem. Chem. Phys.*, 2009 **DOI:** 10.1039/b816187a

A coarse-graining procedure for polymer melts applied to 1,4-polybutadiene
T. Strauch, L. Yelash and W. Paul, *Phys. Chem. Chem. Phys.*, 2009 **DOI:** 10.1039/b818271j

Anomalous waterlike behavior in spherically-symmetric water models optimized with the relative entropy
Aviel Chaimovich and M. Scott Shell, *Phys. Chem. Chem. Phys.*, 2009 **DOI:** 10.1039/b818512c

Coarse-graining dipolar interactions in simple fluids and polymer solutions: Monte Carlo studies of the phase behavior
B. M. Mognetti, P. Virnau, L. Yelash, W. Paul, K. Binder, M. Müller and L. G. MacDowell, *Phys. Chem. Chem. Phys.*, 2009 **DOI:** 10.1039/b818020m

Beyond amphiphiles: coarse-grained simulations of star-polyphile liquid crystalline assemblies
Jacob Judas Kain Kirkensgaard and Stephen Hyde, *Phys. Chem. Chem. Phys.*, 2009 **DOI:** 10.1039/b818032f

Salt exclusion in charged porous media: a coarse-graining strategy in the case of montmorillonite clays
Marie Jardat, Jean-François Dufrêche, Virginie Marry, Benjamin Rotenberg and Pierre Turq, *Phys. Chem. Chem. Phys.*, 2009 **DOI:** 10.1039/b818055e

Improved simulations of lattice peptide adsorption
Adam D. Swetnam and Michael P. Allen, *Phys. Chem. Chem. Phys.*, 2009 **DOI:** 10.1039/b818067a

Curvature effects on lipid packing and dynamics in liposomes revealed by coarse grained molecular dynamics simulations
H. Jelger Risselada and Siewert J. Marrink, *Phys. Chem. Chem. Phys.*, 2009 **DOI:** 10.1039/b818782g

Self-assembling dipeptides: conformational sampling in solvent-free coarse-grained simulation
Alessandra Villa, Christine Peter and Nico F. A. van der Vegt, *Phys. Chem. Chem. Phys.*, 2009 **DOI:** 10.1039/b818144f

Self-assembling dipeptides: including solvent degrees of freedom in a coarse-grained model
Alessandra Villa, Nico F. A. van der Vegt and Christine Peter, *Phys. Chem. Chem. Phys.*, 2009 **DOI:** 10.1039/b818146m

Computing free energies of interfaces in self-assembling systems
Marcus Müller, Kostas Ch. Daoulas and Yuki Norizoe, *Phys. Chem. Chem. Phys.*, 2009 **DOI:** 10.1039/b818111j

Anomalous ductility in thermoset/thermoplastic polymer alloys
Debashish Mukherji and Cameron F. Abrams, *Phys. Chem. Chem. Phys.*, 2009 **DOI:** 10.1039/b818039c

A coarse-grained simulation study of mesophase formation in a series of rod–coil multiblock copolymers
Juho S. Lintuvuori and Mark R. Wilson, *Phys. Chem. Chem. Phys.*, 2009 **DOI:** 10.1039/b818616b

Simulations of rigid bodies in an angle-axis framework
Dwaipayan Chakrabarti and David J. Wales, *Phys. Chem. Chem. Phys.*, 2009 **DOI:** 10.1039/b818054g

Effective force coarse-graining
Yanting Wang, W. G. Noid, Pu Liu and Gregory A. Voth, *Phys. Chem. Chem. Phys.*, 2009 **DOI:** 10.1039/b819182d

Backmapping coarse-grained polymer models under sheared nonequilibrium conditions
Xiaoyu Chen, Paola Carbone, Giuseppe Santangelo, Andrea Di Matteo, Giuseppe Milano and Florian Müller-Plathe, *Phys. Chem. Chem. Phys.*, 2009 **DOI:** 10.1039/b817895j

Energy landscapes for shells assembled from pentagonal and hexagonal pyramids
Szilard N. Fejer, Tim R. James, Javier Hernández-Rojas and David J. Wales, *Phys. Chem. Chem. Phys.*, 2009
DOI: 10.1039/b818062h

Molecular structure and phase behaviour of hairy-rod polymers
David L. Cheung and Alessandro Troisi, *Phys. Chem. Chem. Phys.*, 2009 **DOI:** 10.1039/b818428c

Molecular dynamics study of the effect of cholesterol on the properties of lipid monolayers at low surface tensions
Cameron Laing, Svetlana Baoukina and D. Peter Tieleman, *Phys. Chem. Chem. Phys.*, 2009 **DOI:** 10.1039/b819767a

On using a too large integration time step in molecular dynamics simulations of coarse-grained molecular models
Moritz Winger, Daniel Trzesniak, Riccardo Baron and Wilfred F. van Gunsteren, *Phys. Chem. Chem. Phys.*, 2009 **DOI:** 10.1039/b818713d

The influence of polymer architecture on the assembly of poly(ethylene oxide) grafted C_{60} fullerene clusters in aqueous solution: a molecular dynamics simulation study
Justin B. Hooper, Dmitry Bedrov and Grant D. Smith, *Phys. Chem. Chem. Phys.*, 2009 **DOI:** 10.1039/b818971d

Determination of pair-wise inter-residue interaction forces from folding pathways and their implementation in coarse-grained folding prediction
Sefer Baday, Burak Erman and Yaman Arkun, *Phys. Chem. Chem. Phys.*, 2009 **DOI:** 10.1039/b820801h

Effective force coarse-graining†

Yanting Wang,‡ W. G. Noid,§ Pu Liu and Gregory A. Voth*

Received 29th October 2008, Accepted 15th January 2009
First published as an Advance Article on the web 12th February 2009
DOI: 10.1039/b819182d

An effective force coarse-graining (EF-CG) method is presented in this paper that complements the more general multiscale coarse-graining (MS-CG) methodology. The EF-CG method determines effective pairwise forces between coarse-grained sites by averaging over the atomistic forces between the corresponding atomic groups in configurations sampled from equilibrium all-atom molecular dynamics simulations. The EF-CG method extracts the transferable part of the MS-CG force field at the cost of reduced accuracy in reproducing certain structural properties. Therefore, the EF-CG method provides an alternative to the MS-CG approach for determining CG force fields that give improved transferability but reduced structural accuracy. The EF-CG method is especially suitable for coarse-graining large molecules with high symmetry, such as bulky organic molecules, and for studying complex phenomena across a range of thermodynamic conditions. The connection between the EF-CG and MS-CG approaches as well as the limitations of the EF-CG method are also discussed. Numerical results for neopentane, methanol and ionic liquid systems illustrate the utility of the method.

I. Introduction

With the rapid growth of computational power and simulation methods, computer simulations play an increasingly important role in the investigation of condensed matter and biological systems. First principles simulations based on quantum mechanics very accurately reproduce some molecular properties of interest. However, because they are very computationally demanding, first principles simulations are generally only suitable for simulating systems with hundreds of atoms. Classical atomistic force fields, such as AMBER,[1] OPLS-AA,[2] and CHARMM,[3] have been developed for simulating much larger systems with millions of atoms for tens of nanoseconds. Nevertheless, many interesting phenomena in complex condensed matter and biological systems occur on mesoscopic scales involving many millions of atoms evolving on up to millisecond time scales. In addition, in most cases many of the atomic details involved in a given process are not relevant to the theoretical analysis. Just as atomistic molecular dynamics (MD) simulations have provided a powerful computational methodology for investigating atomic phenomena that cannot be adequately studied with quantum dynamics, ''coarse-grained'' (CG) models that have integrated out much of the atomic detail provide an important tool for investigating mesoscopic and macroscopic phenomena that cannot be adequately studied with atomically detailed models.[4]

Consequently, considerable effort has focused on determining the effective interactions between sites in CG models.[4–8] Many CG models assume that the interactions between CG sites can be represented by particular analytic functional forms and then these functional forms are parameterized either with experimental data or with the results of a more detailed simulation. However, the assumed functional forms may limit the accuracy of such CG approaches. Another class of CG approaches, employing inverse Monte Carlo methods,[9,10] are parameterized to reproduce the radial distribution functions (RDFs) for the CG sites that are observed in equilibrium atomically-detailed simulations. The effective interactions between CG sites are determined through an iterative process that updates the CG force field after each simulation by comparing the pairwise potential of mean forces (PMFs) resulting from a given CG simulation with the PMF measured in atomically detailed simulations. CG models that are parameterized using the inverse Monte Carlo method only consider the target RDFs, namely the final structure, and as a consequence the physical meaning of the resulting interacting potentials is somewhat obscure.

Recently Izvekov and Voth have developed a multiscale coarse-graining (MS-CG) approach[11,12] based on a statistical implementation of the force-matching method[13,14] to systematically determine the low-resolution effective CG force fields from the high-resolution all-atom force fields. The MS-CG approach determines the effective CG interactions between CG sites by optimizing (''matching'') the force on CG sites with a least squares fitting to the total forces on the underlying atomic groups. The MS-CG method has been successfully applied to investigate both liquid[12,15] and biomolecular systems.[11,16–19] Recent work[20,21] has shown that the

Center for Biophysical Modeling and Simulation and Department of Chemistry, University of Utah, 315 S. 1400 E., Rm. 2020, Salt Lake City, Utah 84112-0850, USA.
E-mail: voth@chem.utah.edu; Fax: +1 (1)801 581 4353; Tel: +1 (1)801 581 7272

† Electronic supplementary information (ESI) available: Derivation of eqn (4); comparison of the EF-CG and MS-CG methods. See DOI: 10.1039/b819182d

‡ Present address: Center for Advanced Modeling and Simulation, Idaho National Laboratory, P.O. Box 1625, Idaho Falls, ID 83415-2208, USA.
§ Present address: Department of Chemistry, Pennsylvania State University, University Park, PA 16802, USA.

 This journal is © the Owner Societies 2009

MS-CG method variationally calculates a rigorous optimal approximation to the *many-body* PMF describing the CG sites. If the CG effective potential function only includes central pair potentials, the MS-CG method determines an optimal pair interaction that incorporates critical three-body information through a mechanism[22] that is consistent with well-known theories of the liquid state.[23] Like the inverse Monte Carlo method, the MS-CG method is not limited by assumed functional forms. In contrast, while the inverse Monte Carlo method determines a CG force field that by definition reproduces the pair PMFs, the MS-CG method determines a CG force field that is a systematic approximation to the many-body PMF. Moreover, as discussed by Noid *et al.*,[20] the CG interactions determined by the MS-CG methodology have clear physical meaning.

However, because the many-body PMF depends upon the thermodynamic conditions, effective interactions determined by the MS-CG method at one thermodynamic state point may not necessarily be readily transferable to other state points. Consequently it is important to consider mechanisms for modifying the MS-CG method to optimize the transferability of the effective CG interactions. The present work introduces a new effective force coarse-graining (EF-CG) method that employs the general theoretic framework for the MS-CG method, but identifies an alternative approximation to the many-body PMF that determines effective CG interactions that not only have improved transferability but that are also much more easily computed. Although at a given thermodynamic state point the EF-CG force field may provide a less accurate reproduction of average structure than the more general MS-CG method, the EF-CG interactions may be employed in CG MD simulations under thermodynamic and environmental conditions very different from those of the all-atom MD simulations employed in constructing the CG force fields. In particular, CG simulations employing the EF-CG force field determined at a single thermodynamic state point may be used to investigate systems at different temperatures, different sizes, and also in surface simulations using the CG force fields obtained from a bulk all-atom simulation, albeit with reduced accuracy for a given set of conditions. Moreover, the underlying simplicity of the EF-CG model allows the development of CG force fields that can be readily decomposed into various contributions and manipulated for use in different systems.

This paper is organized as follows. In section II the methodology of the EF-CG approach is described, while in section III the theoretical connection between the EF-CG method and the more general MS-CG methodology is derived; in section IV the EF-CG results for several model systems are shown, and a discussion of these results and associated conclusions are provided in section V.

II. Methods

In this section, the EF-CG methodology for systematically developing effective CG force fields from atomistic force fields is described. The source of error in the EF-CG method is analyzed and may be used as a guide for optimizing the CG strategy. It should be noted that, in comparison to the more general MS-CG approach,[20,21] the EF-CG method relies upon a number of additional assumptions. The EF-CG method can only be applied for systems in which the non-bonded interactions in the underlying model are explicitly pairwise decomposable and the bonded interactions are treated separately. Additionally, while the MS-CG method may be applied for CG models in which certain molecular species have been entirely integrated out (*e.g.*, solvent-free CG models), the EF-CG method assumes that each atom describing the system is involved in exactly one site and that this site is located at the center-of-mass for the atoms involved in the site. Consequently, the present work only considers systems with explicit pairwise non-bonded atomistic force fields for which all molecules will be represented in the CG model. Only molecular dynamics (MD) computer simulations are considered at both the all-atom and CG simulation levels.

Computing effective forces

Because most atomistic force fields model non-bonded interactions with pair potentials, it is highly desirable to develop CG force fields with similar pairwise additivity. Early applications of the MS-CG approach determined the CG pair potentials providing the best fit to the *total* force on each CG site as measured in atomistic MD simulations. This optimization procedure maps contributions from the environment onto the interaction between each pair of CG sites in such a way that the resulting CG potential provides an optimal approximation to the many-body PMF. In contrast, the EF-CG method explicitly computes the total translational force between *two* atomic groups and projects this force onto the radial vector between the centers-of-mass (CMs) for these two atomic groups. The obtained radial forces are then averaged over different orientations so that the effective forces only depend on radial distance between CMs, but do not introduce contributions from the environment. In this way, the effects of averaging over internal and orientational degrees of freedom are confined to the particular pair force being computed and do not "contaminate" other pair forces. Consequently, the EF-CG procedure may provide greater opportunity for deriving transferable CG potentials. This procedure is explained in detail below.

The EF-CG method employs the same coarse-graining strategy as previous applications of the MS-CG method. That is, several bonded atoms are grouped together to define a single dimensionless CG site. The effective non-bonded forces between CG sites are assumed to be pairwise and central, so that the forces between the CG sites are directed along the vector connecting the CMs, and are a function of the radial distance between the CMs. This central pairwise approximation for the non-bonded CG potential is perhaps the most significant source of errors in both the MS-CG and EF-CG approaches. However, this assumption affords tremendous simplification in ensuing MD simulations of the CG model.

For the model systems studied in this paper, the bonded contributions to the CG force fields are obtained separately using the two-step Boltzmann fitting procedure described in ref. 15, after calculating the EF-CG non-bonded forces. The Boltzmann procedure is suitable for systems in which the

bonded interactions are only weakly coupled to the remaining degrees of freedom. For systems with weaker bonded CG interactions, more fitting steps or even alternative methods may be necessary. The following discussion concentrates on obtaining the non-bonded CG forces by the EF-CG method.

Consider a given system described by n atoms with Cartesian coordinates, $\mathbf{r}^n$, and a potential energy function, $u(\mathbf{r}^n)$. The total vector force on atom, i, generally depends upon the coordinates of all the atoms in the system and is determined from the gradient of the potential energy function in the standard way:

$$\mathbf{f}_i(\mathbf{r}^n) = -\partial u(\mathbf{r}^n)/\partial \mathbf{r}_i \tag{1}$$

A CG model described in terms of N sites may be developed by partitioning the n atoms into N atomic groups, so that each atom is involved in one (and only one) group. Each of these atomic groups is then associated with a single CG site and the group of n_I atoms associated with site I is referred to as the set, I_I, of atoms "involved" in the CG site for each $I = 1,\ldots,N$. The Cartesian coordinates, $\mathbf{r}_I^{n_I}$, for the atoms in set I_I then define the Cartesian coordinates for coarse-grained (CG) site I, $\mathbf{R}_I$, through a mapping operator:

$$\mathbf{R}_I = \mathbf{M}_{\mathbf{R}I}(\mathbf{r}_I^{n_I}) = \sum_{i\in I_I} m_i \mathbf{r}_i \Big/ \sum_{i\in I_I} m_i \tag{2}$$

In particular, the EF-CG map defined by eqn (2) places each CG site at the center-of-mass for the associated atomic group.

The non-bonded interactions between the groups of atoms involved in two CG sites, I and J, determine an atomically-detailed CG potential u_{IJ} that is a sum of pair potentials between the atoms involved in the two sites:

$$u_{IJ}(\mathbf{r}_I^{n_I}, \mathbf{r}_J^{n_J}) = \sum_{i\in I_I}\sum_{j\in I_J} u_{ij}(r_{ij}), \tag{3}$$

where r_{ij} is the distance between atoms i and j, and u_{ij} is an atomistic non-bonded pair potential describing the interaction between atoms i and j (*e.g.*, van der Waals (VDW) and/or electrostatic).

For an isotropic homogeneous system, the magnitude of the effective force between CG sites (atomic groups) I and J may be defined in terms of the gradient of the potential in eqn (3),

$$\begin{aligned} F_{IJ}^{\mathrm{EF}}(R) &= -\left\langle \frac{\partial u_{IJ}(\mathbf{r}_I^{n_I}, \mathbf{r}_J^{n_J})}{\partial M_{IJ}(\mathbf{r}_I^{n_I}, \mathbf{r}_J^{n_J})} \right\rangle_{R_{IJ}} \\ &= \langle \hat{\mathbf{M}}_{IJ}(\mathbf{r}_I^{n_I}, \mathbf{r}_J^{n_J}) \cdot \mathbf{f}_{IJ}(\mathbf{r}_I^{n_I}, \mathbf{r}_J^{n_J}) \rangle_{R_{IJ}} \end{aligned} \tag{4}$$

where the subscripted angular brackets denote a conditioned canonical ensemble average over atomistic configurations for which the two groups are separated by a distance R,

$$\langle a(\mathbf{r}^n)\rangle_{R_{IJ}} = \langle a(\mathbf{r}^n)\delta(M_{IJ}(\mathbf{r}_I^{n_I}, \mathbf{r}_J^{n_J}) - R)\rangle / \langle \delta(M_{IJ}(\mathbf{r}_I^{n_I}, \mathbf{r}_J^{n_J}) - R)\rangle, \tag{5}$$

the total force on group I from group J is defined as a sum of non-bonded forces between pairs of atoms

$$\mathbf{f}_{IJ}(\mathbf{r}_I^{n_I}, \mathbf{r}_J^{n_J}) = \sum_{i\in I_I}\sum_{j\in I_J} \mathbf{f}_{ij}(\mathbf{r}_i, \mathbf{r}_j), \tag{6}$$

and the mapping operator in eqn (2) has been used to define a vector between CG sites

$$\mathbf{M}_{IJ}(\mathbf{r}_I^{n_I}, \mathbf{r}_J^{n_J}) = \mathbf{M}_{\mathbf{R}I}(\mathbf{r}_I^{n_I}) - \mathbf{M}_{\mathbf{R}J}(\mathbf{r}_J^{n_J}) \tag{7}$$

$M_{IJ}(\mathbf{r}_I^{n_I}, \mathbf{r}_J^{n_J}) = |\mathbf{M}_{IJ}(\mathbf{r}_I^{n_I}, \mathbf{r}_J^{n_J})|$, and $\hat{\mathbf{M}}_{IJ}(\mathbf{r}_I^{n_I}, \mathbf{r}_J^{n_J}) = \mathbf{M}_{IJ}(\mathbf{r}_I^{n_I}, \mathbf{r}_J^{n_J}) / |\mathbf{M}_{IJ}(\mathbf{r}_I^{n_I}, \mathbf{r}_J^{n_J})|$. An explicit derivation of eqn (4) is given in the ESI.† From eqn (4) it is clear that the EF-CG pair non-bonded force, $F_{IJ}^{\mathrm{EF}}(R)$, is the average projection of the atomic forces between two groups onto the vector connecting the center-of-mass for the two groups given that the two center-of-mass for the two groups are separated by a distance R. Note that according to eqn (6) $\mathbf{f}_{IJ} = -\mathbf{f}_{JI}$ and $\mathbf{f}_{II} = \mathbf{0}$.

The EF-CG pair force in eqn (4) is evaluated at discrete values of R by sampling equilibrium atomistic MD simulations. It should be emphasized that the calculation of the effective force between a pair of CG sites only explicitly considers the specific pair of atomic groups. The presence or absence of other atomic groups only influences the non-bonded EF-CG force field by altering the probabilities of different realizations of the internal coordinates for a given separation of the two atomic groups. This feature improves the transferability of the non-bonded EF-CG force field. Moreover, because CG pairs are considered separately, the calculation of the EF-CG pair forces is computationally inexpensive.

Another feature of the EF-CG method, which can be seen from eqn (4), is that the effective CG forces for different types of interactions can be computed separately. For most chemical and biological systems, the atomistic non-bonded interactions are modeled with only VDW and electrostatic interactions. Because of the linearity of eqn (4), the EF-CG pair force $F_{IJ}^{EF}(R)$ can be obtained by simply adding the VDW and electrostatic effective forces, such that

$$F_{IJ}^{\mathrm{EF}}(R) = F_{IJ}^{\mathrm{EL}}(R) + F_{IJ}^{\mathrm{VDW}}(R) \tag{8}$$

Therefore, the EF-CG approach allows quantitative comparisons of these two types of interactions at the CG level. This relation further ensures the additivity of the EF-CG approach.

Excluded volume effects that prevent atomic groups from overlapping give rise to a "core region" of radius R_c around atomic groups that are not accessible to other groups. Small inter-group separations $R \leq R_c + \delta$ (where δ is some small number) are then only rarely sampled by atomistic MD simulations and, consequently, the EF-CG force field cannot be accurately determined from eqn (4) in or near the boundary of this core region. In order to prevent the overlapping of CG sites in ensuing MD simulations with the EF-CG force field, the force field must be properly extended into the core region. This extension must be sufficiently "hard" to ensure the appropriate excluded volume for the CG sites, while at the same time being sufficiently "soft" to ensure energy conservation is not violated by sites hitting the core region too rapidly. For the effective forces reported in this paper, the pair force is linearly extended to $R_0 = 0$ from R_c. The force value at R_0 is arbitrarily set to be $F_{IJ}^{\mathrm{EF}}(R_0) = F_{IJ}^{\mathrm{EF}}(R_c) + 10$ eV/Å. Nevertheless, our numerical tests have demonstrated that the CG MD simulations are relatively insensitive to the specific methods of extending the force field into the core region.

This journal is © the Owner Societies 2009

Treatment of the Ewald sum

For systems modeled under periodic boundary conditions (PBC), lattice sum techniques based upon the Ewald sum method[24] are frequently employed to calculate the long-range electrostatic interactions. Such lattice summations necessarily introduce forces arising from an infinite number of image atoms. Although it is possible in principle to decompose the Ewald sum into pair contributions and compute the effective electrostatic forces according to eqn (4), because the magnitude of the forces in the lattice sum depends on the geometry of the PBC simulation box, an explicit calculation of the effective electrostatic force based upon the lattice sum may limit the transferability of the resulting EF-CG force field. Rather, in the present work the electrostatic forces between atomic groups considered in eqn (4) are computed directly according to the Coulomb force between atoms i and j:

$$f_{ij}^{\mathrm{EL}}(r_{ij}) = K_{\mathrm{e}} \frac{q_i q_j}{r_{ij}^2} \tag{9}$$

where K_{e} is the appropriate electrostatic constant, and q_i and q_j are the partial charges on atoms i and j, respectively.

The effective electrostatic CG force between CG sites I and J may then obtained by substituting eqn (9) into eqn (4). The ensemble average of eqn (4) is sampled in atomistic MD simulations employing the Ewald method to evaluate electrostatic interactions. The partial charge on each CG site is defined as the sum over all underlying atomic partial charges $Q_I = \sum_{i \in I_I} q_i$. The total electrostatic force between two atomic groups is generally not equal to the point-charge electrostatic force between the groups, $f_{IJ}^{\mathrm{EL}}(R) = K_{\mathrm{e}} \frac{Q_I Q_J}{R^2}$, because asymmetry in the charge distribution introduces contributions from the dipole and higher moments. Nevertheless, the present calculations show that, for the systems considered in this work, the effective electrostatic CG force defined by eqn (4) and (8) is significantly different from the point-charge electrostatic force only at short range (roughly less than 12 Å). Thus the EF-CG electrostatic force can be decomposed in practice into a long-range monopole electrostatic force and a short-range effective multipole electrostatic force. In the CG MD simulations employing the EF-CG force field, the monopole electrostatic force is calculated *via* the Ewald sum, while the short-range contribution from the effective multipole electrostatic force is added to the effective VDW force to define a total effective short-range CG force. Therefore, the effective CG forces used in CG MD simulations may be expressed

$$\begin{aligned} F_{IJ}^{\mathrm{total}}(R) &= F_{IJ}^{\mathrm{EL}}(R) + F_{IJ}^{\mathrm{VDW}}(R) \\ &= F_{IJ}^{\mathrm{point}}(R) + F_{IJ}^{\mathrm{mpole}}(R) + F_{IJ}^{\mathrm{VDW}}(R) \\ &= F_{IJ}^{\mathrm{point}}(R) + F_{IJ}^{\mathrm{short}}(R) \end{aligned} \tag{10}$$

where $F_{IJ}^{\mathrm{point}}(R)$ is the monopole electrostatic interaction treated by the Ewald sum, and $F_{IJ}^{\mathrm{short}}(R)$ is the total short-range interaction between CG sites and includes contributions from both VDW and short-ranged multipole electrostatic interactions.

The above procedure neglects the long-range part of the multipole electrostatic interactions. For atomic groups with high symmetry, this does not lead to significant error, because the electrostatic interaction between two groups is dominated by the monopole (point charge) at large separation. Furthermore, in liquid systems, the Debye screening effect[23] attenuates the effective multipole interaction rather quickly. For asymmetric molecules, the central pairwise approximation employed in the EF-CG force field introduces more significant errors in modeling the interactions between atomic groups than neglecting the long-range multipole electrostatic interactions, as described in the next subsection.

Computational procedure for the EF-CG methodology

The EF-CG forces are constructed in the following steps:

(1) Obtain bonded forces with the Boltzmann fitting method or other approaches from all the sampled configurations.

(2) Read in one sampled configuration.

(3) Calculate different types of effective CG forces for one pair of atomic groups.

(4) Accumulate the obtained CG force values in the corresponding bin.

(5) Repeat steps (3) and (4).

(6) Repeat steps (2) to (5).

(7) Average the accumulated force values and combine the different types of forces together.

(8) Extend the force field into the core region.

(9) Tabulate the CG forces with a suitable force cutoff. Along with the bonded interactions obtained in step (1), the tabulated forces are then ready for CG MD simulations.

Source of errors

In contrast to the MS-CG method, the EF-CG method calculates an effective non-bonded interaction between CG sites directly from the forces between the atomic groups defining the CG sites. Consequently, errors within the calculated EF-CG potentials may be analyzed by considering the forces between atomic groups from the perspective of classical mechanics. In addition to the errors resulting from neglect of multipole interaction, errors in the EF-CG method also result from representing CG sites with structureless point particles.

For a given pair of atomic groups associated with CG sites I and J, the set of underlying atomic coordinates for these groups, $\{\mathbf{r}_I^{n_I}, \mathbf{r}_J^{n_J}\}$, can be described by a set of collective coordinates: the internal vibrational coordinates for each group, d_I and d_J; the relative orientation ω_{IJ} of the two groups, and the distance between the two CMs $R_{IJ} = M_{IJ}(\mathbf{r}_I^{n_I}, \mathbf{r}_J^{n_J})$ as defined after eqn (7). Correspondingly, the total atomistic potential u_{IJ} defined in eqn (3) between these two atomic groups may be represented by potentials describing the internal vibrations $u_{IJ}^{\mathrm{Internal}}$, the rotational torque u_{IJ}^{Torque}, and the translational force u_{IJ}^{Force}, such that

$$\begin{aligned} u_{IJ}(\mathbf{r}_I^{n_I}, \mathbf{r}_J^{n_J}) = u_{IJ}^{\mathrm{Internal}}(d_I, d_J; \omega_{IJ}, R_{IJ}) + u_{IJ}^{\mathrm{Torque}}(\omega_{IJ}; d_I, d_J, R_{IJ}) \\ + u_{IJ}^{\mathrm{Force}}(R_{IJ}; d_I, d_J, \omega_{IJ}) \end{aligned} \tag{11}$$

This collective coordinate representation is illustrated schematically in Fig. 1.

Because the CG model represents each atomic group with a single structureless point particle, the CG model cannot capture the detailed interactions described in eqn (11).

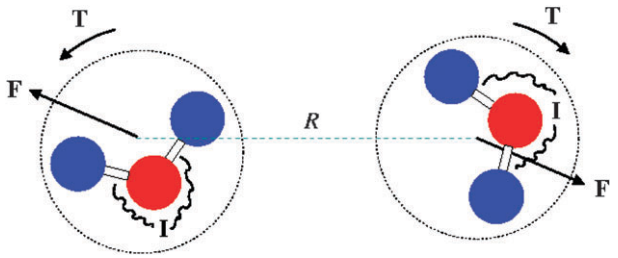

Fig. 1 Schematic illustration of the collective coordinate representation of the force "F" between two atomic groups, with internal modes "I" and torque "T" shown.

Consequently the effects of vibrational degrees of freedom are integrated out in determining the effective CG non-bonded pair forces. Moreover, the orientation for each CG site and corresponding rotational torque cannot be described by the CG model. As a result, even though the instantaneous net force between atomic groups is not necessarily directed along the radial vector between the groups, these forces that are orthogonal to the radial inter-site vector cannot be described in computer simulations of the EF-CG model. The EF-CG potential resulting from integrating out the interactions involving these degrees of freedom corresponds to the effective CG force computed in eqn (4). Note that this interaction is an ensemble average over all possible relative orientations between the two atomic groups.

The approximations described above are the major source of errors in the EF-CG approach and limit its accuracy. In return, they can be used as a guide for the CG strategy to obtain the optimized EF-CG force fields. First, the error introduced by averaging over and neglecting the internal vibrations included in eqn (11) is minimized if the atoms in the CG group are well described as rigid bodies. This approximation also indicates that non-bonded atoms cannot be grouped into one CG site with the EF-CG method, although there exist some other CG approaches, such as the BLOB,[25] designed for this purpose. Second, neglecting the relative orientation of CG sites requires that the atomic groups are as symmetric as possible. Here the "symmetry of an atomic group" refers to the symmetry of the distributions of the VDW parameters and the atomic partial charges of the atomic group.

The rigid body approximation may be satisfied in a liquid state or a biological system with careful design of the CG sites. The approximation of discarding the non-central component of the atomic force is not as bad as it appears. In an homogeneous liquid, because the probabilities of having two vertical forces with the same magnitude but opposite directions are equal, the effective force resulting from projecting the net force along the radial direction is exactly the desired average force. The most significant error of the EF-CG method comes from averaging out the orientation-dependence of the force between atomic groups shown in eqn (11). A non-trivial example is the one-site CG model of water molecules. For two water molecules separated with a fixed distance between their CMs, when the oxygen atom on one water molecule is closer to the hydrogen atoms than to the oxygen atom on the other water molecule, the collective electrostatic interactions are attractive; when the two oxygen atoms are closer, the electrostatic interactions are repulsive. In the one-site CG representation, since the effective force only depends on the radial distance, the two opposite effects will be averaged and cancelled out, and thus cannot be correctly represented in a one-site water EF-CG model.

III. Connection to the general MS-CG theory

MD simulations of a CG model for a given system require an interaction potential describing the effective interactions between the CG sites after averaging over atomistic degrees of freedom. The multiscale coarse-graining (MS-CG) method[20,21] determines this CG force field, $\mathbf{F}$, governing the interactions between the N CG sites through a variational principle for minimizing the functional:

$$\chi^2[\mathbf{F}] = \frac{1}{3N}\left\langle \sum_{I=1}^{N} |\mathbf{F}_I(\mathbf{M}_{\mathbf{R}}^N(\mathbf{r}^n)) - \mathbf{f}_I(\mathbf{r}^n)|^2 \right\rangle \qquad (12)$$

In this equation, $\mathbf{f}_I(\mathbf{r}^n)$ is the net force on the atomic group associated with CG site I in a given atomically detailed structure $\mathbf{r}^n$,[20] $\mathbf{R}^N = \mathbf{M}_{\mathbf{R}}^N(\mathbf{r}^n)$ is the CG representation of this configuration, $\mathbf{F}_I(\mathbf{R}^N)$ is the force on CG site I in the CG representation $\mathbf{R}^N$, and the angular brackets denote a canonical ensemble average evaluated for the atomistic model. The CG force field minimizing the MS-CG functional in eqn (12) is a conditioned canonical ensemble average of the atomistic forces:

$$\mathbf{F}_I(\mathbf{R}^N) = \langle \mathbf{f}_I(\mathbf{r}^n)\rangle_{\mathbf{R}^N}. \qquad (13)$$

where the angular brackets subscripted by $\mathbf{R}^N$ represent an average over all atomistic configurations that map onto the specified CG configuration:

$$\langle a(\mathbf{r}^n)\rangle_{\mathbf{R}^N} = \langle a(\mathbf{r}^n)\delta(\mathbf{M}_{\mathbf{R}}^N(\mathbf{r}^n) - \mathbf{R}^N)\rangle / \langle \delta(\mathbf{M}_{\mathbf{R}}^N(\mathbf{r}^n) - \mathbf{R}^N)\rangle \quad (14)$$

Thus the MS-CG force field minimizing the functional in eqn (12) is a many-body mean force: the MS-CG force on site I in a CG configuration $\mathbf{R}^N$ is the average atomistic force on the same site averaged over all atomistic configurations that map to the given CG configuration. This MS-CG force field corresponds to the force field derived from gradients of the many-body PMF, U^{MS}:

$$\exp[-U^{MS}(\mathbf{R}^N)/k_{\mathrm{B}}T] \propto \langle \delta(\mathbf{M}_{\mathbf{R}}^N(\mathbf{r}^n) - \mathbf{R}^N)\rangle \qquad (15)$$

Because $p_R(\mathbf{R}^N) = \langle \delta(\mathbf{M}_{\mathbf{R}}^N(\mathbf{r}^n) - \mathbf{R}^N)\rangle$ is the probability distribution describing the distribution of CG configurations generated by mapping atomistic configurations that are sampled according to the atomistic distribution, it follows that simulations of a CG model employing the MS-CG potential, U^{MS}, as an interaction potential will sample the CG configuration space with the same probability distribution as CG configurations determined by the atomistic model and CG mapping. Consequently, the MS-CG method provides a variational procedure for calculating the appropriate interaction potential for a CG model that is consistent with a given atomistic model. Moreover, the MS-CG variational principle provides a systematic and rigorous approach for deriving an optimal approximation to this many-body PMF. A more

This journal is © the Owner Societies 2009

detailed and general discussion of the theory underlying the MS-CG method has recently been presented.[20]

As mentioned above, the CG force field defined in eqn (13) expresses the force on the CG site I as a function of the entire CG structure, $\mathbf{R}^N$. It is generally not feasible to either solve or to numerically implement such a many-body interaction potential. Previous work employing the MS-CG method has approximated this many-body potential with central pair potentials describing the non-bonded interactions between CG sites.[11,12,15,16,19,21] Within this approximation the resulting MS-CG force field may be determined by solving a system of linear algebraic equations that are related to the well known Yvon–Born–Green equation and which incorporate critical three-body effects to determine an optimal pair approximation to the many-body PMF defined in eqn (15).[22]

The EF-CG force field introduced in section 2.1 may be derived as a different systematic approximation to the many-body PMF expressed in eqn (13). For the center-of-mass mapping defined in eqn (2), the force on the CG site is simply the sum of the forces on the atoms within the group: $\mathbf{f}_I = \sum_{i \in I_I} \mathbf{f}_i$. Because the atomistic non-bonded interactions have been assumed to be pairwise additive, the total atomistic force on each CG site may be decomposed into a force from bonded interactions, $\mathbf{f}_I^{\text{bon}}$, and a sum of non-bonded pair interactions between atomic groups associated with CG sites:

$$\mathbf{f}_I(\mathbf{r}^n) = \mathbf{f}_I^{\text{bon}}(\mathbf{r}^n) + \sum_{J \neq I} \mathbf{f}_{IJ}(\mathbf{r}_I^{n_I}, \mathbf{r}_J^{n_J}) \tag{16}$$

where $\mathbf{f}_{IJ}$ has been defined in eqn (6). Accordingly, the many-body CG force field defined in eqn (13) may also be formally decomposed into pair interactions:

$$\mathbf{F}_I(\mathbf{R}^N) = \mathbf{F}_I^{\text{bon}}(\mathbf{R}^N) + \sum_{J \neq I} \mathbf{F}_{IJ}(\mathbf{R}^N) \tag{17}$$

where

$$\mathbf{F}_{IJ}(\mathbf{R}^N) = \langle \mathbf{f}_{IJ}(\mathbf{r}_I^{n_I}, \mathbf{r}_J^{n_J}) \rangle_{\mathbf{R}^N} \tag{18}$$

and

$$\mathbf{F}_I^{\text{bon}}(\mathbf{R}^N) = \langle \mathbf{f}_I^{\text{bon}}(\mathbf{r}^n) \rangle_{\mathbf{R}^N} \tag{19}$$

Note that while this decomposition is exact, it is purely formal because each interaction depends upon the coordinates of all N CG sites through the conditional averages defined by eqn (14). This is an important distinction from the assumed pairwise decomposition of the atomistic force field in which each pair interaction is independent of its environment.

The EF-CG non-bonded pair force field may be derived as a mean field approximation to the many-body pair interaction defined in eqn (18):

$$\begin{aligned}\mathbf{F}_{IJ}^{\text{EF}}(\mathbf{R}_I, \mathbf{R}_J) &= \int d\mathbf{X}^N p_R(\mathbf{X}^N) \mathbf{F}_{IJ}(\mathbf{X}_I, \mathbf{X}_J) \delta(\mathbf{X}_I - \mathbf{R}_I) \delta(\mathbf{X}_J - \mathbf{R}_J) \\ &= \frac{\langle \mathbf{f}_{IJ}(\mathbf{r}_I^{n_I}, \mathbf{r}_J^{n_J}) \delta(\mathbf{M}_{\mathbf{R}I}(\mathbf{r}_I^{n_I}) - \mathbf{R}_I) \delta(\mathbf{M}_{\mathbf{R}J}(\mathbf{r}_J^{n_J}) - \mathbf{R}_J) \rangle}{\langle \delta(\mathbf{M}_{\mathbf{R}I}(\mathbf{r}_I^{n_I}) - \mathbf{R}_I) \delta(\mathbf{M}_{\mathbf{R}J}(\mathbf{r}_J^{n_J}) - \mathbf{R}_J) \rangle}\end{aligned} \tag{20}$$

where p_R is the CG configuration probability distribution determined by the atomistic model and CG mapping operators as defined below eqn (15).

According to eqn (20), the EF-CG force field between CG sites I and J with coordinates $\mathbf{R}_I$ and $\mathbf{R}_J$ is the expectation value of the instantaneous net force between the two atomic groups given that the centers-of-mass for the atomic groups are located at the specified coordinates. For systems described by an atomistic potential without an external field, the instantaneous non-bonded force between two atomic groups, $\mathbf{f}_{IJ}(\mathbf{r}_I^{n_I}, \mathbf{r}_J^{n_J})$, is independent of translation and rotation of the system. Consequently it follows from symmetry that, although the instantaneous atomic force is not necessarily directed along the vector between the two CG sites and also depends upon the atomistic degrees of freedom associated with the two sites, the conditional average of this atomic force is directed along this vector and depends only upon the distance between the atomic groups, $R_{IJ} = |\mathbf{R}_{IJ}|$. Therefore under these assumptions, the EF-CG force as defined in eqn (20) may be expressed:

$$\mathbf{F}_{IJ}^{\text{EF}}(\mathbf{R}_I, \mathbf{R}_J) = F_{IJ}^{\text{EF}}(R_{IJ}) \hat{\mathbf{R}}_{IJ} \tag{21}$$

where $F_{IJ}^{\text{EF}}(R)$ is defined by (4). Therefore, the EF-CG non-bonded forces provide a mean field approximation to the exact many-body MS-CG non-bonded pair force defined by eqn (18).

IV. Results and discussion

By averaging over atomistic details, coarse-grained interaction potentials necessarily provide a less accurate model of molecular interactions than atomistic force fields. Based on the error analysis given in section 2.3, it is clear that the EF-CG force field provides an exact model for molecular interactions only for idealized spherically symmetric rigid molecules. For real molecules, the errors introduced by the EF-CG force fields may be predicted by considering the deviations of the molecular geometry from the ideal case. In this section three liquid systems are considered as the test cases for the EF-CG method: the highly symmetric case of one-site neopentane, two-site methanol as an asymmetric case with multiple CG sites and a CG bond, and the $EMIM^+/NO_3^-$ ionic liquid as an example for complex liquids. All of the MD simulations were performed with the DL_POLY simulation package.[26] The systems were coupled to a Nosé–Hoover thermostat[27] for simulations in the constant NVT ensemble as well as a Hoover barostat[28] for simulations in the constant NPT ensemble. In the constant NPT simulations, the pressure was always set to be 1 atm. The long-range electrostatic interactions were computed with the standard Ewald sum algorithm.[24] For all three systems, unless otherwise specified, both the VDW cutoff and the real space cutoff for the Ewald sum were 12 Å. The integration timestep of the MD simulations was 1 fs. The bin size for tabulating the EF-CG force field was chosen to be 0.01 Å for all of the three systems.

One-site neopentane

Since neopentane has a symmetric tetrahedron structure and negligible atomic partial charges, it was selected as a highly symmetric case for testing the EF-CG method. Due to the high molecular symmetry, simulations employing the EF-CG force field were expected to accurately reproduce the structural and

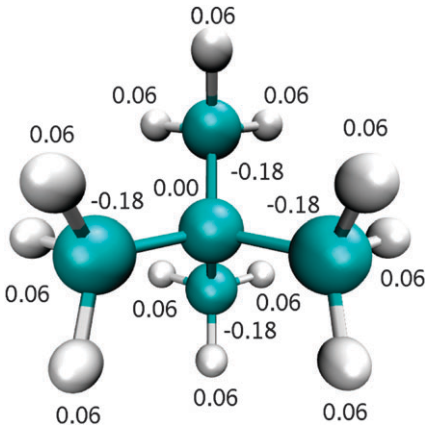

Fig. 2 Molecular structure and atomic partial charges in atomic units for neopentane. In the one-site CG scheme, the whole molecule is coarse-grained to be site A.

thermodynamic properties. The molecular structure and the atomic partial charges of neopentane are shown in Fig. 2. The force field parameters and atomic partial charges for neopentane were taken from the OPLS-AA force field.[2] An atomic model of the liquid neopentane system with 512 molecules was first simulated in the constant *NPT* ensemble at $T = 270$ K for 1 ns. The measured average cubic simulation box size was $L_{neo} = 45.19 \pm 0.12$ Å. All-atom MD simulations of the same system were then performed for 2 ns in the constant *NVT* ensemble at the same temperature and in a cubic simulation box with sides of fixed length, L_{neo}. In total, 1000 configurations were sampled at evenly spaced intervals during this equilibrium run.

The EF-CG approach was then applied to compute the effective CG forces for one-site neopentane from the 1000 sampled all-atom configurations. The effective VDW, long-range electrostatic, and total short-range effective forces (VDW plus multipole electrostatic) obtained by the EF-CG method are shown in Fig. 3. The effective VDW and electrostatic forces were computed according to eqn (4). The total effective force was then computed according to eqn (10). It can be seen that the effective electrostatic force is negligible, since the net charge of a neopentane molecule is zero, and the multipole effect is very small due to the small partial charges of each atom and the symmetric distribution of the partial charges. As a consequence, the total effective force between two neopentane molecules is dominated by the effective VDW force. The cutoff distance for the total effective force is chosen to be 16 Å, beyond which the force value is basically zero.

The total effective CG force field for a one-site model of neopentane was then used to perform constant *NPT* CG MD simulation at $T = 270$ K for 1 ns. All simulation parameters were the same as in the all-atom *NPT* MD simulations. Note that, since the CG sites diffuse faster in a CG MD simulation than in an all-atom MD simulation, the 1 ns of CG simulation sampled as much conformational space as several ns of an all-atom simulation. Moreover, the timestep employed in the CG MD simulation could be increased due to the larger mass of the CG sites. However, for the purposes of comparing the two simulations, the same timestep of 1 fs was employed in both the all-atom and CG MD simulations. The average side length of the cubic simulation box obtained from this constant *NPT* CG MD simulation was $L_{neo}^{CG} = 45.69 \pm 0.13$ Å, which is 1.1% greater than the equilibrium obtained L_{neo} from the all-atom MD simulation. A constant *NVT* CG MD simulation was then simulated at the same temperature with the side length of the cubic simulation box set to L_{neo}^{CG}. The radial distribution function (RDF) obtained from the CG simulation is plotted in Fig. 4 and compared with the RDF obtained from the all-atom MD simulation. There are only very slight differences between the two RDFs, showing that the EF-CG force field is able to accurately rebuild the structural properties of the neopentane system.

The above results for neopentane show that the EF-CG method determines a very good effective CG force field for neopentane. CG MD simulations employing the EF-CG force field reproduce both the structural and thermodynamic properties of the atomistic model with high accuracy. This indicates that the EF-CG method works particularly well for highly symmetric molecules.

Two-site methanol

The above results for a one-site CG model of neopentane verified the accuracy of the EF-CG model for systems with high symmetry and negligible electrostatic interactions. Here, the EF-CG method was tested to develop a CG force field for

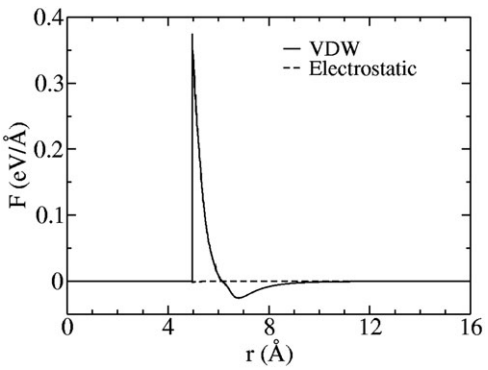

Fig. 3 Effective CG forces for one-site neopentane. Since the electrostatic force (dashed line) is negligible, the total effective force is almost identical to the effective VDW force (solid line).

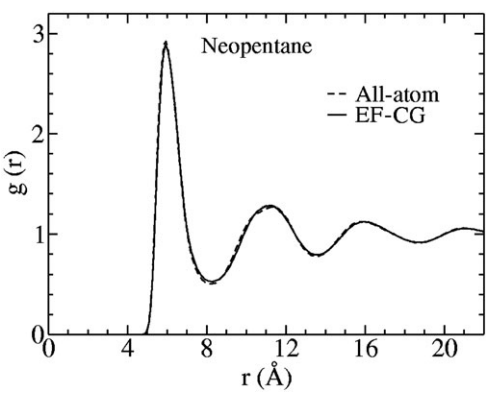

Fig. 4 Radial distribution functions from both the all-atom and the EF-CG MD simulations with 512 neopentane molecules.

This journal is © the Owner Societies 2009

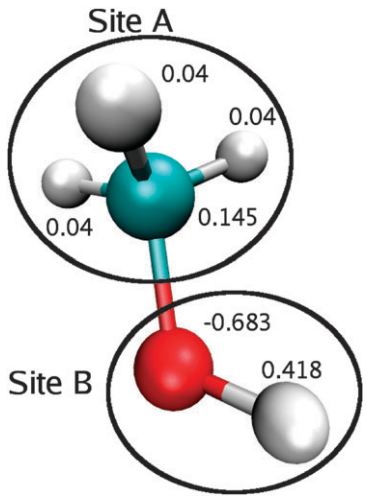

Fig. 5 Molecular structure, atomic partial charges in atomic units, and two-site CG scheme for methanol. According to the atomic partial charges, site A is quite symmetric, while site B is very asymmetric.

a two-site model of methanol. This test case provides a slightly more complicated system because methanol is an asymmetric molecule and the CG model must incorporate bonded interactions and non-zero partial charges on each CG site. The OPLS-AA force field[2] was again used in the all-atom MD simulations of methanol. The molecular structure and atomic partial charges of the methanol molecule are shown in Fig. 5. A liquid system containing 1000 atomically-detailed methanol molecules was simulated first in the constant NPT and then the constant NVT ensemble at $T = 300$ K according to the procedure outlined for neopentane. One thousand configurations were evenly sampled during the 1 ns constant NVT run.

As also shown in Fig. 5, the CH_3 group is represented as CG site A with a net partial charge of 0.265 atomic units, and the OH group as site B with a net partial charge of −0.265. The CG bond between sites A and B was modeled by a harmonic bond $V(r) = \frac{1}{2}k(r - r_0)^2$ and the two parameters for the A–B bond were determined to be $k = 25.5537$ eV Å^{-2} and $r_0 = 1.50779$ Å *via* the two-step Boltzmann fitting procedure[15] using the distribution of distances between the CMs of the

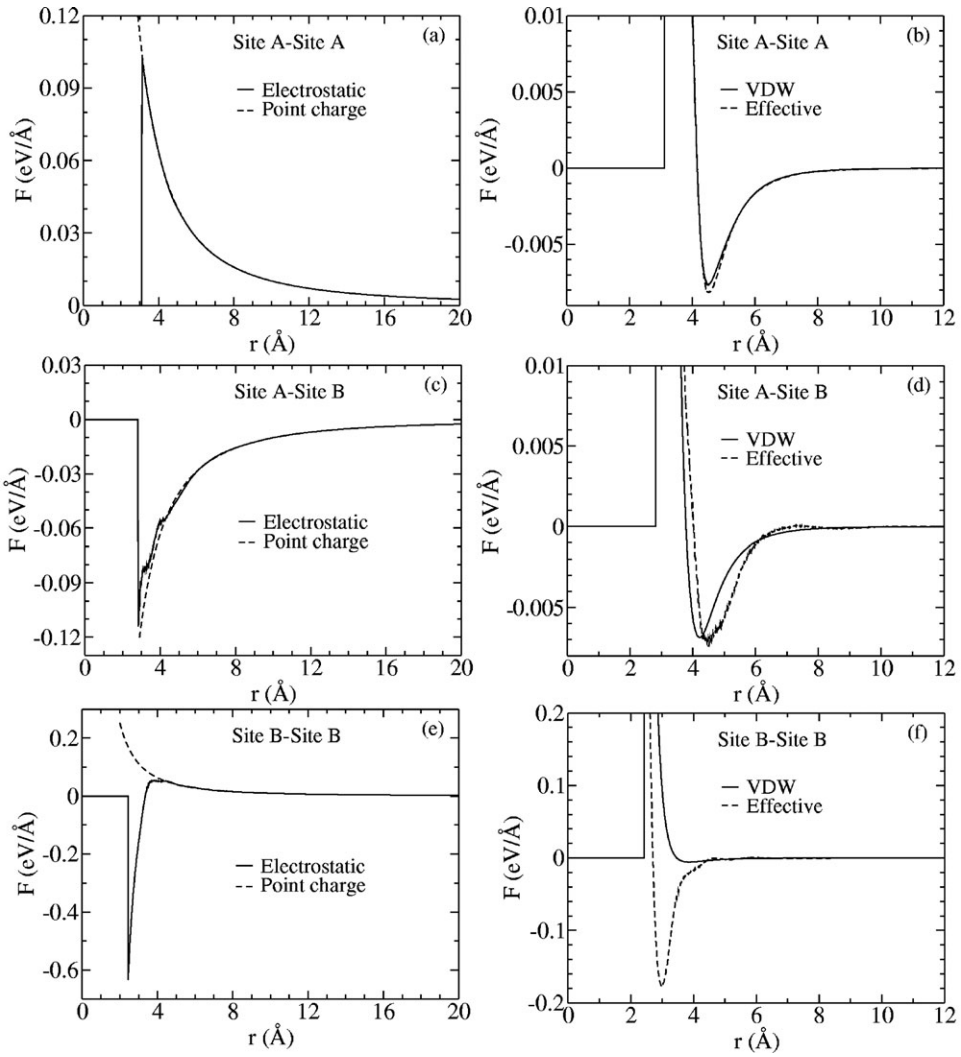

Fig. 6 EF-CG forces for two-site methanol. The difference between the point charge and the effective electrostatic forces are added to the VDW forces to form the total effective short-range CG forces.

two CG sites observed in the atomistic MD simulations. The EF-CG non-bonded force field was computed using the 1000 saved all-atom configurations. The effective CG forces are shown in Fig. 6. From these plots, it is clear that the effective electrostatic interactions, especially those between CG sites B and B, deviate significantly from the Coulomb interaction for point charges, given by eqn (9), at short distances, due to the strong dipole interactions. At medium to long range, the B–B effective electrostatic interaction is well described by the point charge interactions, since the orientation dependence of the electrostatic interactions, *i.e.* the effect of multipole interactions, are very weak when the distance between the CMs is sufficiently large.

The EF-CG force field for the CG two-site methanol model was then employed in corresponding CG MD simulations at $T = 300$ K. The simulated time is 1 ns for each run. The average side length of the cubic simulation box was 40.59 ± 0.13 Å in constant *NPT* CG MD simulations, a relative error of 0.8% with respect to the all-atom value of 40.93 ± 0.11 Å. The RDFs from both the all-atom and CG MD runs are compared in Fig. 7. Although the relation between the RDF and the pair interaction is complicated, the connection between them may be qualitatively explained for the two-site methanol. Since the CH_3 group has a much higher symmetry than the OH group, the RDF for A–A sites obtained in the CG MD run agrees much better with the all-atom result than that for B–B sites. The agreement between the A–B RDF measured in CG and all-atom simulations is intermediate between the above two.

In order to test its transferability for simulations of different system sizes, the EF-CG force field obtained from the system with 1000 methanol molecules was used in the same simulation procedure as before, but for a smaller system with only 343 molecules. The corresponding all-atom simulations were also performed for comparison. The side length of the cubic simulation box for the CG MD simulation is 28.40 ± 0.15 Å,

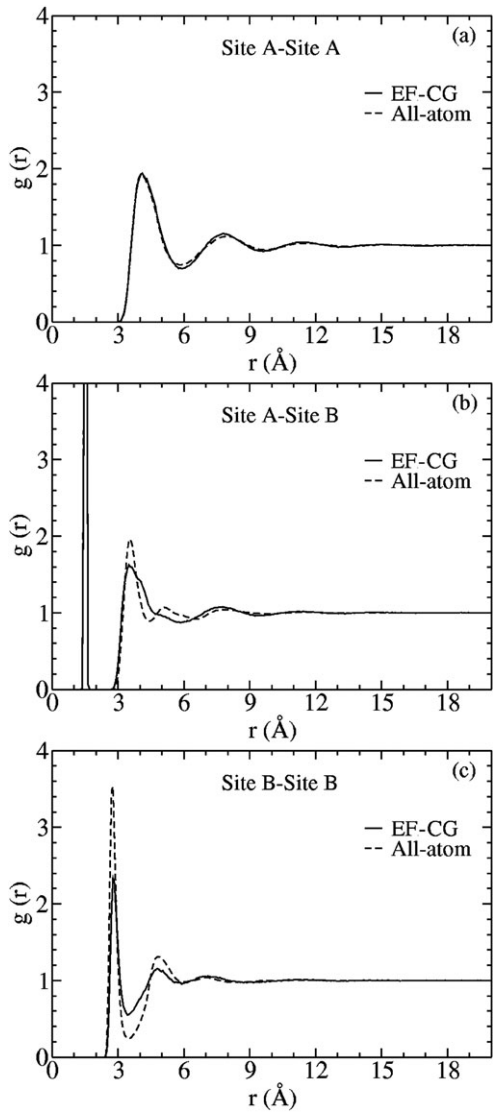

Fig. 7 Radial distribution functions from the all-atom and two-site EF-CG MD simulations of methanol with 1000 molecules.

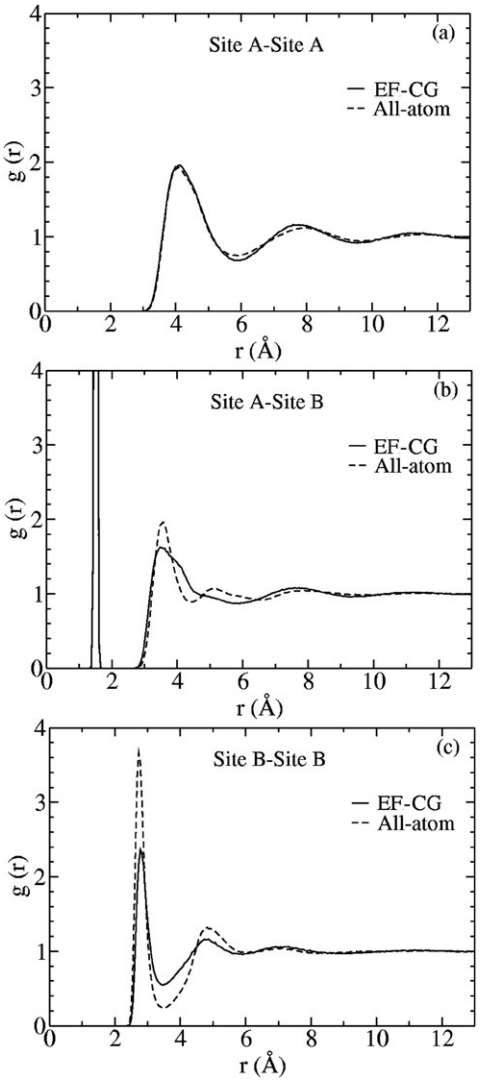

Fig. 8 Radial distribution functions from both the all-atom and the EF-CG MD simulations with 343 methanol molecules. The EF-CG force field was constructed from the all-atom system with 1000 methanol molecules and transferred to the smaller system.

This journal is © the Owner Societies 2009

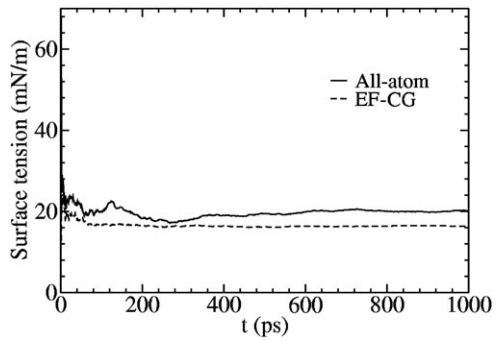

Fig. 9 Running average of the surface tensions of the methanol surface from both the all-atom and the EF-CG MD simulations.

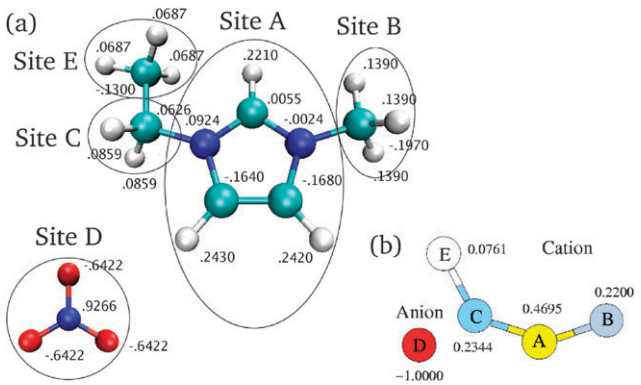

Fig. 10 (a) Molecular structure and atomic partial charges in atomic unit of the $EMIM^+/NO_3^-$ ionic liquid. (b) Corresponding coarse-graining scheme.

a relative error of 1.0% compared with 28.70 ± 0.12 Å for the all-atom MD simulation. The RDFs computed for the CG and atomistic model are compared in Fig. 8 and agree quite well. From these results, it can be concluded that the EF-CG force fields have good transferability between different system sizes.

The transferability of the EF-CG force field between bulk and surface environments was also tested. For both all-atom and CG models of 1000 methanol molecules, two vacuum spaces of the same size as the bulk system were added above and below the methanol fluid along the z-axis. After short equilibration runs, both systems were simulated in the constant NVT ensemble for 1 ns to sample the surface tension. The running averages of the surface tensions computed for both simulations are compared in Fig. 9. The surface tension observed in the all-atom MD simulation is approximately 20 mN m^{-1}, while that in the CG MD simulation is roughly 16 mN m^{-1}. The noticeable difference may be explained by the asymmetry of the CG sites, especially site B. Nevertheless, the surface tension measured in the CG simulation still semi-quantitatively agrees with that measured in all-atom simulation. Considering the sensitivity of the surface tension to the effective CG forces, the EF-CG model can be considered to provide good transferability between bulk and surface environments.

From the results for methanol, it can be concluded that the EF-CG method determines an accurate model for systems composed of molecules with bonded and electrostatic interactions. The errors in CG MD simulations may be understood by considering the geometry of the atomic groups: the more asymmetric the atomic group is, the more significant are the errors in the EF-CG model. These results also verify that the EF-CG force fields may be successfully transferred between different system sizes and the bulk/surface systems.

Ionic liquid

Ionic liquids have important applications in many different fields,[29–32] and have been studied as potential solvents,

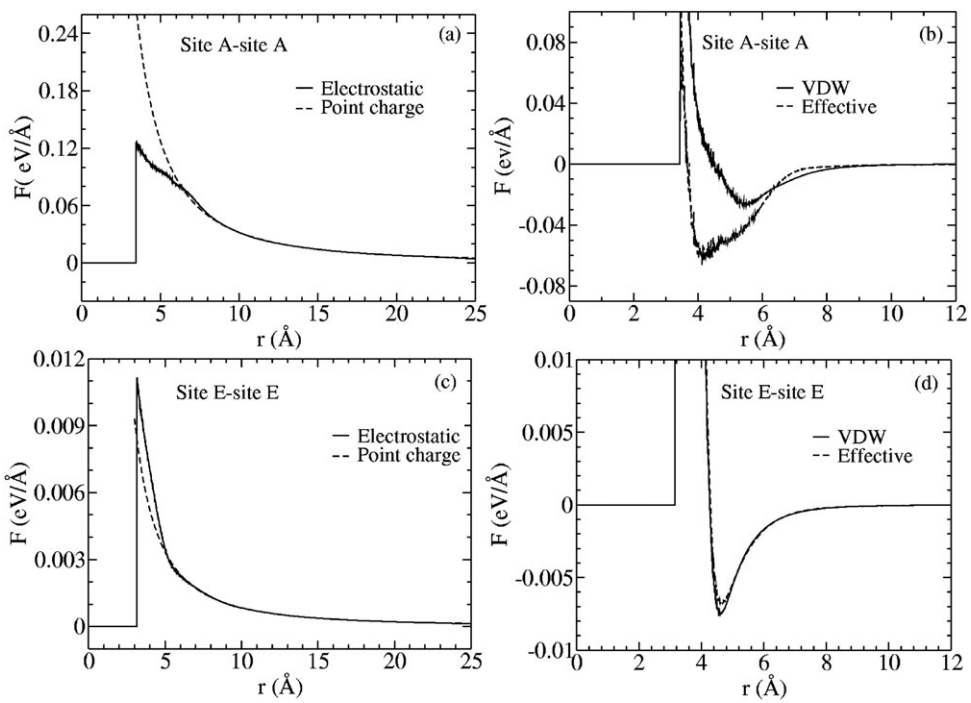

Fig. 11 Sample effective short-range EF-CG forces for the $EMIM^+/NO_3^-$ ionic liquid.

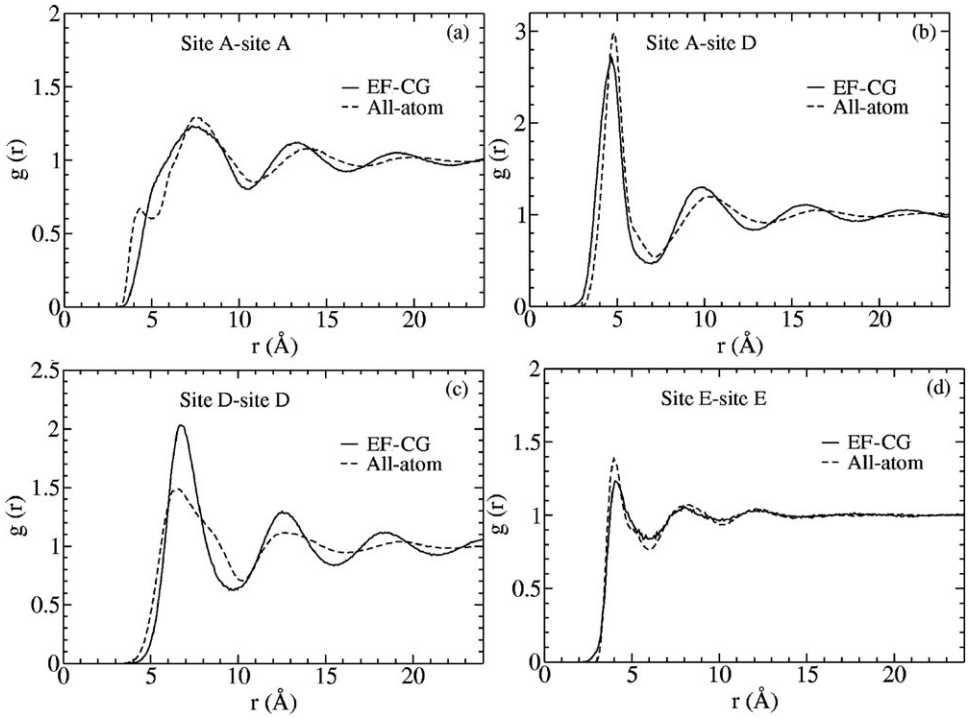

Fig. 12 Sample radial distribution functions for the $EMIM^+/NO_3^-$ ionic liquid at $T = 400$ K from both the all-atom and the EF-CG MD simulations.

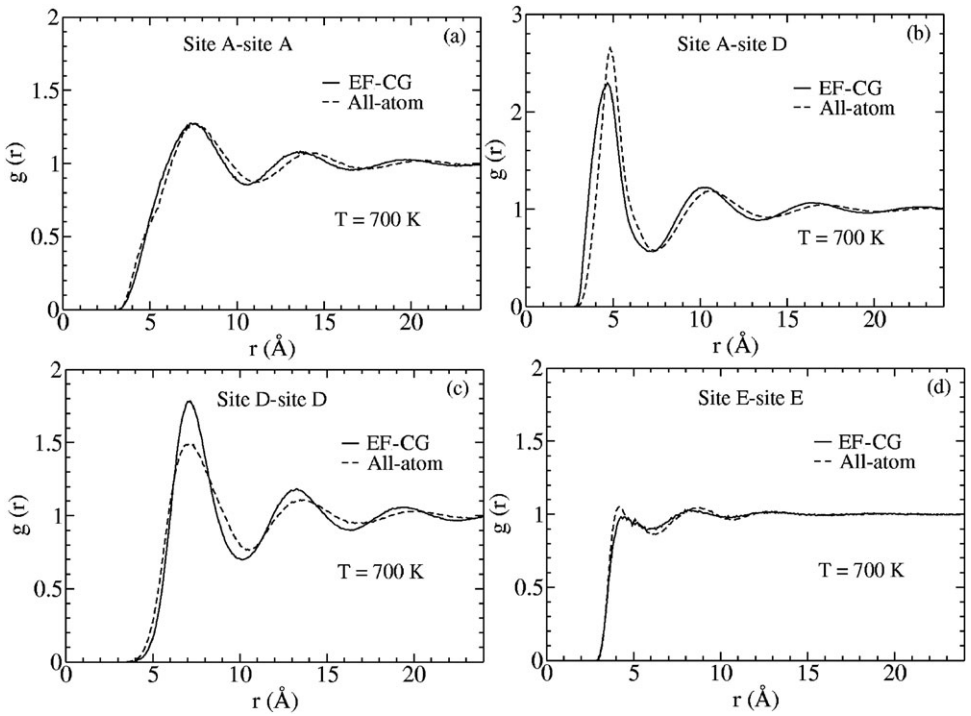

Fig. 13 Sample radial distribution functions for the $EMIM^+/NO_3^-$ionic liquid at $T = 700$ K from both the all-atom and the EF-CG MD simulations. The EF-CG force field was constructed from the all-atom MD simulation at $T = 400$ K and transferred to the higher temperature.

lubricants and energetic materials. The $EMIM^+/NO_3^-$ ionic liquid has been successfully coarse-grained by the MS-CG approach.[15] However, because the MS-CG force fields incorporate entropic effects, at least in principle, the MS-CG interaction depends upon the thermodynamic conditions employed in the parameterization. In this subsection, it is shown that, with little loss in the accuracy of reproducing local structure, the EF-CG force field obtained for this ionic liquid is reasonably transferable between different temperatures, albeit at a reduced degree of accuracy for any given temperature.

 This journal is © the Owner Societies 2009

The molecular structure, atomic partial charges, and the coarse-graining scheme for the $EMIM^+/NO_3^-$ ionic liquid are shown in Fig. 10. The AMBER[1] force field parameters were used in the all-atom MD simulations. Although it has been shown that many-body electric polarization effects are important for ionic liquids,[33] because the EF-CG method has been developed for atomistic force fields with explicitly pairwise additive non-bonded interactions, a non-polarizable model was used in the current study, using force-field parameters listed in ref. 15. An atomically-detailed model of the ionic liquid system with 512 ion pairs was simulated at constant *NPT* for 1 ns followed by a 1 ns simulation at constant *NVT* simulation at $T = 400$ K. The EF-CG force field was then obtained by computing the effective forces based on the 1000 evenly sampled configurations during the *NVT* run. The effective forces between A–A sites and E–E sites are shown in Fig. 11. The strong orientation dependence of the electrostatic interaction between sites A–A at short range causes the total short-range effective force to have a much deeper well than the effective VDW force. Therefore, the interaction between the polar cationic rings is dominated by the electrostatic interactions. In contrast, although the effective electrostatic interaction between E–E sites slightly deviates from a point charge interaction, the effective short-ranged electrostatic interaction is much smaller than the effective VDW interaction. Therefore, the total short-range effective force between sites E–E is very similar to the VDW force.

The EF-CG force fields for the $EMIM^+/NO_3^-$ ionic liquid were then used in CG MD simulations for 512 ion pairs at $T = 400$ K. A 1 ns constant *NPT* CG simulation gave an average side length of the cubic simulation box of 50.63 ± 0.08 Å, a relative error of 1.3% with respect to the average side length of 50.00 ± 0.05 Å determined from all-atom MD simulation. Some of the RDFs from all-atom and CG simulations are compared in Fig. 12. The RDFs for A–A, A–D, D–D, and E–E sites obtained from the CG MD simulations have lost some detailed structure, and the peak positions have shifted slightly to larger distances. The most significant errors appear in the RDF for D–D sites. These errors may be expected to arise due to the planar structures of the A and D sites and the strong correlation of the relative orientations between charged groups in the ionic liquid system.

The EF-CG force field for the $EMIM^+/NO_3^-$ ionic liquid obtained at $T = 400$ K was also used to simulate the same system with 512 ion pairs at $T = 700$ K. The equilibrium side length of the cubic box in the CG run is 53.11 ± 0.16 Å, which represents a relative error of 6.6% from the equilibrium box length of 52.76 ± 0.12 Å measured in the atomistic simulations. Some of the RDFs computed in the all-atom and the EF-CG MD simulations at $T = 700$ K are compared in Fig. 13. The CG and all-atom RDFs agree even better at $T = 700$ K than those at $T = 400$ K shown in Fig. 12. This might be attributed to the fact that, at such a high temperature, the atomic groups, especially the cationic rings and the anions, have more uniform relative orientation distributions. Because the EF-CG force fields are transferable between temperatures, the EF-CG force field parameterized at $T = 400$ K provides almost as accurate a treatment of the interactions of the ionic liquids at $T = 700$ K as an EF-CG force field parameterized at $T = 700$ K.

The example of the $EMIM^+/NO_3^-$ indicates that the EF-CG method also works well with complex fluids, such as ionic liquids, for which the molecules are comparably large and strong electrostatic interactions are present in the system. It also shows that the EF-CG method has good transferability between different temperatures.

V. Conclusions

In this paper it has been shown that the EF-CG method can obtain transferable CG force fields for liquid systems. The EF-CG force fields can reproduce both structural as well as thermodynamic properties of atomistic models with reasonable accuracy. Moreover, the errors resulting from the EF-CG treatment may be predicted by considering the symmetry of the coarse-grained atomic groups.

Coarse-graining approaches generally effectively integrate-out less-important atomic details and thereby enhance the computational efficiency substantially. Both the MS-CG and the EF-CG methods take advantage of the reduced molecular description, larger timesteps, and faster diffusion of CG models to simulate systems much more efficiently than is possible in all-atom MD simulations. However, all CG methods average over high-resolution details, and so can only reproduce certain physical properties with satisfactory accuracy. The EF-CG method was designed to capture the transferable part of the atomistic force fields. Therefore, it is especially suitable in situations when the transferability, for example, between different system sizes, between different temperatures, and between bulk/surface systems, is more important than the detailed accuracy of the structural properties for any given set of conditions.

The simplicity of the EF-CG approach is a considerable advantage of the method. As discussed previously and also described in the ESI,† the MS-CG method requires the[21] solution of high-dimensional linear least squares problem, which can become a challenging numerical task for complex systems. Moreover, because the MS-CG method provides an optimal approximation to a many-body PMF,[20] extensive sampling is required to accurately determine the statistical correlations between the many interactions included in the potential energy function for a complex system. In contrast, because the EF-CG method provides a mean-field approximation to the many-body PMF, so that each pair interaction in the CG force field is calculated separately by simply evaluating the correlation function in eqn (4). The EF-CG method does not require the solution of a high dimensional and potentially ill-conditioned matrix problem and also does not require accurate determination of the statistical correlation between various interactions (albeit at the potential risk of reduced accuracy in the CG force field for a given thermodynamic condition).

The simplicity of the EF-CG method also allows for a relatively simple interpretation and manipulation of the results. The considerations discussed in section II clearly guide the construction of CG mappings and also allow intuitive insight into the errors implicit in the method. Moreover, the EF-CG

pair forces can be readily decomposed into contributions from, *e.g.*, electrostatic or VDW interactions. This simple decomposition suggests that particular contributions to the EF-CG pair forces between a pair of site types can be separately treated and manipulated to predict the EF-CG pair forces between slightly different types of site types in slightly different environments. In fact, recent work has demonstrated that this approach provides a reasonably accurate and highly efficient approach for studying the complex properties of various ionic liquid systems.[34] It is anticipated that this method will continue to be particularly useful for especially large and complex systems for which more accurate methods, such as the MS-CG method, may be difficult to implement numerically.

The thermodynamic properties and the transferability of the EF-CG and MS-CG models for neopentane, methanol and the ionic liquid system are further compared in the ESI.† It is seen, for example, that MS-CG models constructed in bulk cannot be readily applied to perform interface simulations (*e.g.*, calculate the surface tension) without additional modifications. Somewhat surprisingly, the calculations in the ESI demonstrate that both the MS-CG and EF-CG force fields may have reasonable temperature transferability. One motivation for the EF-CG method is that, by simplifying the averaging involved in calculating the CG force field, the EF-CG force field should be less temperature dependent, and thus it should exhibit greater temperature transferability than the MS-CG force field. However, while this is the case sometimes it is not always the case for all of the examples studied. It is possible that the comparable temperature transferability of the two methods may result from the errors intrinsic to the EF-CG method. Regardless, the temperature transferability of each CG method likely depends strongly upon the system studied, the temperature range, and also the choice of mapping from atomistic to CG coordinates.

Besides having reduced accuracy for structural properties, the EF-CG method, which has been developed as a complement to the general MS-CG approach, also has several other limitations that may be addressed in future work. First of all, the EF-CG method can only be applied to coarse-grain atomistic force fields. In contrast the general MS-CG approach can be applied to many systems regardless of the origin of the underlying forces. In comparison to the variational MS-CG method, the EF-CG method relies upon the center-of-mass mapping and treats bonded interactions separately from non-bonded interactions. The EF-CG method cannot effectively decompose many-body interactions into pairwise interactions, as the MS-CG approach does. Consequently, the EF-CG method cannot incorporate the many-body dispersion interactions that may be important for modeling, *e.g.*, polarizable nanoclusters.[35] Finally, in the MS-CG approach, it is possible to eliminate some atoms or entire molecules from the coarse-grained description, while the EF-CG method requires all underlying atoms to be represented explicitly at the CG level.

The central pairwise approximation for the CG forcefield limits the accuracy of both the EF-CG and the MS-CG approaches. However, it may be possible to add orientational degrees of freedom by fixing a local coordinate system on each CG site. In MD simulations, explicit computations of torques are necessary to propagate the orientations. This complexity can be avoided in Monte Carlo simulations, when the local coordinates are rotated randomly. In either case, the computation of non-central interactions is much more demanding than the calculation of central pairwise interactions in CG models. Therefore, CG simulations considering site orientation are only useful when the orientational information is crucial for the target problem.

A promising application of the EF-CG method is for the systematic design of materials. Two examples are the design of ionic liquids to meet customized requirements, such as given viscosity and melting point, and the design of biomaterials to form designated shapes by self-assembly processes. In these applications, the materials must be constructed by developing appropriate building blocks. Thus the transferability and additivity of the effective CG forces are essential.

It may also be helpful to combine the EF-CG and MS-CG methods together. One of the largest challenges in biomolecular simulations is that much of the computational expense arises from modeling a large amount of water, even though the detailed properties of the water molecules are not necessary for many applications. It is possible to eliminate water molecules at the CG level by using the MS-CG method. On the other hand, in some applications, such as the systematic design of biomolecules, the transferability and additivity of CG sites corresponding to biomolecules are necessary. The EF-CG method is suitable for building such CG force fields. By combining the EF-CG and MS-CG approaches together, the effective CG forces may have both good transferability and increased efficiency. This research direction is currently being pursued in our group.

Acknowledgements

This research was supported by the Air Force Office of Scientific Research. W. G. N. acknowledges funding from the National Institutes of Health through a Ruth L. Kirschstein National Research Service Award postdoctoral fellowship (Grant No. 5F32FM076839-02). The authors thank Dr Vinod Krishna, Dr Gary Ayton, and Dr Sergei Izvekov for useful discussions. Allocations of computer time from the XT3 supercomputer allocated in the Pittsburgh Supercomputing Center, through the support of the National Science Foundation, and the Lonestar supercomputer allocated in the Texas Advanced Computing Center are gratefully acknowledged.

References

1 W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *J. Am. Chem. Soc.*, 1995, **117**, 5179.
2 W. L. Jorgensen, D. S. Maxwell and J. Tirado-Rives, *J. Am. Chem. Soc.*, 1996, **118**, 11225.
3 B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, *J. Comput. Chem.*, 1983, **4**, 187.
4 *Coarse-Graining of Condensed Phase and Biomolecular Systems*, ed. G. A. Voth, CRC Press, Boca Raton, 2009.
5 G. S. Ayton, W. G. Noid and G. A. Voth, *Curr. Opin. Struct. Biol.*, 2007, **17**, 192–198.
6 F. Müller-Plathe, *ChemPhysChem*, 2002, **3**, 754.
7 S. O. Nielsen, C. F. Lopez, G. Srinivas and M. L. Klein, *J. Phys.: Condens. Matter*, 2004, **16**, R481.

 This journal is © the Owner Societies 2009

8 V. Tozzini, *Curr. Opin. Struct. Biol.*, 2005, **15**, 144.
9 A. P. Lyubartsev, M. Karttunen, I. Vattulainen and A. Laaksonen, *Soft Mater.*, 2003, **1**, 121.
10 A. P. Lyubartsev and A. Laaksonen, *Phys. Rev. E*, 1995, **52**, 3730.
11 S. Izvekov and G. A. Voth, *J. Phys. Chem. B*, 2005, **109**, 2469.
12 S. Izvekov and G. A. Voth, *J. Chem. Phys.*, 2005, **123**, 134105.
13 S. Izvekov, M. Parrinello, C. J. Burnham and G. A. Voth, *J. Chem. Phys.*, 2004, **120**, 10896.
14 S. Izvekov and G. A. Voth, *J. Phys. Chem. B*, 2005, **109**, 6573.
15 Y. Wang, S. Izvekov, T. Yan and G. A. Voth, *J. Phys. Chem. B*, 2006, **110**, 3564.
16 S. Izvekov and G. A. Voth, *J. Chem. Theory Comput.*, 2006, **2**, 637.
17 J. Zhou, I. F. Thorpe, S. Izvekov and G. A. Voth, *Biophys J.*, 2007, **92**, 4289–4303.
18 I. F. Thorpe, J. Zhou and G. A. Voth, *J. Phys. Chem. B*, 2008, **112**, 13079–13090.
19 Q. Shi, S. Izvekov and G. A. Voth, *J. Phys. Chem. B*, 2006, **110**, 15045.
20 W. G. Noid, J.-W. Chu, G. S. Ayton, V. Krishna, S. Izvekov, G. A. Voth, A. Das and H. C. Andersen, *J. Chem. Phys.*, 2008, **128**, 244114.
21 W. G. Noid, P. Liu, Y. Wang, J.-W. Chu, G. S. Ayton, S. Izvekov, H. C. Andersen and G. A. Voth, *J. Chem. Phys.*, 2008, **128**, 244115.
22 W. G. Noid, J.-W. Chu, G. S. Ayton and G. A. Voth, *J. Phys. Chem. B*, 2007, **111**, 4116–4127.
23 J.-P. Hansen and I. R. McDonald, *The Theory of Simple Liquids*, Academic Press, San Diego, 1986.
24 M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*, Clarendon Press, Oxford, 1987.
25 G. S. Ayton, H. L. Tepper, D. T. Mirijanian and G. A. Voth, *J. Chem. Phys.*, 2004, **120**, 4074.
26 T. R. Forester and W. Smith, *DL_POLY user manual*, CCLRC, Daresbury Laboratory, Daresbury, Warrington UK, 1995.
27 W. G. Hoover, *Phys. Rev. A*, 1985, **31**, 1695.
28 S. Melchionna, G. Ciccotti and B. L. Holian, *Mol. Phys.*, 1993, **78**, 533.
29 P. Wasserscheid and W. Keim, *Angew. Chem., Int. Ed.*, 2000, **39**, 3772.
30 T. Welton, *Chem. Rev.*, 1999, **99**, 2071.
31 R. D. Rogers and K. R. Seddon, *Science*, 2003, **302**, 792.
32 R. P. Singh, R. D. Verma, D. T. Meshri and J. M. Shreeve, *Angew. Chem., Int. Ed.*, 2006, **45**, 3584.
33 T. Yan, C. J. Burnham, M. G. D. Pópolo and G. A. Voth, *J. Phys. Chem. B*, 2004, **108**, 11877.
34 Y. Wang, S. Feng and G. A. Voth, *J. Chem. Theory Comput.*, 2009, in press.
35 H.-Y. Kim, J. O. Sofo, D. Velegol, M. W. Cole and A. Lucas, *Langmuir*, 2007, **23**, 1735–1740.