

Coarse grain to atomistic mapping algorithm: a tool for multiscale simulations

Steven O. Nielsen¹, Bernd Ensing², Preston B. Moore³, and Michael L. Klein⁴

¹Department of Chemistry, University of Texas at Dallas, 2601 North Floyd Road, Richardson, TX 75083-0688

²Department of Chemistry and Applied Biosciences, ETH Zurich USI-Campus, Via Giuseppe Buffi 13, Lugano, CH-6900 Switzerland

³Department of Chemistry and Biochemistry, University of the Sciences in Philadelphia, Philadelphia, PA 19104

⁴Center for Molecular Modeling, Department of Chemistry, University of Pennsylvania, Philadelphia PA 19104-6323

email: steven.nielsen@utdallas.edu

Abstract

A key component to the success of multiscale modeling approaches is the ability to switch between molecular representations involving different levels of detail. In particular, the transition from a coarse to a fine representation is challenging because it requires the creation of information. To address this challenge, an algorithm is presented to fill in the detail when a coarse grain representation of a molecular system is replaced by an atomistic representation. The algorithm consists of minimization on the Lie group $SO(3)$ for every coarse grain site, using the components of the atomistic force field that operate between atoms belonging to different coarse grain sites. This method is maximally efficient because the optimization is done at the coarse level using frozen atomistic library structures corresponding to each kind of coarse grain unit. The algorithm can be applied to any system since its input requirements are simply the force fields and molecular structures of both levels. The role of the algorithm in dual-resolution multiscale simulation methodology is discussed, and the efficacy of the algorithm is demonstrated through its implementation to liquid dodecane.

1. Introduction

The many approaches of modeling chemical and biological systems at the mesoscopic level, which include coarse grain (CG) dynamics, dissipative particle dynamics (DPD), and Brownian dynamics, are increasingly being linked to the underlying molecular description. Historically, these models used a description of the molecular interactions which was rather simple and arbitrary, in part for computational efficiency but also because of limited knowledge concerning the molecular forces. In the past few years, various groups have developed methods to obtain the potential energy terms in these models from either experimental data or from fully atomistic simulation (FA) data. We refer the interested reader to two review articles^{1,2} on this subject and briefly give some examples. Ortiz *et. al.*³ have developed a DPD model for aqueous diblock copolymer systems using experimental surface tension and density data together with computer simulation data – specifically the first and second moments of the bond and angle distributions obtained from atomistic molecular dynamics (MD) simulations. Building the model required the introduction of a density-based atomistic-to-coarse-grain mapping in order to obtain a physically realistic description of the system. This model gave values for the bilayer membrane area expansion modulus and the power-law scaling of the hydrophobic core thickness that were in excellent agreement with experiment. Izvekov *et. al.*⁴ have developed a force-matching method in which the coarse graining of an interparticle force field is constructed directly from the forces obtained from an atomistic MD trajectory. The procedure has been demonstrated for water and methanol⁴, nanoparticles⁵, and phospholipids⁶.

The motivation for making these connections to the molecular details is to obtain more realistic models, in an effort to describe actual systems. The question that arises naturally, then, is how

faithful these models are to the molecular system they are supposed to represent. There are several ways to answer this question, and in doing so attempt to validate the results of a coarse grain or mesoscale study.

Mesoscale methods are employed in situations where treating the full molecular detail is too costly, so validation attempts are only feasible if limited molecular simulations are required. The final configuration from a CG simulation, e.g. of a self-assembly study, could be restored to its full molecular detail, and the resulting system could be simulated to see if this configuration is at least locally stable. A one-dimensional reaction coordinate identified from CG simulations could be targeted for a free energy simulation in full molecular detail, and the free energies and conformations could be compared along the pathway. Such validations of CG results are almost non-existent in the literature. The algorithm presented here is a tool that allows for the molecular detail to be recovered in a straightforward manner. It is hoped that this tool will encourage people to attempt to validate, or at least compare and contrast, their mesoscopic results with limited molecular simulations.

Another use of this algorithm is in atomistic simulations of high molecular weight polymer melts, which typically combine an atomistic to coarse grain, and a coarse grain to atomistic, mapping algorithm to prepare a relaxed initial condition⁷. The long time scales needed to relax a glassy long-chain polymer melt put the construction of an initial condition out of the realm of atomistic simulation. Employing a coarse grain model speeds up this process by many orders of magnitude, and once the chains are relaxed the atomistic details can be reintroduced. Then, only a short equilibration simulation is needed to relax the local degrees of freedom.

Recent developments in multiscale simulation methodology also demand such an algorithm.

Hybrid dual-resolution MD is the type of multiscale simulation in which the $SO(3)$ optimization

technique is particularly advantageous. Recently, a number of groups have made technical advances to couple a molecular system modeled in full atomistic detail (high resolution) with a molecular system modeled at a lower, coarse grain or united atom, level (low resolution)^{8,9,10}. Some of these methods allow matter to dynamically change resolution when it crossed a user-defined boundary. Such a change requires algorithms to convert matter from a coarse to a fine level of representation, and vice versa.

The algorithm presented in this paper uses $SO(3)$ optimization to align molecular fragments corresponding to coarse grain sites. In other words, rigid library structures, obtained from atomistic simulated annealing simulations, are placed at each coarse grain site and rotated in place to minimize an energy function. At the end of this procedure, a globally unfrustrated configuration is obtained. However, local relaxation still needs to be performed on the degrees of freedom which were frozen (the degrees of freedom within each fragment). This is accomplished, for instance, by running a short atomistic simulation.

The approach is based on an algorithm developed by Taylor and Kriegman¹¹ in which a sequence of local parameterizations of the manifold $SO(3)$ is used, rather than relying on a single global parameterization such as the Euler angles. The problems caused by singularities in a global parameterization are thus avoided.

The efficiency of the algorithm comes from two sources, one technical and one conceptual. On the technical side, the use of quaternion arithmetic and the structure of the algorithm make it fast. From a conceptual standpoint, the optimization is performed at the coarse grain level because the molecular fragments corresponding to coarse grain sites are treated as rigid bodies with no internal degrees of freedom. This means that there are only three degrees of freedom per coarse grain site to optimize, namely an element of $SO(3)$. The algorithm is maximally efficient in the

sense that the number of variables to optimize is minimal, and the algorithm for this optimization is extremely fast.

The output from this algorithm consists of rigid molecular fragments positioned at the coarse grain sites, rotated to minimize an energy function consisting of both intra- and inter-molecular terms. In what sense is this configuration a representation of the underlying molecular system? If one insisted that the atomistic configuration be a realization of a particular ensemble (NVT for example), the configuration generated by the algorithm would not be satisfactory. Rather than attempting to provide a mapping algorithm that generates an equilibrated atomistic configuration, we have taken a more pragmatic approach. The mapping algorithm described here quickly generates a globally stable atomistic configuration that further requires very localized relaxation and equilibration. In the context of dual-resolution multiscale simulations, this localized equilibration is inherent in the simulation methodology in the form of a healing region (see the following section) and hence is not required of the mapping algorithm. For wholesale mapping, in which a coarse grain configuration is entirely mapping to its atomic detail, a short atomistic equilibration simulation needs to be performed to both relax the system locally and to equilibrate it in a particular ensemble. For the later case, this two-step procedure is expected to be competitive with one-step algorithms.

The remainder of the paper is organized as follows. First we provide some detail about hybrid dual-resolution methodology, in order to give the reader an appreciation of the utility of the $SO(3)$ algorithm in such a context. Second, the mathematics behind the $SO(3)$ optimization algorithm is introduced. Next, the algorithm is further developed for application to molecular systems.

Following this, bulk liquid dodecane is used as an illustrative example.

2. Hybrid coarse grain – fully atomistic molecular dynamics

In what follows, we briefly describe our implementation of a hybrid MD scheme to model a particular region of a system in atomistic detail, while treating everything outside this region at the computationally less demanding CG resolution. Further technical details will be published elsewhere (manuscript in preparation).

Using the hybrid CG/FA-MD technique, the molecules within a relatively small spatial region interact with each other at the FA level through an atomistic force field. The molecules outside this FA region are instead treated at a CG level, in which several connected atoms are lumped into CG interaction sites (e.g. chemical functional groups of about ten atoms), that interact with each other through a force field designed for CG MD.

It would be cumbersome to have to develop a third “mixed” force field to describe interactions between atoms and CG sites, and fortunately this is not necessary. Instead, the interaction sites of the molecules in the FA region, namely the atom positions, can easily be mapped onto the CG representation of the molecules, after which the mixed FA-CG interactions can be modeled using the same CG force field as used for the CG-CG interactions.

As an illustration, consider a system of methane molecules, partly treated atomistically with the carbon and the four hydrogens as the interaction sites and outside this FA region treated at the united atom level using a single interaction site per CH₄ molecule (see Fig. 1). To map the atomistic representation onto the CG representation, we simply take the center of mass of the atoms of each molecule as the position of the CG site. In fact, we can maintain access to both FA and CG representations of the entire system throughout the simulation, by setting the system up in the FA representation and mapping it onto the CG representation by computing centers of mass. Molecules in the FA region are then evolved in their FA representation, while the

molecules outside the FA region are evolved in their CG representation. For the latter CG molecules, the extra atomistic degrees of freedom are kept frozen and follow their CG site in a static manner. For the mixed interaction between two molecules in the FA and CG regions respectively, we only need to re-compute the center of mass of the molecule in the FA region to know its CG site position and calculate the CG-CG interaction. Then, for the molecule in the FA region, the force resulting from this interaction is distributed in a mass-weighted fashion over its atoms. This way, the molecules in the FA-region “feel” the molecules in the outer CG-region in the averaged CG manner, and likewise the molecules in the CG-region “feel” the molecules in the FA-region as CG molecules.

Since the entire system is represented simultaneously in the FA and in the CG levels of detail throughout the simulation, the evaluation of the forces on each molecule and the evolution of each molecule can be chosen to proceed at either of the two resolutions and, moreover, can switch from one resolution to the other whenever necessary. The latter of course has to take place when a molecule crosses over from the FA region to the CG region, or vice versa.

In practice, however, using a “hard” FA-CG border between the regions causes problems. The atoms of a previously CG molecule can have any orientation and are likely to overlap with other atoms in the FA region when the relatively soft CG interaction is suddenly changed into the atomistic interactions, leading to too high forces and velocities. To overcome this problem an intermediate “healing region” (also called “interface region”, or “switching region”) is introduced in between the FA and CG spatial regions. Molecules in this healing region switch in a smooth fashion from one resolution into the other by treating their interaction partly FA and partly CG. Praprotnik *et. al.*⁸, for example, use force scaling in the healing region so that the total intermolecular force acting between the centers of mass of molecules α and β equals

$$F_{\alpha\beta} = [1 - S(X_\alpha)S(X_\beta)]F_{\alpha\beta}^{CG} + S(X_\alpha)S(X_\beta) \sum_{i\alpha, j\beta} F_{i\alpha, j\beta}^{AA}$$

where i and j are the atoms belonging to CG sites α and β respectively, and $S(X_\alpha)$ is a scaling function that switches smoothly from 0 (in the FA region) to 1 (in the CG region) as a function of the position X_α in the healing region. Note that the interaction between two molecules that find themselves both halfway into the healing region is treated one-quarter atomistically and three-quarters CG.

Our approach is similar, but instead of force scaling, we scale the potentials. The main advantage of scaling the potential is that it allows an estimate of the change in total energy due to the change of the potential when a molecule moves through the healing region.

$$V = \sum_{\alpha\beta} V_{\alpha\beta} + \sum_{i\alpha, j\alpha} V_{i\alpha, j\alpha}^{AA} + V^{\text{inh}}$$

$$V_{\alpha\beta} = S_{\alpha\beta} V_{\alpha\beta}^{CG} + [1 - S_{\alpha\beta}] \sum_{i\alpha, j\beta, \alpha \neq \beta} V_{i\alpha, j\beta}^{AA}$$

The total potential is divided into the pair-potentials spanning more than one CG site, $V_{\alpha\beta}$, the internal FA pair-potentials (bonds and angles) within each CG site, $V_{i\alpha, j\beta}$, and a remainder term V^{inh} that accounts for the excess inherent potential energy of CG molecules that is “integrated out” when moving from the atomistic to the CG representation. The interactions spanning two or more CG sites consist of sums of scaled FA-FA and CG-CG potentials (note the similarity with the first equation), while the second term of internal FA interactions within a single CG site are not scaled in the healing region, but instead are frozen when moving into the CG region.

For our methane example, this means that the intra-molecular bonds and angles are evolved atomistically in the healing region, while the inter-molecular van der Waals and electrostatic interactions are treated partly FA and partly CG, using the scaling factors, $S(r)$. The scaling factor is a polynomial that goes smoothly from zero to unity over a distance from R_1 to R_2

respectively. R_1 and R_2 can, for example, be taken to be the distance from a particular molecule that requires the atomistic treatment, so that a spherical FA region with radius R_1 is defined surrounding the molecule, dressed by a skin of thickness $R_2 - R_1$ defining the healing region, and with a CG region outside the sphere of radius R_2 . Alternatively, a rectangular spatial partitioning can be defined, for example the FA region can be an infinite slab with thickness $2 \times R_1$, flanked on both sides by a healing region of thickness $R_2 - R_1$, with the CG region outside of this. Figure 2 shows these two types of partitioning for a box of one thousand methane molecules.

Figure 1 illustrates the scaling of the interactions between pairs of methane molecules in different regions. Each molecule has a switching function value depending on its position, r , being zero in the FA region, unity in the CG region and an intermediate value in the healing region as shown by the solid black line in figure 1. Each pair interaction is scaled with the maximum scaling factor among the two interacting CG sites

$$S_{\alpha\beta} = \max[S(r_\alpha), S(r_\beta)].$$

Although the healing region overcomes the problem of bad overlaps by switching the atomistic degrees of freedom on in a smooth manner, nevertheless the excess potential energy is converted into kinetic energy that has to be removed, for example by using a thermostat. In particular, coupling the atomistic degrees of freedom to a thermostat is important for randomly oriented molecules moving from the CG region toward the FA region for two reasons: 1) allowing the internal degrees of freedom to thermalize into the desired ensemble and 2) to remove the heat produced during reorientation of the internal degrees of freedom from the energetically unfavorable random orientations assumed in the CG region. However, by using the $SO(3)$ optimization scheme, bad overlaps and energetically unfavorable random orientations in general are largely avoided at the cost of a relatively small overhead. Moreover, boundary effects are

expected to be reduced significantly, allowing for a smaller healing region, which reduces the number of interactions that have to be calculated in both representations.

The second important benefit of repairing the internal random orientations by applying the $SO(3)$ optimization scheme for the molecules in CG region is the expected improvement of structural continuity at the boundary for certain systems. Intuitively, this is most easily understood by considering a box of liquid water with the typical hydrogen bonding network in the FA region and a spherical molecular representation in the CG region. Without the $SO(3)$ optimization, FA water molecules close to the boundary with the CG region cannot seamlessly continue the H-bonded network and will behave as if at an interface to avoid having dangling hydrogens.

3. $SO(3)$ Optimization Algorithm

The algorithm outlined below is intended to minimize a real-valued objective function $E: SO(3) \rightarrow \mathbb{R}$ defined on the set of rotation matrices

$$R \in SO(3) \equiv \{R \in \mathbb{R}^{3 \times 3} : R^t R = I, \det(R) = 1\}.$$

At every point R_0 on the manifold $SO(3)$ we construct a continuous, differentiable mapping between a neighborhood of R_0 on the manifold and an open set in \mathbb{R}^3 ,

$$R(\omega) = R_0 \exp J(\omega), \quad \omega \in \mathbb{R}^3, \quad \|\omega\| < \pi$$

where the skew symmetric operator $J: \mathbb{R}^3 \rightarrow SO(3)$ is defined as

$$J(\omega) = \begin{bmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{bmatrix}.$$

$R(\omega)$ can be computed using the Rodrigues formula, although we will not need to do this. The objective (energy) function can be expanded to quadratic order about R_0 as

$$E(R(\omega)) = E(R_0) + g^t \omega + \omega^t H \omega$$

where g and H are the gradient and the Hessian of the function, respectively, evaluated at the point $\omega = 0$ which corresponds to the rotation matrix R_0 . The conjugate gradient incremental step is

$$\omega_s = -H^{-1} g .$$

This incremental step determines the new rotation matrix as follows:

$$R = R_0 \exp J(\omega_s).$$

The incremental step must lie within the range of the local parameterization, i.e. $\|\omega_s\| < \pi$.

The updating step can be made computationally efficient by representing the rotations by unit quaternions. The relationship between $SO(3)$ and the group of unit quaternions $Sp(1)$ is

$$q = (\cos \theta, \hat{\omega} \sin \theta) \quad , \theta = \|\omega\|/2 .$$

The incremental step corresponds to the quaternion

$$q_s = \left(\cos \frac{\theta}{2}, \omega \frac{\sin(\theta/2)}{\theta} \right) \quad \text{where} \quad \theta = \|\omega\| .$$

With the rotation R_0 expressed as the unit quaternion q_0 , the product of the two rotations, which gives the new rotation matrix, is given by the quaternion multiplication $q_0 q_s$.

It has been shown that this algorithm exhibits quadratic convergence provided that the starting point is sufficiently close to a minimum¹².

4. Algorithm specialized to molecular systems

To apply the algorithm outlined in the previous section to molecular systems, two things are needed. Firstly, an objective function must be chosen which imbues the algorithm with chemical

meaning. This function will provide a measure of the potential energy of the molecular configuration associated with a given rotation matrix. Secondly, the algorithm must be extended to many coupled $SO(3)$ optimizations so that the molecular system is simultaneously and concertedly optimized over all the coarse grain centers. This multi-body extension is in fact straightforward and does not incur any significant computational cost aside from the necessary linear scaling with the number of centers. The nature of the multi-body aspect of the algorithm will become clear in what follows.

Let us now address the choice of an energy function. Only interactions between atoms belonging to different coarse grain units need be considered. This is because the intra-unit degrees of freedom are frozen – the molecular fragment corresponding to a CG site is taken from a library structure obtained from a simulated annealing FA MD simulation. The $SO(3)$ algorithm is designed to find the optimal rotational orientation of each of these fragments, where the center of mass of each fragment is constrained to lie at the location of the CG site representing it, and where no internal relaxation of the intra-fragment degrees of freedom is allowed. This idea is schematically shown in Fig. 3.

We wish to take the functional form and the parameters of all of the terms contributing to the energy function from an underlying atomistic force field. There is no need to invent new potential energy terms when we have well-parameterized ones at our disposal.

The first contribution to the energy function is a bonded term arising from the unconnected bonds in the molecular fragments. This function is harmonic in the interatom distance and is expressed as (see Fig. 4)

$$E(R_1, R_2) = k / 2 (\|r + R_2 v - R_1 u\| - d_0)^2$$

where R_1 and R_2 are the rotation matrices corresponding to coarse grain units 1 and 2, located at

positions COM1 and COM2, respectively. The vector from COM1 to COM2 is denoted r . u represents the vector from COM1 to the atom in coarse grain unit 1 involved in the bond. v represents the vector from COM2 to the atom in coarse grain unit 2 at the other end of the bond. The bond has an equilibrium distance of d_0 and a force constant of k . In order to perform optimization, the gradient must be evaluated. There are six gradient terms associated with this energy function, three for rotation matrix R_1 and three for R_2 . For the ω_x component of R_1 , the gradient is

$$\frac{\partial}{\partial \omega_{x1}} E = -k \frac{(\|r + R_2 v - R_1 u\| - d_0)}{\|r + R_2 v - R_1 u\|} (r + R_2 v + R_1 u) \bullet \frac{\partial}{\partial \omega_{x1}} R_1 u$$

with

$$\frac{\partial}{\partial \omega_{x1}} R_1 u = R_0^1 J(\hat{x}) u$$

where R_0^1 denotes the initial rotation matrix for coarse grain center 1. This last result is computationally important because it means that the J operator only ever acts on three elements, \hat{x} , \hat{y} , and \hat{z} , where,

$$J(\hat{x}) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}, \quad J(\hat{y}) = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}, \quad J(\hat{z}) = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

The atoms in the fragment library which have unconnected bonds can be flagged to help automate the optimization procedure.

A bend term in the potential energy function has the form

$$E(R_1, R_2) = k / 2 (\theta - \theta_0)^2$$

where (see Fig. 5)

$$\theta = \arccos \frac{(R_1 u' - R_1 u) \cdot (r + R_2 v - R_1 u)}{\|R_1 u' - R_1 u\| \|r + R_2 v - R_1 u\|}$$

For concreteness, we have assumed that two of the atoms involved in the bend are located within fragment 1, with the remaining atom in fragment 2. The gradient is straightforward to evaluate by successive applications of the chain rule. The remaining intra-molecular potential energy terms are treated in a similar manner and their gradients are again evaluated by the chain rule.

For example, a torsional contribution, assuming for concreteness, that two of atoms are in fragment 1 and two are in fragment 2, is constructed as follows. In analogy with Fig. 5, we denote the position of the four atoms with respect to their fragment centers of mass as u, u', v, v' and form the four vectors $b_1 = u', b_2 = u, b_3 = r + v, b_4 = r + v'$. We then define three new vectors $a_1 = b_1 - b_2, a_2 = b_3 - b_2, a_3 = b_4 - b_2$, from which the torsion angle is formed using the cross product as

$$\tau = \arccos \left[\frac{(a_1 \times a_2) \cdot (a_2 \times a_3)}{\|a_1 \times a_2\| \|a_2 \times a_3\|} \right].$$

Inter-molecular interactions, and intra-molecular interactions between atoms separated by more than three bonds, are critical for condensed phase systems and must be included in the energy function. Since the crucial contribution of these nonbonded terms is of an excluded volume nature, we take the repulsive part of the Lennard-Jones contribution into the energy function.

The multi-body nature of the minimization procedure proceeds as follows. The potential energy function is written as a sum over all the coarse grain sites, with separate terms arising from bonds, bends, torsions, one-fours, and Lennard-Jones interactions. The gradient of this function is evaluated with respect to each of degree of freedom, namely the three numbers $\omega_x, \omega_y,$ and ω_z for each coarse grain site. This gradient is used to decide upon a global incremental update step

in which all the rotation matrices are simultaneous changed.

According to the previous section, we are supposed to evaluate the Hessian as well as the gradient to compute the update step. However, by employing the Fletcher-Reeves-Polak-Ribiere version of the conjugate gradient algorithm, only the gradient is needed¹³.

5. Algorithm specialized to liquid dodecane

We now proceed to demonstrate how the $SO(3)$ optimization algorithm works on a specific condensed phase system, namely bulk liquid dodecane. The coarse grain representation we adopt for dodecane has four units¹⁴, with each unit composed of three carbon atoms and their associated hydrogen atoms. Two of these units include six hydrogen atoms, and two of them include seven hydrogen atoms. A distinction is made between these cases, giving two distinct types of coarse grain units, called middle (M) and terminal (T). An equilibrated fully atomistic bulk dodecane system at 300 K and 1 bar is subjected to simulated annealing runs in which the temperature is reduced to zero. From this final configuration, one atomic fragment corresponding to M and one corresponding to T are selected, and their atomic coordinates are entered in the fragment library after rigid translation to place each fragment center of mass at the origin.

The energy landscape becomes rougher and the $SO(3)$ optimization problem becomes more and more overdetermined as more terms are included in the energy function. We choose the energy function for liquid dodecane to consist of one bond, four bends, four torsions, and four 1-4s per "join" between intra-molecular CG sites; and all Lennard-Jones repulsions between hydrogen atoms. The potential energy terms corresponding to these contributions are taken from the CHARMM atomistic force field. This selection is schematically shown in Fig. 6.

5.1) *Single configuration*

A single configuration is selected from an equilibrated CG bulk liquid dodecane simulation taken from reference 14. There are 1050 dodecane molecules in the simulation unit cell, at density of 0.74 g/mL. The goal of the $SO(3)$ optimization algorithm is to replace the coarse grain configuration with its atomistic counterpart in a low-energy state. Initially, every CG site is replaced with its fragment library structure, placed so that the center of mass of the fragment is at the CG site location. The minimization steps rotate these fragments about their centers of mass to align the unconnected bonds in the fragments with each other, to remove overlapping nonbonded contacts, and in general to find a favorable energy state for the atomistic system. Since the initial condition is far from a local minimum, the conjugate gradient method is of little use – a steepest descent method is performed in stages as follows. Local minima are plentiful when the full potential energy function is used, and with a bad initial condition a nonbonded contact could override an unconnected bond in terms of directing the motion of the atomistic fragments. To avoid this problem, the energy landscape is refined in two stages. Initially, only the bond and bend terms are included. Optimizing on this smooth energy landscape ensures that all the bonds are connected properly. After convergence, the torsions, one-fours, and Lennard-Jones interactions are turned on and the optimization algorithm is run again. The final configuration is analyzed with the full atomistic field to see how the potential energy is distributed. Using the equipartition theorem, we would expect to have an energy of $1/2k_B T$ per degree of freedom. Solving for T gives the result shown in Table 1 for the intra-molecular degrees of freedom. The bends are too energetic due to an overdetermined set of equations, but these relax immediately upon running a thermostated simulation.

Bonds	T = 294 K
-------	-----------

Bends	T = 1125 K
Torsions	T = 75 K
One-fours	T = 97 K

Table 1. Temperature of the $SO(3)$ optimized structure obtained using the equipartition theorem.

5.2) Incremental conjugate gradient optimization

Once this single configuration has been optimized, we have the option of performing incremental updates on the subsequent configurations under the coarse grain dynamics. Since the CG time step is much larger than the time step used at the atomistic level, the optimized atomistic fragments from the previous CG configuration are no longer appropriate for the updated CG coordinates. However, the conjugate gradient algorithm should work well because the new fragment orientations are not expected to be very different from the ones optimized to the previous CG coordinates. The conjugate gradient algorithm is used with an arbitrary, but very small, convergence tolerance for one hundred consecutive CG coordinates from a CG MD trajectory. The results are summarized in Figs. 7 and 8. Fig. 7 shows the objective energy function value plotted for each instance it is evaluated during the conjugate gradient $SO(3)$ optimization procedure. There are two features worth noticing. First, the conjugate gradient algorithm quickly brings the atomic fragments into line with each new set of CG coordinates. Second, the baseline energy is not constant, but rather smoothly varies with each new set of CG coordinates. This occurs because, for example, the CG bonds are flexible (they have a harmonic form) so that the FA bond lengths of the unconnected fragment bonds, with the center of mass of the atomic fragments fixed to the CG sites, cannot always achieve their minimum energy value. Fig. 8 shows a histogram of the number of conjugate gradient steps needed to converge the

objective function for each of the one hundred sets of CG coordinates. It is seen that roughly ten steps are needed on average.

The final configuration is further analyzed by applied the equipartition theorem to it against the full atomistic force field. The result, shown in Table 2, is similar to that of Table 1.

Bonds	T = 316 K
Bends	T = 1002 K
Torsions	T = 79 K
One-fours	T = 247 K

Table 2. Temperature of the $SO(3)$ optimized structure obtained using the equipartition theorem.

6. Conclusions

The coarse grain to atomistic mapping algorithm described in this manuscript, which is based on $SO(3)$ optimization, is seen to be a mathematically sound, algorithmically efficient, and practical method for recovering the atomistic detail given a reduced level of representation of a molecular system. This algorithm is expected to be of assistance in efforts to validate coarse grain models, a direction that will become increasingly important if coarse grain simulations are intended to model specific chemical systems. Furthermore, this algorithm will play a central role in our efforts to develop hybrid dual-resolution multiscale simulation methodology, in which atomistic and coarse grain degrees of freedom coexist in the simulation cell.

Bibliography

(1) Nielsen, S. O.; Lopez, C. F.; Srinivas, G.; Klein, M. L. *J. Phys.: Condens. Matter* **2004**, *16*, R481.

- (2) Ortiz, V.; Nielsen, S. O.; Klein, M. L.; Discher, D. E. *J. Polym. Sci. Part B: Polym. Phys.* **2006**, *in press*
- (3) Ortiz, V.; Nielsen, S. O.; Discher, D. E.; Klein, M. L.; Lipowsky, R.; Shillcock, J. J. *Phys. Chem. B* **2005**, *109*, 17708.
- (4) Izvekov, S.; Voth, G. A. *J. Chem. Phys.* **2005**, *123*, 134105.
- (5) Izvekov, S.; Violi, A.; Voth, G. A. *J. Phys. Chem. B* **2005**, *109*, 17019.
- (6) Izvekov, S.; Voth, G. A. *J. Phys. Chem. B* **2005**, *109*, 2469.
- (7) Queyroy, S.; Neyertz, S.; Brown, D.; Muller-Plathe, F. *Macromolecules* **2004**, *37*, 7338.
- (8) Praprotnik, M.; Delle Site, L.; Kremer, K. *J. Chem. Phys.* **2005**, *123*, 224106.
- (9) Neri, M.; Anselmi, C.; Cascella, M.; Maritan, A.; Carloni, P. *Phys. Rev. Lett.* **2005**, *95*, 218102.
- (10) Abrams, C. F. *J. Chem. Phys.* **2005**, *123*, 234101.
- (11) Taylor, C. J.; Kriegman, D. J. *Minimization on the Lie group $SO(3)$ and related manifolds*, Tech. Rep. 9405, Center for Systems Science, Dept. of Electrical Engineering, Yale University, New Haven, CT, April 1994.
- (12) Smith, S. *Geometric Optimization Methods for Adaptive Filtering*, PhD thesis, Harvard University, Division of Applied Sciences, Cambridge MA, September 1993.
- (13) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes*; Cambridge University Press: New York, 1992.
- (14) Nielsen, S. O.; Lopez, C. F.; Srinivas, G.; Klein, M. L. *J. Chem. Phys.* **2003**, *119*, 7043.

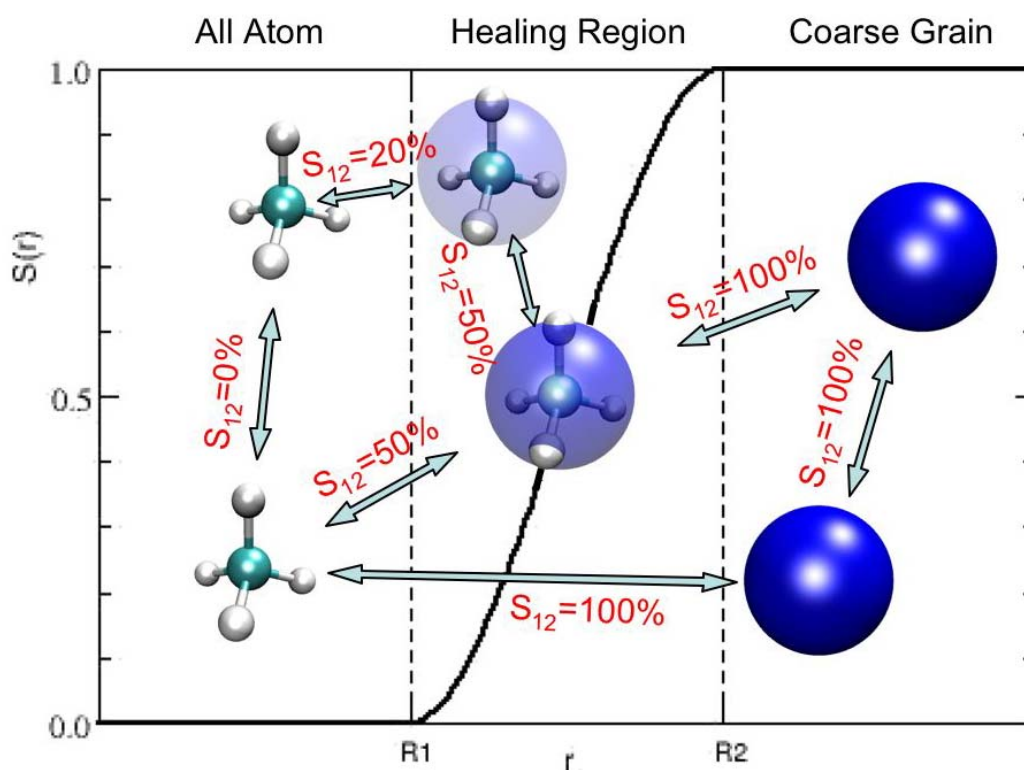


Figure 1. Schematic representation of methane in the atomistic region (at the left), the coarse grain region (at the right), and the connecting healing region (in the middle). The solid black line shows the value of the switching function for a CG site as a function of its position, $S(r)$. The fractional CG character of each pair interaction equals the maximum $S(r)$ among the two interacting CG sites (illustrated in red).

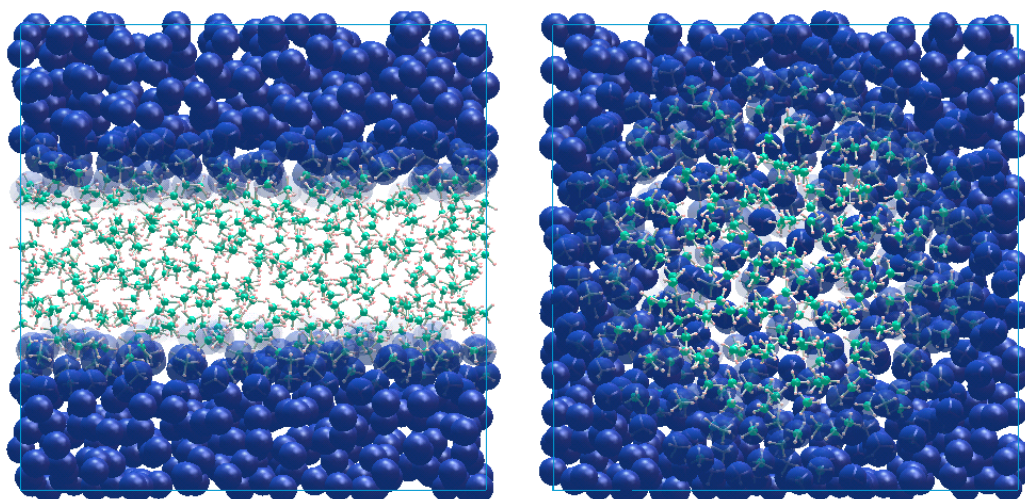


Figure 2. Two examples of the spatial partitioning of a periodic unit cell of 1000 methane molecules into an atomistic region, a coarse grain region, and an intermediate healing region. The transparency of the dark blue CG spheres reflects the values of the switching function $S(r)$. On the left is a rectangular partitioning, and on the right is a spherical partitioning.

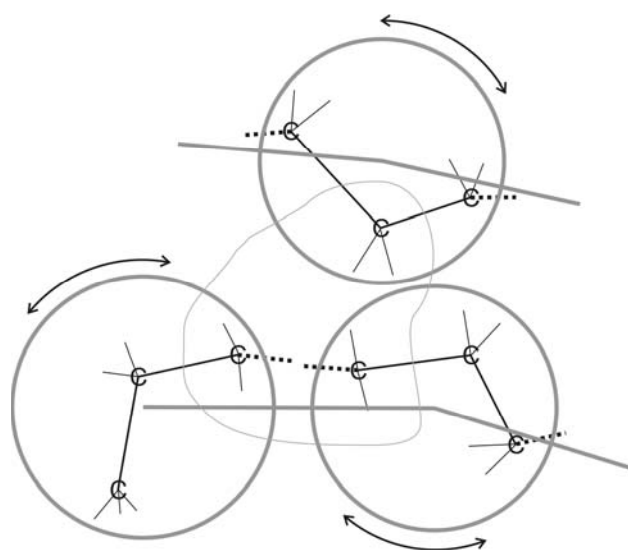


Figure 3. Schematic of the $SO(3)$ optimization algorithm. Atomic fragments are rotated about the centers of mass of the coarse grain units representing them to align the unconnected bonds between fragments, as well as to achieve a low energy state for the other intra- and inter-

molecular interactions at the atomistic level.

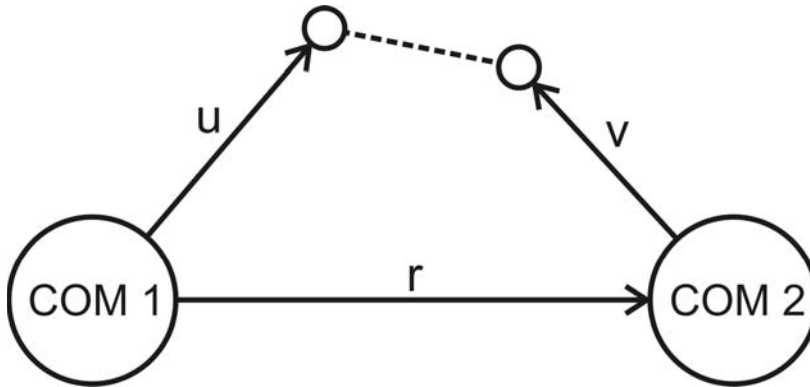


Figure 4. Bond (dotted line) between atomistic fragments corresponding to two neighboring coarse grain sites. Atom u in fragment 1, centered at the coarse grain site $COM1$, interacts with atom v , centered at the coarse grain site $COM2$, through a bonded term in the objective energy function.

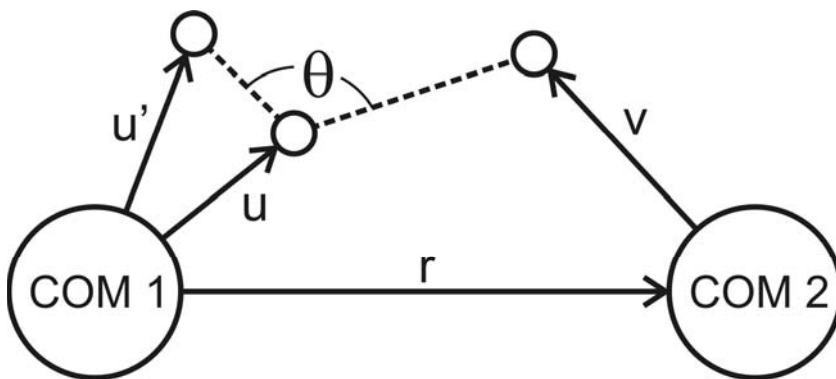


Figure 5. Bend (labeled θ) between atomistic fragments corresponding to two neighboring coarse grain sites. Atoms u and u' in fragment 1, centered at the coarse grain site $COM1$, interact with atom v , centered at the coarse grain site $COM2$, through a bend term in the objective energy function.

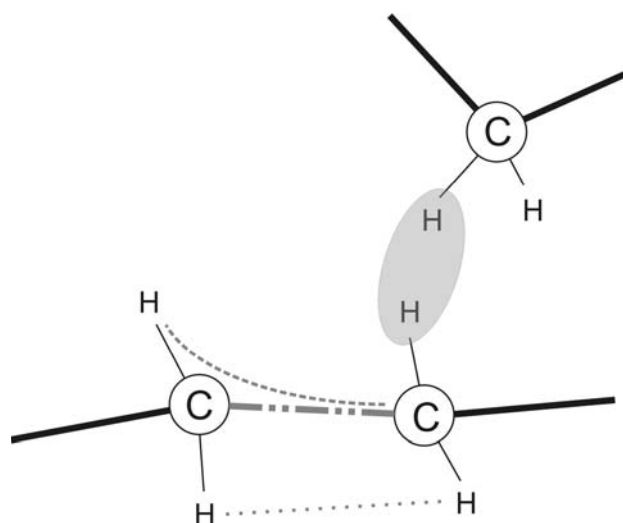


Figure 6. A portion of Fig. 3 is expanded to show some of the terms in the objective energy function. Bonds, bends, torsions, and one-fours are included involving the two carbon atoms belonging to different coarse grain units. Lennard-Jones repulsions between all pairs of hydrogen atoms are also included.

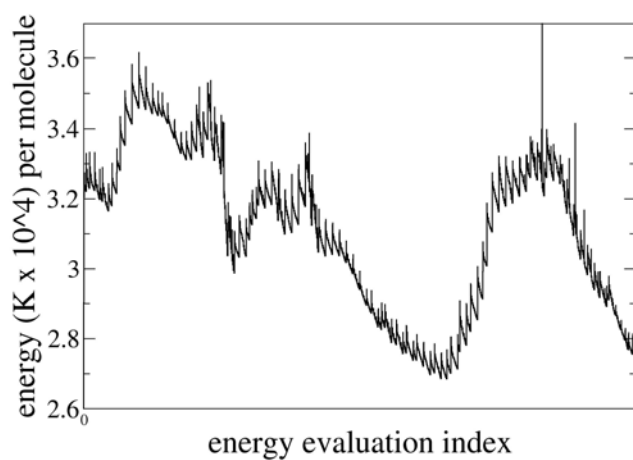


Figure 7. The objective energy function is plotted during the operation of the $SO(3)$ conjugate gradient algorithm acting on one hundred consecutive coordinates from a coarse grain molecular dynamics trajectory of 1050 dodecane molecules at a density of 0.74 g/mL, a temperature of 300 K, and a pressure of 1 bar.

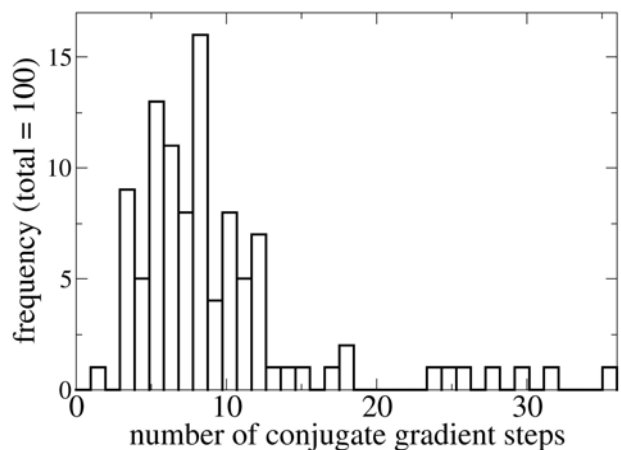


Figure 8. Histogram of the number of conjugate gradient steps needed for the convergence shown in Fig. 7 for one hundred consecutive coordinates from a coarse grain molecular dynamics trajectory. The convergence tolerance is set to be an arbitrary, but very small, fixed number.