Convergence of the Many-Body Expansion for Energy and Forces for Classical Polarizable Models in the Condensed Phase

Omar Demerdash[1] and Teresa Head-Gordon[1,2,3,4*]

[1]Department of Chemistry, [2]Department of Bioengineering, [3]Department of Chemical and Biomolecular Engineering, [4]Chemical Sciences Division, Lawrence Berkeley National Laboratory
University of California, Berkeley, CA 94720

We analyze convergence of energies and forces for the AMOEBA classical polarizable model when evaluated as a many-body expansion (MBE) against the corresponding N-body parent potential in the context of a condensed-phase water simulation. This is in contrast to most MBE formulations based on quantum mechanics, which focus only on convergence of energies for gas-phase clusters. Using a single water molecule as a definition of a body, we find that truncation of the MBE at 3rd order, 3-AMOEBA, captures direct polarization exactly and yields apparent good convergence of the mutual polarization energy. However it renders large errors in the magnitude of polarization forces and requires at least 4th order terms in the MBE to converge toward the parent potential gradient values. We can improve the convergence of polarization forces for 3-AMOEBA by embedding the polarization response of dimers and trimers within a complete representation of the fixed electrostatics of the entire system. We show that the electrostatic embedding formalism helps identify the specific configurations involving linear hydrogen-bonding arrangements that are poorly convergent at the 3-body level. By extending the definition of a body to be a large water cluster, we can reduce errors in forces to yield an approximate polarization model that is up to 10 times faster than the parent potential. The 3-AMOEBA model offers new ways to investigate how the properties of bulk water depend on the degree of connectivity in the liquid.

***Corresponding author**
thg@berkeley.edu

1. INTRODUCTION

The many-body expansion (MBE) of an N-body potential[1-6]

$$U_N \sim U_1 + U_2 + U_3 + ... \tag{1}$$

provides a useful organizing framework for understanding many-body interactions ranging from classical polarization to quantum mechanical energies. Eq. (1) states that the potential energy can be evaluated as the sum of the energies for monomers, dimers, trimers, and so on.

$$U_1 = \sum_{i=1}^{N} U(i) \qquad U_2 = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} U(i,j) - U(i) - U(j)$$

$$U_3 = \sum_{i=1}^{N-2} \sum_{j=i+1}^{N-1} \sum_{k=j+1}^{N} U(i,j,k) - U(i,j) - U(i,k) - U(j,k) + U(i) + U(j) + U(k) \tag{2}$$

where N is the size of the complete system, n is the order of truncation, with the first $n=3$ terms defined in Eq. (2). By differentiating these individual MBE terms with respect to atomic coordinates, it provides a corresponding n-body approximation to the N-body gradients. One of the key advantages afforded by the MBE is that the evaluation of the computationally costly N-body interaction energy is avoided by the calculation of more tractable, as well as independent, subsystems, provided that the expansion can be truncated at low order to good approximation. For these reasons, the MBE has enjoyed a high degree of popularity in the quantum chemical community, where it provides the basis for many of the so-called fragment methods.[7,8,9,10,11,12,13,14,15,16,17,18]

In contrast to most MBE formulations based on quantum mechanics, which focus almost exclusively on convergence of energies for gas-phase clusters, in this work we analyze convergence of the MBE for energies and forces for the AMOEBA classical polarizable model in the condensed phase. We focus on the specific case of liquid water under periodic boundary conditions and using Particle Mesh Ewald (PME) for long-range electrostatics. When a body is one water molecule, and the MBE is truncated at the level of trimers, this formally captures all direct polarization, i.e. where the induced dipoles respond only to the permanent electrostatic field and not the field generated by the other induced dipoles[19-21]. The computational advantages of the direct polarization model are significant since no mutual induction calculations are required, but it also is significantly diminished in accuracy since it fails to account for ~20% of the total polarization response. However, we have shown that the inaccuracies of the direct polarization model can be mostly eliminated by reformulating its parameters using an advanced automated parameterization method called ForceBalance[22], to yield the so-called inexpensive model, iAMOEBA[20].

Truncation of the MBE at the minimum level of n=3 defines the necessary starting point for the many-body *approximation* to mutual polarization itself.[21] Here we extend the MBE to account for the missing mutual polarization not captured by iAMOEBA to define well-defined approximations to the many-body polarization reference potential AMOEBA. We thus consider the approximation to mutual polarization under the 3-body truncation of Eq. (1) to define the 3-

AMOEBA model. For the case when a body is an individual water molecule, we have determined that errors in the polarization energy with respect to the parent potential are small, indicating acceptable convergence of the MBE. However, these small polarization energy errors do not correspond to commensurately small errors in the forces, such that the 3-AMOEBA model exhibits gradient errors approaching ~70-80% with respect to the AMOEBA reference. We therefore introduce an "electrostatic embedding" framework, in which the induced dipole sites of the subsystem are embedded within a larger permanent electrostatic environment, and which leads to insights into how the slowly convergent gradients for the 3-AMOEBA model can be corrected. This analysis has led us to enlarging the fragment size, i.e. by defining a "body" as a much larger cluster of water molecules, thereby yielding a 3-AMOEBA model with greatly reduced errors in forces.

The paper is outlined as follows. In Section 2 we introduce the theoretical details of the AMOEBA polarization model and the systematic hierarchy of electrostatic embedding schemes. In Section 3 we provide results on the convergence of energies and forces of the MBE based on small water cluster fragments under the different embedding schemes, which in turn shows how to improve convergence of the forces by fragmenting the liquid into larger clusters. In Section 4 we briefly summarize the OpenMP/MPI results for 3-AMOEBA in which computational speedups of 3X-10X are realized with respect to the parent AMOEBA potential. Section 5 comprises a summary and conclusion.

2. THEORY

Since AMOEBA's only many-body interaction involves the inducible dipole, we can express the total many-body energy as the polarization energy, U_{pol}, given by the product of a vector of induced-dipole atomic sites, $\underline{\mu}_{ind}$, and permanent electrostatic field elements, $\underline{E}$, corresponding to all atomic sites:

$$U_{pol} = -\frac{1}{2}\left(\underline{\mu}_{ind}^{(d)}\right)^T \underline{E}^{(p)} \tag{3a}$$

where

$$E_{i,\gamma}^{(p)} = \sum_{j}^{13N} T_{ij,\gamma} M_j^{(p)} \qquad \gamma=x,y,z \qquad i=1,\ldots,N \tag{3b}$$

and the elements of μ_{ind} are defined as

$$\mu_{i,\gamma}^{(d)} = \alpha_{i,\gamma\delta}\left(\sum_{j}^{13N} T_{ij,\gamma} M_j^{(d)} + \sum_{j'}^{N} T'_{ij,\gamma}\mu_{j'\delta}\right) \qquad \gamma,\delta = x,y,z \qquad i=1,\ldots,N \tag{3c}$$

where $\alpha_{i,\gamma\delta}$ is the polarizability of atom i, T_{ij} is the rank-two interaction tensor between atoms i and j containing derivatives of $1/r_{ij}$ according to the permanent multipole expansion, and $M_j^{(d)}$ are the permanent multipole moments; the T and (rank-one) M tensors encompass the 13 permanent multipole moments for the AMOEBA potential (q, μ_x, μ_y, μ_z, Q_{xx}, Q_{xy}, Q_{xz}, Q_{yx}, Q_{yy}, Q_{yz}, Q_{zx}, Q_{zy}, Q_{zz}). T' contains just derivative terms of $1/r_{ij}$ corresponding just to the dipole term in the

multipole expansion. The superscript (*p*) or (*d*) refers to special scaling factors used for electrostatic interactions in AMOEBA[23]. The nature of *p* scaling is to scale all relevant polarization interactions to zero for all atom pairs separated by one or two covalent bonds, as is done in many standard molecular mechanics force fields. By contrast *d* scaling refers to polarization interactions that are set to zero within a polarization group, which is specific to the AMOEBA force field [24,25,23]. This will become relevant to the different embedding schemes developed below. In the most general case, the polarizability is an anisotropic rank-two tensor; in the AMOEBA force field, however, it is approximated as an isotropic quantity where $\alpha_{xx}=\alpha_{yy}=\alpha_{zz}$, and off-diagonal elements are zero.

The first term in Eq. (3c) corresponds to the direct polarization response, in which a polarizable site couples with the permanent moments of two other sites

$$\mu_{1,\gamma}^{(d)} = \alpha_{1,\gamma}\left(T_{12,\gamma}M_2^{(s)} + T_{13,\gamma}M_3^{(s)}\right) \qquad \gamma = x,y,z \qquad s=d \text{ or } p \tag{4a}$$

$$U_{direct} = -\frac{1}{2}\sum_{\beta=x,y,z}\sum_{\gamma=x,y,z}\begin{pmatrix}\alpha_{1,\gamma}T_{12,\beta}M_2^{(s)}T_{12,\gamma}M_2^{(s)} + \\ \alpha_{1,\gamma}T_{13,\beta}M_3^{(s)}T_{13,\gamma}M_3^{(s)} + \\ \alpha_{1,\gamma}T_{23,\beta}M_3^{(s)}T_{23,\gamma}M_3^{(s)}\end{pmatrix} \qquad \gamma = x,y,z \qquad s=d \text{ or } p \tag{4b}$$

Truncation at the level of trimers in Eq. (1) is thus formally exact to the *N*-molecule formulation of the energy for a system in which inducible dipoles only interact with the permanent field, as already stated. The last term in Eq. (3c) represents the electric field at atom *i* due to the induced dipoles at all other atomic sites, μ_j, where T_{ij}' is the interaction tensor between atoms *i* and *j* containing derivatives of $1/r_{ij}$ according to the inducible dipole-dipole interactions. The mutual polarization is usually solved numerically via self-consistent field (SCF) calculations[26,27,28] methods, although we have recently halved the number of SCF cycles through a new extended Lagrangian (EL) formation known as iEL/SCF.[29]

Under the 3-body approximation to the MBE, Eqs. (3) and (4) still apply. But now what differs is that only a subset of sites, *n*, corresponding just to the subsystem sites undergo mutual induction, in contrast to the full system of *N* sites. Therefore, the summation in the second term in Eq. (3c) would run over *n*, instead of *N*. However, the MBE is not uniquely defined until we specify the environmental conditions under which we are evaluating the individual fragments, i.e. the electric field via Eq. (3b). This requires a formulation of an electrostatic embedding scheme that, as we will show below, strongly affects the convergence profile of energy *and* forces, which is quantified below.

In order to easily express gradients under the different embedding schemes, Eqs. (3) are better formulated as

$$U_{pol} = -\frac{1}{2}\sum_{i}^{3N}\sum_{j}^{3N}E_i^{(d)}C_{ij}^{-1}E_j^{(p)} \tag{5}$$

in which we define a new matrix C

$$C_{ij} = \alpha_{ij}^{-1} - T'_{ij} \qquad \text{for} \qquad i,j=1,\ldots 3N \tag{6}$$

We now introduce notation that uses subscripts to indicate vector and tensor dimensions reflecting the degrees of freedom of the whole system, *N,* or the monomer, dimer, or trimer subsystem, *n*, and superscripts that describe the size of the permanent electrostatic embedding environment, essentially, the number of permanent multipole sites giving rise to the permanent electrostatic field. For the AMOEBA potential we would thus express Eq. (5) as

$$U_{pol,N}^{N,N} = -\frac{1}{2} \underline{E}_{3N}^{T(d,N)} \underline{\underline{C}}_{3Nx3N}^{-1} \underline{E}_{3N}^{(p,N)} \tag{7a}$$

where

$$E_i^{(s,N)} = \sum_j^{13N} T_{ij} M_j^{(s)} \qquad \text{for } i=1,\ldots 3N \quad \text{and} \quad s=d \text{ or } p \tag{7b}$$

Eqs. (7) describe the evaluation of the full AMOEBA potential in which all interactions are accounted for in the full *N*-dimension system. Note that we include the scaling factors explicitly, since the embedding schemes are based on the size of the permanent electrostatic environment used in either the d-scaled or p-scaled field. Using this notation, we can now consider the interplay of the truncation of the many-body expansion and the definition of the electrostatic embedding environment simultaneously.

Next, we consider three types of embedding schemes for the evaluation of fragments. The simplest electrostatic embedding scheme for the 3-AMOEBA model is one wherein the polarization energy of a given monomer, dimer, or trimer is evaluated in vacuum; that is, the induced dipoles of the subsystem are generated by the permanent electrostatic field of only the subsystem fixed multipoles, isolated from the electrostatic environment of the other permanent sites in the full *N* system. The polarization energy for a given monomer, dimer, or trimer is then

$$U_n^{n,n} = -\frac{1}{2} \underline{E}_{3n}^{T(d,n)} \underline{\underline{C}}_{3nx3n}^{-1} \underline{E}_{3n}^{(p,n)} \tag{8a}$$

where

$$E_i^{(s,n)} = \sum_j^{13n} T_{ij} M_j^{(s)} \qquad \text{for } i=1,\ldots 3n \quad \text{and} \quad s=d \text{ or } p \tag{8b}$$

At the level of the 3-body truncation of the MBE, Eqs. (8) by definition capture the direct polarization contribution exactly (Eqs. 4), and approximates the mutual polarization.

The corresponding gradient of the *kth* atom with respect to Cartesian position can be expressed as follows (further details on the gradient derivation are given in [23], although we have introduced important corrections here):

$$\frac{\partial U_n^{n,n}}{\partial \gamma_k} = -\frac{1}{2}\begin{pmatrix} \frac{\partial}{\partial \gamma_k}\left(\underline{\underline{T}}_{3nx13n}\underline{M}_{13n}^{(d,n)}\right)^T \underline{\underline{C}}_{3nx3n}^{-1}\underline{\underline{T}}_{3nx13n}\underline{M}_{13n}^{(p,n)} + \\ \left(\underline{\underline{T}}_{3nx13n}\underline{M}_{13n}^{(d,n)}\right)^T \frac{\partial}{\partial \gamma_k}\left(\underline{\underline{C}}_{3nx3n}^{-1}\right)\underline{\underline{T}}_{3nx13n}\underline{M}_{13n}^{(p,n)} + \\ \left(\underline{\underline{T}}_{3nx13n}\underline{M}_{13n}^{(d,n)}\right)^T \underline{\underline{C}}_{3nx3n}^{-1}\frac{\partial}{\partial \gamma_k}\left(\underline{\underline{T}}_{3nx13n}\underline{M}_{13n}^{(p,n)}\right) \end{pmatrix} \qquad \gamma=x,y,z \text{ and } k=1,..n. \tag{9a}$$

or equivalently

$$\frac{\partial U_n^{vacuum}}{\partial \gamma_k} = -\frac{1}{2}\begin{pmatrix} \frac{\partial}{\partial \gamma_k}\left(\underline{\underline{T}}_{3nx13n}\underline{M}_{13n}^{(d,n)}\right)^T \cdot \underline{\mu}_{3n}^{(p,n)} + \\ \underline{\mu}_{3n}^{T(d,n)} \cdot \frac{\partial}{\partial \gamma_k}\left(\underline{\underline{T}}_{3nx13n}\underline{M}_{13n}^{(p,n)}\right) + \\ \underline{\mu}_{3n}^{T(d,n)} \cdot \frac{\partial}{\partial \gamma_k}\left(\underline{\underline{T}}'_{3nx3n}\right) \cdot \underline{\mu}_{3n}^{(p,n)} \end{pmatrix} \qquad \gamma=x,y,z \text{ and } k=1,..n. \tag{9b}$$

When considering the special case of a water molecule as the definition of a "body", the monomer energy under vacuum embedding is zero, since $E_1^{(p,n)} = E_1^{(d,n)} = 0$, so that only dimers and trimers contribute to the 3-body approximation to polarization in the condensed phase. In fact, monomers would contribute in the general case of a larger fragment with at least 3 consecutive covalent bonds $E_1^{(p,n)} \neq 0$, and/or with multiple polarization groups within a fragment $E_1^{(d,n)} \neq 0$. We return to this point below.

In the next embedding scheme we describe the total polarization energy as

$$U_{pol,n}^{N,n} = -\frac{1}{2}\underline{E}_{3n\subset 3N}^{T(d,N)}\underline{\underline{C}}_{3nx3n}^{-1}\underline{E}_{3n}^{(p,n)} \tag{10a}$$

where

$$E_i^{(d,N)} = \sum_j^{13N} T_{ij}M_j^{(d)} \qquad \text{for } i=1,\ldots 3n\subset 3N \tag{10b}$$

such that the partial derivative can be expressed as:

$$\frac{\partial U_n^{N,n}}{\partial \gamma_k} = -\frac{1}{2}\begin{pmatrix} \frac{\partial}{\partial \gamma_k}\left(\underline{\underline{T}}_{3nx13N}\underline{M}_{13N}^{(d,N)}\right)^T \cdot \underline{\mu}_{3n}^{(p,n)} + \\ \underline{\mu}_{3n}^{T(d,N)} \cdot \frac{\partial}{\partial \gamma_k}\left(\underline{\underline{T}}_{3nx13n}\underline{M}_{13n}^{(p,n)}\right) + \\ \underline{\mu}_{3n}^{T(d,N)} \cdot \frac{\partial}{\partial \gamma_k}\left(\underline{\underline{T}}'_{3nx3n}\right) \cdot \underline{\mu}_{3n}^{(p,n)} \end{pmatrix} \qquad \gamma=x,y,z \text{ and } k=1,..N. \tag{10c}$$

The superscript used for the energy U is designed to describe the induction of an inducible dipole of an *n*-size fragment due to all *N* permanent sites dotted into the local electric field of the *n*-size fragment. An important point is that although Eq. (10c) formally applies to all sites *N*, the last 2 terms are nonzero only if *k* is one of the *n* sites of the given dimer or trimer. Again, for water the

monomer energy is zero since $E_1^{(p,n)} = 0$, so that only dimers and trimers contribute to the 3-body approximation to polarization in the condensed phase.

In a related embedding scheme to Eq. (10), we can also describe the total polarization energy as

$$U_{pol,n}^{n,N} = -\frac{1}{2}\underline{E}_{3n}^{T(d,n)}\underline{\underline{C}}_{3nx3n}^{-1}\underline{E}_{3n\subset 3N}^{T(p,N)} \tag{11a}$$

where

$$E_i^{(p,N)} = \sum_j^{13N} T_{ij} M_j^{(p)} \qquad \text{for } i=1,\ldots 3n\subset 3N \tag{11b}$$

and the partial derivative can be expressed as:

$$\frac{\partial U_n^{N,n}}{\partial \gamma_k} = -\frac{1}{2}\begin{pmatrix} \frac{\partial}{\partial \gamma_k}\left(\underline{\underline{T}}_{3nx13n}\underline{M}_{13n}^{(d,n)}\right)^T \cdot \underline{\mu}_{3n}^{(p,N)} + \\ \underline{\mu}_{3n}^{T(d,n)} \cdot \frac{\partial}{\partial \gamma_k}\left(\underline{\underline{T}}_{3nx13N}\underline{M}_{13N}^{(p,N)}\right) + \\ \underline{\mu}_{3n}^{T(d,n)} \cdot \frac{\partial}{\partial \gamma_k}\left(\underline{\underline{T}}'_{3nx3n}\right) \cdot \underline{\mu}_{3n}^{(p,N)} \end{pmatrix} \qquad \gamma=x,y,z \text{ and } k=1,..\text{N}. \tag{11c}$$

This describes the induction of an inducible dipole due to the *n*-size fragment's local permanent sites dotted into the electric field generated by the complete *N* permanent sites. In this case, in analogy to the gradient in the embedding scheme of Eq. (10), although $k=1,..N$, only the second term is nonzero for all *k*, as this derivative term has dependency on all *N* permanent electrostatic sites. Again, the monomer energy is zero for water since $E_1^{(d,n)} = 0$, so that only dimers and trimers contribute to the 3-body approximation to polarization in the condensed phase. Although the two schemes that are defined in Eqs. (10) and Eqs. (11) are numerically equivalent for the special case of water and sulfate under the AMOEBA model, in fact they are two distinct embedding schemes for the general AMOEBA case of fragments that have more than one polarization group or groups that are large enough to contain sites separated by 3 covalent bonds.

Finally, we consider complete N-dimensional embedding

$$U_{pol,n}^{N,N} = -\frac{1}{2}\underline{E}_{3n\subset 3N}^{T(d,N)}\underline{\underline{C}}_{3nx3n}^{-1}\underline{E}_{3n\subset 3N}^{T(p,N)} \tag{12a}$$

where

$$E_i^{(s,N)} = \sum_j^{13N} T_{ij} M_j^{(s)} \qquad \text{for } i=1,\ldots 3n\subset 3N \quad \text{and} \quad s=p \text{ or } d. \tag{12b}$$

and the partial derivative can be expressed as:

$$\frac{\partial U_n^{N,N}}{\partial \gamma_k} = -\frac{1}{2}\begin{pmatrix} \frac{\partial}{\partial \gamma_k}\left(\underline{\underline{T}}_{3nx13N}\underline{M}_{13N}^{(d,N)}\right)^T \cdot \underline{\mu}_{3n}^{(p,N)} + \\ \underline{\mu}_{3n}^{T(d,N)} \cdot \frac{\partial}{\partial \gamma_k}\left(\underline{\underline{T}}_{3nx13N}\underline{M}_{13N}^{(p,N)}\right) + \\ \underline{\mu}_{3n}^{T(d,N)} \cdot \frac{\partial}{\partial \gamma_k}\left(\underline{\underline{T}}'_{3nx3n}\right) \cdot \underline{\mu}_{3n}^{(p,N)} \end{pmatrix} \qquad \gamma=x,y,z \text{ and } k=1,..N. \qquad (12c)$$

This final embedding approach describes the induction of an inducible dipole of an *n*-size fragment due to all *N* permanent sites dotted into the total electric field due to all *N* permanent sites.

All calculations were performed with a modified version of the TINKER modeling software package[30] in which we have implemented our 3-AMOEBA polarization model based on truncation of the many-body expansion at trimers, but under different electrostatic embedding schemes formally outlined above and with different sizes of fragments to describe a body.

3. RESULTS

Figures 1a and 1b show the convergence profile of 2-body and 3-body polarization energies with respect to distance for all dimers and trimers, respectively, comprising a 216 water molecule MD snapshot under the vacuum embedding scheme, in which a body is defined as an individual water molecule. We use a 3-body distance metric that sums the shortest 2 oxygen-oxygen distances among the 3 water molecules, while the 2-body distance metric is simply the oxygen-oxygen distance. Under vacuum embedding, the pair polarization energy is fully attractive at all distances and converges quickly with distance, whereas the trimer energy shows cooperative and anti-cooperative interactions, but it too decays reasonably rapidly with the distance metric. Table 1 also provides the errors in evaluation of the energies using 2-AMOEBA and 3-AMOEBA under vacuum embedding with respect to the AMOEBA full *N*-body mutually polarized energy. As already discussed, direct polarization is fully captured up through 3-body terms (Table S1), and it is evident that the truncation of the MBE at the level of trimers for mutual polarization energies appear acceptable under vacuum embedding, yielding relatively small errors of ~1-3%. It is by this energy metric that most MBE applications are judged for their convergence under vacuum embedding.

Table 1 also provides errors in the forces, both in terms of magnitude and direction (defined as an inner product) for 3-AMOEBA compared to the AMOEBA model using vacuum embedding. While the error in energies for bulk water is seen to be relatively small, and force direction is captured almost perfectly, the vacuum embedding result for force magnitudes gives large errors of ~30-60% for the small neutral water systems in Table 1 with respect to the AMOEBA polarization forces. Figure 1c plots all differences in the norm of the force between 3-AMOEBA and the parent AMOEBA model for all pairs of atoms in the 216 water box; errors are seen to be largest at short interatomic distances.

Furthermore, the gradient errors under vacuum embedding give rise to significant inaccuracies in the related internal virial and single-time step velocities, since both depend on

forces, and such errors would significantly degrade energy minimization schemes or ensemble properties calculated in a molecular dynamics simulation. This is especially a problem in the NPT ensemble due to the fact that pressure has both an internal virial as well as an ideal gas contribution directly related to the square of the velocities by the equipartition theorem in the standard kinetic theory of gases:

$$P = \frac{1}{3V}\left[\sum_{i=1}^{N} m\underline{v}_i \cdot \underline{v}_i + \langle W \rangle\right] \qquad \text{where} \qquad W = \sum_{i=1}^{N} \underline{r}_i \cdot \underline{F}_i \tag{13}$$

in which the velocities are propagated by the forces. The situation for vacuum embedding worsens dramatically for the divalent anion-water system in which there is severe degradation in everything: energies, force magnitude and force direction, and internal virial, with respect to the parent full polarization model (Table 1 and Table S1).

The question is then whether a more sophisticated electrostatic embedding scheme would improve convergence of the MBE for force magnitudes to 3-body terms or even fewer. We first consider the evaluation of the mutual polarization energies under the asymmetric *N,n* embedding scheme, which describes the induction of an inducible dipole of an *n*-size fragment due to all *N* permanent sites dotted into the local electric field of the *n*-size fragment. Figure 2a and 2b shows that both cooperative and anti-cooperative interactions are now evident at 2-body polarization, whereas before with simple *n,n*, embedding the 2-body interaction is purely cooperative. A nice result emerges that the direct polarization is exact at the 2-body level, such that a direct polarization model can be refactored into sums over independent pairs (Table S1). Table 1 demonstrates that the energy errors under the asymmetric *N,n* embedding for the water systems have been reduced to ~0.5-1.5% while the RMS error in the gradient magnitudes are reduced to 9-24% with respect to the AMOEBA polarization forces, a substantial improvement over the vacuum embedding scheme. Although the dianion water cluster has improved substantially as well, especially with respect to force direction, the errors in energies and force magnitudes are more slowly convergent compared to the neutral water systems. Table S2 shows that errors in velocities and internal virial have also been reduced.

The full *N,N* embedding approach describes the induction of an inducible dipole of an *n*-size fragment due to all *N* permanent sites dotted into the total electric field due to all *N* permanent sites. Figures 3a and 3b show that the 2- and 3-body energy contributions are smaller, and Table 1 demonstrates that the energy errors have been reduced to < 0.5% for all water systems and less than < 0.9% for the sulfate-water system, while the RMS error in the gradient magnitudes are reduced to 3-10% with respect to the AMOEBA polarization forces. Here the errors in the virial and the velocities are also at their lowest compared with the other electrostatic embedding schemes, as anticipated from the corresponding improvement in the gradient (Table S2). Furthermore the extreme asymmetric polarization of the sulfate-water cluster gives comparable error to the homogenous bulk result, which is encouraging for complex solutions or ions at the air-water interface for example.

The *N,n* or *N,N* embedding schemes by their very nature, wherein the polarization in the 1-, 2-, or 3-body molecular subsystems is a function of the electrostatic field generated by sites

external to the subsystem, imply that higher-order terms in the MBE may be accounted for implicitly. Indeed, this is suggested by the fact that the 2-body polarization energy profile as a function of intermolecular distance displays both cooperative and anti-cooperative effects under *N,n* or *N,N* embedding, while 2-body vacuum embedding shows only cooperative interactions and is purely attractive. The importance of these implicit higher-order effects may be most clearly manifested in the force magnitude, which shows better convergence to the standard AMOEBA model under the *N,n* scheme and even better under *N,N* embedding. It clearly indicates that while 3-AMOEBA energies appear convergent in vacuum, force magnitudes require additional terms in the MBE.

To confirm that the improvement in the forces under embedding is due to the ability to capture higher-order effects at lower order in the expansion, we carried out the MBE to 4th order for vacuum embedding for the tractable 17-water and 216-water molecule cases, and indeed force magnitude errors are comparable to the complete *N,N* embedding (Table 1). Interestingly, the 216-water molecule case shows larger errors in vacuum embedding calculated to 4th order in the MBE than in *N,N* embedding. This suggests that *N,N* embedding is accounting for terms even higher than 4th order implicitly, even when the polarization is evaluated at a lower order in the MBE.

Even so, while the *N,n*, *n,N*, and *N,N* embedding schemes afford much greater accuracy in the calculation of the polarization force magnitude in particular, they come at an intractable computational cost due to the need to evaluate derivative terms with respect to all *N* degrees of freedom that generated the permanent electrostatic field for each subsystem configuration. However, what the embedding schemes lead to is a better understanding of what is the underlying error for the mutual polarization response in vacuum – namely the neglect of the permanent electrostatic environment and how this impinges most dramatically on the quality of the forces. In light of the fact that vacuum embedding is desirable from the standpoint of being computationally tractable, we were motivated to compare its associated forces more closely with those yielded by the higher quality, but computationally more cumbersome, *N,N* embedding model.

Given that one of the defining unique features of water is its hydrogen bonding network, and the recognition that hydrogen bonding arises from many-body effects such as polarization, we sought to determine whether the largest errors in energy and forces could be classified based on geometric descriptors related to hydrogen bonding (Figure 4). First we sought to evaluate the discrepancies in the forces between atoms *i* and *j,* generated by vacuum embedding as function of intermolecular distance, r_{ij}, using the *N,N* embedding scheme as the reference for both 2-body and 3-body polarization. It is evident from Figure 5a that the largest difference in gradients between the more accurate N,N embedding and less accurate vacuum embedding, $\vec{f}_{ij}^{N,N} - \vec{f}_{ij}^{n,n}$, involve the oxygen–hydrogen bonded pair of the donor water (r_{ij}=1.0 Å) and the pair comprising the donor hydrogen and oxygen of the acceptor molecule (r_{ij}~1.8-2.0 Å), i.e. a hydrogen bond. However, at short range there also exist many points corresponding to interatomic forces that are in very good agreement, suggesting that there must be metrics other than interatomic distance

that can better distinguish atomic pairs with large force differences under the two embedding schemes. To further classify these configurations, we next evaluated both intermolecular distance as well as the donor hydrogen-oxygen acceptor angle and examined how the difference in forces between pairs of atoms in *N,N* vs. vacuum embedding varies with these two hydrogen bonding metrics. In Figure 5b, we see that force errors are largest in magnitude for pairs of atoms that are associated with a dimer whose hydrogen bonding angle and distance parameters are near the ideal range, that is, a hydrogen-oxygen distance just under 2.0 Å and a hydrogen bonding angle close to linear.

Since the classification of $\vec{f}_{ij}^{N,N} - \vec{f}_{ij}^{n,n}$ by geometric parameters related to hydrogen bonding was informative in the 2-body case, we then extended this to force differences in the 3-body calculation. Here, the situation is more complicated owing to the greater number of possible hydrogen bonding angles and distances. For each trimer, we determined the 3-possible hydrogen bond distance/angle pairs, and examined how the force error varies as a function of each of these pairs. Indeed, when we plot the 3-body force errors as a function of any given pair of hydrogen bond distances/angles, we observe 2 distinct geometric signatures. The first corresponds to the feature at the ideal hydrogen bond distance that is equivalent to what was observed when evaluating the 2-body case (Figure 6a). We can also see a second peak at an angle of about 1.6-2.0 radians (~90-115°) and a distance of about 4 Å, which is close to the expected distance/angle pair in the trimer that extends the hydrogen-bonded network to favorable linear hydrogen bond configurations into the second shell of neighbors (Figure 6b). It is worth mentioning that 85% and 98% of these identified configurations in Figures 5 and 6 are directly responsible for the largest differences in the 2-body and 3-body energies, respectively, i.e. evaluated as intermolecular energies between atom *i* and *j* under N,N and vacuum embedding ($U_{ij}^{N,N} - U_{ij}^{n,n}$).

Hence while the total polarization energy errors appear to be small at the level of 3-body truncation, the scalar energy is almost worthless as a convergence criterion, whereas the gradient is a better indicator that in fact the MBE is far from convergence to be an adequate approximation to the N-body potential. What is interesting about liquid water is that the lack of MBE convergence is defined by a highly unique set of geometric configurations that are very few in number, a number that is dwarfed by the sheer factorial count of the pairs and/or triplets configurations that need to be evaluated. Taken together, the differences in polarization energies and forces calculated under vacuum embedding vs. complete embedding for liquid water show that the metric for MBE convergence is a collective one, i.e. linear hydrogen bonds are supported by an expanded hydrogen bonded network. In network theory terms, an associated liquid like water is above a percolation threshold since each water molecule makes more than the critical minimum, n_{HB}* ~1.5 hydrogen bonds.[31,32] The picture that emerges from the embedding analysis is that for a truncated MBE based on single molecule of water as a body, the resulting model fluid is below a necessary percolation threshold, and that higher order terms in the MBE are necessary to describe the condensed phase.

An alternative way to realize something akin to the N,N embedding result is then to require that the body must in and of itself approach a percolation threshold. With this realization, it becomes immediately evident that we can investigate convergence of energy and gradients under vacuum embedding by fragmenting the liquid into large clusters. The question of how large these clusters need to be is reported in Table 2, and again emphasizes how slowly the gradient converges as cluster body size enlarges to hundreds to thousands to tens of thousands of water molecules before gradient error subsides to that close to complete fixed electrostatic embedding. In future work we will use this 3-AMOEBA cluster body model to investigate how sensitive are structural, thermodynamic and dynamic properties to this new collective measure of MBE convergence.

4. HYBRID MPI/OPENMP PARALLELIZATION OF 3-AMOEBA

The MBE approximation yields computational savings since the polarization of each individual fragment is calculated independently to allow for trivial parallelization with minimal communication overhead. Moreover the rapid decay of the polarization energy with fragment separation distance permits the use of distance cutoffs, ultimately reducing the cost of 3-AMOEBA mutual polarization to be calculated as an *O(N)* scaling calculation.

We use a hybrid MPI/OpenMP replicated data strategy implemented on a modern CPU hardware architecture, where a number of CPU cores share memory on a single node, allowing for OpenMP parallelism within the node; there is a further level of parallelism that takes place among nodes that do not share memory, wherein the MPI tasks are distributed among nodes according to an optimized load-balancing scheme. We have reported on our MPI/OpenMP implementation of iAMOEBA in other work[33], which yields speedups of 4-8 compared to the optimized OpenMP implementation in the Tinker7 software package. Correspondingly, our hybrid OpenMP/MPI parallel implementation of 3-AMOEBA, whose timings and speedups are reported in Table 3 for the large cluster version in this paper, demonstrate that it is up to an order of magnitude faster than the OpenMP optimized AMOEBA model. Details of the parallel implementation will be reported elsewhere.

5. DISCUSSION AND CONCLUSION

The MBE approach using a single water molecule as a body in the context of a condensed phase classical polarization calculation starkly reveals that while energy appears reasonably convergent when truncated at the level of 3-body interactions in vacuum, the corresponding magnitude of the gradients show a much slower convergence profile, requiring additional terms in the MBE of at least 4^{th} order. As an alternative to including higher order MBE terms, we explored a hierarchical progression of permanent electrostatic embedding schemes that effectively recapture more of the complete polarization response. The improvement in the successive embedding schemes are especially dramatic for the asymmetric environment of a highly charged sulfate divalent anion in a small water cluster embedded in the gas phase- a very challenging test of the polarization approximations. Although the numerical results are specific to the AMOEBA force

field, the model is general enough to show how the total polarization converges for energies and forces for any polarizable force field.

We find that even though 3-AMOEBA energy errors appear small with respect to the parent AMOEBA potential, they are directly tied to *critical* configurations of waters involved in linear hydrogen bonds in which the derivatives with respect to their atomic centers are largest in error. This in turn leads to errors in the computed virial that results in a simulated density in the NPT ensemble that yields a value of 1.08 g/cc. This can be understood by the fact that good (i.e. linear) hydrogen-bonds are required to create an expanded network that recovers a lower density of the water liquid, a condensed phase feature not well supported under the vacuum embedding model. Nonetheless, it appears that we can classify these configurations, and develop criteria for when they are under-polarized or over-polarized, to provide corrections at the 3-body level under vacuum embedding which may yield a good approximation to the parent N-body potential. It would be desirable to correct the 3-body terms for the reason that the combinatorial explosion of including 4-body terms in the MBE is not only costly, but can degrade the numerical solutions due to loss-of-precision problems arising from the dramatically larger number of terms in the expansion.[34] We can avoid both of these problems through a completely different solution by replacing a single water molecule as the defining body in the MBE by a large water cluster to approximate the complete embedding result. Even so, while energies can be converged to ~1/100th of a percent, gradients can still be in error by 15-20%. As is true for any of the fractionation definitions, we can obtain substantial speedups for the MBE relative to the complete potential, by as much as a factor of 10X.

It is also useful to consider what we have leaned about the MBE approach to classical polarization relative to quantum mechanical treatments[35,36,37,38,39,40,41,42,43,17,] including those that use some type of embedding scheme[14,10,13]. While the QM-MBE combined with an advanced embedding scheme has been used to predict crystal energetics[15,44], water clusters and ice structures[45,46], and bulk salt and surface adsorption[38,37], the focus has been almost exclusively on convergence of the energies and not the corresponding (nuclear) gradients. While there are substantial differences between quantum and classical potentials, the result pertaining to small energy errors and slow convergence in the accuracy of gradients should be relatively independent of the level of theory employed. Furthermore, it is naïve to only consider a single scalar property for convergence when the nature of the system may require a completely different convergence metric, such as a collective variable as we have suggested for liquid water. Therefore, the convergence of nuclear gradients would seem to be a critically important analysis for understanding the quality of QM-MBE approaches, and we hope that the corrective measures we have introduced here, i.e. approximating the permanent electrostatic fields influence on the many-body interaction, might find analogies in the quantum based approaches to the MBE. This message may be resonating already in the QM community where recent work has shown that for water clusters that 4-body terms cannot be neglected, and that various embedding schemes in the literature actually afford no systematic improvement when larger water clusters are considered.[47]

ACKNOWLEDGEMENTS. We thank the National Science Foundation grant CHE-1363320 for support of this work.

REFERENCES

(1) Hankins, D.; Moskowitz, J. *J. Chem. Phys.* **1970**, *53*, 4544.
(2) Stoll, H.; Preuss, H. *Theoretica Chimica Acta* **1977**, *46*, 11.
(3) Leverentz, H.; Maerzke, K.; Keasler, S.; Siepmann, J.; Truhlar, D. *Phys Chem Chem Phys* **2012**, *14*, 7669.
(4) Richard, R.; Herbert, J. *J. Chem. Phys.* **2012**, *137*.
(5) Gordon, M.; Fedorov, D.; Pruitt, S.; Slipchenko, L. *Chem Rev* **2011**, *112*.
(6) Fedorov, D. G.; Kitaura, K. In *The Fragment Molecular Orbital Method: Practical Applications to Large Molecular Systems*; Fedorov, D. G., Kitaura, K., Eds.; CRC: Boca Raton, FL, 2009, p 5.
(7) Day, P. N.; Jensen, J. H.; Gordon, M. S.; Webb, S. P.; Stevens, W. J.; Krauss, M.; Garmer, D.; Basch, H.; Cohen, D. *J Chem Phys* **1996**, *105*, 1968.
(8) Gao, J. L. *J Phys Chem B* **1997**, *101*, 657.
(9) Gao, J. L. *J Chem Phys* **1998**, *109*, 2346.
(10) Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. *Chem Phys Lett* **1999**, *313*, 701.
(11) Gordon, M. S.; Freitag, M. A.; Bandyopadhyay, P.; Jensen, J. H.; Kairys, V.; Stevens, W. J. *J Phys Chem A* **2001**, *105*, 293.
(12) Zhang, D. W.; Xiang, Y.; Zhang, J. Z. H. *J Phys Chem B* **2003**, *107*, 12039.
(13) Dahlke, E. E.; Truhlar, D. G. *J Chem Theory Comput* **2007**, *3*, 46.
(14) Fedorov, D. G.; Kitaura, K. *J Phys Chem A* **2007**, *111*, 6904.
(15) Wen, S. H.; Beran, G. J. O. *J Chem Theory Comput* **2011**, *7*, 3733.
(16) Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. V. *Chem Rev* **2012**, *112*, 632.
(17) Richard, R. M.; Lao, K. U.; Herbert, J. M. *Acc. Chem. Res.* **2014**, *47*, 2828.
(18) Medders, G. R.; Götz, A. W.; Morales, M. A.; Bajaj, P.; Paesani, F. *J. Chem. Phys.* **2015**, *143*, 104102
(19) Dykstra, C. E. *Chem. Rev.* **1993**, *93*, 2339.
(20) Wang, L. P.; Head-Gordon, T.; Ponder, J. W.; Ren, P.; Chodera, J. D.; Eastman, P. K.; Martinez, T. J.; Pande, V. S. *J. Phys. Chem. B* **2013**, *117*, 9956.
(21) Demerdash, O.; Yap, E. H.; Head-Gordon, T. *Ann. Rev. Phys. Chem.* **2014**, *65*, 149.
(22) Wang, L.-P. https://simtk.org/home/forcebalance/, Stanford University, March 2014, 2013.
(23) Ren, P. Y.; Wu, C. J.; Ponder, J. W. *J. Chem. Theory Comput.* **2011**, *7*, 3143.
(24) Ren, P. Y.; Ponder, J. W. *J. Comp. Chem.* **2002**, *23*, 1497.
(25) Ren, P. Y.; Ponder, J. W. *J. Phys. Chem. B* **2003**, *107*, 5933.
(26) Young, D. *Iterative Solutions of Large Linear Systems.*; Academic Press: New York, 1971.
(27) Wang, W.; Skeel, R. D. *J. Chem. Phys.* **2005**, *123*, 164107.
(28) Pulay, P. *Chem. Phys. Lett.* **1980**, *73*, 393.
(29) Albaugh, A.; Demerdash, O.; Head-Gordon, T. *J. Chem. Phys.* **2015**, *143*, 174104.
(30) Ponder, J. W.; TINKER7.0 ed. 2014.
(31) Geiger, A.; Stillinger, F. H.; Rahman, A. *J. Chem. Phys.* **1979**, *70*, 4185.
(32) Geiger, A.; Stanley, H. E. *Phys. Rev. Lett.* **1982**, *49*, 1749.

(33) Albaugh, A.; Boateng, H. A.; Bradshaw, R. T.; Demerdash, O.; Dziedzic, J.; Mao, Y.; Margul, D. T.; Swails, J.; Zeng, Q.; Case, D. A.; Eastman, P.; Essex, J. W.; Head-Gordon, M.; Pande, V. S.; Ponder, J. W.; Shao, Y.; Skylaris, C.-K.; Todorov, I. T.; Tuckerman, M. E.; Head-Gordon, T. *J Phys Chem B (Feature article)* **2016**, *submitted*.
(34) Richard, R. M.; Lao, K. U.; Herbert, J. M. *J. Chem. Phys.* **2014**, *141*, 014108.
(35) Huang, C.; Libisch, F.; Peng, Q.; Carter, E. A. *J. Chem. Phys.* **2014**, *140*.
(37) Huang, C.; Pavone, M.; Carter, E. A. *J. Chem. Phys.* **2011**, *134*.
(38) Huang, C.; Carter, E. A. *J. Chem. Phys.* **2011**, *135*.
(39) Huang, P.; Carter, E. A. *J. Chem. Phys.* **2006**, *125*.
(40) Govind, N.; Wang, Y. A.; Carter, E. A. *J. Chem. Phys.* **1999**, *110*, 7677.
(41) Goodpaster, J. D.; Ananth, N.; Manby, F. R.; Miller, T. F. *J. Chem. Phys.* **2010**, *133*.
(42) Goodpaster, J. D.; Barnes, T. A.; Manby, F. R.; Miller, T. F. *J. Chem. Phys.* **2012**, *137*.
(43) Manby, F. R.; Stella, M.; Goodpaster, J. D.; Miller, T. F. *J. Chem. Theory Comput.* **2012**, *8*, 2564.
(44) Beran, G.; Wen, S.; Nanda, K.; Huang, Y.; Heit, Y. In *Prediction and Calculation of Crystal Structures: Methods and Applications*; Aspuru-Guzik, A., Atahan-Evrenk, S., Eds.; Springer: 2014, p 59.
(45) Gillan, M. J.; Alfe, D.; Bygrave, P. J.; Taylor, C. R.; Manby, F. R. *J. Chem. Phys.* **2013**, *139*.
(46) Bygrave, P. J.; Allan, N. L.; Manby, F. R. *J. Chem. Phys.* **2012**, *137*.
(47) Lao, K. U.; Liu, K.-Y.; Richard, R. M.; Herbert, J. M. *J. Chem. Phys.* **2016**, *144*, 164105.
(48) Tuckerman, M.; Berne, B. J.; Martyna, G. J. *J. Chem. Phys.* **1992**, *97*, 1990.
(49) Felicíssimo, V. C.; Guimarães, F. F.; Gel'mukhanov, F.; Cesar, A.; Ågren, H. *J. Chem. Phys.* **2005**, *122*, 094319.

TABLES

Table 1. Errors in the many-body expansion under different electrostatic embedding schemes. Energy units are kcal/mole. Note that the 2-AMOEBA and 3-AMOEBA (N-induce, n-field) model is equivalent to the 3-AMOEBA (n-induce, N-field) model for the special case of water and the small sulfate dianion; in general these would be distinct schemes.

System	**17 waters**	**216 waters**	**512 waters**	**$SO_4^{2-}(H_2O)_7$**
Model	**Energy in kcal/mol (% Error with respect to AMOEBA reference)**			
AMOEBA	-96.698	-653.705	-1854.959	-47.428
2-AMOEBA (vacuum)	-49.420 (-48.9%)	-440.810 (-32.6%)	-1260.705 (-32.0%)	-80.312 (69.3%)
2-AMOEBA (N-induce, n-field)	-79.130 (-18.2%)	-582.504 (-10.9%)	-1652.721 (-10.9%)	-57.195 (20.6%)
2-AMOEBA (N-induce, N-field)	-90.385 (-6.5%)	-629.887 (-3.6%)	-1786.217 (-3.7%)	-48.390 (2.0%)
3-AMOEBA (vacuum)	-92.300 (-4.5%)	-661.765 (1.2%)	-1883.844 (1.6%)	-40.790 (-14.0%)
3-AMOEBA (N-induce, n-field)	-95.234 (-1.5%)	-655.975 (0.4%)	-1864.997 (0.5%)	-45.300 (-4.5%)
3-AMOEBA (N-induce, N-field)	-96.187 (-0.5%)	-654.364 (0.1%)	-1857.775 (0.2%)	-47.011 (-0.9%)
4-AMOEBA (vacuum)	-96.944 (0.25%)	-654.857 (0.18%)		
Model	**Gradient Inner Product (Gradient RMS % Error) with respect to AMOEBA reference**			
2-AMOEBA (vacuum)	0.9769 (403.3%)	0.9436 (226.4%)	0.9280 (255.0%)	0.7866 (471.6%)
2-AMOEBA (N-induce, n-field)	0.9984 (181.4%)	0.9965 (88.9%)	0.9945 (99.7%)	0.9745 (137.0%)
2-AMOEBA (N-induce, N-field)	0.9998 (77.4%)	0.9996 (34.6%)	0.9993 (39.8%)	0.9965 (34.0%)
3-AMOEBA (vacuum)	0.9990 (59.2%)	0.9980 (30%)	0.9974 (38.1%)	0.4320 (2202%)
3-AMOEBA (N-induce, n-field)	0.9999 (23.8%)	0.9998 (9.4%)	0.9997 (12.3%)	0.9981 (34.1%)
3-AMOEBA (N-induce, N-field)	1.0000 (9.7%)	1.0000 (3.1%)	1.0000 (4.5%)	0.9998 (9.1%)
4-AMOEBA (vacuum)	1.0000 (6.66%)	1.0000 (5.0%)		

Table 2. Errors in the many-body expansion under vacuum embedding with large cluster bodies. Energy units are kcal/mole.

System	1600 waters	7000 waters	32000 waters	96000 waters	288,000 waters
Model	**Energy in kcal/mol**				
AMOEBA	-5653.02	-28996.41	-178749.25	-417889.64	-1184234.37
	(% Error with respect to AMOEBA reference)				
3-AMOEBA (N/10)	-0.01%	0.05%	0.01%	-0.03%	0.09%
3-AMOEBA (N/20)	-0.53%	-0.19%	-0.06%	-0.06%	-0.007%
3-AMOEBA (N/30)	-0.77%	-0.54%	-0.13%	-0.15%	-0.09%
3-AMOEBA (1)	-2.60%	-1.22%	-1.70%	-0.66%	-2.33%
Model	**Gradient Inner Product (Gradient RMS % Error) with respect to AMOEBA reference**				
3-AMOEBA (N/10)	0.997 (37.1%)	0.999 (18.0%)	1.000 (20.2%)	0.999 (20.4%)	1.000 (16.1%)
3-AMOEBA (N/20)	0.995 (51.8%)	0.998 (33.9%)	0.999 (29.4%)	0.999 (24.6%)	0.999 (19.3%)
3-AMOEBA (N/30)	0.993 (62.2%)	0.997 (46.0%)	0.999 (34.9%)	0.999 (29.0%)	0.999 (22.3%)
3-AMOEBA (1)	0.992 (67.4%)	0.993 (68.8%)	0.994 (77.8%)	0.993 (70.4%)	0.992 (75.7%)

Table 3. *CPU Performance of TINKER7 MD Simulations for AMOEBA and 3-AMOEBA Models.* CPU timings are reported on various small to large water boxes ranging from 4,800-864,000 atoms for the mutual polarization model AMOEBA and the different fragmentation clusters under the many-body approximations for 3-AMOEBA. The simulation protocol consisted of a standard RESPA[48] integrator based on the separation of intra- and intermolecular forces with respective time steps of 0.1fs and 2.0 fs, a 9.0 Å vdW cutoff, and a 7.0 Å real space cutoff for particle-mesh Ewald (PME) electrostatics. Timings are reported in nanoseconds/day and are referenced with respect to the best AMOEBA OpenMP implementation in TINKER7. Timing results were obtained on a Cray XC30 using 12-core Intel "Ivy Bridge" processor at 2.4 GHz, 24 cores per node. All 3-AMOEBA results were fixed at 3600 cores.

System	AMOEBA	3-AMOEBA (N/30)		3-AMOEBA (N/20)		3-AMOEBA (N/10)	
# waters	**(ns/day)**	ns/day	**Speed up**	ns/day	**Speed up**	ns/day	**Speed up**
1600	1.63	6.28	**3.85**	8.04	**4.93**	5.54	**3.40**
7000	0.36	2.56	**7.11**	1.41	**3.90**	1.04	**2.87**
32000	0.07	0.54	**7.54**	0.44	**6.07**	0.23	**3.18**
96000	0.02	0.18	**9.18**	0.12	**5.86**	0.05	**2.51**
288000	0.005	0.06	**10.79**	0.04	**7.67**	0.01	**1.78**

FIGURE CAPTIONS

Figure 1. *Decay of polarization with respect to distance under the many body expansion for the 2-body and 3-body energy using vacuum embedding.* (a) Under this embedding scheme 2-body polarization is completely attractive, with (b) cooperative and anti-cooperative behavior only evident at the 3-body level in which direct polarization is exact. (c) shows the differences in force magnitudes of atom pairs between vacuum embedding and the full N-body AMOEBA model as a function of interatomic distance.

Figure 2. *Decay of polarization with respect to distance under the many body expansion for the 2-body and 3-body energy using asymmetric N,n-dimensional embedding.* Under this embedding scheme both (a) 2-body and (b) 3-body polarization shows cooperative and anti-cooperative behavior.

Figure 3. *Decay of polarization with respect to distance under the many body expansion for the 2-body and 3-body energy using symmetric N,N-dimensional embedding.* Under this embedding scheme both (a) 2-body and (b) 3-body polarization shows cooperative and anti-cooperative behavior but now energy errors are smaller than in Figure 2.

Figure 4. *Geometric variables that described the water dimer.* The variables that describe the ideal hydrogen bond displayed as a function of distance and the theta angle corresponding to a tilt of the plane of the acceptor with respect to the donor. Reprinted with permission [49]; copyright 2005 American Institute of Physics.

Figure 5. *Difference in 2-body forces* $\vec{f}_{ij}^{N,N} - \vec{f}_{ij}^{n,n}$ *between atom pairs under N.N and vacuum embedding.* (a) as a function of interatomic distance and (b) as a function of both donor-hydrogen-acceptor angle and acceptor-hydrogen distance. We display just the x-component of the force difference, since the y- and z-components display equivalent behavior.

Figure 6. *Differences in 3-body forces* $\vec{f}_{ij}^{N,N} - \vec{f}_{ij}^{n,n}$ *between atom pairs under N.N and vacuum embedding*. (a) as a function of both donor-hydrogen-acceptor angle and acceptor-hydrogen distance. (b) Peak at 1.6-2.0 radians and ~4 Å in (a) displayed with respect to a different donor-hydrogen-acceptor angle. We display just the x-component of the force difference, since the y- and z-components display equivalent behavior.

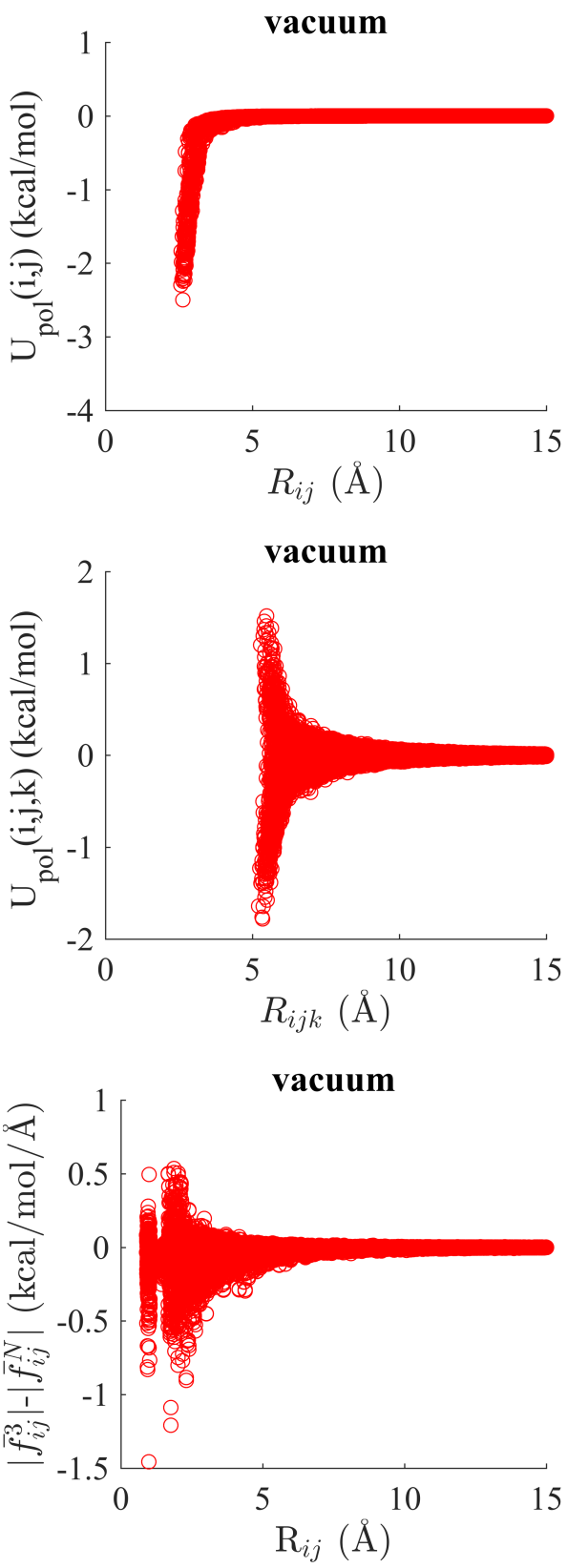

Figure 1. Demerdash and Head-Gordon

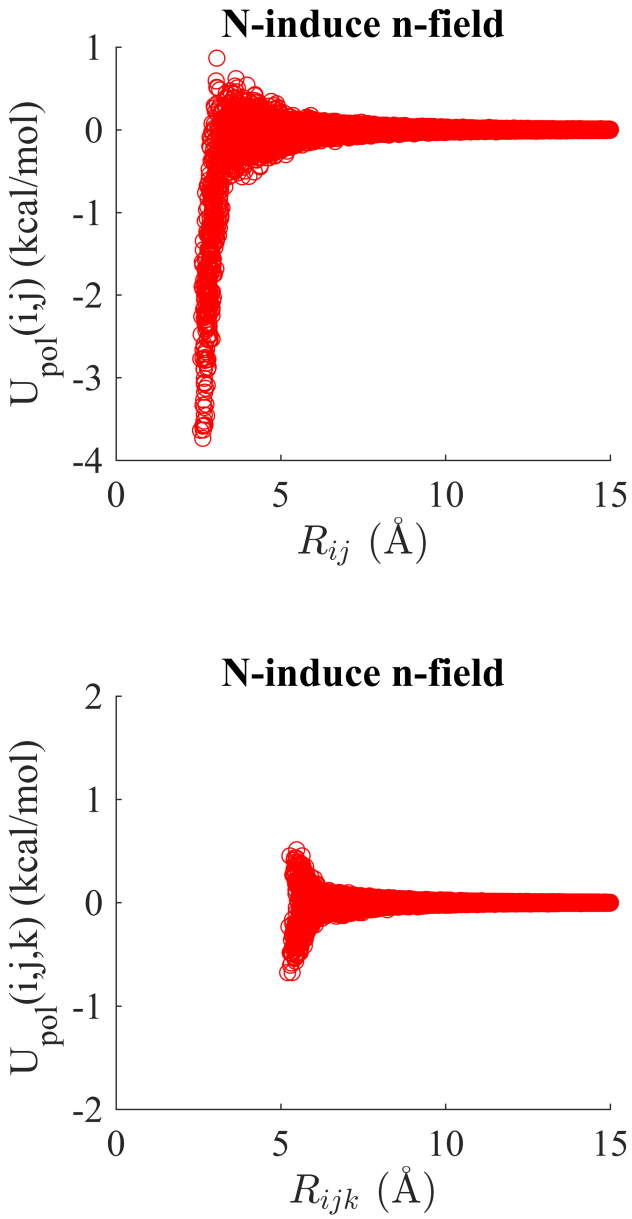

Figure 2. Demerdash and Head-Gordon

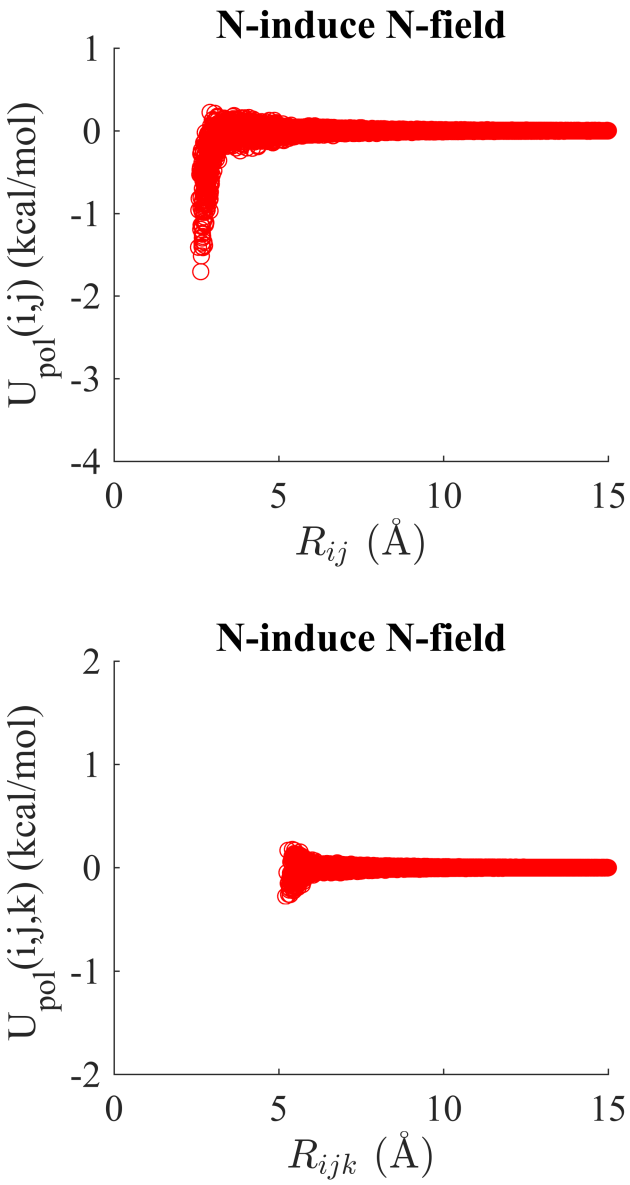

Figure 3. Demerdash and Head-Gordon

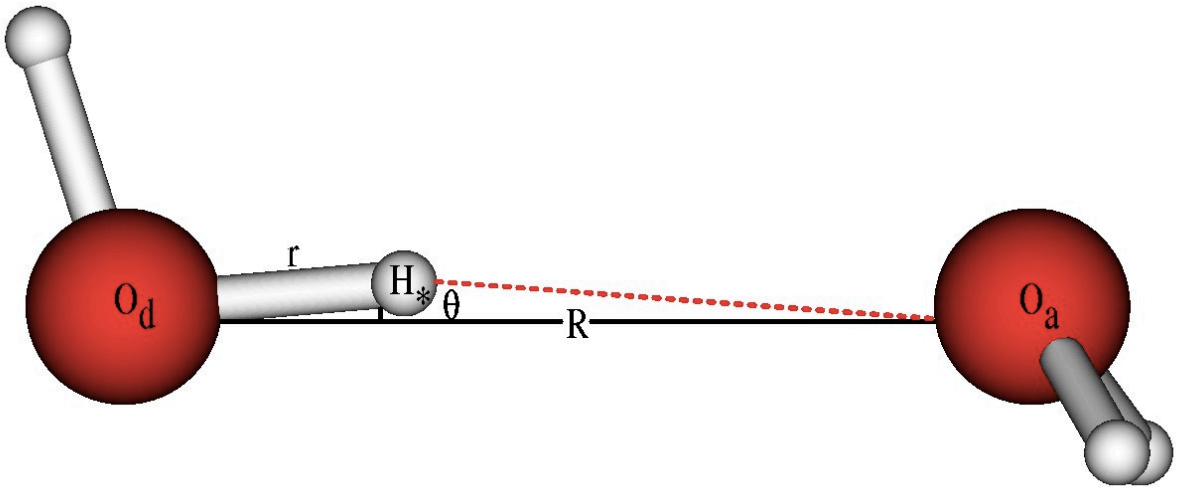

Figure 4. Demerdash and Head-Gordon

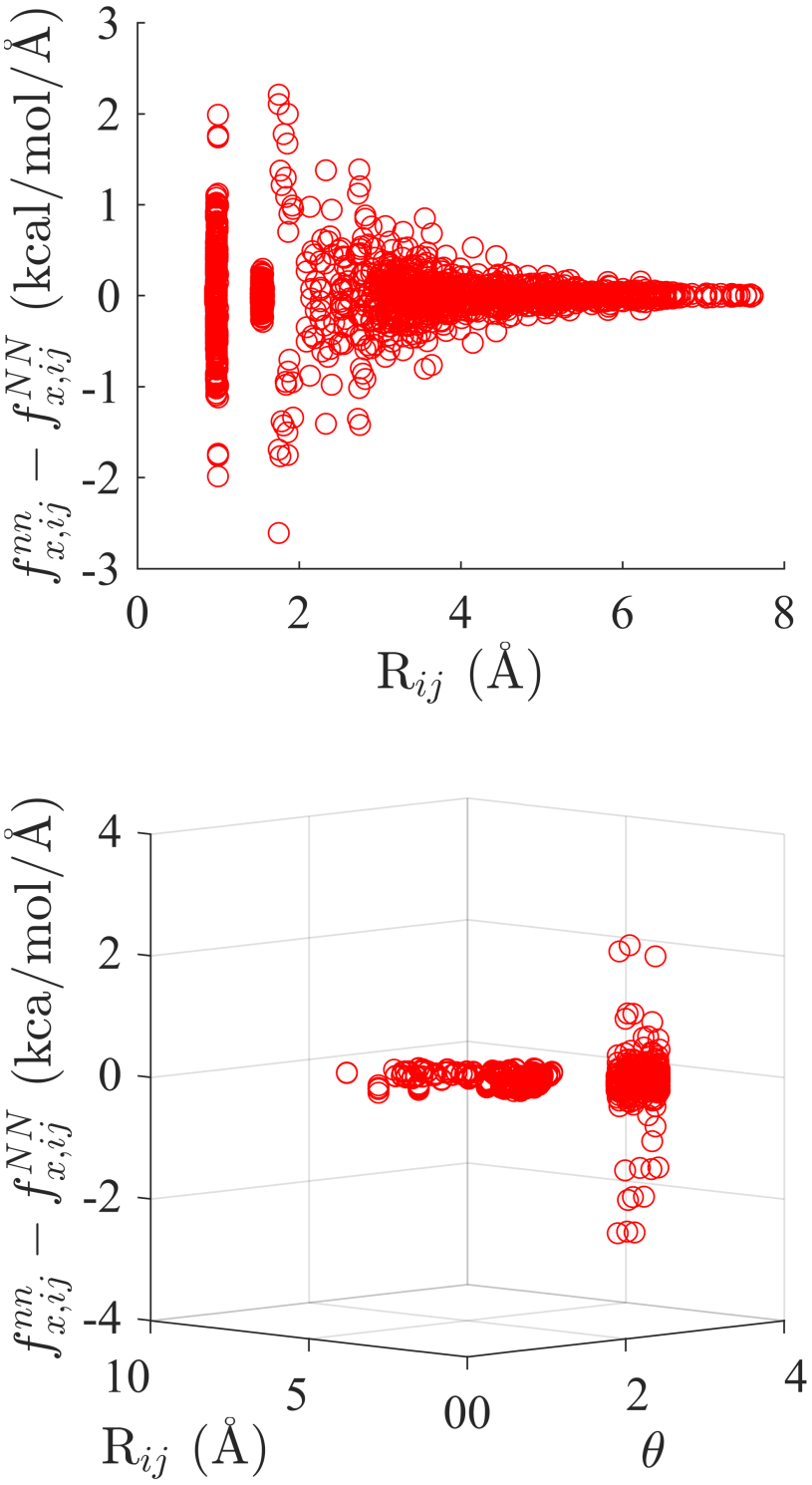

Figure 5. Demerdash and Head-Gordon

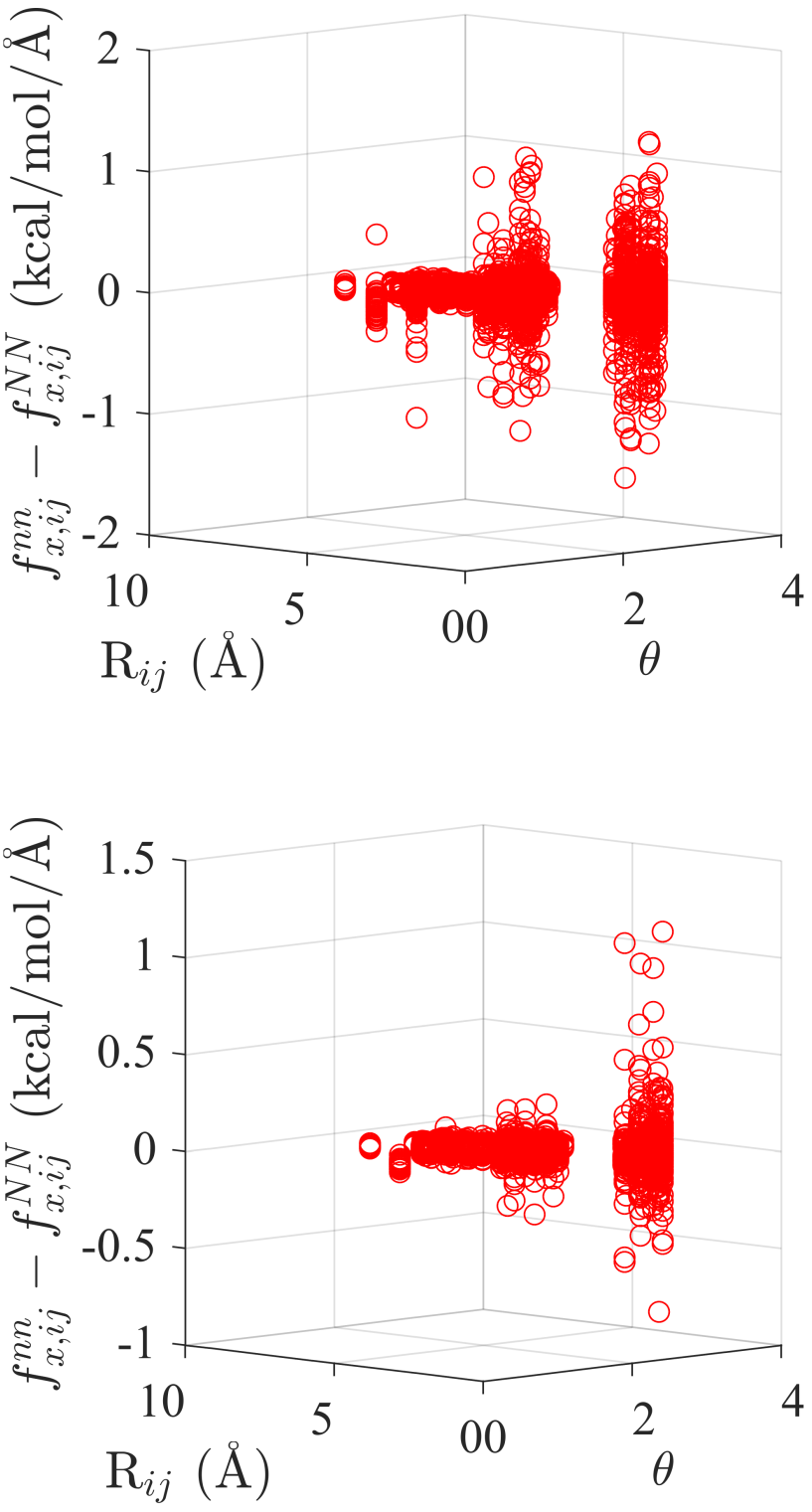

Figure 6. Demerdash and Head-Gordon