JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

# Capturing Many-Body Interactions with Classical Dipole Induction Models

Chengwen Liu,[†] Rui Qi,[†] Qiantao Wang,[‡] J.-P. Piquemal,[†,§,∥] and Pengyu Ren*[,†]

[†]Department of Biomedical Engineering, The University of Texas at Austin, Austin, Texas 78712, United States

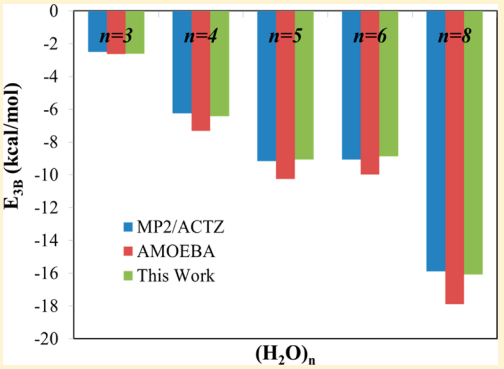
[‡]Key Laboratory of Drug Targeting and Drug Delivery System of Education Ministry, West China School of Pharmacy, Sichuan University, Chengdu, Sichuan 610041, China

[§]Laboratoire de Chimie Théorique, Sorbonne Universités, UPMC, UMR 7616 CNRS, Paris 75252, France

[∥]Institut Universitaire de France, Paris Cedex 05, 75231, France

Supporting Information

**ABSTRACT:** The nonadditive many-body interactions are significant for structural and thermodynamic properties of condensed phase systems. In this work we examined the many-body interaction energy of a large number of common organic/biochemical molecular clusters, which consist of 18 chemical species and cover nine common organic elements, using the Møller−Plesset perturbation theory to the second order (MP2) [Møller et al. *Phys. Rev.* **1934**, *46*, 618.]. We evaluated the capability of Thole-based dipole induction models to capture the many-body interaction energy. Three models were compared: the original model and parameters used by the AMOEBA force field, a variation of this original model where the damping parameters have been reoptimized to MP2 data, and a third model where the damping function form applied to the permanent electric field is modified. Overall, we find the simple classical atomic dipole models are able to capture the 3- and 4-body interaction energy across a wide variety of organic molecules in various intermolecular configurations. With modified Thole models, it is possible to further improve the agreement with MP2 results. These models were also tested on systems containing metal/halogen ions to examine the accuracy and transferability. This work suggests that the form of damping function applied to the permanent electrostatic field strongly affects the distance dependence of polarization energy at short intermolecular separations.




## 1. INTRODUCTION

Molecular simulations are widely used to study the structural, energetic, and thermodynamic properties of condensed phase systems. It is essential for underlying classical potentials, or force fields, to accurately describe the intra- and intermolecular interactions. In the past decade, a major advancement in classical force fields has been the explicit incorporation of the many-body effect, to improve their accuracy and transferability.[2] The many-body interactions can significantly affect various chemical/physical properties of matters. The importance of incorporating explicit polarization has been demonstrated on studying the well-known many anomalous properties of water at various physical conditions;[3−5] molecular structure and polarizability of organic compounds;[6−10] solvation free energy of inorganic salt ions;[11−14] and binding affinity of ligand-protein complexes.[11,15]

In the current polarizable force fields, the many-body interaction energy is explicitly treated through the introduction of electronic polarization mainly in three different ways: (1) fluctuating charge models,[5,16−18] where charge fluctuating is applied to represent the system response to the electrostatic potential; (2) Drude oscillator models,[4,19−21] where the Drude particles on polarizable sites are used to describe the response to the surroundings; and (3) atomic induced dipole models,[3,6,22−24] where induced dipoles are employed to respond to surrounding electrostatic field. Force fields based on these models have been developed for general molecular simulations.[2,25−27]

The AMOEBA (atomic multipole optimized energetics for biomolecular simulation) force field utilizes the induced dipole model, where a point dipole is induced at each polarizable site according to the electric field experienced by that site. Molecular polarization is achieved via iterative dipole induction with distributed atomic polarizabilities based on Thole's damping scheme.[3,6] In this simple yet effective scheme, damping is crucial as it prevents polarization catastrophe at a close distance of two polarizable sites and allows reproducing anisotropic molecular polarizability. The Thole-based polarization model in the AMOEBA potential was parametrized to reproduce molecular polarizability as well as total molecular interaction energy (e.g., water cluster association energy).[3] It

**Received:** March 3, 2017
**Published:** May 8, 2017

DOI: 10.1021/acs.jctc.7b00225
*J. Chem. Theory Comput.* 2017, 13, 2751−2761

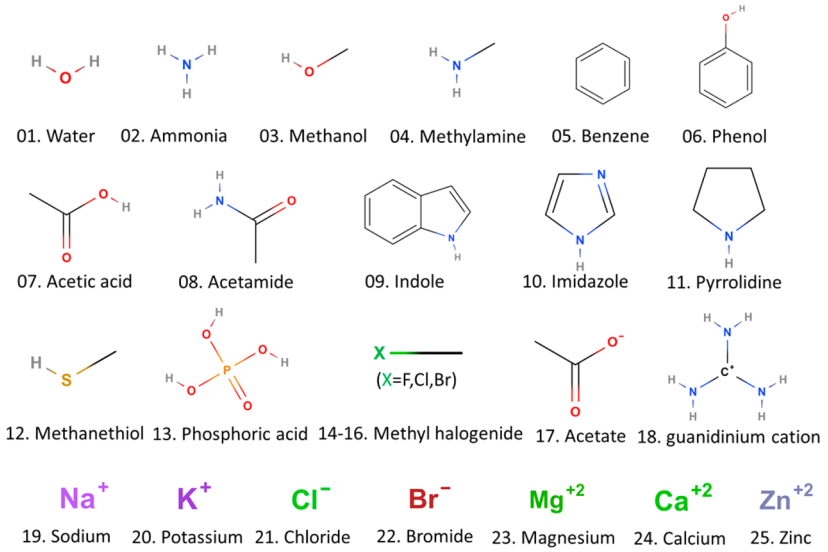



**Figure 1.** An overview of all the chemical species to form molecular clusters of many-body decomposition databases. Trimers and tetramers were constructed from these species. For water, larger clusters up to octamer were included. The composition of molecular clusters can be seen in the following sections and the SI. These chemical formula images were generated via MOLVIEW (http://molview.org).

was shown for water clusters that the current AMOEBA polarization energy gives a reasonable trend of many-body interactions while overestimating 3- and 4-body interaction energy ($E_{3B}$ and $E_{4B}$) consistently.[28,29]

Here we systematically examine the performance of classical polarization models to explore the possibility to improve the polarization energy in the current AMOEBA potential. The nonadditive many-body interaction energy can arise from induction (polarization), exchange-repulsion, and dispersion.[30] The nonadditivity of exchange-repulsion is often believed to be less prominent in most cases[30] except for the hydrated metal cation complexes.[31] While there have been efforts to model the $E_{3B}$ with polarization and dispersion in organic molecular systems,[32−34] improvements of biomolecular force fields have been focusing on incorporating many-body polarization. *In this study, we directly compare the many-body polarization energy from classical models to the MP2 total $E_{3B}$ and $E_{4B}$ of common molecular systems including ions.* In addition to the current AMOEBA polarization model, two new models were examined: (1) a variation of the current AMOEBA model where the damping parameters have been directly optimized to MP2 many-body interaction energy and (2) a model where the permanent field damping function is modified from that of Thole.

The rest of this paper is organized as follows: in the Methods section, AMOEBA polarization framework will be briefly revisited. The many-body decomposition databases that are used to parametrize and validate these models, together with two new models, will also be described in this section. Model parametrization and comprehensive comparisons of the polarization models with MP2 results are made in the Results and Discussion section. The Conclusion is drawn in the last section.

## 2. METHODS

**2.1. AMOEBA Polarization Framework.** Different from the simple fixed-point-charge (partial charge) force fields, such as AMBER[35] and CHARMM,[36] the AMOEBA force field uses atomic point multipoles (monopole, dipole, and quadrupole) to describe the electrostatic potential and field on each atom site.

Polarization is explicitly treated by mutual induction of dipoles at polarizable sites (located at atomic centers). A point dipole moment is induced at each polarizable site according to the electric field experienced by that site

$$\mu_i^{ind} = \alpha_i(E_i^{dir} + E_i^{mut}) \tag{1}$$

where $\alpha_i$ is the atomic polarizability on site $i$; $E_i^{dir}$ is the "*direct*" electric field generated by permanent multipoles of other sites; and $E_i^{mut}$ is the "*mutual*" field generated by induced dipoles of other sites. The $E_i^{dir}$ and $E_i^{mut}$ are expressed as

$$E_i^{dir} = \sum_j T_{ij} M_j \tag{2}$$

$$E_i^{mut} = \sum_{j \neq i} T_{ij}^{11} \mu_j^{ind} \tag{3}$$

where $T_{ij}$ in eq 2 is the multipole-multipole interaction matrix;[6] and $M_j$ is the polytensor of permanent multipoles (charge, dipole, and quadrupole). In eq 3 $T_{ij}^{11}$ is the dipole−dipole interaction matrix, and $\mu_j^{ind}$ is the induced dipole moment of site $j$. The induced dipole on each polarizable site is solved iteratively to obtain the converged dipole values.

With self-consistent field (SCF) converged induced dipole, the polarization energy can be obtained through

$$U_{ele}^{ind} = -\frac{1}{2} \sum_i (\mu_i^{ind})^T E_i^{dir} \tag{4}$$

To ensure the finite nature of the intermolecular induction effect, Thole used a damping scheme in which a "smeared" charge distribution replaces one of the point dipoles, and thus dipole interactions are damped.[37] As a result, the dipole interaction energy approaches a finite value instead of becoming infinite as the atomic separation approaches zero. Thole's scheme is very successful in reproducing dipole molecular polarizability tensors for a broad range of organic molecules using element-based isotropic atomic polarizabilities.[37,38] This scheme has been adopted by the AMOEBA force field to model polarization energy.[3,6] The electric fields due to both the permanent multipoles and the induced dipoles are


DOI: 10.1021/acs.jctc.7b00225
J. Chem. Theory Comput. 2017, 13, 2751−2761

Table 1. Overview of the Many-Body Decomposition Databases[c]

| purpose of use | database | species | data point | brief description |
|---|---|---|---|---|
| parametrization | Water-4568 | 1 | 4 | water clusters; cyclic tetramer, cyclic pentamer, prism hexamer, and S4-symmetric octamer |
| | Tetramers | 11 | 26 | organic tetramers; 10 homo- and 16 heteroclusters[a] |
| validation | Trimers-Distance[b] | 11 | 124 | organic trimers; 11 homotrimers with various separations[a] |
| | Water6-Extra | 1 | 7 | water hexamers; 7 structures other than cyclic configuration |
| | 3B-69 set[b] | 5 | 14 | organic trimers of Beran and co-workers[32,33a] |
| | Trimers-Transferability[b] | 8 | 15 | organic trimers used to test model transferability[a] |
| | metal/halogen ions[b] | 8 | 39 | metal and halogen ion−water clusters; 11 tetramers and 7 trimers with each having 4 intermolecular separations |

[a]Hereafter "organic" means all species in Figure 1 except for ions. [b]Only the $E_{3B}$ was provided for the trimers in these sets. [c]The number of chemical species and data points (each point includes both $E_{3B}$ and $E_{4B}$ except for trimers) was listed. A detailed description of the database, molecular graphics, and XYZ coordinates were given in the SI.

damped using the same function in the current AMOEBA model. This is accomplished by modifying the interaction $T$ matrices in the corresponding orders (only $T_\alpha$ is shown here; see the higher order $T$ matrices in the Supporting Information, SI).

$$T_\alpha = -\lambda_3 \frac{R_\alpha}{R^3} \quad (\alpha = x, y, z) \tag{5}$$

The form of $\lambda_3$ that the current AMOEBA uses is

$$\lambda_3(r) = 1 - e^{-au^3(r)} \tag{6}$$

where $u(r) = r_{ij}/(\alpha_i\alpha_j)^{1/6}$ is the scaled distance between sites $i$ and $j$; and $r_{ij}$ and $\alpha_i$ are the real distance and atomic polarizability, respectively. The factor $a$ is the dimensionless width parameter of the smeared charge distribution and effectively controls the damping strength. The universal damping factor was determined to be 0.39 for both the *direct* and *mutual* part in the current AMOEBA by considering the molecular polarizabilities and total associate energy of water clusters up to hexamers.[3]

**2.3. Many-Body Decomposition Databases.** Databases of a series of molecular clusters were constructed to parametrize and validate the polarization models. The $E_{3B}$ and $E_{4B}$, obtained from MP2 based *ab initio* decomposition (see the SI), were employed in this work.

*2.3.1. Chemical Species Involved in Databases.* In total 25 representative chemical species are included in these databases (Figure 1), containing common solvents, organic compounds, amino acid side-chain analogs, and inorganic salt ions. Nine common chemical elements (C, H, O, N, P, S, F, Cl, and Br) and seven monoatom ions ($Na^+$, $K^+$, $Mg^{2+}$, $Ca^{2+}$, $Zn^{2+}$, $Cl^-$, and $Br^-$) are covered.

*2.3.2. Database Overview.* Here we give an overview of the databases (see Table 1), and a detailed description can be seen in the SI. The parametrization databases include Water-4568 and Tetramers sets. They are composed of the first 11 chemical species listed in Figure 1. The equilibrium geometry clusters are formed in the parametrization databases. The validation databases include the Trimers-Distance set, where the intermolecular separations were systematically varied to generate closer distance configurations (Figure 2), as well as other molecular clusters formed by the last 14 chemical species listed in Figure 1.

*2.3.3. Procedure To Generate Short-Distance Trimers.* Trimers with short intermolecular distances were generated according to the following procedure. To take water as an example, water trimer was first fully optimized to its MP2 equilibrium geometry (shown in Figure 2). By stepwise reducing the "equilibrium" intermolecular distances, $d_1$ and $d_2$, a series of water trimers were generated. For other molecules, the same procedure was applied to generate the configurations with varied intermolecular distances. For the $M(H_2O)_2$ trimer (where M = $Na^+$, $K^+$, $Mg^{2+}$, $Ca^{2+}$, and $Zn^{2+}$), $d_1$ and $d_2$ were selected as the two M···O distances. Considering the computational cost of MP2 many-body decomposition calculations, a smaller number of representative intermolecular separations were selected to generate the trimers for other molecules than water. Finally, 124 trimers were constructed in total for the Trimers-Distance set.

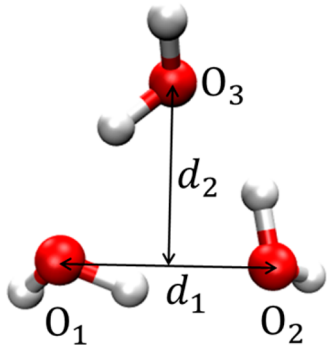



**Figure 2.** A schematic illustration of water trimers with different intermolecular distances. Three oxygen atoms of waters were denoted as $O_1$, $O_2$, and $O_3$. $d_1$ is the distance between $O_1$ and $O_2$, and $d_2$ is the distance between $O_3$ and the midpoint of $O_1$ and $O_2$. These two distances were systematically varied.

**2.4. Variation of AMOEBA Polarization Models.** The current AMOEBA polarization model emphasizes the aspect of molecular polarizability and can well capture the total interaction energy of water clusters. However, it has not been systematically examined beyond water clusters. In addition, the current implementation was shown to overestimate $E_{3B}$ and $E_{4B}$ for water clusters according to MP2/aug-cc-pvtz calculations.[29] Therefore, we explore possible improvements of the current model that can better capture the many-body interactions, including the distance and oriental dependence. We propose two new variations of AMOEBA polarization model:

**Model 1:** this model is the same as the current AMOEBA but with a damping parameter ($a$ in eq 6) "reoptimized" to match the MP2 $E_{3B}$ and $E_{4B}$ energy. A "universal" damping factor of 0.34 is used by Model 1 (except for divalent metal


DOI: 10.1021/acs.jctc.7b00225
J. Chem. Theory Comput. 2017, 13, 2751−2761

Table 2. Parameters Used in the Three Polarization Models for Organic Molecules and Metal/Halogen Ions[c]

| | AMOEBA[a] | | Model 1[a] | | | Model 2[a] | | |
|---|---|---|---|---|---|---|---|---|
| | $\alpha_{atom}$ | $a_{dir}(=\alpha_{mut})$ | $\alpha_{atom}$ | $a_{dir}$ | $a_{mut}$ | $\alpha_{atom}$ | $a_{dir}$ | $a_{mut}$ |
| organics[b] | | 0.390 | | 0.340 | | | 0.750 | 0.390 |
| $Na^+$ | 0.120 | 0.390 | 0.060 | 0.340 | | 0.060 | 0.750 | 0.390 |
| $K^+$ | 0.780 | 0.390 | 0.780 | 0.340 | | 0.780 | 0.750 | 0.390 |
| $Mg^{2+}$ | 0.080 | 0.095 | 0.040 | 0.060 | 0.340 | 0.040 | 0.350 | 0.390 |
| $Ca^{2+}$ | 0.550 | 0.159 | 0.750 | 0.280 | 0.340 | 0.750 | 0.750 | 0.390 |
| $Zn^{2+}$ | 0.260 | 0.210 | 0.260 | 0.340 | | 0.260 | 0.750 | 0.390 |
| $Cl^-$ | 4.000 | 0.390 | 4.000 | 0.340 | | 4.000 | 0.450 | 0.390 |
| $Br^-$ | 5.650 | 0.390 | 5.650 | 0.340 | | 5.650 | 0.450 | 0.390 |

[a]AMOEBA and Model 1 use the same damping function in both *direct* and *mutual* induction (eq 6); Model 2 uses a different damping function for the *direct* induction (eq 7). [b]Molecules 1−18 in Figure 1. All atomic polarizabilities were kept the same as AMOEBA09[6] except for O and C in O=C, which were increased from 0.837/1.334 to 1.45/1.75 Å$^3$ in Model 1s and 2. [c]These parameters include atomic polarizability ($\alpha_{atom}$) (in Å$^3$) and damping factors ($a_{dir}$ and $a_{mut}$).

ions). The molecular polarizabilities are rechecked since the *mutual* damping is affected during this process.

**Model 2:** this model utilizes a new damping function for *direct* (permanent) field and keeps the current Thole damping function of AMOEBA for the *mutual* induction. This was done as the distance dependence of total polarization energy is mostly determined by the *direct* induction. On the other hand, in the Thole damping scheme, the anisotropic molecular response is determined by the *mutual* induction alone, with just a few isotropic atomic polarizabilities.

For both models, we retain the same *atomic polarizability parameters* used by the current AMOEBA force field. The only exceptions are from $CH_3COOH$ and $CH_3CONH_2$, where the atomic polarizabilities of carbon and oxygen on the carbonyl group need to be slightly modified to better match the molecular polarizability and MP2 many-body energy.

By examining the distance dependence behavior of the $E_{3B}$ on the Trimers-Distance set, we found that the following function gave the best performance with an appropriate damping factor. Thus, the damping function for the *direct* induction in Model 2 is

$$\lambda_3'(r) = 1 - e^{-au(r)^{3/2}} \quad (7)$$

where $u(r)$ is the scaled distance and the same as that in eq 6. The damping functions for higher order multipole interactions can be easily derived through the chain rule relationships, and their mathematical forms are given in the SI.

These new models were trained on the parametrization database described above, to mainly identify a single best "universal" damping factor. All these parameters are described in the Parametrization section and Table 2.

**2.5. Computational Details.** Structures of the pure water clusters, including $(H_2O)_n$ ($n$ = 3, 4, and 5, the cyclic configurations were selected), $(H_2O)_8$ ($S_4$-symmetric), and eight $(H_2O)_6$, were taken from Wang et al.[39] without further optimization. The initial structures of ammonia, benzene, and methanol trimers and tetramers and their mixed tetramers were from the literature.[40−45] The remaining clusters were generated by duplicating and translating monomers in three-dimensional Cartesian space. These structures were optimized by using HF/3-21G followed by the MP2/6-31+G(d) level of theory in the GAUSSIAN 09 package.[46] The regularized SAPT(DFT) calculations were performed on 14 dimer pairs using CAMCASP[47,48] and NWCHEM 4.2[49] packages. The PBE0[50] functional and the aug-cc-pvtz[51] basis set were employed. The regularized parameter $\eta$ was chosen to be 3.0 as suggested by Missquitta.[48] For water dimers, the absolutely localized molecular orbital (ALMO)[52,53] EDA was carried out in the QCHEM 4.2 package[54] at the HF/6-311++G(2d,2p) level of theory.

The many-body interaction energy decomposition was performed in the QCHEM 4.2 package.[54] Single point energy of each cluster and its subclusters at the MP2/aug-cc-pvtz[51] level were calculated for pure water clusters. Due to the expensive computational cost, for the clusters other than pure water, the RI-MP2 (TRIM)[55] method combining with the cc-pvtz[56] basis set was used. The auxiliary basis set was chosen to be rimp2-cc-pvtz as recommended.[54] Due to the availability of the basis set, 6-311++G(2d,2p) was used for $K^+$-water and $Ca^{2+}$-water clusters, and the G3Large basis set[57] was used for $Zn^{2+}$-water clusters. The molecular polarizabilities for the 36 molecules from the S101 database[58] were calculated using GAUSSIAN 09 at the $\omega$B97XD[59]/aug-cc-pvtz[51] level of theory. The *exact* polarizability tensors from GAUSSIAN were diagonalized to obtain three eigenvalues.

For the classical models, the *polarization* energy from AMOEBA calculations was used to compute the $E_{3B}$ and $E_{4B}$ as the remaining contribution is pairwise additive. Model 2 was embedded in TINKER source files by modifying the interaction $T$ matrices according to the damping functions we proposed. The ANALYZE program was used to obtain the polarization energy of each molecular cluster. The atomic multipole (charge, dipole, and quadrupole) and polarizability parameters were taken from previous charge penetration work.[58] These parameters were constructed/assigned by POLTYPE,[60] which is an automatic tool to generate the AMOEBA force field parameters for small molecules. The same molecular geometries were used in both MP2 and classical model calculations to obtain the $E_{3B}$ and $E_{4B}$.

## 3. RESULTS AND DISCUSSION

The performance of the three polarization models will be examined based on the MP2 many-body interaction energy in section 3.1 Parametrization and section 3.2 Validations. In section 3.3, the polarization energy from our models and two quantum EDA methods will be compared.

**3.1. Parametrization.** The parameters of the current AMOEBA and two new models are tabulated in Table 2. The main parameter in these classical polarization models is the damping factor ($a$), which is determined by comparing the $E_{3B}$ from the models to MP2 values of the parametrization databases (Water-4568 and Tetramers sets). Root mean square


DOI: 10.1021/acs.jctc.7b00225
J. Chem. Theory Comput. 2017, 13, 2751−2761

Table 3. Statistics Evaluation of the Three Polarization Models on the $E_{3B}$ and $E_{4B}$ of Water-4568 and Tetramers Sets[b]

| | | $E_{3B}$ | | | $E_{4B}$ | | |
|---|---|---|---|---|---|---|---|
| database | statistics[a] | AMOEBA | Model 1 | Model 2 | AMOEBA | Model 1 | Model 2 |
| Water-4568 | MUE | 1.27 | 0.12 | 0.17 | 0.56 | 0.35 | 0.36 |
| | MSE | −1.27 | −0.11 | −0.02 | −0.56 | −0.35 | −0.36 |
| | RMSE | 1.34 | 0.17 | 0.17 | 0.58 | 0.37 | 0.37 |
| Tetramers | MUE | 0.37 | 0.22 | 0.30 | 0.18 | 0.13 | 0.13 |
| | MSE | −0.25 | 0.09 | 0.19 | −0.11 | −0.08 | −0.07 |
| | RMSE | 0.45 | 0.30 | 0.36 | 0.23 | 0.17 | 0.17 |

[a]MUE: mean unsigned error; MSE: mean signed error; RMSE: root-mean-square error. [b]All statistics was calculated using the $E_{3B}$ and $E_{4B}$ from MP2/aug-cc-pvtz (for Water-4568) and RI-MP2/cc-pvtz (for Tetramers) as the references.

error (RMSE) of the $E_{3B}$ from each model with respect to the MP2 was used to optimize the parameters. The atomic polarizabilities were mostly kept the same as those in the current AMOEBA (Table 2).

In Model 1, the optimal damping factor was determined to be 0.34, which means a slightly stronger damping on the electrostatic field (both permanent and mutual) is applied than that of the current AMOEBA. The molecular polarizabilities of all 36 molecules from the S101 database[58] were re-examined with the new damping factor in the *mutual* induction. The molecular polarizabilities from ωB97XD/aug-cc-pvtz were employed as the reference since they well reproduce the available experimental data. For all 36 molecules, this damping factor (0.34) gives an RMSE value of 1.03 Å$^3$ in terms of the three components with respect to those of DFT, while the RMSE value of the current AMOEBA is 0.98 Å$^3$ (see the SI for each component value). This is expected as the molecular polarizability, unlike polarization energy, has a relatively weak dependence on the damping factor.[3]

Other parametrization strategies were also considered. In Model 1, instead of using the same damping parameter for both *direct* and *mutual* induction, we also experimented: 1) two different damping parameters for *direct* and *mutual* induction and 2) the element-based damping factor for the *direct* induction with a common damping parameter for the *mutual* induction. Based on the Tetramers set, these two strategies essentially gave the same results as the simpler one-parameter Model 1, and thus only Model 1 is presented here.

In Model 2, the damping of *mutual* induction (both the function and parameter) was kept the same as the current AMOEBA, and thus the molecular polarizabilities remain unchanged. Only the *direct* damping factor was parametrized, and the optimal damping factor was determined to be 0.75.

For the metal/halogen ion systems, especially the divalent cations, both the damping factors, of two models (Models 1 and 2) and atomic polarizabilities, were optimized to capture the many-body interaction energy.

Using the MP2 $E_{3B}$ and $E_{4B}$ as the reference, the performance of three polarization models is compared in Table 3. The individual $E_{3B}$ and $E_{4B}$ are plotted in Figure 3. Model 1 and Model 2 both show better agreement with the MP2 $E_{3B}$ and $E_{4B}$ than the current AMOEBA. For example, for the Water-4568 set, the $E_{3B}$ RMSE value is reduced from 1.34 (the current AMOEBA) to 0.17 kcal/mol for both Model 1 and Model 2, with an ∼8-fold of improvement. Other statistics such as mean unsigned error (MUE) and mean signed error (MSE) are all reduced. For the Tetramers set, the improvement is not as apparent, largely due to the relatively weak many-body interaction energy. As for the individual $E_{3B}$ and $E_{4B}$ in these two sets (Figure 3), significant improvement of two new

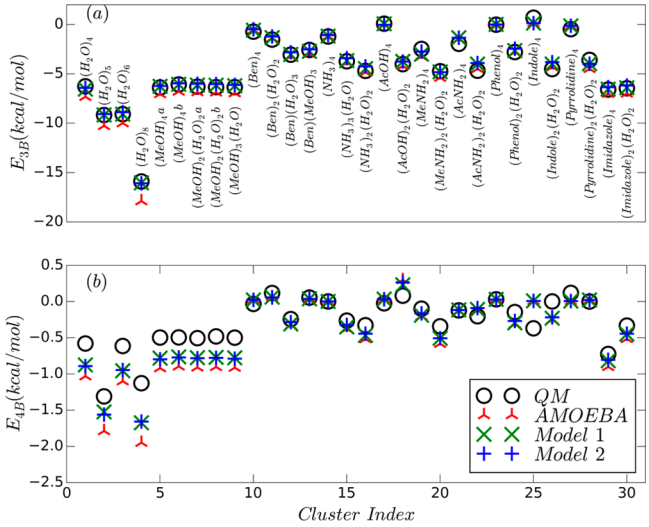



Figure 3. Plots of (a) the $E_{3B}$ and (b) $E_{4B}$ calculated using the three polarization models and QM methods (MP2/aug-cc-pvtz for Water-4568 and RI-MP2/cc-pvtz for the Tetramers set). Each $(MeOH)_4$ and $(MeOH)_2(H_2O)_2$ cluster has two different structures.

models can be seen on the clusters containing water and/or methanol molecules. For the $E_{4B}$ there is still noticeable deviation, but they are very small in magnitude with the majority of them being less than 0.5 kcal/mol. While it is not surprising that the two new models did well as they are now explicitly optimized to the $E_{3B}$ and $E_{4B}$, the current AMOEBA model also shows a great overall trend across the organic set.

**3.2. Validations.** *3.2.1. Organic Clusters. Trimers-Distance Set.* With the damping parameters derived above, the distance dependence of the $E_{3B}$ was examined using the three polarization models on the Trimers-Distance set, which contains 124 configurations of 11 homotrimers. We found here that the $E_{3B}$ distance dependence strongly relies on the *direct* damping function. Both the current AMOEBA and Model 1 have similar distance dependence behavior of the $E_{3B}$ due to the same damping function in both (Figure 4). The smaller damping factor (0.34) in Model 1 leads to less negative $E_{3B}$ than the current AMOEBA with a damping factor of 0.39 for all intermolecular separations. Thus, Model 1 better captures the $E_{3B}$ of equilibrium clusters than the current AMOEBA does (consistent with Table 3 and Figure 3). Model 1 was parametrized on the MP2 $E_{3B}$ data of equilibrium-geometry clusters (Water-4568 and Tetramers sets). Thus, the $E_{3B}$ deviates from that of MP2 for the short intermolecular separations. Both AMOEBA and model display incorrect distance dependence behavior. Model 2, with a modified *direct* damping function (eq 7), is able to reproduce the MP2 (or RI-


DOI: 10.1021/acs.jctc.7b00225
J. Chem. Theory Comput. 2017, 13, 2751−2761

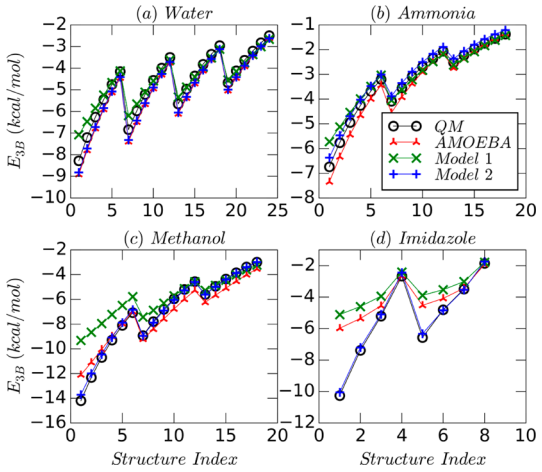



**Figure 4.** Plots of the $E_{3B}$ distance dependence calculated from three polarization models, MP2/aug-cc-pvtz (for water) and RI-MP2/cc-pvtz (for other molecules) for selected trimer systems: (a) Water, (b) Ammonia, (c) Methanol, and (d) Imidazole. The right-most structure indices represent the equilibrium structure for each trimer. The left side of the *x*-axis indicates the smaller intermolecular distances than the right.

MP2) $E_{3B}$ with noticeable improvement. This can be reflected by the smaller error and better correlation between Model 2 and MP2 results than other two models (Table 4).

**Table 4. Statistics Evaluation of the Three Polarization Models on the $E_{3B}$ and $E_{4B}$ of the Trimers-Distance Set[b]**

| statistics[a] | AMOEBA | Model 1 | Model 2 |
|---|---|---|---|
| MUE | 0.41 | 0.51 | 0.24 |
| MSE | −0.02 | 0.43 | 0.09 |
| RMSE | 0.67 | 1.01 | 0.37 |
| $R^2$ | 0.95 | 0.92 | 0.98 |

[a] $R^2$: Pearson correlation coefficient. All statistic errors are in kcal/mol. [b] All statistics was calculated using the $E_{3B}$ from MP2/aug-cc-pvtz (for water trimers) and RI-MP2/cc-pvtz (for other trimers) as the references.

It is interesting to note that the experimented models use *direct* damping functions with power to the distance being 2 or 1, and appropriate damping parameters will also provide good distance dependence. The corresponding RMSE values are 0.55 and 0.60 kcal/mol for the power being 2 and 1, respectively (see the SI).

*Water6-Extra Set.* Seven additional water hexamers were employed to evaluate the three polarization models. These isomers have different hydrogen bonding networks that represent the majority of the hydrogen bonding patterns in bulk water and thus are suitable for testing the directionality and transferability of the models. Comparing to the MP2/aug-cc-pvtz $E_{3B}$ data (Table 5), the RMSE values of the current AMOEBA, Model 1, and Model 2 are 1.36, 0.58, and 0.14, respectively, where an ∼10-fold of improvement can be obtained from AMOEBA to Model 2. While Model 1 shows better agreement than AMOEBA on the absolute magnitudes of $E_{3B}$ and $E_{4B}$, both Models 1 and 2 show similar correlation with MP2 results, due to the use of the same damping function but a different damping factor. The data suggests that Model 2 systematically improved the description of the MP2 $E_{3B}$ among different water hexamer configurations.

*3B-69 Set.* Fourteen clusters consisting of 5 molecules from the 3B-69 database were selected to test the new polarization models. Note that the RMSD of the $E_{3B}$ given by MP2 and CCSD methods[32,33] is only 0.06 kcal/mol for the 14 clusters (Table 6). Overall, all three models display excellent transferability on this set. It is interesting to point out that several sets of trimers show positive $E_{3B}$, and all three Thole-based classical models are able to capture the sign (Table 6, numbers in bold). Similar conclusions are found when the CCSD values are used as the reference (see parentheses). Our models fitted to MP2 $E_{3B}$ data capture not only the $E_{3B}$ for low/medium dispersion molecules but also high dispersion molecules (such as Benzene).[32,33] The performance of these three models on the 3B-69 set is consistent with that on the Tetramers set.

*Trimers-Transferability Set.* Transferability of the polarization models, mainly the universal damping parameter, was evaluated using equilibrium trimers of additional organic molecules. In addition to the C, H, O, and N elements that appear in our parametrization databases, S, P, F, Cl, and Br elements were also included in this test. Polarizability parameters of these molecules were taken from AMOEBA09 without changes, and thus the damping factor was tested for transferability to these new elements. Charge states of the clusters were not limited to neutral, and both positively and negatively charged trimers were included (see Figure 5). The RMSE values of three models with respect to the MP2 $E_{3B}$ are 0.31, 0.25, and 0.22 kcal/mol for the current AMOEBA, Model 1, and Model 2, respectively. The largest error for all three models is on the anionic $(AcO^-)(H_2O)_2$ cluster: 1.04, 0.85, and 0.73 kcal/mol for three models (see the SI). For Model 2, it was found that using a smaller damping parameter (stronger

**Table 5. $E_{3B}$ and $E_{4B}$ Obtained from MP2/aug-cc-pvtz and Three Polarization Models on Seven Water Hexamers[a]**

| | $E_{3B}$ | | | | $E_{4B}$ | | | |
|---|---|---|---|---|---|---|---|---|
| hexamers | MP2 | AMOEBA | Model 1 | Model 2 | MP2 | AMOEBA | Model 1 | Model 2 |
| cage | −9.22 | −10.63 | −9.92 | −9.39 | −0.65 | −1.02 | −0.94 | −0.86 |
| bag | −10.41 | −11.59 | −10.78 | −10.37 | −1.35 | −1.73 | −1.60 | −1.52 |
| cyclic-chair | −11.59 | −12.85 | −12.01 | −11.36 | −2.05 | −2.41 | −2.24 | −2.10 |
| book-1 | −10.35 | −11.78 | −10.99 | −10.40 | −1.33 | −1.72 | −1.60 | −1.49 |
| book-2 | −10.11 | −11.58 | −10.80 | −10.23 | −1.23 | −1.64 | −1.52 | −1.42 |
| cyclic-boat-1 | −11.20 | −12.55 | −11.74 | −11.06 | −1.88 | −2.27 | −2.11 | −1.98 |
| cyclic-boat-2 | −11.19 | −12.58 | −11.78 | −11.10 | −1.87 | −2.26 | −2.10 | −1.97 |
| RMSE(kcal/mol) | | 1.36 | 0.58 | 0.14 | | 0.38 | 0.25 | 0.15 |
| $R^2$ | | 0.984 | 0.981 | 0.995 | | 0.999 | 0.999 | 0.999 |

[a] The RMSE and Pearson correlation coefficient ($R^2$) values were calculated using MP2 data as the reference.



DOI: 10.1021/acs.jctc.7b00225
J. Chem. Theory Comput. 2017, 13, 2751−2761

Table 6. $E_{3B}$ of Five Trimers in the 3B-69 Set Calculated from the Three Polarization Models and Two *Ab Initio* Methods (MP2/CBS and CCSD/CBS)[b]

| molecules | isomers | MP2 | CCSD | AMOEBA | Model 1 | Model 2 |
|---|---|---|---|---|---|---|
| Water | a | −1.39 | −1.39 | −1.56 | −1.40 | −1.36 |
| | **b** | **1.07** | **1.08** | **1.23** | **1.07** | **1.05** |
| | c | −2.47 | −2.42 | −2.90 | −2.66 | −2.48 |
| AcOH | **a** | **0.52** | **0.54** | **0.42** | **0.37** | **0.42** |
| | b | −0.94 | −0.92 | −1.01 | −0.81 | −1.06 |
| | c | −0.25 | −0.21 | −0.19 | −0.13 | −0.20 |
| $AcNH_2$ | a | −0.24 | −0.09 | −0.16 | −0.07 | −0.10 |
| | **b** | **0.53** | **0.58** | **0.79** | **0.68** | **0.69** |
| | c | −0.85 | −0.86 | −0.71 | −0.62 | −0.64 |
| Imidazole | a | −0.77 | −0.66 | −0.76 | −0.69 | −0.70 |
| | **b** | **0.21** | **0.27** | **0.20** | **0.20** | **0.20** |
| | c | −1.63 | −1.64 | −1.70 | −1.53 | −1.54 |
| Benzene | a | −0.05 | 0.05 | *a* | *a* | *a* |
| | **b** | **0.08** | **0.15** | **0.09** | **0.09** | **0.08** |
| | c | −0.06 | −0.03 | −0.09 | −0.09 | −0.07 |
| RMSE(D) (kcal/mol) | | | 0.06 | 0.16(0.17) | 0.12(0.12) | 0.10(0.09) |

[a]For isomer *a* of the Benzene trimer, the coordinates are incomplete in the original paper.[33] [b]RMSE values were calculated using MP2 or CCSD (in parentheses) data as the reference. Isomers with positive $E_{3B}$ values were highlighted in bold.

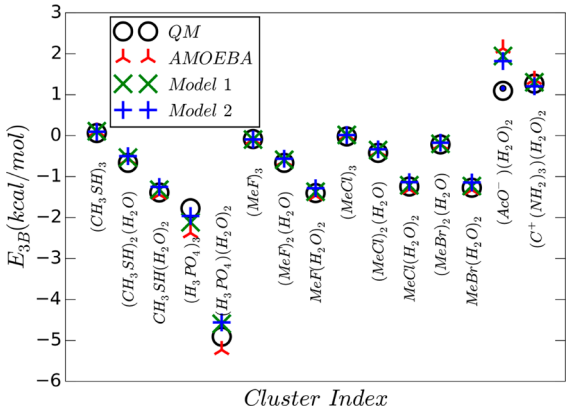



**Figure 5.** $E_{3B}$ calculated from three polarization models and RI-MP2/cc-pvtz for the Trimers-Transferability set. The solid blue dot on $(AcO)^-(H_2O)_2$ shows the $E_{3B}$ calculated from Model 2 with a damping factor of 0.45 for the *direct* induction between the oxygen in the carbonyl group and other sites.

damping) between the anionic carbonyl oxygen and other sites can greatly improve the agreement. For example, using a *direct* damping of 0.45 (instead of 0.75) between the anionic oxygen atoms of $AcO^-$ and other atoms reduced the error to 0.06 kcal/mol. The same trend was found for other monovalent anions $Cl^-$ and $Br^-$ as discussed next (also see Table 2). In summary, the models developed from a small set of molecular clusters show satisfactory accuracy and transferability on a wide range of chemical species.

*3.2.2. Metal/Halogen Ions.* A previous study[61] based on a Gaussian electron density force field indeed demonstrated that for metal ions, where strong fields are involved, the many-body interactions are from both electron polarization and Pauli exclusion effects. It is, therefore, interesting to examine the capability of our classical polarization models, with limited modifications, to capture the many-body interactions in such systems.

Here we focus on the clusters formed by metal/halogen ions and water molecules. The current AMOEBA model overall

Table 7. $E_{3B}$ and $E_{4B}$ Obtained from the RI-MP2 and Polarization Models[b]

| | $E_{3B}$ (kcal/mol) | | | | $E_{4B}$ (kcal/mol) | | | |
|---|---|---|---|---|---|---|---|---|
| clusters | QM | AMOEBA | Model 1 | Model 2 | QM | AMOEBA | Model 1 | Model 2 |
| $Na^+(H_2O)_3$ | 4.26 | 4.85 | 4.63 | 4.61 | −0.08 | −0.14 | −0.13 | −0.13 |
| $K^+(H_2O)_3$ | 3.04 | 2.98 | 3.09 | 2.85 | −0.06 | −0.07 | −0.08 | −0.07 |
| $Mg^{2+}(H_2O)_3$ | 26.17 | 23.36 | 22.41 | 21.58 | −1.53 | −0.92 | −0.87 | −0.84 |
| $Ca^{2+}(H_2O)_3$ | 13.10 | 10.89 | 12.90 | 12.27 | −0.59 | −0.32 | −0.43 | −0.42 |
| $Zn^{2+}(H_2O)_3$_a | 46.66 | 31.73 | 34.48 | 32.05 | −3.48 | −1.55 | −1.77 | −1.65 |
| $Zn^{2+}(H_2O)_3$_b | 43.87 | 24.70 | 25.63 | 24.24 | −3.99 | −1.11 | −1.17 | −1.11 |
| $Cl^-(H_2O)_3$ | 1.05 | 2.75 | 2.73 | 2.31 | −0.03 | −0.09 | −0.11 | −0.08 |
| $Br^-(H_2O)_3$ | 0.44 | 1.47 | 1.53 | 1.49 | 0.00 | −0.01 | −0.03 | −0.02 |
| $(NaCl)_2$ | 26.69 | 34.25 | 32.19 | 24.58 | −0.49 | 1.36 | 1.89 | 1.69 |
| $NaCl(H_2O)_2$_a | 5.57 | 6.29 | 6.55 | 5.81 | −0.68 | −0.41 | −0.41 | −0.39 |
| $NaCl(H_2O)_2$_b | 12.10 | 14.05 | 14.13 | 12.66 | −0.67 | −0.33 | −0.38 | −0.38 |
| RMSE(kcal/mol)[a] | | 2.95 | 2.44 | 1.81 | | 0.67 | 0.84 | 0.78 |

[a]Zn-water clusters are excluded in RMSE calculations; $Zn^{2+}(H_2O)_3$ and $NaCl(H_2O)_2$ each has two structures each; Zinc ion is in the plane composed of three oxygen atoms in $Zn^{2+}(H_2O)_3$_a while out of that plane in $Zn^{2+}(H_2O)_3$_b. See the graphics of these structures in the SI. [b]RMSE values were given using QM (RI-MP2) data as the reference.



DOI: 10.1021/acs.jctc.7b00225
J. Chem. Theory Comput. 2017, 13, 2751−2761

matches MP2 results on the $E_{3B}$ and $E_{4B}$ of these tetramers with better agreements on monovalent ions than on divalent cations (Table 7). The two new models, where the ionic polarizability and damping factors were modified (see Table 2 for parameters), improved the overall performance on divalent ion systems. For example, Model 1 and Model 2 reduce the RMSE of the $E_{3B}$ from 2.95 (the current AMOEBA) to 2.44 and 1.81 kcal/mol, respectively. The only exception lies in the $Zn^{2+}$-water clusters, where our models give less positive $E_{3B}$ and less negative $E_{4B}$. The reason probably can be ascribed to the lack of charge-transfer term in our models. Our previous study shows that the charge-transfer of $Zn(H_2O)_n$ complexes decreases (less negative) as the number of water molecules increases.[14] Based on the $E_{3B}$ and $E_{4B}$ decomposition approaches described in the SI (eqs S4−S6), there would be a positive correction for the $E_{3B}$ and a negative correction for the $E_{4B}$ if the charge-transfer contribution were included. However, it was found that in the bulk simulations of the Zn-water system, the many-body interactions are dominated by polarization. Without the inclusion of an explicit CT term, one should avoid overfitting to the QM charge-transfer energy in small clusters.[14]

All three models well capture the distance dependence of the $E_{3B}$ for monovalent ions, such as $Na^+$, $K^+$, $Cl^-$, and $Br^-$ (Figure 6). It is noted that in the *linear* structure of $M(H_2O)_2$ (M = $Mg^{2+}$, $Ca^{2+}$, and $Zn^{2+}$) clusters (where O···M···O is linear), the $E_{3B}$ can be well captured by classical models for $Mg^{2+}$ and $Ca^{2+}$ but *not* for $Zn^{2+}$. In our model, nearly *zero* induced dipole was found on the metal site in $M(H_2O)_2$ (M = $Mg^{2+}$, $Ca^{2+}$, and $Zn^{2+}$) trimer. The difference between $Zn^{2+}$ and $Mg^{2+}(Ca^{2+})$ is likely due to the lacking of hyperpolarizability (such as induced quadrupole) and charge transfer of transition metals.[62]

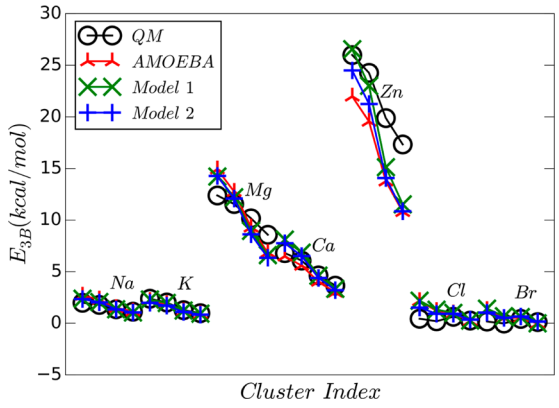



**Figure 6.** Plot of the $E_{3B}$ distance dependence calculated by RI-MP2 and the three polarization models for $M(H_2O)_2$ systems, where M = $Na^+$, $K^+$, $Mg^{2+}$, $Ca^{2+}$, $Zn^{2+}$, $Cl^-$, and $Br^-$ as labeled in the figure. For $M(H_2O)_2$ (M = $Mg^{2+}$, $Ca^{2+}$, and $Zn^{2+}$), the angle ∠OMO is 109°28′, which is a tetrahedral angle.

**3.3. Polarization Energy from EDA.** Through the above discussion of molecular clusters beyond dimer, we have seen that the many-body interactions can be well captured by the classical polarization models, with Model 2 showing a better distance dependence behavior of the $E_{3B}$. Via many EDA methods, the intermolecular interaction energy is usually decomposed into physically meaningful components, such as electrostatics, exchange-repulsion, dispersion, and induction. For the induction component, it remains debated to be further separated into polarization and charge transfer term,[63] although there have been such attempts through EDA methods such as ALMO (absolutely localized molecular orbital),[52,53] BLW (block-localized wave function),[64,65] CSOV (constrained space orbital variation),[66,67] RVS (reduced variational space),[68] and regularized SAPT.[48] It is interesting to compare, in retrospect, the results of classical models with these EDA methods.

Here we compared the polarization energy calculated using our models with the "true induction" from regularized SAPT(DFT) and polarization energy from ALMO method. It is noted that the results from regularized SAPT(DFT) and ALMO agree with each other in the near-equilibrium intermolecular distances but deviate in short separations, where ALMO gives more negative polarization energy than the regularized SAPT(DFT). It should be pointed out that the $\delta^{HF}$ correction, a term to capture some higher-order terms not explicitly evaluated by SAPT,[69] is excluded in regularized SAPT(DFT) calculations. In our models, the current AMOEBA and Model 1 give incorrect distance dependence behavior especially at the short O···H intermolecular distances (Figure 7). Model 2 shows the best trend comparing to the results from

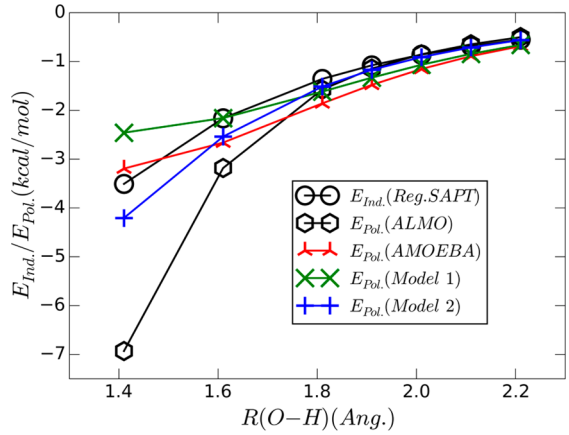



**Figure 7.** Plot of the polarization (or induction) energy obtained from the three polarization models and two QM EDA methods (regularized SAPT(DFT) and ALMO) for water dimer at 7 separations. The $\delta^{HF}$ correction is excluded in the regularized SAPT(DFT) calculation.

the regularized SAPT(DFT). The polarization energy from Model 2 includes part of the $\delta^{HF}$ correction energy, as the $\delta^{HF}$ term from SAPT2+/aug-cc-pvtz (see the SI for details) can be as large as −5.70 kcal/mol for the water dimer whose O···H distance is 1.41 Å and 0.71 kcal/mol for equilibrium distance (2.01 Å). Additional comparisons of the three models and regularized SAPT(DFT) were made for small organic molecules and ion−water systems (see the SI). Again, among the three models, Model 2 shows the best trend over the regularized SAPT(DFT) "true induction" energy on these dimers. In summary, Model 2 qualitatively agrees with the regularized SAPT(DFT) on the polarization term, while the current AMOEBA and Model 1 show much deviation to the regularized SAPT(DFT) as well as ALMO EDA.

## 4. CONCLUSION

In this work, we show that the many-body interactions ($E_{3B}$ and $E_{4B}$) of organic molecular clusters can be well captured by classical polarization models. Overall, the current AMOEBA model and two new models are able to reasonably describe the three- and four-body energy for a wide range of organic


DOI: 10.1021/acs.jctc.7b00225
J. Chem. Theory Comput. 2017, 13, 2751−2761

molecular clusters in different configurations. In these simple models, universal parameters controlling the damping strength perform well for all organic species tested in this study. Comparing to the current AMOEBA, as expected, the two new models that have been explicitly fitted to the $E_{3B}$ show better agreement with the MP2 results. Model 2, where the damping function for *direct* induction due to permanent field was modified, best reproduced the distance dependence of the $E_{3B}$. These models are also able to capture all the positive many-body interactions according to the MP2 calculations. In a physical sense, these results clearly show that instead of using the smeared charge distribution given by the damping function of the current AMOEBA, a different charge distribution is needed to well capture the distance dependence of many-body interactions.

The point dipole induction models were also applied to the molecular clusters involving monovalent ions and divalent cations ($Mg^{2+}$, $Ca^{2+}$, and $Zn^{2+}$). We found that, as with previous studies, the divalent cations required different damping strengths from those of organic clusters. Furthermore, results on $Zn^{2+}$-water clusters indicate the importance of the charge-transfer and hyperpolarizability (induced quadrupole) which are currently missing in our treatment. For highly ionic species and molecular systems such as van der Waals clusters, additional many-body contributions should be considered explicitly.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.7b00225.

Many-body decomposition databases, details of the polarization models, numerical results obtained in this work, and structure graphics of all the molecular clusters (PDF)

Cartesian coordinate files (ZIP)

## ■ AUTHOR INFORMATION


### Corresponding Author

*E-mail: pren@mail.utexas.edu.

### ORCID

Chengwen Liu: 0000-0002-3930-7793

J.-P. Piquemal: 0000-0001-6615-9426



### Funding

The authors are grateful for support from the Robert A. Welch Foundation (F-1691) and the National Institutes of Health (R01GM106137 and R01GM114237). Funding from French CNRS through a PICS grant between UPMC and UT Austin is acknowledged. Q.W. is grateful for support by the National Natural Science Foundation of China (81602954).


### Notes

The authors declare no competing financial interest.

## ■ REFERENCES


(1) Møller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **1934**, *46*, 618.

(2) Ren, P.; Chun, J.; Thomas, D. G.; Schnieders, M. J.; Marucho, M.; Zhang, J.; Baker, N. A. Biomolecular electrostatics and solvation: a computational perspective. *Q. Rev. Biophys.* **2012**, *45*, 427.

(3) Ren, P.; Ponder, J. W. Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation. *J. Phys. Chem. B* **2003**, *107*, 5933.

(4) Lamoureux, G.; MacKerell, A. D.; Roux, B. A simple polarizable model of water based on classical Drude oscillators. *J. Chem. Phys.* **2003**, *119*, 5185.

(5) Stern, H. A.; Rittner, F.; Berne, B. J.; Friesner, R. A. Combined fluctuating charge and polarizable dipole models: Application to a five-site water potential function. *J. Chem. Phys.* **2001**, *115*, 2237.

(6) Ren, P.; Wu, C.; Ponder, J. W. Polarizable Atomic Multipole-Based Molecular Mechanics for Organic Molecules. *J. Chem. Theory Comput.* **2011**, *7*, 3143.

(7) Hagberg, D.; Karlstrom, G.; Roos, B. O.; Gagliardi, L. The coordination of uranyl in water: A combined quantum chemical and molecular simulation study. *J. Am. Chem. Soc.* **2005**, *127*, 14250.

(8) Harder, E.; Anisimov, V. M.; Whitfield, T. W.; MacKerell, A. D.; Roux, B. Understanding the dielectric properties of liquid amides from a polarizable force field. *J. Phys. Chem. B* **2008**, *112*, 3509.

(9) Wang, J. M.; Cieplak, P.; Li, J.; Hou, T. J.; Luo, R.; Duan, Y. Development of Polarizable Models for Molecular Mechanical Calculations I: Parameterization of Atomic Polarizability. *J. Phys. Chem. B* **2011**, *115*, 3091.

(10) Wang, J. M.; Cieplak, P.; Li, J.; Wang, J.; Cai, Q.; Hsieh, M. J.; Lei, H. X.; Luo, R.; Duan, Y. Development of Polarizable Models for Molecular Mechanical Calculations II: Induced Dipole Models Significantly Improve Accuracy of Intermolecular Interaction Energies. *J. Phys. Chem. B* **2011**, *115*, 3100.

(11) Jiao, D.; Golubkov, P. A.; Darden, T. A.; Ren, P. Calculation of protein-ligand binding free energy by using a polarizable potential. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 6290.

(12) Jiao, D.; King, C.; Grossfield, A.; Darden, T. A.; Ren, P. Y. Simulation of $Ca^{2+}$ and $Mg^{2+}$ solvation using polarizable atomic multipole potential. *J. Phys. Chem. B* **2006**, *110*, 18553.

(13) Wu, J. C.; Piquemal, J. P.; Chaudret, R.; Reinhardt, P.; Ren, P. Y. Polarizable Molecular Dynamics Simulation of Zn(II) in Water Using the AMOEBA Force Field. *J. Chem. Theory Comput.* **2010**, *6*, 2059.

(14) Wu, J. C.; Piquemal, J.-P.; Chaudret, R.; Reinhardt, P.; Ren, P. Polarizable Molecular Dynamics Simulation of Zn(II) in Water Using the AMOEBA Force Field. *J. Chem. Theory Comput.* **2010**, *6*, 2059.

(15) Roux, C.; Bhatt, F.; Foret, J.; de Courcy, B.; Gresh, N.; Piquemal, J. P.; Jeffery, C. J.; Salmon, L. The reaction mechanism of type I phosphomannose isomerases: New information from inhibition and polarizable molecular mechanics studies. *Proteins: Struct., Funct., Genet.* **2011**, *79*, 203.

(16) Chen, J.; Hundertmark, D.; Martínez, T. J. A unified theoretical framework for fluctuating-charge models in atom-space and in bond-space. *J. Chem. Phys.* **2008**, *129*, 214113.

(17) Patel, S.; Brooks, C. L. CHARMM fluctuating charge force field for proteins: I parameterization and application to bulk organic liquid simulations. *J. Comput. Chem.* **2004**, *25*, 1.

(18) Patel, S.; Mackerell, A. D.; Brooks, C. L. CHARMM fluctuating charge force field for proteins: II Protein/solvent properties from molecular dynamics simulations using a nonadditive electrostatic model. *J. Comput. Chem.* **2004**, *25*, 1504.

(19) Lopes, P. E. M.; Roux, B.; MacKerell, A. D. Molecular modeling and dynamics studies with explicit inclusion of electronic polarizability: theory and applications. *Theor. Chem. Acc.* **2009**, *124*, 11.

(20) Lopes, P. E. M.; Huang, J.; Shim, J.; Luo, Y.; Li, H.; Roux, B.; MacKerell, A. D. Polarizable Force Field for Peptides and Proteins Based on the Classical Drude Oscillator. *J. Chem. Theory Comput.* **2013**, *9*, 5430.

(21) Lemkul, J. A.; Huang, J.; Roux, B.; MacKerell, A. D. An Empirical Polarizable Force Field Based on the Classical Drude Oscillator Model: Development History and Recent Applications. *Chem. Rev.* **2016**, *116*, 4983.

(22) Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J.-P. Anisotropic, Polarizable Molecular Mechanics Studies of Inter- and Intramolecular Interactions and Ligand−Macromolecule Complexes. A Bottom-Up Strategy. *J. Chem. Theory Comput.* **2007**, *3*, 1960.

(23) Cieplak, P.; Dupradeau, F.-Y.; Duan, Y.; Wang, J. Polarization effects in molecular mechanical force fields. *J. Phys.: Condens. Matter* **2009**, *21*, 333102.




DOI: 10.1021/acs.jctc.7b00225
*J. Chem. Theory Comput.* 2017, 13, 2751−2761

(24) Xie, W.; Gao, J. Design of a Next Generation Force Field: The X-POL Potential. *J. Chem. Theory Comput.* **2007**, *3*, 1890.
(25) Shi, Y.; Ren, P.; Schnieders, M.; Piquemal, J.-P. Polarizable Force Fields for Biomolecular Modeling. In *Reviews in Computational Chemistry*; John Wiley & Sons, Inc.: 2015; Vol. *28*, p 51, DOI: 10.1002/9781118889886.ch2.
(26) Huang, J.; Lopes, P. E. M.; Roux, B.; MacKerell, A. D. Recent Advances in Polarizable Force Fields for Macromolecules: Microsecond Simulations of Proteins Using the Classical Drude Oscillator Model. *J. Phys. Chem. Lett.* **2014**, *5*, 3144.
(27) Cisneros, G. A.; Karttunen, M.; Ren, P.; Sagui, C. Classical Electrostatics for Biomolecular Simulations. *Chem. Rev.* **2014**, *114*, 779.
(28) Kumar, R.; Wang, F.-F.; Jenness, G. R.; Jordan, K. D. A second generation distributed point polarizable water model. *J. Chem. Phys.* **2010**, *132*, 014309.
(29) Christie, R. A.; Jordan, K. D. n-Body Decomposition Approach to the Calculation of Interaction Energies of Water Clusters. In *Intermolecular Forces and Clusters II*; Wales, D. J., Ed.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2005; p 27, DOI: 10.1007/430_003.
(30) Stone, A. *The Theory of Intermolecular Forces*; Oxford University Press: United Kingdom, 2016; p 190.
(31) Chaudret, R.; Gresh, N.; Parisel, O.; Piquemal, J.-P. Many-body exchange-repulsion in polarizable molecular mechanics. I. orbital-based approximations and applications to hydrated metal cation complexes. *J. Comput. Chem.* **2011**, *32*, 2949.
(32) Huang, Y.; Beran, G. J. O. Reliable prediction of three-body intermolecular interactions using dispersion-corrected second-order Møller-Plesset perturbation theory. *J. Chem. Phys.* **2015**, *143*, 044113.
(33) Řezáč, J.; Huang, Y.; Hobza, P.; Beran, G. J. O. Benchmark Calculations of Three-Body Intermolecular Interactions and the Performance of Low-Cost Electronic Structure Methods. *J. Chem. Theory Comput.* **2015**, *11*, 3065.
(34) Beran, G. J. O.; Wen, S.; Nanda, K.; Huang, Y.; Heit, Y. Accurate and Robust Molecular Crystal Modeling Using Fragment-Based Electronic Structure Methods. In *Prediction and Calculation of Crystal Structures: Methods and Applications*; Atahan-Evrenk, S., Aspuru-Guzik, A., Eds.; Springer International Publishing: Cham, 2014; p 59, DOI: 10.1007/128_2013_502.
(35) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
(36) MacKerell, A. D.; Wiorkiewicz-Kuczera, J.; Karplus, M. An all-atom empirical energy function for the simulation of nucleic acids. *J. Am. Chem. Soc.* **1995**, *117*, 11946.
(37) Thole, B. T. Molecular polarizabilities calculated with a modified dipole interaction. *Chem. Phys.* **1981**, *59*, 341.
(38) van Duijnen, P. T.; Swart, M. Molecular and Atomic Polarizabilities: Thole's Model Revisited. *J. Phys. Chem. A* **1998**, *102*, 2399.
(39) Wang, L.-P.; Head-Gordon, T.; Ponder, J. W.; Ren, P.; Chodera, J. D.; Eastman, P. K.; Martinez, T. J.; Pande, V. S. Systematic Improvement of a Classical Molecular Model of Water. *J. Phys. Chem. B* **2013**, *117*, 9956.
(40) Slipchenko, L. V.; Gordon, M. S. Water−Benzene Interactions: An Effective Fragment Potential and Correlated Quantum Chemistry Study. *J. Phys. Chem. A* **2009**, *113*, 2092.
(41) Boyd, S. L.; Boyd, R. J. A Density Functional Study of Methanol Clusters. *J. Chem. Theory Comput.* **2007**, *3*, 54.
(42) Malloum, A.; Fifen, J. J.; Dhaouadi, Z.; Engo, S. G. N.; Jaidane, N.-E. Structures and relative stabilities of ammonia clusters at different temperatures: DFT vs. ab initio. *Phys. Chem. Chem. Phys.* **2015**, *17*, 29226.
(43) Mandal, A.; Prakash, M.; Kumar, R. M.; Parthasarathi, R.; Subramanian, V. Ab Initio and DFT Studies on Methanol−Water Clusters. *J. Phys. Chem. A* **2010**, *114*, 2250.
(44) Matisz, G.; Kelterer, A.-M.; Fabian, W. M. F.; Kunsági-Máté, S. Coordination of Methanol Clusters to Benzene: A Computational Study. *J. Phys. Chem. A* **2011**, *115*, 10556.
(45) Takeuchi, H. Structural Features of Small Benzene Clusters $(C_6H_6)_n$ ($n \leq 30$) As Investigated with the All-Atom OPLS Potential. *J. Phys. Chem. A* **2012**, *116*, 10172.
(46) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09, Revision A. 02*; Gaussian, Inc.: Wallingford, CT, USA, 2009.
(47) Misquitta, A. J.; Szalewicz, K. Symmetry-adapted perturbation-theory calculations of intermolecular forces employing density-functional description of monomers. *J. Chem. Phys.* **2005**, *122*, 214109.
(48) Misquitta, A. J. Charge Transfer from Regularized Symmetry-Adapted Perturbation Theory. *J. Chem. Theory Comput.* **2013**, *9*, 5313.
(49) Valiev, M.; Bylaska, E. J.; Govind, N.; Kowalski, K.; Straatsma, T. P.; Van Dam, H. J. J.; Wang, D.; Nieplocha, J.; Apra, E.; Windus, T. L.; de Jong, W. A. NWChem: A comprehensive and scalable open-source solution for large scale molecular simulations. *Comput. Phys. Commun.* **2010**, *181*, 1477.
(50) Adamo, C.; Barone, V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys.* **1999**, *110*, 6158.
(51) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.* **1992**, *96*, 6796.
(52) Khaliullin, R. Z.; Head-Gordon, M.; Bell, A. T. An efficient self-consistent field method for large systems of weakly interacting components. *J. Chem. Phys.* **2006**, *124*, 204105.
(53) Azar, R. J.; Head-Gordon, M. An energy decomposition analysis for intermolecular interactions from an absolutely localized molecular orbital reference at the coupled-cluster singles and doubles level. *J. Chem. Phys.* **2012**, *136*, 024103.
(54) Shao, Y.; Gan, Z.; Epifanovsky, E.; Gilbert, A. T. B.; Wormit, M.; Kussmann, J.; Lange, A. W.; Behn, A.; Deng, J.; Feng, X.; Ghosh, D.; Goldey, M.; Horn, P. R.; Jacobson, L. D.; Kaliman, I.; Khaliullin, R. Z.; Kuś, T.; Landau, A.; Liu, J.; Proynov, E. I.; Rhee, Y. M.; Richard, R. M.; Rohrdanz, M. A.; Steele, R. P.; Sundstrom, E. J.; Woodcock, H. L.; Zimmerman, P. M.; Zuev, D.; Albrecht, B.; Alguire, E.; Austin, B.; Beran, G. J. O.; Bernard, Y. A.; Berquist, E.; Brandhorst, K.; Bravaya, K. B.; Brown, S. T.; Casanova, D.; Chang, C.-M.; Chen, Y.; Chien, S. H.; Closser, K. D.; Crittenden, D. L.; Diedenhofen, M.; DiStasio, R. A.; Do, H.; Dutoi, A. D.; Edgar, R. G.; Fatehi, S.; Fusti-Molnar, L.; Ghysels, A.; Golubeva-Zadorozhnaya, A.; Gomes, J.; Hanson-Heine, M. W. D.; Harbach, P. H. P.; Hauser, A. W.; Hohenstein, E. G.; Holden, Z. C.; Jagau, T.-C.; Ji, H.; Kaduk, B.; Khistyaev, K.; Kim, J.; Kim, J.; King, R. A.; Klunzinger, P.; Kosenkov, D.; Kowalczyk, T.; Krauter, C. M.; Lao, K. U.; Laurent, A. D.; Lawler, K. V.; Levchenko, S. V.; Lin, C. Y.; Liu, F.; Livshits, E.; Lochan, R. C.; Luenser, A.; Manohar, P.; Manzer, S. F.; Mao, S.-P.; Mardirossian, N.; Marenich, A. V.; Maurer, S. A.; Mayhall, N. J.; Neuscamman, E.; Oana, C. M.; Olivares-Amaya, R.; O'Neill, D. P.; Parkhill, J. A.; Perrine, T. M.; Peverati, R.; Prociuk, A.; Rehn, D. R.; Rosta, E.; Russ, N. J.; Sharada, S. M.; Sharma, S.; Small, D. W.; Sodt, A.; Stein, T.; Stück, D.; Su, Y.-C.; Thom, A. J. W.; Tsuchimochi, T.; Vanovschi, V.; Vogt, L.; Vydrov, O.; Wang, T.; Watson, M. A.; Wenzel, J.; White, A.; Williams, C. F.; Yang,


DOI: 10.1021/acs.jctc.7b00225
J. Chem. Theory Comput. 2017, 13, 2751−2761

J.; Yeganeh, S.; Yost, S. R.; You, Z.-Q.; Zhang, I. Y.; Zhang, X.; Zhao, Y.; Brooks, B. R.; Chan, G. K. L.; Chipman, D. M.; Cramer, C. J.; Goddard, W. A.; Gordon, M. S.; Hehre, W. J.; Klamt, A.; Schaefer, H. F.; Schmidt, M. W.; Sherrill, C. D.; Truhlar, D. G.; Warshel, A.; Xu, X.; Aspuru-Guzik, A.; Baer, R.; Bell, A. T.; Besley, N. A.; Chai, J.-D.; Dreuw, A.; Dunietz, B. D.; Furlani, T. R.; Gwaltney, S. R.; Hsu, C.-P.; Jung, Y.; Kong, J.; Lambrecht, D. S.; Liang, W.; Ochsenfeld, C.; Rassolov, V. A.; Slipchenko, L. V.; Subotnik, J. E.; Van Voorhis, T.; Herbert, J. M.; Krylov, A. I.; Gill, P. M. W.; Head-Gordon, M. Advances in molecular quantum chemistry contained in the Q-Chem 4 program package. *Mol. Phys.* **2015**, *113*, 184.

(55) DiStasio, R. A.; Jung, Y.; Head-Gordon, M. A Resolution-Of-The-Identity Implementation of the Local Triatomics-In-Molecules Model for Second-Order Møller−Plesset Perturbation Theory with Application to Alanine Tetrapeptide Conformational Energies. *J. Chem. Theory Comput.* **2005**, *1*, 862.

(56) Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007.

(57) Mayhall, N. J.; Raghavachari, K.; Redfern, P. C.; Curtiss, L. A.; Rassolov, V. Toward accurate thermochemical models for transition metals: G3Large basis sets for atoms Sc−Zn. *J. Chem. Phys.* **2008**, *128*, 144122.

(58) Wang, Q.; Rackers, J. A.; He, C.; Qi, R.; Narth, C.; Lagardere, L.; Gresh, N.; Ponder, J. W.; Piquemal, J.-P.; Ren, P. General Model for Treating Short-Range Electrostatic Penetration in a Molecular Mechanics Force Field. *J. Chem. Theory Comput.* **2015**, *11*, 2609.

(59) Chai, J.-D.; Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615.

(60) Wu, J. C.; Chattree, G.; Ren, P. Automation of AMOEBA polarizable force field parameterization for small molecules. *Theor. Chem. Acc.* **2012**, *131*, 131−1138.

(61) Chaudret, R.; Gresh, N.; Narth, C.; Lagardère, L.; Darden, T. A.; Cisneros, G. A.; Piquemal, J.-P. S/G-1: An ab Initio Force-Field Blending Frozen Hermite Gaussian Densities and Distributed Multipoles. Proof of Concept and First Applications to Metal Cations. *J. Phys. Chem. A* **2014**, *118*, 7598.

(62) Masia, M.; Probst, M.; Rey, R. On the performance of molecular polarization methods. I. Water and carbon tetrachloride close to a point charge. *J. Chem. Phys.* **2004**, *121*, 7362.

(63) Stone, A. J. Natural Bond Orbitals and the Nature of the Hydrogen Bond. *J. Phys. Chem. A* **2017**, *121*, 1531.

(64) Mo, Y.; Gao, J.; Peyerimhoff, S. D. Energy decomposition analysis of intermolecular interactions using a block-localized wave function approach. *J. Chem. Phys.* **2000**, *112*, 5530.

(65) Mo, Y.; Song, L.; Lin, Y. Block-Localized Wavefunction (BLW) Method at the Density Functional Theory (DFT) Level. *J. Phys. Chem. A* **2007**, *111*, 8291.

(66) Bagus, P. S.; Illas, F. Decomposition of the chemisorption bond by constrained variations: Order of the variations and construction of the variational spaces. *J. Chem. Phys.* **1992**, *96*, 8962.

(67) Bagus, P. S.; Hermann, K.; Bauschlicher, C. W., Jr. A new analysis of charge transfer and polarization for ligand−metal bonding: Model studies of $Al_4CO$ and $Al_4NH_3$. *J. Chem. Phys.* **1984**, *80*, 4378.

(68) Stevens, W. J.; Fink, W. H. Frozen fragment reduced variational space analysis of hydrogen bonding interactions. Application to the water dimer. *Chem. Phys. Lett.* **1987**, *139*, 15.

(69) Moszynski, R. Symmetry-adapted perturbation theory for the calculation of Hartree-Fock interaction energies. *Mol. Phys.* **1996**, *88*, 741.



DOI: 10.1021/acs.jctc.7b00225
*J. Chem. Theory Comput.* 2017, 13, 2751−2761