WIREs Wiley Interdisciplinary Reviews

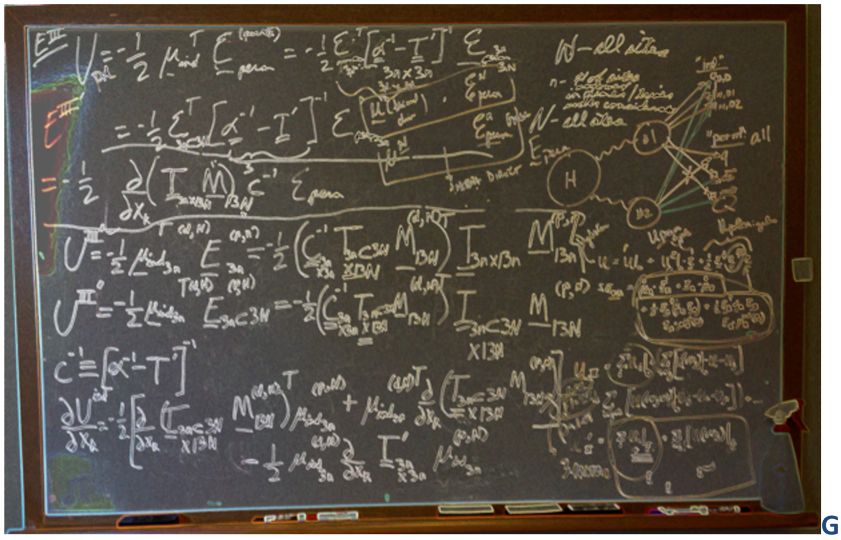
Advanced Models for Water Simulations

Omar Demerdash[1,2], Lee-Ping Wang[5], and Teresa Head-Gordon[1-4*]
[1]Kenneth S. Pitzer Center for Theoretical Chemistry, University of California, Berkeley
Departments of [2]Chemistry, [3]Bioengineering, [4]Chemical and Biomolecular Engineering, University of California, Berkeley USA
[5]Department of Chemistry, University of California, Davis, One Shields Ave. Davis, CA 95616 USA

ABSTRACT

Molecular simulations of water using classical, molecular mechanic potential energy functions have enjoyed a 50-year history of development, and much has been learned regarding their parameterization and the essential physics that must be captured in order to reproduce water properties across the phase diagram and across system sizes, from the dimer to the condensed phase. While pairwise-additive force fields using fixed, point charge-based electrostatics have dominated this history owing to computational cost, their limitations in transferability are being recognized, owing particularly to the lack of many-body effects, as well as an inherent difficulty in capturing quantum mechanical effects that become important at short intermolecular separation. This has spurred an impressive development of novel functional forms and parameterization schemes to account for such effects, especially the leading many-body effect of polarization. This review will discuss recent efforts in the development of advanced models of water, particularly with regard to important details of their parameterization from quantum mechanical or experimental data, the development of novel functional forms including machine learning-based models, and algorithms that reduce the computational cost of polarization dramatically, permitting them to potentially become competitive with pairwise-additive models as the standby of condensed-phase simulation. These technical developments are appraised based on their ability to impact numerical calculations on water, particularly the condensed phase, and it is hoped that this article provides a clear connection between the essential physics captured by the model and their fitness across a range of environments.

*corresponding author



Graphical/Visual Abstract: Advanced classical potential energy surfaces for water are formulating new functional forms for non-bonded interactions, such as hydrogen-bonding, for greater accuracy in future computational studies of water across its phase diagram.

INTRODUCTION

In the development of force fields for molecular simulation, gas phase water clusters, liquid bulk water, and the ice phases tend to be the first testbed for whether these new approximations to molecular interactions are a more accurate description of the underlying potential energy surface. The level of accuracy that is required will of course depend on the application. For example, the characterization of the mechanism for auto-ionization in water[1,2] or the proton transfer reaction[3,4] by definition will depend on *ab initio* molecular dynamics and models for nuclear quantum effects[5-8]. When electron rearrangements and quantum fluctuations are not active or central, then classical models can be robust for almost everything else – conformational energies, structural properties, as well bulk transport properties – since classical Hamiltonians implicitly incorporate these quantum mechanical effects through effective parameterization of the piecewise nature of the molecular mechanics functional form as given in Eq. (1).

$$U = U_{\mathrm{VAL}} + U_{\mathrm{PAULI}} + U_{\mathrm{DISP}} + U_{\mathrm{ELEC}} + \boldsymbol{U_{CP}} + \boldsymbol{U_{POL}} + \boldsymbol{U_{CT}} \tag{1}$$

For standard and widely available empirical water force fields, the VAL(ence) or water geometric nuclear framework is either held rigid or is composed of stiff harmonic terms that permit only small fluctuations around the equilibrium bond-lengths and bond angle, appropriate to the classical assumption where bond making and bond breaking are prohibited. For many years classical water models have primarily relied on the pairwise-additive approximation for the remaining non-bonded interactions, if they are represented at all. This is manifested by PAULI and DISP terms that represent the inherently many-body exchange-repulsion and London dispersion forces, respectively, and which are often combined to yield a simpler two-body potential such as the Lennard-Jones or buffered 14-7 functional form due to Halgren[9]. The ELEC interactions pertain to permanent electrostatics that are generally described in terms of a truncated point multipole expansion, typically using just point charges. But some of the most recent larger gains in accuracy and improved transferability have been the improvement of general permanent electrostatics through inclusion of higher-order permanent multipoles and incorporation of true many-body electrostatic effects such as polarization (POL). We are currently witnessing the emergence of charge penetration (CP) corrections to permanent electrostatics, charge transfer (CT), and many-body exchange and dispersion functions that may improve our understanding and ultimate description of hydrogen-bonding that is responsible for water's many remarkable properties.

In order to gain the full advantage of these advanced classical potential energy surfaces for water, there are three accompanying theoretical needs to fulfill their promise. The first is the ability to define an appropriate functional form for these non-bonded interactions; the translation of inherently quantum mechanical interactions into a model functional form is a trade-off among the practical considerations of the computational expense, keeping the free parameters to a minimum, and avoiding "overcounting" at short-range where interactions are less decomposable. The second is how to effectively parameterize these new functional forms for maximum transferability; at present there are largely three competing, or perhaps complementary, approaches for determining free parameters- least squares optimization, machine learning, and fitting to the individual terms of Eq. (1) through guidance from an energy decomposition of the quantum mechanical energy. Finally, the increase in model complexity means that computational cost of the energy and their force terms also become more expensive, and new algorithms are needed to solve them. In this review we consider the current state of the art in these areas and where we envision there will be future developments, illustrated using a number of advanced models that are being actively used in water simulations.

ADVANCED POTENTIAL ENERGY SURFACES FOR WATER

Over the last ~10-15 years, next-generation water models have been developed that incorporate many-body effects that are largely lacking in standard water force fields that assume pairwise-additivity of the non-

covalent interactions such as the early SPC models by Berendsen et al[10] and the TIP models introduced by Jorgensen and co-workers[11]. Next generation fixed charge models included optimization of water parameters under an Ewald electrostatic embedding scheme such as TIP4P-Ew[12] and followed later by the TIP4P/2005[13] model. Incorporation of many-body effects in principle enables more accurate modeling of molecular properties across water's phase diagram, as well as affording greater accuracy and transferability for heterogeneous aqueous solutions and interfaces. It may seem like a daunting prospect to sort through the host of different water models. However, one can glean a few key observations on the relative merits of the different advanced water models by paying attention to key essential features of the functional form used, the level of QM quantum mechanics (QM) theory and/or condensed phase data on which the molecular mechanics (MM) model is parameterized, and a recognition that short intermolecular separations where electron densities of the different species overlap is where QM effects, particularly due to exchange-repulsion, start to dominate and become more difficult for MM potentials to capture.[14]

Polarizable Models. Probably the most studied intermolecular interaction that has been added to water force fields is the many-body effect arising from polarization.[5,15-32] Polarization usually receives special attention, as it decays more slowly than dispersion, exchange-repulsion, or charge transfer with a $1/R^4$ dependence, so that it is the next important level of electrostatics beyond the permanent electrostatic field. There are a plethora of polarization models for water, and three main approaches have emerged to calculating polarization in empirical force fields: the fluctuating charge method[17,20,21]; the Drude-oscillator approach[15,23,33,34]; and the well-studied induced dipole method[16,24,25,31,32,35-39]. The fluctuating charge and Drude oscillator approaches are unique from the induced dipole model in that they are essentially attempts to extend previous fixed, atom-centered charge models to accommodate polarization. By contrast, the induced dipole model incorporates multipole moments beyond the point charge in a formalism where the natural link between the higher order permanent multipoles and the polarizabilities is clear from the fact they are terms of a Taylor expansion of the energy in the electric field $\bar{E}$

$$U = \vec{E}\frac{\partial U}{\partial \vec{E}}\bigg|_{\vec{E}=0} + \frac{1}{2!}\vec{E}^2\frac{\partial^2 U}{\partial \vec{E}^2}\bigg|_{\vec{E}=0} + \frac{1}{3!}\vec{E}^3\frac{\partial^3 U}{\partial \vec{E}^3}\bigg|_{\vec{E}=0} + \ldots$$
$$= \vec{\mu}\cdot\vec{E} + \frac{1}{2!}\alpha\vec{E}^2 + \frac{1}{3!}\beta\vec{E}^2 + \ldots \qquad (2)$$

where $\vec{\mu}$ is the permanent dipole moment, α is the dipole polarizability, and β is the dipole hyperpolarizability. It should be noted that Drude models for polarization could also be used with higher order multipoles, although its not typical in most Drude polarization models for water, and many induced dipole models for water only use point charges for the permanent electrostatics.

The polarizable AMOEBA model is based on truncation of Eq. (2) using atom-centered point multipoles up through quadrupoles and point inducible dipoles, which are damped at short-range by effectively smearing out the induced dipole to avoid the "polarization catastrophe" whereby atomic sites at short separation distances polarize each other to infinity. AMOEBA, like many polarization models, uses the Thole smeared charge distribution for damping polarization[40]:

$$\rho = \frac{3a}{4\pi}\exp\left(\frac{-ar_{ij}^3}{\left(\alpha_i\alpha_j\right)^{1/2}}\right) \qquad (3)$$

where r_{ij} is the distance between atomic sites i and j, α_i and α_j are their polarizabilities, and a is a dimensionless width parameter that effectively controls the strength of the damping.

Many-body polarization has demonstrably improved accuracy and transferability of advanced water models by reproducing a number of water properties which were not explicitly fit during the parameterization process including viscosity, self-diffusion constant, and surface tension at room temperature, as well as the ice phases. Another case in point is the IR vibrational spectra for liquid water[41],

which we discuss as a more detailed example here. In simulations of the infrared spectrum of liquid water, the bonding vibrations are typically poorly reproduced by classical force fields due to their lack of accounting for zero point energies and/or charge transfer. However at the lower frequencies probed by THz spectroscopy two prominent features at ~200 cm^{-1} and ~650 cm^{-1} have been identified as collective intermolecular vibrations and librational motions of the hydrogen-bonded network for water, respectively. For many years traditional classical force fields based on non-polarizable force fields struggled to reproduce in particular this intermolecular hydrogen-bonding vibrational signature, illustrated using the SPC/Fw water model shown in Figure 1. When analyzed by *ab initio* molecular dynamics methods using the well characterized PBE functional – which was able to find agreement with the far infrared feature of the experiment – this failure of classical force fields was thought to be attributable to lack of charge transfer.[42] However, classical water models that include polarization are certainly capable of capturing this feature, also shown in Figure 1 for the TTM3-F (discussed in more detail below), and the iAMOEBA and AMOEBA14 models (see Side Bar 1). The primary point is that *no* dynamical quantities were included in the parameter training set of iAMOEBA (which only directly captures direct polarization)[43] and AMOEBA14[32], but were reproduced nonetheless, confirmed by showing that the peak at 200cm^{-1} disappears altogether when polarization interactions are turned off in the simulation of the AMOEBA14 water model.

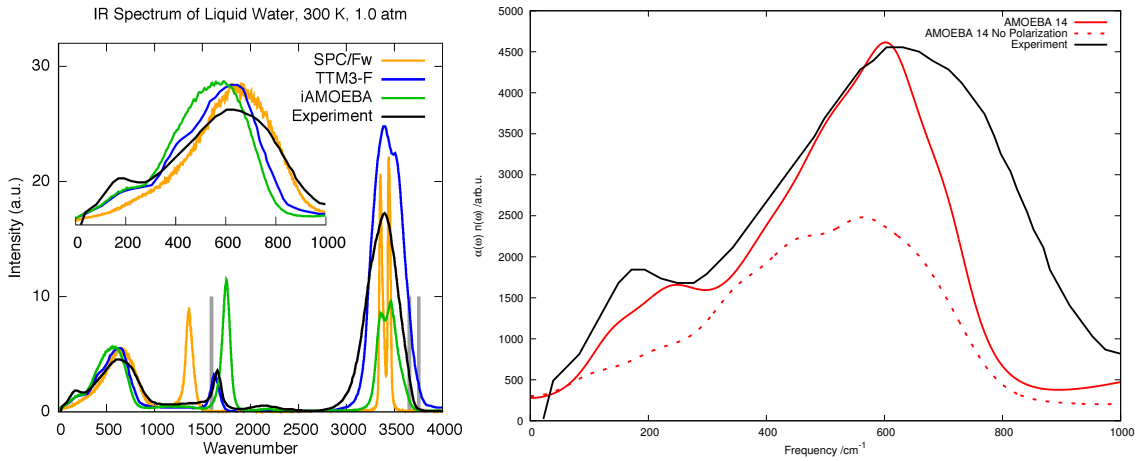


Figure 1. *IR spectra of liquid water from experiment (black) and compared to different classical water models.* (a) using the SPC/Fw, TTM3-F, and iAMOEBA models. Gray bars represent gas phase vibrational frequencies from experiment. Inset: Magnification of the far IR region (< 1000 wavenumber). *Reprinted with permission from ([43]); copyright 2013 American Chemical Society* (b) THz experimental spectra (arbitrary units) of pure bulk water compared to polarizable AMOEBA14 (solid red line) and when polarization interactions are removed (dashed red). *Reproduced from ([44]) with permission from the Royal Society of Chemistry.*

Side Bar 1: Direct and Mutual Polarization Water Models based on the AMOEBA Models

The AMOEBA03 water model, developed by Ren and Ponder[24], has a functional form that includes: full intramolecular flexibility with parameters fitted to gas phase vibrational frequencies, a buffered 14-7 potential centered on both oxygen and hydrogen atoms with parameters fitted to reproduce gas-phase and liquid-phase properties, permanent atomic multipoles up through quadrupoles computed via distributed multipole analysis, and atomic polarizabilities that incorporates Thole damping factor, in which water cluster binding energies were fitted for dimer structures up through the hexamer. In validation studies, AMOEBA03 produced good although not consistent agreement with experiment for thermodynamic, kinetic, and structural properties of liquid water, The iAMOEBA model[43], introduced ten years later, revisited the optimization of AMOEBA parameters with two important modifications: the direct polarization approximation was introduced, thereby omitting all interactions amongst induced dipoles and

removing the need for SCF cycles, and the ForceBalance program was used to optimize the parameters using a more extensive experimental and *ab initio* data set. As the direct approximation changes the form of the interaction, reparameterization of the model was needed to recover quantitative accuracy.

The iAMOEBA model[43] with optimized parameters met or exceeded the AMOEBA03 model in most gas-phase and condensed-phase properties (Figure 2); an extended suite of validation studies showed that iAMOEBA predicts a relatively accurate melting point (261 K) and qualitatively correct phase diagram of high-pressure ices. However, the iAMOEBA approximation leads to a reduction of accuracy in the binding energies of larger water clusters, where the total binding energy is underestimated by ~7% on average compared to AMOEBA03 which predicts a smaller error of ~4%.[31] ForceBalance was also applied to reparameterize the mutual polarization AMOEBA03 model using the iAMOEBA data set, resulting in the AMOEBA14 model, which yielded overall improved agreement with experimental properties.[32] We also developed the uAMOEBA single-site polarizable water model[45] in which the multipoles and induced dipoles were removed from the H atoms, and the remaining parameters optimized using ForceBalance and the same data set; the removal of polarization degrees of freedom from H atoms has precedent in the point dipole[39] and Drude model literature[23]. The uAMOEBA water model features an improvement in the computational efficiency of 3-5 with an accuracy comparable to AMOEBA03, which could be a promising avenue toward speeding up biomolecular simulations that incorporate polarization.[45]

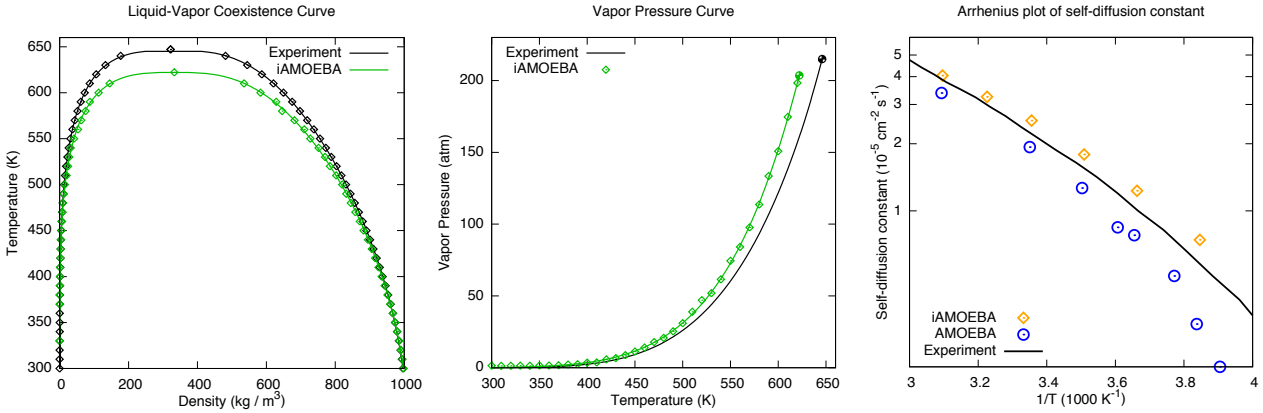


Figure 2. *Comparison of water properties of iAMOEBA water model against experiment*. (a) Arrhenius plot of self- diffusion constant of liquid water vs. temperature, which includes a comparison to AMOEBA03 (b) Liquid-vapor coexistence curve and (c) vapor pressure curve of the iAMOEBA model. Reprinted with permission from ([43]); copyright 2013 American Chemical Society

Ab initio derived water potentials for the condensed phase. Some of the earliest ab initio-derived water potentials are based on so-called fragment methods, exemplified by the effective fragment potential (EFP)[46-52] and X-pol[53-55], in which the MM parameters are derived from QM calculations on individual sub-systems such as monomers, dimers, etc. X-pol relies on a Hartree product of monomer wavefuctions calculated using a semi-empirical single-determinant level of theory, with the addition of 1-electron terms arising from the charges of the other fragments, which themselves are iterated to self consistency. The Hartree product of monomers disobeys anti-symmetry, and a 2-body correction from dimer calculations is added to account for the missing exchange; in the context of the X-pol water model, XP3P, a Lennard-Jones term is added to account for missing exchange and dispersion. X-pol gives excellent agreement with ambient densities and heats of vaporization, adequate diffusion constants, but reports a excessively high dielectric and a density-vs.-temperature profile that is similar to that of the fixed charged TIP models.[54]

EFP similarly obtains its MM parameters directly from *ab initio* calculations, where the noncovalent terms among the rigid fragments consist of electrostatics, polarization, and exchange-repulsion derived from

Hartree-Fock calculations on the monomers, but in its modern form has been supplemented by charge-transfer and dispersion, the later calculated at the MP2 level of theory. The electrostatics and polarization are described by point distributed multipoles at atom centers and bond midpoints and polarizability tensors centered on localized molecular orbital centroids. Electrostatics are damped by an exponential term to account for charge penetration, and polarization is similarly damped to account for exchange-polarization coupling. Exchange-repulsion is expressed with a term based on the orbital overlap between monomers and many-body charge-transfer is approximated to be pairwise-additive function of the orbital overlap and potential exerted by one monomer on the other. Lastly, dispersion is described using a series of C_n/r^n terms where $n \geq 6$ and the C_n coefficents are derived from frequency-dependent polarizability tensor calculation on the fragments, and exchange-dispersion coupling is accomplished through a damping term that is a function of the orbital overlap [48]. Few condensed phase properties have been reported for EFP, primarily radial distribution functions, although recently EFP was shown to yield a melting temperature that was too high[56].

Continued advances in computational power have enabled the development of force fields with less empiricism and based on more accurate levels of QM theory, namely the gold standard for electron correlation coupled cluster singles, doubles, and perturbative triples, CCSD(T), extrapolated to the complete basis set (CBS) limit. This class of models began with the series of Thole-type models (TTM) by Xantheas and co-workers[22,57-61]. Here, and in the models that follow, two- and three-body MM terms are fit to the corresponding CCSD(T)/CBS terms calculated on the water dimer and trimer energy surfaces. The TTM functional forms are relatively simple, with polarization based on isotropic polarizabilities and Thole-type damping, electrostatics using exponentially damped point charges without higher-order multipoles, a Lennard-Jones term for the exchange-repulsion and dispersion, and the flexible intramolecular degrees of freedom parameterized from the spectroscopically accurate functional form developed by Partridge and Schwenke[62]. Both TTM3-F and TTM4-F achieve high accuracy on a number of condensed-phase water properties, through benefit of a cancellation of errors in their 2- and 3-body terms.

More recently, the TTM approach of fitting 2- and 3-body terms to the corresponding 2- and 3-body CCSD(T)/CBS energies has been extended by others, and is exemplified by the CC-pol[63-67], WHBB[68-71], HBB2-pol[72], and most recently, MB-pol[73-76]. Like the TTM models, these rely on isotropic polarizabilities with Thole-type damping with point charges for the electrostatics. The first improvement is more trivial in the inclusion of a dispersion term with Tang-Toennies damping, in contrast to the undamped term of the TTM models. However, the major unique feature is in how the short-range effects are captured. Presumably, these are the effects that predominate in the regime of intermolecular orbital overlap, ascribed to short-ranged effects like exchange-repulsion, charge penetration, and charge transfer that have historically proved difficult to describe using MM functional forms. In contrast to the usual approach of assigning a distinct term to each of these types of short-ranged, many-body terms, the short-ranged 2- and 3-body effects are described collectively by 2- and 3-body permutationally invariant polynomial terms consisting of Born-Mayer-like exponential monomial terms and/or Born-Mayer exponential terms multiplied by r or r^{-1}.[73-76] The 2- and 3-body polynomials can comprise hundreds or even greater than one thousand monomial terms. Moreover, instead of simple dependence on distance between atom centers, additional interaction centers are optimized, effectively accounting for anisotropy in a manner that does not assume any *a priori* notions of where exactly the sites should be located. Lastly, since these terms are meant to capture short-ranged effects, they are only calculated in a small distance range and are smoothly switched off within a cutoff region, beyond which only the simple isotropic Thole-type polarization, point-charge electrostatics, and simple 2-body dispersion are in effect.[73-76]

At this point it behooves us to examine the salient features of models like TTM and MB-pol that recommend their use. First and foremost is the parameterization based on CCSD(T)/CBS. However for MB-pol a central new concept is the recognition that formulations of the short-ranged 2- and 3-body energetics are difficult to capture with single terms corresponding to specific interactions (such as in Eq. (1)), but rather may be better handled by a sum of terms, each of which has the *roughly correct* exponential dependency

(sometimes multiplied by an r-dependency) that follows the general trend of how short-range terms decay in general. In addition, anisotropy is captured through the use of additional sites, but it is not assigned *a priori* based on pre-conceived notions or chemical intuition. Therefore the long-ranged electrostatics, polarization, and dispersion may be kept simpler, since the anisotropy is recognized as a short-ranged effect that is suitably captured in the short-ranged terms. At present, MB-pol achieves unprecedented accuracy in describing water properties from the dimer to the condensed phase and is perhaps one of the all-around best MM water models to date, albeit at a cost that is ~50X that of the AMOEBA force field. However, the large number of polynomial terms in the short-ranged part of the potential will inhibit transferability and application to heterogeneous solution systems, thereby requiring a system-by-system formulation of the MB-Pol potential. The first aqueous system of water-halide solutions for MB-Pol has been completed with notable success[77], but patience will be required for extensions of MB-pol to any arbitrary system of interest.

The future of ab initio derived water potentials. There exist a number of *ab initio*-derived models where the parameters are prescribed *a priori* as in the more familiar empirical force fields utilizing Eq. (1), including the ASP[78,79], NEMO[80-82], SIBFA[83-85], and GEM[14,86] models. Compared with the nearly 50-year history of empirical fixed-charge force fields that started with Lifson and Warshel[87], these are in their infancy. The original ASP model[79] is parameterized from dimer calculations using intermolecular perturbation theory (IMPT), in which electrostatics are described with atom-centered point distributed multipoles, and polarizabilities with atom-centered anisotropic polarizabilities. Owing to the parameterization from dimers, exchange-repulsion, charge-transfer, and dispersion are described by pairwise-additive terms, but are unique in their approach to capturing short-range anisotropy using orientation-dependent shape functions. The NEMO potential, parameterized from HF and MP2 data, similarly models its electrostatics through distributed point multipoles through the quadrupole, and polarization through anisotropic polarizability tensors, and additionally includes quadrupolar polarizability. Dispersion is calculated using a damped C_m/R^n potential, where $6 \leq n \leq 8$ and the C_n are expressed as explicit functions of the polarizabilities. The exchange-repulsion is modeled with a fairly simple, isotropic Born-Mayer exponential form[88,89].

The SIBFA model has a very sophisticated functional form consisting of permanent point multipole electrostatics, anisotropic polarization with short-range attenuation to capture exchange-polarization coupling, elaborate extensions to the description of many-body exchange-repulsion, and many-body charge-transfer with a functional dependency on the electrostatic potential (itself a function of the permanent electrostatics and many-body polarization)[85]. A recent extension of SIBFA is the Gaussian electrostatic model (GEM)[14,86], which contains the same terms as SIBFA, but instead recognizes the finite extent of electron densities, replacing the point multipole description with true static electron densities. In turn, the repulsion term is modified from the original SIBFA as well, and it is taken as the overlap between these densities. In the most recent developments by the Schmidt group[90], the overlap-based prescription of exchange-repulsion has been extended and applied to the damping of the electrostatics and polarization with the rationale that short-range damping is a manifestation of overlap between the electron densities of the separate species. In contrast to the Born-Mayer type functional forms with only an exponential dependency, these newer "beyond Born-Mayer" forms are functionally dependent on an exponential multiplied by a quadratic polynomial in the interatomic separation r[90].

The aforementioned *ab initio*-derived models are only starting to be tested in their ability to reproduce condensed-phase properties of water, since the increased fidelity to the true electron structure comes with increased computational cost. Some of these, such as SIBFA, have not even been enabled for molecular dynamics owing to the lack of analytic gradients until very recently (personal communication). Recently, GEM has been enabled for MD by utilizing the AMOEBA description of polarization and dispersion, GEM*, yielding a model that unfortunately predicts an understructured oxygen-oxygen radial distribution function (RDF) and an overstructured oxygen-hydrogen RDF[37]. On a positive note, GEM* was able to correctly predict trends in the relative energies of water hexamers[14]. It is expected that once

analytical gradient terms appear for all terms in the SIBFA potential, including its native model of polarization, dispersion, and charge-transfer, that these terms will be in turn incorporated into the original GEM, enabling a calculation of condensed-phase water properties. Nevertheless, it should be emphasized that these are early results, and these models have not had the benefit of years of fine-tuning that empirical pairwise-additive, fixed-charge force fields have enjoyed.

OPTIMIZATION APPROACHES TO DETERMINING FREE PARAMETERS

All empirical force fields for water to date inevitably suffer from inaccuracies in the simplifying assumptions underlying the classical functional forms that are used, lack of transferability of parameters, failure to implicitly account for missing effects in the potential, and other shortcomings inherent in the fact that force fields are empirical in nature and rely on fitting to a mixture of quantum mechanical and sometimes condensed-phase experimental data. The success of molecular mechanical force fields for water, especially the simplest pairwise-additive ones, rests on a delicate cancellation of errors among the energetic terms, and an ability to implicitly account for the missing many-body effects such as charge-transfer, and true many-body Pauli repulsion and dispersion. Despite many examples in which advanced potentials for water succeed due to their improved physics, there are also areas of failure in which they are outperformed by their fixed charge counterparts.

While on the face of it such failures of advanced potentials seem to be at odds with what should be a more accurate and transferable model, in fact there are several reasons for the current state of affairs. One is the sheer amount of time that has been devoted to optimizing the pairwise additive classical force fields, and secondly their greater computational tractability permits the necessary sampling to pinpoint their problems. In addition, (i) the advanced functional forms are more difficult to parameterize, since, although they are typically parameterized in an automated fashion targeting QM data from clusters, they also rely on some hand-tuning of their parameters to extrapolate the model to reproduce bulk properties, (ii) they are fit to data like total energies or electrostatic potentials that are only indirectly connected to their piecewise functions, and (iii) they typically rely on but do not demonstrate how cancellation of errors occurs among the molecular interactions accounted for (exchange repulsion, electrostatics, and polarization) or that are missing (charge transfer and charge penetration). Thus the optimization approach of their parameters is a critical area for success of next generation water models.

Energy decomposition analysis for improving water models. It would be highly useful guidance for force field parameterization to benchmark against a theoretical method that is able to ascertain the quality of individual terms of the force field as per Eq. (1). Energy decomposition analysis (EDA)[28,91-99] affords a way to determine the relative contributions of several physically meaningful non-bonded energy terms out of the QM interaction energy, e.g. permanent electrostatics, Pauli repulsion, polarization, dispersion, etc. Although the asymptotic components of any EDA method are uniquely defined[100] for a given electronic structure method, their definitions in the overlapping regions (i.e. water-water interactions in the first solvation shell for example) will differ among different decomposition approaches. However, any well-defined EDA can yield a reasonable and chemically sensible separation of energy components in the overlapping regime – exactly what is required for reasonable force field terms in the same regime. Therefore, by comparing the corresponding terms between an EDA scheme and a force field, one can obtain insight into the strengths and weaknesses of MM formulations, and further develop revised functional forms and/or parameters that in principle should yield substantial improvement in water properties.

There are already successful efforts in this direction such as the effective fragment potential (EFP) method[46-50,52], and some of the most popular EDA schemes are based on a perturbative approach via symmetry-adapted perturbation theory (SAPT)[101-107] that are guiding force field parameters for AMOEBA. In fact, the parameterization of some of the more advanced force fields such as SIBFA and GEM are often guided by EDAs such as the restricted variational space (RVS)[108], constrained space orbital variation (CSOV)[109,110], and most recently SAPT[90,101,102,104,105,107,111] methods. More recently a variational formulation,

such as the second-generation absolutely localized molecular orbitals (ALMO) using density functional theory (DFT) [80,104,10], are currently being used to guide next generation water (and other chemical) potential energy surface models. We believe that using variational EDAs offers advantages over the popular SAPT such as simplicity of terms and avoidance of perturbation theory, and, when used with accurate low-cost density functionals[112-114], is also very computationally efficient. Side Bar 2 describes how ALMO-EDA was used to analyze how well the AMOEBA water model reproduces the 2-body[115] as well as 3-body[116] energies in the distance scans for the genuinely many-body terms of QM energetics, including modified Pauli repulsion, dispersion, polarization, and charge transfer contributions. Since AMOEBA's only many-body term arises from Thole-damped polarization, the analysis must address not only how successfully it renders agreement with the corresponding ALMO-EDA polarization, but whether the 2-body and 3-body sum of ALMO's modified Pauli repulsion, dispersion and charge transfer terms are captured by 3-body polarization or whether it is spread "incoherently" across, for example, the 2- and 3-body polarization contributions or accounted for in strictly two-body terms. This illustrates how future models might be tuned when EDA decomposition data is combined with sophisticated least squared optimization methods such as ForceBalance or machine learning methods, which are described next.

Sidebar 2: Future Water Models Based on Guidance from Energy Decomposition Analysis. We have used ALMO-EDA to assess the quality of the non-covalent terms in the polarizable force field AMOEBA03[25] for the water dimer, water trimer, and a range of water-ion dimer and trimer systems.[115,116] To illustrate its usefulness for water models, the breakdown of AMOEBA's total energy into the total polarization energy contribution for the water trimer, and its further breakdown into 2-body and 3-body polarization, is shown in Figure 3. Compared to the high quality ωB97X-V DFT benchmark the overall total intermolecular energy curve for the AMOEBA water trimer is underbound throughout the entire range of distances. Further breakdown of the many-body polarization into a many-body expansion, expected to converge quickly for an insulator such as water, the 2-body polarization shows excellent agreement with the ALMO-EDA, and AMOEBA's 3-body polarization appears to capture 3-body polarization explicitly and 3-body charge transfer implicitly. Thus the total energy error over the distance scan is attributable to the permanent electrostatics using point multipoles that are excessively repulsive due to lack of charge penetration, and the pairwise additive 14-7 van der Waals wall that is insufficiently softened to correct for that, with perhaps inadequate capturing of 2-body charge transfer. EDA calculations and proposed improvements to the basic AMOEBA model are now beginning to appear in the literature.[115-119]

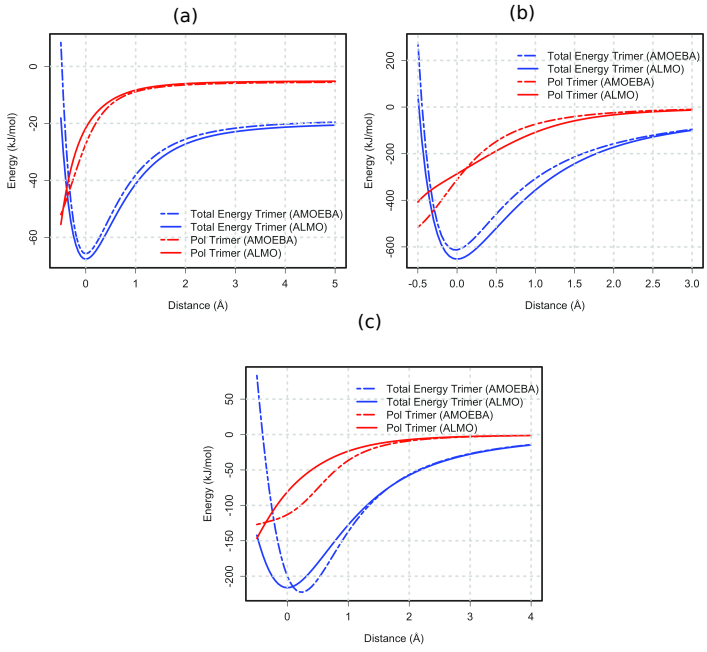


Figure 3. *Comparison of the ALMO-EDA decomposition of the intermolecular energy profile against AMOEBA03[25] for the water trimer.* (a) Total energy and total polarization energy for AMOEBA against the ωB97X-V DFT benchmark and its decomposition using ALMO-EDA for polarization. (b) The 2-body polarization energy for one of the three pairs in the trimer. (c) The 3-body polarization as well as the sum of ALMO's 3-body polarization and charge transfer terms. The distance coordinate corresponds to displacement from equilibrium from the

reference geometries. Reprinted from [116], with the permission of AIP Publishing.

Automated parameterization methods. The parameterization of water models may incorporate training data from diverse experimental and *ab initio* theoretical data sources. In the parameterization procedure, the model is used to simulate physical quantities that are directly compared to the training data, and the parameters are adjusted iteratively to make the differences as small as possible. Experimental data sources are uniquely abundant for water, and include measured values of physical properties including thermodynamic, kinetic, structural, interfacial, and phase change properties across a wide range of temperatures and pressures.[120-127]. Empirical equations of state fitted to the experimental data provide a convenient means for retrieving accurate values for many of these properties at specified temperature and pressure values. Certain physical properties such as the liquid density are particularly well-suited for direct comparison between simulation and experiment; other properties such as the heat of vaporization require *ad hoc* corrections for approximations or assumptions made in the water model or simulation method. In fact, simulations and models that incorporate more physical detail (for example, electronic polarization) have an advantage in that their simulated properties are more directly comparable to the training data, and fewer *ad hoc* corrections are needed (for example, the self-polarization correction used in developing SPC/E has been re-examined in more recent work). When developing water models intended for classical Hamiltonian simulations, the size of nuclear quantum effects on different experimental properties must be considered; the enthalpy of vaporization and isobaric heat capacity have significant quantum effects requiring corrections. For example, the development of TIP4P-Ew required adjusting the experimental target for heat of vaporization and isobaric heat capacity to reflect how the population of high-frequency vibrational modes depends on temperature and phase;[12] this procedure was reproduced in the parameterization of iAMOEBA, AMOEBA14 and uAMOEBA. Even with the modified target values, the fully flexible models tend to overestimate the heat capacity because the high-frequency ($\hbar\omega >> k_B T$) degrees of freedom are not frozen out, as in the case of a quantum system.

Theoretical data sources include *ab initio* calculated values of total potential energies, nuclear gradients, and interaction energies for small water clusters.[128] Calculated electronic properties such as multipole moments and higher-order response properties such as vibrational frequencies may also be used. EDA, described in the previous section, is particularly useful for parameterizing physically motivated potential *terms* in a water model; when used alongside other data sources, the EDA guards against overfitting of model parameters to the total properties of the system. The approximations in the *ab initio* method, the empirical model and the classical simulation imply that the optimized model *should* deviate somewhat from the training data, and this comparison becomes increasingly fraught with more approximate empirical models. Explicit polarization is important for quantitative comparisons to *ab initio* data in the gas phase; fixed-charge models rely on *ad hoc* schemes to approximate polarization in a mean-field sense, which are difficult to improve upon systematically.

The choice of training data is only one dimension of variability in the space of possible parameterization strategies; two other dimensions are the choice of parameters being optimized (including restraints on these parameters), and the optimization method being used. The development of a water model involves producing the training data set, running simulations, and fitting parameters; overall this is a task with many interconnected components that is arduous to carry out and even more difficult to reproduce. The parameterization workflow is usually accomplished using scripts to glue the required components together; a relatively early example is a *tcsh* script for simplex optimization by Faller and coworkers.[129] More recently, several parameterization programs have been made available for further generality and reproducibility; these include ForceBalance (developed by one of us),[31,53] *potfit* by Brommer and coworkers,[130] and Wolf(2)Pack by Hulsmann and coworkers.[131] We also note related research in the AMOEBA, AMBER and CHARMM simulation communities that provide automated programs for parameterizing new molecules by following fixed procedures; these methods are not directly applicable to water or developing novel functional forms.

ForceBalance is a software package for systematic and reproducible model parameterization that has been used to develop a series of water models; these include a polarizable model based on QTPIE (charge transfer),[31] iAMOEBA (direct induced dipoles),[31] AMOEBA14 (mutual induced dipoles),[32] uAMOEBA (single-site mutual induced dipoles),[45] as well as TIP3P-FB and TIP4P-FB (fixed charge).[53] In addition, ForceBalance was used to develop AMOEBA vdW parameters for organochlorine compounds,[74] AMBER-style protein force field parameters,[132] GROMOS-style parameters for phospholipid bilayers,[133] semiempirical parameters for liquid water,[134] and auxiliary grids for the tensor hypercontraction (THC) approximation of MP2.[135] The code introduces three key abstractions that help to accommodate diverse model parameterization workflows:

- The *force field* is a convenient way to represent a plain text or XML file containing numerical values to be optimized, and provides a method for writing copies of the file with modified values. Importantly, the force field allows functional relationships between parameters, as well as constraints and rescaling factors; these are often needed for parameters with physical meanings and which may have very different orders of magnitude depending on the unit system.
- The *engine* is an interface to the simulation software package that implements the model, which can be done using APIs (when available) or the operating system. Engine implementations include OpenMM, AMBER, TINKER, Gromacs, and Psi4.
- The *target* represents an observable that can be calculated using the model and directly compared to a stored reference value; the *objective function* is a weighted sum of least-squares errors from multiple targets, plus a regularization term that penalizes parameter overfitting.

In an optimization cycle (presented graphically in Figure 4), the current values of optimization parameters are passed to the force field object to create a parameter file. The targets then call the engine functions (and by extension, the external codes) to evaluate the observables needed to compute the objective function as well as its derivatives. An optimization algorithm then predicts the next set of optimization parameters to minimize the objective function. ForceBalance implements several optimization algorithms including interfaces to methods in the *SciPy* package for scientific computing; in practice the best performance is obtained from a natively implemented quasi-Newton algorithm that uses the first derivatives of the properties. We have not found evidence for multiple minima in the parameter space for any of the model development projects, though this must be kept in mind whenever gradient-based optimization workflows are used.

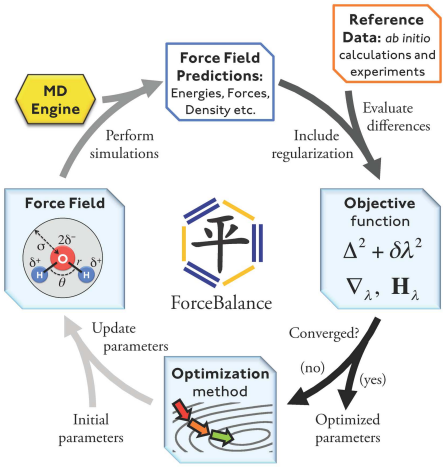


Figure 4. *Steps of the ForceBalance optimization cycle.* The initial force field parameters (lower left) are used to perform simulations using molecular dynamics (MD) software (upper left). The objective function is computed as a least-squares function of the differences between simulation results and reference data (upper right). The optimization method updates the parameters in order to minimize the objective function (bottom right).

In order to use the quasi-Newton optimizer, ForceBalance requires first derivatives of all calculated properties with respect to the parameters being optimized. Derivatives of single-point properties (e.g. energies and gradients) are carried out via finite difference; simulated thermodynamic properties are more

challenging due to the high computational cost and statistical uncertainty inherent to running a simulation. ForceBalance implements semi-analytic expressions for efficiently obtaining parametric derivatives of many thermodynamic properties without needing to run multiple simulations. A statistical mechanical fluctuation formula[136] provides the parametric derivatives of a general thermodynamic property A as:

$$\frac{\partial\langle A\rangle}{\partial\lambda} = \langle\frac{\partial A}{\partial\lambda}\rangle - \frac{1}{k_B T}\left(\langle A\frac{\partial E}{\partial\lambda}\rangle - \langle A\rangle\langle\frac{\partial E}{\partial\lambda}\rangle\right) \quad (4)$$

where λ is the model parameter, $\langle\cdot\rangle$ the ensemble average using the current value of λ, and E the potential energy. Because A and E can be evaluated individually for trajectory frames in the simulation, the quantities on the RHS may be evaluated in a post-processing step by making small changes in λ and looping over the trajectory frames. In practice, this approach is highly effective in accurately fitting thermodynamic properties of water; we typically use six experimental properties (density, heat of vaporization, thermal expansion coefficient, isothermal compressibility, isobaric heat capacity, dielectric constant) over a wide temperature and pressure range. This data, in combination with a large and multifaceted *ab initio* data set, can be accurately fitted using ForceBalance and the AMOEBA functional form. Looking towards the future, we will incorporate EDA into ForceBalance, which we expect will lead to models with improved accuracy and transferability.

Machine Learning approaches to parameterization. Machine learning (ML), broadly defined, consists of training a general model using a large data set in order to make predictions outside the training sata set. Driven by the burgeoning availability of large data sets and increased computational capabilities, ML methods have significantly improved over the last ten years and made major impacts in science and beyond. In the context of molecular simulations, ML – specifically, supervised learning – is used to build a model that predicts physical properties (e.g. potential energies) from the molecular structure, by training on an *ab initio* data set where the target outputs are known.[137] The model parameters are fitted by minimizing a least-squares function of the errors between the model output and training data, similar to the ForceBalance and other optimization procedures discussed above. However, in contrast to physically motivated optimization models, ML models are highly flexible with the ability to fit almost any data, but often with a trade-off that the individual parts of the model may have no direct physical interpretability.

One archetype of ML model is the artificial neural network (ANN); one simple example of which is the multilayer perceptron (MLP). The basic element of the MLP is the *node* or *neuron* – a nonlinear function maps multiple inputs to one output. The nodes are organized into *layers*, where the outputs of one layer are inputs to the next one. The input layer consists of the geometric parameters of the cluster of nodes (called *features*), and is followed by one or more hidden layers, with the definition of "deep learning" referring to many hidden layers. Each hidden node computes the output variable y from input variables x_i using a nonlinear function such as

$$y = (1 + \exp\left[(a - \textstyle\sum_i w_i x_i)/\sigma\right])^{-1} \quad (5)$$

where a, σ, and w_i are adjustable parameters, and the sum is over the number of inputs. The sigmoidal form of the function ensures the output goes smoothly from 0 to 1 as the weighted sum $\sum_i w_i x_i$ increases beyond the threshold value a, roughly mimicking the biological function of a neuron. The final output is the physical property or data representation to be predicted. For computing basic Boolean operations such as the simple XOR function, the parameters in Eq. (5) are easily derived to define a "decision plane" that separates the "on" from the "off" solutions. However, for more complex problems we can't write down a solution for parameters that correctly determines the mapping of the input space {x} to output space {y}, i.e. the determination of the decision hyperplane. In order to find this hyperplane the ANN is provided some representative examples in which to learn the mapping. If we are to maximize the fidelity of this mapping, then it requires minimization of the deviation, D, of the predicted output, y, from the observed output, O:

$$D = \frac{1}{2}\sum_{\mu}^{M}\sum_{i}^{N}\left[O_i^{\mu} - y\left(a_i^{\mu}, \sigma_i^{\mu}, w_i^{\mu}\right)\right]^2 \tag{6}$$

where μ is a sum over the M examples, and i is the sum over the N output units. Hebb's rule provides a way of varying weights and thresholds to maximize fidelity of the network to learn the input/output mapping from example

$$\delta w_{ik} = \varepsilon\left[O_i^{\mu} - y\left(a_i^{\mu}\right)\right]x_k \tag{7}$$

where ε is the "learning intensity", but the astute reader will recognize this as just steepest descents. Thus the basic formulation of a feedforward-back propagation ANN is to ensure that the training set is composed of data examples that are representative of the mappings between inputs {x} and the observations, O, and the ANN encoding of input and output should not be so opaque that the learning process is hampered. Because each node has independent parameters, the model is highly flexible and general for fitting of parameters. Other kinds of ANNs include those that employ radial basis functions (RBFs); here the final output y is computed from the feature vector $\boldsymbol{x}$ as: $y = \sum_i w_i \exp[-\beta_i \|\boldsymbol{x} - \boldsymbol{c}_i\|^2]$, where w_i, β_i and $\boldsymbol{c}_i$ are fitting parameters and the sum runs over the chosen number of RBFs. The Gaussian function is used here as an example but other functions that depend on distance may be used; the output can roughly be interpreted as a weighted sum over "centers" where the contributions depend on the distance from the feature vector to each center.

Gaussian process (GP) regression, or kriging, is another important class of ML model that may be regarded as a type of interpolation.[138,139] The central concept is a probability distribution of functions of the feature space. If we draw a random function $f(\boldsymbol{x})$ from this distribution, the probability of observing some value of the property y at $\boldsymbol{x}$ is a Gaussian random variable with a mean μ and variance σ^2. The central assumption is that pairs of observed values (e.g. y_x and y_z, observed at $\boldsymbol{x}$ and $\boldsymbol{z}$ respectively) are *correlated* and decay with distance, which is reasonable if we assume the functions are smooth on a characteristic length scale ξ_d (d indexes the dimensionality of the feature space). This is mathematically described as:

$$\mathrm{Cov}[f(\boldsymbol{x}), f(\boldsymbol{z})] = \exp[-\textstyle\sum_d \xi_d |x_d - z_d|^{p_d}] \tag{8}$$

where both ξ_d and p_d are adjustable parameters. Finding the parameters of the GP model involves maximizing a *likelihood function* of the model parameters, given that the training data set has already been observed (the set of values y_i at the feature vectors $\boldsymbol{x}_i$); in practice, determining these parameters requires inverting a matrix with dimensionality equal to the size of the training data set. To evaluate the model prediction for a new data point $\boldsymbol{x}^*$, we maximize another likelihood function of $y(\boldsymbol{x}^*)$, given the current values of model parameters and observations in the training data set. The result is given by

$$y_{\max}(\boldsymbol{x}^*) = \mu + \boldsymbol{r}^T\boldsymbol{R}^{-1}(\boldsymbol{y} - \mu) \tag{9a}$$

where

$$r_i = \mathrm{Cov}[f(\boldsymbol{x}_i), f(\boldsymbol{x}^*)],\ R_{ij} = \mathrm{Cov}\left[f(\boldsymbol{x}_i), f\left(\boldsymbol{x}_j\right)\right], \tag{9b}$$

and $\boldsymbol{y}$ is the array of observations from the training data. The GP regression model has been used by Brookes, Demerdash and Head-Gordon to correct for missing higher order many-body forces for water[140] in the context of the many-body expansion of AMOEBA known as 3-AMOEBA[141].

An early ANN model of the water dimer potential surface was introduced by No and coworkers.[80] Popelier and coworkers applied several ML approaches to accurately describe the environmental dependence of multipole moments of water molecules in clusters up to the hexamer.[139,142] Behler and coworkers developed ANN models to fit the short-range part of the intermolecular interactions and fitted energies for neutral clusters containing up to 16 molecules,[143] as well as a number of protonated water clusters;[144] more recently these simulations have been applied in the condensed phase to study aqueous solutions of NaOH.[145] We expect that ML models will continue to make an impact in the simulation of

water, perhaps in combination with physically motivated models; the combined application of many-body expansions with ANN potentials has been explored recently.[146]

NEW ALGORITHMS FOR SOLVING MANY-BODY POLARIZATION

Concurrent to the development of an advanced water model is the equally important need to improve the computational efficiency of its calculations through better methodology. Historically the polarization solution for the point induced dipole model are solved through self-consistent field (SCF) iterative solvers, such as successive over-relaxation (SOR)[147], preconditioned conjugate gradient (PCG)[148], or direct inversion in the iterative subspace (DIIS)[149] methods. More recent approaches have improved upon the computational cost of these standard SCF solvers. One such example is the truncated conjugate gradient (TCG) method, which minimizes the number of matrix-vector multiplications and is amenable to scaling on modern high-performance computing platforms.[150] The 'optimized perturbation theory' (OPT), which cleverly uses a perturbation of the polarization potential, is truncated at a tractable order and is then empirically fit to approximately recover the fully converged result[151-153], all of which have been tested on bulk water systems.

By contrast, Drude and fluctuating charge models for polarization are typically solved through an extended Lagrangian (EL) formulation to treat polarization with negligible cost compared to the SCF approaches[23,34,38]. In the case of Drude oscillators the EL equation of motion is based on a mass repartitioning between the parent atom and its Drude oscillator, with the goal of making the Drude mass small enough to obey the Born Oppenheimer condition. Even so, a basic EL approach using thermalized "hot" Drude oscillators can be plagued with problems of accuracy since the effective polarization vector fluctuates around an average orientation that does not conform to the true electric field vector, and/or problems of stability in the context of a MD trajectory that forces the reduction of the time step to be unacceptably short. Instead, Lamoureux and Roux developed an EL approach whereby the polarization degrees of freedom are kept cold at a temperature T* relative to the temperature of the real degrees of freedom, T, such that T*<<T.[23] Based on this two temperature canonical or isothermal isobaric ensemble (NVT,T* or NPT,T*), the EL(T,T*) schemes were found to be stable on the 1.0-2.0 fs timescale with much better accuracy for the polarizable SPC water model (PSPC)[38].

In contrast to these SCF and EL schemes, we have adapted a time-reversible formulation of *ab initio* dynamics introduced by Niklasson and colleagues[154-158] to the problem of classical polarization[159-161]. Our "inertial EL/SCF" (iEL/SCF) method is a hybrid of an extended Lagrangian and an SCF solution, in which an extended set of auxiliary induced dipoles is introduced and dynamically integrated so as to serve as a time-reversible initial guess for the SCF solution of the real induced dipoles[159] as given in Eq. (10)

$$\mathcal{L}_{hybrid}^{dipole} = \frac{1}{2}\sum_{i=1}^{N} m_i \dot{\vec{r}}_i^2 + \frac{1}{2}\sum_{i=1}^{N} m_{\mu,i}\, \dot{\vec{\mu}}_i^2 - U\left(\vec{r}^{\boldsymbol{N}}, \vec{\mu}_{SCF}^{\boldsymbol{N}}\right) - \frac{1}{2}\omega^2 \sum_{i=1}^{N} m_{\mu,i} \left(\vec{\mu}_{SCF,i} - \vec{\mu}_i\right)^2 \qquad (10)$$

The iEL/SCF method was shown to drop the number of SCF iterations by half for the AMOEBA polarizable model for water[159], and reduces the number of SCF cycles from ~15-20 to ~3-5 for a small box of water using linear scaling DFT in Onetep[162]. In 2017 we introduced a new iEL/SCF method that completely eliminates the need for any SCF iterations, while still using a standard length time step of 1.0 fs for point induced dipoles, illustrated with the AMOEBA model, which we call iEL/0-SCF (i.e. no self consistent field iterations)[161] method. Figure 5 confirms that the properties of calculating mutual polarization with iEL/0-SCF is equivalent to the quality of a tightly converged SCF solution, and is effectively as fast as using a multi-time stepping method with an outer time step of 2 fs. We have recently extended the iEL/0-SCF approach to Drude polarization illustrated with the PSPC polarizable water model[160]. In this case we were able to extend the standard molecular dynamics time step to 6 fs – a factor of 6X increase in time steps compared to standard EL(T,T*) approaches.

The import of this recent work on new solutions to many-body polarization is as follows: it is now possible to evaluate an important many-body effect – polarization – but at a computational cost of a direct polarization model[31,163], i.e. primarily the cost of the chosen model for the permanent electrostatics.

Furthermore, the difference between the PSPC and AMOEBA water models illustrate an important design choice for including polarization. Because the PSPC model is a rigid model, with simple point charge permanent electrostatics, and with no Drude polarization on light hydrogen centers, the time step can be pushed close to an order of magnitude longer. By contrast in order to accurately integrate the forces arising from the fast varying electric fields from permanent dipoles and (especially) quadrupoles, combined with their presence on hydrogens with flexible bonds to oxygen, means that the numerical integration time step must be greatly reduced. Therefore the advanced classical model design choices effects how much statistical sampling is possible.

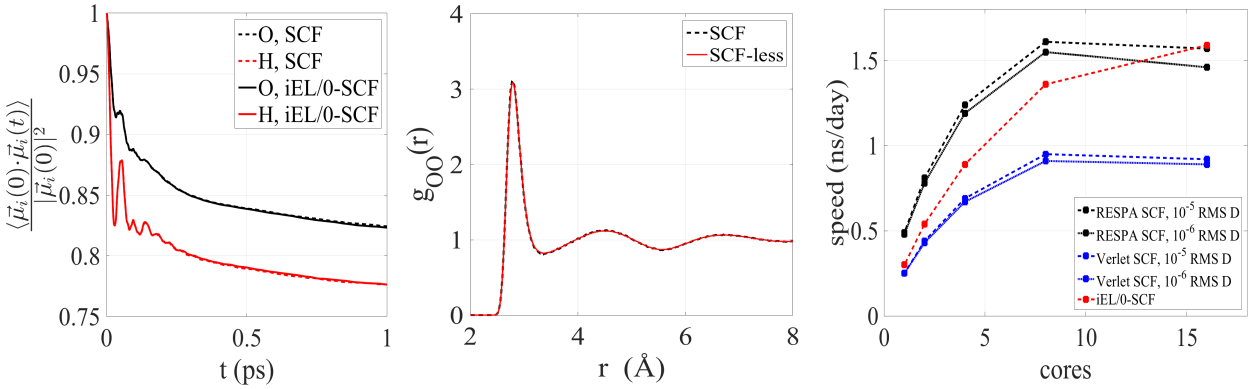


Figure 5: *Comparisons of the standard preconditioned conjugate gradient SCF solver at 10^{-6} RMSD convergence and the SCF-less method for AMOEBA water.* (a) Time autocorrelation function of the induced dipoles for oxygen and hydrogen; (b) Oxygen-oxygen radial distribution function; (c) simulation speed-up in nanoseconds per day for OpenMP scaling as a function of the number of cores for a box of 512 water molecules in the NVT ensemble at 298.0 K. Reprinted with permission from ([161]); copyright 2017 American Chemical Society.

CONCLUSION

Major effort is underway to develop improved MM models of water that seek to address the shortcomings of classical, pairwise-additive fixed charge, manifested most clearly in their difficulty in describing heterogeneous systems and the properties of water across the phase diagram. Historically, such advanced force fields, many of which include the leading-order many-body effect of polarization, have faced obstacles in their widespread adoption owing to computational cost and difficulty in their parameterization that have precluded their widespread use. The purpose of this review is to underscore the major advances in the development of advanced molecular mechanics water models in their parameterization, prescription of functional form, and computational efficiency that are rendering them competitive with standard pairwise-additive fixed charge force fields.

First, we introduce the standard functional forms used to capture the leading-order many-body effect missing from pairwise-additive potentials embodied in full mutual polarization, exemplified by the AMOEBA model. Aside from the noted advantages of polarization in allowing for transferability, we underscore the distinct ability of polarization to capture IR spectroscopic features of the cooperative hydrogen-bonding network, which pairwise-additive potentials cannot recover. While the original AMOEBA model[24] demonstrated notable inconsistencies in its ability to model condensed-phase properties, reparameterization efforts using the ForceBalance[31,53] algorithm have generated models that show remarkable accuracy across the phase diagram, even yielding a computationally efficient polarization model, iAMOEBA[43], that responds only to the permanent electrostatic, or direct, field, eliminating the need for expensive iterative SCF calculations. ForceBalance exemplifies a novel set of approaches towards

optimizing parameters in a systematic fashion by allowing multiple training targets, from oligomeric to condensed-phase properties, to be fit to simultaneously.

In addition to models that rely on parameterization approaches relying on experimental and *ab initio* data, on the other end of the spectrum are models that are parameterized entirely on *ab initio* data, either total QM energies or EDA schemes. EDA schemes afford a breakdown of total QM energies into physicochemically sensible contributions, and can be especially helpful in guiding the parameterization of potentials in regions where intermolecular orbital overlap, and therefore quantum mechanical effects such as exchange-repulsion, becomes non-negligible. A number of force fields in which elaborate functional forms are prescribed for each of the distinct non-covalent contributions are being developed.

An interesting approach towards the formulation of *ab initio*-based MM functional forms recognizes 1) that the ability of the *ab initio* reference calculation to capture electron correlation is critical; and 2) that short-ranged, QM-dominated 2- and 3-body effects, particularly owing to charge transfer, exchange-repulsion, and charge penetration, may inherently be difficult to capture with the standard approach of matching a single physical effect to a distinct functional form; and 3) that anisotropy is important at short-range and should be determined systematically instead of by potentially erroneous chemical intuition. These approaches recognize that such QM-dominant effects may be expressed collectively as an expansion in a basis, each of whose terms represent approximately the known exponential or distance-times-exponential decay at short range. This family of potentials have culminated in the development most recently of MB-pol[58-60], which achieves unprecedented accuracy for water from the dimer to the condensed-phase. An additional crucial feature of such models is that since the QM effects that are difficult to model are short-ranged, the prescriptions for the long-ranged electrostatics, polarization, and dispersion may be kept relatively simple. However while powerful artillery, the MB-Pol water model is not very mobile in its deployment on arbitrary chemical system beyond pure water and simple halide-water systems. Yet another set of approach towards capturing the complexity at short-range are the machine-learned methods that recognize that a MM prescription faithful to electronic structure perhaps may not be rendered easily in a human-readable functional form as is traditionally used in force fields.

Lastly, and very crucially, we show that the major impediment to the adoption of polarizable models, the computational cost of solving for the dipoles, is now being overcome with novel computational techniques. Very recent developments in that area include truncated conjugate gradient methods[150], perturbative methods[151-153], and methods that render the EL approach stable and robust, reducing[159,162] or avoiding SCF[160,161] entirely. These computational efficiencies hold the promise of enabling advanced polarizable force fields to become competitive with and be treated on equal footing with traditional pairwise-additive force fields for water.

ACKNOWLEDGMENTS

THG and OND thank the National Science Foundation Grant No. CHE-1363320 and CHE-1665315 for support of this work. LPW acknowledges support from the ACS Petroleum Research Fund, Award Number 58158-DNI6-0.

REFERENCES

(1) Geissler, P. L.; Dellago, C.; Chandler, D.; Hutter, J.; Parrinello, M. Autoionization in Liquid Water. *Science* **2001,** *291* (5511), 2121.

(2) Marx, D. Proton Transfer 200 Years after von Grotthuss: Insights from Ab Initio Simulations. *ChemPhysChem* **2006,** *7* (9), 1848.

(3) Tuckerman, M.; Laasonen, K.; Sprik, M.; Parrinello, M. Ab initio molecular dynamics simulation of the solvation and transport of hydronium and hydroxyl ions in water. *The Journal of Chemical Physics* **1995,** *103* (1), 150.

(4) Schmitt, U. W.; Voth, G. A. The computer simulation of proton transport in water. *The Journal of Chemical Physics* **1999,** *111* (20), 9361.
(5) Paesani, F.; Iuchi, S.; Voth, G. A. Quantum effects in liquid water from an ab initio-based polarizable force field. *Journal of Chemical Physics* **2007,** *127* (7), 074506.
(6) Habershon, S.; Markland, T. E.; Manolopoulos, D. E. Competing quantum effects in the dynamics of a flexible water model. *The Journal of Chemical Physics* **2009,** *131* (2), 024501.
(7) Ceriotti, M.; Fang, W.; Kusalik, P. G.; McKenzie, R. H.; Michaelides, A.; Morales, M. A.; Markland, T. E. Nuclear Quantum Effects in Water and Aqueous Systems: Experiment, Theory, and Current Challenges. *Chem Rev* **2016,** *116* (13), 7529.
(8) Marsalek, O.; Markland, T. E. Ab initio molecular dynamics with nuclear quantum effects at classical cost: Ring polymer contraction for density functional theory. *The Journal of Chemical Physics* **2016,** *144* (5), 054112.
(9) Halgren, T. A. Representation of Vanderwaals (Vdw) Interactions in Molecular Mechanics Force-Fields - Potential Form, Combination Rules, and Vdw Parameters. *J Am Chem Soc* **1992,** *114* (20), 7827.
(10) van der Spoel, D.; van Maaren, P. J.; Berendsen, H. J. C. A systematic study of water models for molecular simulation: Derivation of water models optimized for use with a reaction field. *The Journal of Chemical Physics* **1998,** *108* (24), 10220.
(11) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **1983,** *79* (2), 926.
(12) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *Journal of Chemical Physics* **2004,** *120* (20), 9665.
(13) Vega, C.; Abascal, J. L. F.; Sanz, E.; MacDowell, L. G.; McBride, C. Can simple models describe the phase diagram of water? *J Phys-Condens Mat* **2005,** *17* (45), S3283.
(14) Cisneros, G. A.; Wikfeldt, K. T.; Ojamae, L.; Lu, J. B.; Xu, Y.; Torabifard, H.; Bartok, A. P.; Csanyi, G.; Molinero, V.; Paesani, F. Modeling Molecular Interactions in Water: From Pairwise to Many Body Potential Energy Functions. *Chemical Reviews* **2016,** *116* (13), 7501.
(15) Sprik, M.; Klein, M. A polarizable model for water using distributed charge sites. *Journal of Chemical Physics* **1988,** *89* (12), 7556.
(16) Warshel, A.; Kuwajima, S. Incorporating Electric Polarizabilities in Water-Water Interaction Potentials. *Journal of Physical Chemistry* **1990,** *94*, 460.
(17) Rick, S.; Stuart, S.; Berne, B. Dynamical fluctuating charge force-fields: Application to liquid water. *Journal of Chemical Physics* **1994,** *101* (7), 6141.
(18) Gao, J. L. Toward a molecular orbital derived empirical potential for liquid simulations. *J Phys Chem B* **1997,** *101* (4), 657.
(19) Chialvo, A. A.; Cummings, P. T. Molecular-based modeling of water and aqueous solutions at supercritical conditions. *Adv Chem Phys* **1999,** *109*, 115.
(20) Stern, H. A.; Kaminski, G. A.; Banks, J. L.; Zhou, R. H.; Berne, B. J.; Friesner, R. A. Fluctuating charge, polarizable dipole, and combined models: Parameterization from ab initio quantum chemistry. *J Phys Chem B* **1999,** *103* (22), 4730.
(21) Chen, B.; Xing, J.; Siepmann, J. I. Development of Polarizable Water Force Fields for Phase Equilibrium Calculations. *The Journal of Physical Chemistry B* **2000,** *104* (10), 2391.
(22) Burnham, C. J.; Xantheas, S. S. Development of transferable interaction models for water. IV. A flexible, all-atom polarizable potential (TTM2-F) based on geometry dependent charges derived from an ab initio monomer dipole moment surface. *Journal of Chemical Physics* **2002,** *116* (12), 5115.
(23) Lamoureux, G.; MacKerell, A. D.; Roux, B. A simple polarizable model of water based on classical Drude oscillators. *Journal of Chemical Physics* **2003,** *119* (10), 5185.
(24) Ren, P. Y.; Ponder, J. W. Polarizable atomic multipole water model for molecular mechanics simulation. *J Phys Chem B* **2003,** *107* (24), 5933.

(25) Ren, P.; Ponder, J. W. Temperature and pressure dependence of the AMOEBA water model. *J Phys Chem B* **2004,** *108*, 13427.
(26) Yu, H. B.; Hansson, T.; van Gunsteren, W. F. Development of a simple, self-consistent polarizable model for liquid water. *Journal of Chemical Physics* **2003,** *118* (1), 221.
(27) Benjamin, K. M.; Schultz, A. J.; Kofke, D. A. Virial Coefficients of Polarizable Water Applications to Thermodynamic Properties and Molecular Clustering. *Journal of Physical Chemistry C* **2007,** *111*, 16021.
(28) Bauer, B. A.; Patel, S. Properties of water along the liquid-vapor coexistence curve via molecular dynamics simulations using the polarizable TIP4P-QDP-LJ water model. *Journal of Chemical Physics* **2009,** *131* (8).
(29) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A., Jr.et al. Current status of the AMOEBA polarizable force field. *J Phys Chem B* **2010,** *114* (8), 2549.
(30) Lee, A. J.; Rick, S. W. The effects of charge transfer on the properties of liquid water. *Journal of Chemical Physics* **2011,** *134* (18).
(31) Wang, L. P.; Chen, J. H.; Van Voorhis, T. Systematic Parametrization of Polarizable Force Fields from Quantum Chemistry Data. *Journal of Chemical Theory and Computation* **2013,** *9* (1), 452.
(32) Laury, M. L.; Wang, L. P.; Pande, V. S.; Head-Gordon, T.; Ponder, J. W. Revised Parameters for the AMOEBA Polarizable Atomic Multipole Water Model. *J Phys Chem B* **2015,** *119* (29), 9423.
(33) Sprik, M. Computer simulation of the dynamics of induced polarization fluctuations in water. *The Journal of Physical Chemistry* **1991,** *95* (6), 2283.
(34) Lopes, P. E.; Roux, B.; Mackerell, A. D., Jr. Molecular modeling and dynamics studies with explicit inclusion of electronic polarizability. Theory and applications. *Theor Chem Acc* **2009,** *124* (1-2), 11.
(35) Piquemal, J. P.; Williams-Hubbard, B.; Fey, N.; Deeth, R. J.; Gresh, N.; Giessner-Prettre, C. Inclusion of the ligand field contribution in a polarizable molecular mechanics: SIBFA-LF. *J Comput Chem* **2003,** *24* (16), 1963.
(36) Cisneros, G. A. Application of Gaussian Electrostatic Model (GEM) Distributed Multipoles in the AMOEBA Force Field. *J Chem Theory Comput* **2012,** *8* (12), 5072.
(37) Duke, R. E.; Starovoytov, O. N.; Piquemal, J. P.; Cisneros, G. A. GEM*: A Molecular Electronic Density-Based Force Field for Molecular Dynamics Simulations. *J Chem Theory Comput* **2014,** *10* (4), 1361.
(38) Van Belle, D.; Couplet, I.; Prevost, M.; Wodak, S. J. Calculations of electrostatic properties in proteins. *Journal of Molecular Biology* **1987,** *198* (4), 721.
(39) Tran, K. N.; Tan, M.-L.; Ichiye, T. A single-site multipole model for liquid water. *J. Chem. Phys.* **2016,** *145*, 034501.
(40) Thole, B. T. Molecular Polarizabilities Calculated with a Modified Dipole Interaction. *Chem Phys* **1981,** *59* (3), 341.
(41) Liu, H.; Wang, Y.; Bowman, J. M. Quantum calculations of the IR spectrum of liquid water using ab initio and model potential and dipole moment surfaces and comparison with experiment. *The Journal of Chemical Physics* **2015,** *142* (19), 194502.
(42) Heyden, M.; Sun, J.; Funkner, S.; Mathias, G.; Forbert, H.; Havenith, M.; Marx, D. Dissecting the THz spectrum of liquid water from first principles via correlations in time and space. *Proceedings of the National Academy of Sciences of the United States of America* **2010,** *107* (27), 12068.
(43) 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* (34), 9956.
(44) Esser, A.; Belsare, S.; Marx, D.; Head-Gordon, T. Mode specific THz spectra of solvated amino acids using the AMOEBA polarizable force field. *Phys Chem Chem Phys* **2017,** *19* (7), 5579.
(45) Qi, R.; Wang, L. P.; Wang, Q. T.; Pande, V. S.; Ren, P. Y. United polarizable multipole water model for molecular mechanics simulation. *Journal of Chemical Physics* **2015,** *143* (1), 12.

(46) Ghosh, D.; Kosenkov, D.; Vanovschi, V.; Williams, C. F.; Herbert, J. M.; Gordon, M. S.; Schmidt, M. W.; Slipchenko, L. V.; Krylov, A. I. Noncovalent Interactions in Extended Systems Described by the Effective Fragment Potential Method: Theory and Application to Nucleobase Oligomers. *Journal of Physical Chemistry A* **2010,** *114* (48), 12739.
(47) Gordon, M.; Freitag, M.; Bandyopadhyay, P.; Jensen, J.; Kairys, V.; Stevens, W. The effective fragment potential method: A QM-based MM approach to modeling environmental effects in chemistry. *JOURNAL OF PHYSICAL CHEMISTRY A* **2001,** *105* (2), 293.
(48) Gordon, M.; Mullin, J.; Pruitt, S.; Roskop, L.; Slipchenko, L.; Boatz, J. Accurate Methods for Large Molecular Systems. *J Phys Chem B* **2009,** *113* (29), 9646.
(49) Gordon, M. S.; Smith, Q. A.; Xu, P.; Slipchenko, L. V. In *Annual Review of Physical Chemistry, Vol 64*; Johnson, M. A.;Martinez, T. J., Eds., 2013; Vol. 64.
(50) Gurunathan, P. K.; Acharya, A.; Ghosh, D.; Kosenkov, D.; Kaliman, I.; Shao, Y. H.; Krylov, A. I.; Slipchenko, L. V. Extension of the Effective Fragment Potential Method to Macromolecules. *J Phys Chem B* **2016,** *120* (27), 6562.
(51) Mullin, J.; Roskop, L.; Pruitt, S.; Collins, M.; Gordon, M. Systematic Fragmentation Method and the Effective Fragment Potential: An Efficient Method for Capturing Molecular Energies. *JOURNAL OF PHYSICAL CHEMISTRY A* **2009,** *113* (37), 10040.
(52) Pruitt, S. R.; Bertoni, C.; Brorsen, K. R.; Gordon, M. S. Efficient and Accurate Fragmentation Methods. *Accounts of Chemical Research* **2014,** *47* (9), 2786.
(53) Gao, J. L.; Truhlar, D. G.; Wang, Y. J.; Mazack, M. J. M.; Loffler, P.; Provorse, M. R.; Rehak, P. Explicit Polarization: A Quantum Mechanical Framework for Developing Next Generation Force Fields. *Accounts of Chemical Research* **2014,** *47* (9), 2837.
(54) Han, J.; Mazack, M. J. M.; Zhang, P.; Truhlar, D. G.; Gao, J. Quantum mechanical force field for water with explicit electronic polarization. *The Journal of Chemical Physics* **2013,** *139* (5), 054503.
(55) Xie, W.; Orozco, M.; Truhlar, D. G.; Gao, J. X-Pol Potential: An Electronic Structure-Based Force Field for Molecular Dynamics Simulation of a Solvated Protein in Water. *Journal of chemical theory and computation* **2009,** *5* (3), 459.
(56) Brorsen, K. R.; Willow, S. Y.; Xantheas, S. S.; Gordon, M. S. The Melting Temperature of Liquid Water with the Effective Fragment Potential. *The Journal of Physical Chemistry Letters* **2015,** *6* (18), 3555.
(57) Burnham, C. J.; Xantheas, S. S. Development of transferable interaction models for water. III. Reparametrization of an all-atom polarizable rigid model (TTM2-R) from first principles. *Journal of Chemical Physics* **2002,** *116* (4), 1500.
(58) Burnham, C. J.; Xantheas, S. S. Development of transferable interaction models for water. I. Prominent features of the water dimer potential energy surface. *Journal of Chemical Physics* **2002,** *116* (4), 1479.
(59) Xantheas, S. S.; Burnham, C. J.; Harrison, R. J. Development of transferable interaction models for water. II. Accurate energetics of the first few water clusters from first principles. *Journal of Chemical Physics* **2002,** *116* (4), 1493.
(60) Burnham, C. J.; Anick, D. J.; Mankoo, P. K.; Reiter, G. F. The vibrational proton potential in bulk liquid water and ice. *Journal of Chemical Physics* **2008,** *128* (15).
(61) Fanourgakis, G. S.; Xantheas, S. S. Development of transferable interaction potentials for water. V. Extension of the flexible, polarizable, Thole-type model potential (TTM3-F, v. 3.0) to describe the vibrational spectra of water clusters and liquid water. *Journal of Chemical Physics* **2008,** *128* (7).
(62) Partridge, H.; Schwenke, D. W. The determination of an accurate isotope dependent potential energy surface for water from extensive ab initio calculations and experimental data. *Journal of Chemical Physics* **1997,** *106* (11), 4618.
(63) Bukowski, R.; Szalewicz, K.; Groenenboom, G. C.; van der Avoird, A. Predictions of the properties of water from first principles. *Science* **2007,** *315* (5816), 1249.

(64) Bukowski, R.; Szalewicz, K.; Groenenboom, G. C.; van der Avoird, A. Polarizable interaction potential for water from coupled cluster calculations. I. Analysis of dimer potential energy surface. *Journal of Chemical Physics* **2008,** *128* (9).
(65) Bukowski, R.; Szalewicz, K.; Groenenboom, G. C.; van der Avoird, A. Polarizable interaction potential for water from coupled cluster calculations. II. Applications to dimer spectra, virial coefficients, and simulations of liquid water. *Journal of Chemical Physics* **2008,** *128* (9).
(66) Cencek, W.; Szalewicz, K.; Leforestier, C.; van Harrevelt, R.; van der Avoird, A. An accurate analytic representation of the water pair potential. *Phys Chem Chem Phys* **2008,** *10* (32), 4716.
(67) Gora, U.; Cencek, W.; Podeszwa, R.; van der Avoird, A.; Szalewicz, K. Predictions for water clusters from a first-principles two- and three-body force field. *Journal of Chemical Physics* **2014,** *140* (19).
(68) Huang, X. C.; Braams, B. J.; Bowman, J. M. Ab initio potential energy and dipole moment surfaces of (H2O)(2). *Journal of Physical Chemistry A* **2006,** *110* (2), 445.
(69) Wang, Y. M.; Bowman, J. M. Towards an ab initio flexible potential for water, and post-harmonic quantum vibrational analysis of water clusters. *Chem Phys Lett* **2010,** *491* (1-3), 1.
(70) Wang, Y. M.; Bowman, J. M. Ab initio potential and dipole moment surfaces for water. II. Local-monomer calculations of the infrared spectra of water clusters. *Journal of Chemical Physics* **2011,** *134* (15).
(71) Wang, Y. M.; Shepler, B. C.; Braams, B. J.; Bowman, J. M. Full-dimensional, ab initio potential energy and dipole moment surfaces for water. *Journal of Chemical Physics* **2009,** *131* (5).
(72) Babin, V.; Medders, G. R.; Paesani, F. Toward a Universal Water Model: First Principles Simulations from the Dimer to the Liquid Phase. *J Phys Chem Lett* **2012,** *3* (24), 3765.
(73) Babin, V.; Leforestier, C.; Paesani, F. Development of a "First Principles" Water Potential with Flexible Monomers: Dimer Potential Energy Surface, VRT Spectrum, and Second Virial Coefficient. *Journal of Chemical Theory and Computation* **2013,** *9* (12), 5395.
(74) Babin, V.; Medders, G. R.; Paesani, F. Development of a "First Principles" Water Potential with Flexible Monomers. II: Trimer Potential Energy Surface, Third Virial Coefficient, and Small Clusters. *Journal of Chemical Theory and Computation* **2014,** *10* (4), 1599.
(75) Medders, G. R.; Babin, V.; Paesani, F. Development of a "First-Principles" Water Potential with Flexible Monomers. III. Liquid Phase Properties. *Journal of Chemical Theory and Computation* **2014,** *10* (8), 2906.
(76) Reddy, S. K.; Straight, S. C.; Bajaj, P.; Pham, C. H.; Riera, M.; Moberg, D. R.; Morales, M. A.; Knight, C.; Gotz, A. W.; Paesani, F. On the accuracy of the MB-pol many-body potential for water: Interaction energies, vibrational frequencies, and classical thermodynamic and dynamical properties from clusters to liquid water and ice. *Journal of Chemical Physics* **2016,** *145* (19).
(77) Bajaj, P.; Gotz, A. W.; Paesani, F. Toward Chemical Accuracy in the Description of Ion-Water Interactions through Many-Body Representations. I. Halide-Water Dimer Potential Energy Surfaces. *Journal of Chemical Theory and Computation* **2016,** *12* (6), 2698.
(78) Millot, C.; Soetens, J. C.; Costa, M. T. C. M.; Hodges, M. P.; Stone, A. J. Revised anisotropic site potentials for the water dimer and calculated properties. *Journal of Physical Chemistry A* **1998,** *102* (4), 754.
(79) Millot, C.; Stone, A. J. Towards an Accurate Intermolecular Potential for Water. *Mol Phys* **1992,** *77* (3), 439.
(80) Engkvist, O.; Forsberg, N.; Schutz, M.; Karlstrom, G. A comparison between the NEMO intermolecular water potential and ab initio quantum chemical calculations for the water trimer and tetramer. *Mol Phys* **1997,** *90* (2), 277.
(81) Holt, A.; Bostrom, J.; Karlstrom, G.; Lindh, R. A NEMO Potential that Includes the Dipole-Quadrupole and Quadrupole-Quadrupole Polarizability. *J Comput Chem* **2010,** *31* (8), 1583.
(82) Holt, A.; Karlstrom, G. Inclusion of the quadrupole moment when describing polarization. The effect of the dipole-quadrupole polarizability (vol 29, pg 2033, 2008). *J Comput Chem* **2008,** *29* (14), 2485.

(83) Gresh, N. Model, multiply hydrogen-bonded water oligomers (N = 3-20). How closely can a separable, ab initio-grounded molecular mechanics procedure reproduce the results of supermolecule quantum chemical computations? *Journal of Physical Chemistry A* **1997,** *101* (46), 8680.
(84) Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J. P. Anisotropic, polarizable molecular mechanics studies of inter- and intramoecular interactions and ligand-macromolecule complexes. A bottom-up strategy. *Journal of Chemical Theory and Computation* **2007,** *3* (6), 1960.
(85) 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* (14), 2949.
(86) Cisneros, G. A.; Elking, D.; Piquemal, J. P.; Darden, T. A. Numerical fitting of molecular properties to Hermite Gaussians. *Journal of Physical Chemistry A* **2007,** *111* (47), 12049.
(87) Lifson, S.; Warshel, A. Consistent force field for calculations of conformations, vibrational spectra, and enthalpies of cycloalkane and n-alkane molecules. *J. Chem. Phys.* **1969,** *49*, 5116.
(88) Born, M.; Mayer, J. E. Zur Gittertheorie der Ionenkristalle. *Zeitschrift für Phys* **1932,** *75*, 1.
(89) Buckingham, R. A. The classical equation of state of gaseous helium,neon,and argon. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **1938,** *168*, 264.
(90) Van Vleet, M. J.; Misquitta, A. J.; Stone, A. J.; Schmidt, J. R. Beyond Born–Mayer: Improved Models for Short-Range Repulsion in ab Initio Force Fields. *Journal of Chemical Theory and Computation* **2016,** *12* (8), 3851.
(91) Kitaura, K.; Morokuma, K. NEW ENERGY DECOMPOSITION SCHEME FOR MOLECULAR-INTERACTIONS WITHIN HARTREE-FOCK APPROXIMATION. *INTERNATIONAL JOURNAL OF QUANTUM CHEMISTRY* **1976,** *10* (2), 325.
(92) Chen, W.; Gordon, M. Energy decomposition analyses for many-body interaction and applications to water complexes. *JOURNAL OF PHYSICAL CHEMISTRY* **1996,** *100* (34), 14316.
(93) Mitoraj, M.; Michalak, A.; Ziegler, T. A Combined Charge and Energy Decomposition Scheme for Bond Analysis. *JOURNAL OF CHEMICAL THEORY AND COMPUTATION* **2009,** *5* (4), 962.
(94) Horn, P. R.; Head-Gordon, M. Alternative definitions of the frozen energy in energy decomposition analysis of density functional theory calculations. *Journal of Chemical Physics* **2016,** *144* (8).
(95) Jeziorski, B.; Moszynski, R.; Szalewicz, K. Perturbation-Theory Approach to Intermolecular Potential-Energy Surfaces of Van-Der-Waals Complexes. *Chemical Reviews* **1994,** *94* (7), 1887.
(96) Misquitta, A.; Podeszwa, R.; Jeziorski, B.; Szalewicz, K. Intermolecular potentials based on symmetry-adapted perturbation theory with dispersion energies from time-dependent density-functional calculations. *JOURNAL OF CHEMICAL PHYSICS* **2005,** *123* (21), 214103.
(97) Khaliullin, R. Z.; Cobar, E. A.; Lochan, R. C.; Bell, A. T.; Head-Gordon, M. Unravelling the origin of intermolecular interactions using absolutely localized molecular orbitals. *Journal of Physical Chemistry A* **2007,** *111* (36), 8753.
(98) Reed, A.; Curtiss, L.; Weinhold, F. INTERMOLECULAR INTERACTIONS FROM A NATURAL BOND ORBITAL, DONOR-ACCEPTOR VIEWPOINT. *CHEMICAL REVIEWS* **1988,** *88* (6), 899.
(99) Glendening, E.; Landis, C.; Weinhold, F. Natural bond orbital methods. *WILEY INTERDISCIPLINARY REVIEWS-COMPUTATIONAL MOLECULAR SCIENCE* **2012,** *2* (1), 1.
(100) Stone, A. J. *The Theory of Intermolecular Forces*; Clarendon Press: Oxford,UK, 1997.
(101) Tafipolsky, M.; Ansorg, K. Toward a Physically Motivated Force Field: Hydrogen Bond Directionality from a Symmetry-Adapted Perturbation Theory Perspective. *Journal of Chemical Theory and Computation* **2016,** *12* (3), 1267.
(102) McDaniel, J. G.; Schmidt, J. R. In *Annual Review of Physical Chemistry, Vol 67*; Johnson, M. A.;Martinez, T. J., Eds., 2016; Vol. 67.
(103) Schmidt, J. R.; Yu, K.; McDaniel, J. G. Transferable Next-Generation Force Fields from Simple Liquids to Complex Materials. *Accounts of Chemical Research* **2015,** *48* (3), 548.
(104) McDaniel, J. G.; Schmidt, J. R. First-Principles Many-Body Force Fields from the Gas Phase to Liquid: A "Universal" Approach. *J Phys Chem B* **2014,** *118* (28), 8042.

(105) McDaniel, J. G.; Schmidt, J. R. Physically-Motivated Force Fields from Symmetry-Adapted Perturbation Theory. *Journal of Physical Chemistry A* **2013,** *117* (10), 2053.
(106) Yu, K.; Schmidt, J. R. Many-body effects are essential in a physically motivated CO2 force field. *Journal of Chemical Physics* **2012,** *136* (3).
(107) McDaniel, J. G.; Schmidt, J. R. Robust, Transferable, and Physically Motivated Force Fields for Gas Adsorption in Functionalized Zeolitic Imidazolate Frameworks. *Journal of Physical Chemistry C* **2012,** *116* (26), 14031.
(108) 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* (1), 15.
(109) Bagus, P. S.; Hermann, K.; Bauschlicher, C. W. A New Analysis of Charge-Transfer and Polarization for Ligand-Metal Bonding - Model Studies of Al4co and Al4nh3. *Journal of Chemical Physics* **1984,** *80* (9), 4378.
(110) Bagus, P. S.; Illas, F. Decomposition of the Chemisorption Bond by Constrained Variations - Order of the Variations and Construction of the Variational Spaces. *Journal of Chemical Physics* **1992,** *96* (12), 8962.
(111) Misquitta, A. J.; Stone, A. J. Ab Initio Atom-Atom Potentials Using CAMCASP: Theory and Application to Many-Body Models for the Pyridine Dimer. *Journal of Chemical Theory and Computation* **2016,** *12* (9), 4184.
(112) Mardirossian, N.; Head-Gordon, M. omega B97X-V: A 10-parameter, range-separated hybrid, generalized gradient approximation density functional with nonlocal correlation, designed by a survival-of-the-fittest strategy. *Phys Chem Chem Phys* **2014,** *16* (21), 9904.
(113) Mardirossian, N.; Head-Gordon, M. Mapping the genome of meta-generalized gradient approximation density functionals: The search for B97M-V. *Journal of Chemical Physics* **2015,** *142* (7).
(114) Mardirossian, N.; Head-Gordon, M. omega B97M-V: A combinatorially optimized, range-separated hybrid, meta-GGA density functional with VV10 nonlocal correlation. *Journal of Chemical Physics* **2016,** *144* (21).
(115) Mao, Y.; Demerdash, O.; Head-Gordon, M.; Head-Gordon, T. Assessing Ion–Water Interactions in the AMOEBA Force Field Using Energy Decomposition Analysis of Electronic Structure Calculations. *Journal of Chemical Theory and Computation* **2016,** *12* (11), 5422.
(116) Demerdash, O.; Mao, Y.; Liu, T.; Head-Gordon, M.; Head-Gordon, T. Assessing many-body contributions to intermolecular interactions of the AMOEBA force field using energy decomposition analysis of electronic structure calculations. *The Journal of Chemical Physics* **2017,** *147* (16), 161721.
(117) Qi, R.; Wang, Q.; Ren, P. General van der Waals potential for common organic molecules. *Bioorganic & Medicinal Chemistry* **2016,** *24* (20), 4911.
(118) Liu, C.; Qi, R.; Wang, Q.; Piquemal, J. P.; Ren, P. Capturing Many-Body Interactions with Classical Dipole Induction Models. *Journal of Chemical Theory and Computation* **2017,** *13* (6), 2751.
(119) Wang, Q. T.; Rackers, J. A.; He, C.; Qi, R.; Narth, C.; Lagardere, L.; Gresh, N.; Ponder, J. W.; Piquemal, J. P.; Ren, P. Y. General Model for Treating Short-Range Electrostatic Penetration in a Molecular Mechanics Force Field. *Journal of Chemical Theory and Computation* **2015,** *11* (6), 2609.
(120) Kell, G. S. Density, Thermal Expansivity, and Compressibility of Liquid Water from 0 Degrees to 150 Degreesc - Correlations and Tables for Atmospheric-Pressure and Saturation Reviewed and Expressed on 1968 Temperature Scale. *J Chem Eng Data* **1975,** *20* (1), 97.
(121) Wagner, W.; Pruss, A. The IAPWS formulation 1995 for the thermodynamic properties of ordinary water substance for general and scientific use. *Journal of Physical and Chemical Reference Data* **2002,** *31* (2), 387.
(122) Hura, G.; Sorenson, J. M.; Glaeser, R. M.; Head-Gordon, T. A high-quality x-ray scattering experiment on liquid water at ambient conditions. *Journal of Chemical Physics* **2000,** *113* (20), 9140.

(123) Soper, A. K. The radial distribution functions of water and ice from 220 to 673 K and at pressures up to 400 MPa. *Chem Phys* **2000,** *258* (2-3), 121.
(124) Skinner, L. B.; Huang, C.; Schlesinger, D.; Pettersson, L. G.; Nilsson, A.; Benmore, C. J. Benchmark oxygen-oxygen pair-distribution function of ambient water from x-ray diffraction measurements with a wide Q-range. *J Chem Phys* **2013,** *138* (7), 074506.
(125) Brookes, D. H.; Head-Gordon, T. Family of Oxygen-Oxygen Radial Distribution Functions for Water. *J Phys Chem Lett* **2015,** *6* (15), 2938.
(126) Clark, G. N. I.; Cappa, C. D.; Smith, J. D.; Saykally, R. J.; Head-Gordon, T. The structure of ambient water. *Mol Phys* **2010,** *108* (11), 1415.
(127) Maréchal, Y. The molecular structure of liquid water delivered by absorption spectroscopy in the whole IR region completed with thermodynamics data. *Journal of Molecular Structure* **2011,** *1004* (1–3), 146.
(128) Fanourgakis, G. S.; Aprà, E.; Xantheas, S. S. High-level ab initio calculations for the four low-lying families of minima of (H2O)20. I. Estimates of MP2/CBS binding energies and comparison with empirical potentials. *The Journal of Chemical Physics* **2004,** *121* (6), 2655.
(129) Faller, R.; Schmitz, H.; Biermann, O.; Muller-Plathe, F. Automatic parameterization of force fields for liquids by simplex optimization. *J Comput Chem* **1999,** *20* (10), 1009.
(130) Brommer, P.; Gahler, F. Potfit: effective potentials from ab initio data. *Modelling and Simulation in Materials Science and Engineering* **2007,** *15* (3), 295.
(131) Hulsmann, M.; Kirschner, K. N.; Kramer, A.; Heinrich, D. D.; Kramer-Fuhrmann, O.; Reith, D. In *Foundations of Molecular Modeling and Simulation*; Snurr, R. Q.;Adjiman, C. S.;Kofke, D. A., Eds.; Springer-Verlag Singapore Pte Ltd: Singapore, 2016, DOI:10.1007/978-981-10-1128-3_4 10.1007/978-981-10-1128-3_4.
(132) Wang, L. P.; McKiernan, K. A.; Gomes, J.; Beauchamp, K. A.; Head-Gordon, T.; Rice, J. E.; Swope, W. C.; Martinez, T. J.; Pande, V. S. Building a More Predictive Protein Force Field: A Systematic and Reproducible Route to AMBER-FB15. *J Phys Chem B* **2017,** *121* (16), 4023.
(133) McKiernan, K. A.; Wang, L. P.; Pande, V. S. Training and Validation of a Liquid-Crystalline Phospholipid Bilayer Force Field. *Journal of Chemical Theory and Computation* **2016,** *12* (12), 5960.
(134) Welborn, M.; Chen, J. H.; Wang, L. P.; Van Voorhis, T. Why Many Semiempirical Molecular Orbital Theories Fail for Liquid Water and How to Fix Them. *J Comput Chem* **2015,** *36* (12), 934.
(135) Kokkila Schumacher, S. I. L.; Hohenstein, E. G.; Parrish, R. M.; Wang, L. P.; Martinez, T. J. Tensor Hypercontraction Second-Order Moller-Plesset Perturbation Theory: Grid Optimization and Reaction Energies. *Journal of Chemical Theory and Computation* **2015,** *11* (7), 3042.
(136) Bourasseau, E.; Haboudou, M.; Boutin, A.; Fuchs, A. H.; Ungerer, P. New optimization method for intermolecular potentials: Optimization of a new anisotropic united atoms potential for olefins: Prediction of equilibrium properties. *Journal of Chemical Physics* **2003,** *118* (7), 3020.
(137) Behler, J. Perspective: Machine learning potentials for atomistic simulations. *Journal of Chemical Physics* **2016,** *145* (17), 9.
(138) Kolb, B.; Marshall, P.; Zhao, B.; Jiang, B.; Guo, H. Representing Global Reactive Potential Energy Surfaces Using Gaussian Processes. *Journal of Physical Chemistry A* **2017,** *121* (13), 2552.
(139) Handley, C. M.; Hawe, G. I.; Kell, D. B.; Popelier, P. L. A. Optimal construction of a fast and accurate polarisable water potential based on multipole moments trained by machine learning. *Phys Chem Chem Phys* **2009,** *11* (30), 6365.
(140) Brookes, D. H.; Demerdash, O.; Head-Gordon, T. Machine Learning Corrections for Errors in 2- and 3-body Approximations to N-body Forces. *submitted* **2017**.
(141) Demerdash, O. N.; Head-Gordon, T. Convergence of the Many-Body Expansion for Energy and Forces for Classical Polarizable Models in the Condensed Phase. *J. Chem. Theory Comput.* **2016,** *in press*.
(142) Davie, S. J.; Di Pasquale, N.; Popelier, P. L. A. Incorporation of local structure into kriging models for the prediction of atomistic properties in the water decamer. *J Comput Chem* **2016,** *37* (27), 2409.

(143) Morawietz, T.; Behler, J. A Density-Functional Theory-Based Neural Network Potential for Water Clusters Including van der Waals Corrections. *Journal of Physical Chemistry A* **2013,** *117* (32), 7356.
(144) Natarajan, S. K.; Morawietz, T.; Behler, J. Representing the potential-energy surface of protonated water clusters by high-dimensional neural network potentials. *Phys Chem Chem Phys* **2015,** *17* (13), 8356.
(145) Hellstrom, M.; Behler, J. Proton-Transfer-Driven Water Exchange Mechanism in the Na+ Solvation Shell. *J Phys Chem B* **2017,** *121* (16), 4184.
(146) Yao, K.; Herr, J. E.; Parkhill, J. The many-body expansion combined with neural networks. *Journal of Chemical Physics* **2017,** *146* (1), 9.
(147) Young, D. *Iterative Solutions of Large Linear Systems.*; Academic Press: New York, 1971.
(148) Wang, W.; Skeel, R. D. Fast evaluation of polarizable forces. *Journal of Chemical Physics* **2005,** *123*, 164107.
(149) Pulay, P. Convergence Acceleration of Iterative Sequences. The Case of SCF Iteration. *Chem Phys Lett* **1980,** *73*, 393.
(150) Aviat, F.; Levitt, A.; Stamm, B.; Maday, Y.; Ren, P.; Ponder, J. W.; Lagardère, L.; Piquemal, J.-P. Truncated Conjugate Gradient: An Optimal Strategy for the Analytical Evaluation of the Many-Body Polarization Energy and Forces in Molecular Simulations. *Journal of Chemical Theory and Computation* **2017,** *13* (1), 180.
(151) Simmonett, A. C.; IV, F. C. P.; Ponder, J. W.; Brooks, B. R. An empirical extrapolation scheme for efficient treatment of induced dipoles. *The Journal of Chemical Physics* **2016,** *145* (16), 164101.
(152) Simmonett, A. C.; Pickard, F. C.; Shao, Y.; Cheatham, T. E., III; Brooks, B. R. Efficient treatment of induced dipoles. *Journal of Chemical Physics* **2015,** *143* (7).
(153) Simmonett, A. C.; Pickard, F. C. t.; Schaefer, H. F., 3rd; Brooks, B. R. An efficient algorithm for multipole energies and derivatives based on spherical harmonics and extensions to particle mesh Ewald. *J Chem Phys* **2014,** *140* (18), 184101.
(154) Niklasson, A. M. Extended Born-Oppenheimer molecular dynamics. *Phys Rev Lett* **2008,** *100* (12), 123004.
(155) Niklasson, A. M.; Cawkwell, M. J. Generalized extended Lagrangian Born-Oppenheimer molecular dynamics. *J Chem Phys* **2014,** *141* (16), 164123.
(156) Niklasson, A. M.; Steneteg, P.; Odell, A.; Bock, N.; Challacombe, M.; Tymczak, C.; Holmström, E.; Zheng, G.; Weber, V. Extended Lagrangian Born–Oppenheimer molecular dynamics with dissipation. *The Journal of chemical physics* **2009,** *130* (21), 214109.
(157) Niklasson, A. M.; Tymczak, C. J.; Challacombe, M. Time-reversible Born-Oppenheimer molecular dynamics. *Phys Rev Lett* **2006,** *97* (12), 123001.
(158) Niklasson, A. M.; Tymczak, C. J.; Challacombe, M. Time-reversible ab initio molecular dynamics. *J Chem Phys* **2007,** *126* (14), 144103.
(159) Albaugh, A.; Demerdash, O.; Head-Gordon, T. An efficient and stable hybrid extended Lagrangian/self-consistent field scheme for solving classical mutual induction. *J Chem Phys* **2015,** *143* (17), 174104.
(160) Albaugh, A.; Head-Gordon, T. A New Method for Treating Drude Polarization in Classical Molecular Simulation. *J. Chem. Theory. Comput.* **2017**, DOI:10.1021/acs.jctc.7b00838 10.1021/acs.jctc.7b00838.
(161) Albaugh, A.; Niklasson, A. M. N.; Head-Gordon, T. Accurate Classical Polarization Solution with No Self-Consistent Field Iterations. *J Phys Chem Lett* **2017,** *8* (8), 1714.
(162) Vitale, V.; Dziedzic, J.; Albaugh, A.; Niklasson, A. M.; Head-Gordon, T.; Skylaris, C. K. Performance of extended Lagrangian schemes for molecular dynamics simulations with classical polarizable force fields and density functional theory. *J Chem Phys* **2017,** *146* (12), 124115.
(163) Demerdash, O.; Yap, E. H.; Head-Gordon, T. Advanced Potential Energy Surfaces for Condensed Phase Simulation. *Annu Rev Phys Chem* **2014,** *65*, 149.