Advanced models for water simulations

Omar Demerdash,1,2 Lee-Ping Wang3 and Teresa Head-Gordon1,2,4,5*

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 discusses 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. © 2017 Wiley Periodicals, Inc.

How to cite this article:


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 water1,2 or the proton transfer reaction3,4 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 as bulk transport properties—since classical Hamiltonians implicitly incorporate these quantum mechanical effects through effective parameterization of the piecewise

*Correspondence to: thg@berkeley.edu

1Kenneth S. Pitzer Center for Theoretical Chemistry, University of California, Berkeley, CA, USA
2Department of Chemistry, University of California, Berkeley, CA, USA
3Department of Chemistry, University of California, Davis, Davis, CA, USA
4Department of Bioengineering, University of California, Berkeley, CA, USA
5Department of Chemical and Biomolecular Engineering, University of California, Berkeley, CA, USA

Conflict of interest: The authors have declared no conflicts of interest for this article.
nature of the molecular mechanics (MM) functional form as given in Eq. (1).

\[
U = U_{\text{VAL}} + U_{\text{PAULI}} + U_{\text{DISP}} + U_{\text{ELEC}} + U_{\text{CP}} + U_{\text{POL}} + U_{\text{CT}}
\]  

(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 nonbonded interactions, if they are represented at all. This is manifested by PAULI and DISP(ersion) 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(trostatic) 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 nonbonded 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 (ML), 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 the computational cost of the energy and their force terms also becomes 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 force fields that assume pairwise-additivity of the noncovalent interactions such as the early simple point charge (SPC) models by Berendsen et al.10 and the transferable intermolecular potential (TIP) models introduced by Jorgensen et al.11 Next-generation fixed-charge models included optimization of water parameters under an Ewald electrostatic embedding scheme such as TIP4P-Ew12 and followed later by the TIP4P200513 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 quantum mechanics (QM) theory and/or condensed phase data on which the 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 POL.5,15–32 POL usually receives special attention, as it decays more slowly than dispersion, exchange–repulsion, or CT with a 1/R4 dependence, so that it is the next important level of
electrostatics beyond the permanent electrostatic field. There are plethoras of POL models for water, and three main approaches have emerged to calculating POL in empirical force fields: the fluctuating charge method\textsuperscript{17,20,21}, the Drude-oscillator approach\textsuperscript{15,23,33,34}, and the well-studied induced dipole method.\textsuperscript{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 POL. 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 multipole moments and the polarizabilities is clear from the fact they are terms of a Taylor expansion of the energy in the electric field \(E\).

\[
U = E \left. \frac{\partial U}{\partial E} \right|_{E=0} + \frac{1}{2!} E \left. \frac{\partial^2 U}{\partial E^2} \right|_{E=0} + \frac{1}{3!} E^2 \left. \frac{\partial^3 U}{\partial E^3} \right|_{E=0} + \ldots
\]

\[
= \bar{\mu} \cdot E + \frac{1}{2!} \alpha E^2 + \frac{1}{3!} \beta E^3 + \ldots
\]  

(2)

where \(\bar{\mu}\) is the permanent dipole moment, \(\alpha\) is the dipole polarizability, and \(\beta\) is the dipole hyperpolarizability. It should be noted that Drude models for POL could also be used with higher order multipoles, although it is not typical in most Drude POL models for water, and many induced dipole models for water only use point charges for the permanent electrostatics.

The polarizable atomic multipole optimized energetics for biomolecular applications (AMOEBA) model is based on truncation of Eq. (2) at the second-order term 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 'POL catastrophe' whereby atomic sites at short separation distances polarize each other to infinity. AMOEBA, like many POL models, uses the Thole smeared charge distribution for damping POL\textsuperscript{40}:

\[
\rho = \frac{3a}{4\pi} \exp \left( \frac{-ar_i^3}{(\alpha_i \alpha_j)^{1/2}} \right)
\]

(3)

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

Many-body POL 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,\textsuperscript{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 CT. However, at the lower frequencies probed
by THz spectroscopy two prominent features at ~200
and ~650 cm$^{-1}$ have been identified as collective inter-
molecular vibrations and librational motions of the
hydrogen-bonded network for water, respectively. For
many years, traditional classical force fields based on
nonpolarizable force fields struggled to reproduce in
particular this intermolecular hydrogen-bonding vibra-
tional signature, illustrated using the SPC/Fw water
model shown in Figure 1. When analyzed by ab initio
molecular dynamics methods using the well character-
ized PBE functional—which was able to find agree-
ment with the far infrared feature of the experiment—
this failure of classical force fields was thought to be
attributable to lack of CT.42 However, classical water
models that include POL are certainly capable of cap-
turing this feature, also shown in Figure 1 for the
TTM3-F (discussed in more detail below), and the
iAMOEBA and AMOEBA14 models (see Box 1). The
primary point is that no dynamical quantities were
included in the parameter training set of iAMOEBA
(which only directly captures direct POL)$^{43}$ and
AMOEBA14,$^{32}$ but were reproduced nonetheless, con-
firmed by showing that the peak at 200 cm$^{-1}$ disap-
ppears altogether when POL interactions are turned off
in the simulation of the AMOEBA14 water model.
The iAMOEBA model$^{43}$ with optimized param-
eters 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 clus-
ters, 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 POL
AMOEBA03 model using the iAMOEBA data set,
resulting in the AMOEBA14 model, which yielded
overall improved agreement with experimental prop-
erties.$^{32}$ 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
POL degrees of freedom from H atoms has precedent
in the point dipole$^{39}$ and Drude model literature.$^{23}$
The uAMOEBA water model features an

**BOX 1**

**DIRECT AND MUTUAL POL 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 fre-
cuencies, a buffered 14-7 potential centered on
both oxygen and hydrogen atoms with param-
eters fitted to reproduce gas-phase and liquid-
phase properties, permanent atomic multipoles
up through quadrupoles computed via distrib-
uted multipole analysis, and atomic polarizabil-
ities that incorporates Thole damping factor, in
which water cluster binding energies were
fitted for dimer structures up through the hex-
amer. In validation studies, AMOEBA03 pro-
duced good although not consistent agreement
with experiment for thermodynamic, kinetic,
and structural properties of liquid water. The
iAMOEBA model,$^{43}$ introduced 10 years later,
revisited the optimization of AMOEBA param-
ters with two important modifications: the
direct POL approximation was introduced,
thereby omitting all interactions among
induced dipoles and removing the need for SCF
cycles, and the ForceBalance program was used
to optimize the parameters using a more exten-
sive 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.

**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,$^{55}$ in which the MM parameters are derived
from QM calculations on individual subsystems such
as monomers, dimers, and so on. X-pol relies on a
Hartree product of monomer wavefunctions calculated
using a semi-empirical single-determinant level of the-
ory, with the addition of one-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 two-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-versus-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, POL, 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 POL 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 CP, and POL is similarly damped to account for exchange–POL coupling. Exchange–repulsion is expressed with a term based on the orbital overlap between monomers and many-body charge-transfer is approximated to be a 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 $C_n$ coefficients 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 (RDFs), 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 et al.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 POL 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 two- and three-body terms.

More recently, the TTM approach of fitting two- and three-body terms to the corresponding two- and three-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, CP, and CT 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 two- and three-body effects are described collectively by two- and three-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 two- and three-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 \textit{a priori} notions of where exactly the sites should be located. Lastly, as 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 Born–Mayer POL, point-charge electrostatics, and simple two-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 two- and three-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 \textit{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 they are not assigned \textit{a priori} based on preconceived notions or chemical intuition. Therefore, the long-ranged electrostatics, POL, and dispersion may be kept simpler, because 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 \( \sim50\times \) 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 \textit{Ab Initio}-Derived Water Potentials

There exist a number of \textit{ab initio}-derived models where the parameters are prescribed \textit{a priori} as in the more familiar empirical force fields utilizing Eq. (1), including the anisotropic site-site potential (ASP),78,79 non-empirical molecular orbital (NEMO),80–82 sum of interactions between fragments \textit{ab initio} computed (SIBFA),83–85 and GEM (Gaussian electrostatic model)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 model79 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 POL through anisotropic polarizability tensors, and additionally includes quadrupolar polarizability. Dispersion is calculated using a damped \( C_n/R^n \) 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 POL with short-range attenuation to capture exchange–POL 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 POL.\textsuperscript{85} A recent extension of SIBFA is the GEM,\textsuperscript{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,\textsuperscript{90} the overlap-based prescription of exchange–repulsion has been extended and applied to the damping of the electrostatics and POL 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 \( r \).\textsuperscript{90}

The aforementioned \textit{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 POL and dispersion, GEM\textsuperscript{a}, yielding a model that unfortunately predicts an understructured oxygen–oxygen RDF and an overstructured oxygen-hydrogen RDF.\textsuperscript{37} On a positive note, GEM\textsuperscript{a} was able to correctly predict trends in the relative energetics of water hexamers.\textsuperscript{14} It is expected that once analytical gradient terms appear for all terms in the SIBFA potential, including its native model of POL, 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, (1) 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, (2) they are fit to data like total energies or electrostatic potentials that are only indirectly connected to their piecewise functions, and (3) they typically rely on but do not demonstrate how cancellation of errors occurs among the molecular interactions accounted for (exchange–repulsion, electrostatics, and POL) or that are missing (CT and CP). 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)\textsuperscript{28,91–99} affords a way to determine the relative contributions of several physically meaningful nonbonded energy terms out of the QM interaction energy, e.g., permanent electrostatics, Pauli repulsion, POL, dispersion, and so on. Although the asymptotic components of any EDA method are uniquely defined\textsuperscript{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 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)51–57 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),8 methods. More recently, a variational formulation, such as the second-generation absolutely localized molecular orbitals (ALMO) using density functional theory (DFT)10,80,104 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. Box 2 describes how ALMO-EDA was used to analyze how well the AMOEBA water model reproduces the two-body,115 as well as three-body,116 energies in the distance scans for the genuinely many-body terms of QM energetics, including modified Pauli repulsion, dispersion, POL, and CT contributions (Figure 3). As AMOEBA’s only many-body term arises from Thole-damped POL, the analysis must address not only how successfully it renders agreement with the corresponding ALMO-EDA POL, but whether the two- and three-body sum of ALMO’s modified Pauli repulsion, dispersion and CT terms are captured by three-body POL or whether it is spread ‘incoherent’ across, e.g., the two- and three-body POL contributions or accounted for in strictly two-body terms. This illustrates how future models might be tuned when EDA decomposition data are combined with sophisticated least-squared optimization methods such as ForceBalance or ML methods, which are described next.

### Box 2

**FUTURE WATER MODELS BASED ON GUIDANCE FROM EDA**

We have used ALMO-EDA to assess the quality of the noncovalent terms in the polarizable force field AMOEOBA0325 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 AMOEOBA’s total energy into the total POL energy contribution for the water trimer, and its further breakdown into two- and three-body POL, is shown in Figure 3. Compared to the high quality oB97X-V DFT benchmark the overall total intermolecular energy curve for the AMOEOBA water trimer is underbound throughout the entire range of distances. Upon further breakdown of the many-body POL into a many-body expansion, expected to converge quickly for an insulator such as water, the two-body POL shows excellent agreement with the ALMO-EDA, and AMOEOBA’s three-body POL appears to capture three-body POL explicitly and three-body CT 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 CP, and the pairwise-additive 14-7 van der Waals wall that is insufficiently softened to correct for that, with perhaps inadequate capturing of two-body CT. EDA calculations and proposed improvements to the basic AMOEOBA model are now beginning to appear in the literature.115–119

**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 (e.g., electronic POL) have an advantage in that their simulated properties are more directly comparable to the training data, and fewer ad hoc corrections are needed (e.g., the self-POL correction used in developing SPC/E has been reexamined 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\textsuperscript{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 ($\hbar \omega \gg k_B T$) degrees of freedom are not frozen out, as in the case of a quantum system.

Theoretical data sources include \textit{ab initio} calculated values of total potential energies, nuclear gradients, and interaction energies for small water clusters.\textsuperscript{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 \textit{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 \textit{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 POL is important for quantitative comparisons to \textit{ab initio} data in the gas phase; fixed-charge models rely on ad hoc schemes to approximate POL 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 \texttt{tcsb} script for simplex optimization by Faller and coworkers.\textsuperscript{129} More recently, several parameterization programs have been made available for further generality and reproducibility; these include
ForceBalance (developed by one of us),\textsuperscript{31,53} potfit by Brommer et al.,\textsuperscript{130} and Wolf(2)Pack by Hulsmann et al.\textsuperscript{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 (CT),\textsuperscript{31} iAMOEBA (direct induced dipoles),\textsuperscript{31} AMOEBA14 (mutual induced dipoles),\textsuperscript{32} uAMOEBA (single-site mutual induced dipoles),\textsuperscript{45} as well as TIP3P-FB and TIP4P-FB (fixed charge).\textsuperscript{53} In addition, ForceBalance was used to develop AMOEBA vdW parameters for organochlorine compounds,\textsuperscript{74} AMBER-style protein force field parameters,\textsuperscript{132} GROMOS-style parameters for phospholipid bilayers,\textsuperscript{133} semiempirical parameters for liquid water,\textsuperscript{134} and auxiliary grids for the tensor hypercontraction (THC) approximation of MP2.\textsuperscript{135}

The code introduces three key abstractions that help to accommodate diverse model parameterization workflows:

1. 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.

2. 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.

3. 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.

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\textsuperscript{136} provides the parametric derivatives of a general thermodynamic property $A$ as:

$$
\frac{\partial A}{\partial \lambda} \approx \frac{\partial A}{\partial \lambda} \frac{1}{k_B T} \left( A \frac{\partial E}{\partial \lambda} - A \frac{\partial A}{\partial \lambda} \right)
$$

where $\lambda$ is the model parameter, $\langle \cdot \rangle$ the ensemble average using the current value of $\lambda$, and $E$ the potential energy. Because $A$ and $E$ can be evaluated individually for trajectory frames in the simulation, the quantities on the right-hand side may be evaluated in a post-processing step by making small changes in $\lambda$ 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, and dielectric constant) over a wide temperature and pressure range. These data, in combination with a large and multifaceted ab initio data set, can be accurately fitted using ForceBalance and the AMOEBA functional form. Looking toward to the future, we will incorporate EDA into ForceBalance, which we expect will lead to models with improved accuracy and transferability.

**ML 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 data set. Driven by the burgeoning availability of large data sets and increased computational capabilities, ML methods have significantly improved over the last 10 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.\textsuperscript{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 archetypal 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

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

where $a$, $\sigma$, and $w_i$ are adjustable parameters, and the sum is 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 cannot 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, $O$:

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

where $\mu$ 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.

$$
\delta w_{i k} = \varepsilon [ O_i^\mu - y(a_i^\mu)] x_k
$$
where $\varepsilon$ 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 $x$ as: $y = \sum_i w_i \exp[-\beta |x - c_i|^2]$, where $w_i$, $\beta$, and $c_i$ 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. The central concept is a probability distribution of functions of the feature space. If we draw a random function $f(x)$ from this distribution, the probability of observing some value of the property $y$ at $x$ is a Gaussian random variable with a mean $\mu$ and variance $\sigma^2$. The central assumption is that pairs of observed values (e.g., $y_x$ and $y_z$, observed at $x$ and $z$ respectively) are correlated and decay with distance, which is reasonable if we assume the functions are smooth on a characteristic length scale $\xi_d$ (where indexes the dimensionality of the feature space). This is mathematically described as:

$$\text{Cov}[f(x), f(z)] = \exp \left[-\sum_d \xi_d |x_d - z_d|^p_d \right]$$

where both $\xi_d$ 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 $x_i$); 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 $x'$, we maximize another likelihood function of $y(x')$, given the current values of model parameters and observations in the training data set. The result is given by

$$y_{\text{max}}(x^*) = \mu + r^TR^{-1}(y - \mu)$$

where

$$r_i = \text{Cov}[f(x_i), f(x^*)]$$

and $y$ 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 (D. H. Brookes et al., unpublished data) in the context of the many-body expansion of AMOEBA known as 3-AMOEBA.

An early ANN model of the water dimer potential surface was introduced by No et al. 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. 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, as well as a number of protonated water clusters; more recently these simulations have been applied in the condensed phase to study aqueous solutions of NaOH. 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.

NEW ALGORITHMS FOR SOLVING MANY-BODY POL

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 POL solution for the point induced dipole model are solved through self-consistent field (SCF) iterative solvers, such as successive over-relaxation (SOR), preconditioned conjugate gradient (PCG), or direct inversion in the iterative subspace (DIIS) 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.

The `optimized perturbation theory’ (OPT), which cleverly uses a perturbation of the POL potential, is truncated at a tractable order and is then empirically fit to approximately recover the fully
comparisons of the standard preconditioned conjugate gradient SCF solver at 10 molecules in the NVT ensemble at 298.0 K (Reprinted with permission from Ref 160. Copyright 2017 American Chemical Society).

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 Ref 160. Copyright 2017 American Chemical Society).

FIGURE 5 | Comparisons of the standard preconditioned conjugate gradient SCF solver at 10 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 Ref 160. Copyright 2017 American Chemical Society).

Even so, a basic EL approach using thermalized hot Drude oscillators can be plagued with problems of accuracy since the effective POL 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 POL 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).158

In contrast to these SCF and EL schemes, we have adapted a time-reversible formulation of ab initio dynamics introduced by Niklasson and colleagues153 to the problem of classical POL.158–160 Our ‘inertial EL/SCF’ (iEL/SCF) method is a hybrid of an EL 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 dipoles138 as given in Eq. (10)

![Dipole hybrid expression](image)

The iEL/SCF method was shown to drop the number of SCF iterations by half for the AMOEBA polarizable model for water,158 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.161 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 the iEL/0-SCF (i.e., no self consistent field iterations)160 method. Figure 5 confirms that the properties of calculating mutual POL with iEL/0-SCF is equivalent to the quality of a tightly converged SCF solution, and is effectively as fast as using a multitime stepping method with an outer time step of 2 fs. We have recently extended the iEL/0-SCF approach to Drude POL illustrated with the PSPC polarizable water model.159 In this case, we were able to extend the standard molecular dynamics time step to 6 fs—a factor of 6× increase in time steps compared to standard EL(T, T*) approaches.

The import of this recent work on new solutions to many-body POL is as follows: it is now
possible to evaluate an important many-body effect—POL—but at a computational cost of a direct POL model,\textsuperscript{31,162} 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 POL. Because the PSPC model is a rigid model, with simple point charge permanent electrostatics, and with no Drude POL 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 affect how much statistical sampling is possible.

**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 POL, 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 MM 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 POL, exemplified by the AMOEBA model. Aside from the noted advantages of POL in allowing for transferability, we underscore the distinct ability of POL to capture IR spectroscopic features of the cooperative hydrogen-bonding network, which pairwise-additive potentials cannot recover. While the original AMOEBA model\textsuperscript{24} demonstrated notable inconsistencies in its ability to model condensed-phase properties, reparameterization efforts using the ForceBalance \textsuperscript{31,53} algorithm have generated models that show remarkable accuracy across the phase diagram, even yielding a computationally efficient POL model, iAMOEBA,\textsuperscript{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 toward 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 \textit{ab initio} data, on the other end of the spectrum are models that are parameterized entirely on \textit{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 nonnegligible. A number of force fields in which elaborate functional forms are prescribed for each of the distinct noncovalent contributions are being developed.

An interesting approach toward the formulation of \textit{ab initio}-based MM functional forms recognizes (1) that the ability of the \textit{ab initio} reference calculation to capture electron correlation is critical; and (2) that short-ranged, QM-dominated two- and three-body effects, particularly owing to CT, exchange–repulsion, and CP, 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,\textsuperscript{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, POL, 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 systems beyond pure water and simple halide-water systems. Yet another set of approaches toward capturing the complexity at short-range are the machine-learned methods that recognize that an 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 induced dipoles, is now being overcome with novel computational techniques. Very recent developments in that area include TCG methods,149 perturbative methods,150–152 and methods that render the EL approach stable and robust, reducing158,161 or avoiding SCF159,160 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


127. Maréchal Y. The molecular structure of liquid water delivered by absorption spectroscopy in the whole IR


