



Advanced models for water simulations

Omar Demerdash,^{1,2} Lee-Ping Wang³ and Teresa Head-Gordon^{1,2,4,5*}

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:

WIREs Comput Mol Sci 2017, e1355. doi: 10.1002/wcms.1355

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 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).

$$U = U_{\text{VAL}} + U_{\text{PAULI}} + U_{\text{DISP}} + U_{\text{ELEC}} + U_{\text{CP}} + U_{\text{POL}} + U_{\text{CT}} \quad (1)$$

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 water 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 TIP4P/2005¹³ 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 $1/R^4$ 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^{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 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 multipoles and the polarizabilities is clear from the fact they are terms of a Taylor expansion of the energy in the electric field \vec{E} .

$$U = \vec{E} \cdot \frac{\partial U}{\partial \vec{E}} \Big|_{\vec{E}=0} + \frac{1}{2!} \vec{E} \cdot \frac{\partial^2 U}{\partial \vec{E}^2} \Big|_{\vec{E}=0} + \frac{1}{3!} \vec{E} \cdot \frac{\partial^3 U}{\partial \vec{E}^3} \Big|_{\vec{E}=0} + \dots$$

$$= \vec{\mu} \cdot \vec{E} + \frac{1}{2!} \alpha \vec{E}^2 + \frac{1}{3!} \beta \vec{E}^3 + \dots \quad (2)$$

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 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 POL⁴⁰:

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

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 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,⁴¹ which we discuss as a more

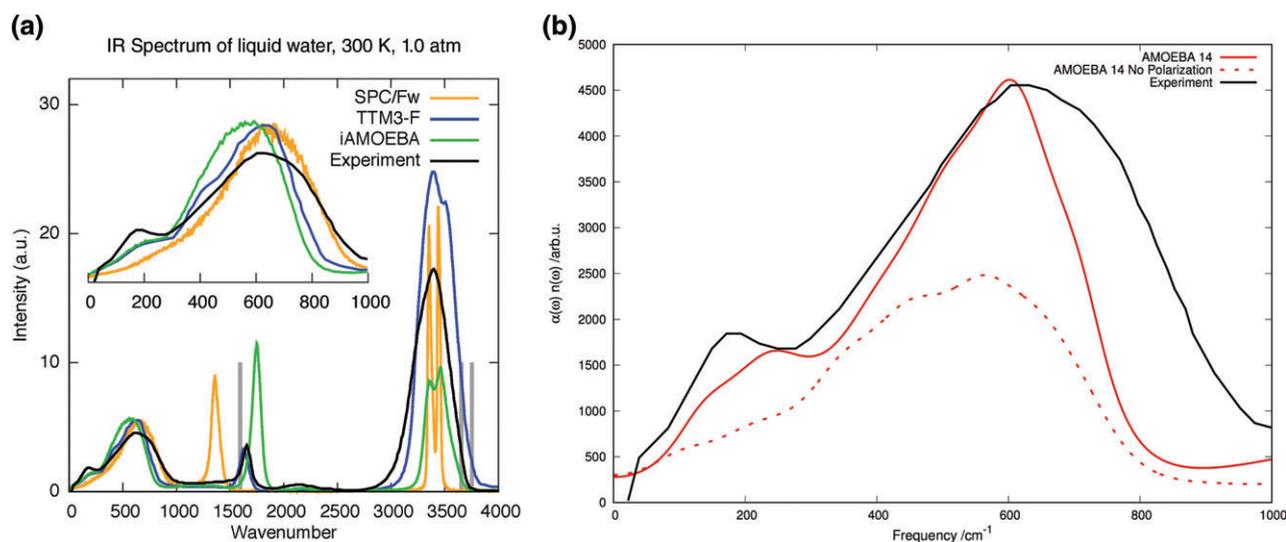


FIGURE 1 | IR spectra of liquid water from experiment (black) and compared to different classical water models (a) using the SPC/Fw, TTM3-F, and iAMOEBA models. Gray bars represent gas phase vibrational frequencies from experiment. Inset: Magnification of the far IR region (<1000 wavenumber) (Reprinted with permission from Ref 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) (Reprinted with permission from Ref 44. Copyright 2017 Royal Society of Chemistry).

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 $\sim 650\text{ cm}^{-1}$ 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 nonpolarizable 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 CT.⁴² However, classical water models that include POL 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 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, confirmed by showing that the peak at 200 cm^{-1} disappears altogether when POL interactions are turned off in the simulation of the AMOEBA14 water model.

The iAMOEBa model⁴³ 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 $\sim 7\%$ on average compared to AMOEBA03 which predicts a smaller error of $\sim 4\%$.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 properties.³² We also developed the uAMOEBa single-site polarizable water model⁴⁵ 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 dipole³⁹ 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 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 10 years later, revisited the optimization of AMOEBA parameters 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 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.

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 POL.⁴⁵

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 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 theory, 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,

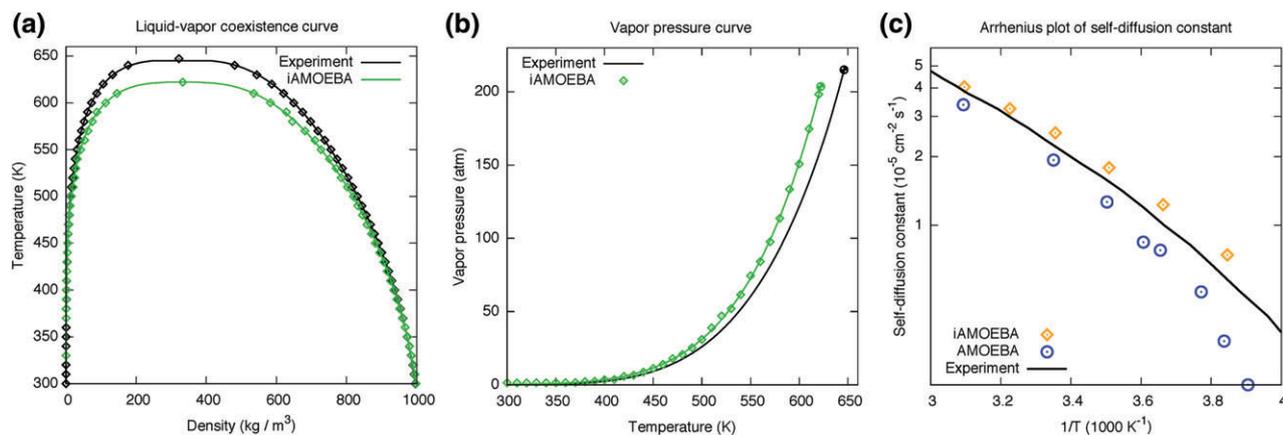


FIGURE 2 | Comparison of water properties of iAMOEBa water model against experiment. (a) Arrhenius plot of self-diffusion constant of liquid water versus temperature, which includes a comparison to AMOEBa03, (b) liquid–vapor coexistence curve, and (c) vapor pressure curve of the iAMOEBa model (Reprinted with permission from Ref 43. Copyright 2013 American Chemical Society).

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 *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 Thole-type 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 *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 *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 $\sim 50\times$ 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 *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 anisotropic site-site potential (ASP),^{78,79} non-empirical molecular orbital (NEMO),^{80–82} sum of interactions between fragments *ab initio* computed (SIBFA),^{83–85} and GEM (Gaussian electrostatic model)^{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 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_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 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).⁸⁵ A recent extension of SIBFA is the 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 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 .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 POL and dispersion, GEM*, yielding a model that unfortunately predicts an understructured oxygen–oxygen 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 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)^{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 defined¹⁰⁰ 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)^{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),^{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-body¹¹⁵ as well as three-body¹¹⁶ 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 'incoherently' 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 AMOEBA03²⁵ 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 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 ω B97X-V DFT benchmark the overall total intermolecular energy curve for the AMOEBA 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 AMOEBA'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 AMOEBA 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

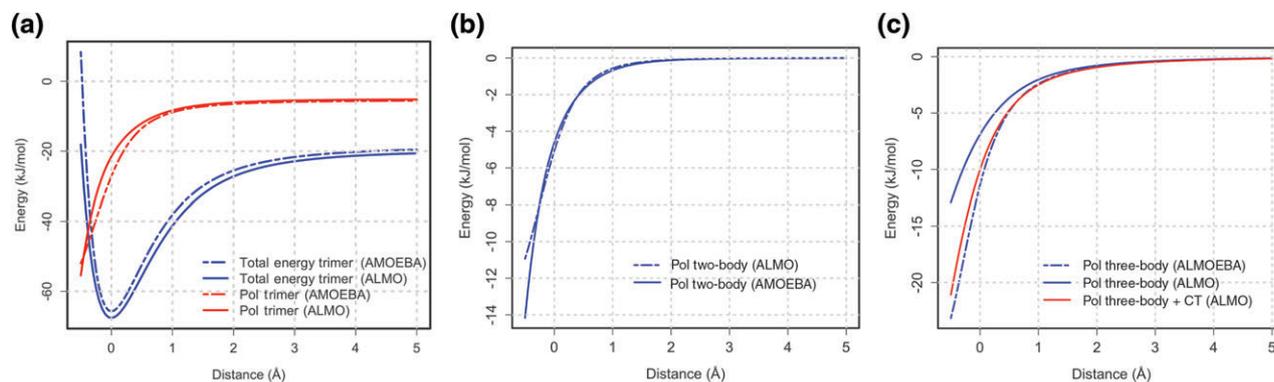


FIGURE 3 | Comparison of the ALMO-EDA decomposition of the intermolecular energy profile against AMOEBA03²⁵ for the water trimer. (a) Total energy and total polarization energy for AMOEBA against the ω B97X-V DFT benchmark and its decomposition using ALMO-EDA for polarization. (b) The two-body polarization energy for one of the three pairs in the trimer. (c) The three-body polarization as well as the sum of ALMO's three-body polarization and charge-transfer terms. The distance coordinate corresponds to displacement from equilibrium from the reference geometries (Reprinted with permission from Ref 116. Copyright 2017 AIP Publishing).

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 phase¹²; 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 *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 POL is important for quantitative comparisons to *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 *tcsb* 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 et al.,¹³⁰ and Wolf(2)Pack by Hulsmann et al.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),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:

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

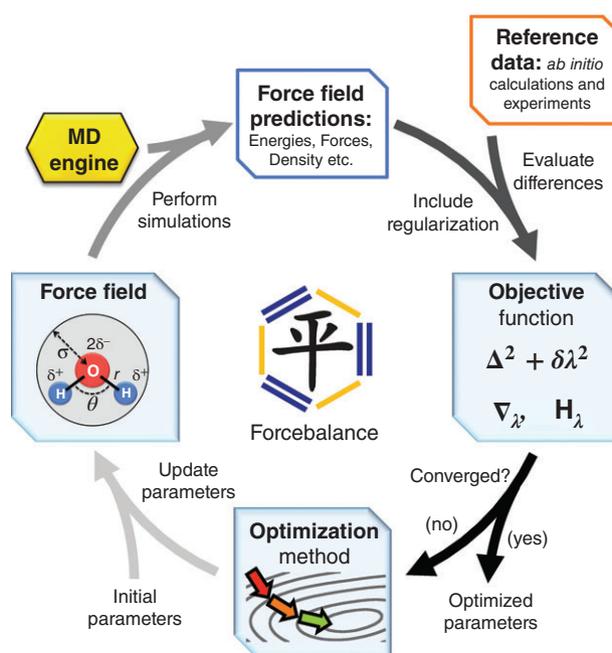


FIGURE 4 | Steps of the ForceBalance optimization cycle. The initial force field parameters (lower left) are used to perform simulations using molecular dynamics (MD) software (upper left). The objective function is computed as a least-squares function of the differences between simulation results and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

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 formula¹³⁶ provides the parametric derivatives of a general thermodynamic property A as:

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

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 right-hand side 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, 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 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.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

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

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 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 :

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

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.

$$\delta w_{ik} = \epsilon [O_i^{\mu} - y(a_i^{\mu})] x_k \quad (7)$$

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 x as: $y = \sum_i w_i \exp[-\beta_i x - c_i^2]$, where w_i , β_i 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.^{138,139} 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 μ and variance σ^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 ξ_d (d indexes the dimensionality of the feature space). This is mathematically described as:

$$\text{Cov}[f(x), f(z)] = \exp\left[-\sum_d \xi_d |x_d - z_d|^{p_d}\right] \quad (8)$$

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 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

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

where

$$r_i = \text{Cov}[f(x_i), f(x^*)], R_{ij} = \text{Cov}[f(x_i), f(x_j)] \quad (9b)$$

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.¹⁴⁰

An early ANN model of the water dimer potential surface was introduced by No et al.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.^{139,141} 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,¹⁴² as well as a number of protonated water clusters¹⁴³; more recently these simulations have been applied in the condensed phase to study aqueous solutions of NaOH.¹⁴⁴ 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.¹⁴⁵

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),¹⁴⁶ preconditioned conjugate gradient (PCG),¹⁴⁷ or direct inversion in the iterative subspace (DIIS)¹⁴⁸ 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.¹⁴⁹ 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

converged result,^{150–152} all of which have been tested on bulk water systems.

By contrast, Drude and fluctuating charge models for POL are typically solved through an extended Lagrangian (EL) formulation to treat POL 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 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).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^{153–157} 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 dipoles¹⁵⁸ as given in Eq. (10)

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

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

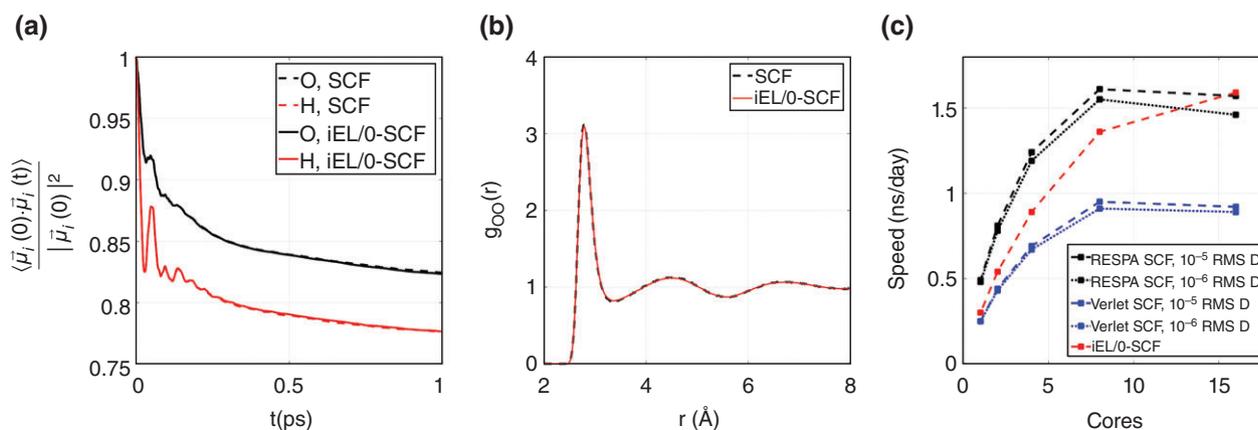


FIGURE 5 | Comparisons of the standard preconditioned conjugate gradient SCF solver at 10⁻⁶ RMSD convergence and the SCF-less method for AMOEBA water. (a) Time autocorrelation function of the induced dipoles for oxygen and hydrogen; (b) Oxygen–oxygen radial distribution function; (c) simulation speed-up 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 Ref 160. Copyright 2017 American Chemical Society).

possible to evaluate an important many-body effect—POL—but at a computational cost of a direct POL model,^{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 model²⁴ 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 POL 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 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 *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 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 *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 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,^{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 system 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, reducing^{158,161} or avoiding SCF^{159,160} 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

1. Geissler PL, Dellago C, Chandler D, Hutter J, Parrinello M. Autoionization in liquid water. *Science* 2001, 291:2121–2124.
2. Marx D. Proton transfer 200 years after von Grothuss: insights from ab initio simulations. *ChemPhysChem* 2006, 7:1848–1870.
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. *J Chem Phys* 1995, 103:150–161.
4. Schmitt UW, Voth GA. The computer simulation of proton transport in water. *J Chem Phys* 1999, 111:9361–9381.
5. Paesani F, Iuchi S, Voth GA. Quantum Effects in liquid water from an ab initio-based polarizable force field. *J Chem Phys* 2007, 127:074506.
6. Habershon S, Markland TE, Manolopoulos DE. Competing quantum effects in the dynamics of a flexible water model. *J Chem Phys* 2009, 131:024501.
7. Ceriotti M, Fang W, Kusalik PG, McKenzie RH, Michaelides A, Morales MA, Markland TE. Nuclear quantum effects in water and aqueous systems: experiment, theory, and current challenges. *Chem Rev* 2016, 116:7529–7550.
8. Marsalek O, Markland TE. Ab initio molecular dynamics with nuclear quantum effects at classical cost: ring polymer contraction for density functional theory. *J Chem Phys* 2016, 144:054112.
9. Halgren TA. Representation of Vanderwaals (Vdw) interactions in molecular mechanics force-fields—potential form, combination rules, and Vdw parameters. *J Am Chem Soc* 1992, 114:7827–7843.
10. van der Spoel D, van Maaren PJ, Berendsen HJCA. Systematic study of water models for molecular simulation: derivation of water models optimized for use with a reaction field. *J Chem Phys* 1998, 108:10220–10230.
11. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 1983, 79:926–935.
12. Horn HW, Swope WC, Pitner JW, Madura JD, Dick TJ, Hura GL, Head-Gordon T. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J Chem Phys* 2004, 120:9665–9678.
13. Vega C, Abascal JLF, Sanz E, MacDowell LG, McBride C. Can simple models describe the phase diagram of water? *J Phys-Condens Mat* 2005, 17: S3283–S3288.
14. Cisneros GA, Wikfeldt KT, Ojamae L, Lu JB, Xu Y, Torabifard H, Bartok AP, Csanyi G, Molinero V, Paesani F. Modeling molecular interactions in water: from pairwise to many body potential energy functions. *Chem Rev* 2016, 116:7501–7528.
15. Sprik M, Klein MA. Polarizable model for water using distributed charge sites. *J Chem Phys* 1988, 89:7556–7560.
16. Warshel A, Kuwajima S. Incorporating electric Polarizabilities in water-water interaction potentials. *J Phys Chem* 1990, 94:460.
17. Rick S, Stuart S, Berne B. Dynamical fluctuating charge force-fields: application to liquid water. *J Chem Phys* 1994, 101:6141–6156.
18. Gao JL. Toward a molecular orbital derived empirical potential for liquid simulations. *J Phys Chem B* 1997, 101:657–663.
19. Chialvo AA, Cummings PT. Molecular-based modeling of water and aqueous solutions at supercritical conditions. *Adv Chem Phys* 1999, 109:115.

20. Stern HA, Kaminski GA, Banks JL, Zhou RH, Berne BJ, Friesner RA. Fluctuating charge, polarizable dipole, and combined models: parameterization from ab initio quantum chemistry. *J Phys Chem B* 1999, 103:4730–4737.
21. Chen B, Xing J, Siepmann JI. Development of polarizable water force fields for phase equilibrium calculations. *J Phys Chem B* 2000, 104:2391–2401.
22. Burnham CJ, Xantheas SS. 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. *J Chem Phys* 2002, 116:5115.
23. Lamoureux G, MacKerell AD, Roux BA. Simple polarizable model of water based on classical Drude oscillators. *J Chem Phys* 2003, 119:5185–5197.
24. Ren PY, Ponder JW. Polarizable atomic multipole water model for molecular mechanics simulation. *J Phys Chem B* 2003, 107:5933–5947.
25. Ren P, Ponder JW. Temperature and pressure dependence of the AMOEBA water model. *J Phys Chem B* 2004, 108:13427–13437.
26. Yu HB, Hansson T, van Gunsteren WF. Development of a simple, self-consistent polarizable model for liquid water. *J Chem Phys* 2003, 118:221–234.
27. Benjamin KM, Schultz AJ, Kofke DA. Virial coefficients of polarizable water applications to thermodynamic properties and molecular clustering. *J Phys Chem C* 2007, 111:16021–16027.
28. Bauer BA, Patel S. Properties of water along the liquid-vapor coexistence curve via molecular dynamics simulations using the polarizable TIP4P-QDP-LJ water model. *J Chem Phys*, 2009, 131:084709.
29. Ponder JW, Wu C, Ren P, Pande VS, Chodera JD, Schnieders MJ, Haque I, Mobley DL, Lambrecht DS, DiStasio RA Jr, et al. Current status of the AMOEBA polarizable force field. *J Phys Chem B* 2010, 114:2549–2564.
30. Lee AJ, Rick SW. The effects of charge transfer on the properties of liquid water. *J Chem Phys* 2011, 134:184507.
31. Wang LP, Chen JH, Van Voorhis T. Systematic Parametrization of polarizable force fields from quantum chemistry data. *J Chem Theory Comput* 2013, 9:452–460.
32. Laury ML, Wang LP, Pande VS, Head-Gordon T, Ponder JW. Revised parameters for the AMOEBA polarizable atomic multipole water model. *J Phys Chem B* 2015, 119:9423–9437.
33. Sprik M. Computer simulation of the dynamics of induced polarization fluctuations in water. *J Phys Chem* 1991, 95:2283–2291.
34. Lopes PE, Roux B, Mackerell AD Jr. Molecular modeling and dynamics studies with explicit inclusion of electronic polarizability. Theory and applications. *Theor Chem Acc* 2009, 124:11–28.
35. Piquemal JP, Williams-Hubbard B, Fey N, Deeth RJ, Gresh N, Giessner-Prettre C. Inclusion of the ligand field contribution in a polarizable molecular mechanics: SIBFA-LF. *J Comput Chem* 2003, 24:1963–1970.
36. Cisneros GA. Application of Gaussian electrostatic model (GEM) distributed multipoles in the AMOEBA force field. *J Chem Theory Comput* 2012, 8:5072–5080.
37. Duke RE, Starovoytov ON, Piquemal JP, Cisneros GA. GEM*: a molecular electronic density-based force field for molecular dynamics simulations. *J Chem Theory Comput* 2014, 10:1361–1365.
38. Van Belle D, Couplet I, Prevost M, Wodak SJ. Calculations of electrostatic properties in proteins. *J Mol Biol* 1987, 198:721–735.
39. Tran KN, Tan M-L, Ichiye TA. Single-site multipole model for liquid water. *J Chem Phys* 2016, 145:034501.
40. Thole BT. Molecular polarizabilities calculated with a modified dipole interaction. *Chem Phys* 1981, 59:341–350.
41. Liu H, Wang Y, Bowman JM. Quantum calculations of the IR spectrum of liquid water using ab initio and model potential and dipole moment surfaces and comparison with experiment. *J Chem Phys* 2015, 142:194502.
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. *Proc Natl Acad Sci USA* 2010, 107:12068–12073.
43. Wang LP, Head-Gordon T, Ponder JW, Ren P, Chodera JD, Eastman PK, Martinez TJ, Pande VS. Systematic improvement of a classical molecular model of water. *J Phys Chem B* 2013, 117:9956–9972.
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, 19:5579–5590.
45. Qi R, Wang LP, Wang QT, Pande VS, Ren PY. United polarizable multipole water model for molecular mechanics simulation. *J Chem Phys* 2015, 143:12.
46. Ghosh D, Kosenkov D, Vanovschi V, Williams CF, Herbert JM, Gordon MS, Schmidt MW, Slipchenko LV, Krylov AI. Noncovalent interactions in extended systems described by the effective fragment potential method: theory and application to Nucleobase oligomers. *J Phys Chem A* 2010, 114:12739–12754.
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. *J Phys Chem A* 2001, 105:293–307.
48. Gordon M, Mullin J, Pruitt S, Roskop L, Slipchenko L, Boatz J. Accurate methods for large molecular systems. *J Phys Chem B* 2009, 113:9646–9663.
 49. Gordon MS, Smith QA, Xu P, Slipchenko LV. Accurate first principles model potentials for intermolecular interactions. *Annu Rev Phys Chem* 2013, 64:553–578.
 50. Gurunathan PK, Acharya A, Ghosh D, Kosenkov D, Kaliman I, Shao YH, Krylov AI, Slipchenko LV. Extension of the effective fragment potential method to macromolecules. *J Phys Chem B* 2016, 120:6562–6574.
 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. *J Phys Chem A* 2009, 113:10040–10049.
 52. Pruitt SR, Bertoni C, Brorsen KR, Gordon MS. Efficient and accurate fragmentation methods. *Acc Chem Res* 2014, 47:2786–2794.
 53. Gao JL, Truhlar DG, Wang YJ, Mazack MJM, Loffler P, Provorse MR, Rehak P. Explicit polarization: a quantum mechanical framework for developing next generation force fields. *Acc Chem Res* 2014, 47:2837–2845.
 54. Han J, Mazack MJM, Zhang P, Truhlar DG, Gao J. Quantum mechanical force field for water with explicit electronic polarization. *J Chem Phys* 2013, 139:054503.
 55. Xie W, Orozco M, Truhlar DG, Gao J. X-pol potential: an electronic structure-based force field for molecular dynamics simulation of a solvated protein in water. *J Chem Theory Comput* 2009, 5:459–467.
 56. Brorsen KR, Willow SY, Xantheas SS, Gordon MS. The melting temperature of liquid water with the effective fragment potential. *J Phys Chem Lett* 2015, 6:3555–3559.
 57. Burnham CJ, Xantheas SS. Development of transferable interaction models for water. III. Reparametrization of an all-atom polarizable rigid model (TTM2-R) from first principles. *J Chem Phys* 2002, 116:1500–1510.
 58. Burnham CJ, Xantheas SS. Development of transferable interaction models for water. I. Prominent features of the water dimer potential energy surface. *J Chem Phys* 2002, 116:1479–1492.
 59. Xantheas SS, Burnham CJ, Harrison RJ. Development of transferable interaction models for water. II. Accurate energetics of the first few water clusters from first principles. *J Chem Phys* 2002, 116:1493–1499.
 60. Burnham CJ, Anick DJ, Mankoo PK, Reiter GF. The vibrational proton potential in bulk liquid water and ice. *J Chem Phys* 2008, 128:154519.
 61. Fanourgakis GS, Xantheas SS. 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. *J Chem Phys* 2008, 128:074506.
 62. Partridge H, Schwenke DW. The determination of an accurate isotope dependent potential energy surface for water from extensive ab initio calculations and experimental data. *J Chem Phys* 1997, 106:4618–4639.
 63. Bukowski R, Szalewicz K, Groenenboom GC, van der Avoird A. Predictions of the properties of water from first principles. *Science* 2007, 315:1249–1252.
 64. Bukowski R, Szalewicz K, Groenenboom GC, van der Avoird A. Polarizable interaction potential for water from coupled cluster calculations. I. Analysis of dimer potential energy surface. *J Chem Phys* 2008, 128:094314.
 65. Bukowski R, Szalewicz K, Groenenboom GC, 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. *J Chem Phys* 2008, 128:094314.
 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, 10:4716–4731.
 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. *J Chem Phys* 2014, 140:194101.
 68. Huang XC, Braams BJ, Bowman JM. Ab initio potential energy and dipole moment surfaces of (H₂O)₂. *J Chem Phys A* 2006, 110:445–451.
 69. Wang YM, Bowman JM. Towards an ab initio flexible potential for water, and post-harmonic quantum vibrational analysis of water clusters. *Chem Phys Lett* 2010, 491:1–10.
 70. Wang YM, Bowman JM. Ab initio potential and dipole moment surfaces for water. II. Local-monomer calculations of the infrared spectra of water clusters. *J Chem Phys* 2011, 134:244313.
 71. Wang YM, Shepler BC, Braams BJ, Bowman JM. Full-dimensional, ab initio potential energy and dipole moment surfaces for water. *J Chem Phys* 2009, 131:054511.
 72. Babin V, Medders GR, Paesani F. Toward a universal water model: first principles simulations from the dimer to the liquid phase. *J Phys Chem Lett* 2012, 3:3765–3769.

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. *J Chem Theory Comput* 2013, 9:5395–5403.
74. Babin V, Medders GR, Paesani F. Development of a "first principles" water potential with flexible monomers. II: Trimer potential energy surface, third virial coefficient, and small clusters. *J Chem Theory Comput* 2014, 10:1599–1607.
75. Medders GR, Babin V, Paesani F. Development of a "first-principles" water potential with flexible monomers. III. Liquid phase properties. *J Chem Theory Comput* 2014, 10:2906–2910.
76. Reddy SK, Straight SC, Bajaj P, Pham CH, Riera M, Moberg DR, Morales MA, Knight C, Gotz AW, 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. *J Chem Phys* 2016, 145:194504.
77. Bajaj P, Gotz AW, Paesani F. Toward chemical accuracy in the description of ion-water interactions through many-body representations. I. Halide-water dimer potential energy surfaces. *J Chem Theory Comput* 2016, 12:2698–2705.
78. Millot C, Soetens JC, Costa MTCM, Hodges MP, Stone AJ. Revised anisotropic site potentials for the water dimer and calculated properties. *J Chem Phys A* 1998, 102:754–770.
79. Millot C, Stone AJ. Towards an accurate intermolecular potential for water. *Mol Phys* 1992, 77:439–462.
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, 90:277.
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, 31:1583–1591.
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, 29:2485–2486.
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? *J Chem Phys A* 1997, 101:8680–8694.
84. Gresh N, Cisneros GA, Darden TA, Piquemal JP. Anisotropic, polarizable molecular mechanics studies of inter- and intramolecular interactions and ligand-macromolecule complexes. A bottom-up strategy. *J Chem Theory Comput* 2007, 3:1960.
85. Chaudret R, Gresh N, Parisel O, Piquemal JP. Many-body exchange-repulsion in polarizable molecular mechanics. I. Orbital-based approximations and applications to hydrated metal Cation complexes. *J Comput Chem* 2011, 32:2949–2957.
86. Cisneros GA, Elking D, Piquemal JP, Darden TA. Numerical fitting of molecular properties to Hermite Gaussians. *J Chem Phys A* 2007, 111:12049–12056.
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 JE. Zur Gittertheorie der Ionenkristalle. *Z Phys* 1932, 75:1–18.
89. Buckingham RA. The classical equation of state of gaseous helium, neon, and argon. *Proc R Soc A Math Phys Eng Sci* 1938, 168:264–283.
90. Van Vleet MJ, Misquitta AJ, Stone AJ, Schmidt JR. Beyond Born–Mayer: improved models for short-range repulsion in ab initio force fields. *J Chem Theory Comput* 2016, 12:3851–3870.
91. Kitaura K, Morokuma K. New energy decomposition scheme for molecular-interactions within Hartree-Fock approximation. *Int J Quantum Chem* 1976, 10:325–340.
92. Chen W, Gordon M. Energy decomposition analyses for many-body interaction and applications to water complexes. *J Chem Phys* 1996, 100:14316–14328.
93. Mitoraj M, Michalak A, Ziegler TA. Combined charge and energy decomposition scheme for bond analysis. *J Chem Theory Comput* 2009, 5:962–975.
94. Horn PR, Head-Gordon M. Alternative definitions of the frozen energy in energy decomposition analysis of density functional theory calculations. *J Chem Phys* 2016, 144:084118.
95. Jeziorski B, Moszynski R, Szalewicz K. Perturbation-theory approach to intermolecular potential-energy surfaces of van-der-Waals complexes. *Chem Rev* 1994, 94:1887.
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. *J Chem Phys* 2005, 123:214103.
97. Khaliullin RZ, Cobar EA, Lochan RC, Bell AT, Head-Gordon M. Unravelling the origin of intermolecular interactions using absolutely localized molecular orbitals. *J Chem Phys A* 2007, 111:8753–8765.
98. Reed A, Curtiss L, Weinhold F. Intermolecular interactions from a natural bond orbital, donor-acceptor viewpoint. *Chem Rev* 1988, 88:899–926.
99. Glendening E, Landis C, Weinhold F. Natural bond orbital methods. *Wiley Interdiscip Rev Comput Mol Sci* 2012, 2:1–42.

100. Stone AJ. *The Theory of Intermolecular Forces*. Oxford: Clarendon Press; 1997.
101. Tafipolsky M, Ansorg K. Toward a physically motivated force field: hydrogen bond directionality from a symmetry-adapted perturbation theory perspective. *J Chem Theory Comput* 2016, 12:1267–1279.
102. McDaniel JG, Schmidt JR. Next-generation force fields from symmetry-adapted perturbation theory. *Annu Rev Phys Chem* 2016, 67:467–488.
103. Schmidt JR, Yu K, McDaniel JG. Transferable next-generation force fields from simple liquids to complex materials. *Acc Chem Res* 2015, 48:548–556.
104. McDaniel JG, Schmidt JR. First-principles many-body force fields from the gas phase to liquid: a "universal" approach. *J Phys Chem B* 2014, 118:8042–8053.
105. McDaniel JG, Schmidt JR. Physically-motivated force fields from symmetry-adapted perturbation theory. *J Chem Phys A* 2013, 117:2053–2066.
106. Yu K, Schmidt JR. Many-body effects are essential in a physically motivated CO₂ force field. *J Chem Phys* 2012, 136:034503.
107. McDaniel JG, Schmidt JR. Robust, transferable, and physically motivated force fields for gas adsorption in functionalized zeolitic Imidazolate frameworks. *J Phys Chem C* 2012, 116:14031–14039.
108. Stevens WJ, Fink WH. Frozen fragment reduced variational space analysis of hydrogen-bonding interactions—application to the water dimer. *Chem Phys Lett* 1987, 139:15–22.
109. Bagus PS, Hermann K, Bauschlicher CWA. New analysis of charge-transfer and polarization for ligand-metal bonding—model studies of Al₄co and Al₄nh₃. *J Chem Phys* 1984, 80:4378–4386.
110. Bagus PS, Illas F. Decomposition of the chemisorption bond by constrained variations—order of the variations and construction of the variational spaces. *J Chem Phys* 1992, 96:8962–8970.
111. Misquitta AJ, Stone AJ. Ab initio atom-atom potentials using CAMCASP: theory and application to many-body models for the pyridine dimer. *J Chem Theory Comput* 2016, 12:4184–4208.
112. Mardirossian N, Head-Gordon M. omega B97X-V: a 10-parameter, range-separated hybrid, generalized gradient approximation density functional with non-local correlation, designed by a survival-of-the-fittest strategy. *Phys Chem Chem Phys* 2014, 16:9904–9924.
113. Mardirossian N, Head-Gordon M. Mapping the genome of meta-generalized gradient approximation density functionals: the search for B97M-V. *J Chem Phys* 2015, 142:074111.
114. Mardirossian N, Head-Gordon M. omega B97M-V: a combinatorially optimized, range-separated hybrid, meta-GGA density functional with VV10 nonlocal correlation. *J Chem Phys* 2016, 144:214110.
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. *J Chem Theory Comput* 2016, 12:5422–5437.
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. *J Chem Phys* 2017, 147:161721.
117. Qi R, Wang Q, Ren P. General van der Waals potential for common organic molecules. *Bioorg Med Chem* 2016, 24:4911–4919.
118. Liu C, Qi R, Wang Q, Piquemal JP, Ren P. Capturing many-body interactions with classical dipole induction models. *J Chem Theory Comput* 2017, 13:2751–2761.
119. Wang QT, Rackers JA, He C, Qi R, Narth C, Lagardere L, Gresh N, Ponder JW, Piquemal JP, Ren PY. General model for treating short-range electrostatic penetration in a molecular mechanics force field. *J Chem Theory Comput* 2015, 11:2609–2618.
120. Kell GS. Density, thermal Expansivity, and compressibility of liquid water from 0 degrees to 150 degrees—correlations and tables for atmospheric-pressure and saturation reviewed and expressed on 1968 temperature scale. *J Chem Eng Data* 1975, 20:97–105.
121. Wagner W, Pruss A. The IAPWS formulation 1995 for the thermodynamic properties of ordinary water substance for general and scientific use. *J Phys Chem Ref Data Monogr* 2002, 31:387–535.
122. Hura G, Sorenson JM, Glaeser RM, Head-Gordon TA. High-quality x-ray scattering experiment on liquid water at ambient conditions. *J Chem Phys* 2000, 113:9140–9148.
123. Soper AK. The radial distribution functions of water and ice from 220 to 673 K and at pressures up to 400 MPa. *Chem Phys* 2000, 258:121–137.
124. Skinner LB, Huang C, Schlesinger D, Pettersson LG, Nilsson A, Benmore CJ. Benchmark oxygen-oxygen pair-distribution function of ambient water from x-ray diffraction measurements with a wide Q-range. *J Chem Phys* 2013, 138:074506.
125. Brookes DH, Head-Gordon T. Family of oxygen-oxygen radial distribution functions for water. *J Phys Chem Lett* 2015, 6:2938–2943.
126. Clark GNI, Cappa CD, Smith JD, Saykally RJ, Head-Gordon T. The structure of ambient water. *Mol Phys* 2010, 108:1415–1433.
127. Maréchal Y. The molecular structure of liquid water delivered by absorption spectroscopy in the whole IR

- region completed with thermodynamics data. *J Mol Struct* 2011, 1004:146–155.
128. Fanourgakis GS, Aprà E, Xantheas SS. 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. *J Chem Phys* 2004, 121:2655–2663.
129. Faller R, Schmitz H, Biermann O, Muller-Plathe F. Automatic parameterization of force fields for liquids by simplex optimization. *J Comput Chem* 1999, 20:1009–1017.
130. Brommer P, Gahler F. Potfit: effective potentials from ab initio data. *Model Simul Mater Sci Eng* 2007, 15:295–304.
131. Hulsmann M, Kirschner KN, Kramer A, Heinrich DD, Kramer-Fuhrmann O, Reith D. Optimizing molecular models through force-field parameterization via the efficient combination of modular program packages. In: Snurr RQ, Adjiman CS, Kofke DA, eds. *Foundations of Molecular Modeling and Simulation*. Singapore: Springer-Verlag Singapore Pte Ltd; 2016 https://doi.org/10.1007/978-981-10-1128-3_4.
132. Wang LP, McKiernan KA, Gomes J, Beauchamp KA, Head-Gordon T, Rice JE, Swope WC, Martinez TJ, Pande VS. Building a more predictive protein force field: a systematic and reproducible route to AMBER-FB15. *J Phys Chem B* 2017, 121:4023–4039.
133. McKiernan KA, Wang LP, Pande VS. Training and validation of a liquid-crystalline phospholipid bilayer force field. *J Chem Theory Comput* 2016, 12:5960–5967.
134. Welborn M, Chen JH, Wang LP, Van Voorhis T. Why many Semiempirical molecular orbital theories fail for liquid water and how to fix them. *J Comput Chem* 2015, 36:934–939.
135. Kokkila Schumacher SIL, Hohenstein EG, Parrish RM, Wang LP, Martinez TJ. Tensor Hypercontraction second-order Moller-Plesset perturbation theory: grid optimization and reaction energies. *J Chem Theory Comput* 2015, 11:3042–3052.
136. Bourasseau E, Haboudou M, Boutin A, Fuchs AH, Ungerer P. New optimization method for intermolecular potentials: optimization of a new anisotropic united atoms potential for olefins: prediction of equilibrium properties. *J Chem Phys* 2003, 118:3020–3034.
137. Behler J. Perspective: machine learning potentials for atomistic simulations. *J Chem Phys* 2016, 145:9.
138. Kolb B, Marshall P, Zhao B, Jiang B, Guo H. Representing global reactive potential energy surfaces using Gaussian processes. *J Chem Phys A* 2017, 121:2552–2557.
139. Handley CM, Hawe GI, Kell DB, Popelier PLA. Optimal construction of a fast and accurate polarisable water potential based on multipole moments trained by machine learning. *Phys Chem Chem Phys* 2009, 11:6365–6376.
140. Demerdash ON, 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, 12:3884–3893.
141. Davie SJ, Di Pasquale N, Popelier PLA. Incorporation of local structure into kriging models for the prediction of atomistic properties in the water decamer. *J Comput Chem* 2016, 37:2409–2422.
142. Morawietz T, Behler JA. Density-functional theory-based neural network potential for water clusters including van der Waals corrections. *J Chem Phys A* 2013, 117:7356–7366.
143. Natarajan SK, Morawietz T, Behler J. Representing the potential-energy surface of protonated water clusters by high-dimensional neural network potentials. *Phys Chem Chem Phys* 2015, 17:8356–8371.
144. Hellstrom M, Behler J. Proton-transfer-driven water exchange mechanism in the Na⁺ solvation shell. *J Phys Chem B* 2017, 121:4184–4190.
145. Yao K, Herr JE, Parkhill J. The many-body expansion combined with neural networks. *J Chem Phys* 2017, 146:9.
146. Young D. *Iterative Solutions of Large Linear Systems*. New York: Academic Press; 1971.
147. Wang W, Skeel RD. Fast evaluation of polarizable forces. *J Chem Phys* 2005, 123:164107.
148. Pulay P. Convergence acceleration of iterative sequences. The case of SCF iteration. *Chem Phys Lett* 1980, 73:393–398.
149. Aviat F, Levitt A, Stamm B, Maday Y, Ren P, Ponder JW, 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. *J Chem Theory Comput* 2017, 13:180–190.
150. Simmonett AC, IV FCP, Ponder JW, Brooks BR. An empirical extrapolation scheme for efficient treatment of induced dipoles. *J Chem Phys* 2016, 145:164101.
151. Simmonett AC, Pickard FC, Shao Y, Cheatham TE III, Brooks BR. Efficient treatment of induced dipoles. *J Chem Phys* 2015, 143:074115.
152. Simmonett AC, Pickard FCT, Schaefer HF 3rd, Brooks BR. An efficient algorithm for multipole energies and derivatives based on spherical harmonics and extensions to particle mesh Ewald. *J Chem Phys* 2014, 140:184101.
153. Niklasson AM. Extended born-Oppenheimer molecular dynamics. *Phys Rev Lett* 2008, 100:123004.
154. Niklasson AM, Cawkwell MJ. Generalized extended Lagrangian born-Oppenheimer molecular dynamics. *J Chem Phys* 2014, 141:164123.
155. Niklasson AM, Steneteg P, Odell A, Bock N, Challacombe M, Tymczak C, Holmström E,

- Zheng G, Weber V. Extended Lagrangian born–Oppenheimer molecular dynamics with dissipation. *J Chem Phys* 2009, 130:214109.
156. Niklasson AM, Tymczak CJ, Challacombe M. Time-reversible born–Oppenheimer molecular dynamics. *Phys Rev Lett* 2006, 97:123001.
157. Niklasson AM, Tymczak CJ, Challacombe M. Time-reversible ab initio molecular dynamics. *J Chem Phys* 2007, 126:144103.
158. 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, 143:174104.
159. Albaugh A, Head-Gordon T. A new method for treating Drude polarization in classical molecular simulation. *J Chem Theory Comput* 2017. <https://doi.org/10.1021/acs.jctc.7b00838>.
160. Albaugh A, Niklasson AMN, Head-Gordon T. Accurate classical polarization solution with no self-consistent field iterations. *J Phys Chem Lett* 2017, 8:1714–1723.
161. Vitale V, Dziedzic J, Albaugh A, Niklasson AM, Head-Gordon T, Skylaris CK. Performance of extended Lagrangian schemes for molecular dynamics simulations with classical polarizable force fields and density functional theory. *J Chem Phys* 2017, 146:124115.
162. Demerdash O, Yap EH, Head-Gordon T. Advanced potential energy surfaces for condensed phase simulation. *Annu Rev Phys Chem* 2014, 65:149–174.