An Advanced Force Field for Many-Body Polarization and Charge Transfer for Water and Ions

Joseph P. Heindel,[†,‡] Selim Sami,[†] and Teresa Head-Gordon[*,†,‡,¶]

†*Kenneth S. Pitzer Theory Center and Department of Chemistry, University of California, Berkeley, California 94720, United States*

‡*Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, United States*

¶*Departments of Bioengineering and Chemical and Biomolecular Engineering, University of California, Berkeley, California 94720, United States*

E-mail: thg@berkeley.edu

Abstract

Solvated atomic ions such as alkaline metals and halides play critical roles in the regulation of biological systems, environmental chemistry, and energy materials. This work analyzes the many-body energies of water and atomic ions for non-additivities in each term of an energy decomposition analysis (EDA). We then construct an advanced force field which quantitatively reproduces each term of the EDA thereby enabling accuracy and transferability. This is achieved through the introduction of new many-body models for polarization and charge transfer in the formulation of a very high quality water force field, which generalizes to provide accurate dimer surfaces and three-body polarization and charge transfer for solvated alkali metal cations and halide anions. We also utilize a new one-body potential that accounts for intramolecular polarization by

including an electric field-dependent correction to the Morse potential. This one-body potential dramatically improves the accuracy of forces as well as spectroscopic signatures for hydrogen-bonded O−H stretches.

Introduction

Force fields (FFs) are approximations to the quantum mechanical (QM) potential energy surface, in which the model design goal is to predict structure, dynamics, and thermodynamics of any molecular system. Although pairwise additive FFs remain popular due to their insight and computational efficiency, they are inherently limited by their inability to describe the inherent many-body contributions of the QM energy, thereby reducing their accuracy in property predictions and transferability to new chemistry. For example, it has long been appreciated that non-additive interactions are critical for describing hydrogen-bonded interactions in water,[1] and the many-body energy contributions for water interacting with ions is quite large, typically around 15% of the non-bonded energy, and the cooperativity effects can be either stabilizing or destabilizing depending on ion type and binding motif.[2,3] And yet many simulations involving ions or ionized molecules are carried out by treating the ion as a scaled point charge with simple functional forms for dispersion and repulsion that cannot describe such complexity.[4,5]

Recently there has been a paradigm shift in non-reactive many-body FF development by combining energy decomposition analysis (EDA)[6–8] with the many-body expansion (MBE)[9–11] to better control the accuracy and transferability of advanced FF models. The MBE of the non-bonded energy[2,3,10] allows one to gain insights into how much non-additivity is present in the QM energy (and forces), while the EDA decomposes the total energy (and forces) into physically motivated contributions from pairwise additive terms such as electrostatics with charge penetration and Pauli repulsion, and many-body contributions that arise from polarization, charge transfer, and dispersion. For example, our recent (many-body) MB-UCB force field for water[12] and extensions to monovalent and divalent alkali metal and

halide ions[13] describe ion-water interactions through a systematic buildup of 2- and 3-body interactions whose molecular interactions are formulated using the ALMO-EDA energy decomposition scheme.[14] By reproducing the EDA term-by-term, we can ensure that the force field will be transferable across the phase diagram of a homogeneous system and, ideally, to new heterogenous systems.

The EDA approach to FF development also provides chemical insights in the relative weights of particular energy terms, for instance, charge transfer versus polarization, and better separation of the Pauli repulsion and dispersion energies. While many-body potentials for water strive for higher accuracy, models such as AMOEBA+, HIPPO,[15] MB-Pol, and q-AQUA differ in their use of EDA, MBE, and most critically, their functional forms used for the many-body energy and forces. Historically, there have been two main approaches to including polarization in force fields: fluctuating charges[16] and induced dipoles.[17] There have also been attempts to unify these approaches allowing for both charge rearrangements and induced dipoles[18] in order to also capture the charge transfer interaction.[16] But with the advent of variational EDA techniques,[8,19,20] it is now clear that the charge transfer energy scales exponentially and hence cannot be modelled by fluctuating charges or artificial induced dipole models alone.

This work takes a large step forwards in the construction of many-body force fields capable of not just describing water, but handling very strong interactions such as those between water and ions, which has remained elusive and is an open research problem dating back decades. This is possible through the introduction of advanced functional forms for the many-body polarization and charge transfer contributions, and the advancement of a one-body potential that accounts for intramolecular polarization. In particular, we introduce a combined fluctuating charge (FQ) and induced dipole model of electronic polarization that couples to our model for many-body charge transfer (CT) that we will show better reproduces all terms from EDA, and allows for explicit forward and backward transfer of charge between molecules. We also eliminate the need to treat intramolecular and intermolecular

polarization separately through a recently reported one-body potential,[?] further modified by a field-dependent correction to the Morse potential, thereby dramatically improving the accuracy of electrostatic and polarization forces. The resulting FQCT (fluctuating charges and charge transfer) model shows excellent accuracy against EDA data, and additional validation data for water clusters and ion-water energies, and reproduces the structure-frequency correspondence expected for hydrogen-bonded O−H stretches.[21]

Theory

The FQCT force field described in this work will model each of the terms in Eq. 1 using ALMO-EDA to separate the total non-bonded interaction energy into individual contributions[19,20]

$$E_{\text{int}} = E_{\text{elec}} + E_{\text{Pauli}} + E_{\text{disp}} + E_{\text{pol}} + E_{\text{CT}} \tag{1}$$

where E_{elec}, E_{Pauli}, E_{disp}, E_{pol}, and E_{CT} correspond to the contributions from the permanent electrostatics, Pauli repulsion, dispersion, polarization, and charge transfer, respectively. ALMO-EDA is described elsewhere,[19,20] but we provide two important clarifications here. First the most appropriate choice for E_{elec} is the quasi-classical expression, which depends only on the geometry of individual monomers,[22] and we use the fragment electric response function approach (at the dipole plus quadrupole level) to evaluate the polarization, ensuring a well-defined basis set limit.[23] All the ALMO-EDA calculations are performed at the level of ωB97X-V DFT functional[24] and def2-QZVPPD basis set[25] using the Q-Chem software package.[26] Note that we will use a convention of referring to all energy terms in the force field with a V and all energy terms from electronic structure with an E.

Density Overlap Model

Our approach adopts ideas from the density overlap hypothesis[27–31] which states that the short-range contributions to intermolecular interactions is proportional to the electron den-

sity overlap. In order for this idea to be amenable to force fields, one must use overlaps of atom-centered densities. The form of the charge density used in our model is,

$$\rho(r) = \frac{Qb^3}{8\pi} e^{-br} + Z\delta(r) \tag{2}$$

where Q is the charge associated with the model electron density, Z is the effective nuclear charge of the atom, and b defines the width of the Slater density. The delta function, $\delta(r)$, means the core is treated as a point particle. One can show that the overlap, S_{ii}^{ρ}, of two identical Slater-like atomic densities at different locations, $\rho_i(\boldsymbol{r}_i)$ and $\rho_i(\boldsymbol{r}_j)$, is,

$$S_{ii}^{\rho} = \frac{\pi D^2}{b_{ii}^3} P(b_{ii} r_{ij}) \exp(-b_{ii} r_{ij}) \tag{3}$$

The above overlap expression is only strictly true for the exponential tail of the Slater density and for identical atoms. The overlap between atoms with different densities, S_{ρ}^{ij}, has a more complicated form, but it has been shown that setting $b_{ij} = \sqrt{b_i b_j}$ allows the expression for S_{ii}^{ρ} to be used for different atom types to a good approximation.[30] The polynomial prefactor in the overlap is,

$$P(b_{ij} r_{ij}) = \frac{1}{3}(b_{ij} r_{ij})^2 + b_{ij} r_{ij} + 1 \tag{4}$$

where, again, we will use the combination rule $b_{ij} = \sqrt{b_i b_j}$ unless otherwise noted.

The density overlap hypothesis has been advanced by Misquitta and others[32,33] based on iterated stockholder atoms which can be used to define Slater-like densities for atoms in molecules, as well as by van Vleet *et al.* in the MASTIFF force field.[30,31] Rackers *et al.* utilize a similar idea in the HIPPO model[15] but rather than relying on density overlap, they treat the Slater function as an orbital and are able to derive models of Pauli repulsion, charge penetration, and even dispersion. Because HIPPO is derived from a model orbital, the damping functions which prevent singularities in various short-range energetic contributions arise naturally. We find the HIPPO approach to both Pauli repulsion and electrostatics to be physically principled and utilize them here without significant modification. In addition,

we utilize both Slater density overlap and Slater orbital overlap in the other energy terms within the FQCT model thereby distinguishing the degree of what is "short-range" for a pair of atoms.

Electrostatics, Pauli, and Dispersion Energies

Permanent Electrostatics. Our description of electrostatics comes from a traditional point multipole approach up to the quadrupoles, and a charge penetration (CP) contribution that modifies the short-range electrostatic energy to be more attractive than the point multipole expansion alone. We isolate the CP energy by taking the total classical electrostatic energy from EDA minus the point multipole interaction energy when using Stone's distributed multipole analysis (DMA)[34,35] out to hexadecapoles on all atoms.

$$E^{CP} = E^{elec}_{EDA} - E^{elec}_{DMA} \tag{5}$$

The advantage of this approach is it allows us to ensure that our multipoles are not biased to compensate for error in the description of charge penetration, and vice versa, which is essential to reproduce the classical electrostatic energy in EDA.

CP is described by treating each atom as having both a positively charged core and negatively charged shell. Considering the interactions of the collection of cores and shells, which are expanded in multipoles, results in the following electrostatic energy expression:

$$V_{elec} = \sum_{i<j} Z_i T_{ij} Z_j + Z_i \boldsymbol{T}^{damp}_{ij} \boldsymbol{M}_j + Z_j \boldsymbol{T}^{damp}_{ji} \boldsymbol{M}_i + \boldsymbol{M}_i \boldsymbol{T}^{overlap}_{ij} \boldsymbol{M}_j \tag{6}$$

The first term in Eq. 6 represents repulsive core-core interactions where $T_{ij} = 1/r_{ij}$ with Z_i the core charge on the ith atom; note that this is not the nuclear charge but an effective nuclear charge. The second and third terms describe attractive core-shell interactions where $\boldsymbol{M}_i$ is a vector whose entries are the components of the multipoles located on that atom. The final term corresponds to the shell-shell interactions. Further details of the damping

functions are provided in the Supplementary Information.

Pauli Repulsion. The original aim of the density overlap model was to model the Pauli repulsion energy formally as an exponential repulsion at short-range.[28,36,37] However, Rackers and Ponder have made a convincing argument that the appropriate functional form for Pauli repulsion is $V^{exch} \propto e^{-b_{ij}r_{ij}}/r_{ij}$.[15,38] While the exponential is the dominant contribution, the factor of $1/r_{ij}$ becomes important at short distance and allows for the Pauli repulsion energy to be expressed as a multipole expansion. Their basic idea is that the Pauli repulsion energy between a pair of atoms is proportional to S^2/r_{ij} where S is the overlap between pseudo-orbitals, defined as $\sqrt{\rho}$ where ρ is the density in Eq. 2.

Therefore, the Pauli repulsion energy can be written as

$$V_{Pauli} = \sum_{i<j} \frac{K_{ij}^{q} S_q^2 + K_{ij}^{\mu} S_\mu^2 + K_{ij}^{\Theta} S_\Theta^2}{r_{ij}} \tag{7}$$

where S^2 contains contributions from charges, dipoles, and quadrupoles as it is found that Pauli repulsion is highly anisotropic. However, producing parameters for a complete multipole expansion tends to result in overfitting when there is not a way to derive the initial multipoles from electronic structure. Therefore, the proportionality constants $K_{ij} = K_i K_j$ are fit instead. Since S^2 takes the form of a damped multipole expansion,[38] these proportionality constants mean multipoles which handle repulsion are proportional to the actual electrostatic multipoles. Hence the calculation of electrostatics and multipolar Pauli repulsion differs only in the choice of damping function thereby reducing computational cost of the two EDA terms.

The expansion of Pauli repulsion in terms of multipoles has an interesting physical interpretation. Namely, as two electron densities begin to overlap, the electrons will be expelled from the internuclear region in order to keep the total system wavefunction antisymmetric. This results in a "hole" in the electron density where nuclei are exposed to one another. In a sense, then, these multipoles describe the magnitude and shape of the depletion of electron

density between two atoms which are near one another.

Dispersion. The dispersion energy uses a damped polynomial interaction given by,

$$V_{disp} = \sum_{i<j} f_6^{TT}(x_{ij}) \frac{C_{6,ij}}{r_{ij}^6} \tag{8}$$

where $C_{6,ij}$ is the dispersion coefficient between atoms i and j which is determined as $C_{6,ij} = \sqrt{C_{6,i} C_{6,j}}$, and $C_{6,i}$ is a parameter fit to the EDA dispersion energy. $f_6^{TT}(x_{ij})$ is the sixth-order Tang-Toennies damping function[39] which was originally derived to damp short-range dispersion,

$$f_n^{TT}(x_{ij}) = 1 - e^{-x_{ij}} \sum_{k=0}^{n} \frac{x_{ij}^k}{k!} \tag{9}$$

The appropriate form of x for the tail of a Slater electron density has been derived before[30] and takes the form,

$$x_{ij} = b_{ij} r_{ij} - \frac{2 b_{ij}^2 r_{ij}^2 + 3 b_{ij} r_{ij}}{b_{ij}^2 r_{ij}^2 + 3 b_{ij} r_{ij} + 3} \tag{10}$$

Note that the TT damping functions, Eq. 9, depends parametrically on the choice of integer n. In their original work, Tang and Toennies show that the appropriate choice of n for dispersion is $n = 6$. This makes the damping function an exponential multiplied by a sixth order polynomial. This polynomial is able to control the r^{-6} scaling of dispersion, while the exponential ensures no damping at long distances. As an aside, one could also use TT damping functions of different orders to control mutual polarization. We have tested this and it works just as well as using $f^{overlap}$.

Polarization, Charge Transfer, and the One-Body Potential

Here we highlight the unique aspects of our FQCT model through the introduction of new physics and functional forms to describe polarization, charge transfer, and the one-body potential. Polarization is handled in a manner that allows for both intramolecular charge fluctuations and induced dipoles, while also using a new approach to modelling charge transfer

which allows for explicit movement of charge between molecules. We show that this naturally describes many-body charge transfer by coupling into the polarization equations. Additionally, we offer a more quantitative way to describe intramolecular polarization through a recent design of the one-body term.

Polarization. While distributed polarization naturally contains both charge-flow and induced dipole contributions,[40] typically the charge-flow contributions are eliminated through localization.[41] Our approach allows for charge flow polarization using a modification of the electronegativity equalization model (EEM) of polarization.[42] In EEM, the energy of a molecule is expanded to second-order as a function of charge while allowing all charges to interact

$$V(\boldsymbol{q}) = \sum_i \chi_i q_i + \frac{1}{2}\sum_i \eta_i q_i^2 + \sum_{i<j} \frac{q_i q_j}{r_{ij}} \tag{11}$$

where χ_i represents the electronegativity of atom i and η_i is the atomic hardness of atom i. By requiring the electronegativity of all atoms to become become equal, new atomic charges are determined by solving a system of linear equations.

There are several known shortcomings of EEM for non-reactive FFs including allowing for long-range transfer of charge between molecules, which is unphysical,[43,44] as well as a change in charge of atoms in a molecule that interferes with the definition of the permanent electrostatics. Our solution to the first problem is to allow charge rearrangements within a molecule but not between molecules. This constraint is enforced by the method of Lagrange multipliers. For the second problem, we drop the linear term in Eq. 11 and focus only on the fluctuation of charge around the reference charge used for the permanent electrostatics. Thus we are equalizing electronegativity around an "already equalized" state, and the change in electronegativity at each atom due to an environment is simply the electric potential at that atom. We can then write the fluctuating charge (FQ) contribution in our model as,

$$V(\delta\boldsymbol{q}) = \frac{1}{2}\sum_i \eta_i \delta q_i^2 + \sum_i \delta q_i V_i + \sum_{i<j} \frac{\delta q_i \delta q_j}{r_{ij}} + \sum_\alpha \lambda_\alpha \sum_{i\in\alpha} \delta q_i \tag{12}$$

We also allow electric fields due to the environment to induce dipoles on all atoms as done previously for other polarization models. The energy of an induced dipole $\boldsymbol{\mu}_i^{ind}$ in an electric field, $\boldsymbol{E}$, including mutual polarization is,

$$V(\boldsymbol{\mu}^{ind}) = -\frac{1}{2}\sum_i \boldsymbol{\mu}_i^{ind} \cdot \boldsymbol{E}_i^{damp} + \sum_{i<j} \boldsymbol{\mu}_i^{ind} \boldsymbol{T}_{ij}^{\mu\mu} \boldsymbol{\mu}_j^{ind} \tag{13}$$

The field $\boldsymbol{E}_i^{damp}$ is the damped electric field generated by a Slater density and $\boldsymbol{T}_{ij}^{\mu\mu}$ is the damped dipole-dipole interaction tensor which is derived from appropriate gradients of $f_{ij}^{overlap}/r_{ij}$. The form of the ij entries of the multipole interaction tensors are as follows:

$$T_{ij}^{qq} = f_1^{overlap} \frac{1}{r_{ij}} \tag{14a}$$

$$\boldsymbol{T}_{ij}^{q\mu} = f_3^{overlap} \frac{-\boldsymbol{r}_{ij}}{r_{ij}^3} \tag{14b}$$

$$\boldsymbol{T}_{ij}^{\mu\mu} = \left(f_5^{overlap} \frac{\boldsymbol{r}_{ij} \otimes \boldsymbol{r}_{ij}}{r_{ij}^5} - f_3^{overlap} \frac{\mathbf{1}}{r_{ij}^3} \right) \tag{14c}$$

The interaction tensors in Eq. 14 are the usual Cartesian multipole interaction tensors, generated by successive gradients of $1/r_{ij}$ where r_{ij} is the distance between two atoms. These tensors are multiplied by the overlap damping function derived from the overlap of two slater pseudo-orbitals.[15]

Normally, the dipole polarizability is treated as a constant in polarizable force fields, but Chung *et al.* have pointed out that polarizabilities can be significantly diminished in the aqueous phase for ions.[45] This effect is not exclusive to ions, but is simply more important in the case of diffuse anions especially. We thus adopt a slightly simplified version of the scheme suggested by Chung *et al.* for making the polarizability dependent on the local environment. We damp the inverse polarizability, $\boldsymbol{\alpha}^{-1}$, as follows

$$\boldsymbol{\alpha}_i^{-1} = R_i \begin{pmatrix} \alpha_{xx,i}^{-1} & 0 & 0 \\ 0 & \alpha_{yy,i}^{-1} & 0 \\ 0 & 0 & \alpha_{zz,i}^{-1} \end{pmatrix} R_i^T + \mathbb{1} \sum_j k_{ij}^{damp} S_{ij}^{\rho} \tag{15}$$

The first term in Eq. 15 is a typical expression of the dipole polarizability in the local axis frame of that atom. $\alpha_{xx,i}$ is the xx component of the dipole polarizability with other entries defined analogously. R_i is the rotation matrix that transforms the local axis system of atom i to the global axis system. The second term defines an environment-dependent isotropic damping of the polarizability. k_{ij}^{damp} is a pair-specific parameter which modulates the increase of the inverse polarizability which is proportional the density overlap, S_{ij}^{ρ}.

Note that this modification of the polarizability describes a completely different effect from the damping of induced electrostatics. In the case of multipolar interactions, the damping arises from the fact that real charge densities have a finite width. The effect modelled in Eq. 15 is the shrinking of atoms which occurs due to antisymmetrization of the wavefunction. The effect is most important for very diffuse atoms, such as I^-, or for very close contacts such as the interaction of Li^+ with H_2O.

What now remains is to determine the values of $\delta\boldsymbol{q}$ and $\boldsymbol{\mu}^{ind}$ which minimize the total energy of the system. In order to do this, we take the derivative with respect to each δq_i and each component of each $\boldsymbol{\mu}_i^{ind}$ and set them all equal to zero. This results in a system of linear equations which can be written succinctly as follows:

$$\begin{pmatrix} \boldsymbol{T}^{qq} & \mathbf{1}_\lambda & \boldsymbol{T}^{q\mu} \\ \mathbf{1}_\lambda^\dagger & 0 & 0 \\ -\boldsymbol{T}^{\mu q} & 0 & \boldsymbol{T}^{\mu\mu} \end{pmatrix} \begin{pmatrix} \delta\boldsymbol{q} \\ \boldsymbol{\lambda} \\ \boldsymbol{\mu} \end{pmatrix} = \begin{pmatrix} -\boldsymbol{V} \\ \boldsymbol{Q} \\ \boldsymbol{E} \end{pmatrix} \tag{16}$$

where $\delta\boldsymbol{q}$ contains the optimally rearranged charges, $\boldsymbol{\lambda}$ are the Lagrange multipliers which enforce charge conservation, and $\boldsymbol{\mu}$ are the induced dipoles. The solution vector in Eq. 16 contains the electric potential, $\boldsymbol{V}$, the total charges of each molecule, $\boldsymbol{Q}$, and the electric

field on each atom $\boldsymbol{E}$. The matrix has several blocks containing the charge-charge ($\boldsymbol{T}^{qq}$), charge-dipole ($\boldsymbol{T}^{q\mu}$), dipole-charge ($\boldsymbol{T}^{\mu q}$), and dipole-dipole interaction tensors ($\boldsymbol{T}^{\mu\mu}$). Note that the diagonal elements of $\boldsymbol{T}^{qq}$ are the atomic hardness η and the 3×3 diagonal blocks of $\boldsymbol{T}^{\mu\mu}$ are the inverse polarizability tensor $\boldsymbol{\alpha}_i^{-1}$. The block $\mathbf{1}_\lambda$ has a column for each molecule in the system. An entry in that column is 1 if the ith atom is in that molecule and zero otherwise. These blocks enforce the charge-conservation constraints for each molecule.

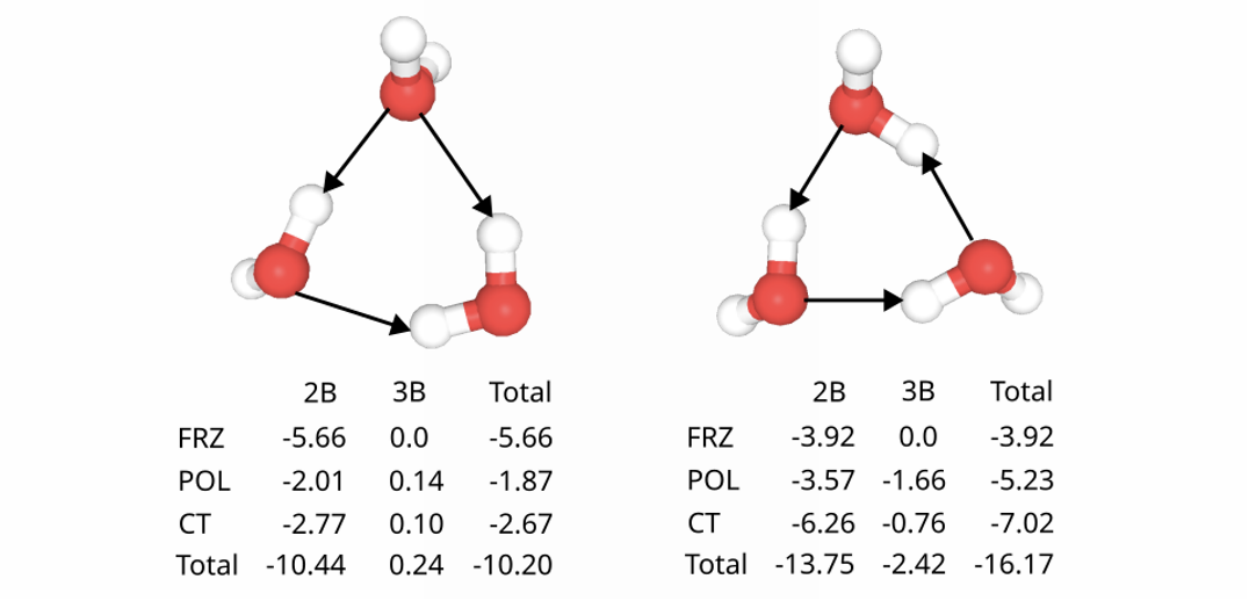
Finally, there is one more term in our polarization model which is designed to only contribute at short range. Specifically, we introduce another term proportional to density overlap,

$$V_{pol,sr} = \sum_{i<j} a_{ij}^{sr} f_4^{TT}(x_{ij}) \left(\frac{1}{2} \frac{\bar{\alpha}_i + \bar{\alpha}_j}{r_{ij}^3} \right)^{4/3} S_{ij}^{\rho} \tag{17}$$

where $\bar{\alpha}_i$ is the mean dipole polarizability of atom i, $f_4^{TT}(x_{ij})$ is the fourth-order Tang-Toennies damping function, defined in Eq. 9, and $a_{ij}^{sr} = a_i^{sr} a_j^{sr}$ is the pairwise parameter fit for this term. While Eq. 17 is more empirical than the rest of the FQCT force field, it seems to capture the effect of quadrupole polarization, which is known to be important in water[46] (which is very computationally expensive to account for compared to the magnitude of the effect). In fact the last term in Eq. 17 tends to have a magnitude for water that is only a bit larger than what would one expect for quadrupole polarization of around 20%. Thus, even though Eq. 17 is not a true model for quadrupole polarization, since it is pairwise-additive while quadrupole polarization has a many-body contribution, it works quite well. Similar terms have been suggested before for capturing the short-range polarization of ions.[47]

Charge Transfer. The true definition of charge transfer involves the transfer of actual charge between molecules,[48] and involves both attractive and repulsive contributions. The attractive component is the energy lowering associated with delocalizing the electron density, while the repulsive contribution arises from any molecule having a non-integer total charge. While there is a penalty for moving charge between molecules, Figure 1 illustrates how hydrogen bonding networks benefit from many-body charge transfer. If we assume that an equal amount of charge is transferred along each hydrogen bond, the hydrogen bonds of the

trimer on the left are organized in such a way that some molecules donate more hydrogen bonds than they receive and vice versa so that some molecules end up with nonzero total charges. This is consistent with the known instability of double acceptor water molecules which do not donate any hydrogen bonds.[49]



	2B	3B	Total
FRZ	-5.66	0.0	-5.66
POL	-2.01	0.14	-1.87
CT	-2.77	0.10	-2.67
Total	-10.44	0.24	-10.20

	2B	3B	Total
FRZ	-3.92	0.0	-3.92
POL	-3.57	-1.66	-5.23
CT	-6.26	-0.76	-7.02
Total	-13.75	-2.42	-16.17

Figure 1: Two water trimers illustrating the importance of both polarization and charge transfer for the stability of water molecules. FRZ corresponds t the frozen contribution: the sum of Pauli, electrostatic, and dispersion enegies. POL is the polarization energy and CT is the charge transfer energy. Energies are computed at the ωB97X-V/def2-QZVPPD level of theory. Total energies are interaction energies and not binding energies. See text for discussion.

In the trimer on the right, each molecule donates and receives the same number of hydrogen bonds. This means each molecule will have a nearly net-zero charge while still benefitting from charge delocalization along each hydrogen bond. If we remove any water molecule from this trimer, the two remaining molecules would have nonzero total charges and hence incur a penalty. The elimination of this penalty when a third molecule is added to the network is exactly what gives rise to a many-body charge transfer stabilization. Indeed, the trimer on the right is a so-called homodromic ring, which results in enhanced cooperativity of the induced dipoles on each molecule.[1]

The ability of hydrogen bond networks to delocalize charge while keeping each fragment very nearly neutral seems to us an under-appreciated aspect of hydrogen bond cooperativity. This can be seen by the large 3-body polarization contribution of -1.66 kcal/mol and non-negligible charge transfer contribution of -0.76 kcal/mol in Figure 2. The fact that water can receive and donate two hydrogen bonds simultaneously makes it uniquely capable of passing charge between molecules while keeping the total charge of each molecule neutral. Furthermore it is interesting to note that the EDA frozen term, i.e. the sum of electrostatics, dispersion, and Pauli repulsion, can end up being smaller than either the polarization or charge transfer contributions. Additionally, although not shown explicitly here, the charge penetration contribution, Eq. 5, to the electrostatic energy of a hydrogen bond is typically about equal to the point-electrostatic contribution.

This begs the question of how so many force fields have been even qualitatively successful while neglecting such large components of the intermolecular energy? It is rather easy to see from Figure 2 that the sum of CT and CP is generally better correlated to Pauli repulsion than either term individually. Hence most force fields implicitly describe CT and CP by having too soft of a repulsive wall when compared against an *ab initio* calculation of Pauli repulsion such as that from EDA.

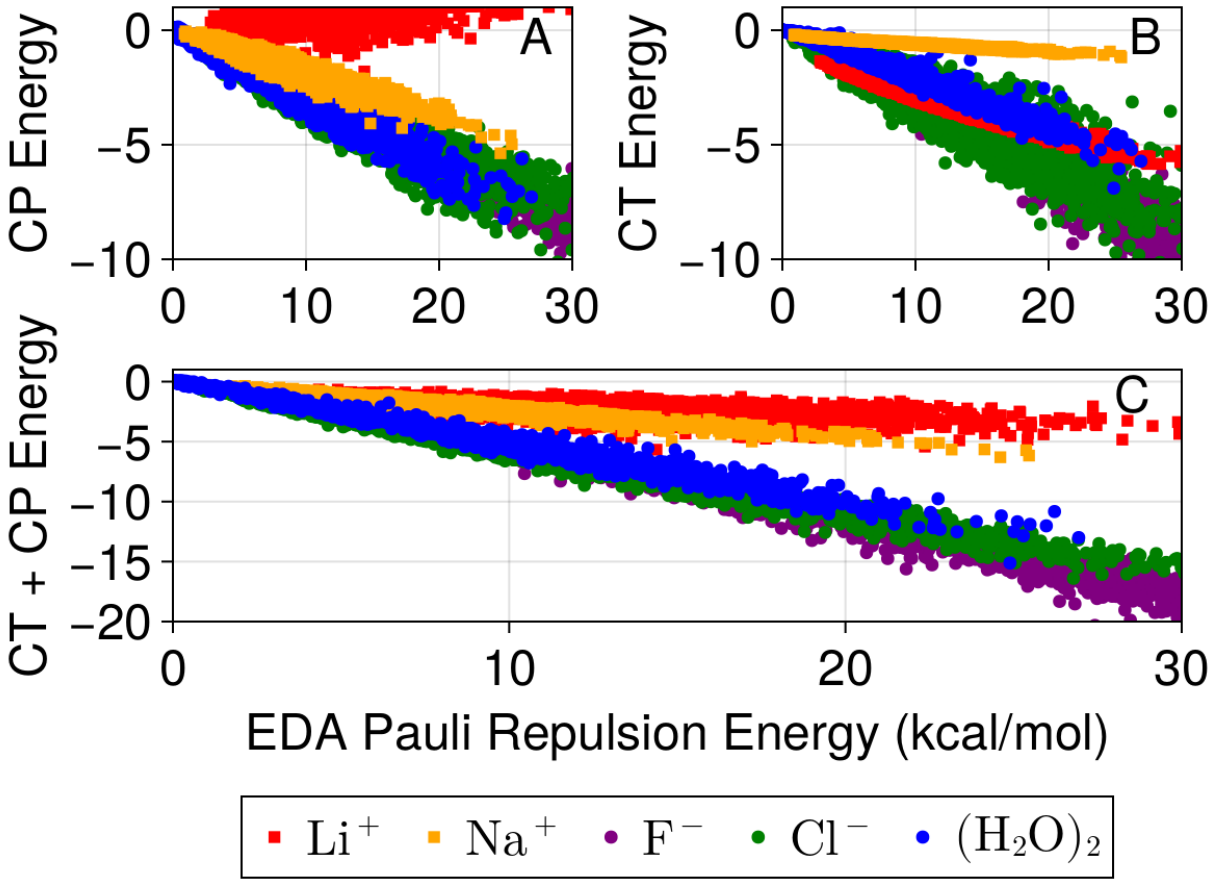


Figure 2: Correlation between (A) charge penetration energy (CP), (B) charge transfer energy (CT), and (C) their sum against Pauli repulsion energy. Energies are computed at the ωB97X-V/def2-QZVPPD level of theory. Labels with ions correspond to ion-water dimers.

The data in Figure 2 make it clear that this correlation is robust for water and ion-water dimers, albeit with different slopes depending on the dimer. Furthermore, Figure 2 shows that adding charge penetration to a force field without adding charge transfer can make the force field worse by destroying a fortuitous cancellation of errors. Of course, if a force field relies on this error cancellation, the many-body contribution to charge transfer will be neglected completely. We consider that Figures 1 and 2 provide ample motivation for the development of a better model of charge transfer, which we present here.

Charge transfer is the most difficult of the terms in EDA to model since there is no classical analogue to the QM charge transfer process involving electron flow.[46] One com-

mon approach to capturing charge transfer is to use a simple exponential dependent on the distance between atoms,[15] which captures the main effect of short-range exponential stabilization due to charge delocalization. But many-body charge transfer is non-negligible and this effect will be completely missed when using just exponentials. Another idea is to treat charge transfer the same way as polarization and solve a set of induced dipole equations as was done in MB-UCB.[12,50] While it has the benefit of capturing many-body charge transfer energies, it does not actually allow for charge to flow between molecules and therefore misses some of the salient physics. It is also ambiguous if the induced dipoles relevant to CT should be treated as real dipoles and allowed to interact with permanent and induced multipoles. These unsatisfactory approaches are also plagued by the fact that charge transfer can be an even larger contribution than polarization, especially at short range. This means the charge transfer energy would be even more susceptible to polarization catastrophes than ordinary polarization, and thus require new damping schemes to ensure the model is stable.

For all of these reasons, we introduce a new approach to describing CT which is enabled by the fact we allow for explicit charge rearrangements in our description of polarization. Our CT model includes both direct and indirect energy contributions. The direct contributions allow for energetic stabilization associated with both forward and backward CT, and are proportional to the density overlap.

$$V_{i\to j}^{CT} = a_{i\to j}^{CT} S_{ij}^{\rho} \tag{18a}$$

$$V_{j\to i}^{CT} = a_{j\to i}^{CT} S_{ij}^{\rho} \tag{18b}$$

$$V_{direct}^{CT} = \sum_{i<j} V_{i\to j}^{CT} + V_{j\to i}^{CT} \tag{18c}$$

We take inspiration from perturbation theory which shows, approximately, that the amount of charge transferred between two molecules is proportional to the energy associated with forward and backward CT.[19,51,52] Therefore, we define the amount of charge transferred from

i to j, $\Delta Q^{CT}_{i\to j}$, and from j to i, $\Delta Q^{CT}_{j\to i}$, as

$$\Delta Q^{CT}_{i\to j} = \frac{V^{CT}_{i\to j}}{\epsilon_{i\to j}} \tag{19a}$$

$$\Delta Q^{CT}_{j\to i} = \frac{V^{CT}_{j\to i}}{\epsilon_{j\to i}} \tag{19b}$$

The proportionality constant between the direct CT energy and the amount of transferred charge is written as $\epsilon_{i\to j}$ to emphasize that this proportionality is related to the difference in energy of an occupied orbital on i and an unoccupied orbital on j.[19] In our model, we choose this to be a pair-specific parameter since it decreases the total number of parameters and avoids having to choose an arbitrary combination rule. By making $\epsilon_{i\to j}$ pair-specific, it can reduce the number of parameters since many pairs do not exchange appreciable amounts of charge, such that the energy gap is effectively infinite. For instance, in water, only the oxygen to hydrogen parameter is relevant.

This approach is novel by allowing charge to explicitly move between fragments. This is achieved by modifying the molecular charge constraints used in Eq. 16. The charge constraint for a fragment A will then take the form,

$$Q^{CT}_A = Q_A + \sum_{i\in A}\sum_{j\notin A} \Delta Q^{CT}_{j\to i} - \Delta Q^{CT}_{i\to j} \tag{20}$$

The charge constraint including charge transfer, Q^{CT}_A, is simply the difference in charge transferred to atom i (in A) and charge transferred from atom i, summed over all atoms in molecule A. These charges will not be optimally distributed, so they will be allowed to relax during the polarization process. This allows us to capture the so-called "re-polarization"[19] effect in which orbitals relax after allowing for occupied-virtual mixing. For example, when charge is transferred from oxygen to hydrogen in a water dimer, the final excess charge will mostly come to rest on the oxygen in the water with net-negative charge. This re-polarization

gives rise to the indirect contributions to charge transfer,

$$V_{indirect}^{CT} = V_{pol}(\Delta Q_{CT}) - V_{pol}(0) \tag{21}$$

The indirect contribution to CT, $V_{indirect}^{CT}$, is defined as the polarization energy with CT, $V_{pol}(\Delta Q_{CT})$, minus the polarization energy wihout CT, $V_{pol}(0)$. This is the term that gives rise to many-body CT.

Because the charge transferred between fragments is proportional to the direct CT contributions, the charge constraints depend on the distance between atoms. This means there is a gradient contribution which multiplies the lagrange multipliers with the gradient of $\Delta Q_{i\to j}^{CT}$ and $\Delta Q_{j\to i}^{CT}$. This is not difficult or expensive to evaluate, but is an unusual gradient term which must be accounted for in software implementations.

The One-Body Potential. The deformation energy for a single water molecule is constructed following a protocol we have recently published.[?] The one-body potential consists of a Morse potential, cosine angle potential, a bond-bond coupling term, and bond-angle coupling term.

$$V_{bond} = D_{\mathrm{OH}}\left[1 - \exp(-\alpha(R - R_e))\right]^2 \tag{22}$$

$$V_{bb} = k_{bb}(R_1 - R_e)(R_2 - R_e) \tag{23}$$

$$V_{angle} = \frac{k_a}{2}(\cos\theta - \cos\theta_e)^2 \tag{24}$$

$$V_{ba} = k_{ba}(R - R_e)(\cos\theta - \cos\theta_e) \tag{25}$$

where D_{OH} is the dissociation energy of the O−H bond in water, R_e is the equilibrium bond length in water, and $\alpha = \sqrt{k_e/2D}$ determines the curvature of the potential as is evident from the fact it is written in terms of the harmonic force constant, k_e. The two O−H stretches in water are coupled linearly in Eq. 23 via a single bond-coupling parameter, k_{bb}. The angle potential is harmonic in $\cos\theta$ where θ is the HOH angle and θ_e is the equilibrium

angle in water, as seen in Eq. 24, and Eq. 25 shows that the angle and bond potentials are linearly coupled by a single parameter, k_{ba}. The parameters are fit to reproduce the CCSD(T)/aug-cc-pV5Z Hessian at the corresponding equilibrium geometry. Note that this is the only term for which we do not use ωB97X-V/def2-QZVPPD as a reference, simply because CCSD(T)/aug-cc-pV5Z is closer to the experimental water monomer geometry.

As will be shown, the polarizability derivatives of the water monomer are not possible to reproduce using just atomic dipole polarizabilities. Our model, however, includes fluctuating charges which improve the polarizability derivatives considerably, although the agreement with polarizability derivatives computed from electronic structure are still flawed. In the same way the dipole derivatives can be reproduced accurately by including charge flux in a model,[53] we have implemented geometry-dependent atomic hardness parameters, η.

$$\eta_{\mathrm{H1}} = \eta_{\mathrm{H}} \left(\frac{R_e}{R_{\mathrm{OH},1}} \right)^{k^\eta} \left(\frac{R_e}{R_{\mathrm{OH},2}} \right)^{k^\eta_{bb}} + k^\eta_a(\theta - \theta_e) \tag{26}$$

In Eq. 26, the atomic hardness of a particular hydrogen, η_{H1}, is modified based on the length of both O−H bonds, $R_{\mathrm{OH},1}$ and $R_{\mathrm{OH},2}$, and the angle θ. The parameters k^η, k^η_{bb}, and k^η_a describe the magnitude of change in atomic hardness and are fit to reproduce the polarizability derivatives computed from electronic structure. This particular functional form was chosen to be well-behaved when either bond is elongated, such that when the hardness of atom H1 is decreased, it increases the polarizability along that bond. This is a source of so-called electrical anharmonicity and hence contributes to the large, positive second dipole derivative associated with hydrogen-bonded water molecules.[54] Note that this term adds negligible cost to the force field evaluation since we already compute the derivatives with respect to each internal coordinate when computing the deformation energy.

Like electrostatics, polarization parameters need to be constrained to give physically meaningful parameters. Specifically, in addition to EDA energies, we include the polarizability and polarizability derivatives at the ωB97X-V/def2-QZVPPD equilibrium geometry

of water in the fitting process. The loss function we minimize against is,

$$L_{pol} = \sqrt{\frac{\sum_{i=1}^{N}(V_i^{FF} - E_i^{EDA})^2}{N}} + w_1||\boldsymbol{\alpha}^{FF} - \boldsymbol{\alpha}^{EDA}|| + w_2||\frac{\partial \boldsymbol{\alpha}^{FF}}{\partial \boldsymbol{r}} - \frac{\partial \boldsymbol{\alpha}^{EDA}}{\partial \boldsymbol{r}}|| \tag{27}$$

In the above, the first term is the RMSD of the predicted energies, V_i^{FF}, from the EDA energies E_i^{EDA}. The second term is the Frobenius norm of the difference between the computed and predicted molecular polarizabilities, $\boldsymbol{\alpha}$. The third term is the same as the second but for the polarizability derivatives. The weights, w_1 and w_2 are set to 1.0 and 0.5 respectively. This, in essence, forces the molecular polarizability to be reproduced exactly while allowing for some error in the polarizability derivatives which are much more difficult to reproduce.

The dipole derivatives needed to evaluate the field-dependent Morse potential, Eqs. 28 and 29 are computed from electronic structure by scanning along the O−H bond length of a water monomer. somethings missing here - this is unclear. As will be shown, Pauli repulsion suffers from the same systematic error in forces as the other terms. We do not know of a way to compute the dipole derivatives relevant to Pauli repulsion from electronic structure, so we treat these as fitting parameters. We have already mentioned that our Pauli repulsion model has a natural electrostatic interpretation which provides justification for having a field-dependent contribution to Pauli repulsion. Note that the field used in the case of Pauli repulsion is not the actual electric field, but the field produced by the Pauli multipoles used in evaluating Eq. 7.

In the FQCT model the intramolecular polarization is described by coupling the bonding potential to the environment through the electric field, which makes the polarization energies more accurate and dramatically improves the underlying forces. Furthermore, this term enables us to more accurately reproduce the well-known structure-frequency correspondence in water.[21] We do this by modifying Eq. 22 to also be dependent on the environmental electric field.[21] Specifically, the bond force constant, k_e, and equilibrium bond length, R_e, are coupled to the electric field projected along the bond, $\mathrm{E_{OH}}$, via the first and second dipole

derivatives, $\mu^{(1)}$ and $\mu^{(2)}$. These derivatives are treated as parameters which we compute from electronic structure. The equilibrium bond length, R_e, becomes

$$R_e(\mathrm{E_{OH}}) = R_e^0 + \frac{\mathrm{E_{OH}}\mu^{(1)}}{k_e^0 - \mathrm{E_{OH}}\mu^{(2)}} \tag{28}$$

where R_e^0 is the equilibrium bond length and k_e^0 is the force constant under zero field. The force constant under a nonzero field, $k_e(\mathrm{E_{OH}})$, is then

$$k_e(\mathrm{E_{OH}}) = k_e^0 - 3k_e^0\alpha\left(R_e(\mathrm{E_{OH}}) - R_e^0\right) - \mathrm{E_{OH}}\mu^{(2)} \tag{29}$$

We see in Eq. 29 that the field-dependent Morse potential guarantees the structure-frequency correlation will be respected at least approximately.

Based on the electrostatic interpretation of Pauli repulsion, we also include a contribution to the Pauli energy based on the field-dependent Morse potential, Eqs. 28 and 29. In fact, we find it interesting to observe that when permanent electrostatics are used, the shifts in bond length and frequency are unrealistically large. Allowing Pauli repulsion to also contribute to the field-dependent Morse potential therefore acts to mostly cancel out the effect of permanent electrostatics in shifting the bond length and frequency. This makes physical sense if one considers how each term affects the electron density of another molecule. For example, in the water dimer, electrostatics will tend to pull electron density into the internuclear region, hence weakening the hydrogen-bonded O−H bond. Pauli repulsion, however, arises from antisymmetrization of the total system wavefunction which will tend to expel electron density from the internuclear region and therefore shorten the hydrogen-bonded O−H bond.

By using the above equations we will obtain the correct slope of the structure-frequency correlation, but it will tend to understimate the actual bond length and frequency shifts. The final step to reproduce the bond lengths and vibrational frequencies is to include a contribution from CT. This is motivated by adiabatic EDA calculations, where the largest

contribution to bond elongation and red-shifting occurs on the CT surface.[22] Specifically, we allow both the bond length and force constants to be modified according to the amount of charge transferred into a hydrogen atom, as computed with Eq. 19a. This results in the final expressions used for the bond length and force constant in our environment-dependent bonding potential,

$$R_e(\mathrm{E_{OH}}) = R_e^0 + \frac{\mathrm{E_{OH}}\mu^{(1)}}{k_e^0 - \mathrm{E_{OH}}\mu^{(2)}} + k_{ct}^{(1)}\Delta Q_{CT}^{\mathrm{H}} \tag{30}$$

$$k_e(\mathrm{E_{OH}}) = k_e^0 - 3k_e^0\alpha\left(R_e(\mathrm{E_{OH}}) - R_e^0\right) - \mathrm{E_{OH}}\mu^{(2)} + k_{ct}^{(2)}\Delta Q_{CT}^{\mathrm{H}} \tag{31}$$

The rationale for using the amount of charge accepted by a hydrogen atom, $\Delta Q_{CT}^{\mathrm{H}}$, in Eqs. 30 and 31 is that this charge is transferred into an anti-bonding orbital and hence should have a large effect on the O−H bond in question. Similarly, we ignore the charge transferred out of the oxygen atom since these electrons come from non-bonding orbitals and should therefore minimally affect the O−H bonds in that water. This introduces two additional parameters, $k_{ct}^{(1)}$ and $k_{ct}^{(2)}$, which determine the sensitivity of the bond-length and force constant to CT. The importance of these parameters is that while they have a very small effect on the CT energy, they contribute a sizable effect on the CT forces. In the Supplementary Information we illustrate how the environment-dependent bonding potential improves energies and forces for the water dimer.

Ion-Water and Ion-Ion Potentials Most of the functional forms used in the representation of water are identical for ion-water and ion-ion interactions with some exceptions described here. First, ion-ion interactions at very short-range are large and hence the validity of the combination rules we use for the ion-water interaction breaks down. To remedy this issue, we make all ion-ion parameters pair-specific, and to parameterize the model in such a way that it will be easy to add new solvent-solute parameters. For many molecules pair-specific parameters are unnecessary, but for others, such as ion-ion interactions, they cannot be avoided. We also introduce simple terms which are quadratic in the density overlap.

$$V_{S^2} = \sum_{i=1}^{N_{ion}-1} \sum_{j=i+1}^{N_{ion}} a_{ij} (S_{ij}^{\rho})^2 \tag{32}$$

This term is convenient since $(S_{ij}^{\rho})^2$ tends to only be nonzero at short-range and near the equilibrium ion-ion distance. The form of Eq. 32 is used for the following terms in the ion-ion force field potential: Pauli repulsion, electrostatics, and polarization. Dispersion and charge transfer are both well-behaved enough that pair-specific parameters are sufficient. As we will see, this strategy of adding energy terms proportional to the overlap squared is very effective and we we expect to see this strategy exploited in future force fields. Finally, the ion-ion polarization and charge transfer are the same as the water model, so, as we will show, ion-ion-water many-body interactions are described accurately. **Need to see how true this is.**

Methods

Reference Data. Our model is parameterized using water clusters of size $(H_2O)_n$ with n=2-5. We use 2400 dimers, trimers, tetramers, and pentamers extracted from various minimized cluster geometries. We additionally generated 4800 pseudo-random water dimers based on a Sobol sequence. We follow exactly the same procedure as described elsewhere.[55] Using the same procedure we generated 4800 ion-water dimer geometries for all ion species considered in this study, namely F^-, Cl^-, Br^-, I^-, Li^+, Na^+, K^+, Rb^+, and Cs^+. For all ions, we also ran a 10ps *ab initio* molecular dynamics simulation at 500K with ωB97X-V/def2-TZVPPD to generate more probable ion-water configurations. We then sampled 2400 evenly spaced configurations from this trajectory to be used for parameterization.

Some larger ion-water clusters were also generated by the following procedure. We used the Crest software package[56] which uses the semi-empirical GFN2-XTB[57] method to search for global minima on a potential energy surface. We carried out the Crest global minimum search with five different seed structures generated by taking water clusters, $(H_2O)_n$, n=6-

17, from a water cluster database[58] and replacing one water randomly with one of the ions mentioned. We then took the structures of up to the ten lowest energy minima which had different hydrogen-bond networks and optimized them at the ωB97X-V/def2-TZVPPD level of theory. This resulted in a total of 1044 unique ion-water clusters. These full clusters are used to characterize the ion-water potentials, but we also extracted all possible dimers and trimers from these clusters to be used in fitting of the ion force field parameters.

We then generated clusters of the form $X^-Y^+(H_2O)_n$ for n=4-10 where X^- is any of the halide anions and Y^+ is any of the alkali metals cations. We followed the same protocol as above in generating these clusters except we started from the clusters containing a cation sampled in the previous step. A large number (**insert when finalized**) of trimers and tetramers of the ion pairs, $X^-Y^+(H_2O)_{1,2}$, were then extracted from these clusters for analyis of many-body energies involving pairs of ions rather than just a single ion. All sampled clusters are available with the paper.

All energies used in fitting parameters of the force field are computed at the ωB97X-V/def2-QZVPPD level of theory. In the cases where clusters are optimized at ωB97X-V/def2-TZVPPD, we recompute the energies of those clusters and any derived sub-structures with ωB97X-V/def2-QZVPPD. All distributed multipole calculations were carried out in the Orient program.[59]

Parameterization strategy. We fit each term against only the EDA contribution to that particular energy component. Optimization of parameters is done using simple gradient descent against the root mean-square deviation (RMSD) of predicted and EDA energies. For electrostatics and Pauli repulsion, we only use dimers in the fitting process since electrostatics is strictly pairwise-additive and Pauli repulsion is nearly so. For these terms 200 random water dimers from the datasets described above are used in fitting whereas for other many-body terms we use 200 random water dimers, trimers, tetramers, and pentamers from the datasets described above.

When parameterizing electrostatics, we optimize against two objectives. First, we en-

sure that the dipole derivatives at the equilibrium geometry of water are correct (this can be achieved nearly exactly). Second, we optimize against the distributed multipole electrostatic energy described near Eq. 5. We also include 200 random dimers of $Cl^-(H_2O)$ and $K^+(H_2O)$ when fitting against the distributed multipole electrostatics. This seems to help with optimizing to physically meaningful multipoles. We then freeze the total charges and dipoles on each atom so that the dipole derivatives will remain correct. Next, we fit the value of the core charges, Z and electrostatic exponents, b_{elec}, on each atom with respect to the total electrostatic energy from EDA. We also allow the quadrupoles to relax against the total electrostatic energy as a form of compensation for the lack of higher-order multipoles.

The Pauli repulsion term is first fit against the RMSD of the corresponding EDA energy. The repulsion parameters are then allowed to relax against the total interaction energy and interaction forces for only dimers. Using the forces is essential to get meaningful values of the dipole derivatives used in the field-dependent Morse potential for Pauli repulsion. This procedure essentially results in improved error cancellation which we find is still necessary for a robust force field. We will have more to say about the necessity of error cancellation in force fields later. It should be noted that we only allow the Pauli repulsion to optimize against dimers so that it cannot correct errors in the many-body contributions. Furthermore, we will see that the Pauli repulsion energy still ends up providing an unbiased estimate of the EDA Pauli repulsion energy.

The charge transfer energy and dispersion energies are simply fit against the RMSD from their EDA energies. Dispersion, on the other hand, has a large enough many-body contribution that if only dimers are used in the fitting, one will systematically over-estimate the dispersion energy since many-body dispersion is usually repulsive. There are methods for modeling many-body dispersion, but we have not included such terms in the FQCT model.[31,60]

Results

Water Monomer Properties

In the construction of this model, we have gone to great lengths to ensure that the model reproduces as many properties of the water monomer as possible. For instance, it was first pointed out by Fanourgakis and Xantheas that reproducing the dipole surface of water is essential for capturing the opening of the bend angle of water in the condensed phase.[61] This requires a model that reproduces the dipole derivatives of water to correctly predict the opening of the bend angle as water clusters become larger.

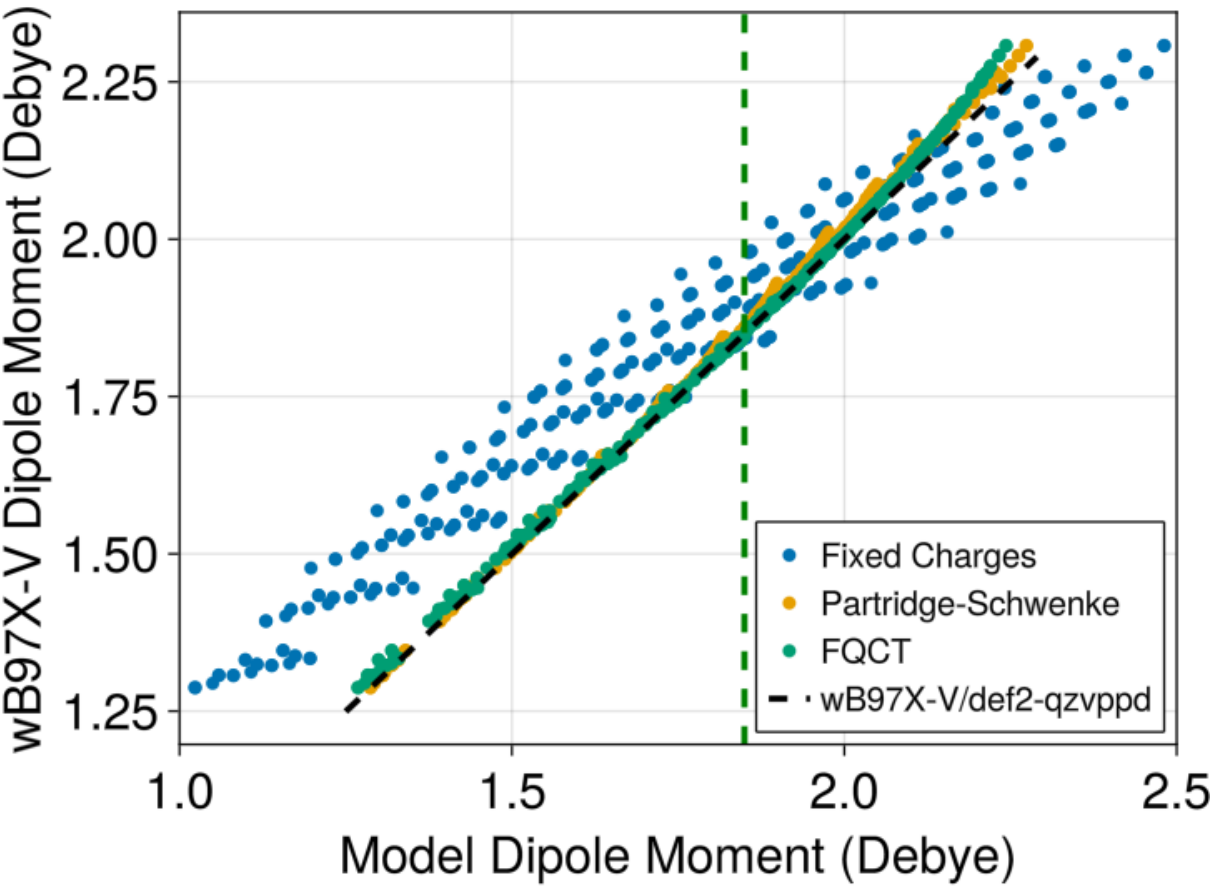


Figure 3: The dipole surface of water for all structures taken from a 3-D scan of water internal coordinates with a deformation energy less than 20 kcal/mol. The black dashed line shows the values computed with ωB97X-V/def2-QZVPPD. The green dashed line corresponds to the experimental gas-phase dipole moment of water of 1.85 Debye.

In Figure 3 we make a comparison between the dipole surface of the FQCT model (green), a dipole surface with fixed charges and dipoles that optimally reproduce the EDA electrostatic energy of the water dimer (blue), the dipole surface using the Partridge-Schwenke (PS) model[62] (orange), against the reference surface computed with ωB97X-V/def2-QZVPPD. Clearly, fixed charge force fields completely fail to reproduce the dipole surface of water whereas the dipole surface associated with the PS water monomer surface is exact over a wide range of energies by construction.[62] Figure 3 shows that the FQCT dipole surface is nearly exact as the PS dipole surface, reproducing the dipole derivatives of water at its equilibrium geometry to five decimal places in atomic units.

Another important property for accurate energetics and transferability, especially for interactions with ions, is the molecular polarizability. The polarizability is evaluated for a water molecule with the y-axis as the bisector of the HOH angle and the z-axis normal to the plane of the water molecule. Our model reproduces the ωB97X-V/def2-QZVPPD molecular polarizability up to three decimal places in bohr^3, with $\alpha_{xx} = 10.0321$, $\alpha_{yy} = 9.65958$, and $\alpha_{zz} = 9.40921$.

The last monomer property we explicitly aimed to reproduce are the polarizability derivatives of gas-phase water at its equilibrium geometry. The polarizability derivatives control the intensity of peaks measured with Raman spectroscopy but are rarely discussed in the construction of water models. For example, at least one water model has been constructed for the express purpose of reproducing Raman spectra,[63] but the polarizability derivatives were not reported. Additionally, since water distorts when interacting with other molecules, reproducing the polarizability derivatives indicates how well the molecular polarizability at distorted geometries will be reproduced.

In Table 1, we report the polarizability derivatives of the FQCT model as well as the polarizability derivatives for an identically parameterized model which does not include fluctuating charges, and compare them to the QM reference model. Because charge fluctuations for water contribute to in-plane polarization, we find that the model gives much better xx,

xy, and yy polarizability derivatives than one which just uses anisotropic dipole polarizabilities. Note that the polarizability derivatives in Table 1 are also improved considerably by the geometry-dependent atomic hardness described in Eq. 26. This indicates that one of the main reasons water models have historically predicted Raman intensities very poorly[64] is the lack of fluctuating charges in the polarization process.

Table 1: Polarizability derivatives of water computed at the ωB97X-V/def2-QZVPPD equilibrium geometry in three different ways. The first entry is computed with ωB97X-V/def2-QZVPPD, the second with FQCT, and the third with the same model but using parameters optimized without fluctuating charges. The Atom column tells both the atom and component of its position we take the derivative with respect to. The water monomer has its bisector aligned with the y-axis and the z-axis is normal to the water molecule plane. Note that the derivatives for the second hydrogen are identical to the first but with opposite sign. The xz and yz entries are omitted since they are small and reproduced to three decimal places by both models.

Polarizability Derivatives of Water (bohr^2)

Atom	xx	xy	yy	zz
Ox	-	4.04/3.99/-0.13	-	-
Oy	5.15/6.07/2.04	-	4.45/4.85/-2.03	1.50/0.0/0.0
Oz	-	-	-	-
Hx	-4.61/-4.80/0.78	-2.02/-1.99/0.06	-2.53/-1.68/-0.78	-1.39/0.0/0.0
Hy	-2.57/-3.04/-1.02	-1.68/-1.66/-0.08	-2.22/-2.43/1.02	-0.75/0.0/0.0
Hz	-	-	-	-

While some the polarizability derivatives in Table 1 are reproduced very accurately for the in-plane components, the zz polarizability derivatives are not as well described. It is important to note that one can achieve better agreement with reference polarizability derivatives if that is the only quantity one aims to reproduce, but these models tend to result in inadequate polarization energies. The zz polarizability derivatives of water are an interesting case since they can be reproduced accurately when using intramolecular induced dipole interactions, but this tends to make the polarization energies worse and will make polarization catastrophes more likely for ion-water systems. We can also reproduce the zz polarizability derivatives by allowing the z-component of atomic polarizabilities to be geometry-dependent. We decided not to do this since it adds additional complexity with a

fairly small benefit for water, but for a system like benzene which predominantly interacts via π interactions we will take that additional step to recover accurate out-of-plane polarizability derivatives in the future.

The zz polarizability derivatives of water are an interesting case since they can be reproduced accurately when using intramolecular induced dipole interactions, but this tends to make the polarization energies worse and will make polarization catastrophes more likely for ion-water systems. We could certainly reproduce the zz polarizability derivatives by allowing the z-component of atomic polarizabilities to be geometry-dependent. We decided not to do this since it adds additional complexity with a fairly small benefit for water. For a system like benzene, which predominantly interacts via π interactions, we will take that additional step to recover accurate out-of-plane polarizability derivatives in the future.

Water Intermolecular Interactions

Next we turn to the performance of FQCT on water-water intermolecular properties. We start with the reproduction of the QM energies and ALMO-EDA energy components presented in Supplementary Table S1 which presents the mean absolute errors (MAEs) over each cluster size. In general, the MAEs obtained by our model are excellent, with errors no larger than 0.25 kcal/mol. The reason for this is illustrated in Figure 4 which shows the correlation of errors in all attractive terms against Pauli repulsion both with and without error fitting against ALMO-EDA. Figure 4A shows that just by fitting each EDA term independently, our model naturally correlates attractive and repulsive errors such that when repulsion gets stronger, all of the attractive terms also get stronger, indicating that some amount of error cancellation is guaranteed. Thus the FQCT model is much more effective at taking advantage of error cancellation as seen in Figure 4B because Pauli repulsion and all other terms are naturally correlated. This is easily seen in Figure 2 where the sum of charge penetration and charge transfer are linearly correlated to Pauli repulsion over a very wide energy range.

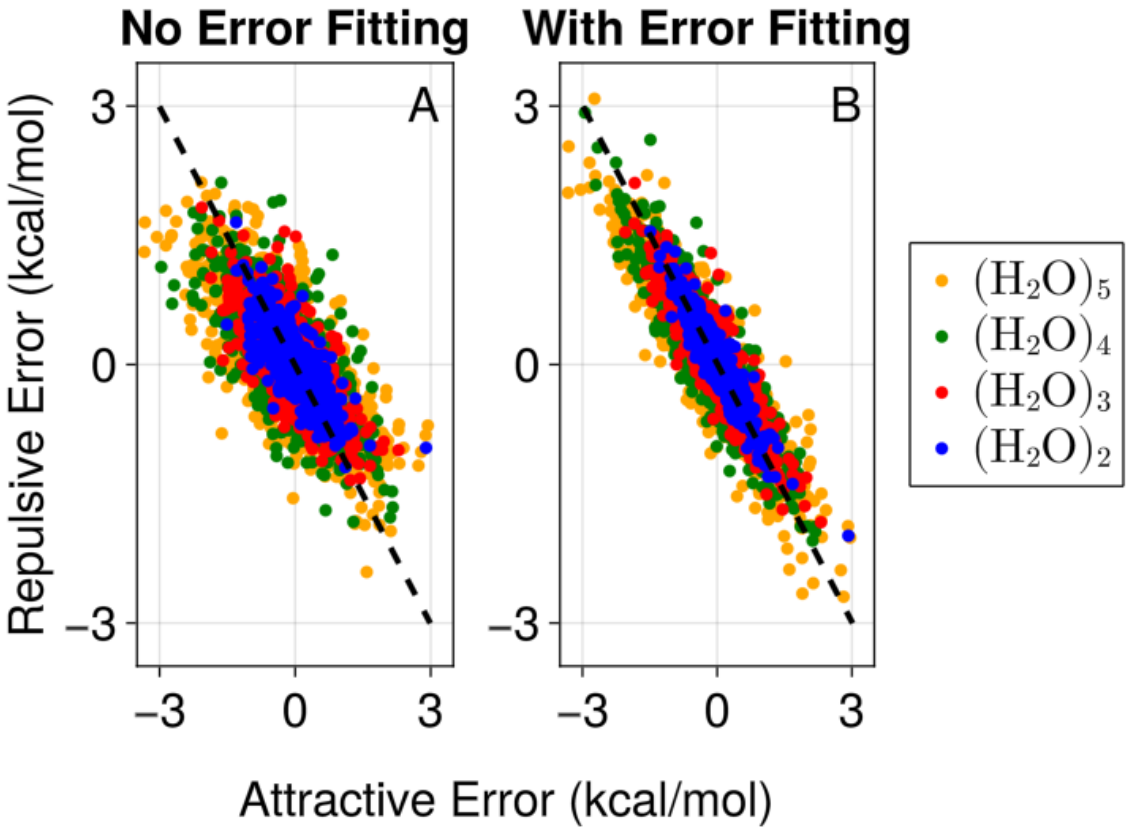


Figure 4: Correlation of errors in Pauli repulsion against all attractive interactions from EDA including dispersion, electrostatics and CP, polarization, and charge transfer. Figure A shows the error correlation without any error fitting, where each EDA term is fit independently. Figure B shows the same plot after allowing the Pauli repulsion to relax against the interaction energy for dimers only to improve error cancellation. See section on parameterization for more details.

One of the major goals of the FQCT model is to quantitatively reproduce the many-body contributions to both polarization and charge transfer. To assess how well we have achieved this, we computed the three-body contribution to both polarization and charge transfer with water trimers not used in parameterizing the water model. Figure 5 shows that we obtain excellent agreement with electronic structure for the major three-body contributions to the energy. Indeed, the model manages to capture both repulsive and attractive three-body contributions to both polarization and charge transfer despite repulsive three-body contributions to either of those quantities being absent in ordinary water clusters.[10]

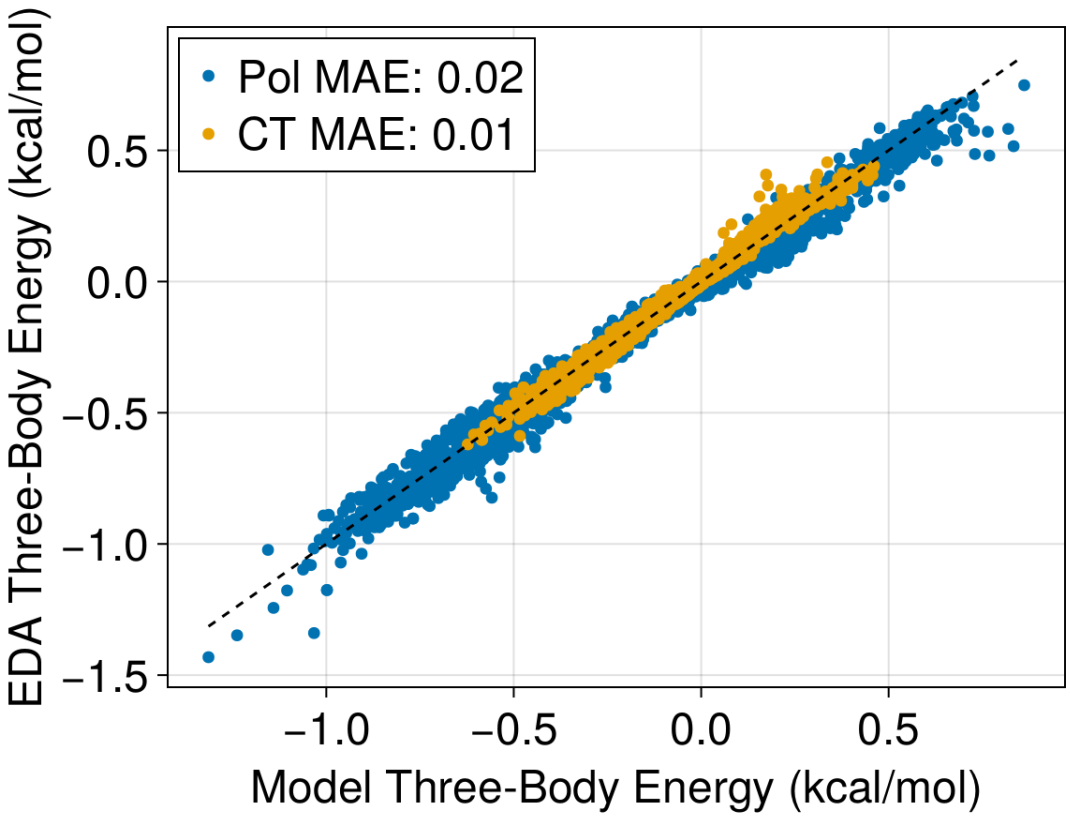


Figure 5: Correlation plot of the three-body contribution to polarization and charge transfer as computed by our model and with ωB97X-V/def2-QZVPPD. All water trimers in this plot are drawn from ion-water clusters and therefore span a wide range of configurations, some of which are atypical of liquid water. Since trimers drawn from a cluster may be disonnected, we enforce that the trimer have an absolute three-body contribution of at least 0.02 kcal/mol.

We have also computed the energy of many water clusters for which high-quality CCSD(T) benchmark energies are available,[65] and compare the energies predicted by our model with several other force fields in Table 2. We also include ωB97X-V/def2-QZVPPD in Table 2 since this is the reference method for our force field. It is clear that this level of theory compares very favorably to the CCSD(T)/CBS and MP2/CBS numbers used as reference.

It is worth noting that both q-AQUA and MB-Pol include fits of the 2-body, 3-body, and in the case of q-AQUA 4-body contributions to the energy at CCSD(T). Therefore, the fact that our model built for transferability is able to achieve an MAE over all clusters that is as accurate as these fitted models is encouraging. Interestingly, MB-Pol seems to have a tendency to underestimate the binding energy of clusters as they get larger. Presumably

this shortcoming has been corrected in the updated version of MB-Pol[66] which is reportedly at least as accurate as q-AQUA. As highlighted in the methods section, the electrostatics and Pauli repulsion used in our model are nearly the same as those used in HIPPO.[15] It is very encouraging, therefore, to see that our model improves on the MAEs of the HIPPO model rather substantially. One of the main reasons we are able to achieve better MAEs is

Table 2: Comparison of various advanced force fields and ωB97X-V/def2-QZVPPD against benchmark cluster energies.[65] The reference energies are mostly CCSD(T)/CBS values but some are MP2/CBS values; see original paper for further details on structures.[65] We use the FF optimized geometries to evaluate energies, and any FF energies which could not be found in the literature are left blank. The bottom row shows the mean absolute error per molecule for all available energies.

Comparison of Methods on Benchmark Water Cluster Binding Energies

$(H_2O)_n$	Isomer	q-AQUA	MB-Pol	FQCT	ωB97X-V	HIPPO	Ref.
$(H_2O)_2$		-4.97	-4.96	-4.90	-5.00	-4.96	-4.99
$(H_2O)_3$		-15.73	15.69	-15.35	-15.77	-15.77	-15.77
$(H_2O)_4$		-27.35	-27.18	-27.34	-27.75	-26.69	-27.39
$(H_2O)_5$		-35.71	-35.55	-36.36	-36.51	-34.58	-35.9
$(H_2O)_6$	Prism	-46.21	-45.94	-45.95	-46.53	-46.15	-46.2
$(H_2O)_6$	Cage	-45.94	-45.66	-45.70	-46.30	-45.39	-45.9
$(H_2O)_6$	Book	-45.21	-44.89	-45.58	-45.95	-44.25	-45.4
$(H_2O)_6$	Ring	-43.71	-43.66	-45.31	-45.07	-42.54	-44.3
$(H_2O)_7$		-57.71	-57.11	-57.63	-58.08	-	-57.4
$(H_2O)_8$	D_2d	-73.32	-72.38	-72.67	-73.58	-71.55	-73.0
$(H_2O)_8$	S_4	-72.93	-72.06	-72.75	-73.55	-71.56	-72.9
$(H_2O)_9$	D_2dDD	-82.87	-81.40	-82.28	-83.00	-	-83.0
$(H_2O)_{10}$		-94.72	-92.53	-93.81	-94.50	-	-94.6
$(H_2O)_{11}$	43'4	-104.23	-103.93	-103.03	-103.77	-100.23	-104.6
$(H_2O)_{16}$	Antiboat	-164.87	-160.89	-163.26	-164.20	-159.63	-164.6
$(H_2O)_{16}$	4444-a	-163.10	-162.28	-163.32	-164.28	-161.84	-164.2
$(H_2O)_{16}$	4444-b	-162.54	-161.08	-163.18	-163.84	-161.56	-164.1
$(H_2O)_{16}$	Boat a	-164.53	-161.43	-162.99	-164.51	-159.36	-164.4
$(H_2O)_{16}$	Boat b	-164.31	-160.86	-163.09	-164.35	-159.43	-164.2
$(H_2O)_{17}$	Sphere	-177.56	-171.75	-175.26	-175.78	-170.68	-175.7
$(H_2O)_{20}$	ES Prism	-212.49	-207.84	-211.35	-211.98	-	-214.2
$(H_2O)_{20}$	FS Prism	-210.63	-206.96	-209.20	-210.12	-	-211.9
$(H_2O)_{20}$	Fused Cubes	-208.07	-207.45	-208.85	-209.90	-	-210.6
$(H_2O)_{20}$	Pentag. Dodec.	-199.79	-195.22	-200.29	-201.22	-	-200.8
$(H_2O)_{25}$	Isomer 2	-276.50	-266.04	-269.60	-270.26	-	-276.3
MAE/n		**0.040**	**0.156**	**0.079**	**0.051**	**0.194**	-

that our many-body energies, as shown in Figure 5, are both accurate and nearly unbiased, while the HIPPO model systematically underestimates the many-body contributions to water cluster energies,[15] which is likely attributable to the absence of many-body charge transfer. Interestingly, for larger clusters our model begins to slightly underestimate the energies compared to the benchmark references, but so does ωB97X-V/def2-QZVPPD, which is the best our force field can possibly perform.

In addition to accurate energetics, many-body FFs should also produce accurate geometries. Table 3 shows the root mean-squared deviation (RMSD) of water cluster structures optimized with various FFs and ωB97X-V/def2-QZVPPD compared to previously reported structures optimized with either CCSD(T) or MP2.[65] The average RMSD for our model manages to outperform MB-Pol and is comparable even to q-AQUA. Notably, ωB97X-V/def2-QZVPPD is even more accurate on this benchmark than q-AQUA which provides further evidence of the high accuracy of this density functional.

In order for a FF to be useful for theoretical spectroscopy, it must respect the relationships between structure and vibrational frequencies. In the case of water, this manifests as a linear relationship between the change in bond length and change in O−H stretching frequency,[21] also known as Badger's rule.[67] In the course of constructing the FQCT FF, we tested Badger's rule and if it satisfied the expected slope $\Delta\omega$ vs ΔR_e of ≈ -19 cm^{-1}/.001Å.[21] We found instead that the slope was ≈ -10 cm^{-1}/.001Å(See Fig. S2) which is in agreement with a slope of ≈ -11 cm^{-1}/.001Åfound by Boyer *et al.* when they use a field-independent Morse oscillator with parameters appropriate to water. To obtain the correct slope value requires an environmental electric field dependence,[21] and motivates our bonding potential parameters to be modulated by the field along the O-H bond, as given by Eqs. 28 and 29.

We computed the necessary dipole derivatives from a simple O−H scan and found the parameters $\mu^{(1)} = 0.1654$ and $\mu^{(2)} = -0.01246$. If we do the same calculation with FQCT, we get $\mu^{(1)} = 0.1658$ and $\mu^{(2)} = -0.0104$. This indicates that as long as a force field has an accurate dipole surface, the dipole derivatives needed to compute the field-dependence

Table 3: Comparison of various advanced force fields and ωB97X-V/def2-QZVPPD against benchmark cluster structures.[65] The reference structures are optimized at either CCSD(T)/aug-cc-pVDZ or MP2/aug-cc-pVTZ. See original paper for further details on structures.[65] The bottom row shows the root mean-squared deviation (RMSD) in angstrom for all available structures.

Comparison of Methods on Benchmark Water Cluster Structures

$(H_2O)_n$	Isomer	q-AQUA	MB-Pol	FQCT	ωB97X-V
$(H_2O)_2$		0.005	0.008	0.014	0.005
$(H_2O)_3$		0.010	0.014	0.021	0.008
$(H_2O)_4$		0.008	0.024	0.017	0.006
$(H_2O)_5$		0.013	0.059	0.047	0.008
$(H_2O)_6$	Prism	0.010	0.035	0.021	0.009
$(H_2O)_6$	Cage	0.013	0.027	0.024	0.018
$(H_2O)_6$	Book	0.010	0.029	0.042	0.009
$(H_2O)_6$	Ring	0.013	0.043	0.031	0.010
$(H_2O)_7$		0.016	0.041	0.027	0.025
$(H_2O)_8$	D_2d	0.006	0.041	0.018	0.004
$(H_2O)_8$	S_4	0.007	0.019	0.017	0.005
$(H_2O)_9$	D_2dDD	0.089	0.116	0.037	0.052
$(H_2O)_{10}$		0.012	0.049	0.022	0.010
$(H_2O)_{11}$	43'4	0.034	0.065	0.029	0.017
$(H_2O)_{16}$	Antiboat	0.023	0.064	0.036	0.017
$(H_2O)_{16}$	4444-a	0.039	0.038	0.026	0.015
$(H_2O)_{16}$	4444-b	0.040	0.049	0.026	0.029
$(H_2O)_{16}$	Boat a	0.023	0.038	0.025	0.016
$(H_2O)_{16}$	Boat b	0.028	0.057	0.047	0.016
$(H_2O)_{17}$	Sphere	0.039	0.063	0.032	0.022
$(H_2O)_{20}$	ES Prism	0.042	0.056	0.039	0.024
$(H_2O)_{20}$	FS Prism	0.047	0.050	0.028	0.023
$(H_2O)_{20}$	Fused Cubes	0.067	0.050	0.027	0.029
$(H_2O)_{20}$	Pentag. Dodec.	0.034	0.066	0.035	0.018
$(H_2O)_{25}$	Isomer 2	0.029	0.049	0.037	0.038
RMSD (Å)		**0.026**	**0.046**	**0.029**	**0.017**

of a Morse potential can be computed directly from the force field. The final result of the structure-frequency correlation in water using the FQCT model is shown in Figure 6. We consider this an excellent result given the simplicity of the field-dependent Morse potential, especially since it requires no free parameters.

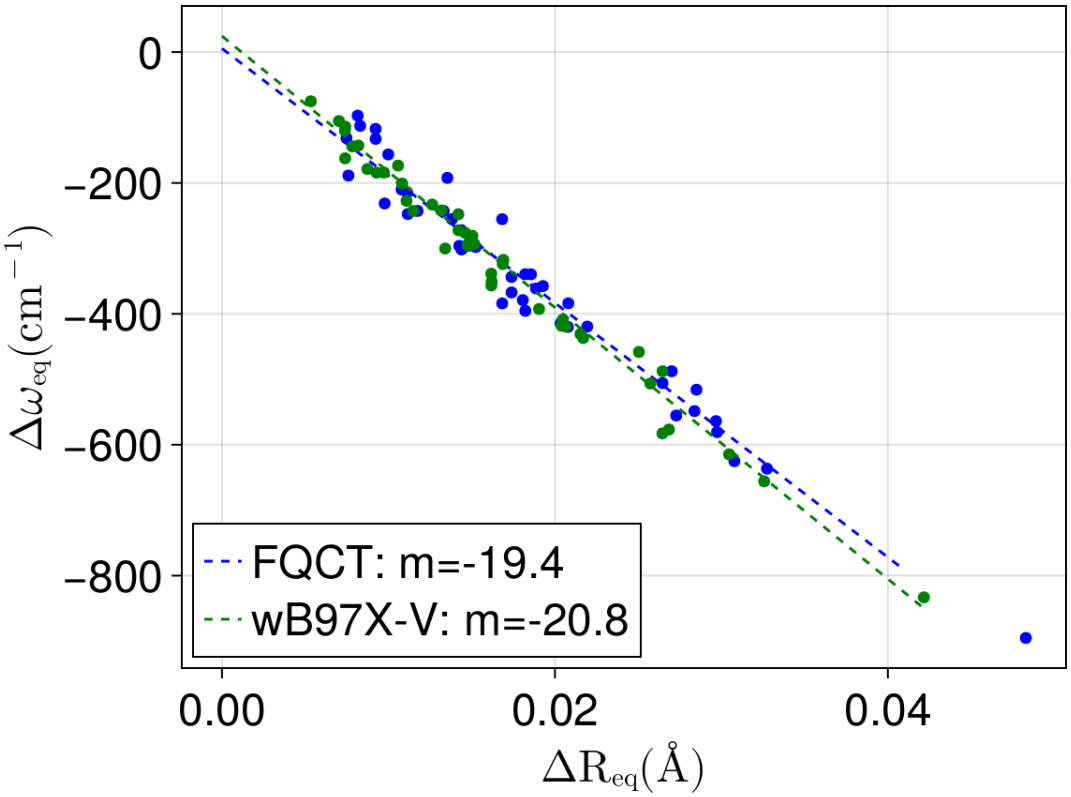


Figure 6: Correlation of $\Delta\omega$ vs ΔR_e over a collection of small water clusters using FQCT and ωB97X-V/def2-QZVPPD. The water clusters are low-energy structures of $(H_2O)_{2\text{-}6}$ available with the paper. The linear fits are not constrained to pass through zero which explains the slightly large slopes compared to previous work.[21]

Ion-Water Model

Accurately reproducing many-body energies is a very stringent test of a water model, but this is especially true for ion-water systems where polarizable force fields have historically struggled. We now turn our attention to the performance of this model on ion-water clusters. First, we consider simple scans for each of the ion-water dimers considered in this study. Figure 7 show the total energy curve of FQCT as a solid colored line and the QM reference energy curve as a dashed line of the same color for each ion. In the Supplementary Information, we include a variation on these plots where we show the error in each individual EDA component of the scan.

Perhaps the most important feature of the dimer scans in Figure 7 is the long-range

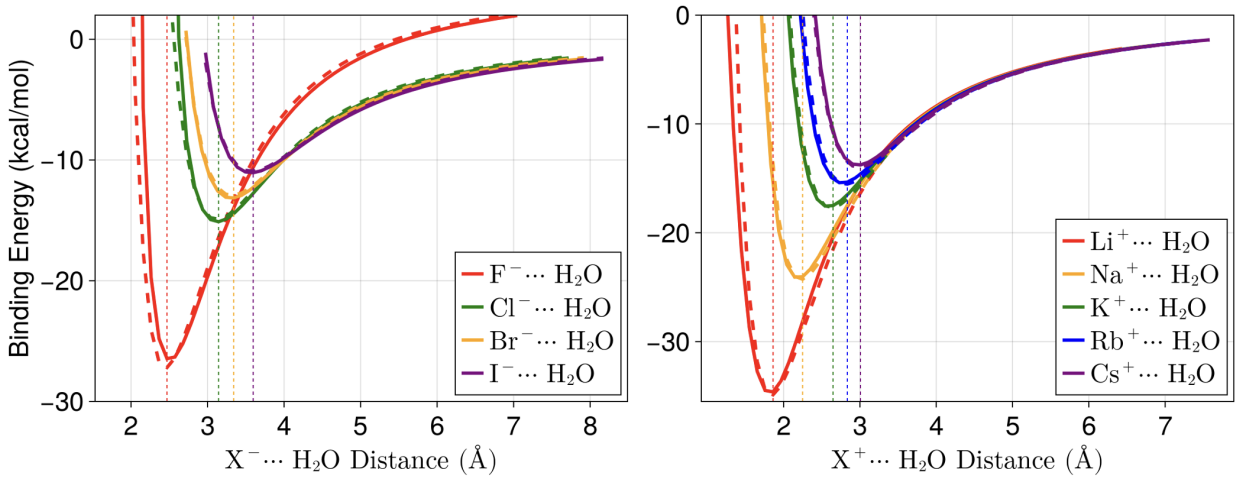


Figure 7: Scans of the $X^-(H_2O)$ for cations and anions. (A) (X^-=F^-, Cl^-, Br^-, and I^-) and (B) Scans of the $X^+(H_2O)$ (X^+=Li^+, Na^+, K^+, Rb^+, and Cs^+) potentials where the geometry is held fixed at the equilibrium structure of each dimer. The solid curve corresponds to FQCT and the dashed curve corresponds to ωB97X-V/def2-QZVPPD. Vertical dashed lines show the positions of each minimum with ωB97X-V/def2-QZVPPD. Binding energies and harmonic frequencies are reported in Table 4.

attractive part of the potential which we are able to capture very accurately in all cases. Unsurprisingly, the largest errors are found for F^- and Li^+ which are the most challenging ions considered in this work due to their extremely short-range (nearly covalent) interactions. Another important feature of the ion-water dimer potential is the repulsive wall since in the condensed phase, many-body stabilization shortens the oxygen-ion distance beyond the dimer equilibrium distance. Looking at the errors of each EDA term inside the equilibrium distance in Supplementary Figure X, it is clear that the Pauli repulsion and charge transfer tends to be underestimated while electrostatics and polarization are overestimated but these errors are small and compensated for by a net error cancellation. In fact the binding energies of ion-water interactions with FQCT are extremely accurate compared to the DFT reference, with the largest error of 0.7 kcal/mol for $F^-(H_2O)$, while the rest of the ions have binding energies accurate to within a few tenths of a kcal/mol.

Since ion-water interactions are strong, they can result in large red-shifts of the underlying vibrational frequencies. Table 4 provides the frequencies and binding energies of FQCT compared to ωB97X-V in which the FF model is generally quite accurate. The O−H

Table 4: Comparison of FQCT frequencies (cm^{-1}) and binding energies (kcal/mol) against ωB97X-V/def2-QZVPPD for all ion-water dimers discussed in this paper.

Ion-Water Dimer Vibrational Frequencies

$X^{+/-}(H_2O)$	Method	NM1	NM2	NM3	NM4	NM5	NM6	D_e
$F^-(H_2O)$	FQCT	365	723	1328	1806	2156	3842	-26.7
	ωB97X-V	384	569	1144	1702	2233	3916	-27.2
$Cl^-(H_2O)$	FQCT	197	395	767	1719	3427	3893	-15.1
	ωB97X-V	193	341	726	1678	3417	3919	-14.9
$Br^-(H_2O)$	FQCT	157	360	718	1713	3507	3894	-13.2
	ωB97X-V	151	288	657	1673	3521	3916	-12.9
$I^-(H_2O)$	FQCT	129	303	643	1701	3591	3895	-11.1
	ωB97X-V	118	220	579	1668	3623	3911	-10.9
$Li^+(H_2O)$	FQCT	301	512	535	1669	3677	3801	-34.9
	ωB97X-V	392	524	554	1681	3815	3882	-34.9
$Na^+(H_2O)$	FQCT	268	316	434	1679	3737	3848	-24.2
	ωB97X-V	307	367	437	1677	3830	3902	-24.3
$K^+(H_2O)$	FQCT	220	306	367	1680	3767	3873	-17.6
	ωB97X-V	213	359	369	1673	3833	3910	-17.7
$Rb^+(H_2O)$	FQCT	180	304	342	1680	3776	3881	-15.4
	ωB97X-V	178	347	351	1671	3836	3914	-15.6
$Cs^+(H_2O)$	FQCT	160	306	316	1679	3785	3889	-13.8
	ωB97X-V	157	327	339	1668	3836	3916	-14.0

stretching modes, NM5, are fairly accurate for all anions especially considering the strength of interactions involved. This highlights that reproducing the structure-frequency correspondence shown in Figure 6 transfers to interactions with anions. Is this true - it looks as good as anions to me....On the other hand, the O−H stretching modes for cation-water dimers are red-shifted too much compared to ωB97X-V. The bending mode for anion-water dimers, NM4, is systematically blue-shifted, while this problem with the bending frequencies does not occur for the cations. This is because the anion attracts the water hydrogen atom causing the HOH angle to close more than it should. This is likely attributable to the lack of a field-dependent contribution to the bending potential, which would counter-act this tendency by instead opening the HOH angle. In the future, we may explore the addition of a field-dependent contribution to the bending potential. **cite allen**

Just as with water, many-body contributions to both polarization and charge transfer are

important for ion-water systems. While on a relative basis their contributions in ion-water systems are smaller than in water due to the dominance of electrostatic interactions, that does not mean they can be neglected. In Figure 8 we show the correlation of three-body polarization and charge transfer energies computed from FQCT and EDA. These energies are computed from 400 trimers drawn from ion-water clusters which we generated for this work. Figures 8 demonstrate that our models of polarization and charge transfer are transferable even to very strong interactions.

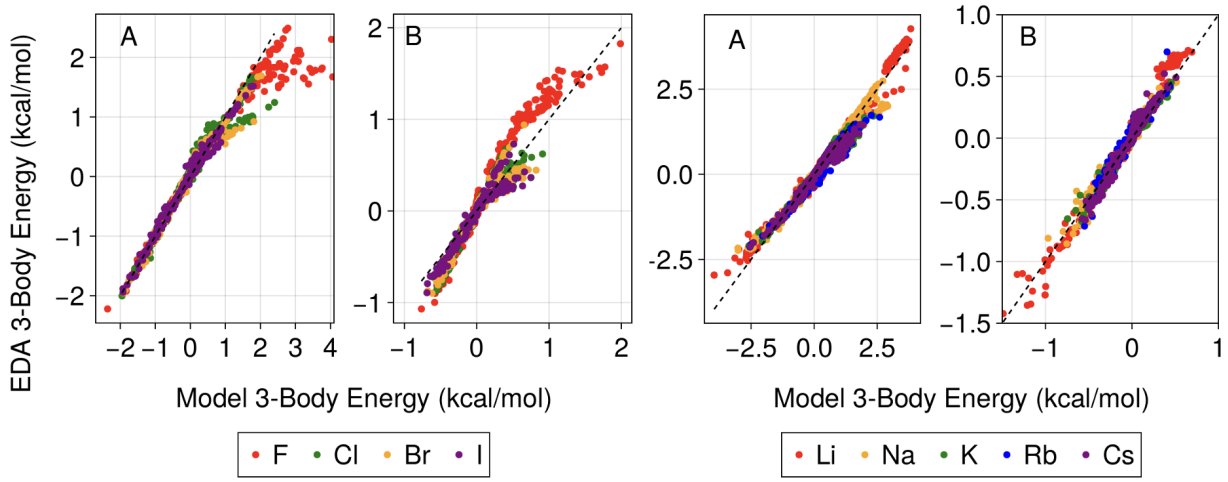


Figure 8: Correlation of three-body polarization and charge transfer energies between FQCT and EDA for anions and cations. All trimers are drawn from ion-water clusters optimized with ωB97X-V/def2-TZVPPD and energies are computed with the def2-QZVPPD basis set. (A) The 3-body polarization MAEs over the halide trimers are 0.208, 0.072, 0.060, and 0.046 kcal/mol for F^-, Cl^-, Br^-, and I^-, respectively. (B) The 3-body charge transfer MAEs over the halide trimers are 0.129, 0.068, 0.056, and 0.044 kcal/mol for F^-, Cl^-, Br^-, and I^-, respectively. (C) The 3-body polarization MAEs over the alkali trimers are 0.122, 0.105, 0.080, 0.123, and 0.078 kcal/mol for Li^+, Na^+, K^+, Rb^+, and Cs^+, respectively. (D) The 3-body charge transfer MAEs over the alkali trimers are 0.050, 0.024, 0.021, 0.033, and 0.029 kcal/mol for Li^+, Na^+, K^+, Rb^+, and Cs^+, respectively.

In Figure 8A for water-halides, the attractive contributions to many-body polarization are captured exceptionally well, while some additional scatter arises for repulsive three-body polarization. The accuracy of many-body charge transfer with water-halides is similar regardless of whether the energy is attractive or repulsive as seen from Figure 8B. While Figure 8C finds that many-body polarization is somewhat worse for the aqueous alkali ions than for

water-halides, it is still quite accurate with an MAE of 0.1 kcal/mol. The greater difficulty in modeling cation-water polarization is likely due to the highly polarizable oxygen atom which cations interact with directly. Figure 8D is another illustration of the transferability and general accuracy of our charge transfer model. Interestingly, the errors in three-body charge transfer decrease going from Li^+ to Na^+ to K^+, which is expected since the magnitude of the energy also decreases. The MAEs for Rb^+ and Cs^+ increase to about 0.033 kcal/mol and the correlation is slightly skewed. In any case, many-body charge transfer is rather small for these ions and FQCT is still able to capture the trend accurately.

Ion-Ion Potentials

Aqueous ionic solutions often exhibit equilibria between ion-ion contact pairs, solvent-separated ion pairs, and fully solvated ions. The relative abundance of these states depends sensitively on the nature of the ion-ion interaction itself.[68] Let us begin, then, by looking at the ion-ion dimer potentials in Figure 9.

Discussion

Most polarizable force fields include intramolecular polarization by allowing the induced dipoles to interact regardless of whether they are in the same molecule or not. In our experience with EDA, this tends to result in worse agreement between the model and reference polarization energies. This is not surprising given that multipoles located within a couple bond lengths are too close for electrostatics to give a good representation of the relevant interactions. There is no doubt, however, that intramolecular polarization should be considered in some way.

The necessity of coupling the bonding potential to the environment to accurately reproduce structure-frequency relationships in a force field is a new observation as far as we are aware. We consider this an excellent result given the simplicity of the field-dependent morse

potential, especially since it requires no free parameters.

When combined with our new model for the one-body potential, we also ensure that the force field reproduces all physically relevant monomer properties including the dipole moment, dipole derivatives, molecular polarizability, and polarizability derivatives.

The damping functions generated by the usual Thole damping procedure,[69] are exponentials multiplied by first-, second-, and third-order polynomials for charge-charge, charge-dipole, and dipole-dipole interactions, respectively.[69] These polynomials are smaller than the scaling of mutual polarization, which is roughly the square of a similar permanent multipole interaction. That is, the field due to a permanent dipole decays as r_{ij}^{-3}. The dipole induced by this field interacts with other induced dipoles with an energy that is proportional to the magnitude of the induced dipole and via a field that decays as r_{ij}^{-3}. Since the magnitude of an induced dipole is proportional to the permanent applied field, this means mutual polarization of induced dipoles decays as $\sim r_{ij}^{-6}$. The damping functions generated from the overlap of slater orbitals are third-, fourth-, and fifth-order polynmials multiplied by an exponential for charge-charge, charge-dipole, and dipole-dipole interactions, respectively. These damping functions are therefore better able to control the polarization energy than the Thole damping functions used in other force fields. In fact, the small orders of the polynomials in Thole damping likely explains much of the historic difficulty of controlling polarization between ions and water.[70,71]

The necessity of coupling the bonding potential to the environment to accurately reproduce structure-frequency relationships has been shown to reproduce the O-H signatures in a recent Raman theory for water,[?] and has been applied to force field for a first time here. This approach is easily extensible to other force fields and should immediately improve spectroscopic predictions. Additionally, the ability of FQCT to generate accurate structures is useful for cases where the force field is used to generate configurations for further analysis with electronic structure. As an aside, many force fields use harmonic bond potentials and the field-perturbed quantum harmonic oscillator is exactly solvable, so a similar modification

can be made in that case.

Conclusions

We have described new approaches to modeling polarization and charge transfer which we parameterize against a very accurate energy decomposition analysis. In doing so, we have highlighted the importance of many-body charge transfer for aqueous systems. Our new model of charge transfer is able to quantitatively capture many-body charge transfer for both water and ion-water clusters. This model of charge transfer is enabled by the use of fluctuating charges in the polarization model.

We also show that fluctuating charges greatly improve the accuracy of polarizability derivatives. Accurate polarizability derivatives are essential for computing Raman spectra, which polarizable force fields have historically modeled very poorly. We therefore expect this model to be useful for theoretical spectroscopy among many other uses. To that end, we also show that our model can reproduce the structure-frequency correspondence central to hydrogen-bonded vibrations. Specifically, we found that accurately predicting the red-shift and bond elongations associated with hydrogen bonds required a field-dependent contribution to the bonding potential. The dipole derivatives which control the field-dependent bond potential can be computed from electronic structure or directly from the force field as long as the force field has an accurate dipole surface.

By using physically motivated damping functions, we are able to control the polarization catastrophes which have otherwise inhibited accurate polarizable force fields for ions being developed. Indeed, we expect that this force field represents a turning point in the ability to model strong, short-range interactions with physical models. One can, of course, fit error corrections on top of this force field. We expect that achieving such high accuracy with a physical model will enable systems that are otherwise out of reach to be studied. These might include concentrated ionic solutions and divalent ions.

The most recent progress, arguably, has been made by explicitly fitting the terms of the MBE for ion-water interactions.[72,73] This approach has the drawback that generalizing the procedure to multi-component systems, such as a solution containing multiple types of ions, is made difficult by the combinatorial explosion of terms which need to be fit.

Even worse, one can only exclude charge transfer and charge penetration from force fields because these energies are strongly correlated to the Pauli repulsion (see Fig. 2). This correlation is not guaranteed to be consistent between systems, however, which may explain part of the historical difficulty in producing water models which generalize to heterogeneous systems.

Acknowledgement

We acknowledge support from the U.S. National Science Foundation through Grant No. CHE-2313791. Computational resources were provided by the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility operated under Contract DE-AC02-05CH11231.

Supporting Information Available

TODO

References

(1) Xantheas, S. S. Cooperativity and hydrogen bonding network in water clusters. *Chemical Physics* **2000**, *258*, 225–231.

(2) Heindel, J. P.; Xantheas, S. S. The many-body expansion for aqueous systems revisited: II. Alkali metal and halide ion–water interactions. *Journal of chemical theory and computation* **2021**, *17*, 2200–2216.

(3) Herman, K. M.; Heindel, J. P.; Xantheas, S. S. The many-body expansion for aqueous systems revisited: III. Hofmeister ion–water interactions. *Physical Chemistry Chemical Physics* **2021**, *23*, 11196–11210.

(4) Leontyev, I.; Stuchebrukhov, A. Accounting for electronic polarization in non-polarizable force fields. *Physical Chemistry Chemical Physics* **2011**, *13*, 2613–2626.

(5) Bedrov, D.; Piquemal, J.-P.; Borodin, O.; MacKerell Jr, A. D.; Roux, B.; Schroeder, C. Molecular dynamics simulations of ionic liquids and electrolytes using polarizable force fields. *Chemical reviews* **2019**, *119*, 7940–7995.

(6) Szalewicz, K. Symmetry-adapted perturbation theory of intermolecular forces. **2012**, *2*, 254–272.

(7) Patkowski, K. Recent developments in symmetry-adapted perturbation theory. **2020**, e1452.

(8) Mao, Y.; Loipersberger, M.; Horn, P. R.; Das, A.; Demerdash, O.; Levine, D. S.; Veccham, S. P.; Head-Gordon, T.; Head-Gordon, M. From intermolecular interaction energies and observable shifts to component contributions and back again: A tale of variational energy decomposition analysis. **2021**, *72*, 641–666.

(9) Demerdash, O.; Mao, Y.; Liu, T.; Head-Gordon, M.; Head-Gordon, T. Assessing many-body contributions to intermolecular interactions of the AMOEBA force field using energy decomposition analysis of electronic structure calculations. *The Journal of chemical physics* **2017**, *147*, 161721.

(10) Heindel, J. P.; Xantheas, S. S. The many-body expansion for aqueous systems revisited: I. Water–water interactions. *Journal of Chemical Theory and Computation* **2020**, *16*, 6843–6855.

(11) Heindel, J. P.; Herman, K. M.; Xantheas, S. S. Many-body effects in aqueous systems: Synergies between interaction analysis techniques and force field development. *Annual Review of Physical Chemistry* **2023**, *74*, 337–360.

(12) Das, A. K.; Urban, L.; Leven, I.; Loipersberger, M.; Aldossary, A.; Head-Gordon, M.; Head-Gordon, T. Development of an advanced force field for water using variational energy decomposition analysis. *Journal of chemical theory and computation* **2019**, *15*, 5001–5013.

(13) Das, A. K.; Liu, M.; Head-Gordon, T. Development of a many-body force field for aqueous alkali metal and halogen ions: an energy decomposition analysis guided approach. *Journal of Chemical Theory and Computation* **2022**, *18*, 953–967.

(14) Mao, Y.; Loipersberger, M.; Horn, P. R.; Das, A.; Demerdash, O.; Levine, D. S.; Prasad Veccham, S.; Head-Gordon, T.; Head-Gordon, M. From intermolecular interaction energies and observable shifts to component contributions and back again: A tale of variational energy decomposition analysis. *Annual review of physical chemistry* **2021**, *72*, 641–666.

(15) Rackers, J. A.; Silva, R. R.; Wang, Z.; Ponder, J. W. Polarizable water potential derived from a model electron density. *Journal of chemical theory and computation* **2021**, *17*, 7056–7084.

(16) Rick, S. W.; Stuart, S. J.; Berne, B. J. Dynamical fluctuating charge force fields: Application to liquid water. *The Journal of chemical physics* **1994**, *101*, 6141–6156.

(17) Applequist, J. A multipole interaction theory of electric polarization of atomic and molecular assemblies. *The Journal of chemical physics* **1985**, *83*, 809–826.

(18) Stern, H. A.; Rittner, F.; Berne, B.; Friesner, R. A. Combined fluctuating charge and polarizable dipole models: Application to a five-site water potential function. *The Journal of chemical physics* **2001**, *115*, 2237–2251.

(19) Khaliullin, R. Z.; Cobar, E. A.; Lochan, R. C.; Bell, A. T.; Head-Gordon, M. Unravelling the origin of intermolecular interactions using absolutely localized molecular orbitals. *The Journal of Physical Chemistry A* **2007**, *111*, 8753–8765.

(20) Horn, P. R.; Mao, Y.; Head-Gordon, M. Probing non-covalent interactions with a second generation energy decomposition analysis using absolutely localized molecular orbitals. *Physical Chemistry Chemical Physics* **2016**, *18*, 23067–23079.

(21) Boyer, M. A.; Marsalek, O.; Heindel, J. P.; Markland, T. E.; McCoy, A. B.; Xantheas, S. S. Beyond Badger's rule: The origins and generality of the structure–spectra relationship of aqueous hydrogen bonds. *The Journal of Physical Chemistry Letters* **2019**, *10*, 918–924.

(22) Mao, Y.; Horn, P. R.; Head-Gordon, M. Energy decomposition analysis in an adiabatic picture. *Physical Chemistry Chemical Physics* **2017**, *19*, 5944–5958.

(23) Horn, P. R.; Head-Gordon, M. Polarization contributions to intermolecular interactions revisited with fragment electric-field response functions. **2015**, *143*, 114111.

(24) Mardirossian, N.; Head-Gordon, M. Exploring the limit of accuracy for density functionals based on the generalized gradient approximation: Local, global hybrid, and range-separated hybrid functionals with and without dispersion corrections. *J Chem Phys* **2014**, *140*.

(25) Rappoport, D.; Furche, F. Property-optimized Gaussian basis sets for molecular response calculations. *The Journal of chemical physics* **2010**, *133*.

(26) Epifanovsky, E.; Gilbert, A. T. B.; Feng, X.; Lee, J.; Mao, Y.; Mardirossian, N.; Pokhilko, P.; White, A. F.; Coons, M. P.; Dempwolff, A. L.; Gan, Z.; Hait, D.; Horn, P. R.; Jacobson, L. D.; Kaliman, I.; Kussmann, J.; Lange, A. W.; Lao, K. U.; Levine, D. S.; Liu, J.; McKenzie, S. C.; Morrison, A. F.; Nanda, K. D.; Plasser, F.;

Rehn, D. R.; Vidal, M. L.; You, Z.-Q.; Zhu, Y.; Alam, B.; Albrecht, B. J.; Aldossary, A.; Alguire, E.; Andersen, J. H.; Athavale, V.; Barton, D.; Begam, K.; Behn, A.; Bellonzi, N.; Bernard, Y. A.; Berquist, E. J.; Burton, H. G. A.; Carreras, A.; Carter-Fenk, K.; Chakraborty, R.; Chien, A. D.; Closser, K. D.; Cofer-Shabica, V.; Dasgupta, S.; de Wergifosse, M.; Deng, J.; Diedenhofen, M.; Do, H.; Ehlert, S.; Fang, P.-T.; Fatehi, S.; Feng, Q.; Friedhoff, T.; Gayvert, J.; Ge, Q.; Gidofalvi, G.; Goldey, M.; Gomes, J.; González-Espinoza, C. E.; Gulania, S.; Gunina, A. O.; Hanson-Heine, M. W. D.; Harbach, P. H. P.; Hauser, A.; Herbst, M. F.; Hernández Vera, M.; Hodecker, M.; Holden, Z. C.; Houck, S.; Huang, X.; Hui, K.; Huynh, B. C.; Ivanov, M.; Jasz, A.; Ji, H.; Jiang, H.; Kaduk, B.; Kähler, S.; Khistyaev, K.; Kim, J.; Kis, G.; Klunzinger, P.; Koczor-Benda, Z.; Koh, J. H.; Kosenkov, D.; Koulias, L.; Kowalczyk, T.; Krauter, C. M.; Kue, K.; Kunitsa, A.; Kus, T.; Ladjánszki, I.; Landau, A.; Lawler, K. V.; Lefrancois, D.; Lehtola, S., et al. Software for the frontiers of quantum chemistry: An overview of developments in the Q-Chem 5 package. *J. Chem. Phys.* **2021**, *155*, 084801.

(27) Kim, Y. S.; Kim, S. K.; Lee, W. D. Dependence of the closed-shell repulsive interaction on the overlap of the electron densities. *Chemical Physics Letters* **1981**, *80*, 574–575.

(28) Wheatley, R. J.; Price, S. L. An overlap model for estimating the anisotropy of repulsion. *Molecular Physics* **1990**, *69*, 507–533.

(29) Gavezzotti, A. Calculation of intermolecular interaction energies by direct numerical integration over electron densities. I. Electrostatic and polarization energies in molecular crystals. *The Journal of Physical Chemistry B* **2002**, *106*, 4145–4154.

(30) Van Vleet, M. J.; Misquitta, A. J.; Stone, A. J.; Schmidt, J. R. Beyond Born–Mayer: Improved models for short-range repulsion in ab initio force fields. *Journal of chemical theory and computation* **2016**, *12*, 3851–3870.

(31) Van Vleet, M. J.; Misquitta, A. J.; Schmidt, J. New angles on standard force fields: Toward a general approach for treating atomic-level anisotropy. *Journal of Chemical Theory and Computation* **2018**, *14*, 739–758.

(32) Misquitta, A. J.; Stone, A. J.; Fazeli, F. Distributed multipoles from a robust basis-space implementation of the iterated stockholder atoms procedure. *Journal of Chemical Theory and Computation* **2014**, *10*, 5405–5418.

(33) Misquitta, A. J.; Stone, A. J. ISA-Pol: distributed polarizabilities and dispersion models from a basis-space implementation of the iterated stockholder atoms procedure. *Theoretical Chemistry Accounts* **2018**, *137*, 1–20.

(34) Stone, A. J. Distributed multipole analysis, or how to describe a molecular charge distribution. *Chemical Physics Letters* **1981**, *83*, 233–239.

(35) Stone, A. J.; Alderton, M. Distributed multipole analysis: methods and applications. *Molecular Physics* **1985**, *56*, 1047–1064.

(36) Wallqvist, A.; Karlström, G. A new non-empirical force field for computer simulations. *Chem. Scr. A* **1989**, *29*, 1989.

(37) Gordon, J. H. J. M. S. An approximate formula for the intermolecular Pauli repulsion between closed shell molecules. *molecular physics* **1996**, *89*, 1313–1325.

(38) Rackers, J. A.; Ponder, J. W. Classical Pauli repulsion: An anisotropic, atomic multipole model. *The Journal of chemical physics* **2019**, *150*.

(39) Tang, K.; Toennies, J. P. An improved simple model for the van der Waals potential based on universal damping functions for the dispersion coefficients. *The Journal of chemical physics* **1984**, *80*, 3726–3741.

(40) Stone, A. Distributed polarizabilities. *Molecular Physics* **1985**, *56*, 1065–1082.

(41) Ruth Le Sueur, C.; Stone, A. J. Localization methods for distributed polarizabilities. *Molecular Physics* **1994**, *83*, 293–307.

(42) Mortier, W. J.; Ghosh, S. K.; Shankar, S. Electronegativity-equalization method for the calculation of atomic charges in molecules. *Journal of the American Chemical Society* **1986**, *108*, 4315–4320.

(43) Chen, J.; Martínez, T. J. QTPIE: Charge transfer with polarization current equalization. A fluctuating charge model with correct asymptotics. *Chemical physics letters* **2007**, *438*, 315–320.

(44) Chen, J.; Hundertmark, D.; Martínez, T. J. A unified theoretical framework for fluctuating-charge models in atom-space and in bond-space. *The Journal of chemical physics* **2008**, *129*.

(45) Chung, M. K.; Wang, Z.; Rackers, J. A.; Ponder, J. W. Classical Exchange Polarization: An Anisotropic Variable Polarizability Model. *The Journal of Physical Chemistry B* **2022**, *126*, 7579–7594.

(46) Herman, K. M.; Stone, A. J.; Xantheas, S. S. Accurate calculation of many-body energies in water clusters using a classical geometry-dependent induction model. **2023**,

(47) Sheng, X.; Tang, K. T.; Toennies, J. P. A semiempirical potential for alkali halide diatoms with damped interactions I. Rittner potential. *Physical Chemistry Chemical Physics* **2022**, *24*, 24823–24833.

(48) Thirman, J.; Engelage, E.; Huber, S. M.; Head-Gordon, M. Characterizing the interplay of Pauli repulsion, electrostatics, dispersion and charge transfer in halogen bonding with energy decomposition analysis. *Physical Chemistry Chemical Physics* **2018**, *20*, 905–915.

(49) Kirov, M. V.; Fanourgakis, G. S.; Xantheas, S. S. Identifying the most stable networks in polyhedral water clusters. *Chemical Physics Letters* **2008**, *461*, 180–188.

(50) Wang, W.; Yan, D.; Cai, Y.; Xu, D.; Ma, J.; Wang, Q. General Charge Transfer Dipole Model for AMOEBA-Like Force Fields. *Journal of Chemical Theory and Computation* **2023**, *19*, 2518–2534.

(51) Khaliullin, R. Z.; Bell, A. T.; Head-Gordon, M. Analysis of charge transfer effects in molecular complexes based on absolutely localized molecular orbitals. *The Journal of chemical physics* **2008**, *128*.

(52) Khaliullin, R. Z.; Bell, A. T.; Head-Gordon, M. Electron donation in the water–water hydrogen bond. *Chemistry–A European Journal* **2009**, *15*, 851–855.

(53) Liu, C.; Piquemal, J.-P.; Ren, P. Implementation of geometry-dependent charge flux into the polarizable AMOEBA+ potential. *The journal of physical chemistry letters* **2019**, *11*, 419–426.

(54) McCoy, A. B. The role of electrical anharmonicity in the association band in the water spectrum. *The Journal of Physical Chemistry B* **2014**, *118*, 8286–8294.

(55) Misquitta, A. J.; Welch, G. W.; Stone, A. J.; Price, S. L. A first principles prediction of the crystal structure of C6Br2ClFH2. *Chemical Physics Letters* **2008**, *456*, 105–109.

(56) Pracht, P.; Bohle, F.; Grimme, S. Automated exploration of the low-energy chemical space with fast quantum chemical methods. *Physical Chemistry Chemical Physics* **2020**, *22*, 7169–7192.

(57) Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB—An accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions. *Journal of chemical theory and computation* **2019**, *15*, 1652–1671.

(58) Rakshit, A.; Bandyopadhyay, P.; Heindel, J. P.; Xantheas, S. S. Atlas of putative minima and low-lying energy networks of water clusters n= 3–25. *The Journal of chemical physics* **2019**, *151*.

(59) Stone, A. J.; Dullweber, A.; Engkvist, O.; Fraschini, E.; Hodges, M. P.; Meredith, A.; Nutt, D.; Popelier, P.; Wales, D. ORIENT, version 4.6. *University of Cambridge, England* **2002**,

(60) Anatole von Lilienfeld, O.; Tkatchenko, A. Two-and three-body interatomic dispersion energy contributions to binding in molecules and solids. *The Journal of chemical physics* **2010**, *132*.

(61) Fanourgakis, G. S.; Xantheas, S. S. The flexible, polarizable, thole-type interaction potential for water (TTM2-F) revisited. *The Journal of Physical Chemistry A* **2006**, *110*, 4100–4106.

(62) Partridge, H.; Schwenke, D. W. The determination of an accurate isotope dependent potential energy surface for water from extensive ab initio calculations and experimental data. *The Journal of Chemical Physics* **1997**, *106*, 4618–4639.

(63) Sidler, D.; Meuwly, M.; Hamm, P. An efficient water force field calibrated against intermolecular THz and Raman spectra. *The Journal of chemical physics* **2018**, *148*.

(64) Hamm, P. 2D-Raman-THz spectroscopy: A sensitive test of polarizable water models. *The Journal of Chemical Physics* **2014**, *141*.

(65) Herman, K. M.; Xantheas, S. S. An extensive assessment of the performance of pairwise and many-body interaction potentials in reproducing ab initio benchmark binding energies for water clusters n= 2–25. *Physical Chemistry Chemical Physics* **2023**, *25*, 7120–7143.

(66) Zhu, X.; Riera, M.; Bull-Vulpe, E. F.; Paesani, F. MB-pol (2023): Sub-chemical accuracy for water simulations from the gas to the liquid phase. *Journal of Chemical Theory and Computation* **2023**, *19*, 3551–3566.

(67) Badger, R. M. A relation between internuclear distances and bond force constants. *The Journal of Chemical Physics* **1934**, *2*, 128–131.

(68) Roy, S.; Baer, M. D.; Mundy, C. J.; Schenter, G. K. Marcus theory of ion-pairing. *Journal of chemical theory and computation* **2017**, *13*, 3470–3477.

(69) Thole, B. T. Molecular polarizabilities calculated with a modified dipole interaction. *Chemical Physics* **1981**, *59*, 341–350.

(70) Jiao, D.; King, C.; Grossfield, A.; Darden, T. A.; Ren, P. Simulation of Ca2+ and Mg2+ solvation using polarizable atomic multipole potential. *The journal of physical chemistry B* **2006**, *110*, 18553–18559.

(71) Mason, P. E.; Wernersson, E.; Jungwirth, P. Accurate description of aqueous carbonate ions: an effective polarization model verified by neutron scattering. *The Journal of Physical Chemistry B* **2012**, *116*, 8145–8153.

(72) Zhuang, D.; Riera, M.; Zhou, R.; Deary, A.; Paesani, F. Hydration structure of Na+ and K+ ions in solution predicted by data-driven many-body potentials. *The Journal of Physical Chemistry B* **2022**, *126*, 9349–9360.

(73) Caruso, A.; Zhu, X.; Fulton, J. L.; Paesani, F. Accurate modeling of bromide and iodide hydration with data-driven many-body potentials. *The Journal of Physical Chemistry B* **2022**, *126*, 8266–8278.

TOC Graphic

Some journals require a graphical entry for the Table of Contents. This should be laid out "print ready" so that the sizing of the text is correct.
Inside the `tocentry` environment, the font used is Helvetica 8 pt, as required by *Journal of the American Chemical Society*.
The surrounding frame is 9 cm by 3.5 cm, which is the maximum permitted for *Journal of the American Chemical Society* graphical table of content entries. The box will not resize if the content is too big: instead it will overflow the edge of the box.
This box and the associated title will always be printed on a separate page at the end of the document.

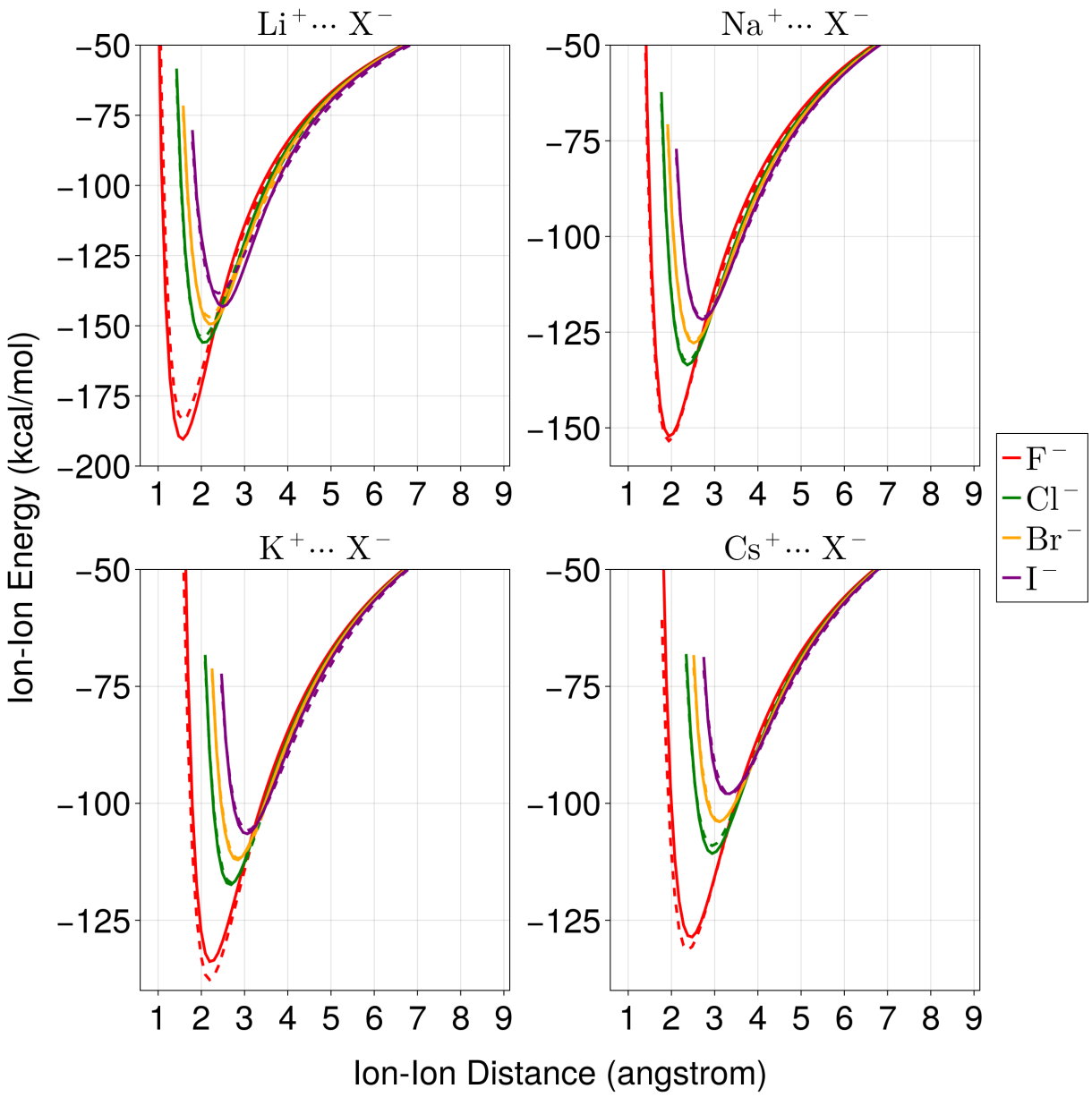


Figure 9: Scans along all ion-ion pairs with FQCT (solid lines) and ωB97X-V/def2-QZVPPD (dashed lines).