

A Status Report on “Gold Standard” Machine-Learned Potentials for Water

Qi Yu,* Chen Qu,* Paul L. Houston,* Apurba Nandi, Priyanka Pandey, Riccardo Conte, and Joel M. Bowman*





Cite This: *J. Phys. Chem. Lett.* 2023, 14, 8077–8087



Read Online

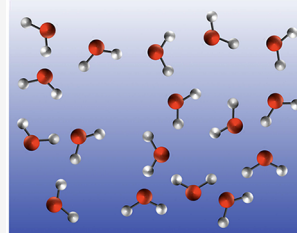
ACCESS |

 Metrics & More

 Article Recommendations

ABSTRACT: Owing to the central importance of water to life as well as its unusual properties, potentials for water have been the subject of extensive research over the past 50 years. Recently, five potentials based on different machine learning approaches have been reported that are at or near the “gold standard” CCSD(T) level of theory. The development of such high-level potentials enables efficient and accurate simulations of water systems using classical and quantum dynamical approaches. This Perspective serves as a status report of these potentials, focusing on their methodology and applications to water systems across different phases. Their performances on the energies of gas phase water clusters, as well as condensed phase structural and dynamical properties, are discussed.

CCSD(T)-level water potential



Arguably, the most important component of nuclear dynamics simulations (broadly defined) of molecules, clusters, and condensed phase molecular systems is the potential energy surface (PES), also termed the force field (FF). We recall that a “first-principles” PES makes explicit use of Born–Oppenheimer separation of electronic and nuclear motion. The former describes the PES and the latter the nuclear dynamics. Since the PES is derived from electronic motion, it represents the most quantum mechanical aspect of the overall simulation. While it would be ideal to describe nuclear dynamics quantum mechanically, this is prohibitive for applications involving condensed phases. Fortunately, classical molecular dynamics (MD) and path-integral motivated MD approaches provide an accurate description of the dynamics currently of interest, such as the diffusion constant. For thermodynamic and dynamic properties of hydrogenic molecules and materials, rigorous path-integral methods are widely employed to capture substantial nuclear quantum effects. For non-hydrogenic materials, classical MD is generally adequate. In any case, the ultimate accuracy of dynamics and statistical mechanical simulations depends critically on the accuracy of the PES. We conclude this introductory paragraph by noting that the current “gold standard” quantum chemistry approach for the PES is the coupled-cluster method, with CCSD(T) being the current workhorse version. So this method is used synonymously with “gold standard”. However, this level of electronic structure theory is also very computer-intensive with scaling of $O(N^7)$, where N is the size of the electronic space. So clearly this method cannot currently be

extensively applied to molecules or clusters of more than roughly ten atoms.

These general remarks are applicable to PESs describing water, from the smallest water dimer cluster to systems encompassing hundreds of water monomers. The motivation behind simulating the unique properties of water is widely recognized and does not require reiterating. Thus, it is not surprising that there are over 50 PESs/FFs available that describe noncovalent water interactions. One general aim of these PESs is to treat an arbitrary number of monomers and perform the use of simple functional forms. The historically important ones, which are still the mostly widely used, are based on simple electrostatic models with two-body non-covalent potentials with parameters determined empirically. These potentials are limited in the underlying chemical physics and, of course, are not “first-principles”. Such approaches have been intensively investigated in the past decade using density functional theory (DFT) and initially using direct dynamics, which by-passes the need for a functional representation of the PES.^{1–4} DFT is the only electronic structure method that is feasible for these calculations; however, it still involves a large cost in computational effort. Some interesting efforts to

Received: June 30, 2023

Accepted: August 28, 2023

improve DFT using the many-body CCSD(T)-based MB-pol have been reported.⁵ DFT-based machine learning potentials have also recently appeared.^{6–10}

Ideally, one would like to have the best of both worlds, namely, an analytical PES for water, based on *ab initio* electronic energies (and possibly forces). The “holy grail” of this effort would be for the energies to be at the gold-standard CCSD(T) level. There has been great progress toward achieving this goal very recently using machine-learning approaches, and that is the focus of this Perspective. To be clear about this progress, and to preface the outline of the paper, we state at the outset that the progress is based on two different approaches, neither of which is based on direct fitting of CCSD(T) energies (and forces) for tens or hundreds of water monomers, as this would be totally infeasible to do. One approach is the many-body expansion, and the other, more recent one is based on transfer learning to bring a DFT-based potential close to the CCSD(T) level. We begin with the many-body expansion approach.

MANY-BODY APPROACHES

q-AQUA. To begin, we note that the strict many-body expansion (MBE) for the total energy of N water monomers is given by

$$V(1, \dots, N) = \sum_{i=1}^N V_{1-b}(i) + \sum_{i>j}^N V_{2-b}(i, j) + \sum_{i>j>k}^N V_{3-b}(i, j, k) + \sum_{i>j>k>l}^N V_{4-b}(i, j, k, l) + \dots \quad (1)$$

where V_{1-b} denotes the 1-body potential, i.e., the potential for the isolated monomer, and V_{2-b} , V_{3-b} , V_{4-b} , etc. are the *intrinsic* 2-, 3-, 4-body, etc. interactions. This representation, truncated at the 4-b term, was used extensively in seminal studies of the binding energies at the MP2 and MP4 level of theory for moderate sized clusters, from the dimer to the hexamer in 1994.¹¹ From this early work, it was concluded that the expansion converged at the 3-body level and that 4-body terms were “negligible”. This conclusion was essentially confirmed in recent work by Heindel and Xantheas,^{12,13} who extended the analysis to clusters as large as the 21-mer. With basis set superposition errors eliminated, it was concluded that the 4-body interactions account for at most one percent of the total binding energy. The slight change in the assessment of the four-body interaction from “negligible” to “very small” is probably a result of the increasing number of 4-b terms for the larger clusters studied in the more recent study. This important aspect of the MBE was discussed in detail in a 2010 *Frontiers Article*.¹⁴ It was noted that, although the number of four-body interactions for N monomers goes as $O(N^4)$, this is quantitatively a major overestimate because the four-body interaction decays quickly with monomer separation and most four-body interactions are negligible.

In 2010, the precise fitting of 2-body (six atoms) and 3-body (nine atoms) interactions became achievable through the application of permutationally invariant polynomial (PIP) machine-learning regression.¹⁵ Recently, the PIP approach was extended to handle systems with more than 10 atoms,¹⁶ and this extension was successfully applied to the 12-atom 4-body water interaction.¹⁷ This advancement enabled the development of a many-body potential, termed as q-AQUA, which incorporates up to the 4-body interaction using precise PIP

fitting of extensive data sets of CCSD(T) energies.¹⁸ For further details regarding the data sets, fitting details, precision metrics, and numerous tests, we direct interested readers to the aforementioned paper. Before discussing the key performance highlights of the q-AQUA potential for water clusters and liquid water, we remark on the strategy adopted to describe the n -body potentials in the long-range regime. In the case of the 2-body interaction, the long-range behavior is represented by the dipole–dipole interaction, utilizing the dipole moment of the isolated water monomer. This approach is justified, since the induction effects that enhance the dipole moment predominantly occur in the short-range regime. The dipole–dipole interaction was obtained based on a high-level dipole moment surface for the flexible water monomer.¹⁹ Instead of using the known expression for this interaction, we used the widely used Coulomb expression

$$V_{2-b}^{\text{long-range}} = \sum_{i=1}^3 \sum_{j=1}^3 \frac{q_i q_j}{r_{ij}} \quad (2)$$

where q_i and q_j are the partial charges on the i th atom of monomer 1 and the j th atom of monomer 2, and r_{ij} is the distance between the atoms. The partial charge on each atom of the water monomer is uniquely obtained from the monomer dipole moment surface. It is perhaps worth emphasizing that these partial charges are not empirically determined, as they are in the widely used rigid-monomer TIP4P and SPC potentials^{20,21} and their extensions for flexible monomers, q-SPC²² and q-TIP4P/F.²³ Of course this interaction must be “damped” to zero in the range where it matches direct CCSD(T) 2-b energies; details of this damping are given in ref 18.

The long-range 3-body and 4-body interactions are due to induction, which can in principle be obtained using standard, but costly approaches. Since these interactions are both weaker and shorter ranged than the 2-body interaction, our approach was to just rely on fitting the CCSD(T) energies directly and then damp the fit to zero over a finite range. To demonstrate these effects, we show comparisons of the q-AQUA 2-body, 3-body, and 4-body potentials with direct CCSD(T) energies for attractive 1d cuts in Figures 1 and 2, respectively. First, consider the 2-body comparison in Figure 1: the dipole–dipole interaction merges with the CCSD(T) energies at long-range

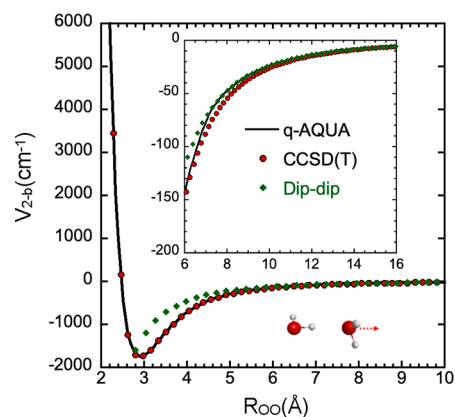


Figure 1. Comparison of the 2-body fit (black line) and direct CCSD(T) energies (red circles) for an attractive cut. The isolated monomer dipole–dipole interaction is also indicated (green diamonds).

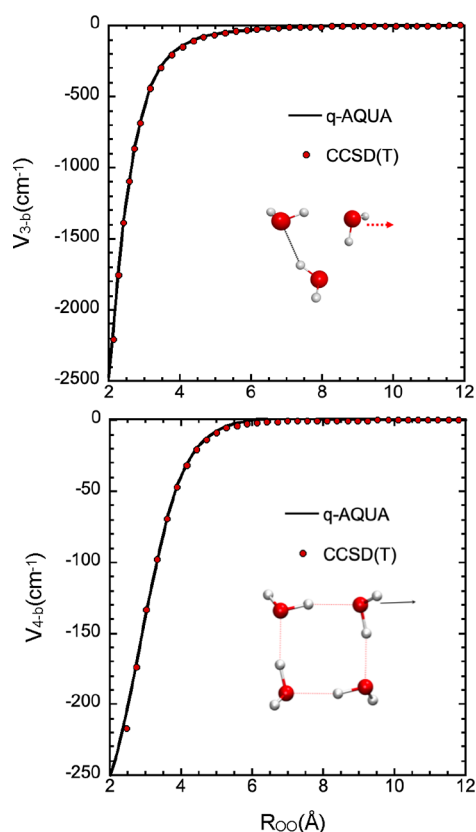


Figure 2. Comparison of 3-b (top) and 4-b (bottom) q-AQUA potentials with direct CCSD(T) energies for an attractive cut.

but becomes less accurate at short monomer distances. Interestingly, and as expected, it increasingly underestimates the CCSD(T) energies at a shorter range where polarization effects become increasingly important. Exchange–repulsion interaction effects dominate at distances less than around 3 Å, and the rapid rise in energy occurs at around 2.6 Å. In q-AQUA, the PIP 2-body fit is switched smoothly to the dipole–dipole interaction in the range of 6.5–7.8 Å. Thus, q-AQUA is in excellent agreement with the CCSD(T) energies in both short and long ranges.

Corresponding comparisons for the q-AQUA 3-body and 4-body interactions are shown in Figure 2. Several important points can be highlighted regarding these interactions. The energies associated with the 3-body and 4-body interactions are smaller in magnitude compared to the 2-body interactions. The magnitude of 4-body interactions is generally smaller than the 3-body ones, which also aligns with expectations. Additionally, the range of these interactions is shorter than that of the 2-body interactions, as anticipated. The q-AQUA fits are specifically dampened to zero over different ranges for the 3-body and 4-body interactions, allowing for flexibility and user-defined control over these ranges. These potentials exhibit a rapid shift toward negative values at shorter distances, particularly at 2 Å, where the 2-body interaction is highly repulsive but the 3-body and 4-body interactions are both attractive. The latter does become repulsive at shorter distances.¹⁷ Lastly, it is worth noting that the 1-body term in q-AQUA corresponds to the spectroscopically accurate potential for H₂O.²⁴

It is important to emphasize that the precise fitting of 2-body and 3-body electronic energies for flexible monomers was done

more than a decade ago, beginning with the WHBB potential,²⁵ which fit thousands of 2-b CCSD(T) and 3-b MP2 energies using PIPs, followed by MB-pol^{26,27} which fit thousands of 2-b and 3-b CCSD(T) energies using PIPs and CC-pol-flex which fit thousands of 2-b CCSD(T) energies using a multiparameter, physically based functional form for the potential.²⁸ MB-pol is the most accurate of this group of three potentials as it uses CCSD(T) data for the 2-b and 3-b energies. It has been used successfully in many applications.^{27,29} As we mention below, MB-pol is a “hybrid” water potential which makes corrections to the TTM4-F potential.³⁰ Finally, we note that the size and extent of the CCSD(T) energies used in the new q-AQUA potential are larger than those used in previous potentials. Also, we note again that q-AQUA includes CCSD(T) 4-b energies, unlike those of earlier potentials.

To summarize thus far, q-AQUA is currently the only strictly CCSD(T)-based many (up to 4)-body potential for water that accommodates an arbitrary number of monomers. The feasibility to obtain tens of thousands of *ab initio* CCSD(T) energies for 2–4 body interactions together with precise and efficient PIP fits to the data sets has made this possible. These PIP fits also provide analytical gradients, enhancing the computational efficiency of the approach. q-AQUA has demonstrated its accuracy in describing the energies of hexamer isomers and other water clusters.¹⁸ Additionally, properties of liquid water, such as radial distribution functions and self-diffusion constants, obtained from NVT classical and path integral simulations employing q-AQUA, have exhibited an accuracy comparable to those obtained using other potentials.

q-AQUA-pol. Δ -Machine learning is a general term in ML that describes a method to bring an ML representation of a property, determined using low-level theory, to be close to high-level theory, e.g., the CCSD(T) level. For example, the property might be a DFT-based potential energy surface, or it might be a polarizable water force field. Δ -Machine learning has been extensively applied to DFT-based potentials and energies.^{31–37} This subsection describes the very recent realizations of this goal for a water potential, where the underlying low-level potential is the well-known polarizable TTM3-F potential.³⁸ Our first effort in this direction³⁹ utilized TTM2.1, a semiempirical, many-body force field for water;⁴⁰ however, no applications to condensed phase properties were explored in that study. More recently, we utilized this approach to correct the TTM3-F potential³⁸ up to 4-body terms. This latest water potential, derived from the same data sets used to develop q-AQUA, is referred to as q-AQUA-pol.⁴¹

The expression for the Δ -machine learning q-AQUA-pol potential is a specific example of the general one given previously,³⁹ namely

$$V_{\text{q-AQUA-pol}} = V_{\text{TTM3-F}} + \sum_{i>j}^N \Delta V_{2-b}(i, j) + \sum_{i>j>k}^N \Delta V_{3-b}(i, j, k) + \sum_{i>j>k>l}^N \Delta V_{4-b}(i, j, k, l) + \dots \quad (3)$$

These correction terms, ΔV_{n-b} , are the differences between the CCSD(T) and TTM3-F n -body energies. Ideally, these are short-ranged if TTM3-F³⁸ does provide a quantitatively accurate description of the long-range interactions. As has

been shown in ref 41, the polarizable TTM3-F force field behaves quantitatively accurately in the long-range against CCSD(T) reference data. Thus, in q-AQUA-pol, we only conducted PIP fits to short-range 2-, 3-, and 4-body interaction corrections. The data set used in q-AQUA-pol is the same as that in q-AQUA where the electronic energies extend to sufficiently high energies. The resulting PES reaches energies well beyond the zero-point energy. This enables the possibilities of quantum simulations to investigate the structural and transport properties of water, ranging from clusters to the condensed phase. Finally, the PES should be invariant with respect to the permutation of monomers, and each monomer should also be invariant with respect to the interchange of two H atoms. This requirement is inherently incorporated through the use of permutationally invariant polynomials in ML fits.

We note that MB-pol^{26,29} also falls under this category of MB approach with a polarizable FF, where the FF is TTM4-F.³⁰ The latest version, MB-pol(2023),⁴² also makes use of extensive CCSD(T) energies from the q-AQUA data set.¹⁸ This updated version is indeed more accurate than previous versions of MB-pol and is certainly at the state-of-the-art.

Selected Results from q-AQUA and q-AQUA-pol. Here we examine the q-AQUA and q-AQUA-pol potentials for simulations of bulk water properties, as well as the energies of the centrally important isomers of the water hexamer. Specifically, classical molecular dynamics (MD), path integral molecular dynamics (PIMD), and ring polymer molecular dynamics (RPMD)^{23,43} were used to calculate both static and dynamic properties of liquid water. All the MD simulations were performed with the i-PI software.⁴⁴ The computational details about these calculations are provided in the Supporting Information of the original papers.^{18,41}

In panel A of Figure 3, the OO radial distribution functions (RDF) obtained from NVT MD and PIMD simulations at 298 K using the q-AQUA potential are presented. The classical MD

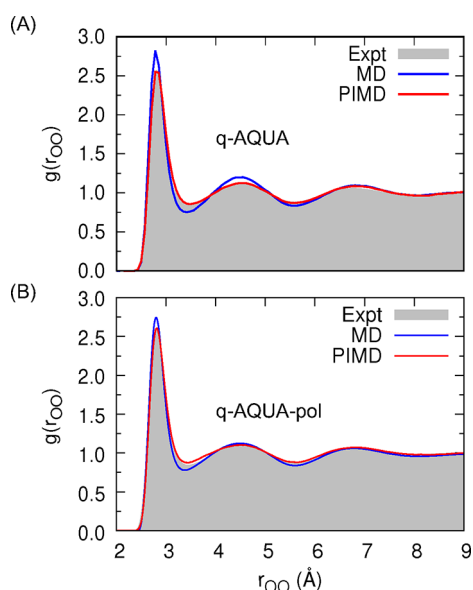


Figure 3. OO radial distribution function from classical (blue) and path integral (red) molecular dynamics simulations at 298 K using q-AQUA (panel A) and q-AQUA-pol potentials (panel B). The experimental data are from refs 45 and 46. Simulation data are from refs 18 and 41.

simulations yield reasonable predictions for the peak positions of the RDF compared to experimental measurements. However, there is a discrepancy in the amplitudes of the peaks that do not align well with the experimental data. On the other hand, the OO RDF calculated from PIMD simulations exhibits significantly improved agreement with both the peak positions and amplitudes observed in experiments. Similar trends are observed in panel B of Figure 3 from NPT simulations at 298 K and 1 atm, utilizing the q-AQUA-pol potential. The classical MD approach tends to predict more localized water structures, while the inclusion of nuclear quantum effects in PIMD simulations results in a more delocalized OO radial distribution function. The RDFs obtained with just the 2-b terms in q-AQUA and q-AQUA-pol were reported in refs 18 and 41. In those cases, inaccuracies were observed, particularly with the second hydration peak appearing at larger OO distances. Interestingly, as the level of truncation increases from 2-body to 3-body and finally to 4-body interactions, the peaks in the RDFs shift toward shorter OO distances. This suggests the presence of an effective additional attraction when considering higher-order n -body interactions. Figure 4 shows the oxygen–oxygen–oxygen

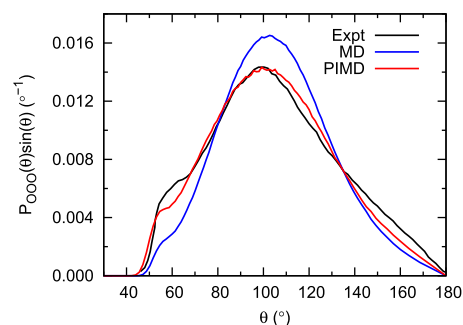


Figure 4. Oxygen–oxygen–oxygen triplet angular distribution functions from classical and path integral molecular dynamics simulations at 298 K and 1 atm using the q-AQUA-pol potential. The experimental data are taken from ref 48. Reproduced from ref 41. Copyright 2023 American Chemical Society.

(OOO) triplet angular distribution functions of liquid water at 298 K and 1 atm, obtained from simulations with the q-AQUA-pol potential. Consistent with the observations in the OO radial distribution functions, the results from PIMD simulations exhibit considerably better agreement with experimental data compared to the classical MD results. The inclusion of nuclear quantum effects in PIMD simulations improves the accuracy of capturing the triplet angular distribution of water molecules, highlighting the importance of considering quantum effects in accurately describing the behavior of liquid water.

Additional NPT classical MD and PIMD simulations using q-AQUA-pol potential were performed at 1 atm with temperatures from 238 to 340 K. The calculated densities of liquid water are presented in Figure 5. In the high-temperature region, both classical MD and PIMD simulations predict densities that reasonably agree with experimental measurements. As the temperature decreases, the differences between the classical MD predictions and experimental data become more pronounced, although the overall trend still aligns well. This is expected since classical MD neglects the consideration of nuclear quantum effects, which have been demonstrated to be important in water density simulations.⁴⁹ To illustrate the

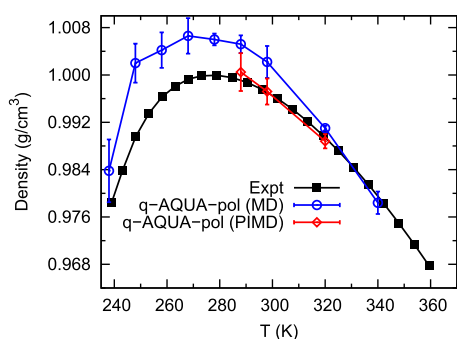


Figure 5. Temperature-dependence of the density of liquid water at 1 atm. The experimental data are taken from refs 50 and 51. Reproduced from ref 41. Copyright 2023 American Chemical Society.

impact of nuclear quantum effects, PIMD NPT simulations were conducted at 288, 298, and 320 K. As observed in Figure 5, the predicted water densities from PIMD simulations exhibit quantitative agreement with experimental data. Furthermore, the improvement over classical MD results becomes increasingly significant as the temperature decreases. Again, this highlights the crucial role of nuclear quantum effects in accurately describing the density of water and emphasizes the advantage of employing PIMD simulations to capture these effects.

The tetrahedral order parameter distribution as a function of the temperature is depicted in Figure 6. As seen, as

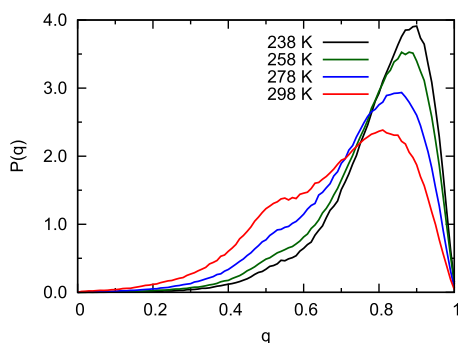


Figure 6. Probability distribution of the tetrahedral order parameter q at different temperatures from classical MD simulations. Reproduced from ref 41. Copyright 2023 American Chemical Society.

temperature decreases to the supercooled liquid regime, the distribution narrows significantly, indicating more tetrahedral H-bonding structures. In the next section, we show this distribution at 298 K for q-AQUA and one NN-TL potential along with q-AQUA-pol. We also defer presenting results for the self-diffusion constant using q-AQUA and q-AQUA-pol to that section.

Next, we consider the accuracy of q-AQUA and q-AQUA-pol for the centrally important water hexamer binding energies shown in Figure 7. Panel A shows the electronic dissociation energies for eight isomers. It is evident that q-AQUA-pol achieves nearly perfect agreement with the CCSD(T)/CBS results, which is an unprecedented level of accuracy for a water potential. q-AQUA is also highly accurate, except for the three highest energy isomers. Both potentials exhibit significantly higher accuracy compared to TTM3-F.³⁸ However, it should be noted that q-AQUA-pol, being a Δ -ML correction to TTM3-F, benefits from the many-body polarization effects

beyond the 4-body level present in TTM3-F, which accounts for its improved accuracy for the chair and two boat isomers. Panels B, C, and D provide detailed comparisons of the many-body contributions to the electronic dissociation energies up to the 4-b level. It can be observed that the 4-b energies have the largest magnitude for the chair and boat isomers.

In summary thus far, both q-AQUA and q-AQUA-pol demonstrate high accuracy in predicting condensed phase properties of water as well as the binding energies of water hexamer isomers. However, q-AQUA-pol incorporates polarization effects beyond the 4-b level through TTM3-F, thereby enhancing its accuracy.

Before discussing other approaches to bring a low-level water potential to the CCSD(T) level, it is of interest to examine the corrections to TTM3-F term by term. This is shown for the OO radial distribution function at 298 K using NVT calculations (Figure 8). As seen, the RDF using TTM3-F is in good agreement with experiment. However, adding the CCSD(T) 2-b correction ΔV_{2-b} results in an RDF that is significantly less accurate than the TTM3-F one. By adding the 3-b correction agreement with experiment is very good. The 4-b correction makes a small contribution to this property. These results are shown in order to make the following point. Namely, we advise caution when correcting a force field using just a 2-b *ab initio* correction, even at the CCSD(T) level. Doing so may result in a decrease in accuracy. Based on extensive work on the FF for water, for which *ab initio* 3-b interactions are essential, we recommend investigating these interactions for correcting FFs.

TRANSFER AND Δ -LEARNING

Δ -Learning has also been applied by another group to develop a coupled-cluster level PES of water,⁵³ although it differs from q-AQUA-pol and MB-pol(2023) in that the Δ is not learned in the many-body manner. In addition to Δ -learning, transfer learning (TL) is a method to modify the optimized parameters of a machine-learned potential trained on, say, low-level DFT energies/forces using a relatively small training set of CCSD(T) energies.⁵⁴ This approach was recently applied for a neural network (NN) potential for water.⁵⁵ Both PESs use the basic Tersoff form of the potential, i.e., as a sum of atomic energies each of which is element-specific and given by a general density expression that contains radial and angular components.⁵⁶ Behler and Parrinello made a major significant modification to this basic form by training a NN for the atomic energy using DFT energies.⁵⁷

In the recent TL PES, this form of the potential was trained on DFT or Hartree–Fock energies and forces of 16 water monomers in a periodic box based on configurations selected using an active learning procedure.⁵⁵ Then the transfer-learning approach was applied using a small data set of (periodic) CCSD(T) energies for 16 monomers. The NN- Δ -ML water potential⁵³ was first trained on domain-based local pair natural orbital (DLPNO) MP2 energies and forces, using finite clusters of 64 water monomers sampled from AIMD. Then the difference between the DLPNO-CCSD(T) and DLPNO-MP2 interaction is fit by a second NN as the correction to the MP2-level NN. This PES additionally makes use of fixed-charged electrostatic interactions that are damped as usual in the short-range and repulsive Yukawa potentials. The charges are -0.8 for O and 0.4 for H, exactly the ones in the first generation (1981) fixed charged model, TIP.⁵⁸ The interested reader is referred to refs 53 and 55 for more details.

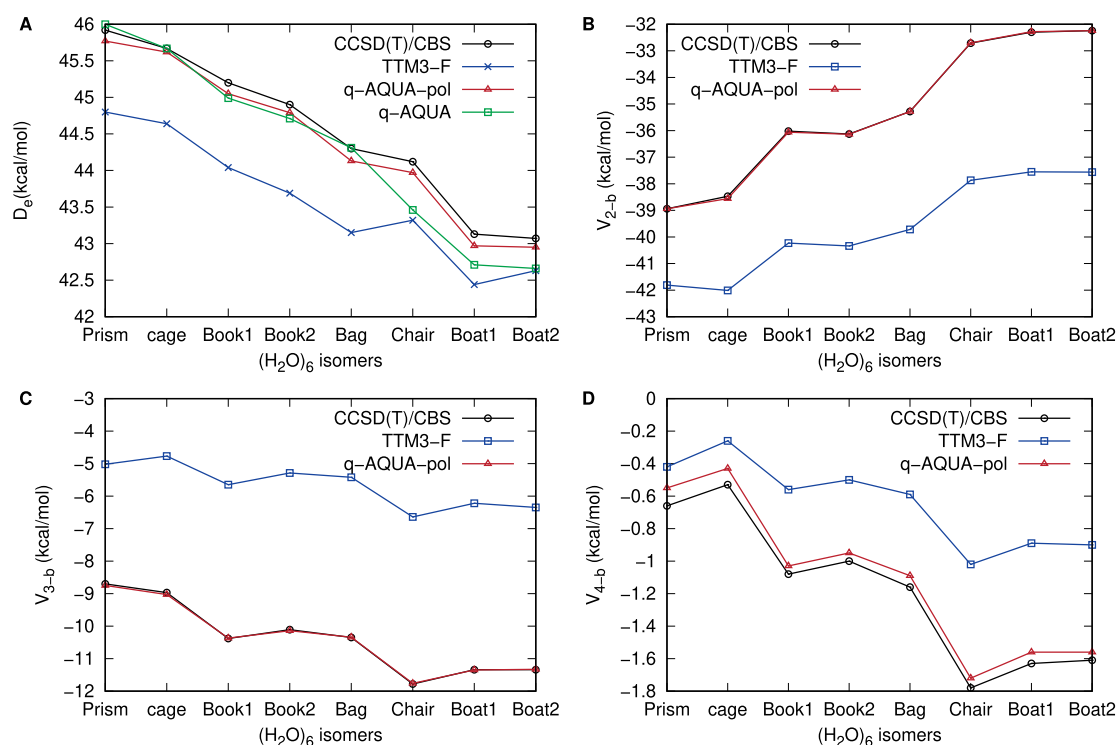


Figure 7. Binding energies (A), 2-body energies (B), 3-body energies (C), and 4-body energies (D) for water hexamer isomers from TTM3-F, q-AQUA (panel A), q-AQUA-pol, and benchmark CCSD(T) calculations (data taken from ref 52).

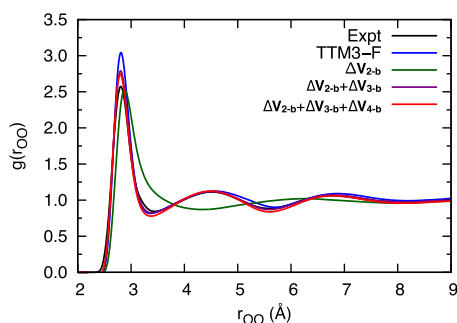


Figure 8. Effects of 2-b, 3-b, and 4-b corrections on the OO radial distribution function of liquid water at 298 K from classical MD simulations. The experimental data are taken from refs 45 and 46.

The differences between these two approaches and the many-body expansion (MBE) approach used in q-AQUA, q-AQUA-pol, MB-pol, and MB-pol(2023) are noteworthy. These approaches directly fit the energies and forces of relatively large water clusters (16 monomers in a periodic box and 64 monomers), sampled from large-scale AIMD or on-the-fly PIMD simulations. In contrast, the MBE approach involves fitting electronic energies of small clusters, specifically the dimer of the 2-body interaction, the trimer for the 3-body interaction, and the tetramer for the 4-body interaction. These small cluster data sets are extensive, and the associated energies are typically computed using the CCSD(T) method. The data sets for small water clusters cover a broader range of energies than those obtained from AIMD and PIMD simulations at around 300 K, which are used in those fits. While this difference may not be crucial for MD and PIMD simulations conducted at room temperature, it can become significant for quantum simulations that explore high-energy regions of the PES. Interested readers can refer to a perspective on the

differences in data sets obtained from MD simulations at 300 K and higher energies for more details on this topic.⁵⁹ Overall, the Δ -ML and TL approaches directly incorporate information from large-scale simulations, while the MBE approach focuses on fitting small cluster energies at the CCSD(T) level. Both approaches have their advantages and considerations, and the choice is dependent on the specific requirements in the fitting architecture.

The above remarks notwithstanding, it does appear that the MBE and Δ -ML/TL potentials produce results of similar accuracy for several condensed phase properties. For example, Figure 9 demonstrates that both the MBE potentials (q-AQUA and q-AQUA-pol) and the TL potential (NN-TL) yield similar descriptions of the local structure of water molecules, characterized by the tetrahedral order parameter. Such good agreement among all three potentials indicates convergence between the MBE and TL approaches in predicting liquid-phase water structures. As shown in Table 1, the self-diffusion

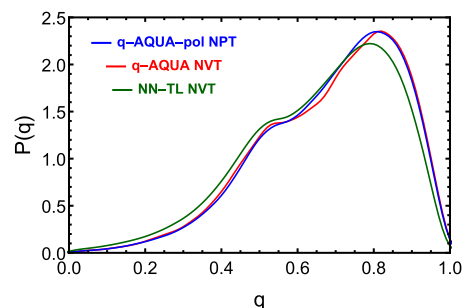


Figure 9. Probability distribution of the tetrahedral order parameter q at room temperature from q-AQUA-pol, q-AQUA, and recent NN-transfer-learned potential with CCSD(T) accuracy.⁵⁵ Reproduced from ref 41. Copyright 2023 American Chemical Society.

Table 1. Self-Diffusion Coefficient, D ($\text{\AA}^2/\text{ps}$), of Liquid Water at 298 K from Classical and Quantum Simulations with Different Potentials

| Potential | Classical | Path integral | Expt. ^a |
|-------------------------------|---------------|---------------|--------------------|
| q-AQUA | 0.145 ± 0.012 | 0.226 ± 0.020 | 0.230 |
| q-AQUA-pol | 0.185 ± 0.004 | 0.233 ± 0.027 | |
| NN- Δ -ML ^b | | 0.244 ± 0.002 | |
| NN-TL ^c | 0.221 ± 0.006 | 0.230 ± 0.008 | |

^aFrom refs 60 and 61. ^bFrom ref 53. ^cFrom ref 55, and simulations were conducted at 300 K.

coefficients, a measure of the mobility of water molecules, also exhibit good agreement with experimental values for these different types of potentials. It is important to note that the nuclear quantum effects play a significant role in the dynamical properties of water, including self-diffusion. In the case of q-AQUA and q-AQUA-pol potentials, the inclusion of NQE through the TRPMD method leads to significantly larger self-diffusion coefficients compared to classical MD simulations. This is consistent with the understanding that NQE weakens the hydrogen bonding network and promotes molecular motion in liquid water. Interestingly, the NN-TL potential exhibits a less pronounced impact of NQE on self-diffusion coefficients. This may arise from the competing effects of intramolecular and intermolecular NQE in liquid water, as discussed in ref 23. The intramolecular NQE weakens covalent bonds and enhances molecular dipole moments, while intermolecular NQE weakens hydrogen bonds and forms a less structured hydrogen-bonded network. In the case of q-AQUA and q-AQUA-pol, the dominance of intermolecular NQE leads to larger diffusion coefficients. However, in the NN-TL potential, there appears to be a balance between these two types of NQE, resulting in only a slight increase in diffusion coefficients.

Thus far, we have shown that the efforts at transfer learning were successful in predicting condensed-phase properties of water including radial distribution function, self-diffusion coefficient, and tetrahedral order distribution functions. The specific choice between the MBE and TL approaches may depend on factors such as computational efficiency, desired level of accuracy, and the ability to capture different aspects of water's behavior. The computational scaling advantages of NN-TL potentials, achieved through the use of the sum of atomic energies, make them appealing for large molecular and materials systems. However, there are certain aspects of these new potentials that have yet to be fully explored, such as their performance on water clusters. Experimental studies have extensively investigated water clusters, providing important tests for theoretical models and yielding valuable insights into the nature of hydrogen bonding in water.^{62–66} It is reasonable to speculate that the accuracy of the NN-TL potentials may not be on par with that of q-AQUA and q-AQUA-pol when it comes to describing water clusters. This is primarily due to the lack of direct training on clusters in the DFT and TL fitting process. The absence of explicit consideration for the unique properties and behavior of water clusters during PES development could potentially limit the accuracy of NN-TL potentials in this regime. To gain a more comprehensive understanding of the performance of NN-TL potentials on water clusters, it would be beneficial to assess their capabilities in reproducing the structural, energetic, and spectroscopic properties observed in the experimental studies.

■ TIMING

Finally, it is worth considering the computational cost associated with using q-AQUA and q-AQUA-pol potentials for dynamics simulations of a 256-water system. All tests were conducted using single or multiple cores of a 2.4 GHz Intel Xeon processor. The timing results for the q-AQUA and q-AQUA-pol potentials are summarized in Table 2.

Table 2. Computational Costs of the q-AQUA and q-AQUA-pol Potentials for Energy and Gradient Calculations of a 256-Water System

| | | q-AQUA | | | |
|------------------|--------|---------------------|--------------------|-------------------------------|--------------------|
| | | Time for energy (s) | | Time for energy +gradient (s) | |
| Component | Number | 1 core | 8 cores | 1 core | 8 cores |
| V_{1-b} | 256 | 0.002 | 0.002 ^a | 0.003 | 0.003 ^a |
| V_{2-b} | 32640 | 0.23 | 0.02 | 0.72 | 0.08 |
| V_{3-b} | 84051 | 0.42 | 0.05 | 1.94 | 0.26 |
| V_{4-b} | 115922 | 1.26 | 0.17 | 4.34 | 0.52 |
| Total | | 2.00 | 0.35 | 7.12 | 0.97 |
| | | q-AQUA-pol | | | |
| | | Time for energy (s) | | Time for energy +gradient (s) | |
| Component | Number | 1 core | 8 cores | 1 core | 8 cores |
| TTM3-F | / | NA | NA ^b | 0.27 | 0.07 ^b |
| ΔV_{2-b} | 3816 | 0.11 | 0.02 | 0.36 | 0.06 |
| ΔV_{3-b} | 20790 | 0.65 | 0.23 | 2.20 | 0.39 |
| ΔV_{4-b} | 28786 | 0.30 | 0.04 | 1.05 | 0.14 |
| Total | | 1.33 | 0.36 | 3.88 | 0.66 |

^a1-b terms are not parallelized. ^bThe current TTM3-F force field code calculates the energy and gradient simultaneously by default.

As shown in Table 2, the computation of 4-body interactions is the most computationally intensive part of the q-AQUA potential, accounting for more than half of the computation time in both energy and gradient calculations. It is important to note that the number of final calculated four-body terms, 115922, is only a small fraction of the total four-body interactions, which would be on the order of 256^4 in a 256-water system. This is due to the fast damping of 4-body interaction to 0 when the monomer distances exceed 7.5 \AA in the q-AQUA potential. By directly dampening high-order interactions beyond this distance, significant computational savings are achieved. Additionally, efficient OpenMP parallelization allows for a speed-up of up to 6 times when using 8 cores. Furthermore, the cost of gradient calculations is observed to be approximately twice the cost of energy calculations, benefiting from the efficient implementation of reverse differentiation in the PIP approach.¹⁶

Similar trends in the computational cost are observed for the q-AQUA-pol potential. The overall cost of energy or energy +gradient calculations with q-AQUA-pol is approximately half that of q-AQUA. This is primarily because the total number of correction terms (ΔV_{2-b} , ΔV_{3-b} , ΔV_{4-b}) is much smaller than the corresponding n -body interactions in q-AQUA. The implementation of finite-range switching functions enables the selection of a short O–O region for corrections, resulting in a reduced number of these interaction terms. For example, only 20790 ΔV_{3-b} terms need to be calculated with a maximum O–O distance smaller than 7.0 \AA . The accuracy of the TTM3-F force field in describing long-range many-body interactions

allows for a safe selection of this short O–O region for corrections.

It is evident from Table 2 that the calculation of ΔV_{3-b} interaction terms accounts for the majority of the total computation time in q-AQUA-pol. A smaller cutoff range leads to a significant reduction in the number of 3-b correction terms. However, this reduction comes at the expense of some accuracy. We investigate the impact of the cutoff maximum O–O distance of 3-b interactions on the cost and accuracy of the q-AQUA-pol potential. The calculated computational cost associated with energy and gradient calculations in q-AQUA-pol for the same 256-water system over a range of maximum O–O distances are given in Table 3. Note that the cost for the

Table 3. Computational Cost of the 3-b Interaction of q-AQUA-pol Potential for Energy and Gradient Calculations of a 256-Water System

| Maximum r_{OO} (Å) | ΔV_{3-b} term | Time for energy (s) (1 core) | Time for energy+gradient (s) (1 core) |
|----------------------|-----------------------|------------------------------|---------------------------------------|
| 6.0 | 8266 | 0.30 | 0.99 |
| 6.3 | 10972 | 0.37 | 1.25 |
| 6.5 | 13369 | 0.43 | 1.49 |
| 6.7 | 15825 | 0.50 | 1.73 |
| 7.0 | 20790 | 0.65 | 2.20 |

ΔV_{2-b} correction terms and TTM3-F force field remains the same as in Table 2 throughout these calculations. It can be seen from Table 3 that if we only take 3-b interactions with a maximum O–O distance smaller than 6.0 Å into account, the total number of ΔV_{3-b} correction terms is reduced to 8266, and the cost of 3-b interaction calculation for the energy alone and for the energy and gradient speed up by a factor of 2.2.

To investigate the influence of the cutoff distance on condensed phase properties, we conducted classical MD simulations of liquid water at a temperature of 298 K and pressure of 1 atm using the q-AQUA-pol potential. Figure 10

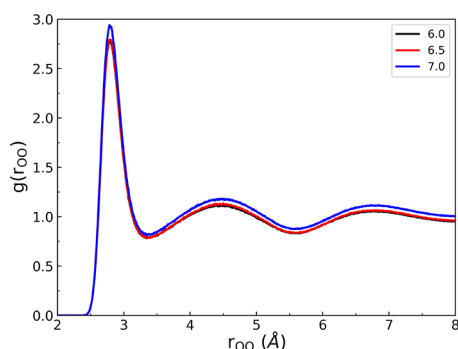


Figure 10. OO radial distribution function from classical MD simulations at 298 K using q-AQUA-pol potential for three range parameters indicated (Å) for the 3-b interaction.

displays the radial distribution functions (RDFs) for oxygen–oxygen (O–O) pairs obtained from the MD simulations with three different cutoff distances for the three-body (3-b) interactions. Although the overall RDF for a cutoff distance of 7 Å exhibits a slightly higher magnitude compared to the other two cutoff values, the positions of the first and second peaks, which correspond to H-bonded neighbors and non-H-bonded water molecules, respectively, remain consistent across all three cutoff distances. This indicates that the local structure

and H-bonding network of liquid water are preserved regardless of the range chosen for the 3-b interactions. Furthermore, we analyzed the oxygen–oxygen–oxygen (O–O–O) triplet angular distribution function $P_{OOO}(\theta)$ and tetrahedral order parameter at room temperature for three different cutoff distances in order to determine the influence of the cutoff distance. The corresponding results are shown in Figures 11 and 12, respectively. It is evident from these figures

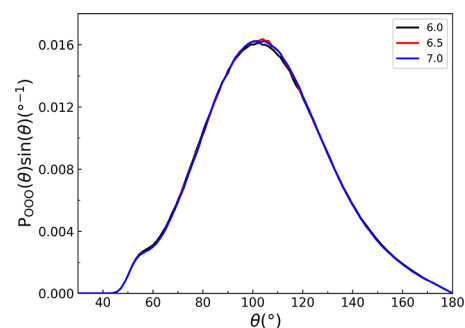


Figure 11. Oxygen–oxygen–oxygen triplet angular distribution functions of liquid water at 298 K from classical MD simulations for three range parameters indicated (in Å) for the 3-b interaction.

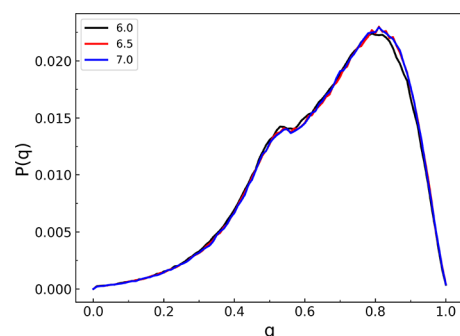


Figure 12. Probability distribution of the tetrahedral order parameter q at 298 K from classical MD simulations for three range parameters indicated (in Å) for the 3-b interaction.

that the computed angular distribution function and tetrahedral order parameter are nearly independent of the chosen cutoff distance. Again, this suggests that the orientational ordering and local structure of liquid water are not significantly affected by the specific range used for the 3-body interactions.

By utilizing multiple cores and adjusting the ranges of the 3-b and 4-b interactions, the computational timings can be greatly modified. It is important to note that the choice of cutoff distances should be carefully considered, depending on the property of interest. The convergence of results can be verified by varying the ranges until the desired properties are satisfactorily converged. Overall, these findings demonstrate the flexibility of the q-AQUA-pol potential in terms of adjusting the computational parameters to achieve accurate and efficient simulations of liquid water.

SUMMARY AND WHAT'S NEXT

The status of the CCSD(T)-based general potentials for water is that they have “arrived”. These basically come in three versions. One, q-AQUA, is based on a strict many-body representation truncated at the 4-b term. The second is where

we employ a Δ -ML approach, where a sophisticated polarizable potential is corrected using the MB approach. This was demonstrated by q-AQUA-pol which corrected TTM3-F up to the 4-b terms. MB-pol(2023) falls into this category but uses TTM4-F. These “bottom up” potentials produce highly accurate results from clusters to the condensed phase. The third version is based on ML fitting of DFT energies of 16 and 64 monomers obtained from molecular and path integral molecular dynamics in NVT simulations at 300 K. These initial fits are then corrected using Δ -learning and transfer learning with a limited number of CCSD(T) energies. Results for the condensed phase properties are accurate and comparable to those obtained with q-AQUA, q-AQUA-pol, and MB-pol(2023). However, no specific tests on water clusters using these Δ -learning and transfer learning potentials have been reported at this stage.

It is worth mentioning that another CCSD(T)-based potential for water is being developed by Zhang and collaborators, utilizing the fundamental invariant neural network (FINN) approach^{67,68} along with extensive high-quality data sets. The preliminary version of this many-body potential includes fits on 220,000 CCSD(T)/CBS 2-body data and 430,000 CCSD(T)/aVTZ 3-body data. This potential has shown promising results and has been successfully applied to torsional tunneling splitting calculations of water trimers.⁶⁹

Looking ahead, these potentials can be used in many possible applications and hopefully with some confidence in their predictive accuracy. Examples from our group include diffusion Monte Carlo calculations, which were already reported for q-AQUA and q-AQUA-pol to small clusters,^{18,41} as well as semiclassical initial value representation dynamics calculations for spectroscopy and vibrational wave functions.^{70,71} To be run successfully, these methods need a potential energy surface that is reliable at energies even higher than the zero-point one. Of course, this requires that these potentials be made available. Source code for q-AQUA and q-AQUA-pol are available on github.^{47,72}

AUTHOR INFORMATION

Corresponding Authors

Qi Yu – Department of Chemistry and Cherry L. Emerson Center for Scientific Computation, Emory University, Atlanta, Georgia 30322, United States; orcid.org/0000-0002-2030-0671; Email: qyu28@emory.edu

Chen Qu – Independent Researcher, Toronto, Ontario M9B 0E3, Canada; orcid.org/0000-0001-8889-4851; Email: szquchen@gmail.com

Paul L. Houston – Department of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853, United States; Department of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, Georgia 30332, United States; orcid.org/0000-0003-2566-9539; Email: PLH2@cornell.edu

Joel M. Bowman – Department of Chemistry and Cherry L. Emerson Center for Scientific Computation, Emory University, Atlanta, Georgia 30322, United States; orcid.org/0000-0001-9692-2672; Email: jmbowma@emory.edu

Authors

Apurba Nandi – Department of Chemistry and Cherry L. Emerson Center for Scientific Computation, Emory University, Atlanta, Georgia 30322, United States;

Department of Physics and Materials Science, University of Luxembourg, L-1511 Luxembourg City, Luxembourg; orcid.org/0000-0002-6191-5584

Priyanka Pandey – Department of Chemistry and Cherry L. Emerson Center for Scientific Computation, Emory University, Atlanta, Georgia 30322, United States

Riccardo Conte – Dipartimento di Chimica, Università degli Studi di Milano, 20133 Milano, Italy; orcid.org/0000-0003-3026-3875

Complete contact information is available at: <https://pubs.acs.org/10.1021/acs.jpcllett.3c01791>

Notes

The authors declare no competing financial interest.

Biographies

Qi Yu received his B.S. from University of Science and Technology of China in 2014 and his Ph.D. from Emory University in 2019 under the supervision of Prof. Joel M. Bowman. He was a postdoctoral associate in the Hammes-Schiffer lab at Yale University from 2019 to 2022.

Chen Qu received his B.S. from University of Science and Technology of China in 2011. He obtained his Ph.D. from Emory University in 2017, under the supervision of Prof. Joel M. Bowman. During 2019 to 2022, he worked as a postdoctoral associate at National Institute of Standards and Technology.

Paul L. Houston received a B. A. from Yale University and an M.S. and Ph.D. from the Massachusetts Institute of Technology. Following postdoctoral work with C. Bradley Moore at UC Berkeley, he joined the faculty at Cornell University, where he is now Peter J. W. Debye Professor of Chemistry, Emeritus. From 2007 through 2013 he was Dean of the College of Sciences at Georgia Institute of Technology.

Apurba Nandi received his B.Sc. from Jadavpur University and M.Sc. from Indian Institute of Technology (IIT) Kanpur. He obtained his Ph.D. from Emory University under the supervision of Prof. Joel M. Bowman in 2022. He is currently working as a postdoctoral fellow with Prof. Alexandre Tkatchenko at the University of Luxembourg.

Priyanka Pandey received a B.Sc from CSJM University in 2014 and an M.Sc in chemistry from the University of Delhi in 2016. She obtained her Ph.D. from Indian Institute of Technology Kanpur in 2022 under the supervision of Prof. Srihari Keshavamurthy. She is currently working as a postdoctoral fellow with Prof. Joel M. Bowman at Emory University.

Riccardo Conte received his M.Sc. from the University of Pisa and his diploma and Ph.D. from Scuola Normale Superiore di Pisa. After postdoctoral appointments with Eli Pollak at the Weizmann Institute of Science, Joel M. Bowman at Emory University, and Michele Ceotto at the University of Milan, he is currently assistant professor of theoretical chemistry at the University of Milan.

Joel M. Bowman received an A.B. from the University of California, Berkeley, and Ph.D. from California Institute of Technology. He is currently Samuel Candler Dobbs Professor of Chemistry at Emory University.

ACKNOWLEDGMENTS

J.M.B. thanks the ARO, DURIP grant (W911NF-14-1-0471), for funding a computer cluster where most of the calculations were performed and current financial support from NASA (80NSSC20K0360). R.C. thanks Università degli Studi di Milano under grant PSR2021-DIP-005-RCONT.

REFERENCES

- (1) Sun, J.; Ruzsinszky, A.; Perdew, J. P. Strongly Constrained and Appropriately Normed Semilocal Density Functional. *Phys. Rev. Lett.* **2015**, *115*, No. 036402.
- (2) Ruiz Pestana, L.; Mardirossian, N.; Head-Gordon, M.; Head-Gordon, T. Ab initio Molecular Dynamics Simulations of liquid Water using High Quality Meta-GGA Functionals. *Chem. Sci.* **2017**, *8*, 3554–3565.
- (3) Marsalek, O.; Markland, T. E. Quantum Dynamics and Spectroscopy of Ab Initio Liquid Water: The Interplay of Nuclear and Electronic Quantum Effects. *J. Phys. Chem. Lett.* **2017**, *8*, 1545–1551.
- (4) Ruiz Pestana, L.; Marsalek, O.; Markland, T. E.; Head-Gordon, T. The Quest for Accurate Liquid Water Properties from First Principles. *J. Phys. Chem. Lett.* **2018**, *9*, 5009–5016.
- (5) Dasgupta, S.; Lambros, E.; Perdew, J. P.; Paesani, F. Elevating Density Functional Theory to Chemical Accuracy for Water Simulations Through a Density-Corrected Many-Body Formalism. *Nat. Commun.* **2021**, *12*, 6359.
- (6) Morawietz, T.; Singraber, A.; Dellago, C.; Behler, J. How van der Waals interactions determine the unique properties of water. *Proc. Natl. Acad. Sci.* **2016**, *113*, 8368–8373.
- (7) Zhang, L.; Han, J.; Wang, H.; Car, R.; E, W. Deep Potential Molecular Dynamics: A Scalable Model with the Accuracy of Quantum Mechanics. *Phys. Rev. Lett.* **2018**, *120*, 143001.
- (8) Yao, Y.; Kanai, Y. Temperature dependence of nuclear quantum effects on liquid water via artificial neural network model based on SCAN meta-GGA functional. *J. Chem. Phys.* **2020**, *153*, No. 044114.
- (9) Zhang, C.; Tang, F.; Chen, M.; Xu, J.; Zhang, L.; Qiu, D. Y.; Perdew, J. P.; Klein, M. L.; Wu, X. Modeling liquid water by climbing up Jacob's ladder in density functional theory facilitated by using deep neural network potentials. *J. Phys. Chem. B* **2021**, *125*, 11444–11456.
- (10) Liu, J.; Lan, J.; He, X. Toward High-level Machine Learning Potential for Water Based on Quantum Fragmentation and Neural Networks. *J. Phys. Chem. A* **2022**, *126*, 3926–3936.
- (11) Xantheas, S. S. Ab Initio Studies of Cyclic Water Clusters ($H_2O)_N$, $N = 1-6$. 2. Analysis of Many-Body Interactions. *J. Chem. Phys.* **1994**, *100*, 7523.
- (12) Heindel, J. P.; Xantheas, S. S. The Many-Body Expansion for Aqueous Systems Revisited: I. Water–Water Interactions. *J. Chem. Theory Comput.* **2020**, *16*, 6843–6855.
- (13) Heindel, J. P.; Herman, K. M.; Aprà, E.; Xantheas, S. S. Guest–Host Interactions in Clathrate Hydrates: Benchmark MP2 and CCSD(T)/CBS Binding Energies of CH_4 , CO_2 , and H_2S in $(H_2O)_{20}$ Cages. *J. Phys. Chem. Lett.* **2021**, *12*, 7574–7582.
- (14) Wang, Y. M.; Bowman, J. M. Towards an Ab Initio Flexible Potential for Water, and Post-Harmonic Quantum Vibrational Analysis of Water Clusters. *Chem. Phys. Lett.* **2010**, *491*, 1.
- (15) Braams, B. J.; Bowman, J. M. Permutationally Invariant Potential Energy Surfaces in High Dimensionality. *Int. Rev. Phys. Chem.* **2009**, *28*, 577–606.
- (16) Houston, P. L.; Qu, C.; Nandi, A.; Conte, R.; Yu, Q.; Bowman, J. M. Permutationally Invariant Polynomial Regression for Energies and Gradients, Using Reverse Differentiation, Achieves Orders of Magnitude Speed-up with High Precision Compared to Other Machine Learning Methods. *J. Chem. Phys.* **2022**, *156*, No. 044120.
- (17) Nandi, A.; Qu, C.; Houston, P. L.; Conte, R.; Yu, Q.; Bowman, J. M. A CCSD(T)-Based 4-Body Potential for Water. *J. Phys. Chem. Lett.* **2021**, *12*, 10318–10324.
- (18) Yu, Q.; Qu, C.; Houston, P. L.; Conte, R.; Nandi, A.; Bowman, J. M. q-AQUA: a Many-body CCSD(T) Water Potential, Including 4-body Interactions, Demonstrates the Quantum Nature of Water from Clusters to the Liquid Phase. *J. Phys. Chem. Lett.* **2022**, *13*, 5068–5074.
- (19) Lodi, L.; Tennyson, J.; Polyansky, O. L. A Global, High Accuracy Ab Initio Dipole Moment Surface for the Electronic Ground State of the Water Molecule. *J. Chem. Phys.* **2011**, *135*, No. 034113.
- (20) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (21) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The missing term in effective pair potentials. *J. Phys. Chem.* **1987**, *91*, 6269–6271.
- (22) Paesani, F.; Zhang, W.; Case, D. A.; Cheatham, T. E.; Voth, G. A. An accurate and Simple Quantum Model for Liquid Water. *J. Chem. Phys.* **2006**, *125*, 184507.
- (23) Habershon, S.; Markland, T. E.; Manolopoulos, D. E. Competing Quantum Effects in the Dynamics of Flexible Water Model. *J. Chem. Phys.* **2009**, *131*, No. 024501.
- (24) 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. *J. Chem. Phys.* **1997**, *106*, 4618.
- (25) Wang, Y. M.; Huang, X. C.; Shepler, B. C.; Braams, B. J.; Bowman, J. M. Flexible, Ab Initio Potential, and Dipole Moment Surfaces for Water. I. Tests and Applications for Clusters up to the 22-mer. *J. Chem. Phys.* **2011**, *134*, No. 094509.
- (26) Babin, V.; Medders, G. R.; Paesani, F. Development of a “First Principles” Water Potential with Flexible Monomers. II: Trimer Potential Energy Surface, Third Virial Coefficient, and Small Clusters. *J. Chem. Theory Comput.* **2014**, *10*, 1599.
- (27) Babin, V.; Medders, G. R.; Paesani, F. Toward a Universal Water Model: First Principles Simulations from the Dimer to the Liquid Phase. *J. Phys. Chem. Lett.* **2012**, *3*, 3765.
- (28) Gora, U.; Cencek, W.; Podeszwa, R.; van der Avoird, A.; Szalewicz, K. Predictions for Water Clusters from a First-Principles Two- and Three-Body Force Field. *J. Chem. Phys.* **2014**, *140*, 194101.
- (29) Reddy, S. K.; Straight, S. C.; Bajaj, P.; Huy Pham, C.; Riera, M.; Moberg, D. R.; Morales, M. A.; Knight, C.; Götz, A. W.; Paesani, F. On the Accuracy of the MB-pol Many-body Potential for Water: Interaction Energies, Vibrational Frequencies, and Classical Thermodynamic and Dynamical Properties from Clusters to Liquid water and Ice. *J. Chem. Phys.* **2016**, *145*, 194504.
- (30) Burnham, C. J.; Anick, D. J.; Mankoo, P. K.; Reiter, G. F. The Vibrational Proton Potential in Bulk Liquid Water and Ice. *J. Chem. Phys.* **2008**, *128*, 154519.
- (31) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; von Lilienfeld, O. A. Big Data Meets Quantum Chemistry Approximations: The Δ -Machine Learning Approach. *J. Chem. Theory Comput.* **2015**, *11*, 2087–2096.
- (32) Zaspel, P.; Huang, B.; Harbrecht, H.; von Lilienfeld, O. A. Boosting Quantum Machine Learning Models with a Multilevel Combination Technique: Pople Diagrams Revisited. *J. Chem. Theory and Comput.* **2019**, *15*, 1546–1559.
- (33) Dral, P. O.; Owens, A.; Dral, A.; Csányi, G. Hierarchical Machine Learning of Potential Energy Surfaces. *J. Chem. Phys.* **2020**, *152*, 204110.
- (34) Zhu, J.; Vuong, V. Q.; Sumpter, B. G.; Irlé, S. Artificial Neural Network Correction for Density-Functional Tight-Binding Molecular Dynamics Simulations. *MRS Commun.* **2019**, *9*, 867–873.
- (35) Nandi, A.; Qu, C.; Houston, P. L.; Conte, R.; Bowman, J. M. Δ -machine Learning for Potential Energy Surfaces: A PIP Approach to Bring a DFT-based PES to CCSD(T) Level of Theory. *J. Chem. Phys.* **2021**, *154*, No. 051102.
- (36) Qu, C.; Houston, P. L.; Conte, R.; Nandi, A.; Bowman, J. M. Breaking the Coupled Cluster Barrier for Machine-Learned Potentials of Large Molecules: The Case of 15-Atom Acetylacetone. *J. Phys. Chem. Lett.* **2021**, *12*, 4902–4909.
- (37) Liu, Y.; Li, J. Permutation-Invariant-Polynomial Neural-Network-Based Δ -Machine Learning Approach: A Case for the HO_2 Self-Reaction and Its Dynamics Study. *J. Phys. Chem. Lett.* **2022**, *13*, 4729–4738.
- (38) Fanourgakis, G. S.; Xantheas, S. S. Development of Transferable Interaction Potentials for Water. V. Extension of the Flexible, Polarizable, Thole-Type Model Potential (TTM3-F, v. 3.0) to Describe the Vibrational Spectra of Water Clusters and Liquid Water. *J. Chem. Phys.* **2008**, *128*, No. 074506.

- (39) Bowman, J. M.; Qu, C.; Conte, R.; Nandi, A.; Houston, P. L.; Yu, Q. Δ -Machine Learned Potential Energy Surfaces and Force Fields. *J. Chem. Theory and Comput.* **2023**, *19*, 1–17.
- (40) Fanourgakis, G. S.; Xantheas, S. S. The Flexible, Polarizable, Thole-type Interaction Potential for Water (TTM2-F) Revisited. *J. Phys. Chem. A* **2006**, *110*, 4100–4106.
- (41) Qu, C.; Yu, Q.; Houston, P. L.; Conte, R.; Nandi, A.; Bowman, J. M. Interfacing q-AQUA with a Polarizable Force Field: The Best of Both Worlds. *J. Chem. Theory Comput.* **2023**, *19*, 3446–3459.
- (42) 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. *J. Chem. Theory Comput.* **2023**, *19*, 3551–3556.
- (43) Miller, T. F.; Manolopoulos, D. E. Quantum Diffusion in Liquid Water from Ring Polymer Molecular Dynamics. *J. Chem. Phys.* **2005**, *123*, 154504.
- (44) Kapil, V.; Rossi, M.; Marsalek, O.; Petraglia, R.; Litman, Y.; Spura, T.; Cheng, B.; Cuzzocrea, A.; Meißner, R. H.; Wilkins, D. M.; et al. i-PI 2.0: A Universal Force Engine for Advanced Molecular Simulations. *Comput. Phys. Commun.* **2019**, *236*, 214–223.
- (45) Skinner, L. B.; Huang, C.; Schlesinger, D.; Pettersson, L. G. M.; Nilsson, A.; Benmore, C. J. Benchmark Oxygen-oxygen Pair-distribution Function of Ambient Water from X-ray Diffraction Measurements with a Wide Q-range. *J. Chem. Phys.* **2013**, *138*, No. 074506.
- (46) Skinner, L. B.; Benmore, C. J.; Neufeind, J. C.; Parise, J. B. The Structure of Water Around the Compressibility Minimum. *J. Chem. Phys.* **2014**, *141*, 214507.
- (47) q-AQUA. <https://github.com/jmbowma/q-AQUA>, 2022.
- (48) Soper, A.; Benmore, C. Quantum Differences Between Heavy and Light Water. *Phys. Rev. Lett.* **2008**, *101*, No. 065502.
- (49) Ceriotti, M.; Fang, W.; Kusalik, P. G.; McKenzie, R. H.; Michaelides, A.; Morales, M. A.; Markland, T. E. Nuclear quantum effects in water and aqueous systems: Experiment, Theory, and Current Challenges. *Chem. Revs.* **2016**, *116*, 7529–7550.
- (50) Wagner, W.; Pruß, A. The IAPWS Formulation 1995 for the Thermodynamic Properties of Ordinary Water Substance for General and Scientific Use. *J. Phys. Chem. Ref. Data* **2002**, *31*, 387–535.
- (51) Harvey, A. Properties of Ice and Supercooled Water. *CRC Handbook of Chemistry and Physics*; CRC Press: Boca Raton, FL, 2019.
- (52) Bates, D. M.; Tschumper, G. S. CCSD(T) Complete Basis Set Limit Relative Energies for Low-Lying Water Hexamer Structures. *J. Phys. Chem. A* **2009**, *113*, 3555–3559.
- (53) Daru, J.; Forbert, H.; Behler, J.; Marx, D. Coupled Cluster Molecular Dynamics of Condensed Phase Systems Enabled by Machine Learning Potentials: Liquid Water Benchmark. *Phys. Rev. Lett.* **2022**, *129*, 226001.
- (54) Käser, S.; Meuwly, M. Transfer-Learned Potential Energy Surfaces: Toward Microsecond-Scale Molecular Dynamics Simulations in the Gas Phase at CCSD(T) Quality. *J. Chem. Phys.* **2023**, *158*, 214301.
- (55) Chen, M. S.; Lee, J.; Ye, H.-Z.; Berkelbach, T. C.; Reichman, D. R.; Markland, T. E. Data-Efficient Machine Learning Potentials from Transfer Learning of Periodic Correlated Electronic Structure Methods: Liquid Water at AFQMC, CCSD, and CCSD(T) Accuracy. *J. Chem. Theory Comput.* **2023**, *19*, 4510–4519.
- (56) Tersoff, J. Modeling Solid-State Chemistry: Interatomic Potentials for Multicomponent Systems. *Phys. Rev. B* **1989**, *39*, 5566–5568.
- (57) Behler, J.; Parrinello, M. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Phys. Rev. Lett.* **2007**, *98*, 146401.
- (58) Jorgensen, W. L. Quantum and Statistical Mechanical Studies of Liquids. 10. Transferable Intermolecular Potential Functions for Water, Alcohols, and Ethers. Application to Liquid Water. *J. Am. Chem. Soc.* **1981**, *103*, 335–340.
- (59) Bowman, J. M.; Qu, C.; Conte, R.; Nandi, A.; Houston, P. L.; Yu, Q. The MD17 Datasets From the Perspective of Datasets for Gas-Phase “Small” Molecule Potentials. *J. Chem. Phys.* **2022**, *156*, 240901.
- (60) Mills, R. Self-diffusion in Normal and Heavy Water in the Range 1–45. *J. Phys. Chem.* **1973**, *77*, 685–688.
- (61) Holz, M.; Heil, S. R.; Sacco, A. Temperature-dependent Self-diffusion Coefficients of Water and Six Selected Molecular Liquids for Calibration in Accurate ^1H NMR PFG Measurements. *Phys. Chem. Chem. Phys.* **2000**, *2*, 4740–4742.
- (62) Keutsch, F.; Saykally, R. Water clusters: Untangling the Mysteries of the Liquid, One Molecule at a Time. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 10533–10540.
- (63) Buck, U.; Pradzynski, C. C.; Zeuch, T.; Dieterich, J. M.; Hartke, B. A Size Resolved Investigation of Large Water Clusters. *Phys. Chem. Chem. Phys.* **2014**, *16*, 6859–6871.
- (64) Ch’ng, L. C.; Samanta, A. K.; Czako, G.; Bowman, J. M.; Reisler, H. Experimental and Theoretical Investigations of Energy Transfer and Hydrogen-Bond Breaking in the Water Dimer. *J. Am. Chem. Soc.* **2012**, *134*, 15430.
- (65) Pérez, C.; Muckle, M. T.; Zaleski, D. P.; Seifert, N. A.; Temelso, B.; Shields, G. C.; Kisiel, Z.; Pate, B. H. Structures of Cage, Prism, and Book Isomers of Water Hexamer from Broadband Rotational Spectroscopy. *Science* **2012**, *336*, 897–901.
- (66) Schwan, R.; Qu, C.; Mani, D.; Pal, N.; van der Meer, L.; Redlich, B.; Leforestier, C.; Bowman, J. M.; Schwaab, G.; Havenith, M. Observation of the Low-Frequency Spectrum of the Water Dimer as a Sensitive Test of the Water Dimer Potential and Dipole Moment Surfaces. *Angew. Chem., Int. Ed.* **2019**, *58*, 13119–13126.
- (67) Shao, K.; Chen, J.; Zhao, Z.; Zhang, D. H. Communication: Fitting potential energy surfaces with fundamental invariant neural network. *J. Chem. Phys.* **2016**, *145*, No. 071101.
- (68) Chen, R.; Shao, K.; Fu, B.; Zhang, D. H. Fitting potential energy surfaces with fundamental invariant neural network. II. Generating fundamental invariants for molecular systems with up to ten atoms. *J. Chem. Phys.* **2020**, *152*, 204307.
- (69) Zhu, Y.-C.; Yang, S.; Zeng, J.-X.; Fang, W.; Jiang, L.; Zhang, D. H.; Li, X.-Z. Torsional Tunneling Splitting in a Water Trimer. *J. Am. Chem. Soc.* **2022**, *144*, 21356–21362.
- (70) Conte, R.; Nandi, A.; Qu, C.; Yu, Q.; Houston, P. L.; Bowman, J. M. Semiclassical and VSCF/VCI Calculations of the Vibrational Energies of trans-and gauche-Ethanol Using a CCSD (T) Potential Energy Surface. *J. Phys. Chem. A* **2022**, *126*, 7709–7718.
- (71) Aieta, C.; Bertaina, G.; Micciarelli, M.; Ceotto, M. Representing Molecular Ground and Excited Vibrational Eigenstates with Nuclear Densities Obtained from Semiclassical Initial Value Representation Molecular Dynamics. *J. Chem. Phys.* **2020**, *153*, 214117.
- (72) q-AQUA-pol. <https://github.com/jmbowma/q-AQUA-pol>, 2023.