JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Perspective

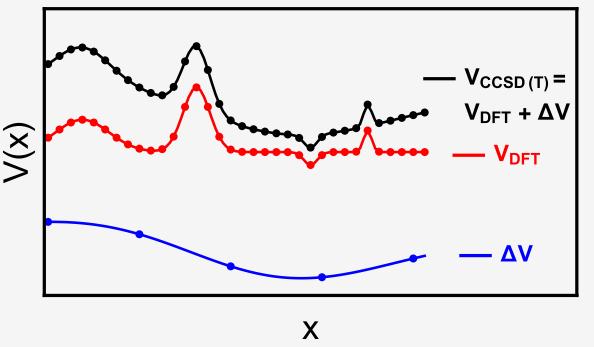
Δ-Machine Learned Potential Energy Surfaces and Force Fields

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

Cite This: https://doi.org/10.1021/acs.jctc.2c01034 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: There has been great progress in developing machine-learned potential energy surfaces (PESs) for molecules and clusters with more than 10 atoms. Unfortunately, this number of atoms generally limits the level of electronic structure theory to less than the "gold standard" CCSD(T) level. Indeed, for the well-known MD17 dataset for molecules with 9−20 atoms, all of the energies and forces were obtained with DFT calculations (PBE). This Perspective is focused on a Δ-machine learning method that we recently proposed and applied to bring DFT-based PESs to close to CCSD(T) accuracy. This is demonstrated for hydronium, *N*-methylacetamide, acetyl acetone, and ethanol. For 15-atom tropolone, it appears that special approaches (e.g., molecular tailoring, local CCSD(T)) are needed to obtain the CCSD(T) energies. A new aspect of this approach is the extension of Δ-machine learning to force fields. The approach is based on many-body corrections to polarizable force field potentials. This is examined in detail using the TTM2.1 water potential. The corrections make use of our recent CCSD(T) datasets for 2-b, 3-b, and 4-b interactions for water. These datasets were used to develop a new fully ab initio potential for water, termed q-AQUA.



INTRODUCTION

There has been dramatic progress in machine-learned potentials (MLPs) over the past 10−15 years for isolated molecules, many-body non-covalent clusters, and applications to solid materials. Assessments of the many machine learning (ML) methods that have been developed (e.g., artificial neural networks, kernel methods, polynomial regression, etc.) have only recently appeared.[1,2] Aside from the obvious importance of doing such assessments, the stress here is on the datasets used in MLPs, namely, the widely used MD-17 datasets.[3] These are all DFT (PBE) energies and forces for molecules ranging in size from ethanol to aspirin. By contrast, from the field of gas-phase dynamics, PESs for molecules with four to seven atoms are based on the "gold standard" CCSD(T) method.[4] To the best of our knowledge, the semiglobal CCSD(T)-based PES for the 10-atom formic acid dimer is the largest system for which this level of theory has been applied.[5] Recent reactive CCSD(T)-based PESs from Czako and co-workers for several nine-atom reactions are close seconds.[6−8] Of course, there are good reasons for the DFT/CCSD(T) dichotomy, the major one being that DFT scales as $O(N^3)$ whereas CCSD(T) scales as $O(N^7)$, where N is the size of the system.

Our group has been active for some time in developing MLPs using permutationally invariant polynomials (PIPs)[2,4,9−15] for "small molecules"; some early examples include CH_5^+ and $H_5O_2^+$, and later ones include malonaldehyde and the 10-atom formic acid dimer.[5] More recently, we have reported PIP PESs for molecules containing 10−15 atoms using enhancements to the PIP basis.[16−18] The PIP method was recently evaluated[2] against the ML methods assessed in ref 1. PIPs were shown to be as precise as the most precise methods but to run much faster both for energies and gradients with fast backward differentiation.

The general target of our PESs is to be accurate at large enough energies to enable rigorous diffusion Monte Carlo (DMC) calculations of the zero-point energy (ZPE) and wave function,[19−22] quantum vibrational and scattering calculations, and quasiclassical trajectory (QCT) calculations of chemical reaction dynamics.[23,24] To this end, our general strategy is to run DFT-based direct dynamics with total energies initiated from different minima and saddle point barriers. (For a recent discussion of this approach, see ref 25.) Often, a DFT-based PIP PES is fit to this dataset. However, the general objective is to use tens of thousands of DFT configurations to determine CCSD(T) energies and then to fit those. Of course, this approach becomes impractical for large molecules (more than 10−15 atoms), where the computational cost per configuration at the CCSD(T) level of theory becomes too high. The advantage of DFT in scaling allows the size of molecules to be larger than the 10−15 atom limit for CCSD(T), so there is

Received: October 18, 2022

https://doi.org/10.1021/acs.jctc.2c01034
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

strong motivation to investigate applying ML methods to bring DFT (or MP2)-based PESs to the CCSD(T) level. This is the subject of this Perspective, where we focus on our specific version[26] of Δ-machine learning (Δ-ML).[27,28]

Our approach is to construct a high-level (typically CCSD(T)) PES starting from a lower-level MP2 or DFT one using a correction that is a fit to a small number of high-level ab initio energies. Explicitly, the corrected high-level PES, denoted $V_{LL\rightarrow CC}$, is given by

$$V_{LL\rightarrow CC} = V_{LL} + \Delta V_{CC-LL} \quad (1)$$

where V_{LL} is the lower-level PES and ΔV_{CC-LL} is the correction PES. We note that previously the basic idea of using energy differences between high and low levels of ab initio theory to correct a PES can be found in refs 29 and 30 for the OH + H_2 and F + H_2 reactions, respectively.

In the next section we discuss applications of Δ-ML to a variety of molecular systems, starting with a pedagogical example of the small-molecule PES for H_3O^+, for which we have reported a CCSD(T)-based PIP PES based on tens of thousands of CCSD(T) energies.[31] Then we move to more challenging molecules, namely, *N*-methylacetamide, ethanol, and acetylacetone for which we have reported Δ-ML PESs. Then we conclude that section of the Perspective with a discussion of tropolone, for which we have reported a DFT-based PES[18] but for which a Δ-ML PES is still a work in progress. The third section describes a new direction for us, namely, using Δ-ML to correct a force field. Preliminary work in this direction was recently reported with the goal of correcting the four-body (4-b) water interaction in the MB-pol potential;[32] the interested reader is referred to ref 33 for details. In this Perspective, we demonstrate the Δ-ML approach for the polarizable water potential TTM2.1.[34] The Perspective closes with a Discussion and Summary.

Δ-ML PESs FOR A VARIETY OF MOLECULAR SYSTEMS

Hydronium Ion. Hydronium is a centrally important cation in chemistry and has been extensively studied spectroscopically due to its large tunneling splittings for the ground and excited vibrational states governed by the barrier separating the symmetric double well potentials. The splittings are very sensitive to the barrier height, so this small molecule has served as an excellent test for the Δ-ML method.[26] We present highlights of that work here.

Because there was not a pre-existing DFT-based PES, we developed one using configurations from our previously reported PES, which was a fit to 32 142 CCSD(T)/aug-cc-pVQZ energies.[31] Details were reported previously.[26]

One important aspect of the Δ-ML method is to examine the fidelity of ΔV_{CC-LL} using different training datasets. This was done using 1000, 500, 250, and 125 configurations. The test datasets consist of the remaining data from the total of 32 142 configurations. The root-mean-square errors (RMSEs) between the $V_{LL\rightarrow CC}$ and CCSD(T) energies, given in Table 1, are similar for all of the training datasets. The result for the training set of 125 energy differences are encouraging, where the RMSE is just 32 cm^{-1} for test energies up to 23 000 cm^{-1}. In this case the PIP basis for ΔV_{CC-LL} contains only 51 terms. Timings for ΔV_{CC-LL} for this and other molecules are given in a subsequent section.

Plots of $V_{LL\rightarrow CC}$ versus the direct CCSD(T) energy for the training set of 500 points and the corresponding test data are shown in Figure 1. Overall the correction PES reproduces the direct energies precisely. As expected, the absolute errors increase with energy and do become large for very high energies. If needed, one can always reduce these errors by adding the high-energy data points into the training dataset. Beyond the energy RMSE, we reported a test of this Δ-ML PES for the equilibrium geometries of both the global minimum and transition state (TS) and the normal mode frequencies. This corrected PES, $V_{LL\rightarrow CC}$, produced results in excellent agreement with the direct CCSD(T) ones and also provided a significant improvement compared to the DFT PES in terms of barrier height, normal mode frequencies, and geometric parameters. Perhaps most impressive is the excellent accuracy that was achieved even with the smallest training dataset of 100 energies. The interested reader is referred to the original paper for details.[26]

Table 1. RMS Error (RMSE) between Direct CCSD(T) and $V_{LL\rightarrow CC}$ Energies (in cm^{-1}) with the Indicated Number of Test Configurations (N_{test}) for H_3O^+, where Training on ΔV_{CC-LL} Was Done for Various Numbers of Training Data (N_{train})[26]

N_{train}	N_{test}	RMSE
1000[a]	31142	55
500[b]	31642	28
250[c]	31892	51
125[d]	32017	32

[a] ΔV_{CC-LL} was trained with maximum polynomial order of 7 and a basis size of 348. [b] ΔV_{CC-LL} was trained with maximum polynomial order of 6 and a basis size of 196. [c] ΔV_{CC-LL} was trained with maximum polynomial order of 5 and a basis size of 103. [d] ΔV_{CC-LL} was trained with maximum polynomial order of 4 and a basis size of 51.

***N*-Methylacetamide.** We next consider the 12-atom *N*-methylacetamide (NMA) molecule, which is of central interest as the smallest molecule in the peptide linkage in polypeptides and proteins. For us, the interest is the challenge to correct a DFT PES that contains the *cis* and *trans* isomers, two isomerization saddle points, and two threefold methyl torsional potentials. This is clearly a more complex PES than the one for H_3O^+ and more representative of PESs of larger molecules.

Our first report of DFT-based PESs for NMA examined full and fragmented PIP basis sets.[35,36] The idea of using a fragmented basis to extend the PIP approach to molecules with more than 10 atoms was first illustrated for NMA. The full basis for a maximum polynomial order of 3 has 8040 linear coefficients. The fragmented PIP basis, also with a maximum polynomial order of 3, contains 6121 coefficients. The fits were done using 6607 energies and the corresponding 237 852 gradient components, giving a total dataset size of 244 459. These data were obtained from direct dynamics using the B3LYP/cc-pVDZ level of theory. Clearly, obtaining a dataset of this size from CCSD(T) calculations is not feasible, and Δ-ML was used in order to bring this DFT-based PES close to CCSD(T) quality.

For the training and testing we calculated a total of 5430 CCSD(T)-F12/aug-cc-pVDZ energies. Training of ΔV_{CC-LL} was done on 4696 data points of the difference between direct CCSD(T) and DFT absolute energies. Testing of $V_{LL\rightarrow CC}$ was done on 734 energies. The PIP basis to fit ΔV_{CC-LL} was generated using MSA software with the reduced permutational symmetry of 31111113. This notation corresponds to the

https://doi.org/10.1021/acs.jctc.2c01034
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

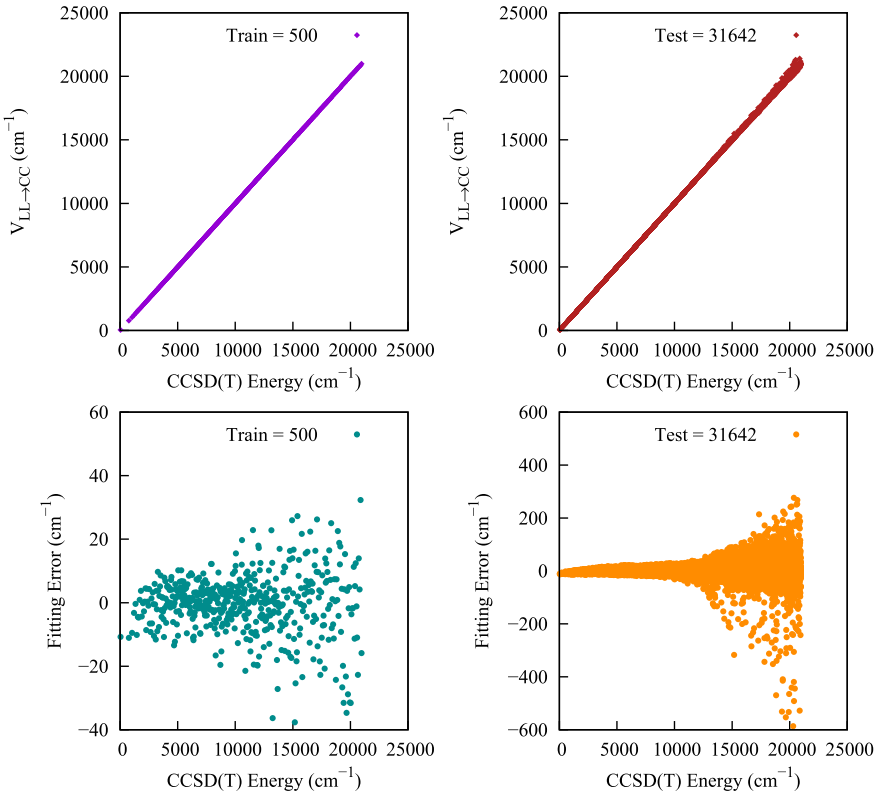


Figure 1. (top) Plots of energies of H_3O^+ from $V_{LL\rightarrow CC}$ vs direct CCSD(T) ones for the indicated datasets. The plot labeled "Train" corresponds to the configurations used in the training of ΔV_{CC-LL}, and the plot labeled "Test" is for the remaining configurations. (bottom) Corresponding fitting errors relative to the minimum energy. Reproduced with permission from ref 26. Copyright 2021 the authors of ref 26, published under license by AIP Publishing.

molecule class $A_3BCDEFGH_3$, where "A" and "H" are hydrogen atoms and "B" and "G" are carbon atoms, etc. This is the minimum symmetry that describes the identity of the hydrogen atoms within a methyl group. This is essential to get the threefold torsional barrier. The maximum polynomial order is 2. This basis leads to 569 linear coefficients. The fitting RMSE of this ΔV_{CC-LL} was 57 cm^{-1}. Plots of $V_{LL\rightarrow CC}$ versus the direct CCSD(T) energy for the training and test data are shown in Figure 2. The RMSEs between the $V_{LL\rightarrow CC}$ and direct CCSD(T) energies for the training and test datasets are 57 and 147 cm^{-1}, respectively.

Further tests of the Δ-ML PES were performed for geometry optimization and normal mode analyses for the *cis* and *trans* isomers, and details can be found in ref 26. Here we note that the DFT PES, which predicts an incorrect minimum for the *trans* isomer, is successfully corrected using the ΔV_{CC-LL} PES. Specifically, the DFT torsion angle of one methyl rotor (NH−CH_3) is shifted by 60° relative to the CCSD(T) structure. In addition, the torsional potentials of the two methyl rotors of both the *trans* and *cis* isomers of NMA are improved substantially from the previously reported DFT PES results.[36]

Detailed comparisons of the partially relaxed torsional barriers are given in Table 2. As can be seen, there are large differences between the DFT PES and CCSD(T) results for the NH−CH_3 rotors for both the *cis* and *trans* isomers. Overall, the Δ-ML PES barriers are significantly closer in energy to the CCSD(T) ones. To the best of our knowledge there is no experimental determination of these torsional barriers. However, a previous report of a torsional barrier of 24 cm^{-1} for acetamide[37] is consistent with the small barriers of 34 and 74 cm^{-1} (Δ-ML PES) for *trans*-NMA. For more details, interested readers are referred to the original paper.[26]

Ethanol. Ethanol is an important case to consider next for several reasons. First, it was the focus of two studies assessing numerous MLP methods, mentioned already.[1,2] Second, it is scientifically of fundamental interest because it has two nearly isoenergetic conformers (*trans* and *gauche*) and two different methyl rotors. It is also of great applied interest in fields as diverse as combustion and astrochemistry and in the condensed phase as a solvent.

In our very recent paper,[38] we observed that the ground state of ethanol resembles the *trans* conformer but that substantial delocalization to the *gauche* conformer is present based on rigorous quantum calculations of the vibrational zero-point state using a newly developed CCSD(T) PES. The new PES was obtained by means of a Δ-ML approach starting from a pre-existing low-level DFT surface.[2] The DFT dataset was taken from our previously reported "MDQM21" dataset,[2] where a total of 11 000 energies and their corresponding gradients were generated from ab initio molecular dynamics (AIMD) simulations at the B3LYP/6-311+G(d,p) level of theory. The DFT PES (V_{LL}) was a fit using 8500 DFT data that spanned the energy range of 0−35 000 cm^{-1}, with a maximum polynomial order of 4 and 321111 permutational symmetry, which led to a total of 14 752 PIPs in the fitting basis set. The fitting RMSEs for energies and gradients were 40 and 73 cm^{-1} bohr^{-1}, respectively.

This low-level DFT surface was brought to the CCSD(T) level of theory using a relatively small number of ab initio CCSD(T) energies. A dataset of 2319 geometries was sparsely

https://doi.org/10.1021/acs.jctc.2c01034
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

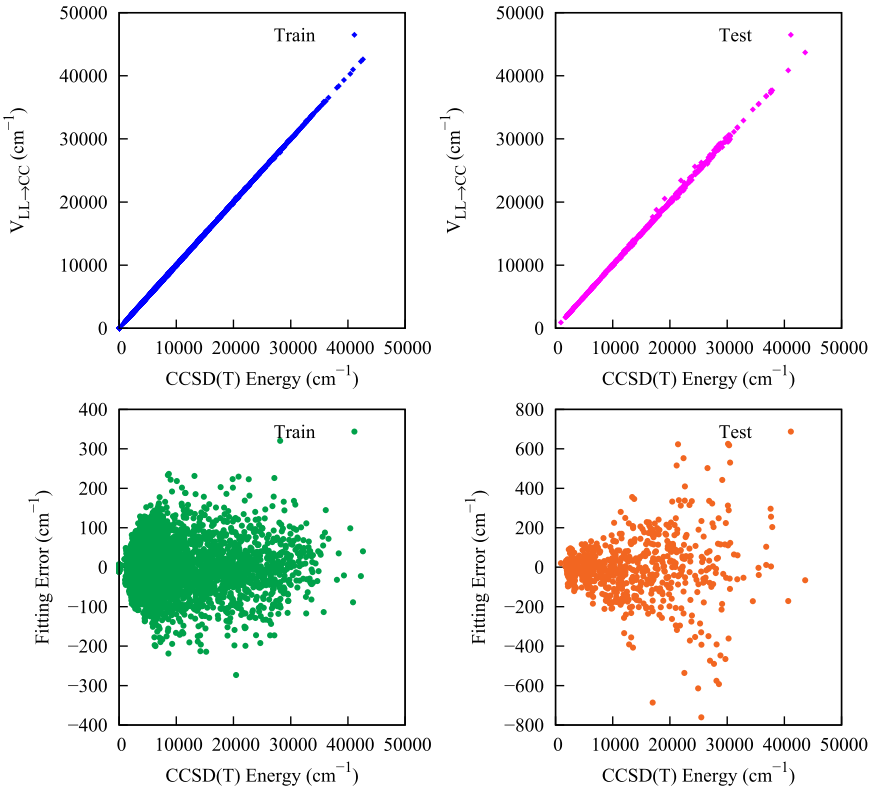


Figure 2. (top) Plots of energies of *N*-methylacetamide from $V_{LL\rightarrow CC}$ vs direct CCSD(T) ones for the indicated datasets. The plot labeled "Train" corresponds to the configurations used in the training of ΔV_{CC-LL}, and the plot labeled "Test" is for the remaining configurations. (bottom) Corresponding fitting errors relative to the minimum energy. Reproduced with permission from ref 26. Copyright 2021 the authors of ref 26, published under license by AIP Publishing.

Table 2. Comparison of Torsion Barriers of the Methyl Rotors CH_3-NH and CH_3-CO (in cm^{-1}) for *Trans* and *Cis* Isomers of *N*-Methylacetamide[26]

method	CH_3-NH	CH_3-CO
	trans-NMA	
DFT PES	256	37
Δ-ML PES	34	74
CCSD(T)	42	103
	cis-NMA	
DFT PES	61	361
Δ-ML PES	153	366
CCSD(T)	148	303

selected from the "MDQM21" DFT dataset, and CCSD(T)-F12a/aug-cc-pVDZ energy computations were performed at those geometries. To develop the correction PES, we trained ΔV_{CC-LL} on the difference between the CCSD(T) and DFT absolute energies of 2069 geometries and tested the obtained surface on the remaining 250 geometries.

It was observed that ΔV_{CC-LL} does not vary as strongly as V_{LL} with respect to the nuclear configuration. Therefore, low-order polynomial fitting was employed to fit this correction PES. The maximum polynomial order of 2 with permutational symmetry 321111 was used to fit the training dataset, leading to a total of 208 unknown linear coefficients (equivalent to the number of terms in the PIP fitting basis set). We added this correction ΔV_{CC-LL} to the low-level DFT PES, V_{LL}, to obtain the CCSD(T) energies. Plots of $V_{LL\rightarrow CC}$ versus the corresponding direct CCSD(T) energy for the training set of 2069 points and the test set of 250 points are shown in Figure 3, as recently reported.[38] As can be seen, there is overall excellent precision; however, we see a few larger errors. The RMSEs between the $V_{LL\rightarrow CC}$ and direct CCSD(T) energies for the training and test datasets are 49 and 63 cm^{-1}, respectively.

Next, this PES was examined for general fidelity by performing a geometry optimization and a normal mode frequency calculation for both the *trans* and *gauche* isomers and their two isomerization saddle point geometries. We found excellent accuracy with respect to the direct CCSD(T)-F12a calculations. We obtained PES optimized energies within 5 cm^{-1} of the direct CCSD(T)-F12a calculation and found that the *trans* isomer is lower in energy by 38 cm^{-1}. This PES obtained barrier heights for *trans*−*gauche* isomerization with respect to eclipsed and syn TSs that agree well with the corresponding direct ab initio values as well the experimental barrier heights.[39] Another comparison to experiment was the torsional barrier for the methyl rotor as a function of the torsional angle for both the *trans* and *gauche* isomers as well as two TSs. The agreement was very good.

Apart from these electronic energy investigations and comparisons, we moved to consider nuclear quantum effects. As a remarkable quantum nuclear application of the PES, we presented the results of DMC and semiclassical AS SCIVR calculations of the ZPEs for both the *trans* and *gauche* isomers as well as singly deuterated isotopologues. In addition, it is well-known that a DMC calculation is a very challenging test to examine the quality of a PES in extended regions of the configuration space. A common issue in PES fitting is the unphysical behavior in the extrapolated regions where the fitting dataset lacks data, and this is dramatically manifested by large negative energy values with respect to the global-

https://doi.org/10.1021/acs.jctc.2c01034
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

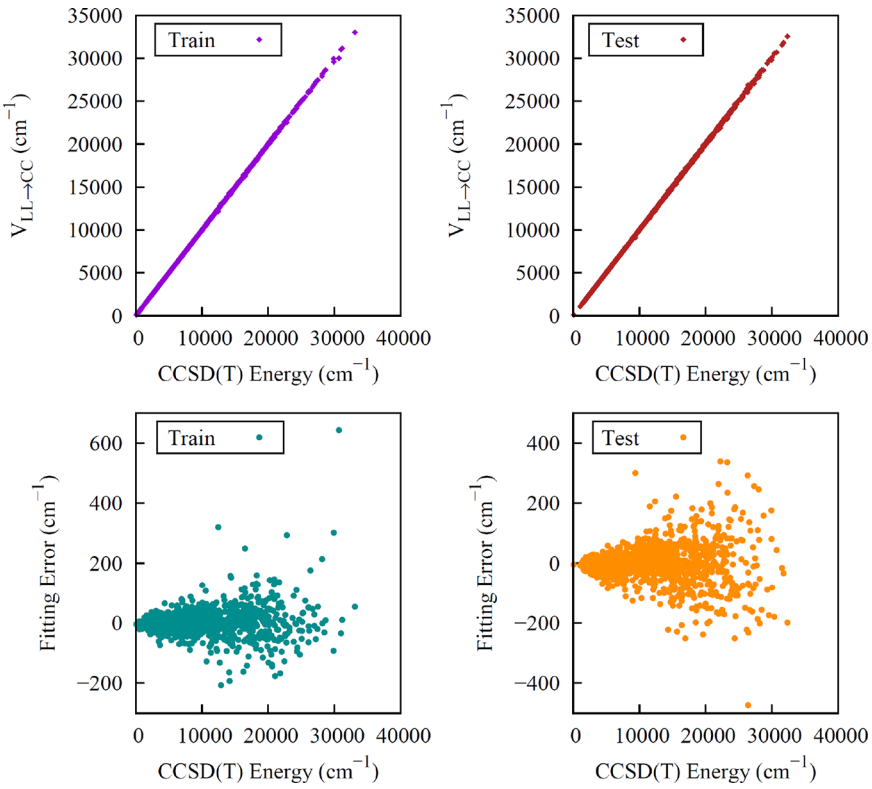


Figure 3. (top) Plots of energies of CH_3CH_2OH from $V_{LL\rightarrow CC}$ vs direct CCSD(T) ones for the indicated datasets. The plot labeled "Train" corresponds to the configurations used in the training of ΔV_{CC-LL}, and the plot labeled "Test" is for the set of remaining configurations. (bottom) Corresponding fitting errors relative to the minimum energy. From ref 38. CC BY 4.0.

minimum energy. These are called "holes" in the PES. Generally, we have observed that holes occur for highly repulsive configurations, that is short internuclear distances. Therefore, one goal of presenting DMC calculations was also to demonstrate that our PES correctly describes the high-energy regions of ethanol and is therefore suitable for quantum approaches that need to sample these regions.

The DMC results showed that the ground-state wave function has a *trans* fingerprint even when starting from the *gauche* conformer, an observation that revealed the ground state to be of the *trans* type with a leak to the *gauche* conformer. This PES was further employed to study the motions of the $-CH_3$ and $-OH$ rotors at the quantum-mechanical level. Rigorous DMC and 1D discrete variable representation (DVR) results were in very good agreement, and the computed DVR wave functions confirm the presence of the "leak" effect. Furthermore, the previously suggested geared motion of the rotors was also confirmed by our calculations, and the 2D model of the torsions based on potential cuts through the full-dimensional potential provides reasonable energy levels compared to experiment.

Acetylacetone. The simplest reactions are isomerizations, and of these, the symmetric double well isomerizations are the most studied. Malonaldehyde is perhaps the most studied both experimentally and theoretically.[40] Acetylacetone (AcAc), which is the next molecule we consider, does have an analogous H atom transfer between two oxygen atoms. However, to the best of our knowledge the corresponding tunneling splitting has not been reported experimentally.

Here we focus on an application of the Δ-ML approach to AcAc, starting with a fragmented PIP basis that we used to obtained a new PES.[41] We benefited from the dataset used for a neural-network-based PES for AcAc,[42] but we decided to augment the dataset for that fit with additional MP2/aVTZ energies and gradients. The Δ-ML approach was then used to bring this surface up to LCCSD(T)-F12 accuracy.

A PES fit based on fragmentation of the AcAc PIP basis into four fragments with a maximum polynomial order of 3 was used for the V_{LL} (MP2) surface. The details of this approach have been described previously.[17,35,36,41] The permutational symmetries and atom numbers (using the scheme in Figure 4)

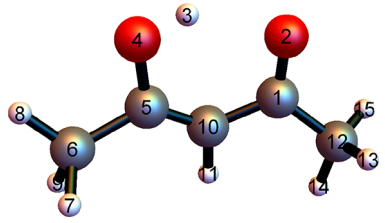


Figure 4. Numbering scheme used for the 4-(9,11,11,8) basis set. Reproduced with permission from ref 41. Copyright 2020 The PCCP Owner Societies.

are as follows: symmetry {1, 1, 1, 1, 1, 1, 1, 1} with atoms {1, 2, 3, 4, 5, 6, 10, 11, 12}; symmetry {3, 1, 1, 1, 1, 1, 1, 1} with atoms {13, 14, 15, 1, 2, 3, 4, 5, 10, 11, 12}; symmetry {3, 1, 1, 1, 1, 1, 1, 1} with atoms {7, 8, 9, 1, 2, 3, 4, 5, 6, 10, 11}; and symmetry {3, 3, 1, 1} with atoms {7, 8, 9, 13, 14, 15, 6, 12}. We refer to this fragmentation as 4-(9,11,11,8), where the numbers in parentheses represent the numbers of atoms in the four fragments. Along with performance details, properties of this basis set and its fit to the dataset are shown in Table 3. Fits

https://doi.org/10.1021/acs.jctc.2c01034
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

were inverse-energy-weighted, and gradients were weighted by a factor of $^1/_3$ relative to energies.

Table 3. 4-(9,11,11,8) Basis Set for AcAc[a]

property	value
monomials	3609
polynomials	24030
no. of fitted points	5454
wRMSE (pot)	22
wRMSE (grad)	16
time (pot/grad)	0.43/26.61

[a]Weighted RMSE (wRMSE) values are in cm^{-1} and cm^{-1} $bohr^{-1}$ for potentials and gradient components, respectively. Times are in seconds for an average over 10 tests using a 2.7 GHz Intel Core i7 processor. Each test involved evaluation of 5000 configurations, and the time listed is that required for evaluation of all 5000 configurations.

The database used to calculate the difference potential ΔV_{CC-LL} was generated by MOLPRO[43] calculations to obtain LCCSD(T)-F12/cc-pVTZ energies at 2151 geometries taken from the original MP2 database. We chose 1100 geometries nearest to the global minimum (denoted GM) and the saddle point for the H-atom transfer (denoted SP), and another 1051 were taken by choosing a random integer that indicated their position on a list, discarding choices whose energies were more than 30 000 cm^{-1} and stopping the selection when the requisite number of choices had been made. This selection produced a new PES with a barrier for H-atom transfer that is much closer to the direct LCCSD(T) one than the MP2 one, for which the barrier is underestimated by more than 1 kcal/mol. We additionally found that training with even as few as 430 energies from this database of 2151 LCCSD(T) energies still resulted in a PES with a barrier in much better agreement with the LCCSD(T) one.[41] Tunneling splittings due to H-atom transfer were calculated using this new PES, providing improved estimates over previous ones obtained using an MP2-based PES.

We also performed two benchmark calculations at the GM and SP. These two calculations found the optimum geometries and energies and determined as well the harmonic vibrational frequencies and normal coordinates. While the LCCSD(T)-F12 calculations for just the energy at a single geometry took approximately 30 min using 12 cores of the 2.4 GHz Intel Xeon processors, the full optimization and frequency calculations took on the order of 73 days using the same number of processors. This computational cost certainly underscores the infeasibility of obtaining LCCSD(T) calculations for a full AcAc PES.

In order to take account of the small dataset (i.e., a maximum of 2151 energies), the fit to the difference potential has to limit the number of terms in the PIP basis to be significantly less than 2151 in order to avoid overfitting. We used a PIP basis of maximum polynomial order of 2 and a symmetry designation of {1,2,5,7}, meaning that the two oxygens were allowed to permute, the five carbons were allowed to permute, and seven of the hydrogens were allowed to permute. The last hydrogen, which was not allowed to permute with the other hydrogens, was the one that is transferred between the two oxygen atoms. This basis contains just 85 PIPs and thus 85 linear coefficients to be determined by standard least-squares regression. A bonus of this small basis set is that we can examine small training datasets without concerns about overfitting. We do note that the RMS fitting error for the full database of 2151 energies is 95 cm^{-1} (0.27 kcal/mol). Other metrics of the performance of the Δ-ML approach were presented in the Supporting Information of ref 44. For example, the harmonic frequencies of the GM and SP are similar for the lower-level and Δ-ML surfaces, with the exceptions of the high-frequency OH stretch and the imaginary frequency, where the Δ-ML values are a significant improvement over those for the low-level PES. In addition, the mean absolute errors (MAEs) of the frequencies were significantly improved by the Δ-ML method.

The seven low-lying stationary points (including the GM and SP) have the energies shown in Table 4. TS(T)-I/II/III

Table 4. Energies (in cm^{-1}) of the Seven Stationary Points Relative to the Global Minimum (GM) Obtained Using the Indicated Methods; Numbers in Parentheses for the Two Δ-ML PESs Are the Sizes of the Training Datasets[44]

stationary point	LCCSD(T)[a]	$V_{LL\rightarrow CC}$ (1935)	$V_{LL\rightarrow CC}$ (430)	V_{LL}
GM	0	0	0	0
SP	1234	1218	1219	745
TS(T)-I	123	165	154	160
TS(T)-II	488	477	481	399
TS(T)-III	581	627	623	541
TS(HT)-I	1434	1299	1306	820
TS(HT)-II	1645	1359	1374	864

[a]From LCCSD(T)-F12 calculations at the MP2-optimized geometries.

are three saddle points with respect to the torsion of the two methyl rotors, and TS(HT)-I/II are two higher-order saddle points with imaginary frequencies in both the H-transfer motion and the methyl torsion. In nearly all cases, the energies of the stationary points are better captured by the $V_{LL\rightarrow CC}$ PESs than by the V_{LL} one.

We used the average of 10 DMC calculations to determine the ZPE of the corrected PES, which was 26 741 ± 7 cm^{-1}. The energy of the excited state for the H-transfer motion from 10 fixed-node DMC calculations is 26773 ± 10 cm^{-1}. Consequently, the tunneling splitting is about 32 cm^{-1}. By comparison, the splitting we obtained using the MP2-based PES (i.e., V_{LL}) is 160 cm^{-1}. Such a significant decrease in the tunneling splitting is expected because the barrier height of the $V_{LL\rightarrow CC}$ PES (1218 cm^{-1}) is significantly higher than the V_{LL} barrier height (745 cm^{-1}).

Although we performed the DMC calculations for the singly deuterated isotopologue of AcAc, the energies of the ground state and excited state are so close that the splitting is smaller than the uncertainty in the DMC calculations. Thus, we could not obtain a reliable estimate of the tunneling splitting for the deuterated AcAc using DMC.

We also applied an approximate 1D approach to obtain the tunneling splittings. The two mass-scaled 1D potentials (one for H transfer and the other for D transfer) are shown in Figure 5. These potentials were developed in a two-step procedure. First, potentials were obtained with the two methyl rotors restricted to their structures at the saddle point, since it is not practical to get correct relaxation of these using the Q_{im} approach. Thus, the barrier heights from these not fully relaxed potentials are 179 cm^{-1} less than the LCCSD(T) value. Therefore, in the second step a small scaling ("morphing") of

https://doi.org/10.1021/acs.jctc.2c01034
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

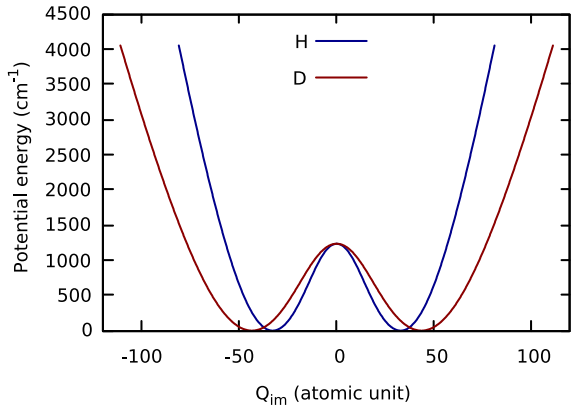


Figure 5. One-dimensional $V(Q_{im})$ paths for H and D transfer in AcAc. The barrier heights have been "morphed" to agree with the LCCSD(T)-F12 value. From ref 44. CC BY 4.0.

these potentials was performed to produce the correct barrier height. Using this 1D approach, the ground-state tunneling splittings are 37.9 and 8.2 cm^{-1} for H and D, respectively; the H splitting is in reasonably good agreement with the value of 32 cm^{-1} obtained from the DMC calculation, and the D splitting is also consistent with the fact that it is smaller than the uncertainty of the DMC calculations.

It is clearly seen from this AcAc example that the Δ-ML approach can indeed bring the PES closer to the CCSD(T) level of accuracy, especially as applied to the barrier height for H-atom transfer.

Tropolone. Tropolone, like AcAc, is a 15-atom molecule, although it is cyclic in structure, and also like AcAc, it is of interest to both theoretical and experimental chemistry because of its hydrogen transfer dynamics. The small mass of hydrogen allows it to tunnel through the barrier of a double well, resulting in a splitting of vibrational levels. Several experimental studies have been reported,[45−53] and among them, Vaccaro and co-workers recently used two-color resonant four-wave mixing techniques to measure the splitting in tropolone as a function of vibrational level.[54] Although there have been several previous theoretical approaches to this problem,[55−58] these new measurements beg for a detailed explanation, one that no doubt will require an accurate potential energy surface as well as a quantum dynamics approach for determining the splitting and its dependence on the vibrational mode. Developing a full-dimensional PES of tropolone is a challenging task because it is a 15-atom system with a symmetric double well H-transfer barrier; the surface needs to maintain this permutational symmetry.

Recently we reported a full-dimensional PES of tropolone at the DFT/B3LYP/6-31+G(d) level of theory using PIP fitting bases.[18] This was a fit to 6601 energies and their corresponding gradients. Both full and fragmented PIP (three different fragmented scheme) approaches[35,36] were employed to fit this tropolone surface. This PES was analyzed for geometry optimization and normal mode frequency calculations for GM and SP geometries. We obtained the H-transfer barrier height at about 2131 cm^{-1}, which is underestimated by 1.22 kcal/mol or 426 cm^{-1} with respect to the reported CCSD(T)/aug-cc-pVDZ result.[54] This PES was applied to compute the tunneling splitting using a fast 1D Q_{im} path tunneling method[59,60] in which the potential along the imaginary-frequency normal mode of the H atom TS (Q_{im}) is fully relaxed with respect to all other modes. We obtained tunneling splittings of around 3−5 cm^{-1} depending on the fit, whereas the experimental value is 0.96 cm^{-1}.[51,54] As can be seen, the splitting value was much larger than the experimental value, mostly due to the error in the barrier height. In addition, a full-dimensional tunneling splitting was computed via an Instanton calculation that indicated a splitting of 2.56 cm^{-1} (from a private communication with Jeremy Richardson), which is still higher than the experimental value. Considering these results, one concludes that the DFT surface is not accurate enough to estimate the H-transfer barrier height and tunneling dynamics of tropolone. A correction to this DFT PES is needed to achieve the CCSD(T) level of accuracy.

Because of the very steep scaling of CCSD(T) theory (N^7, where N is the number of basis functions), the Δ-ML approach is a method that uses a small number (several hundreds to thousand) of CCSD(T) energies to correct a PES based on DFT electronic energies and gradients. For systems with more than 10 atoms, it becomes arduous job to compute CCSD(T) energies for a thousand geometries using a Dunning basis. For example, a single-point energy computation of the 15-atom tropolone molecule at the CCSD(T)/aug-cc-pVTZ level takes about 2063 min on a single 16-core machine. Therefore, correcting the DFT tropolone PES by using, say, 1000 CCSD(T) energies becomes computationally very expensive and even prohibitive. Of course, a local method such as the LCCSD(T) method applied for AcAc could be used. However, our research in progress uses a fragmented approach. Several fragmentation-based methods have been proposed during the last two decades for treating large molecules. These methods have been extensively tested and benchmarked for obtaining the molecular energies, gradients, and Hessian matrix. An overview of such fragmented methods is given in ref 61. One such fragmented method, namely, the molecular tailoring approach (MTA), has been developed by Gadre and co-workers and appraised for various large molecules and molecular clusters.[62−66] Very recently, this MTA method was successfully applied to develop a full-dimensional PES for the AcAc molecule, and this first-ever full-dimensional PES of a 15-atom system is reported at the CCSD(T)/aVTZ level of theory.[67] From this PES the H-transfer barrier is estimated as 3.02 kcal/mol, in excellent agreement with the benchmarked barrier of 3.19 kcal/mol. Therefore, This MTA approach is being taken to correct the DFT tropolone surface in collaboration with Gadre group.

PES Timing Comparison. Table 5 shows the computation times for evaluation of the low-level DFT-based PES and the correction PES (ΔV_{CC-LL}). These calculations were carried out on a single core of an Intel Xeon 2.40 GHz processor with 64 GB of RAM. Clearly, the computation of ΔV_{CC-LL} is much faster than the computation of the V_{LL} PES because less-complex and lower-order PIPs are employed to fit the correction PES. Thus, the additional cost to bring the DFT-

Table 5. Timings (in s) for 100 000 Potential Evaluations for Hydronium Ion, *N*-Methylacetamide (NMA), and Ethanol

molecule	V_{LL}	ΔV_{CC-LL}	$V_{LL\rightarrow CC}$
H_3O^+	0.04	0.01	0.05
NMA	2.05	0.12	2.17
ethanol	3.12	0.06	3.18

https://doi.org/10.1021/acs.jctc.2c01034
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

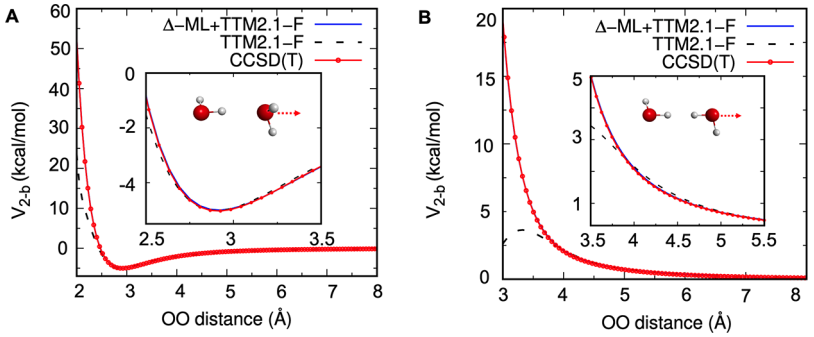


Figure 6. Comparisons of the Δ-ML-corrected TTM2.1-F 2-b potential and direct CCSD(T)/CBS energies for (A) an attractive cut and (B) a repulsive cut.

based PES to the CCSD(T) level of accuracy is a small fraction of the cost of evaluating the DFT PES.

Δ-MACHINE LEARNING POLARIZABLE FORCE FIELDS

The above material is largely a review of our recent work on Δ-ML for PESs of isolated molecules. This section is a perspective on extending Δ-ML to force fields (FFs) describing large molecules and many-atom or many-molecule nonreactive interactions. Such FFs are pervasive in computational chemistry, biology, and materials research. The extension is applied and tested extensively for a sophisticated FF for water. As noted above, this approach was recently presented by us[33] with the specific goal of correcting the 4-b water interaction in the MB-pol potential.

Polarizable force fields with a focus on efficiency have been reviewed recently.[68−70] These force fields are many-body, at least with regard to how the polarizability is treated. They all have two-body terms, such as a Lennard-Jones or exp-6 potential. This general form immediately suggests the following expression for a many-body Δ correction:[33]

$$V_{\Delta\text{-ML+MB-FF}} = V_{\text{MB-FF}} + \sum_{i>j}^{N} \Delta V_{2\text{-b}}(i,j) + \sum_{i>j>k}^{N} \Delta V_{3\text{-b}}(i,j,k) + \sum_{i>j>k>l}^{N} \Delta V_{4\text{-b}}(i,j,k,l) + \cdots \quad (2)$$

where $V_{\text{MB-FF}}$ is the force field and the $\Delta V_{n\text{-b}}$ are the many-body corrections to the MB-FF many-body terms. These are given by the difference between CCSD(T) and MB-FF n-body (n-b) interaction energies. In general the n-b interaction energy is obtained from a cluster of n monomers. For example, the 2-b interaction is obtained by calculating the total energy of the dimer (two monomers) and subtracting all of the 1-b interactions from the total energy. It should be noted that for simplicity we assume that an accurate 1-b term (e.g., the single monomer) is given in the MB-FF. It should also be noted that if the MB-FF is given by an MB expansion, the corrections $\Delta V_{n\text{-b}}$ are obvious term by term.

Here we investigate this approach for the TTM2.1-F force field for water[34] up to 4-b interactions. In this case, $V_{\text{MB-FF}}$ is $V_{\text{TTM2.1}}$ and the 1-b term is the accurate isolated H_2O potential of Partridge and Schwenke.[71] Specifically, we corrected the 2-b, 3-b, and 4-b interactions targeting the CCSD(T) level of accuracy. The datasets of CCSD(T) energies for 2-b, 3-b, and 4-b interactions were recently reported by us in developing a new water potential, q-AQUA,[72] that is exactly truncated at the 4-b level. This potential was shown to be very accurate from clusters to the condensed phase. The role of 4-b interactions has been well-established in the literature, and the interested reader is referred to recent refs 73 and 74 and references cited therein. Based on this work, we recently created a 4-b CCSD(T) energy dataset[74] that was fit using PIPs.[2,9,10,75] That fit, together with PIP fits to extensive new datasets for 2-b and 3-b CCSD(T) energies, constitute the q-AQUA potential, as noted already. However, it is certainly of interest to incorporate higher-body interactions, and that is achieved using the current approach, which relies on TTM2.1 for 5- and higher-body interactions.

In principle, these datasets and the PIP fitting approach can be used to create the $\Delta V_{n\text{-b}}$ needed here. This was done initially, but the fitting error of the dominant $\Delta V_{2\text{-b}}$ interaction was 30 cm^{-1}. This was mainly due to the major failure of the TTM2.1 2-b interaction at short range, as shown below. Reducing the fitting error is straightforward by increasing the maximum polynomial order of the PIP basis, but this comes at a big increase in the computation time. Instead, we represent $\Delta V_{2\text{-b}}$ as the direct difference between q-AQUA and TTM2.1 $V_{2\text{-b}}$ interactions. For the 3-b and 4-b interactions, PIP fits to $\Delta V_{3\text{-b}}$ and $\Delta V_{4\text{-b}}$ were done. Briefly, the Δ-ML 3-b correction potential, $\Delta V_{3\text{-b}}$, was fitted from a database of 42 710 reference energies of water trimer structures selected from the 3-b database in ref 72 with maximum O−O distance in the range of [2.0, 7.5] Å. For each trimer structure, the $\Delta V_{3\text{-b}}$ correction energy was computed as the difference between BSSE-corrected CCSD(T)-F12a/aVTZ and TTM2.1-F 3-b energies. Using fourth-order 222111-symmetry PIPs, the fitting RMSE for the whole dataset is 9 cm^{-1}. For the Δ-ML 4-b correction potential, $\Delta V_{4\text{-b}}$, we employed the same PIP bases as described in q-AQUA 4-b and the Δ-ML 4-b to MB-pol[33] using a dataset of 3692 tetramer energies computed at the CCSD(T)-F12/haTZ level of theory. The fitting RMSE for this correction potential is 6.3 cm^{-1}. More details of the $\Delta V_{3\text{-b}}$ and $\Delta V_{4\text{-b}}$ PES fits are provided in the Supporting Information. Below we present systematic assessments of the performance of the 2-b, 3-b, and 4-b interactions from CCSD(T) energies, TTM2.1, and $\Delta V_{n\text{-b}}$. Large differences between the former two are seen.

Interaction Potentials. The first results focus on the large 2-b interaction, which is the interaction of highest magnitude. Figure 6 shows two cuts of the 2-b potential as a function of the O−O distance, where comparisons are made between the CCSD(T)/CBS points, the TTM2.1-F interaction potential, and the Δ-ML corrected TTM2.1-F potential. As can be seen, TTM2.1 becomes rapidly inaccurate as the O−O distance

https://doi.org/10.1021/acs.jctc.2c01034
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

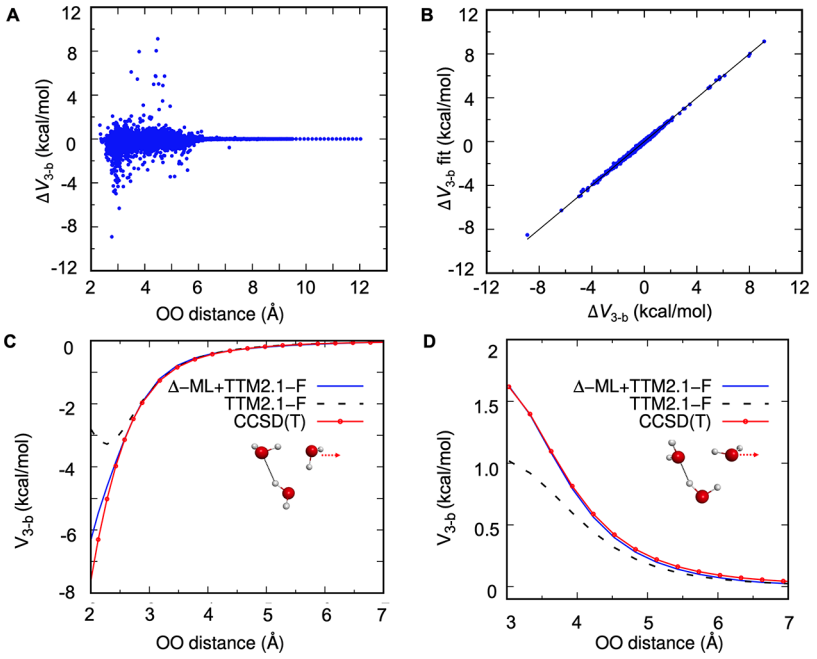


Figure 7. (A) Distribution of $\Delta V_{3\text{-b}}$ correction energies ($V_{3\text{-b}}^{\text{CCSD(T)}} - V_{3\text{-b}}^{\text{TTM2.1-F}}$) vs O−O distance. (B) Correlation between fitted $\Delta V_{3\text{-b}}$ correction energies and reference data. (C, D) Comparisons of the Δ-ML-corrected TTM2.1-F 3-b potential, the TTM2.1-F 3-b potential, and the direct CCSD(T) energies for (C) an attractive cut and (D) a repulsive cut. All of the CCSD(T) energies were calculated at CCSD(T)-F12/aVTZ level of theory.

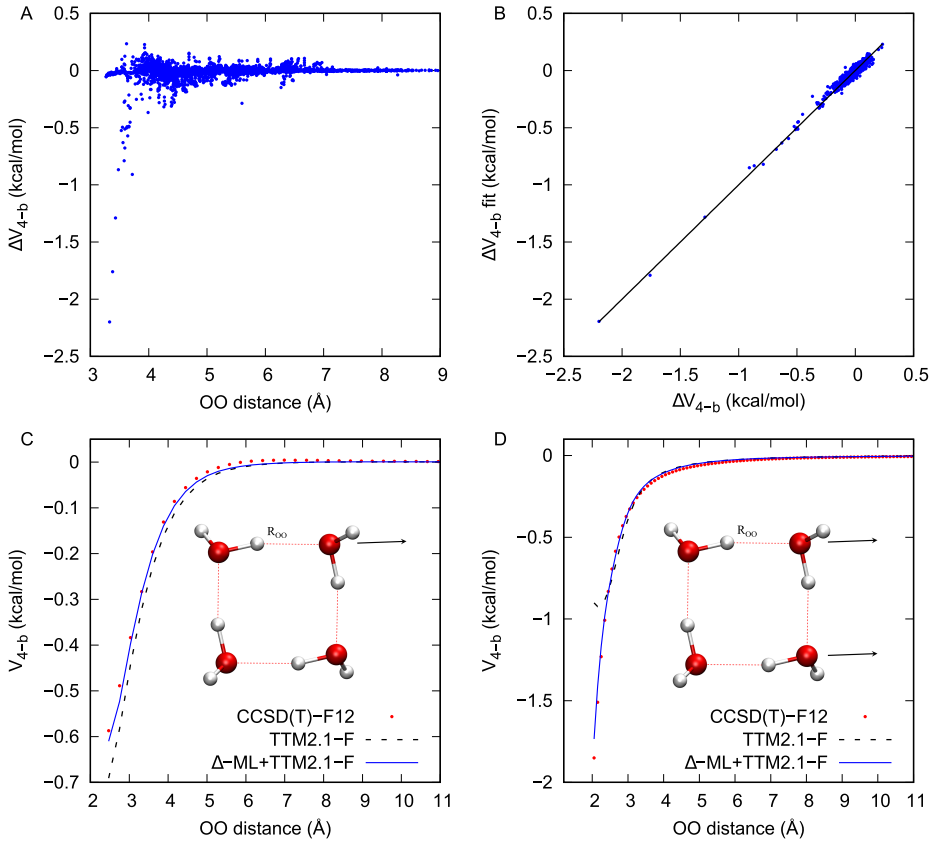


Figure 8. (A) Distribution of $\Delta V_{4\text{-b}}$ correction energies ($V_{4\text{-b}}^{\text{CCSD(T)}} - V_{4\text{-b}}^{\text{TTM2.1-F}}$) as a function of the maximum O−O distance in a tetramer. (B) Correlation between fitted $\Delta V_{4\text{-b}}$ energies and reference data. (C, D) Comparisons of the Δ-ML corrected TTM2.1-F 4-b potential, the TTM2.1-F 4-b potential, and the direct CCSD(T)-F12 energies for (C) a monomer−trimer cut and (D) a dimer−dimer cut.

decreases for both geometries shown. It is also clear that the corrected 2-b is in excellent agreement with direct CCSD(T)/CBS energies. It should be noted that for clusters and the bulk there are relatively few 2-b interactions, and therefore, the

https://doi.org/10.1021/acs.jctc.2c01034
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 6. 2-b, 3-b, 4-b, and Total Dissociation Energies (in kcal/mol) for Various Water Hexamer Isomers

isomer	CCSD(T)[a]	Δ-ML + TTM2.1-F	TTM2.1-F	CCSD(T)[b]	Δ-ML + TTM2.1-F	TTM2.1-F
		D_e			2-b Energy	
Prism	45.92	45.90	43.95	−38.94	−39.07	−38.80
Cage	45.67	45.59	44.63	−38.47	−38.49	−39.15
Book 1	45.20	45.04	44.28	−36.02	−35.97	−37.51
Book 2	44.90	44.78	44.10	−36.13	−36.04	−37.69
Bag	44.30	44.30	43.38	−35.28	−35.38	−37.09
Ring	44.12	44.01	43.38	−32.71	−32.59	−35.04
Boat 1	43.13	43.18	42.82	−32.30	−32.27	−34.82
Boat 2	43.07	43.12	43.05	−32.24	−32.21	−34.89
MAE	–	0.07	0.84	–	0.07	1.64
		3-b Energy			4-b Energy	
Prism	−8.70	−8.73	−6.96	−0.66	−0.58	−0.69
Cage	−8.97	−9.04	−7.29	−0.53	−0.47	−0.62
Book 1	−10.38	−10.36	−8.01	−1.08	−1.07	−1.05
Book 2	−10.11	−10.15	−7.85	−1.00	−1.00	−0.98
Bag	−10.35	−10.36	−7.80	−1.16	−1.09	−1.03
Ring	−11.78	−11.81	−8.91	−1.78	−1.70	−1.52
Boat 1	−11.34	−11.44	−8.69	−1.63	−1.57	−1.42
Boat 2	−11.34	−11.41	−8.79	−1.61	−1.57	−1.43
MAE	/	0.05	2.33	/	0.05	0.12
		Higher-Body (>4-b) Energy				
Prism	0.06	0.06	0.06			
Cage	0.01	0.03	0.03			
Book 1	−0.04	−0.06	−0.06			
Book 2	−0.02	−0.04	−0.04			
Bag	−0.01	−0.06	−0.06			
Ring	−0.20	−0.23	−0.23			
Boat 1	−0.17	−0.21	−0.21			
Boat 2	−0.17	−0.21	−0.21			
MAE	–	0.03	0.03			

[a]CCSD(T)/CBS data from ref 80. [b]CCSD(T)/CBS data from ref 78.

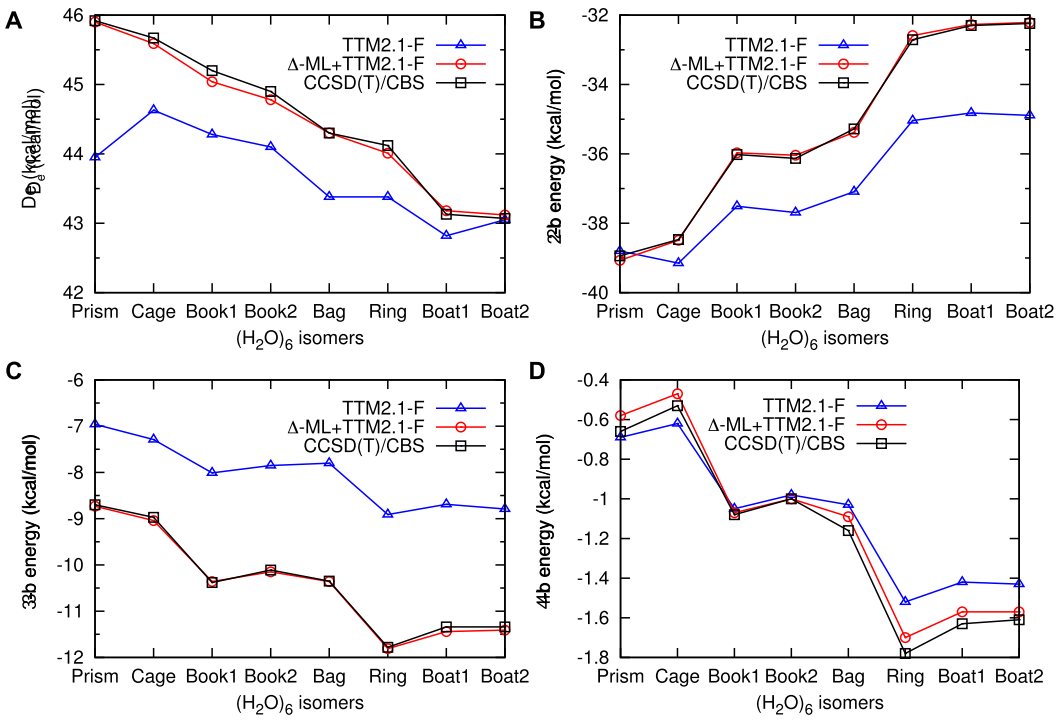


Figure 9. (A) Binding energies (D_e), (B) 2-body energies, (C) 3-body energies, and (D) 4-body energies for water hexamer isomers from TTM2.1-F, Δ-ML-corrected TTM2.1-F, and benchmark CCSD(T) calculations (taken from refs 80 and 78).

https://doi.org/10.1021/acs.jctc.2c01034
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

direct calculation of TTM2.1 and q-AQUA 2-b interactions does not take a significant fraction of the total computational time.

Figure 7A shows a scatter plot of the $\Delta V_{3\text{-b}}$ data (not a fit) versus the maximum O−O distance in a trimer. As can be seen, many of these are less than 1 kcal/mol; however, a large number are several kcal/mol and as large as 10 kcal/mol. These are clearly significant errors in the TTM2.1 3-b interaction. Figure 7B shows the correlation plot of the precise $\Delta V_{3\text{-b}}$ PIP fit (described in detail below). Figure 7C,D shows two 1D cuts when one of the monomers is moved away from the dimer. As one can see, the $\Delta V_{3\text{-b}}$ energy can be very large at short range, indicating the inaccuracy of the TTM2.1-F force field in this range; this is also clearly shown in the two potential cuts. The fitting brings the 3-b interaction into much better agreement with the CCSD(T) energies. Similarly, Figure 8A shows a scatter plot of the $\Delta V_{4\text{-b}}$ data (not a fit) versus the maximum O−O distance in a tetramer, and again, significant errors in the TTM2.1-F 4-b can be seen. Figure 8B shows the correlation plot of the $\Delta V_{4\text{-b}}$ PIP fit. Figure 8C,D shows two 1D cuts when one of the monomers and a dimer are moved away from the remaining water molecules; again, the $\Delta V_{4\text{-b}}$ correction brings TTM2.1-F into better agreement with CCSD(T)-F12 energies, especially at short range.

Energy Analysis of the Water Hexamer and 20-mer Isomers. The isomers of the water hexamer play a major role in both experimental and theoretical studies of water clusters. The lowest-energy isomers are noncyclic.[22,76] The energies of eight isomers have become a standard test of the fidelity of a water FF.[77] A detailed analysis of the electronic energies of the eight isomers is given Table 6. The electronic dissociation energies, D_e, for the corrected FF are in much better agreement with the benchmark CCSD(T) results than the TTM2.1 ones. In particular, the Prism is incorrectly predicted to be number 4 in dissociation. The corrected FF also has unprecedented agreement with the CCSD(T) results compared to MB-pol[78] and is even slightly better than q-AQUA.[72] The detailed *n*-body analysis presented in Table 6 shows the accuracy of the Δ-corrected FF for each *n*, and impressively, for $n > 4$ TTM2.1 provides the additional small term that results excellent accuracy of the Δ-ML-corrected FF. This analysis is presented graphically in Figure 9. The results shown in Table 6 are even more accurate than those from our q-AQUA potential,[72] which is truncated at the 4-b level, and also more accurate than the MB-pol potential.[78]

Recently, Heindel et al. reported benchmark electronic dissociation energies of several 20-mers.[79] This is a stringent test of a force field and was done for TTM2.1-F and also for the MB-pol FF. For the latter, the energies are roughly 5 kcal/mol lower than the benchmark values. This is just 2.5% of the D_e of roughly 200 kcal/mol. A comparison of these benchmark energies with those from TTM2.1-F and the Δ-ML-corrected FF are given in Table 7. Using TTM2.1-F optimized geometries, the dissociation energies are very close to the MP2 benchmark ones. However, it should be noted that upon using the same MP2/aVDZ-optimized geometries, the TTM2.1-F energies are ∼9 kcal/mol higher than the MP2/aV5Z data. This indicates that the geometry optimization using the TTM2.1-F force field results in significant changes in the geometries of these isomers. With the Δ-ML-corrected TTM2.1-F potential, the structures of the three isomers do not undergo significant changes during optimization, and the dissociation energies using the optimized geometries are close to both the MP2 and CCSD(T) benchmark results.

Table 7. Electronic Dissociation Energies (in kcal/mol) of Three $(H_2O)_{20}$ Isomers

isomer	MP2/aV5Z	MP2/CBS	TTM2.1-F	Δ-ML + TTM2.1-F
A3	−202.1	199.2 ± 0.5[a] (−200.8 ± 2.1[b])	−202.2[c] (−193.4[d])	−196.2[c] (−194.7[d])
A2d	−202.1	n.a.	−202.0[c] (−193.4[d])	−198.1[c] (−196.7[d])
9	−201.5	n.a.	−202.2[c] (−193.6[d])	−197.0[c] (−195.6[d])

[a]From ref 79. [b]CCSD(T)/CBS dissociation energy from ref 79. [c]Energies obtained using PES-optimized geometries. [d]Energies obtained using MP2/aVDZ-optimized geometries.

Quantum Simulations of D_0 Energy and IR Spectra of Gas-Phase Clusters. Unconstrained DMC calculations were performed for the ZPEs of the water dimer and trimer using the Δ-CCSD(T)-corrected TTM2.1-F FF. Details of these calculations, as done by us, have been given numerous times, and the interested reader can consult recent papers[81,82] for details. It should be recalled that these are rigorous "exact" quantum calculations and, when combined with the electronic dissociation energy (D_e) and exact ZPE of the water monomer, give the exact dissociation energy D_0. Table 8 shows D_e, the ZPEs of the reactants and products, and the ZPE-corrected dissociation energy (D_0) for the following dissociations: $(H_2O)_2 \rightarrow 2H_2O$, $(H_2O)_3 \rightarrow 2H_2O + H_2O$, and $(H_2O)_3 \rightarrow 3H_2O$. As can be seen, the agreement with experiment is excellent. The agreement is also excellent with DMC calculations of D_0 for the dimer using the HBB2[83] and MB-pol[84] potentials, both of which predict a value of 1101 cm^{-1}. These values are also in excellent agreement with the value of D_0 reported for the dimer[85] using the recent flexible CCpol-8sf potential.[86] For the water trimer, the present results are in excellent agreement with previous DMC calculations using WHBB[87] (2724 cm^{-1}) and MB-pol[84] (2693 cm^{-1}). Also there is very good agreement for dissociation to three monomers, namely, 3854 cm^{-1} for WHBB and 3794 cm^{-1} for MB-pol. We do not show results for TTM2.1, as the electronic dissociation eneriges for the dimer and trimer already show significant errors compared to the above potentials.

We also note that we performed exploratory DMC calculations using 20 000 walkers and 25 000 steps for the hexamer using the new PES and did not find holes (i.e., regions of spuriously large negative energies). The ZPE of the Cage isomer is roughly 70 cm^{-1} lower than that of the Prism isomer (both ZPEs are referenced to the electronic energy of the Prism isomer), in semiquantitative agreement with the results using q-AQUA,[72] though the uncertainties here are much larger due to the smaller numbers of walkers and steps.

Finally, we demonstrate the efficacy of the correction approach for quantum calculations of the IR spectra of two isomers of the hexamer. These spectra have played a major role in the fruitful interaction between theory and experiment.[76,90] In particular, the IR spectra of the higher-energy Ring[91,92] and Book isomers[93,94] have attracted the attention of theorists. Previously, we applied the WHBB potential and DMS in VSCF/VCI calculations of the IR spectra of numerous isomers in the region of the intramolecular bend and stretch, where experiments have been reported.[95] Those earlier results showed very good agreement with experiment and settled a

https://doi.org/10.1021/acs.jctc.2c01034
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 8. Electronic Dissociation Energies (D_e), ZPEs of the Reactants and Products, and ZPE-Corrected Dissociation Energies (D_0) for Water Dimer and Trimer (in cm^{-1})

dissociation	D_e	ZPE (react.)	ZPE (prod.)	D_0	D_0 (expt.)
$(H_2O)_2 \rightarrow 2H_2O$	1739	9912 ± 1	9272 ± 1	1099 ± 2	1105 ± 10[a]
$(H_2O)_3 \rightarrow 2H_2O + H_2O$	3769	15613 ± 4	14548 ± 2	2704 ± 6	2650 ± 150[b]
$(H_2O)_3 \rightarrow 3H_2O$	5508	15613 ± 4	13908 ± 2	3803 ± 6	NA

[a]From ref 88. [b]From ref 89.

controversy about the origin of a strong hydrogen-bonding OH stretch band in the Ring isomer. Here these spectra provide another important demonstration of the success of the current approach. Results are given in Figure 10 for the Ring isomer and Figure 11 for the Book isomer. First consider Figure 10, where the experimental spectrum is divided into two regions due to overlapping experimental spectra of the Ring and Book isomers, as determined by our previous calculations[95] and the present ones using the Δ-ML-corrected PES.

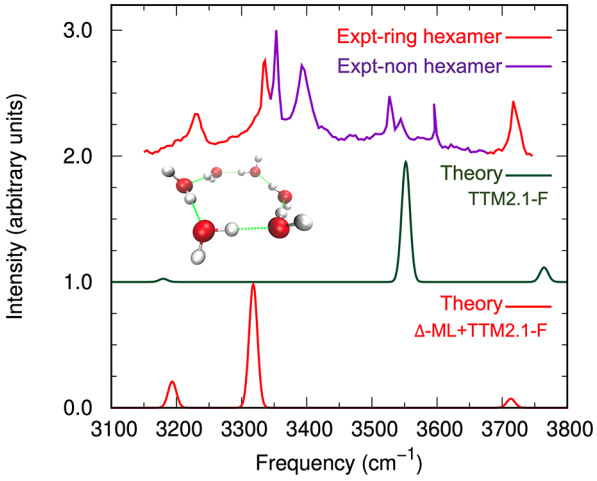


Figure 10. Theoretical VSCF/VCI spectra of the water Ring hexamer using the TTM2.1-F and Δ-ML-corrected TTM2.1-F potentials compared to the experimental spectrum from refs 92 and 97. The purple portion of the experimental spectrum is due to other water clusters smaller than the hexamer.

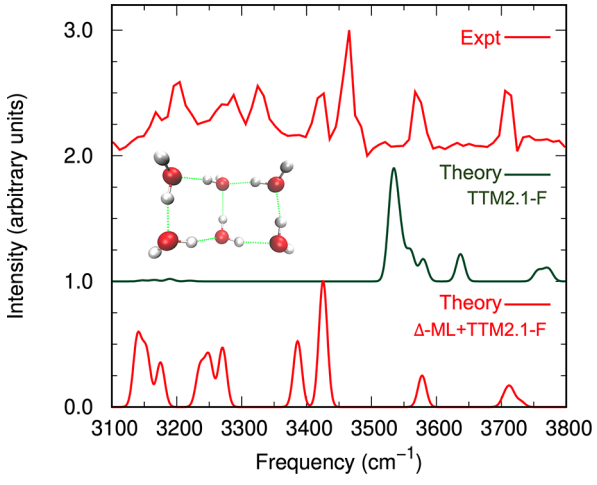


Figure 11. Theoretical VSCF/VCI spectra of the water Book hexamer using the TTM2.1-F and Δ-ML-corrected TTM2.1-F potentials compared with the experimental spectrum from ref 94.

It should be noted that the peak at roughly 3180 cm^{-1} is due to the overtone of the monomer bend and did settle the controversy about the origin of the experimental peak seen at around 3220 cm^{-1}. More details and discussion can be found in ref 95. For the present purpose, we note the poor agreement with experiment using the TTM2.1 potential, with the WHBB DMS, in identical VSCF/VCI calculations. The improvement of that spectrum just using the 2-b correction (not shown) is significant but not in quantitative agreement with the Δ-ML-corrected spectrum. The corresponding comparisons with experiment for the Book isomer are given in Figure 11. As can be seen, theory, with the Δ-corrected potentials, reproduces the complex experimental spectrum well, whereas the spectrum obtained using TTM2.1 does not. For completeness we note that the dipole moment surface is a 1- and 2-body one that has been described and tested previously.[96]

Δ-ML Correction Timings. Here we examine the additional computational cost in making the Δ-ML correction to the TTM2.1-F force field. This is an important aspect of ML FFs, as noted in the Introduction. Table 9 shows the computational cost of the Δ-ML-corrected TTM2.1-F model for both energy and gradient calculations for a 256-water system using a single core or multiple cores of a 2.4 GHz Intel Xeon processor, where the cost for each correction term $\Delta V_{n\text{-b}}$ is also listed. The TTM2.1-F force field code is an optimized and parallelized one that is computationally more efficient than the original TTM2.1-F force field code for gas-phase cluster calculations. The 256-water structure was chosen from an MD simulation of liquid water using our recently developed water potential, q-AQUA.[72] The number of terms given is less than the factorial result due to the use of the finite-range switching function. For example, the total number of 3-b interaction terms in this 256-mer water system is $\binom{256}{3} = 2{,}763{,}520$, but only 3-b interactions with a maximum O−O distance smaller than 7.0 Å are considered for correction, which results in only 20 790 $\Delta V_{3\text{-b}}$ terms. It is also straightforward to do multicore processing (using OpenMP) of the $\Delta V_{n\text{-b}}$ terms, which is

Table 9. Computational Cost of the Δ-ML-Corrected TTM2.1-F Potential for Energy and Gradient Calculations of a 256-Water System

		time for energy (s)		time for energy +gradient (s)	
component	number	one core	eight cores	one core	eight cores
TTM2.1-F	–	1.09	0.28[a]	1.09	0.28[a]
$\Delta V_{2\text{-b}}$	3816	0.25	0.03	0.59	0.07
$\Delta V_{3\text{-b}}$	20790	0.58	0.08	2.47	0.35
$\Delta V_{4\text{-b}}$	28786	0.31	0.04	1.07	0.15
total		2.23	0.43	5.22	0.85

[a]The current TTM2.1-F force field code calculates the gradients by default.

https://doi.org/10.1021/acs.jctc.2c01034
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

clearly effective as shown in Table 9. Using eight cores, the additional time to evaluate the Δ correction is almost negligible for energy calculations and almost 2 times the TTM2.1-F force field for gradient calculations. The corresponding timing for an energy plus gradient calculation on eight cores for the q-AQUA potential is 0.97 s on the same cluster (reported in Table S5 of ref 72). This is similar to that for the present Δ-corrected FF.

To summarize this section, we have shown than Δ-ML can be successfully applied to force fields that are explicitly or implicitly many-body. The Δ-ML method used was at the gold standard CCSD(T) level, where we made use of our recent extensive 2-b, 3-b, and 4-b datasets. Explicit high-dimensional $\Delta V_{n\text{-b}}$ correction terms were determined. The 2-b correction $\Delta V_{2\text{-b}}$ is given explicitly by the 2-b term in our recent q-AQUA potential minus the 2-b interaction in TTM2.1. For the 3-b and 4-b corrections, the $\Delta V_{n\text{-b}}$ are permutationally invariant polynomial fits to 3-b and 4-b energy differences between the corresponding CCSD(T)and TTM2.1 energies. These fits were to high-dimensional datasets, i.e., for the 4-b correction, $\Delta V_{4\text{-b}}$ is 12-atom full-dimensional and is a function of all 66 Morse-transformed internuclear distances. The fits are invariant with respect to all permutations of monomers and with respect to the two H atoms of each monomer.

The new Δ-ML-corrected TTM2.1 force field was shown to be of unprecedented accuracy for the electronic energies of the water hexamer isomers and isomers of the 20-mer. This new force field is also robust for unconstrained diffusion Monte Carlo calculations. These were performed to obtain rigorous dissociation energies of the water dimer and trimer and were shown to be in excellent agreement with previous calculations and experiment. Finally, VSCF/VCI calculations of the IR spectra of the hexamer Ring and Book isomers were shown to be in very good agreement with experiment.

Future work will focus on applications of the CCSD(T)-corrected TTM2.1 force field for the condensed phase and for DMC calculations of large water clusters. These will benefit from needed optimization and parallelization of the TTM2.1 software, which is our next specific goal.

The success of the correction to the TTM2.1 force field leads us to conclude tentatively that the approach will be applicable to general polarizable force fields, which continue to be developed and improved (see, e.g., ref 98).

DISCUSSION

The Δ-ML correction described and demonstrated here certainly appears to be successful. For the isolated molecules, the basic reason for this success is clear from the plots showing that the correction energies are much smaller than the energies (either CCSD(T) or DFT). This means that the correction is small (on the order of tens to a few thousands of wavenumbers). In addition, it is generally smoothly varying. These favorable aspects of the problem mean that a small number of points and a low-order PIP fit to them can achieve a highly precise fit. Clearly, however, the general issue of how to pick configurations for electronic structure calculations also applies to selecting configurations for expensive CCSD(T) calculations. This is a not a simple issue, and there is not a simple response to it. However, as discussed by us in a recent Perspective in the *Journal of Chemical Physics*,[25] the selection of configurations is driven by the goal of the usage of the PES. We generally use PESs for dynamics calculations, ranging from quantum dynamics to high-energy (quasi)classical trajectory calculations of reaction dynamics. These often require both energies at a large number of geometries and also high accuracy, particularly near barriers to reaction. The Δ-ML method allows us to combine lower-level theory (DFT, MP2) for a large number of geometries with higher-level theory (CCSD(T)) at a few geometries so that the accuracy at all geometries is greatly improved. Also, we note that a similar approach, subsequent to ours, was reported by Liu and Li, who developed a Δ-ML potential for the HO_2 self-reaction.[99] Finally, we note that modifying a PES has been done in the past for small molecules. A method using scaled coordinates[100] that has been called "morphing"[101] is one example.

For the application to force fields, the many-body correction approach is qualitatively the generalization of a single-correction PES to a molecular PES to several many-body correction PESs. However, the $\Delta V_{n\text{-b}}$ potentials are based on much larger datasets, which in this case were already available from our recent work in developing the q-AQUA water potential.[72] It should also be noted that the basic approach of blending a force field with ab initio many-body potentials is not new. Again for water, this approach is the basis of the WHBB potential,[102] where 2-b and 3-b PIP PESs are combined with the TTM3F potential[103] for higher-body interactions. The subsequent MB-pol potential[104] combines ab initio PIP 2-b and 3-b PESs with the TTM4F potential.[105] It should be noted that both of these potentials also use the Partridge−Schwenke potential for the monomer,[71] as is done in the present work. Here we have extended this approach to 4-b interactions for the first time. We note that we recently corrected the MB-pol PES by adding a correction to the 4-b interaction.[33]

SUMMARY

This Perspective has focused on a Δ-machine learning method to bring DFT-based potential energy surfaces close to the CCSD(T) level of accuracy. This was demonstrated for H_3O^+, as a pedagogical test case, and then for nine-atom ethanol, 12-atom *N*-methylacetamide, and 15-atom acetylacetone. We have also described work in progress for an application to 15-atom tropolone. A major extension of the Δ-machine learning approach for force fields was demonstrated in detail for the TTM2.1 force field for water. The corrected FF was shown to result in major improvements over the TTM2.1 FF for a variety of important properties of clusters.

Finally, it is important to note that the success of the Δ-machine learning approach described here is limited by the cost of doing a single-point CCSD(T) calculation. This is the case because any correction approach, including transfer learning, requires a minimum of several hundred CCSD(T) calculations. Clearly, there is motivation to correct low-level energies to the CCSD(T) level of accuracy. Fortunately, work in this direction is underway.[106−111]

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.2c01034.

Details of the $\Delta V_{3\text{-b}}$ and $\Delta V_{4\text{-b}}$ PES fits and sample potential cuts for 2-b, 3-b, and 4-b interactions (PDF)

https://doi.org/10.1021/acs.jctc.2c01034
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

AUTHOR INFORMATION

Corresponding Authors

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

Riccardo Conte − *Dipartimento di Chimica, Università Degli Studi di Milano, 20133 Milano, Italy;* orcid.org/0000-0003-3026-3875; Email: riccardo.conte1@unimi.it

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

Qi Yu − *Department of Chemistry, Yale University, New Haven, Connecticut 06520, United States;* orcid.org/0000-0002-2030-0671; Email: q.yu@yale.edu

Authors

Chen Qu − *Independent Researcher, Toronto, Canada 66777*

Apurba Nandi − *Department of Chemistry and Cherry L. Emerson Center for Scientific Computation, Emory University, Atlanta, Georgia 30322, United States;* orcid.org/0000-0002-6191-5584

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.2c01034

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Kristina Herman and Sotiris Xantheas for sending the original version of the TTM2.1-F force field and George Fanourgakis for sending the parallelized version. J.M.B. thanks NASA (80NSSC20K0360) for financial support. R.C. thanks Università degli Studi di Milano ("PSR, Azione A Linea 2 - Fondi Giovani Ricercatori") for support. Q.Y. thanks Professor Sharon Hammes-Schiffer and the National Science Foundation (Grant CHE-1954348) for support.

REFERENCES

(1) Pinheiro, M.; Ge, F.; Ferré, N.; Dral, P. O.; Barbatti, M. Choosing the right molecular machine learning potential. *Chem. Sci.* **2021**, *12*, 14396−14413.

(2) 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*, 044120.

(3) Chmiela, S.; Tkatchenko, A.; Sauceda, H. E.; Poltavsky, I.; Schütt, K. T.; Müller, K.-R. Machine learning of accurate energy-conserving molecular force fields. *Sci. Adv.* **2017**, *3*, e1603015.

(4) Qu, C.; Yu, Q.; Bowman, J. M. Permutationally invariant potential energy surfaces. *Annu. Rev. Phys. Chem.* **2018**, *69*, 151−175.

(5) Qu, C.; Bowman, J. M. An *ab initio* potential energy surface for the formic acid dimer: zero-point energy, selected anharmonic fundamental energies, and ground-state tunneling splitting calculated in relaxed 1−4-mode subspaces. *Phys. Chem. Chem. Phys.* **2016**, *18*, 24835−24840.

(6) Meyer, J.; Tajti, V.; Carrascosa, E.; Győri, T.; Stei, M.; Michaelsen, T.; Bastian, B.; Czakó, G.; Wester, R. Atomistic dynamics of elimination and nucleophilic substitution disentangled for the F^- + CH_3CH_2Cl reaction. *Nat. Chem.* **2021**, *13*, 977−981.

(7) Tajti, V.; Czako, G. Vibrational mode-specific dynamics of the F + CH_3CH_2Cl multi-channel reaction. *Phys. Chem. Chem. Phys.* **2022**, *24*, 8166−8181.

(8) Yin, C.; Tajti, V.; Czakó, G. Full-dimensional potential energy surface development and dynamics for the HBr + C_2H_5 → $Br(^2P_{3/2})$ + C_2H_6 reaction. *Phys. Chem. Chem. Phys.* **2022**, *24*, 24784−24792.

(9) Braams, B. J.; Bowman, J. M. Permutationally invariant potential energy surfaces in high dimensionality. *Int. Rev. Phys. Chem.* **2009**, *28*, 577−606.

(10) *MSA Software with Gradients*, 2019. https://github.com/szquchen/MSA-2.0 (accessed 2019-01-20).

(11) Győri, T.; Czakó, G. Automating the Development of High-Dimensional Reactive Potential Energy Surfaces with the robosurfer Program System. *J. Chem. Theory Comput.* **2020**, *16*, 51−66.

(12) Czakó, G.; Győri, T.; Papp, D.; Tajti, V.; Tasi, D. A. First-Principles reaction dynamics beyond six-atom systems. *J. Phys. Chem. A* **2021**, *125*, 2385−2393.

(13) Lambros, E.; Dasgupta, S.; Palos, E.; Swee, S.; Hu, J.; Paesani, F. General many-body framework for data-driven potentials with arbitrary quantum mechanical accuracy: water as a case study. *J. Chem. Theory Comput.* **2021**, *17*, 5635−5650.

(14) Moberg, D. R.; Jasper, A. W. Permutationally invariant polynomial expansions with unrestricted complexity. *J. Chem. Theory Comput.* **2021**, *17*, 5440−5455.

(15) Moberg, D. R.; Jasper, A. W.; Davis, M. J. Parsimonious Potential Energy Surface Expansions Using Dictionary Learning with Multipass Greedy Selection. *J. Phys. Chem. Lett.* **2021**, *12*, 9169−9174.

(16) Qu, C.; Bowman, J. M. Quantum approaches to vibrational dynamics and spectroscopy: is ease of interpretation sacrificed as rigor increases? *Phys. Chem. Chem. Phys.* **2019**, *21*, 3397−3413.

(17) Conte, R.; Qu, C.; Houston, P. L.; Bowman, J. M. Efficient generation of permutationally invariant potential energy surfaces for large molecules. *J. Chem. Theory Comput.* **2020**, *16*, 3264−3272.

(18) Houston, P. L.; Conte, R.; Qu, C.; Bowman, J. M. Permutationally invariant polynomial potential energy surfaces for tropolone and H and D atom Tunneling Dynamics. *J. Chem. Phys.* **2020**, *153*, 024107.

(19) McCoy, A. B.; Braams, B. J.; Brown, A.; Huang, X.; Jin, Z.; Bowman, J. M. *Ab initio* diffusion monte carlo calculations of the quantum behavior of CH_5^+ in full dimensionality. *J. Phys. Chem. A* **2004**, *108*, 4991−4994.

(20) McCoy, A. B.; Huang, X.; Carter, S.; Landeweer, M. Y.; Bowman, J. M. Full-dimensional vibrational calculations for $H_5O_2^+$ using an *ab initio* potential energy surface. *J. Chem. Phys.* **2005**, *122*, 061101.

(21) McCoy, A. B. Diffusion monte carlo approaches for investigating the structure and vibrational spectra of fluxional systems. *Int. Rev. Phys. Chem.* **2006**, *25*, 77−107.

(22) Wang, Y. M.; Babin, V.; Bowman, J. M.; Paesani, F. The water hexamer: cage, prism, or both. Full dimensional quantum simulations say both. *J. Am. Chem. Soc.* **2012**, *134*, 11116.

(23) Bowman, J. M.; Czakó, G.; Fu, B. High-dimensional ab initio potential energy surfaces for reaction dynamics calculations. *Phys. Chem. Chem. Phys.* **2011**, *13*, 8094−8111.

(24) Czakó, G.; Bowman, J. M. Reaction dynamics of methane with F, O, Cl, and Br on *ab initio* potential energy surfaces. *J. Phys. Chem. A* **2014**, *118*, 2839−2864.

(25) 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.

(26) 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*, 051102.

(27) 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.

(28) Dral, P. O.; Owens, A.; Dral, A.; Csányi, G. Hierarchical machine learning of potential energy surfaces. *J. Chem. Phys.* **2020**, *152*, 204110.
(29) Nguyen, K. A.; Rossi, I.; Truhlar, D. G. A dual-level Shepard interpolation method for generating potential energy surfaces for dynamics calculations. *J. Chem. Phys.* **1995**, *103*, 5522−5530.
(30) Fu, B.; Xu, X.; Zhang, D. H. A hierarchical construction scheme for accurate potential energy surface generation: An application to the F + H_2 reaction. *J. Chem. Phys.* **2008**, *129*, 011103.
(31) Yu, Q.; Bowman, J. M. *Ab initio* potential for $H_3O^+ \rightarrow H^+ + H_2O$: A step to a many-body representation of the hydrated proton? *J. Chem. Theory Comput.* **2016**, *12*, 5284−5292.
(32) Cisneros, G. A.; Wikfeldt, K. T.; Ojamäe, L.; Lu, J.; Xu, Y.; Torabifard, H.; Bartók, A. P.; Csányi, G.; Molinero, V.; Paesani, F. Modeling molecular interactions in water: From pairwise to many-body potential energy functions. *Chem. Rev.* **2016**, *116*, 7501−7528.
(33) Qu, C.; Yu, Q.; Conte, R.; Houston, P. L.; Nandi, A.; Bowman, J. M. A Δ-machine learning approach for force fields, illustrated by a CCSD(T) 4-body correction to the MB-pol water potential. *Digital Discovery* **2022**, *1*, 658−664.
(34) 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.
(35) Qu, C.; Bowman, J. M. Communication: A fragmented, permutationally invariant polynomial approach for potential energy surfaces of large molecules: Application to *N*-methyl acetamide. *J. Chem. Phys.* **2019**, *150*, 141101.
(36) Nandi, A.; Qu, C.; Bowman, J. M. Full and fragmented permutationally invariant polynomial potential energy surfaces for *trans* and *cis* *N*-methyl Acetamide and isomerization saddle points. *J. Chem. Phys.* **2019**, *151*, 084306.
(37) Suenram, R.; Golubiatnikov, G.; Leonov, I.; Hougen, J.; Ortigoso, J.; Kleiner, I.; Fraser, G. Reinvestigation of the microwave spectrum of acetamide. *J. Mol. Spectrosc.* **2001**, *208*, 188−193.
(38) Nandi, A.; Conte, R.; Qu, C.; Houston, P. L.; Yu, Q.; Bowman, J. M. Quantum calculations on a new CCSD(T) machine-learned potential energy surface reveal the leaky nature of gas-phase *trans* and *gauche* ethanol conformers. *J. Chem. Theory Comput.* **2022**, *18*, 5527−5538.
(39) Durig, J.; Larsen, R. Torsional vibrations and barriers to internal rotation for ethanol and 2,2,2-trifluoroethanol. *J. Mol. Struct.* **1990**, *238*, 195−222.
(40) Wang, Y.; Braams, B. J.; Bowman, J. M.; Carter, S.; Tew, D. P. Full-dimensional quantum calculations of ground-state tunneling splitting of malonaldehyde using an accurate ab initio potential energy surface. *J. Chem. Phys.* **2008**, *128*, 224314.
(41) Qu, C.; Conte, R.; Houston, P. L.; Bowman, J. M. Full-dimensional potential energy surface for acetylacetone and tunneling splittings. *Phys. Chem. Chem. Phys.* **2021**, *23*, 7758−7767.
(42) Meuwly, M. Atomistic simulations for reactions and vibrational spectroscopy in the era of machine learning—*quo vadis? J. Phys. Chem. B* **2022**, *126*, 2155−2167.
(43) Werner, H.-J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schütz, M. *MOLPRO, a package of ab initio programs*, ver. 2015.1, 2015; http://www.molpro.net.
(44) 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.
(45) Redington, R. L. H atom and heavy atom tunneling process in tropolone. *J. Chem. Phys.* **2000**, *113*, 2319−2335.
(46) Redington, R. L.; Redington, T. E.; Montgomery, J. M. IR spectra of tropolone(OH) and tropolone(OD). *J. Chem. Phys.* **2000**, *113*, 2304−2318.
(47) Redington, R. L.; Sams, R. L. N_2 pressure broadened Q branch spikes and vibration−contortion−rotation effects in the high resolution FTIR spectrum of tropolone. *Chem. Phys.* **2002**, *283*, 135−151.
(48) Redington, R. L.; Sams, R. L. State-specific spectral doublets in the FTIR spectrum of gaseous tropolone. *J. Phys. Chem. A* **2002**, *106*, 7494−7511.
(49) Redington, R. L.; Redington, T. E.; Blake, T. A.; Sams, R. L.; Johnson, T. J. O18 Effects on the infrared spectrum and skeletal tunneling of tropolone. *J. Chem. Phys.* **2005**, *122*, 224311.
(50) Redington, R. L.; Redington, T. E.; Sams, R. L. Quantum tunneling in the midrange vibrational fundamentals of tropolone. *J. Phys. Chem. A* **2006**, *110*, 9633−9641.
(51) Redington, R. L.; Redington, T. E.; Sams, R. L. Infrared absorption spectra in the hydroxyl stretching regions of gaseous tropolone OHO Isotopomers. *Z. Phys. Chem.* **2008**, *222*, 1197−1211.
(52) Redington, R. L.; Redington, T. E.; Sams, R. L. Tunneling splittings for "O···O stretching" and other vibrations of tropolone isotopomers observed in the infrared spectrum below 800 cm^{-1}. *J. Phys. Chem. A* **2008**, *112*, 1480−1492.
(53) Frost, R. K.; Hagemeister, F. C.; Arrington, C. A.; Zwier, T. S.; Jordan, K. D. Fluorescence-dip infrared spectroscopy of tropolone and tropolone-OD. *J. Chem. Phys.* **1996**, *105*, 2595−1604.
(54) Murdock, D.; Burns, L. A.; Vaccaro, P. H. Vibrational specificity of proton-transfer dynamics in ground-state tropolone. *Phys. Chem. Chem. Phys.* **2010**, *12*, 8285−8299.
(55) Vener, M. V.; Scheiner, S.; Sokolov, N. D. Theoretical study of hydrogen bonding and proton transfer in the ground and lowest excited singlet states of tropolone. *J. Chem. Phys.* **1994**, *101*, 9755−9765.
(56) Guo, Y.; Sewell, T. D.; Thompson, D. L. Semiclassical calculations of tunneling splitting in tropolone. *J. Phys. Chem. A* **1998**, *102*, 5040−5048.
(57) Giese, K.; Kühn, O. The all-cartesian reaction plane hamiltonian: Formulation and application to the H-atom Transfer in tropolone. *J. Chem. Phys.* **2005**, *123*, 054315.
(58) Giese, K.; Petkovic, M.; Naundorf, H.; Kühn, O. Multidimensional quantum dynamics and infrared spectroscopy of hydrogen bonds. *Phys. Rep.* **2006**, *430*, 211−276.
(59) Wang, Y.; Bowman, J. M. One-dimensional tunneling calculations in the imaginary-frequency, rectilinear saddle-point normal mode. *J. Chem. Phys.* **2008**, *129*, 121103.
(60) Wang, Y.; Bowman, J. M. Mode-specific tunneling using the Q_{im} path: Theory and an application to full-dimensional malonaldehyde. *J. Chem. Phys.* **2013**, *139*, 154303.
(61) Collins, M. A.; Bettens, R. P. A. Energy-based molecular fragmentation methods. *Chem. Rev.* **2015**, *115*, 5607−5642.
(62) Gadre, S. R.; Shirsat, R. N.; Limaye, A. C. Molecular tailoring approach for simulation of electrostatic properties. *J. Phys. Chem.* **1994**, *98*, 9165−9169.
(63) Ganesh, V.; Dongare, R. K.; Balanarayan, P.; Gadre, S. R. Molecular tailoring approach for geometry optimization of large molecules: Energy evaluation and parallelization strategies. *J. Chem. Phys.* **2006**, *125*, 104109.
(64) Rahalkar, A. P.; Katouda, M.; Gadre, S. R.; Nagase, S. Molecular tailoring approach in conjunction with MP2 and Ri-MP2 codes: A comparison with fragment molecular orbital method. *J. Comput. Chem.* **2010**, *31*, 2405−2418.
(65) Furtado, J. P.; Rahalkar, A. P.; Shanker, S.; Bandyopadhyay, P.; Gadre, S. R. Facilitating minima search for large water clusters at the MP2 level via molecular tailoring. *J. Phys. Chem. Lett.* **2012**, *3*, 2253−2258.
(66) Sahu, N.; Gadre, S. R. Molecular tailoring approach: A route for *ab initio* rreatment of large clusters. *Acc. Chem. Res.* **2014**, *47*, 2739−2747.
(67) Khire, S. S.; Gurav, N. D.; Nandi, A.; Gadre, S. R. Enabling rapid and accurate construction of CCSD(T)-level potential energy surface of large molecules using molecular tailoring approach. *J. Phys. Chem. A* **2022**, *126*, 1458−1464.
(68) McDaniel, J. G.; Schmidt, J. Next-generation force fields from symmetry-adapted perturbation theory. *Annu. Rev. Phys. Chem.* **2016**, *67*, 467−488.

https://doi.org/10.1021/acs.jctc.2c01034
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

(69) Jing, Z.; Liu, C.; Cheng, S. Y.; Qi, R.; Walker, B. D.; Piquemal, J.-P.; Ren, P. Polarizable force fields for biomolecular simulations: Recent advances and applications. *Annu. Rev. Biophys.* **2019**, *48*, 371−394.
(70) Inakollu, V. S.; Geerke, D. P.; Rowley, C. N.; Yu, H. Polarisable force fields: what do they add in biomolecular simulations? *Curr. Opin. Struct. Biol.* **2020**, *61*, 182−190.
(71) 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.
(72) 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.
(73) 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.
(74) 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.
(75) Conte, R.; Qu, C.; Bowman, J. M. Permutationally invariant fitting of many-body, non-covalent interactions with application to three-body methane−water−water. *J. Chem. Theory Comput.* **2015**, *11*, 1631.
(76) Liu, K.; Brown, M.; Carter, C.; Saykally, R.; Gregory, J.; Clary, D. Characterization of a cage form of the water hexamer. *Nature* **1996**, *381*, 501−503.
(77) Lambros, E.; Paesani, F. How good are polarizable and flexible models for water: Insights from a many-body perspective. *J. Chem. Phys.* **2020**, *153*, 060901.
(78) 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.
(79) 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.
(80) 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.
(81) Zhu, J.; Vuong, V. Q.; Sumpter, B. G.; Irle, S. Artificial neural network correction for density-functional tight-binding molecular dynamics simulations. *MRS Commun.* **2019**, *9*, 867−873.
(82) Qu, C.; Conte, R.; Houston, P. L.; Bowman, J. M. Full-dimensional Potential Energy Surface for Acetylacetone and Tunneling Splittings. *Phys. Chem. Chem. Phys.* **2021**, *23*, 7758−7767.
(83) Shank, A.; Wang, Y.; Kaledin, A.; Braams, B. J.; Bowman, J. M. Accurate *ab initio* and "hybrid" potential energy surfaces, intramolecular vibrational energies, and classical ir spectrum of the water dimer. *J. Chem. Phys.* **2009**, *130*, 144314.
(84) Mallory, J. D.; Mandelshtam, V. A. Diffusion monte carlo studies of MB-pol $(H_2O)_{2-6}$ and $(D_2O)_{2-6}$ clusters: structures and binding energies. *J. Chem. Phys.* **2016**, *145*, 064308.
(85) Wang, X.-G.; Carrington, T. Using monomer vibrational wavefunctions to compute numerically exact (12D) rovibrational levels of water dimer. *J. Chem. Phys.* **2018**, *148*, 074108.
(86) Leforestier, C.; Szalewicz, K.; van der Avoird, A. Spectra of water dimer from a new *ab initio* potential with flexible Monomers. *J. Chem. Phys.* **2012**, *137*, 014305.
(87) Wang, Y.; Bowman, J. M. Communication: Rigorous calculation of dissociation energies (D_0) of the water trimer $(H_2O)_3$ and $(D_2O)_3$. *J. Chem. Phys.* **2011**, *135*, 131101.
(88) Rocher-Casterline, B. E.; Ch'ng, L. C.; Mollner, A. K.; Reisler, H. Communication: Determination of the bond dissociation energy (D_0) of the water dimer, $(H_2O)_2$, by velocity map imaging. *J. Chem. Phys.* **2011**, *134*, 211101.
(89) Ch'ng, L. C.; Samanta, A. K.; Wang, Y.; Bowman, J. M.; Reisler, H. Experimental and theoretical investigations of the dissociation energy (D_0) and dynamics of the water trimer, $(H_2O)_3$. *J. Phys. Chem. A* **2013**, *117*, 7207−7216.
(90) 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.
(91) Nauta, K.; Miller, R. Formation of cyclic water hexamer in liquid helium: The smallest piece of ice. *Science* **2000**, *287*, 293−295.
(92) Burnham, C. J.; Xantheas, S. S.; Miller, M. A.; Applegate, B. E.; Miller, R. E. The formation of cyclic water complexes by sequential ring insertion: Experiment and theory. *J. Chem. Phys.* **2002**, *117*, 1109−1122.
(93) Steinbach, C.; Andersson, P.; Melzer, M.; Kazimirski, J. K.; Buck, U.; Buch, V. Detection of the book isomer from the OH-stretch spectroscopy of size selected water hexamers. *Phys. Chem. Chem. Phys.* **2004**, *6*, 3320−3324.
(94) Diken, E. G.; Robertson, W. H.; Johnson, M. A. The Vibrational spectrum of the neutral $(H_2O)_6$ precursor to the "magic" $(H_2O)_6^-$ cluster anion by argon-mediated, population-modulated electron attachment spectroscopy. *J. Phys. Chem. A* **2004**, *108*, 64−68.
(95) Wang, Y. M.; Bowman, J. M. IR spectra of the water hexamer: Theory, with inclusion of the monomer bend overtone, and experiment are in agreement. *J. Phys. Chem. Lett.* **2013**, *4*, 1104.
(96) Liu, H. C.; Wang, Y. M.; Bowman, J. M. Local-monomer calculations of the intramolecular IR spectra of the cage and prism isomers of $HOD(D_2O)_5$ and HOD and D_2O ice Ih. *J. Phys. Chem. B* **2014**, *118*, 14124.
(97) Douberly, G. Private communication, 2013.
(98) Adjoua, O.; Lagardère, L.; Jolly, L.-H.; Durocher, A.; Very, T.; Dupays, I.; Wang, Z.; Inizan, T. J.; Célerse, F.; Ren, P.; Ponder, J. W.; Piquemal, J.-P. Tinker-HP: Accelerating molecular dynamics simulations of large complex systems with advanced point dipole polarizable force fields using GPUs and multi-GPU systems. *J. Chem. Theory Comput.* **2021**, *17*, 2034−2053.
(99) 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.
(100) Gazdy, B.; Bowman, J. M. An adjusted global potential surface for HCN based on rigorous vibrational calculations. *J. Chem. Phys.* **1991**, *95*, 6309−6316.
(101) Meuwly, M.; Hutson, J. M. Morphing ab initio potentials: A systematic study of Ne−HF. *J. Chem. Phys.* **1999**, *110*, 8338−8347.
(102) Wang, Y. M.; Bowman, J. M. *Ab initio* potential and dipole moment surfaces for water. II. Local-monomer calculations of the infrared spectra of water clusters. *J. Chem. Phys.* **2011**, *134*, 154510.
(103) 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*, 074506.
(104) 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*, 1599−1607.
(105) 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.
(106) Townsend, J.; Vogiatzis, K. D. Data-driven acceleration of the coupled-cluster singles and doubles iterative solver. *J. Phys. Chem. Lett.* **2019**, *10*, 4129−4135.
(107) Dick, S.; Fernandez-Serra, M. Machine learning accurate exchange and correlation functionals of the electronic density. *Nat. Commun.* **2020**, *11*, 3509.

https://doi.org/10.1021/acs.jctc.2c01034
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

(108) Qiao, Z.; Welborn, M.; Anandkumar, A.; Manby, F. R.; Miller, T. F. OrbNet: Deep learning for quantum chemistry using symmetry-adapted atomic-orbital features. *J. Chem. Phys.* **2020**, *153*, 124111.
(109) Ikabata, Y.; Fujisawa, R.; Seino, J.; Yoshikawa, T.; Nakai, H. Machine-learned electron correlation model based on frozen core approximation. *J. Chem. Phys.* **2020**, *153*, 184108.
(110) Ruth, M.; Gerbig, D.; Schreiner, P. R. Machine learning of coupled cluster (T)-energy corrections via delta (Δ)-Learning. *J. Chem. Theory Comput.* **2022**, *18*, 4846−4855.
(111) Dral, P. O.; Zubatiuk, T.; Xue, B.-X. Learning from multiple quantum chemical methods: Δ-learning, transfer learning, co-kriging, and beyond. In *Quantum Chemistry in the Age of Machine Learning*; Dral, P. O., Ed.; Elsevier, 2023; pp 491−507.

https://doi.org/10.1021/acs.jctc.2c01034
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX