JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Article

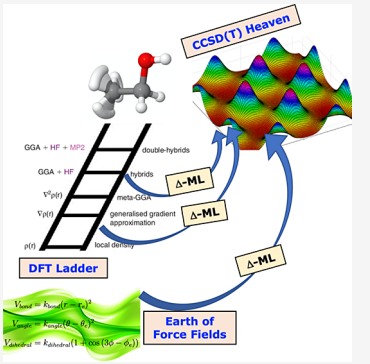
Δ-Machine Learning to Elevate DFT-Based Potentials and a Force Field to the CCSD(*T*) Level Illustrated for Ethanol

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

Cite This: *J. Chem. Theory Comput.* 2024, 20, 8807−8819 Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Progress in machine learning has facilitated the development of potentials that offer both the accuracy of first-principles techniques and vast increases in the speed of evaluation. Recently, Δ-machine learning has been used to elevate the quality of a potential energy surface (PES) based on low-level, e.g., density functional theory (DFT) energies and gradients to close to the gold-standard coupled cluster level of accuracy. We have demonstrated the success of this approach for molecules, ranging in size from H_3O^+ to 15-atom acetyl-acetone and tropolone. These were all done using the B3LYP functional. Here, we investigate the generality of this approach for the PBE, M06, M06-2X, and PBE0 + MBD functionals, using ethanol as the example molecule. Linear regression with permutationally invariant polynomials is used to fit both low-level and correction PESs. These PESs are employed for standard RMSE analysis for training and test data sets, and then general fidelity tests such as energetics of stationary points, normal-mode frequencies, and torsional potentials are examined. We achieve similar improvements in all cases. Interestingly, we obtained significant improvement over DFT gradients where coupled cluster gradients were not used to correct the low-level PES. Finally, we present some results for correcting a recent molecular mechanics force field for ethanol and comment on the possible generality of this approach.



INTRODUCTION

Developing high-dimensional, ab initio-based potential energy surfaces (PESs) is an active area of theoretical and computational research. Major progress has been made in using and developing machine learning (ML) approaches for PESs with more than four atoms, based on fitting thousands of CCSD(*T*) energies[1−5] or forces.[6,7] Some of these ML approaches have used permutationally invariant polynomials (PIPs) or PIPs as inputs to neural network software.[1−5] Of course, there are numerous other ML methods. It is perhaps of interest and relevance to this paper that the precision of a PIP PES for ethanol was shown to be as good as the best performing ML methods and to be substantially faster (factors of 10 or more)[8] than all the ML methods considered, i.e., GAP-SOAP,[9] ANI,[10] DPMD,[11] sGDML,[6,7] PhysNet,[12] KREG,[13] and pKREG.[14] The data set for these comparisons was from the rMD17 database,[15] which uses 500 K direct dynamics based on the PBE0 functional to obtain energies and forces. The metrics used in the "learning curves" were root-mean-square errors in energies and forces. This followed the standard protocol used earlier to assess many ML methods for potentials.[16]

CCSD(*T*) data sets for larger molecules are rare, owing to the steep scaling of CCSD(*T*) calculations of order ∼ N^7, *N* being the number of basis functions. The potential energy surface for the 10-atom formic acid dimer is one example where PESs have been reported at the CCSD(*T*) level, using PIPs[17] and later an atom-centered high-dimensional NN.[18] Complex reactive potentials for 6 and 7-atom chemical reactions, which are fitted to tens of thousands or even hundred thousand CCSD(*T*) energies, have been reported.[19,20] The PIP-based automated ROBOSURFER software[5] has been applied to develop a number of complex PESs for 9-atom chemical reactions.[21,22]

Correcting ab initio-based potential energy surfaces has been a long-standing goal of computational chemistry. Of relevance here are approaches that aim to bring a PES, based on a low-level of electronic theory, typically DFT or MP2 theory, to a higher level such as coupled cluster (CC) theory. In consideration of larger molecules and clusters, where high-level methods are prohibitively expensive, the motivation for doing this is clear. "Δ-machine learning" (Δ-ML) is the method of direct relevance to the present paper. This approach seeks to add a correction to a property obtained using an efficient and thus perforce low-level ab initio theory.[6,7,23−25] A hierarchical Δ-ML method using multiple quantum chemistry

Received: July 27, 2024
Revised: September 17, 2024
Accepted: September 18, 2024
Published: October 3, 2024

https://doi.org/10.1021/acs.jctc.4c00977
J. Chem. Theory Comput. 2024, 20, 8807−8819

methods has been applied to the five-atom CH_3Cl PES.[26] In this sense, the approach is related, in spirit at least, to the correction potential approach mentioned above, when the property is the PES. Other, related methods that utilize a high and lower level of electronic structure theory are mentioned in the Discussion section.

We recently applied a Δ-ML approach, originally given for the three-atom F + H_2,[27] to larger systems.[28−33] Additionally, the approach has also been proposed to correct many-body force fields.[34] In all these examples, the B3LYP functional was used to obtain the low-level PES.

Considering the success of the Δ-ML method with the B3LYP functional,[35,36] it is both interesting and significant to explore whether this straightforward approach can be extended to other functionals and to molecular mechanics, including "classical" force fields (FFs). There is a vast literature on molecular mechanics force fields, and the reader is directed to a recent and relevant (vide infra) paper that surveys this field.[37] While these FFs, which are heavily semi or totally empirical, have made an enormous impact in biomolecular simulations, there is strong motivation to progress from these. Broadly put, there are two approaches that can be undertaken. One is to replace these FFs with strictly ML FFs, based on electronic structure energies and forces for the covalent and noncovalent interactions, and sophisticated treatments of long-range interactions. This is a major challenge for an ML method that aims to deal with hundreds of atoms in a single step. A recent example of this approach by Tkatchenko, Müller and co-workers can be found in ref 38. Of course, invoking the "no free lunch" axiom, this approach is far more demanding in computational effort compared to a classical FF. A second approach is to correct a classical force field. There have been several papers along these lines including one from this group aimed at correcting a sophisticated classical FF for water, by correcting the short-range 2-b, 3-b, and even 4-b interactions.[34] However, while water is essential for life it is not a biomolecule. Other similar approaches, focused on correcting the short-range interactions have also appeared recently.[39] While these approaches may be less computationally demanding than a full ML approach, they are still far more demanding than biomolecular FFs.

A variation of the second approach, which is our focus, is to continue to use the empirical FF expression for the potential, i.e., harmonic bond stretches and bends, periodic torsional potentials, plus simple 2-b noncovalent interactions and long-range electrostatics, and to add a computationally efficient ML correction. To facilitate the goal of efficiency, the ML correction can be applied to some terms, at least, in the classical FF are corrected using ab initio electronic energies. Recently, Meuwly and co-workers,[37,40] investigated correcting the CHARMM classical force field for specific examples. An earlier, but still recent, example of this approach used atomic force matching (AFM), using MP2 theory, to determine classical FF intramolecular terms of ethanol plus the 2-b intermolecular interaction between an ethanol and water molecules.[41] Here, we use this AFM-corrected FF for ethanol to investigate our computationally efficient Δ-ML approach, which substantially improves several key properties of AFM-corrected FF. Most notably, it addresses the harmonic normal-mode frequencies, which are greatly overestimated by this FF for all but the lowest several normal modes.

The paper is organized as follows. A brief review of the Δ-ML approach is provided, along with the essentials of the highly efficient ML linear-regression approach we use with permutationally invariant polynomials. Results and discussion follow, including remarks on the extension of the Δ-ML PIP approach to much larger molecules.

THEORY

Δ-ML Approach. The Δ-ML approach is given by the equation[28]

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

where $V_{LL\rightarrow CC}$ is the corrected PES, V_{LL} is a PES fit to low-level DFT electronic data, and ΔV_{CC-LL} is the correction PES based on high-level coupled cluster energies. As shown in ref 28, we use PIPs to represent the PESs on the right-hand side. This might suggest that the cost to evaluate their sum is twice the cost to evaluate V_{LL}. However, because the correction potential is "small" and more slowly varying over configuration space, relative to V_{LL}, the PIPs expansion for that PES is much smaller than the one for the V_{LL}. Thus, the cost of evaluating that term is much less than the cost to evaluate V_{LL}. If the above were not true, then the Δ-ML would not be a viable approach. We will return to this with specific numbers below for the present application to ethanol. Also, we note that the above equation was given 16 years ago,[27] and so we rediscovered it.

To investigate the efficacy of the Δ-ML approach, four widely used functionals are employed here, M06[42] and M06-2X[42,43] functionals with the 6-311+G(d,p) basis, PBE[44] with the def2-SVP basis, and PBE0[45,46] including many-body dispersion (MBD)[47] with the "intermediate" basis setting.[48] Additionally, we also replace V_{LL} with a classical force field.

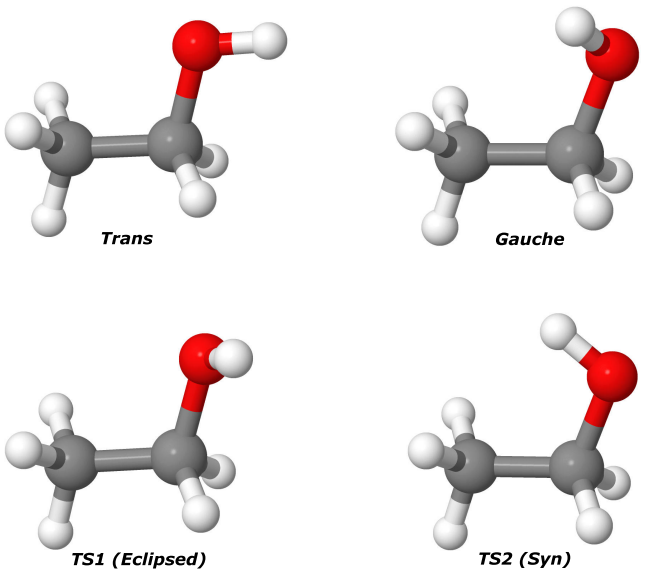


Figure 1. Optimized geometry of *trans* and *gauche* conformers of ethanol and their two isomerization TSs at CCSD(*T*)-F12a/aug-cc-pVDZ level. Reproduced from ref 49. Copyright 2022 American Chemical Society.

Table 1. RMS Fitting Error (in cm^{-1}) of V_{LL} for Training and Test Data Sets

	PBE	M06	M06-2X	B3LYP	PBE0 + MBD
training	45	79	47	40	40
test	56	82	57	51	51

https://doi.org/10.1021/acs.jctc.4c00977
J. Chem. Theory Comput. 2024, 20, 8807−8819

Table 2. RMS Fitting Error (in cm^{-1}) of Correction PESs ΔV_{CC-LL} for Training and Test Data Sets

	PBE	M06	M06-2X	B3LYP	PBE0 + MBD
training	67	53	32	28	26
test	90	61	40	30	30

Previously, it was noted that the difference between CCSD(T) and DFT energies, ΔV_{CC-LL}, does not vary as strongly as V_{LL} with respect to the nuclear configurations and therefore, a small number of high-level electronic energies is adequate to fit the correction PES.

It is not clear if this observation applies, at least semiquantitatively, for classical force fields. In this case, the differences can be much larger, as expected, and indeed verified here for ethanol. We investigate this using a previous data set of 2319 CCSD(T)-F12a/aug-cc-pVDZ electronic energies.[49]

As noted above, the permutationally invariant polynomial (PIP) approach is used to fit both the V_{LL} and ΔV_{CC-LL} PESs. The theory of permutationally invariant polynomial is well established and has been presented in several review articles.[1,2,50−52] In terms of a PIP basis, the potential energy, V, can be written in compact form as

$$V(\boldsymbol{x}) = \sum_{i=1}^{n_p} c_i p_i(\boldsymbol{y}) \tag{2}$$

where c_i are linear coefficients, p_i are PIPs, n_p is the total number of polynomials for a given maximum polynomial order, and $\boldsymbol{y}$ are the collection of Morse variables. For example, $y_{\alpha\beta}$ is given by $\exp(-r_{\alpha\beta}/\lambda)$, where $r_{\alpha\beta}$ is the internuclear distance between atoms α and β. The range (hyper)parameter, λ, equals 2 Bohr; this is the typical value that has been used in many of our PIPs PESs.[2,50] The linear coefficients are obtained using standard least-squares methods for large data sets of electronic energies molecules and gradients.

The Ethanol Force Field. Figure 1 shows conformations of *trans* and *gauche*-ethanol and two saddle point transition states. These are from electronic structure calculations at the CCSD(T)-F12a/aug-cc-pVDZ level.

The molecular mechanics force field we consider is the recent one that was corrected using force matching MP2 gradients computed with triple-ζ-quality basis sets using the Adaptive Force Matching method.[41] The mathematical expression for the total energy of the force field includes intramolecular interaction terms for interactions of atoms that are linked by molecular bonds.

$$V_{FF} = V_{bond} + V_{angle} + V_{dihedral} \tag{3}$$

where V_{bond} and V_{angle} are modeled by the quadratic energy functions, corresponding to the oscillations about an equilibrium bond length and bond angle, based on the Harmonic approximation, and $V_{dihedral}$ is modeled by the cosine function, corresponding to the torsional rotation of four atoms about a central bond.

$$V_{bond} = k_{bond}(r - r_e)^2 \tag{4}$$

$$V_{angle} = k_{angle}(\theta - \theta_e)^2 \tag{5}$$

$$V_{dihedral} = k_{dihedral}(1 + \cos(3\phi - \phi_e)) \tag{6}$$

The value of fitting parameters, k_{bond}, k_{angle}, $k_{dihedral}$, as well as the equilibrium bond lengths and angles are taken from ref 41.

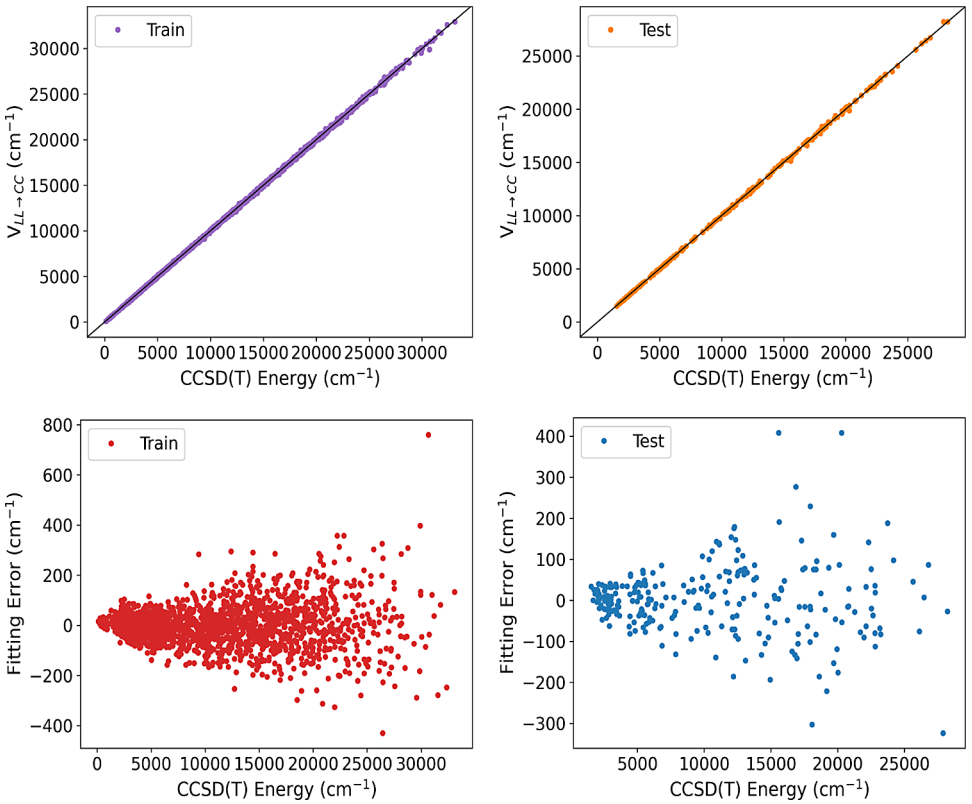


Figure 2. Two upper panels show energies of ethanol from $V_{LL\to CC}$ vs direct CCSD(T) ones for the indicated data sets calculated using the PBE functional. Corresponding fitting errors relative to the minimum energy are given in the lower panels.

https://doi.org/10.1021/acs.jctc.4c00977
J. Chem. Theory Comput. 2024, 20, 8807−8819

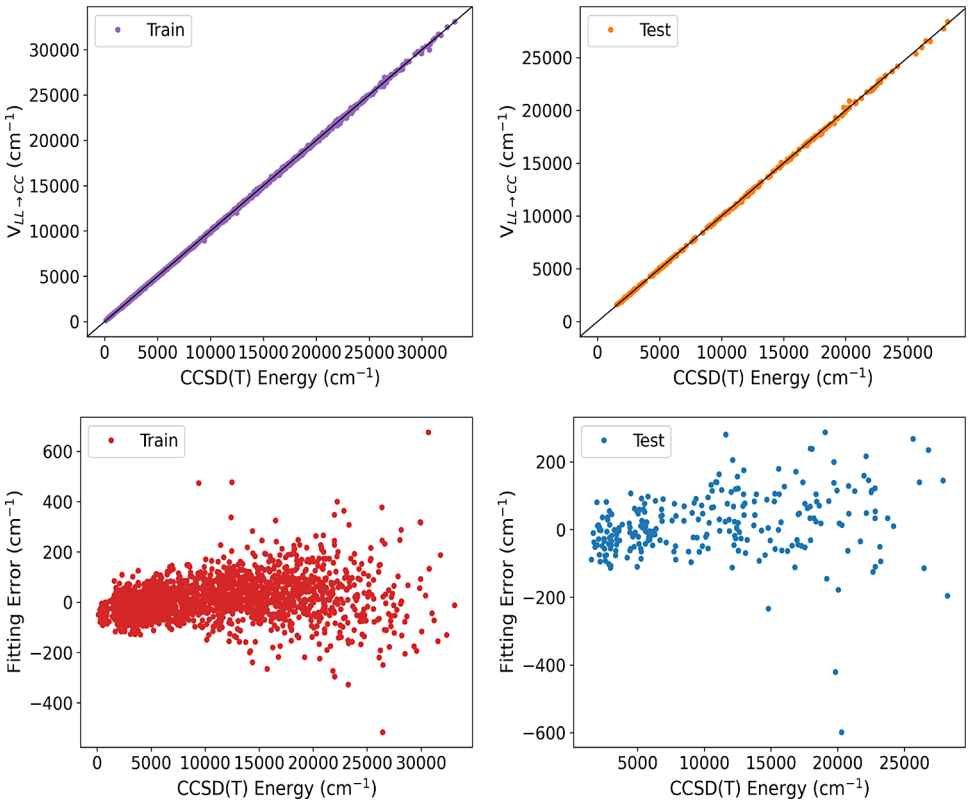


Figure 3. Two upper panels show energies of ethanol from $V_{LL\rightarrow CC}$ vs direct CCSD(T) ones for the indicated data sets calculated using the M06 functional. Corresponding fitting errors relative to the minimum energy are given in the lower panels.

COMPUTATIONAL DETAILS

We use the data set from our recently reported "MDQM21" data set,[8] which includes a total of 11,000 energies and their corresponding gradients generated from ab initio molecular dynamics (AIMD) simulations at B3LYP/6-311+G(d,p) level of theory. This data set was partitioned into a training set of 8,500 energies and gradients and a test set of 2500 energies and gradients. The same training and test data set were used for single point energy and gradients computations at M06/6-311+G(d,p),[42] M06-2X/6-311+G(d,p),[42,43] and PBE/def2-SVP[44] level of theory using MOLPRO[53] quantum chemistry package and at PBE0+MBD[45−47] level of theory with "intermediate" basis setting using the FHI-aims electronic structure package.[48,54]

RESULTS AND DISCUSSION

Δ-ML for DFT Functionals. The low-level PES, V_{LL}, is fitted using PIPs with a maximum polynomial order of 4 (14,752 terms) with permutational symmetry 321111. This notation indicates the three equivalent H atoms of the CH_3-group and the 2 equiv H atoms of the CH_2 group. The range of the energies for training and testing is 0 to roughly 30,000 cm^{-1}. (Note, this energy range is much above the range in the rMD17 data set for ethanol of roughly 8500 cm^{-1}.) The root-mean-square (RMS) fitting errors for training and test data sets are shown in Table 1. We see these fits are not overfit and that the precision is high, given the range of the data set.

Next, we train ΔV_{CC-LL} on the difference between the CCSD(T) and DFT absolute energies at 2069 geometries and test the obtained surface on the remaining 250 geometries. To fit the $\Delta V_{LL\rightarrow CC}$, we have used a maximum polynomial order of 2 with permutational symmetry 321111 for the training data set. This results in a basis size of 208 PIPs generated using our MSA software.[55,56] The RMS training and test errors for the energies of correction PES are shown in Table 2.

Finally, to obtain $V_{LL\rightarrow CC}$ we add the correction ΔV_{CC-LL} to the low-level DFT PES, V_{LL}. The correlation plots of the $V_{LL\rightarrow CC}$ fit for a training set of 2069 points and a test set of 250 points for the PBE and M06 DFT functional are presented in Figures 2 and 3, respectively. The RMS training and test errors for the energies of Δ-corrected PES are shown in Table 3.

Table 3. RMS Fitting Error (in cm^{-1}) of Corrected PESs $V_{LL\rightarrow CC}$ for Training and Test Data Sets

	PBE	M06	M06-2X	B3LYP	PBE0 + MBD
training	78	79	56	53	53
test	87	97	62	52	52

To determine the accuracy of the $V_{LL\rightarrow CC}$ PES for various DFT functionals, we perform the geometry optimization and normal-mode frequency analysis for both *trans* and *gauche* isomers and their two isomerization saddle point geometries (Anti and Syn). The structures of these isomers and the saddle points are shown in Figure 1. The energetics of all four stationary points of ethanol relative to the trans minima, calculated using various DFT functional, are listed in Table 4. The Δ-corrected PES leads to better optimized energetics for all four stationary points across all DFT functionals, as seen in Table 5.

The comparison of harmonic mode frequencies of various Δ-corrected PES calculated using different DFT PESs (V_{LL}) for *trans* ethanol with the corresponding direct CCSD(T) frequencies are shown in Table 6. The overall agreement of

https://doi.org/10.1021/acs.jctc.4c00977
J. Chem. Theory Comput. 2024, 20, 8807−8819

Table 4. Comparison of the Energetics in kcal/mol (cm^{-1}) of All Four Stationary Points of Ethanol Relative to the *trans* Minima from Direct DFT Calculations

isomer	PBE	M06	M06-2X	B3LYP	PBE0 + MBD
trans	0.00 (0)	0.00 (0)	0.00 (0)	0.00 (0)	0.00 (0)
gauche	−0.37 (−129)	0.37 (129)	0.08 (28)	0.05 (18)	0.05 (19)
TS1 (Anti)	1.98 (692)	1.17 (409)	1.16 (407)	1.05 (367)	1.18 (411)
TS2 (Syn)	1.24 (432)	1.69 (591)	1.45 (507)	1.44 (505)	1.13 (395)

Table 5. Comparison of the Energetics in kcal/mol (cm^{-1}) of All Four Stationary Points of Ethanol Relative to the *trans* Minima for Direct CCSD(T) and Δ-ML PESs

		$V_{LL\rightarrow CC}$				
isomer	direct CCSD(T)	LL = PBE	LL = M06	LL = M06-2X	LL = B3LYP	LL = PBE0 + MBD
trans	0.00 (0)	0.00 (0)	0.00 (0)	0.00 (0)	0.00 (0)	0 (0.00)
gauche	0.13 (45)	0.04 (14)	0.21 (73)	0.13 (45)	0.11 (38)	0.14 (51)
TS1 (Anti)	1.09 (381)	1.22 (427)	1.12 (392)	1.08 (378)	1.08 (378)	1.04 (363)
TS2 (Syn)	1.36 (476)	1.41 (493)	1.27 (444)	1.34 (469)	1.35 (472)	1.24 (435)

Table 6. Comparison of Harmonic Frequencies (in cm^{-1}) of DFT PESs and Δ-Corrected PES Computed Using Indicated DFT Functionals and Corresponding Ab Initio Ones (CCSD(T)-F12a/aug-cc-pVDZ) for *Trans*-ethanol

	CCSD(T)	PBE		M06		M06-2X		B3LYP		PBE0 + MBD	
mode	direct	V_{LL}	ΔML	V_{LL}	ΔML	V_{LL}	ΔML	V_{LL}	ΔML	V_{LL}	ΔML
1	222	251	252	245	241	251	245	237	243	241	241
2	274	282	292	289	280	281	274	269	273	275	273
3	413	408	415	419	415	423	418	417	417	420	417
4	813	793	816	791	804	822	818	820	818	817	822
5	907	891	913	908	906	923	911	896	909	916	910
6	1049	1018	1060	1039	1049	1057	1055	1035	1055	1055	1056
7	1115	1103	1120	1115	1106	1134	1115	1094	1115	1131	1116
8	1180	1140	1189	1166	1175	1185	1180	1176	1181	1181	1183
9	1274	1230	1281	1246	1267	1273	1280	1266	1284	1276	1285
10	1300	1245	1293	1281	1293	1310	1302	1299	1302	1303	1303
11	1402	1331	1396	1376	1400	1403	1405	1402	1403	1394	1406
12	1456	1406	1449	1435	1443	1459	1450	1446	1454	1453	1458
13	1484	1408	1485	1451	1480	1489	1488	1483	1488	1475	1488
14	1501	1429	1492	1461	1487	1507	1503	1498	1500	1492	1503
15	1531	1463	1533	1497	1516	1539	1530	1524	1530	1521	1531
16	3001	2878	2994	2960	2985	3030	2994	2978	2995	3000	2993
17	3036	2909	3032	2998	3024	3060	3025	3005	3028	3032	3028
18	3042	2980	3051	3025	3026	3079	3029	3031	3036	3058	3032
19	3122	3073	3141	3115	3109	3149	3110	3098	3120	3142	3115
20	3127	3077	3144	3116	3113	3153	3116	3105	3126	3144	3120
21	3853	3720	3865	3924	3852	3915	3849	3843	3862	3913	3856
MAE		54	9	23	9	16	6	11	4	11	6

these harmonic frequencies with the direct CCSD(T) ones is excellent, as presented in Figure 4. Note that the Δ-corrected PES tends to minimize the gap between the direct-CCSD(T) frequencies and the calculated ones, especially for the high-frequency modes. As depicted in Figure 4, although PBE functional has the highest deviation in frequency from the CCSD(T) values, the correction tends to reduce the deviation within a few cm^{-1}.

Next, we examine the change in the PES gradient after the incorporation of the correction. In order to make a more detailed examination of the errors in gradients, we calculated the cosine of the angle between the direct CCSD(T) and DFT gradient vector as well as corresponding $V_{LL\rightarrow CC}$ PES gradient vector, also the mean absolute difference (magnitude of 27 gradient components for each geometry) between these two gradient vectors for 10 randomly selected geometries. This is shown in Figure 5. As seen, there is a substantial reduction in the errors in the gradient in the Δ-ML corrected PES compared to the DFT PESs. Specially, in case of PBE the gradient differences are much larger as well as the cos θ values. This is especially encouraging as the correction PES, ΔV_{CC-LL}, is trained only on CCSD(T) energies without CCSD(T) gradients. Presumably, including gradient data in the training of ΔV_{CC-LL} would result in a larger reduction in the error. We plan to investigate this in the future.

Next, we compare the PES calculated torsional barrier for the methyl rotor with the direct CCSD(T) level. The methyl rotor torsional potentials (not fully relaxed) for the *trans* minima as a function of the torsional angle are shown in Figure 6. For all the DFT functionals, the results from the Δ-corrected PESs are comparable to the direct ab initio calculations at the CCSD(T) level, as mentioned in Table 7. Note that the methyl

https://doi.org/10.1021/acs.jctc.4c00977
J. Chem. Theory Comput. 2024, 20, 8807−8819

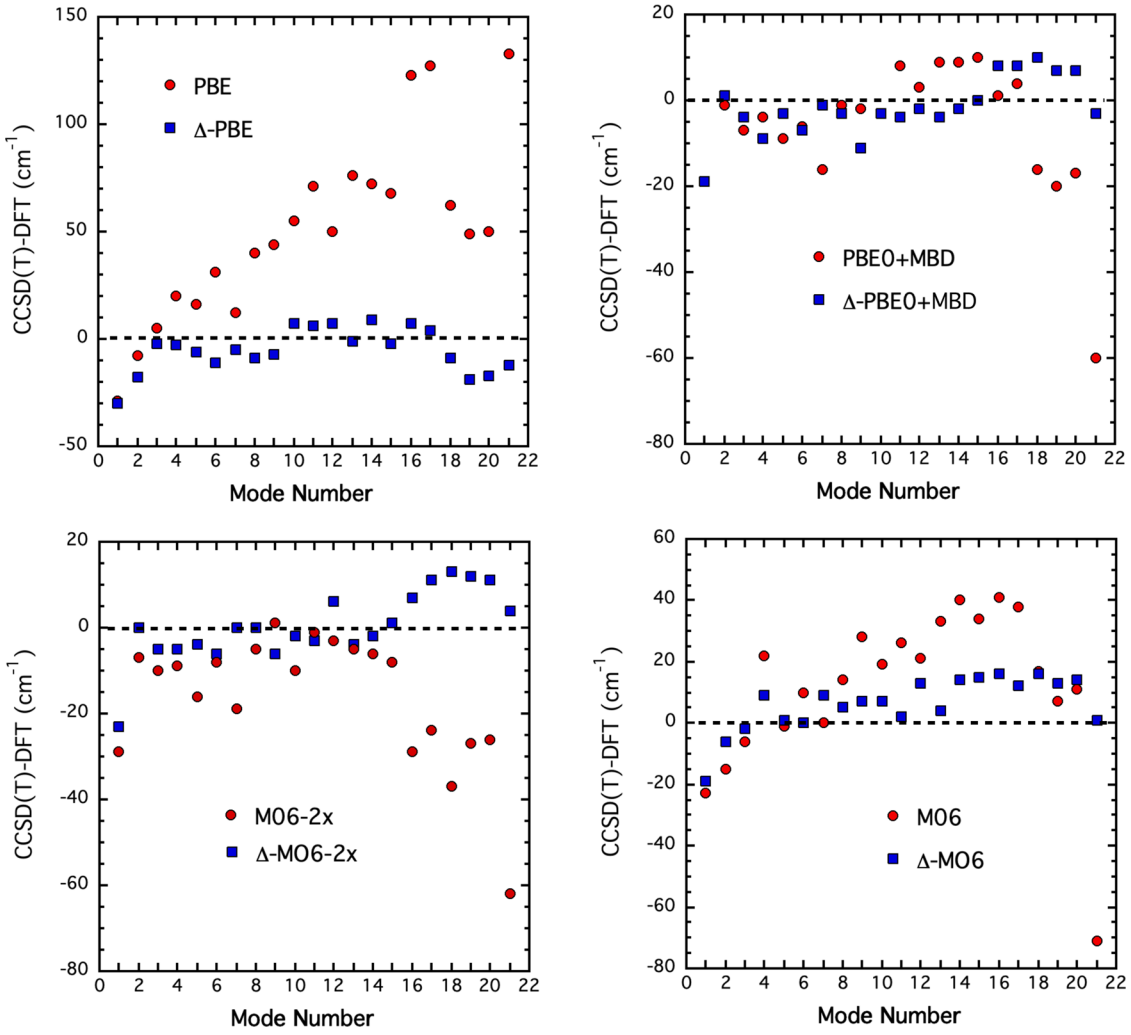


Figure 4. Differences of the CCSD(T) and DFT frequencies (in red) and Δ-corrected frequencies (in blue) for *trans*-ethanol for indicated functionals.

torsional barrier height for the *trans* isomer evaluated from the microwave spectroscopy is 1174 cm^{-1}.[57,58] Similarly, another experimental analysis of the infrared and Raman spectra determined the methyl torsional barriers to be 1185 cm^{-1}.[59]

To conclude this subsection we comment on the additional cost to evaluate ΔV_{CC-LL} relative to the cost to evaluate V_{LL}. As noted above, the PIPs bases for these two PESs contain 208 and 14,752 terms, respectively. So adding the two potentials results in a negligible 1% increase in cost relative to evaluating V_{LL}.

Δ-ML for Force Field. To calculate the Δ-corrected force field potential, we first calculated the force field potential energy using eq 3. Next, we train the Δ-correction PES on the difference between the CCSD(T) and FF energies of 2069 geometries and test the obtained surface on the remaining 250 geometries.[49] To fit the corrected PES, a maximum polynomial order of 2 is used with permutationally symmetry 321111 for the training data set. A plot of $V_{FF\rightarrow CC}$ versus corresponding direct CCSD(T) energies for the training and test sets calculated using the harmonic approximation for the MP2 corrected force field, along with the fitting error, is shown in Figure 7. A huge fitting error for both the training and test sets is found, with RMSE values of 1436 and 2097 cm^{-1}, respectively. The substantial RMSE observed in the Δ-ML corrected force field PES indicates the imprecise fitting.

The harmonic approximation works well for the small oscillation around the equilibrium position but its accuracy decreases for larger amplitude vibrations where anharmonicity becomes significant. Hence, we use a Morse potential as V_{bond} to provide a more realistic representation to the higher bond stretching.

$$V_{bond} = D_e(1 - e^{-\alpha(r-r_e)})^2 \quad (7)$$

Here, the value of α is equal to $\sqrt{k_{bond}/D_e}$ for all bond types, with the dissociation energy D_e provided in Table 8.

A plot of $V_{FF\rightarrow CC}$ versus corresponding direct CCSD(T) energies for the training and test sets, calculated using the Morse potential along with the fitting error, is shown in Figure 8. The fitting error decreases slightly for both the training and test sets compared to Figure 7, with reduced RMSE values of 1089 and 1529 cm^{-1}, respectively. However, these RMSE values are still large enough to produce inaccurate results for the entire data set. Therefore, we attempt to improve the fitting by implementing energy cut-offs across the entire data set. For this purpose, we select two energy cut-offs at 10,000 and 5000 cm^{-1} above the global minimum. For the 10,000 cm^{-1} energy cut off case, the correction PES is trained on the difference

https://doi.org/10.1021/acs.jctc.4c00977
J. Chem. Theory Comput. 2024, 20, 8807−8819

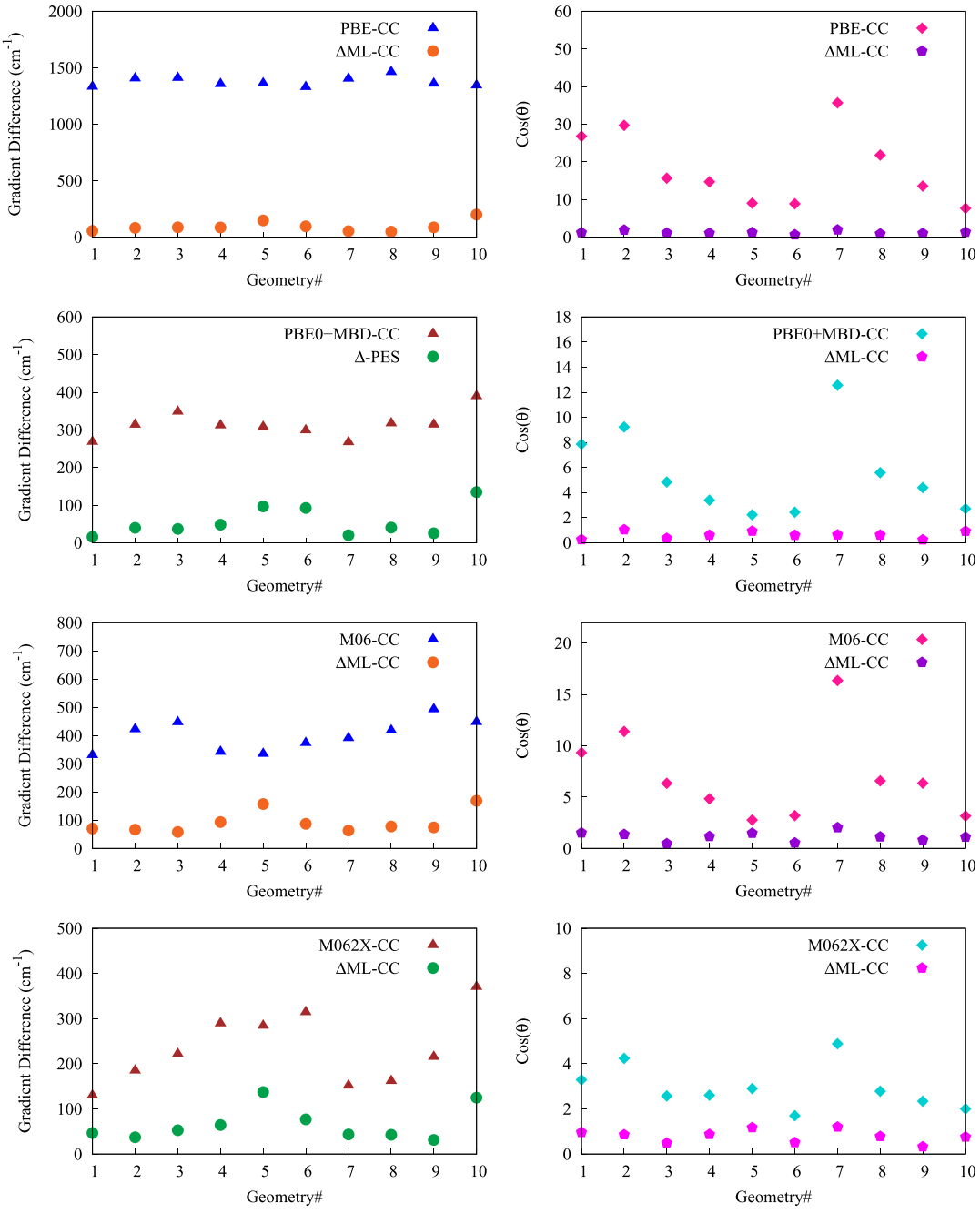


Figure 5. Plot of mean absolute gradient magnitude difference (left panel) and cos θ (right panel), where θ is angle between the direct CCSD(T) and indicated DFT gradients as well as corresponding $V_{LL\rightarrow CC}$ PES gradients for randomly selected 10 ethanol geometries. See the text for details.

between the CCSD(T) and FF absolute energies of 1124 geometries and tested on the remaining 125 geometries. For the 5000 cm^{-1} energy cut off case, the number of training and test geometries are 702 and 65, respectively.

Table 9 presents the RMS errors in Δ-corrected PES computed using the force field for both the training and test data sets. The RMSE values decrease by a factor of 5 for the energy cutoff set at 10,000 cm^{-1} and by a factor of 15 for the data set at 5000 cm^{-1}, compared to the RMSE of the Δ-corrected PES computed using the force field with the harmonic approximation. Since direct CCSD(T) energies of the ethanol isomers and their saddle point transition states are quite small in comparison to the RMSE values, their energy optimization results are random. Note that the time taken to calculate 100,000 data points using the force field is 2.04 and 2.09 s for the harmonic and Morse potentials, respectively. Even after adding the Δ-correction, the time taken to calculate 100,000 data points is 2.15 s. Hence, the force field Δ-ML PES is much faster than the one using the DFT functional; to be precise, it has almost doubled the evaluation speed.

Next, we performed normal-mode analyses for *trans*-ethanol to examine the vibrational frequency predictions of these PESs. The comparisons of harmonic frequencies with the corresponding ab initio frequencies for the *trans*-ethanol are shown in Table 10. As seen in the table, the original FF produces poor results except for the two lowest frequencies. The sets of corrected FFs all show significant improvement at these frequencies. The corrections to the original FF, denoted as the

https://doi.org/10.1021/acs.jctc.4c00977
J. Chem. Theory Comput. 2024, 20, 8807−8819

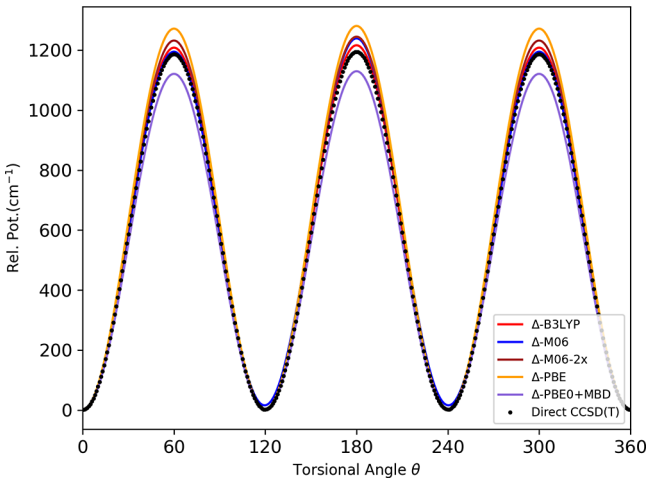


Figure 6. Comparison of torsional potential (not fully relaxed) of the methyl rotor of trans ethanol between direct CCSD(T) and Δ-corrected PES computed using indicated DFT functionals.

Table 7. Barrier Height of the Methyl Rotor Torsional Potential for the Trans Isomer[a]

direct-CCSD(T)	Δ-PBE	Δ-M06	Δ-M06-2X	Δ-B3LYP	Δ-PBE0 + MBD
1194	1272	1195	1232	1208	1121

[a]Energies are in cm^{-1}.

"Harmonic FF", and those where the harmonic stretch modes were replaced by Morse potentials, denoted as the "Morse FF", were evaluated. Overall, the corrected Morse FF results are

Table 8. Intramolecular Potential Parameters of Ethanol Taken from Ref 41

bond type	r_e (Å)	k_{bond} (kcal/mol Å^2)	D_e (kcal/mol)
C−C	1.5204	551.9110	82.69
O−H	0.9609	1056.6764	110.66
C−O	1.4396	577.2346	85.56
C−H	1.0937	742.5561	98.71

superior to those of the corrected harmonic FF. Notably, there are interesting dependencies to the extent of the training data set. Limiting the maximum energy to 5000 cm^{-1} produces the best correction, and this is for the Morse FF. This is probably due to the higher precision for the correction PES for this limited energy range, as shown in Table 9. However, the results using the full range still show a significant improvement over the uncorrected FF, with the mean absolute error (MAE) being approximately five times less than that of the uncorrected FF. The reason for this can be deduced from Figures 7 and 8. As seen, the fitting errors are relatively small for energies up to 10,000 cm^{-1} and then grow rapidly above that energy. Therefore, for properties that are largely determined by energies up to 10,000 cm^{-1}, such as harmonic frequencies and torsional barriers the correction PES trained on this energy range performs well.

Lastly, we analyzed the torsional barrier for the methyl rotor calculated by Δ-ML PES using the force field. The results of the methyl torsional barrier height for the trans isomer, calculated from the various PESs of the force field, are listed in Table 11. As shown in Figure 9, the torsional barrier height for the harmonic force field is much lower than the direct CCSD(T) value. For all the corrected PESs, the barrier height

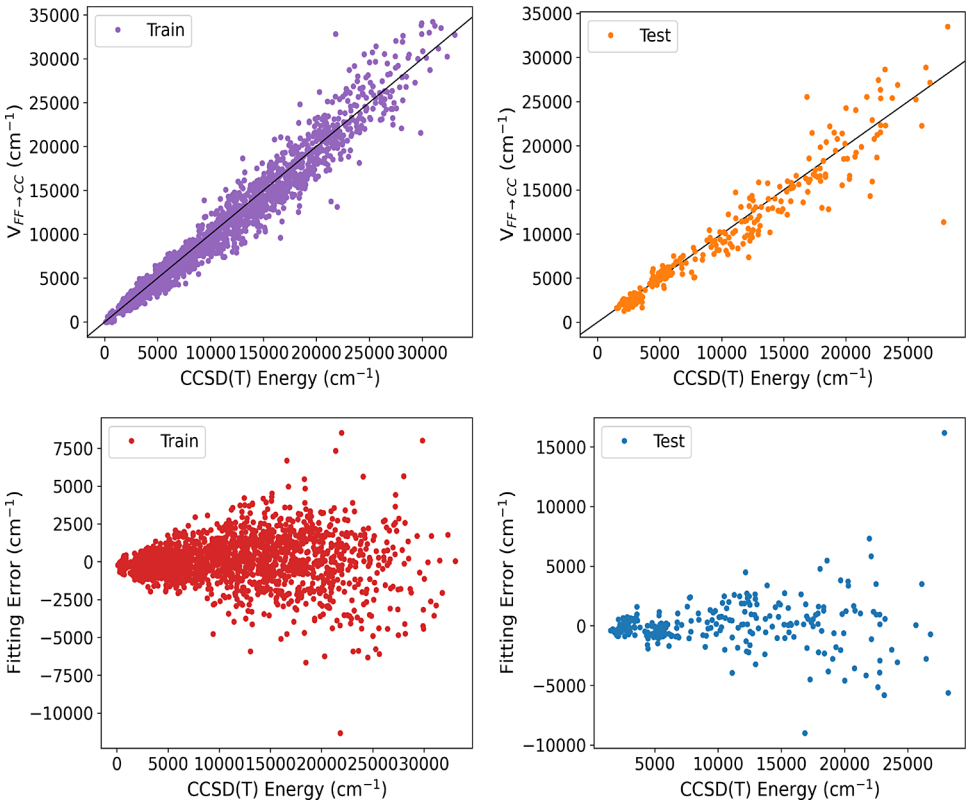


Figure 7. Two upper panels show energies of ethanol from $V_{FF\rightarrow CC}$ vs direct CCSD(T) ones for the indicated data sets calculated using the Harmonic approximation for the MP2 corrected force field. Corresponding fitting errors relative to the minimum energy are given in Table 9.

https://doi.org/10.1021/acs.jctc.4c00977
J. Chem. Theory Comput. 2024, 20, 8807−8819

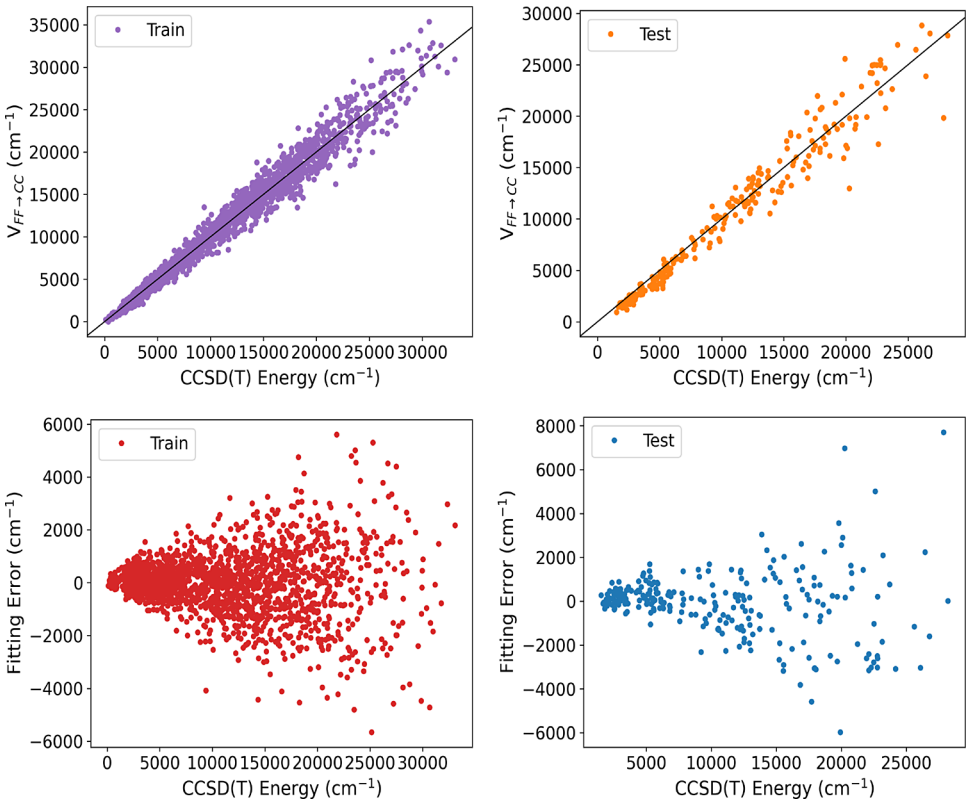


Figure 8. Two upper panels show energies of ethanol from $V_{FF\rightarrow CC}$ vs direct CCSD(T) ones for the indicated data sets calculated using the Morse potential for the MP2 corrected force field. Corresponding fitting errors relative to the minimum energy are given in Table 9.

Table 9. RMS Error in Δ-Corrected Energies (in cm^{-1}) Computed Using the Force Field with the Original Harmonic Stretch and Present Morse Potential Modified Stretch Potentials for Train and Test Data Sets[a]

RMS error	Harmonic FF[b]	Morse FF[b]	Morse FF[c]	Morse FF[d]
train	1436	1089	294	98
test	2097	1529	462	233

[a]These calculations use the fitting basis of 208 terms described in the text. [b]Fit using all CCSD(T) data to roughly 32,000 cm^{-1}. [c]Fit using data at 10,000 cm^{-1}. [d]Fit using data at 5000 cm^{-1}.

improves. The barrier height matches to the direct CCSD(T) value for the Morse FF when the full data set is considered.

Overall, the correction to this classical FF has been successful. And, it is reasonable to ask how the approach taken could be used for general classical FFs, especially for molecules much larger than ethanol. There is not a simple answer to this question, but clearly this is a fruitful area for future work. One preliminary thought is to make use of the simple form of FFs, which is just the generalization of eqs 4−6, and to correct groups of terms instead of the entire FF.

We end this section with some remarks about the history of dual-level (broadly defined) approaches to developing ab initio-based PESs. Perhaps one of the earliest examples dates to 1985, where we would consider a low-level ab initio quartic force field for formaldehyde was reported. In this instance, the harmonic force constants were adjusted to improve and agree with the experiment using vibrational configuration interaction calculations, with all other force constants unchanged.[60] This early approach evolved into a more sophisticated one using the N-mode representation of the potential.[61] By using this

Table 10. Comparison of Harmonic Frequencies (in cm^{-1}) between $V_{FF\rightarrow CC}$ PES Computed at Indicated Force Field and Corresponding Ab Initio Ones (CCSD(T)-F12a/aug-cc-pVDZ) for *Trans*-ethanol

	CCSD(T)	force field	Harmonic FF	Morse FF		
mode	direct	harmonic	ΔML[a]	ΔML[a]	ΔML[b]	ΔML[c]
1	222	234	258	381	245	234
2	274	261	276	385	277	240
3	413	606	400	507	351	402
4	813	1131	631	747	748	779
5	907	1148	829	875	917	960
6	1049	1300	835	992	1087	1083
7	1115	1421	1135	1055	1148	1144
8	1180	1443	1320	1090	1165	1191
9	1274	1753	1322	1158	1224	1298
10	1300	1816	1453	1190	1314	1325
11	1402	1960	1481	1195	1347	1413
12	1456	1994	1564	1266	1390	1454
13	1484	2019	1585	1421	1423	1530
14	1501	2049	1600	1480	1427	1539
15	1531	2142	1786	1484	1508	1657
16	3001	4247	1933	3353	3035	3112
17	3036	4300	2200	3442	3224	3135
18	3042	4398	2256	3468	3295	3173
19	3122	4409	2420	3476	3316	3263
20	3127	4410	2674	3502	3322	3304
21	3853	5129	3524	4099	4145	3930
MAE		624	272	116	111	171

[a]Fit using full data points up to 35,000 cm^{-1}. [b]Fit using data points up to 10,000 cm^{-1}. [c]Fit using data points up to 5000 cm^{-1}.

https://doi.org/10.1021/acs.jctc.4c00977
J. Chem. Theory Comput. 2024, 20, 8807−8819

Table 11. Barrier Height of the Methyl Rotor Torsional Potential Calculated Using Δ-Corrected Force Field for the Trans Isomer[a]

CCSD(T)	force field	Harmonic FF	Morse FF		
direct	harmonic	ΔML[b]	ΔML[b]	ΔML[c]	ΔML[d]
1194	938	1280	1194	1316	1086

[a]Energies are in cm^{-1}. [b]Fit using full data points up to 35,000 cm^{-1}. [c]Fit using data points up to 10,000 cm^{-1}. [d]Fit using data points up to 5000 cm^{-1}.

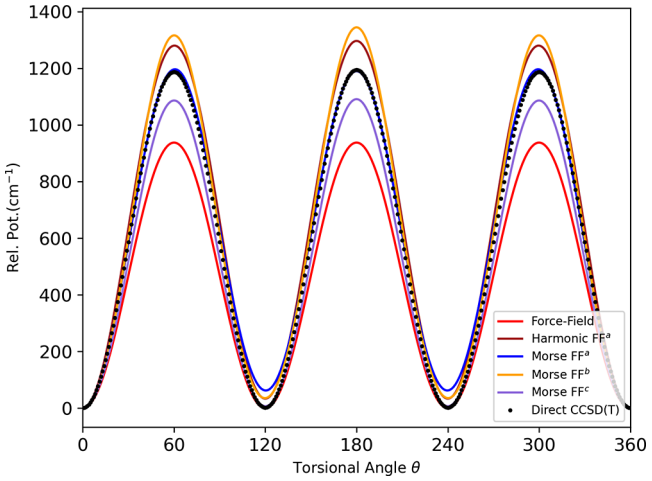


Figure 9. Comparison of torsional potential (not fully relaxed) of the methyl rotor of trans-ethanol between direct CCSD(T) and Δ-corrected PES computed using the force field.

representation, different levels of electronic structure theory are used for the different n-mode coupling terms.[62−65]

For global potentials, the Shepard interpolation method of Collins[66] was extended for chemical reactions using a dual-level method.[67] Specifically, high-level data points are placed along the reaction path, whereas the rest of the configuration space is described by a low-level method.

More recent work using machine learning to correct PESs uses Transfer Learning (TL)[68,69] and the Δ-ML described above. In TL the nonlinear parameters of a model, typically a neural network, trained on low-level data are retrained on sparse high-level data. This approach has been used successfully and extensively by Meuwly and co-workers for a number of applications, including anharmonic vibrational analyses, chemical reaction dynamics, and tunneling splittings.[70−73]

For the present approach, i.e., eq 1, we note that expression has been used to develop reaction PIP-NN PESs.[74,75] In this approach the correction potential ΔV_{CC-LL} is fit using the PIP-NN method,[76] and then used to generate data at the configurations where the low-level calculations were done. Then the sum of the low-level energies and those from ΔV_{CC-LL} are fit, again using the PIP-NN method. This results in a single Δ-corrected PES in contrast to the approach used here where the ΔV_{CC-LL} and the low-level PES are added together. We examined these two approaches for a PIP potential for the formic acid-ammonia dimer.[33] As expected, they yield virtually identical results. As already noted, the additional overhead in using the two-PES approach is small or, as in the present case, negligible compared to the one-PES. So, based on this alone, there is not a strong reason to prefer one approach over the other. However, we do suggest reporting the correction PES ΔV_{CC-LL} whichever approach is used. This will be useful in the event that holes are found in $V_{LL\rightarrow CC}$ or the single PES, say in high energy dynamics calculations.

SUMMARY AND CONCLUSIONS

The generality of the single-step Δ-ML method we proposed and applied using B3LYP to a number of PIP PESs has been demonstrated here for ethanol using other popular DFT functionals. In each case, the Δ-ML method produces a substantial improvement in accuracy compared to the CCSD(T) benchmark results. The most dramatic improvement is observed in the harmonic frequencies, where the DFT PIP PESs produce both significant underestimates and overestimates of the CH and OH-stretch frequencies. Additionally, we achieved significant improvement over DFT gradients without using CC gradient data to correct the PES. An exploratory application of this Δ-ML method to a recent force field (FF) for ethanol was given. Notably, the inaccurate harmonic frequencies at the global minimum from the force field are significantly corrected. The torsional barrier from the FF is also improved using the Δ-ML method. Additionally, the computational cost for the correction is about the same as the cost to evaluate the simple FF.

The new DFT Δ-ML potentials are expected to perform as well for Diffusion Monte Carlo and VSCF/VCI calculations as the original B3LYP Δ-ML one.[49,77] (We remind the interested reader that this corrected PES is available in Supporting Information in ref 77.) However, the performance of the Δ-ML corrected Force Field will need to be investigated for such calculations. In addition, it will also be of interest to test the new Quantum-Monte Carlo based sGDML potential[78] for such calculations.

Finally, this one-step Δ-ML method is very straightforward and can be easily implemented into other ML methods or descriptors. While ethanol molecule is used here as a prototype example, this approach is also applicable to large molecular systems for developing machine-learned force fields as accurate as the CC level.

AUTHOR INFORMATION

Corresponding Authors

Apurba Nandi − *Department of Physics and Materials Science, University of Luxembourg, L-1511 Luxembourg City, Luxembourg;* orcid.org/0000-0002-6191-5584; Email: apurba.nandi@uni.lu

Alexandre Tkatchenko − *Department of Physics and Materials Science, University of Luxembourg, L-1511 Luxembourg City, Luxembourg;* orcid.org/0000-0002-1012-4854; Email: alexandre.tkatchenko@uni.lu

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

Priyanka Pandey − *Department of Chemistry and Cherry L. Emerson Center for Scientific Computation, Emory University, Atlanta, Georgia 30322, United States;* orcid.org/0000-0002-6930-792X

Paul L. Houston − *Department of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853, United*

https://doi.org/10.1021/acs.jctc.4c00977
J. Chem. Theory Comput. 2024, 20, 8807−8819

States; Department of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, Georgia 30332, United States; orcid.org/0000-0003-2566-9539

Chen Qu − *Independent Researcher, Toronto, Ontario M9B0E3, Canada;* orcid.org/0000-0001-8889-4851

Qi Yu − *Department of Chemistry, Fudan University, Shanghai 200438, P. R. China;* orcid.org/0000-0002-2030-0671

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.jctc.4c00977

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

A.N. and A.T. acknowledge support from PHANTASTIC grant INTER/MERA22/16521502/PHANTASTIC. J.M.B. and P.P. acknowledge support from NASA grant 80NSSC22K1167. R.C. acknowledges support from Università degli Studi di Milano under grant PSR2022_DIP_005_PI_R-CONT.

■ ABBREVIATIONS

PES, potential energy surface

■ REFERENCES

(1) 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.

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

(3) Fu, B.; Zhang, D. H. Ab initio potential energy surfaces and quantum dynamics for polyatomic bimolecular reactions. *J. Chem. Theory Comput.* **2018**, *14*, 2289−2303.

(4) Jiang, B.; Li, J.; Guo, H. High-Fidelity Potential Energy Surfaces for Gas-Phase and Gas-Surface Scattering Processes from Machine Learning. *J. Phys. Chem. Lett.* **2020**, *11*, 5120−5131.

(5) 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.

(6) Chmiela, S.; Sauceda, H. E.; Müller, K.-R.; Tkatchenko, A. Towards exact molecular dynamics simulations with machine-learned force fields. *Nat. Commun.* **2018**, *9*, 3887.

(7) Sauceda, H. E.; Chmiela, S.; Poltavsky, I.; Müller, K. R.; Tkatchenko, A. Molecular force fields with gradient-domain machine learning: Construction and application to dynamics of small molecules with coupled cluster forces. *J. Chem. Phys.* **2019**, *150*, 114102.

(8) 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.

(9) Bartók, A. P.; Csányi, G. Gaussian approximation potentials: A brief tutorial introduction. *Int. J. Quantum Chem.* **2015**, *115*, 1051−1057.

(10) Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. *Chem. Sci.* **2017**, *8*, 3192−3203.

(11) 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*, No. 143001.

(12) Unke, O. T.; Meuwly, M. PhysNet: A Neural Network for Predicting Energies, Forces, Dipole Moments, and Partial Charges. *J. Chem. Theory Comput.* **2019**, *15*, 3678−3693.

(13) Dral, P. O.; Owens, A.; Yurchenko, S. N.; Thiel, W. Structure-based sampling and self-correcting machine learning for accurate calculations of potential energy surfaces and vibrational levels. *J. Chem. Phys.* **2017**, *146*, 244108.

(14) Dral, P. O. MLatom: A program package for quantum chemical research assisted by machine learning. *J. Comput. Chem.* **2019**, *40*, 2339−2347.

(15) Christensen, A. S.; von Lilienfeld, O. A. On the role of gradients for machine learning of molecular energies and forces. *Machine Learning: Science and Technology* **2020**, *1*, No. 045018.

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

(17) 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.

(18) Rasheeda, D. S.; Santa Daría, A. M.; Schröder, B.; Mátyus, E.; Behler, J. High-dimensional neural network potentials for accurate vibrational frequencies: the formic acid dimer benchmark. *Phys. Chem. Chem. Phys.* **2022**, *24*, 29381−29392.

(19) Fu, Y.-L.; Lu, X.; Han, Y.-C.; Fu, B.; Zhang, D. H.; Bowman, J. M. Collision-induced and complex-mediated roaming dynamics in the H + C_2H_4 → H_2 + C_2H_3 reaction. *Chem. Sci.* **2020**, *11*, 2148−2154.

(20) Lu, D.; Behler, J.; Li, J. Accurate Global Potential Energy Surfaces for the H + CH_3OH Reaction by Neural Network Fitting with Permutation Invariance. *J. Phys. Chem. A* **2020**, *124*, 5737−5745.

(21) Papp, D.; Tajti, V.; Győri, T.; Czakó, G. Theory Finally Agrees with Experiment for the Dynamics of the Cl + C_2H_6 Reaction. *J. Phys. Chem. Lett.* **2020**, *11*, 4762−4767.

(22) Czakó, G.; Gruber, B.; Papp, D.; Tajti, V.; Tasi, D. A.; Yin, C. First-principles mode-specific reaction dynamics. *Phys. Chem. Chem. Phys.* **2024**, *26*, 15818−15830.

(23) 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.

(24) 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.

(25) Stöhr, M.; Medrano Sandonas, L.; Tkatchenko, A. Accurate Many-Body Repulsive Potentials for Density-Functional Tight Binding from Deep Tensor Neural Networks. *J. Phys. Chem. Lett.* **2020**, *11*, 6835−6843.

(26) Dral, P. O.; Owens, A.; Dral, A.; Csányi, G. Hierarchical machine learning of potential energy surfaces. *J. Chem. Phys.* **2020**, *152*, 204110.

(27) 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*, No. 011103.

(28) 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.

(29) 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.

(30) 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.

(31) Nandi, A.; Nagy, P. R. Combining state-of-the-art quantum chemistry and machine learning make gold standard potential energy

surfaces accessible for medium-sized molecules. *Artif. Intell. Chem.* **2024**, *2*, No. 100036.
(32) Nandi, A.; Laude, G.; Khire, S. S.; Gurav, N. D.; Qu, C.; Conte, R.; Yu, Q.; Li, S.; Houston, P. L.; Gadre, S. R.; Richardson, J. O.; Evangelista, F. A.; Bowman, J. M. Ring-Polymer Instanton Tunneling Splittings of Tropolone and Isotopomers using a Δ-Machine Learned CCSD(T) Potential: Theory and Experiment Shake Hands. *J. Am. Chem. Soc.* **2023**, *145*, 9655−9664.
(33) Houston, P. L.; Qu, C.; Yu, Q.; Pandey, P.; Conte, R.; Nandi, A.; Bowman, J. M.; Kukolich, S. G. Formic Acid−Ammonia Heterodimer: A New Δ-Machine Learning CCSD(T)-Level Potential Energy Surface Allows Investigation of the Double Proton Transfer. *J. Chem. Theor. Comp.* **2024**, *20*, 1821−1828.
(34) 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 Comput.* **2023**, *19*, 1−17.
(35) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785−789.
(36) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(37) Devereux, M.; Boittier, E. D.; Meuwly, M. Systematic improvement of empirical energy functions in the era of machine learning. *J. Comput. Chem.* **2024**, *45*, 1899−1913.
(38) Unke, O. T.; Stöhr, M.; Ganscha, S.; Unterthiner, T.; Maennel, H.; Kashubin, S.; Ahlin, D.; Gastegger, M.; Medrano Sandonas, L.; Berryman, J. T.; Tkatchenko, A.; Müller, K. R. Biomolecular dynamics with machine-learned quantum-mechanical force fields trained on diverse chemical fragments. *Sci. Adv.* **2024**, *10*, No. eadn4397.
(39) Plé, T.; Lagardère, L.; Piquemal, J.-P. Force-field-enhanced neural network interactions: from local equivariant embedding to atom-in-molecule properties and long-range effects. *Chem. Sci.* **2023**, *14*, 12554−12569.
(40) Song, K.; Käser, S.; Töpfer, K.; Vazquez-Salazar, L. I.; Meuwly, M. PhysNet meets CHARMM: A framework for routine machine learning/molecular mechanics simulations. *J. Chem. Phys.* **2023**, *159*, No. 024125.
(41) Rogers, T. R.; Wang, F. Accurate MP2-based force fields predict hydration free energies for simple alkanes and alcohols in good agreement with experiments. *J. Chem. Phys.* **2020**, *153*, 244505.
(42) Zhao, Y.; Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **2008**, *120*, 215−241.
(43) Walker, M.; Harvey, A. J. A.; Sen, A.; Dessent, C. E. H. Performance of M06, M06-2X, and M06-HF Density Functionals for Conformationally Flexible Anionic Clusters: M06 Functionals Perform Better than B3LYP for a Model System with Dispersion and Ionic Hydrogen-Bonding Interactions. *J. Phys. Chem. A* **2013**, *117*, 12590−12600.
(44) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868. , Erratum Phys. Rev. Lett. 1997, 78, 1396.
(45) Perdew, J. P.; Ernzerhof, M.; Burke, K. Rationale for mixing exact exchange with density functional approximations. *J. Chem. Phys.* **1996**, *105*, 9982−9985.
(46) Adamo, C.; Barone, V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys.* **1999**, *110*, 6158−6170.
(47) Tkatchenko, A.; DiStasio, R. A.; Car, R.; Scheffler, M. Accurate and Efficient Method for Many-Body van der Waals Interactions. *Phys. Rev. Lett.* **2012**, *108*, No. 236402.
(48) Blum, V.; Gehrke, R.; Hanke, F.; Havu, P.; Havu, V.; Ren, X.; Reuter, K.; Scheffler, M. Ab initio molecular simulations with numeric atom-centered orbitals. *Comput. Phys. Commun.* **2009**, *180*, 2175−2196.
(49) 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.
(50) Braams, B. J.; Bowman, J. M. Permutationally invariant potential energy surfaces in high dimensionality. *Int. Rev. Phys. Chem.* **2009**, *28*, 577−606.
(51) Bowman, J. M.; Braams, B. J.; Carter, S.; Chen, C.; Czakó, G.; Fu, B.; Huang, X.; Kamarchik, E.; Sharma, A. R.; Shepler, B. C.; Wang, Y.; Xie, Z. Ab-initio-based potential energy surfaces for complex molecules and molecular complexes. *J. Phys. Chem. Lett.* **2010**, *1*, 1866−1874.
(52) Xie, Z.; Bowman, J. M. Permutationally Invariant Polynomial Basis for Molecular Energy Surface Fitting via Monomial Symmetrization. *J. Chem. Theory Comput.* **2010**, *6*, 26−34.
(53) Werner, H.-J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schütz, M. *MOLPRO, version 2015.1, a package of ab initio programs*. 2015; see http://www.molpro.net.
(54) Ren, X.; Rinke, P.; Blum, V.; Wieferink, J.; Tkatchenko, A.; Sanfilippo, A.; Reuter, K.; Scheffler, M. Resolution-of-identity approach to Hartree−Fock, hybrid density functionals, RPA, MP2 and GW with numeric atom-centered orbital basis functions. *New J. Phys.* **2012**, *14*, No. 053020.
(55) *MSA 2.0 Software with Gradients*, 2019. https://github.com/szquchen/MSA-2.0; Accessed: 2019−01−20.
(56) *Video MSA 2.0 Software with Gradients*, 2024. https://scholarblogs.emory.edu/bowman/msa/.
(57) Quade, C. A Note on Internal Rotation−Rotation Interactions in Ethyl Alcohol. *J. Mol. Spectrosc.* **2000**, *203*, 200−202.
(58) Pearson, J. C.; Sastry, K. V. L. N.; Winnewisser, M.; Herbst, E.; De Lucia, F. C. The Millimeter- and Submillimeter-Wave Spectrum of trans-Ethyl Alcohol. *J. Phys. Chem. Ref. Data* **1995**, *24*, 1−32.
(59) 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.
(60) Romanowski, H.; Bowman, J. M.; Harding, L. B. Vibrational energy levels of formaldehyde. *J. Chem. Phys.* **1985**, *82*, 4155−4165.
(61) Carter, S.; Culik, S. J.; Bowman, J. M. Vibrational self-consistent field method for many-mode systems: A new approach and application to the vibrations of CO adsorbed on Cu(100). *J. Chem. Phys.* **1997**, *107*, 10458−10469.
(62) Pflüger, K.; Paulus, M.; Jagiella, S.; Burkert, T.; Rauhut, G. Multi-level vibrational SCF calculations and FTIR measurements on furazan. *Theor. Chem. Acc.* **2005**, *114*, 327−332.
(63) Yagi, K.; Hirata, S.; Hirao, K. Multiresolution potential energy surfaces for vibrational state calculations. *Theor. Chem. Acc.* **2007**, *118*, 681−691.
(64) Sparta, M.; Høyvik, I.-M.; Toffoli, D.; Christiansen, O. Potential Energy Surfaces for Vibrational Structure Calculations from a Multiresolution Adaptive Density-Guided Approach: Implementation and Test Calculations. *J. Phys. Chem. A* **2009**, *113*, 8712−8723.
(65) Schröder, B.; Rauhut, G. From the Automated Calculation of Potential Energy Surfaces to Accurate Infrared Spectra. *J. Phys. Chem. Lett.* **2024**, *15*, 3159−3169.
(66) Ischtwan, J.; Collins, M. A. Molecular potential energy surfaces by interpolation. *J. Chem. Phys.* **1994**, *100*, 8080−8088.
(67) 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.
(68) Pan, S. J.; Yang, Q. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345−1359.
(69) Smith, J. S.; Nebgen, B. T.; Zubatyuk, R.; Lubbers, N.; Devereux, C.; Barros, K.; Tretiak, S.; Isayev, O.; Roitberg, A. E. Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning. *Nat. Commun.* **2019**, *10*, 2903−2906.
(70) Käser, S.; Boittier, E. D.; Upadhyay, M.; Meuwly, M. Transfer Learning to CCSD(T): Accurate Anharmonic Frequencies from

Machine Learning Models. *J. Chem. Theory Comput.* **2021**, *17*, 3687−3699.
(71) Käser, S.; Meuwly, M. Transfer learned potential energy surfaces: accurate anharmonic vibrational dynamics and dissociation energies for the formic acid monomer and dimer. *Phys. Chem. Chem. Phys.* **2022**, *24*, 5269−5281.
(72) 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.
(73) Käser, S.; Richardson, J. O.; Meuwly, M. Transfer Learning for Affordable and High-Quality Tunneling Splittings from Instanton Calculations. *J. Chem. Theory Comput.* **2022**, *18*, 6840−6850.
(74) 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. Chem. Phys. Lett.* **2022**, *13*, 4729−4738.
(75) Song, K.; Li, J. The neural network based Δ-machine learning approach efficiently brings the DFT potential energy surface to the CCSD(T) quality: a case for the OH + CH3OH reaction. *Phys. Chem. Chem. Phys.* **2023**, *25*, 11192−11204.
(76) Jiang, B.; Guo, H. Permutation invariant polynomial neural network approach to fitting potential energy surfaces. *J. Chem. Phys.* **2013**, *139*, No. 054112.
(77) 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.
(78) Slootman, E.; Poltavsky, I.; Shinde, R.; Cocomello, J.; Moroni, S.; Tkatchenko, A.; Filippi, C. Accurate quantum Monte Carlo forces for machine-learned force fields: Ethanol as a benchmark. *J. Chem. Theory Comput.* **2024**, *20*, 6020−6027.

https://doi.org/10.1021/acs.jctc.4c00977
J. Chem. Theory Comput. 2024, 20, 8807−8819