New Permutationally Invariant Polynomial Potential Energy Surfaces for $H_5O_2^+$ with Fast Analytical Gradients calculated Using Reverse Differentiation

Saikiran Kotaru,[†] Chen Qu,[‡] Paul L. Houston,[*,¶] Qi Yu,[§] Riccardo Conte,[‖] Apurba Nandi,[⊥] and Joel M. Bowman[*,†]

[†]*Department of Chemistry and Cherry L. Emerson Center for Scientific Computation, Emory University, Atlanta, Georgia 30322, U.S.A.*

[‡]*Independent Researcher, Toronto, Ontario M9B0E3, Canada*

[¶]*Department of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853, U.S.A.*

[§]*Department of Chemistry, Fudan University, Shanghai 200438, P. R. China*

[‖]*Dipartimento di Chimica, Università Degli Studi di Milano, via Golgi 19, 20133 Milano, Italy*

[⊥]*Department of Physics and Materials Science, University of Luxembourg, L-1511, Luxembourg City, Luxembourg.*

E-mail: plh2@cornell.edu; jmbowma@emory.edu

Abstract

Given the central importance of the protonated water dimer to the study of the hydrated proton, we report new fits to the previous CCSD(T) data set of Huang, Braams and Bowman (HBB) that are more precise and, unlike HBB, provide fast gradients. The new fits like the HBB one, are based on linear regression with permutationally invariant polynomials (PIPs). The fast gradients are provided via reverse differentiation. They cost roughly just three times the cost for an energy call and are roughly 20 times faster than the HBB numerical gradients. The two new PESs are fits to the original HBB datasets up to roughly 60,000 and to 110,000 cm^{-1}. Comparisons to the CCSD(T) benchmarks and to the HBB results for stationary points and Diffusion Monte Carlo ZPEs are reported and show good agreement.

Introduction

The Zundel cation, $H_5O_2^+$, is central to our understanding of the hydrated proton. As such, it has received widespread attention both theoretically and experimentally.[1–19] A systematic study of the dynamical behavior of Zundel contributes to a better understanding of proton transfer process in bulk water, which is fundamentally significant in chemistry and biology.

The first high-level CCSD(T)-based, full dimensional potential and MP2-based dipole moment surfaces for $H_5O_2^+$ were reported in 2005 using permutationally invariant polynomial (PIP) regression.[5] Significant extensions of that PES and DMS followed roughly a decade later by Yu and Bowman.[20] Prior to that work, pioneering work describing an MP2 based-PES for the hydrated proton was reported in 1998.[3] This "OSS" PES was based on elaborating many-component models for the interaction with numerous linear and non-linear parameters. These were optimized by non-linear least squares fitting of MP2 electronic energies and represented the state-of-the-art at the end of the 20th century. The precision shown in Fig. 1 of that paper is certainly below the level routinely seen today and in particular shown here. Indeed it was shown in 2005 that the OSS potential for Zundel is not quantitatively accurate.[5] Multi-state empirical valence bond (MS-EVB) potentials for the hydrated proton have also been reported.[21,22] These potentials are not quantitatively accurate compared to benchmark CCSD(T) energetics and frequencies for protonated water clusters. In contrast, the many-body PIP PES of Yu and Bowman is in excellent accord with such benchmarks.[15,17,20] However, it should be noted that this many-body PES was not trained on the liquid hydrated proton and so it cannot be used directly for such applications. The MB-EVB potentials can be used for these, and a hybrid approach in which snapshots from a molecular dynamics simulation using an MS-EVB potential[23] were used to obtain vibrational spectra using the many-body PIP PES and DMS.[24] Finally, a recent neural network (NN) potential for $H_5O_2^+$ has been reported,[25] based on a limited set of CCSD(T) electronic energies compared to the HBB dataset; we discuss this surface below.

The HBB PES does not provide analytical gradients, and so obtaining those from finite

difference approximations is computationally slow. Even so, this potential has been used for path integral molecular dynamics of the temperature dependence of the structural fluctuations of $H_5O_2^+$ and its isotopomer[26] and "Variational Quantum Monte Carlo with Path Integral Langevin Dynamics" calculations.[27] Here we report two new PIP PESs. These are fits to the CCSD(T)/aVTZ dataset of energies used for the 2005 HBB PES. Briefly, the basis sets of PIPS are obtained via MSA software,[28,29] and they are also post-processed using MATHEMATICA-based software[30] to include fast analytical gradients, both those produced by forward and reverse differentiation.

Methods

Fitting with PIPs

To begin, recall the expression for the potential is,[28,31]

$$V(\mathbf{y}) = \sum_{\alpha=1}^{M} c_\alpha p_\alpha(\mathbf{y}), \tag{1}$$

where the c_α are the linear coefficients, p_α are PIPs and $\mathbf{y}$ are the collection of Morse variables. The PESs use a PIP basis of symmetry A_5B_2. One new PES uses a PIP basis with a maximum polynomial order of 7 resulting in a size of 8717 polynomials and is a fit to the HBB dataset up to roughly 110 000 cm^{-1}; this PES is denoted "PES110K". A second new fit uses a basis of 8001 terms and a subset of the HBB dataset up to roughly 60 000 cm^{-1}; this PES is denoted "PES60K". The range parameter, a, in the Morse variable, exp(-r_{ij}/a), is 3.0 bohr, where r_{ij} is the internuclear distance between atoms i and j. Details of these fits are given below.

A histogram of the data set of 48,199 CCSD(T) energies used in the 2005 paper[5] is shown in Fig. 1. As seen, it extends to very high energies, corresponding to highly distorted structures. Also, there is extensive coverage out to dissociation, namely O-O distances of

300, 150, 75, 35, 20, 17, 14, 11, 9, 8, 7, 6, 5, and 4 bohr.[5] The fit was done, as usual, using

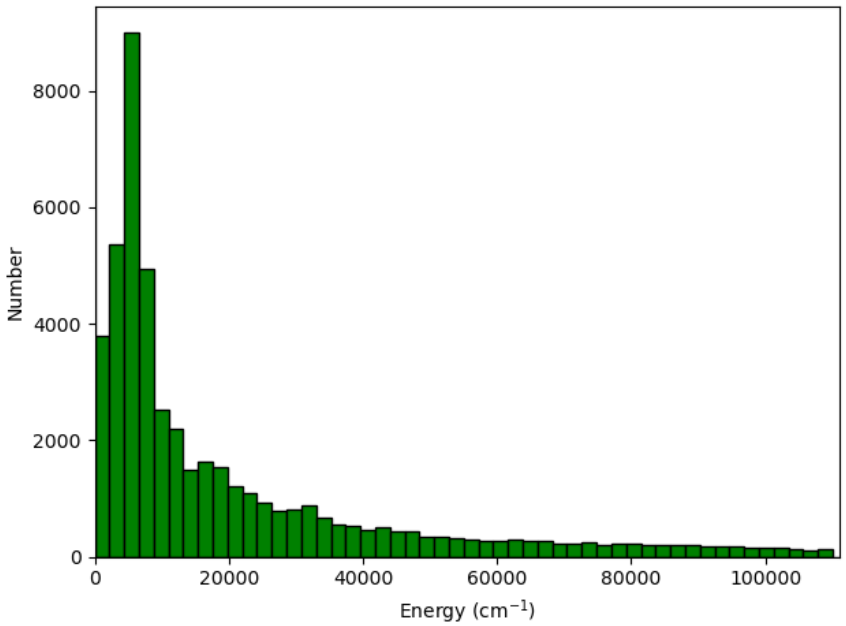


Figure 1: Histogram of CCSD(T) energies relative to the global minimum energy.

least-squares without regularization as described recently.[31] The energies in hartrees were weighted using the weight $w(E) = \Delta/(\mathrm{E}+\Delta)$, where $\Delta = 0.1$ hartree. Several fits, using maximum basis orders of 3-7 were performed using the full range of the data set.

One further fit, along with the 7^{th}-order fit, was selected for development. For this further fit, a PIP basis with 8001 polynomials/coefficients was made as described below. While the 7^{th}-order basis was fit to all CCSD(T)/aug-cc-pVTZ geometries shown in Fig. 1, the 8001 coefficient fit was fit to a truncated set, one with energies up to a maximum of about 60,000 cm^{-1}. We will call the two fits PES110K and PES60K based on their energy range. The PES60K basis set was obtained by adding basis functions to the 6^{th}-order, 2651 coefficient basis using PESPIP software.[30] Because the product of two PIPs is also a PIP, we can be assured that the permutation symmetry is maintained. To determine which PIPs are most important, we evaluate all possible new PIPs by calculating their values for all geometries in the data base and recording the maximum value obtained for each proposed PIP. Recall that the polynomials are Morse variables that decrease with the internuclear distances. Thus,

those PIPs with the largest values will correspond to the atom pairs that are closest to one another and should be most important to keep. We determined these in forming a basis of 8001 PIPs.

Next, we briefly describe the software to use reverse differentiation to produce fast analytical gradients.

Reverse Differentiation and Fast Analytical Gradients

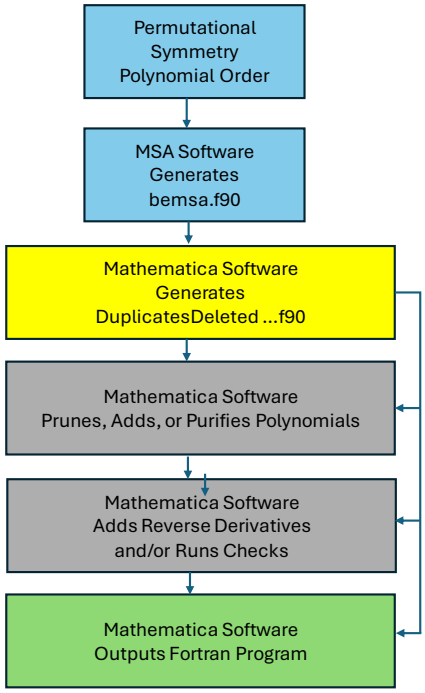


Figure 2: Flow Chart for Mathematica Software

The last steps for completing the basis set are to add methods for calculating analytical gradients. Fig. 2 shows the procedure using available software. After deciding on the desired permutational symmetry and the polynomial order, the user runs MSA software[28,29] (`https://github.com/szquchen/MSA-2.0`). This software generates a Fortran file, bemsa.f90, which contains the list of monomials and polynomials needed for generating the PIP basis. Optionally, the MSA software can provide analytical gradients. The MATHEMATICA[32]

software developed by our group[30] (`https://github.com/PaulLHouston/PESPIP`) reads this file and generates a DuplicatesDeleted....f90 file that can be used by further MATHEMATICA software to prune, add, or purify the polynomial basis and to add fast, analytical reverse derivatives. It then outputs a Fortran file for compilation, fitting, and evaluation of the energy and gradients at desired geometries.

The MATHEMATICA software we developed uses symbolic calculations of the formulae needed to modify the basis or to add gradients. It then assembles these formulae into a Fortran program using its text manipulation features. After compilation, the formulae are evaluated in Fortran. The software is described in detail elsewhere,[30] so that we merely summarize the results concerning the addition of gradients here, emphasizing the difference between "normal analytic" gradients[33,34] and "reverse" (sometimes called "backward" or "back-propagated") gradients.[35,36] The former are obtained by differentiation of both sides of Eq. (1) to determine how the change in potential depends on the change in the p_α with respect to the Cartesian coordinates. Because this calculation needs to be performed for each coordinate at unit cost equal to that for determining the energy, the total cost is $3N$ times that for the energy. In contrast, the "gradient by reverse differentiation" method runs a calculation of the energy and then works backwards from the energy using partial differentiation to determine *all* the $3N$ gradients in a single pass. The total cost is approximately 3-4 times the cost of the energy and is *independent* of the number of atoms.[30,37] We used this reverse differentiation method to add fast analytic gradients to both PES60K and PES110K.

We now clarify an important point: the MATHEMATICA code to generate gradients must be run to generate Fortran output for each permutational symmetry and polynomial order of interest. Thus, there is an overhead on the order of an hour or so to generate the fast derivative method for most problems of interest. Once the basis set and derivative method have been established, however, they can be run without further change. Also, the basis and reverse derivative code can be used for any molecule with the same permutational symmetry and polynomial order.

Although the MATHEMATICA code is normally run on a computer with a MATHEMATICA license, it can also be run using the freely available Wolfram Engine (`https://www.wolfram.com/engine/`) with Jupyter notebook (`https://jupyter.org/`). The user needs to download and install Wolfram Engine, Jupyter, and the Wolfram Language kernel add-on for Jupyter notebooks (`https://github.com/WolframResearch/WolframLanguageForJupyter`). Then the user can 1) download the template from `https://github.com/PaulLHouston/PESPIP`, 2) copy the code provided in "TemplatesAndExamplesV1.2.nb" into the Jupyter notebook, and 3) run the program to generate the Fortran code with fast reverse derivatives included.

Diffusion Monte Carlo

Diffusion Monte Carlo (DMC) calculations were performed using the two fits presented in this paper, PES110K and PES60K. These calculations employ the standard unbiased protocol described in ref 38. Specifically, 10 DMC "trajectories" were run for each PES, and in each DMC trajectory, 30000 walkers were initiated at the global minimum configuration, and were propagated for 55,000 steps, with an imaginary-time step size of 5.0 au. The first 5000 steps were used for equilibration, and the reference energies of the remaining 50,000 steps were collected and averaged to calculate the zero-point energy of the molecule. The standard deviation of the ZPEs from 10 trajectories are calculated to estimate the uncertainty of the DMC calculation.

Results and Discussion

To begin, we present a mainly pedagogical study of the precision of PIP fits for a variety of sizes of PIP bases and fits to the full dataset, which extends to a maximum energy of roughly 110 000 cm^{-1}. We also show a result most relevant for the paper, where the dataset is restricted to energies up to 60,000 cm^{-1}. The former results are shown in Table 1 and graphically in Figure 3. For the largest basis, the unweighted and weighted RMS fitting error are 36 and 25 cm^{-1}, respectively. These are very small and close to the fitting errors reported in 2005.

Table 1: Fitting Precision (cm^{-1}) for Indicated PIP Bases and maximum energy in the dataset

Max Order	Basis Size	max. E cm^{-1}	RMSE	wRMSE
3	59	110k	5598	3590
4	218	110k	1922	1207
5	772	110k	661	410
6	2651	110k	207	128
7, PES110K	8717	110k	36	25
6+, PES60K	8001	60k	13	6

The precision of the fit PES60K using energies up to 60k cm^{-1} is shown in Figure 4. The unweighted rms error is 13.0 cm^{-1}, the mean absolute error (MAE) is 6.9 cm^{-1}, and the R^2 coefficient is 0.999999. In addition to normal analytical gradients, this 60k cm^{-1} fit also provides gradients obtained by reverse differentiation[37] as well as a Hessian calculation based on the reverse differentiation gradients combined with numerical ones. The gradients calculated by reverse differentiation are significantly faster than the normal analytical ones, and both of these are much faster than fully numerical ones.[37]

Timing results using reverse differentiation were measured for the PES60K fit. Using an Intel® Core™ i7-8750H (2.20 GHz) processor, the times for calculating 10,000 energies and 10,000 gradient sets were 1.60 and 3.75 seconds, respectively.

Normal mode analyses of the PES60K fit are compared to the those using the HBB[5] PES in Table 2. For the C_2 minimum, the mean absolute error (MAE) for a comparison of

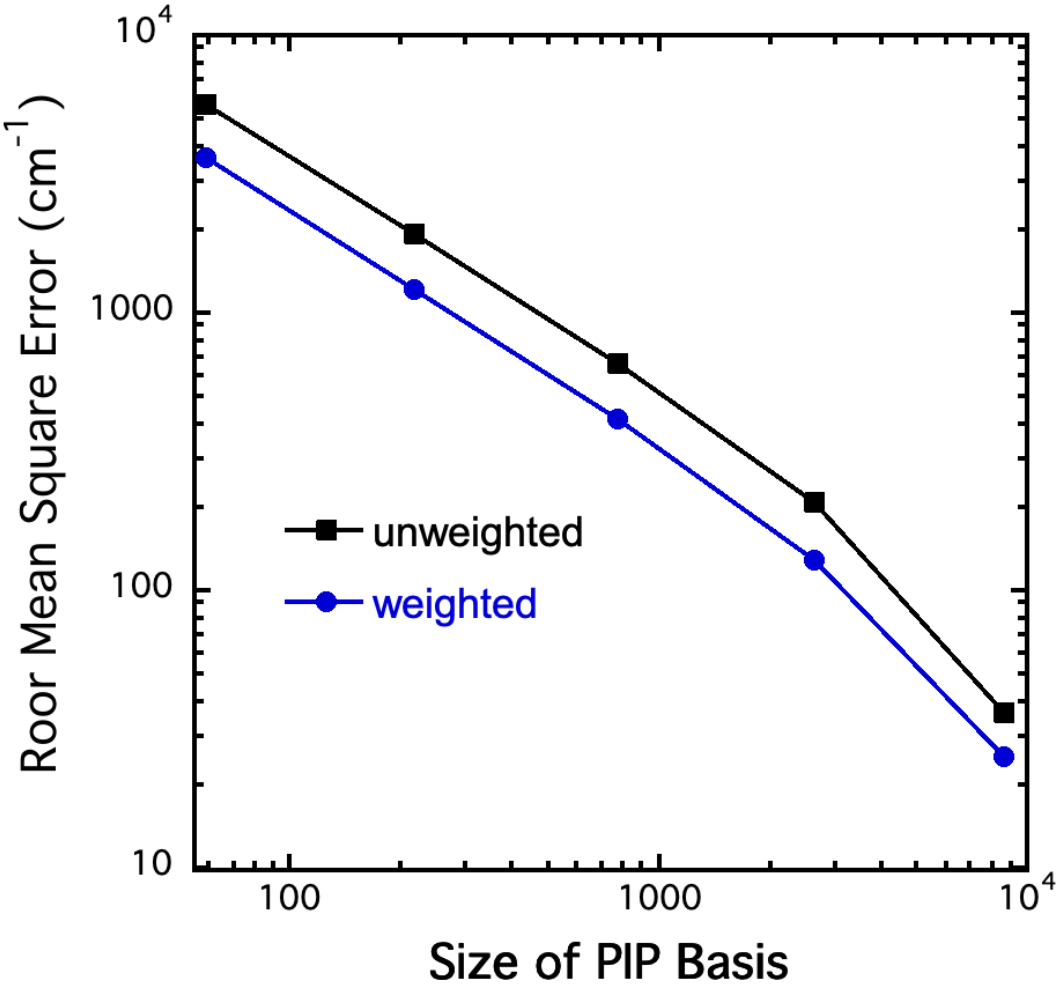


Figure 3: RMSE as a function of basis size for fits to the entire data set of 110,000 energies.

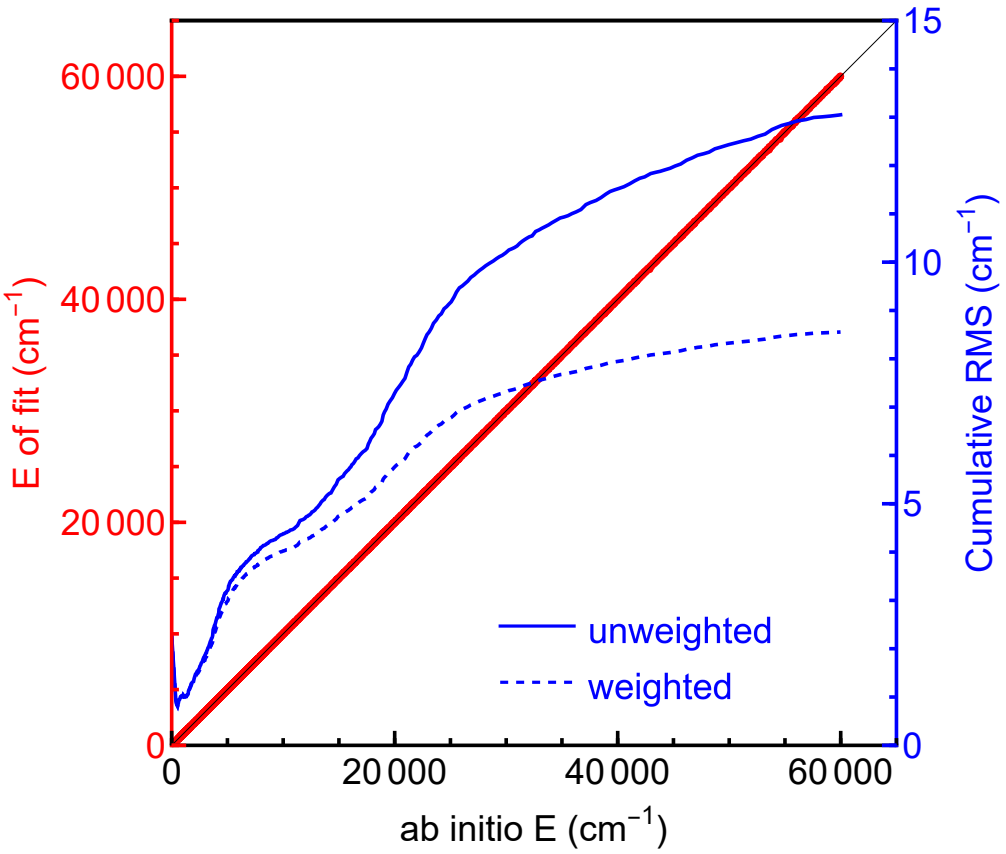


Figure 4: Correlation plot showing the agreement between the ab initio data and the fit, PES60K, to that data using the 8001 coefficient basis.

the harmonic frequencies is 4.3 cm^{-1}, while for the C_s inv transition state, the MAE is 3.3 cm^{-1}. Frequencies and energies for three other low-lying stationary states are also provided. Both surfaces give reasonable agreement with the energies of CCSD(T) benchmark. The agreement for the C_{2h} trans, C_{2v} cis, and C_s inv transition states is especially noteworthy, as these stationary points are not in the training data set.

Table 2: Energies and Harmonic frequencies (cm^{-1}) for low-lying stationary points on the $H_5O_2^+$ PES60K

Energy						
Geom.	C_2 min		C_{2h} trans	C_{2v} cis	C_s inv	
HBB[5]	0		214	445	158	
PES60K	0		217	442	164	
CCSD(T)	0		213	434	164	
Harmonic Frequency						
Mode	HBB[5]	PES60K	PES60K	PES60K	HBB[5]	PES60K
1	170	167	$125i$	$190i$	$326i$	$325i$
2	339	331	342	269	186	185
3	471	460	450	459	429	431
4	532	536	557	475	492	490
5	554	553	616	618	497	490
6	630	629	742	734	591	583
7	861	858	791	793	1206	1210
8	1494	1477	1201	1152	1401	1394
9	1574	1570	1687	1709	1621	1619
10	1720	1725	1763	1725	1688	1691
11	1770	1774	1833	1826	1817	1820
12	3744	3743	3732	3741	3724	3728
13	3750	3750	3733	3751	3789	3792
14	3832	3832	3816	3832	3804	3807
15	3832	3834	3820	3833	3889	3887
	MAE =	4.3			MAE =	3.3

DMC calculations were done to locate holes in the PESs. None were found. The ZPE of $H_5O_2^+$, using PES110K and PES60K, were 12402 ± 2 $cm^{-1}$and 12386 ± 2 cm^{-1}, respectively. The results are in good agreement with the DMC ZPE of 12395 ± 5 cm^{-1} using the HBB potential.[39]

Figure 5 shows the minimum energy pathway for dissociation/association for $H_5O_2^+ \rightleftharpoons$

$H_2O + H_3O^+$ as a function of the O-O distance on an 8001 coefficient PES, PES60K. This "relaxed" pathway was created by starting with the dissociation products, each in their global minimum configuration, at an O-O distance of 30 bohr and performing a simulated annealing trajectory by decreasing the kinetic energy at each step of the trajectory. The conditions of the trajectory were as follows: an initial kinetic energy of 12 cm^{-1}, a time step of 5 a.u., and a loss of 0.01% of the kinetic energy at each time step. In addition, a "brake" was used to prevent excessive kinetic energy from causing the trajectory to leave the minimum energy path; if the kinetic energy exceeded 100 cm^{-1}, it was reduced by 50%. In order to determine the inner wall of the potential cut, a similar trajectory was performed starting at an O-O distance of 3.7 bohr with a time step of 3 a.u., a loss of 0.005% per step, an initial kinetic energy of 0.1 cm^{-1}, and a brake at 6 cm^{-1}. Finally, a trajectory starting at 15 bohr with 4 a.u. step, 0.01% loss, 5 cm^{-1} kinetic energy, and brake at 6 cm^{-1} was found to agree extremely well with the 30 bohr starting trajectory where they overlapped. The resultant potential energy cut shown in Fig. 5 dissociates smoothly and indicates a D_e of 11,933 cm^{-1}, in good agreement with the result on the HBB surface (see ref. 5, Fig. 5) of 11,924 cm^{-1}. In fact, the two PES cuts are nearly indistinguishable. When corrected for zero point energies, these two numbers are in good agreement with the experimental D_0, values.[5,40]

As mentioned in the Introduction, a recent PES for the Zundel cation, called the BBSM PES,[19] has been reported[25] that uses an atom-centered, Behler-Parinello neural network fitting approach.[41,42] It is based on geometries whose energies are calculated using the CCSD(T*)-F12a/aug-cc-VTZ method, which is better corrected for basis set incompleteness error than the CCSD(T)/aug-cc-VTZ method use here and for the HBB potential. The training data set for H_2O and clusters from H_3O^+ to $H_9O_4^+$ was comprised of 49242 calculated energies, and the training and test sets for $H_5O_2^+$ included 8869 and 711 energies, respectively. The RMSE was found to be 0.07 (kJ/mol)/atom for the training set and 0.09 (kJ/mol)/atom for the test set. These correspond to RMSE values of 53 cm^{-1} and 41 cm^{-1}, respectively, and can be compared to the values of 36 cm^{-1} for our (unweighted) PES110K

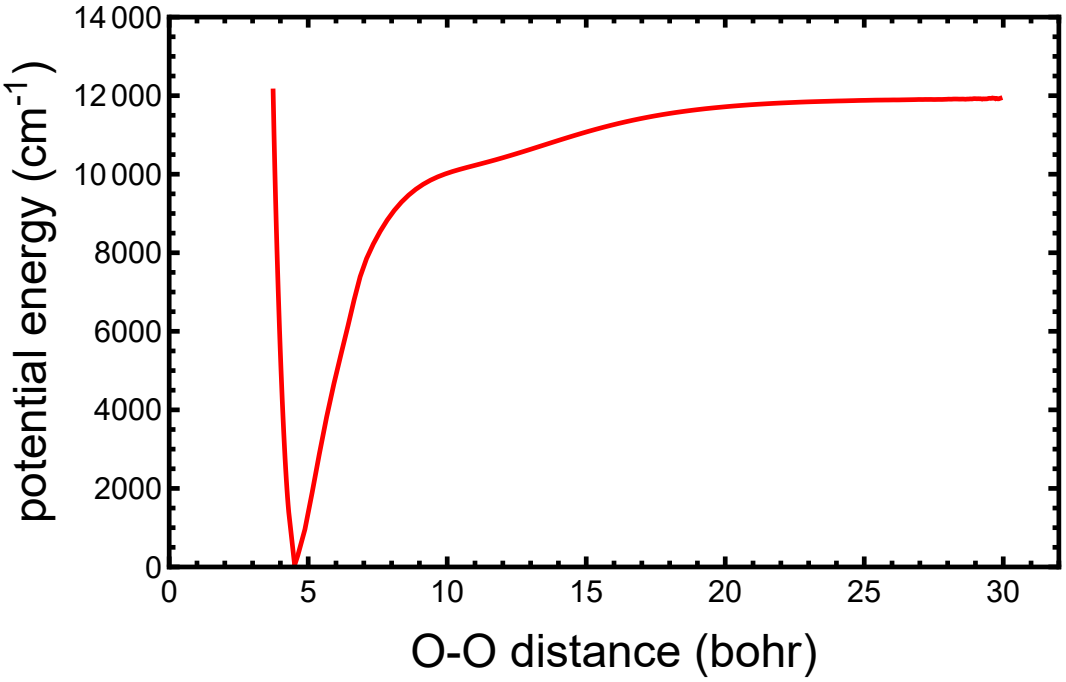


Figure 5: Cut through an 8001 coefficient surface as a function of the O-O distance. At each point along the cut, all other coordinates are at their minimum energy. The O-O distance at the minimum is at 4.5124 bohr, and the dissociation energy is 11,933 cm^{-1}.

with 8717 coefficients and 13 cm^{-1} for our (unweighted) PES60K with 8001 coefficients. However, our PES110K data set consists of 48,199 geometries/energies up to 110,046 cm^{-1}, whereas their training and test set consists of (a total of) 13,600 geometries/energies up to about 28,000 cm^{-1}. Thus it would be more appropriate to compare their PES to our 8001 coefficient PES60K fit, which is based on 43523 geometries/energies extending up to 60,000 cm^{-1}. Based on the above data, our fit to CCSD(T) data appears to be about three times more precise than the BBSM fit to the CCSD(T*)-F12a data. Using the CCSD(T*)-F12a data as the target, ref. 19 reports that the BBSM PES is about three times more accurate when compared to the BBH potential. These statements are not inconsistent because the less precise BBSM fit was trained on the more accurate CCSD(T*)-F12a data, whereas our more precise fit and the HBB one are fit to the less accurate CCSD(T) data. While it would be instructive to compare directly the current surface with the BBSM one in terms of the values of D_e and the potential energy cuts for dissociation along the O-O distance, this

information for the BBSM surface is not available.[19,25] We note that their $H_5O_2^+$ dataset, from the SI of Ref. 19, includes only 29 points with O-O distances larger than 7 bohr, and the largest of these is 11.4 bohr, which on our O-O PES cut would be at only 86% of the dissociation energy. While longer distances from larger clusters were incorporated into the total learning, it seems unlikely that the BBSM surface would correctly describe dissociation of the $H_5O_2^+$ Zundel cation. Of course the objective of their study was to develop a general potential for the hydrated proton using the Behler-Parinello NN fitting method in contrast to our approach which is focused on the Zundel cation. It is worth noting that earlier Qi and Bowman developed a general hydrated proton potential, based on a many-body representation in which the Zundel cation is the 2-b hydronium-water interaction.[17,20] Because this 2-b interaction can extend to large distances, the accurate dissociation of the Zundel cation potential is important.

Summary and Conclusions

We reported two new global PESs for the Zundel ($H_5O_2^+$) cation using permutationally invariant polynomial fitting to a data set of energies up to 110,046 cm^{-1}. Recommended surfaces for fits up energies up to 60,000 cm^{-1} and up to 110,046 cm^{-1} are given by PES110K and PES60K described above. Both PESs are also provided with fast, reverse derivative gradients and are available for download. Diffusion Monte Carlo calculations, comparisons of stationary points, and calculation of the potential energy cut as a function of O-O distances, show good agreement both with results on the HBB surface and with the CCSD(T) results, where available.

Data Availability

The full data set and the source code for the PES60K are available for download at `https://github.com/jmbowma/QM-22`. This is the faster program. Those interested in PES110k

should contact the authors.

Notes

The authors declare no competing financial interests.

Acknowledgment

SK and JMB acknowledge funding from NASA (grant 80NSSC22K1167).

References

(1) Yeh, L. I.; Lee, Y. T.; Hougen, J. T. Vibration - Rotation Spectroscopy of the Hydrated Hydronium Ions $H_5O_2^+$ and $H_9O_4^+$. *J. Mol. Spectrosc.* **1994**, *164*, 473.

(2) Asmis, K. R.; Pivonka, N. L.; Santambrogio, G.; Brummer, M.; Kaposta, C.; Newmark, D. M.; Woste, L. Gas Phase Infrared Spectrum of the Protonated Water Dimer. *Science* **2003**, *299*, 1375–1377.

(3) Ojamäe, L.; Shavitt, I.; Singer, S. J. Potential models for simulations of the solvated proton in water. *The Journal of Chemical Physics* **1998**, *109*, 5547–5564.

(4) Dai, J.; Bačić, Z.; Huang, X.; Carter, S.; Bowman, J. M. A theoretical study of vibrational mode coupling in $H_5O_2^+$. *J. Chem. Phys.* **2003**, *119*, 6571–6580.

(5) Huang, X.; Braams, B. J.; Bowman, J. M. Ab initio Potential Energy and Dipole Moment Surfaces for $H_5O_2^+$. *J. Chem. Phys.* **2005**, *122*, 044308.

(6) Hammer, N. I.; Diken, E. G.; Roscioli, J. R.; Johnson, M. A.; Myshakin, E. M.; Jordan, K. D.; McCoy, A. B.; Huang, X.; Bowman, J. M.; Carter, S. The Vibrational Predissociation Spectra of the $H_5O_2^+$ RG_n(RG=Ar,Ne) Clusters: Correlation of the

Solvent Perturbations in the Free OH and Shared Proton Transitions of the Zundel Ion. *J. Chem. Phys.* **2005**, *122*, 244301.

(7) Huang, X.; Habershon, S.; Bowman, J. M. Comparison of Quantum, Classical, and Ring-polymer Molecular Dynamics Infrared Spectra of $Cl^-(H_2O)$ and $H^+(H_2O)_2$. *Chem. Phys. Lett.* **2008**, *450*, 253 – 257.

(8) Vendrell, O.; Gatti, F.; Meyer, H. D. Dynamics and Infrared Spectroscopy of the Protonated Water Dimer. *Angew. Chem. Int. Ed.* **2007**, *46*, 6918–6921.

(9) Vendrell, O.; Gatti, F.; Meyer, H. D. Strong Isotope Effects in the Infrared Spectrum of the Zundel Cation. *Angew. Chem. Int. Ed.* **2009**, *48*, 352–355.

(10) Niedner-Schatteburg, G. Infrared spectroscopy and ab initio theory of isolated H5O2+: From buckets of water to the Schrödinger equation and back. *Angew. Chem. Int. Ed Engl.* **2008**, *47*, 1008–1011.

(11) Rossi, M.; Ceriotti, M.; Manolopoulos, D. E. How to Remove the Spurious Resonances from Ring Polymer Molecular Dynamics. *J. Chem. Phys.* **2014**, *140*, 234116.

(12) Fournier, J. A.; Wolke, C. T.; Johnson, M. A.; Odbadrakh, T. T.; Jordan, K. D.; Kathmann, S. M.; Xantheas, S. S. Snapshots of Proton Accommodation at a Microscopic Water Surface: Understanding the Vibrational Spectral Signatures of the Charge Defect in Cryogenically Cooled $H^+(H_2O)_n$=2-28 Clusters. *J. Phys. Chem. A* **2015**, *119*, 9425–9440.

(13) Wolke, C. T.; Fournier, J. A.; Dzugan, L. C.; Fagiani, M. R.; Odbadrakh, T. T.; Knorke, H.; Jordan, K. D.; McCoy, A. B.; Asmis, K. R.; Johnson, M. A. Spectroscopic Snapshots of the Proton-transfer Mechanism in Water. *Science* **2016**, *354*, 1131.

(14) 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.

(15) Yu, Q.; Bowman, J. M. High-Level Quantum Calculations of the IR Spectra of the Eigen, Zundel, and Ring Isomers of $H^+(H_2O)_4$ Find a Single Match to Experiment. *J. Am. Chem. Soc.* **2017**, *139*, 10984–10987.

(16) Duong, C. H.; Gorlova, O.; Yang, N.; Kelleher, P. J.; Johnson, M. A.; McCoy, A. B.; Yu, Q.; Bowman, J. M. Disentangling the Complex Vibrational Spectrum of the Protonated Water Trimer, $H^+(H_2O)_3$, with Two-Color IR-IR Photodissociation of the Bare Ion and Anharmonic VSCF/VCI Theory. *J. Phys. Chem. Lett.* **2017**, *8*, 3782–3789.

(17) Heindel, J. P.; Yu, Q.; Bowman, J. M.; Xantheas, S. S. Benchmark Electronic Structure Calculations for $H_3O^+(H_2O)_n$, n=0-5 Clusters and Tests of an Existing 1,2,3-body Potential Energy Surface with a New 4-body Correction. *J. Chem. Theory Comput.* **2018**, *14*, 4553–3566.

(18) Moonkaen, P.; McCoy, A. B. Evaluation of Infrared Intensities Using Diffusion Monte Carlo. *J. Phys. Chem. A* **2025**, *129*, 2705–2717.

(19) Larsson, H. R.; Schröder, M.; Beckmann, R.; Brieuc, F.; Schran, C.; Marx, D.; Vendrell, O. State-resolved infrared spectrum of the protonated water dimer: revisiting the characteristic proton transfer doublet peak. *Chem. Sci.* **2022**, *13*, 11119–11125.

(20) Yu, Q.; Bowman, J. M. Communication: VSCF/VCI Vibrational Spectroscopy of $H_7O_3^+$ and $H_9O_4^+$ using High-level, Many-body Potential Energy Surface and Dipole Moment Surfaces. *J. Chem. Phys.* **2017**, *146*, 121102.

(21) Schmitt, U. W.; Voth, G. A. Multistate Empirical Valence Bond Model for Proton Transport in Water. *J. Phys. Chem. B* **1998**, *102*, 5547–5551.

(22) Kumar, R.; Christie, R. A.; Jordan, K. D. A Modified MSEVB force field for protonated water clusters. *J. Phys. Chem. B* **2009**, *113*, 4111–4118.

(23) Wu, Y.; Chen, H.; Wang, F.; Paesani, F.; Voth, G. A. An improved Multistate Empirical Valence Bond Model for Aqueous Proton Solvation and Transport. *J. Phys. Chem. B* **2008**, *112*, 467.

(24) Yu, Q.; Carpenter, W. B.; Lewis, N. H.; Tokmakoff, A.; Bowman, J. M. High-level VSCF/VCI calculations decode the vibrational spectrum of the aqueous proton. *J. Phys. Chem. B* **2019**, *123*, 7214–7224.

(25) Schran, C.; Behler, J.; Marx, D. Automated Fitting of Neural Network Potentials at Coupled Cluster Accuracy: Protonated Water Clusters as Testing Ground. *J. Chem. Theory Comput.* **2020**, *16*, 88–99.

(26) Suzuki, K.; Tachikawa, M.; Shiga, M. Temperature dependence on the structure of Zundel cation and its isotopomers. *J. Chem. Phys.* **2013**, *138*, 184307.

(27) Mouhat, F.; Sorella, S.; Vuilleumier, R.; Saitta, A. M.; Casula, M. Fully Quantum Description of the Zundel Ion: Combining Variational Quantum Monte Carlo with Path Integral Langevin Dynamics. *J. Chem. Theory Comput.* **2017**, *13*, 2400–2417.

(28) 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.

(29) MSA Software with Gradients. `https://github.com/szquchen/MSA-2.0`, 2019; Accessed: 2019-01-20.

(30) Houston, P. L.; Qu, C.; Yu, Q.; Conte, R.; Nandi, A.; Li, J. K.; Bowman, J. M. PESPIP: Software to fit complex molecular and many-body potential energy surfaces with permutationally invariant polynomials. *J. Chem. Phys.* **2023**, *158*, 044109.

(31) Bowman, J. M.; Qu, C.; Conte, R.; Nandi, A.; Houston, P. L.; Yu, Q. A perspective marking 20 years of using permutationally invariant polynomials for molecular potentials. *J. Chem. Phys.* **2025**, *162*, 180901.

(32) WolframResearch *Mathematica*, version 11.2 or higher, ed.; Wolfram Research, Inc.: Champaign, Illinois, 2017.

(33) Qu, C.; Bowman, J. M. A Fragmented, Permutationally Invariant Polynomial Approach for Potential Energy Surfaces of Large Molecules: Application to N-methyl acetamide. *J. Chem. Phys.* **2019**, *150*, 141101.

(34) Nandi, A.; Qu, C.; Bowman, J. M. Using Gradients in Permutationally Invariant Polynomial Potential Fitting: A Demonstration for CH4 Using as Few as 100 Configurations. *JCTC* **2019**, *15*.

(35) Baydin, A. G.; Pearlmutter, B. A. Automatic Differentiation of Algorithms for Machine Learning. *JMLR: Workshop and Conference Proceedings, ICML 2014 AutoML Workshop* **2014**, 1–7.

(36) Baydin, A. G.; Pearlmutter, B. A.; Radul, A. A.; Siskind, J. M. Automatic Differentiation in Machine Learning: a Survey. *The Journal of Machine Learning Research* **2017**, *18*, 5595–5637.

(37) 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. *The Journal of Chemical Physics* **2022**, *156*, 044120.

(38) Kosztin, I.; Faber, B.; Schulten, K. Introduction to the Diffusion Monte Carlo Method. *Am. J. Phys.* **1996**, *64*, 633–644.

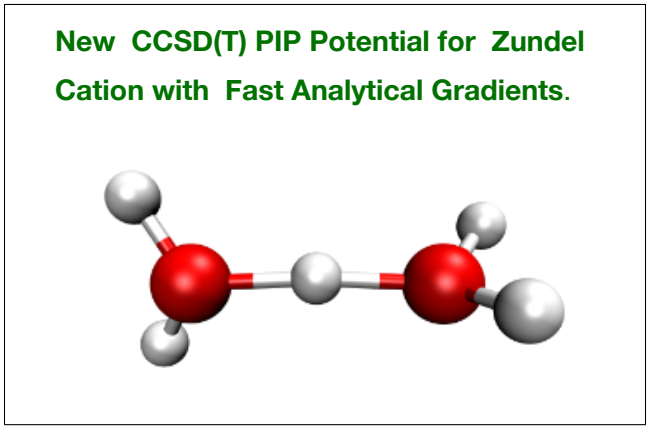
(39) McCoy, A. B.; Huang, X.; Carter, S.; Landeweer, M. Y.; Bowman, J. M. Full-dimensional Vibrational Calculations for $H_5O_2^+$ using an Ab Initio Potential Energy Surface. *J. Chem. Phys.* **2005**, *122*, 061101.

(40) Dalleska, N. F.; Honma, K.; Armentrout, P. B. Stepwise Solvation Enthalpies of Protonated Water Clusters: Collision-Induced Dissociation as and Alternative to Equilibrium Studies. *J. Am. Chem. Soc.* **1993**, *115*, 12125–12131.

(41) Behler, J.; Parrinello, M. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Phys. Rev. Lett.* **2007**, *98*, 146401.

(42) Behler, J. Constructing high-dimensional neural network potentials: A tutorial review. *Int. J. Quantum Chem.* **2015**, *115*, 1032–1050.

TOC Graphic


New CCSD(T) PIP Potential for Zundel
Cation with Fast Analytical Gradients.