SUPPORTING INFORMATION: Fast Near *Ab Initio* Potential Energy Surfaces using Machine Learning[†]

Fenris Lu,[‡] Lixue Cheng,[¶] Ryan J. DiRisio,[‡] Jacob M. Finney,[‡] Mark A. Boyer,[‡] Pattarapon Moonkaen,[‡] Jiace Sun,[¶] Sebastian J. R. Lee,[¶] J. Emiliano Deustua,[¶] Thomas F. Miller, III,[*,¶] and Anne B. McCoy [*,‡]

‡*Department of Chemistry, University of Washington, Seattle, WA 98195, USA*

¶*Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125, United States*

E-mail: tfm@caltech.edu; abmccoy@uw.edu

Phone: 206-543-7464

[†]FL, LC and RJD contributed equally to this work

CONTENTS:

Overview of Diffusion Monte Carlo (DMC)

DMC and the details of our implementation have been described elsewhere.[1–7] In this study, we use both guided and unguided DMC simulations to obtain the ground state energy and wave function for the systems of interest. We also use our recently developed NN-DMC approach,[8] in which we replace the potential energy surface with a neural network potential energy surface for the DMC simulation. This results in significant savings in the computational resources required for the DMC calculations.

For most of this study, we performed unguided DMC simulations. In an unguided DMC simulation, the ground state wave function, Φ_0, is represented by an ensemble of N_w localized functions, which we will refer to as walkers. The density of walkers in a particular region of configuration space provides the amplitude of the ground state wave function at that geometry. The ensemble of walkers explores the potential energy surface of the system of interest through a propagation in imaginary time, $\tau = it/\hbar$, based on the imaginary-time time-dependent Schrödinger equation,

$$\begin{aligned} \Phi_0(\tau+\Delta\tau) &= \exp\left[-(H-V_\mathrm{ref}(\tau))\Delta\tau\right]\Phi_0(\tau) \\ &\approx \exp\left[-\{V(\mathbf{x}_i(\tau))-V_\mathrm{ref}(\tau)\}\Delta\tau\right]\exp\left[-T\Delta\tau\right]\Phi_0(\tau) \end{aligned} \tag{S1}$$

At each time step, $\Delta\tau$, the position, $\mathbf{x}_i(\tau)$, and weight, $w_i(\tau)$, of each of the walkers are updated. Specifically, the coordinates of each of the atoms that are described by the walkers are displaced according to a Gaussian distribution, with a standard deviation of $\hbar\sqrt{\Delta\tau/m_j}$, where m_j is the mass of the atom that is displaced. The weight of the ith walker is updated based on

$$w_i(\tau+\Delta\tau) = \exp\left[-\{V(\mathbf{x}_i(\tau))-V_\mathrm{ref}(\tau)\}\Delta\tau\right] w_i(\tau) \tag{S2}$$

To ensure that a small fraction of walkers do not carry most of the weight, a branching step is introduced. In this step, the weights of the walkers are compared to upper and lower bound thresholds.

All walkers with weights that are smaller than the lower bound threshold are removed from the ensemble. To keep the ensemble size and sum of the weights constant, an equal number of walkers with the highest weight are duplicated, and each of the walkers and their copies are given a weight that is half the original weight of the duplicated walker. After all the low-weight walkers have been removed from the ensemble, walkers that have a weight larger than the upper bound threshold are also duplicated, as described above, and an equal number of walkers with the lowest weights are removed from the simulation.

To minimize the effects of process-to-process communication latency in the potential evaluations, in the DMC simulations of water and CH_5^+ used to obtain the **MOB-ML** energies reported in Table 6, we introduced a small adjustment to the continuous weighting DMC algorithm. In this modified approach, we propagate the coordinates and evaluate the potential energy of each of the walkers for N_τ steps before considering branching. Once the N_τ potential evaluations are complete, we update V_{ref}, the weights, and check for branching at each time step. This introduces an approximation into the DMC algorithm. Although we check for branching after each time step in the simulation, the branching is only applied every N_τ steps. We find that the delayed branching does not impact the overall accuracy of the DMC simulation, as typically fewer than 0.5% of the walkers undergo branching at each time step. By performing a smaller number of total MPI calls, we are able to cut down on the latency overhead involved in node-to-node communication.

For the DMC calculations that used the **NN-PES**, the weights of all walkers are constrained to 1, and the duplication or removal of walkers is achieved by an additional Monte Carlo step.[9] In this case, the ensemble size will fluctuate as the simulation progresses. This technique is referred to as discrete weighting, and the algorithm that allows the weights of the walkers to evolve with τ is called continuous weighting. Finally,

$$V_{\text{ref}}(\tau) = \frac{\sum_{i=1}^{N_{\text{w}}} w_i(\tau) V(\mathbf{x}_i(\tau))}{\sum_{i=1}^{N_{\text{W}}} w_i(\tau)} - \alpha \left[\frac{\sum_{i=1}^{N_{\text{w}}} w_i(\tau) - w_i(\tau = 0)}{\sum_{i=1}^{N_{\text{w}}} w_i(\tau = 0)} \right] \tag{S3}$$

is evaluated, where $\alpha = 0.5/\Delta\tau$. The introduction of the second term in Eq. S3 ensures the sum of the weights of the walkers is roughly constant throughout the simulation. The time-averaged value

of V_{ref} provides the zero-point energy of the system once the simulation has equilibrated.

The main difference between guided DMC simulations and unguided DMC simulations is that in the guided simulations, $f = \Phi_0 \Psi_T$ is represented by the ensemble of walkers, where Ψ_T is the guiding function. This change leads to the potential energy evaluations being replaced by evaluations of the local energy,

$$E_{\text{L}} = \frac{H\Psi_T}{\Psi_T} \tag{S4}$$

When Ψ_T provides a good approximation to Φ_0, the local energy is approximately constant. Using a guiding function also introduces a drift term that moves the walkers away from regions where the amplitude of Ψ_T is small and towards regions with large amplitude.

In several recent studies, we showed that using guiding functions that are direct products of one-dimensional wave functions in the high frequency stretches and, in the case of H_2O, the HOH bend, provide effective guiding functions for H_2O and CH_5^+,[5,6] and we employ the approaches developed in those studies to obtain the MOB-ML zero-point energies that are reported in the second column of Table 6. Finally, descendant weighting is used to obtain projections of the probability amplitude onto coordinates of interest.[3,4,10] The unguided NN-DMC simulations and the DMC simulations used to collect training structures and energies for $C_2H_5^+$ were performed using PyVibDMC, a general-purpose, open source simulation package,[11] while the calculations on water and CH_5^+ that utilized the MOB-ML energies were run with an earlier implementation of DMC.[12]

Numerical Details

Training the MOB-ML Potential Energy Surfaces

The 3000 training and test configurations for H_2O, CH_5^+ and $C_2H_5^+$, and 1000 training and test configurations for 8 small validation molecules are sampled at 50 fs intervals from *ab initio* molec-

ular dynamics (AIMD) trajectories performed with the Q-CHEM 5.0 software package,[13] using the B3LYP[14–17]/6-31G*[18] level of theory. To ensure the full coverage of the potential energy surfaces, a Langevin thermostat[19] at 6003 K is applied for the H_2O AIMD trajectory starting from the optimized H_2O geometry at B3LYP/6-31G* level of theory. For CH_5^+, three AIMD trajectories are performed by starting from the three literature local minima of CH_5^+[20] with a Langevin thermostat at 350 K, and each trajectory provides 1000 sampled configurations. For $C_2H_5^+$ and validation molecules, we followed the same configuration generation protocols described in Ref. 21, and 22 the configurations are sampled from a single 350 K AIMD trajectory starting from the optimized geometry at B3LYP/6-31G* level of theory.

In the case of $C_2H_5^+$, we were concerned that the ground state wave function samples geometries that are substantially higher in energy than the ones that are sampled in the 350 K trajectory. The source of the concern is illustrated in the results plotted in Figure S5(a), where we compare the predicted energies using MOB-ML model trained on 2500 AIMD structures against the CCDS(T) energies of the same structures. The structures were extracted from a ground-state DMC simulation, which was run using the MOB-ML model. As can be seen, there are errors in excess of 500 cm^{-1}, and the errors do not appear to be uniformly distribute about 0 cm^{-1}. We noted that many of the high energy configurations contained large displacements of the CH bond lengths. Based on this observation, we generated 1000 additional stretched geometries of $C_2H_5^+$ structures, which have energies up to 20 000 cm^{-1}. These structures were generated by randomly selecting a structure from the AIMD simulation and adjusting the five C-H distances to randomly selected values between 0.8 Å and 1.3 Å. In displacing these structures, the CC distance was kept constant, as were the HCC angles. Since there are two CH distances involving the bridging hydrogen atom, the one that corresponded to the shorter CH distance was stretched, keeping that HCC' bond angle constant with fixed non-bridged C-H bond orientations and CC distance.

Once the structures have been selected, we follow the same feature generation protocol described in Husch et al.[23] to compute the associated features at density-fitted HF with aug-cc-pVTZ[24] basis set and aug-cc-pVTZ-JKFIT density fitting basis set[25] using ENTOS QCORE.[26] In

this study, valence virtual orbitals are all localized by Intrinsic Bond Orbital method.[27] Valence occupied orbitals are localized by Boys–Foster localization for H_2O, and by Intrinsic Bond Orbital localizations[28,29] for all the rest of molecules, including eight small validation molecules, CH_5^+ and $C_2H_5^+$. Reference pair correlation energies are computed at the level of density-fitted CCSD(T)[30,31] with the aug-cc-pVTZ-JKFIT density fitting basis sets. All these correlation computations are performed with frozen core approximation and full iterative triples treatments using the same LMOs computed by ENTOS QCORE.

Gaussian process regressions (GPRs)[32] with white noise regularized Matérn 5/2 kernel were used to model the diagonal and offdiagonal pair energies separately using GPY 1.9.6 software package.[33] A subset of the generated set of structures for each molecule is used to make up the training set, and the remainder of the structures form the test set. The learning curves for H_2O, CH_5^+ and $C_2H_5^+$ are generated by moving structures from the test into the training set and testing using the structures that remain in the test set, and the MAEs that are evaluated for the test set as the training set is expanded provide the learning curves that are plotted in Figure 8. This procedure is modified slightly for the stretched model for $C_2H_5^+$. In this case, the training set was constructed to include 20% stretched structures. The remaining 80% of the structures being taken from the AIMD tranectory. These structures were also used to train the original $C_2H_5^+$ model. In this way, the final test set for the stretched model contains the 2000 structures used to train the original model as well as 500 stretched geometries.

Finally, we note that since the numbers of valid features are under 150 and will not cause overfitting due to the small sizes of these molecules, all the valid features are used in training without feature selection. The negative log marginal likelihood objective of GPR is optimized with respect to the kernel hyperparameters with a scaled conjugate gradient scheme for 100 steps and then apply the BFGS algorithm until full convergence.[21–23]

DMC Details

The guiding functions used in the guided DMC simulations are products of one-dimensional wave functions of the high frequency vibrations as described in our previous work.[6] The HOH bend is described by a harmonic oscillator with a frequency of 1668 cm^{-1} and a $G-$matrix element[34] of 2.338 amu^{-1} $Å^{-2}$. One-dimensional discrete variable representation (DVR) calculations were used to obtain the r_{OH} and r_{CH} wave functions.[35] The r_{OH} wave function was obtained via a potential scan along the r_{OH} coordinate with 900 r_{OH} bond lengths ranging from 0.26 Å to 1.59 Å. These potential values were then used as the potential function in the one-dimensional DVR calculation. A similar scan was done along the r_{CH} coordinate with 900 r_{CH} bond lengths ranging from 0.53 Å to 2.12 Å, keeping all other r_{CH} bond lengths and HCH angles constant.

All DMC simulations in this study were performed using a time step ($\Delta\tau$) of 1 a.u. The zero-point energies reported in Table 6 are calculated by averaging V_{ref} over the last two-thirds of the simulation time. For the guided **MOB-ML** DMC simulations, the minimum weight threshold was set at 0.01, and the maximum weight threshold is 1% of the ensemble size (e.g. 50 for a 5000 walker simulation). Each DMC simulation is run independently five times. The uncertainty of the reported zero-point energy is the standard deviation of these five simulations. In the H_2O **MOB-ML** guided DMC calculations, we propagated 2304 walkers for 10 000 a.u. and for CH_5^+ we propagated 5120 walkers for 5000 a.u. For the H_2O **NN+(MOB-ML)** unguided DMC simulations, we propagated 60 000 walkers for 50 000 a.u. and for CH_5^+, we propagated 60 000 walkers for 20 000 a.u. We ran analogous **PS** H_2O and **JBB** CH_5^+ calculations to compare energies and wave functions with the **NN+(MOB-ML)** unguided simulations. Finally, for $C_2H_5^+$ **NN+(MOB-ML)** unguided DMC simulations, we propagated 200 000 walkers for 50 000 a.u.

Training the NN+(MOB-ML) Potential Energy Surfaces

We used the Keras API implemented in the TensorFlow library[36] to construct, train, and evaluate the **NN+(MOB-ML)** surface. The neural network structure, hyperparameters and training procedure are identical to previous work.[8] To collect training data for the **NN+(MOB-ML)** surfaces,

we performed two unguided DMC calculations for each system using the **MOB-ML** surface. For one of the DMC simulations, we multiplied all of the masses of each of the atoms by 0.5, and for the other we use standard masses. We propagated 7168 walkers for each DMC simulation. For all simulations used to collect training data, starting at the second time step, we collected all of the walkers and energies every 5 time steps until time step 50. Then, we collected all walkers every 50 time steps. The resultant training data consisted of approximately 8.6×10^5 configurations and energies for H_2O and 1.5×10^6 configurations and energies for CH_5^+, since for H_2O we propagated the walkers for 2500 a.u. and for CH_5^+ we propagated the walkers for 5000 a.u. For $C_2H_5^+$, we ran four data-collecting DMC simulations with 7168 walkers for 2500 time steps. Of these four simulations, two were run with regular masses and the other two were run with the masses multiplied by 0.5. Additionally, for each set of masses, one simulation was run starting with all of the walkers in the minimum energy geometry of $C_2H_5^+$, while the other two simulations started with geometries that were randomly selected from a harmonic ground state wave function. This led to a generation of 2.5×10^6 configurations.

This procedure differs from previous work, where masses as small as one tenth the natural masses were used in the generation of the training data for the **NN-PES**. This change was needed because the Hartree-Fock calculations did not always converge when we started sampling based on these small mass simulations. Based on subsequent analysis, we found that training data collected from the DMC simulation, in which the mass is multiplied by 0.5 and 1, sufficiently covers the regions of the potential needed to develop the **NN-PES**.

From the configurations that are collected for H_2O and CH_5^+, 10 000 configurations are randomly selected for the validation set, while remaining configurations provide the training set. In the case of $C_2H_5^+$, the validation set contained 7.5×10^4 configurations. Training the model with only the training set while evaluating the errors based on both the training and the validation sets allows us to monitor for over-fitting of the model, and we terminate the training when the MAE for the validation set increases. An additional set of 10 000 configurations for H_2O and CH_5^+ and 4×10^5 configurations for $C_2H_5^+$ are randomly selected from only the simulations that were

run with natural masses. These configurations form the test set. The MAE for the test set provides a measure of the expected performance of the model in a production-run ground state DMC simulations. The MAEs of all three sets for each model are reported in Table S2.

Variational Calculation [reproduced from the Supporting Information of Ref. 8]

The calculations of the vibrational levels of water were performed in Jacobi coordinates. While these are not the most efficient coordinates for describing low-lying vibrational levels of water, they have the advantage of a simple kinetic energy operator,

$$\hat{H} = \frac{\hat{p}_r^2}{2\mu_r} + \frac{\hat{P}_R^2}{2\mu_R} + \left(\frac{1}{2\mu_R R^2} + \frac{1}{2\mu_r r^2} \right) \hat{j}^2 + V(R, r, \theta) \tag{S5}$$

where r represents one of the OH bond lengths, with reduced mass μ_r, R provides the distance between the second hydrogen atom and the center of mass of the OH bond described by r, and θ is the angle between $\vec{r}$ and $\vec{R}$. The reduced mass associated with R is

$$\mu_R = \left(\frac{1}{m_{\mathrm{H}}} + \frac{1}{m_{\mathrm{H}} + m_{\mathrm{O}}} \right) \tag{S6}$$

To start, three cuts through the potential were taken, one along each of the three coordinates with the other two coordinates set to their equilibrium values. Each cut was used in a 1D Discrete Variable Representation (DVR) calculation,[37] where a DVR based on the Hermite polynomials was used for R and r and the DVR in θ was based on Legendre Polynomials. For each DVR calculation, 250 DVR points were used. The resulting wave functions were used to obtain potential-optimized DVR points, with 35 in R and r and 30 in θ. These DVR points and the associated kinetic energy terms were used to set up the full Hamiltonian along with a potential cutoff of 35 000 cm^{-1}. With these parameters, we were able to converge the energies of the vibrational states of interest to within 1 cm^{-1}.

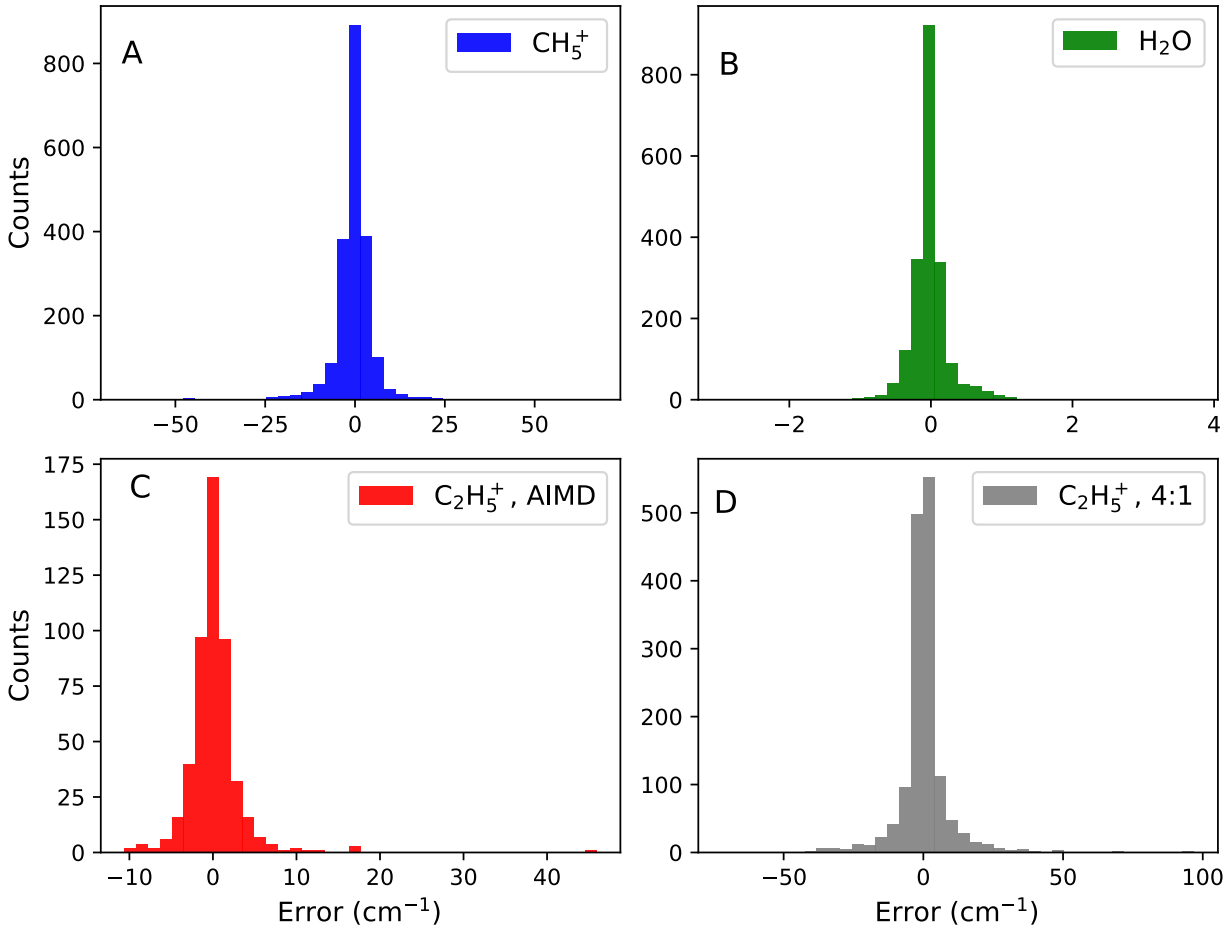

Figure S1: Histograms of the test errors for the final **MOB-ML** models for H_2O, CH_5^+, and $C_2H_5^+$. In the case of $C_2H_5^+$, we show the results for the two MOB-ML models. The AIMD model only included structures from the AIMD trajectory, and the test set consists of the 500 structures from that trajectory that were not included in the training set. For the 4:1 model, the test set consists of 1000 structures from the AIMD trajectory and 500 stretched structures. This is the same data that was used to obtain the mean absolute errors reported in Table S3 and Figure 8.

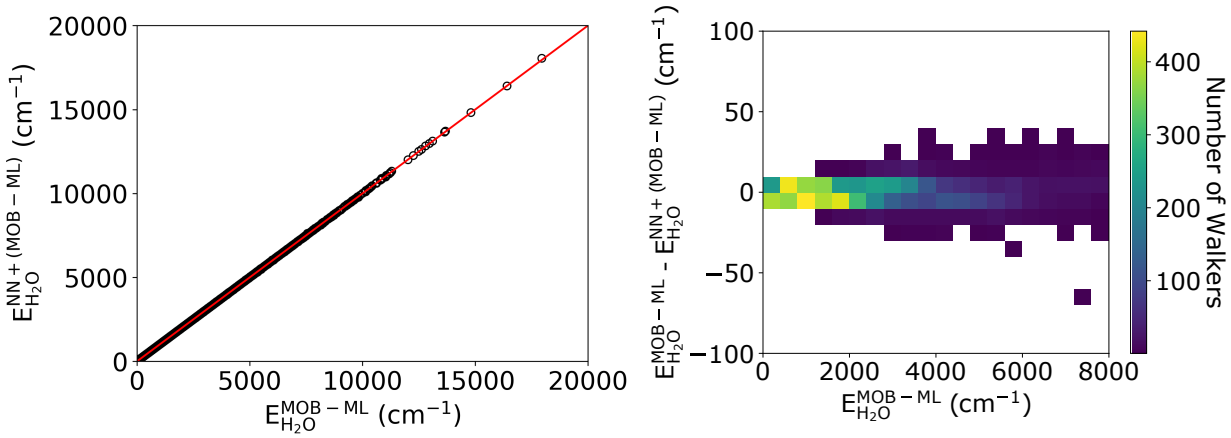

Figure S2: Comparison of the **NN+(MOB-ML)** and **MOB-ML** energies of the **NN+(MOB-ML)** ground state test data set for H_2O. This data is also used to calculate the ground state MAE in Table S2. The predicted **NN+(MOB-ML)** energy plotted as a function of the **MOB-ML** energy (left), and the number of geometries in the test set plotted as a function of the difference between the energies evaluated using the two surfaces and the **MOB-ML** energy (right).

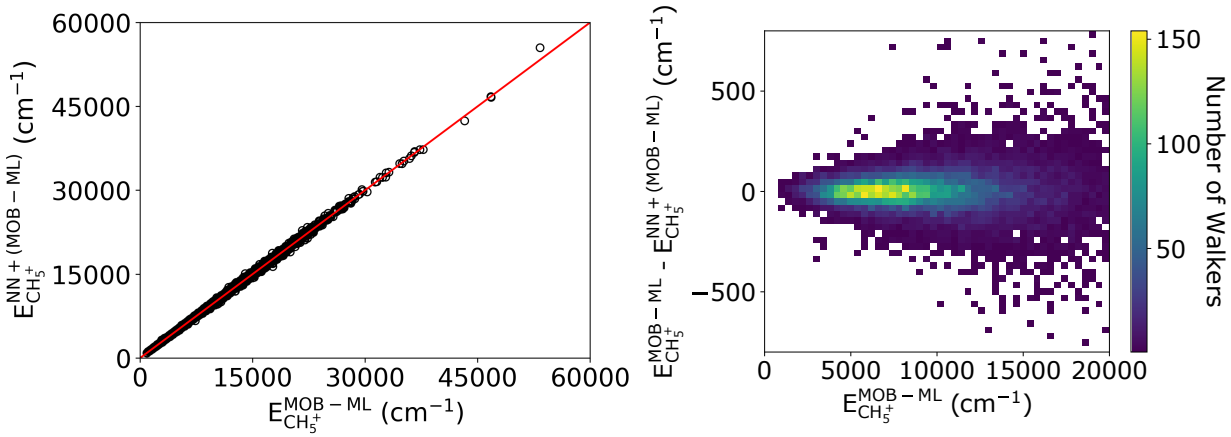

Figure S3: Comparison of the **NN+(MOB-ML)** and **MOB-ML** energies of the **NN+(MOB-ML)** ground state test data set for CH_5^+. This data is also used to calculate the ground state MAE in Table S2. The predicted **NN+(MOB-ML)** energy plotted as a function of the **MOB-ML** energy (left), and the number of geometries in the test set plotted as a function of the difference between the energies evaluated using the two surfaces and the **MOB-ML** energy (right).

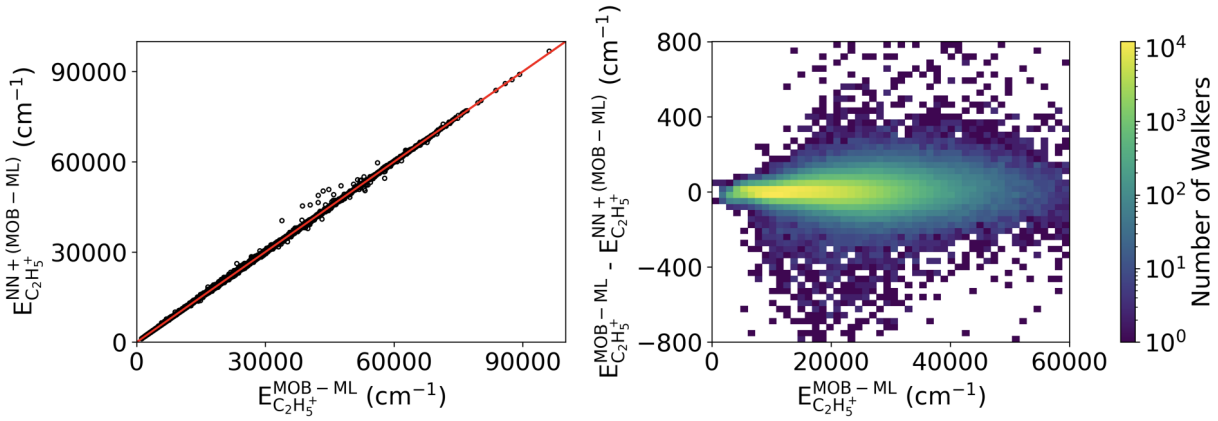

Figure S4: Comparison of the **NN+(MOB-ML)** and **MOB-ML** energies of the **NN+(MOB-ML)** ground state test data set for $C_2H_5^+$ for the MOB-ML model that was trained to structures from the 350 K AIMD trajectory and ones in which the CH distances were modified. This data is also used to calculate the ground state MAE in Table S2. The predicted **NN+(MOB-ML)** energy plotted as a function of the **MOB-ML** energy (left), and the number of geometries in the test set plotted as a function of the difference between the energies evaluated using the two surfaces and the **MOB-ML** energy (right).

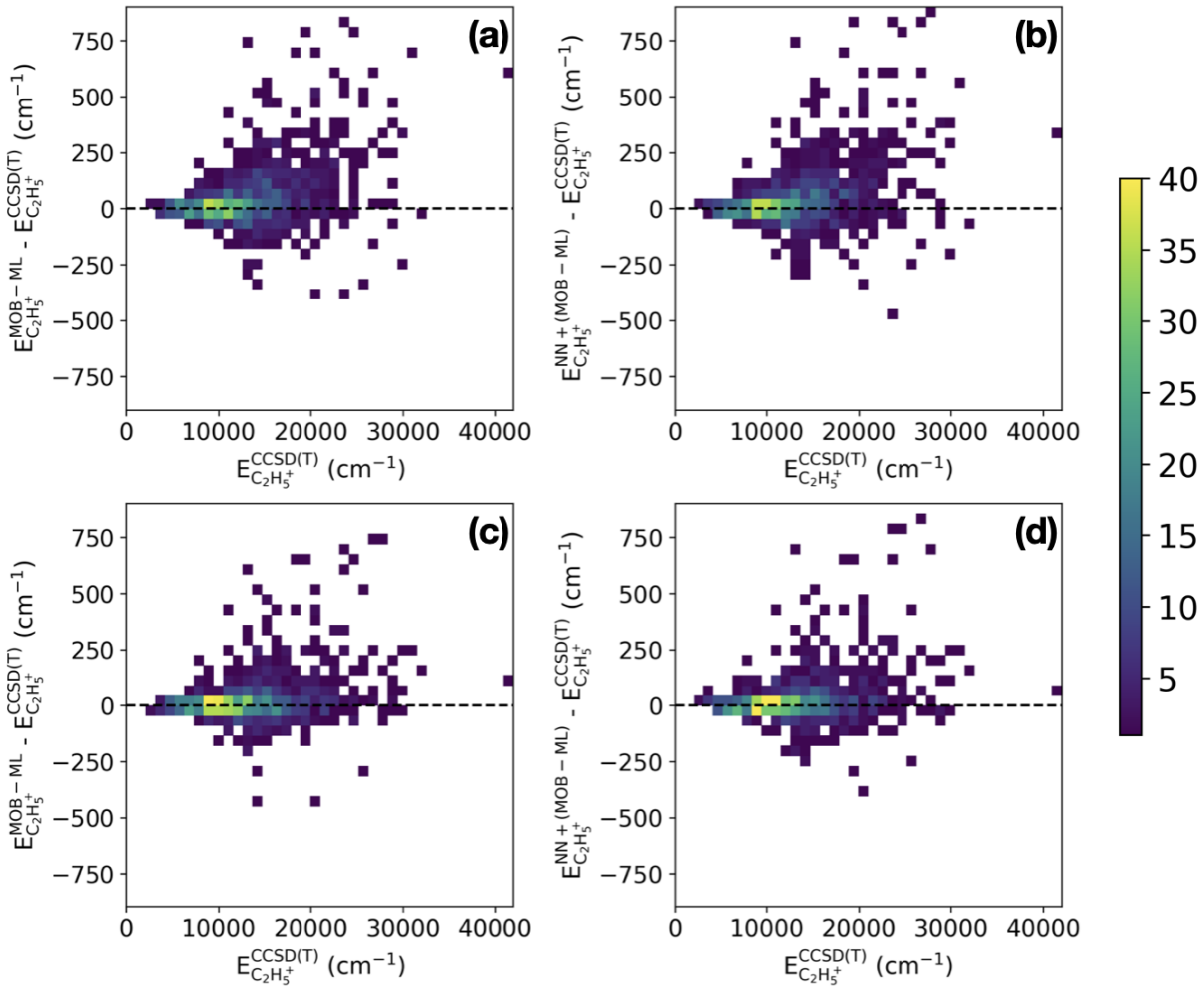

Figure S5: The comparison of the CCSD(T) energies and those obtained a) and c) based on the **MOB-ML** surface and b) and d) the **NN+(MOB-ML)** surface for energies 1000 structures that were extracted from a ground-state DMC simulation. The results obtained for the MOB-ML model that was trained to the structures from the 350 K AIMD trajectory are shown in panels a) and b), while panels c) and d) provides results obtained with the model that was trained to both AIMD and stretched structures.

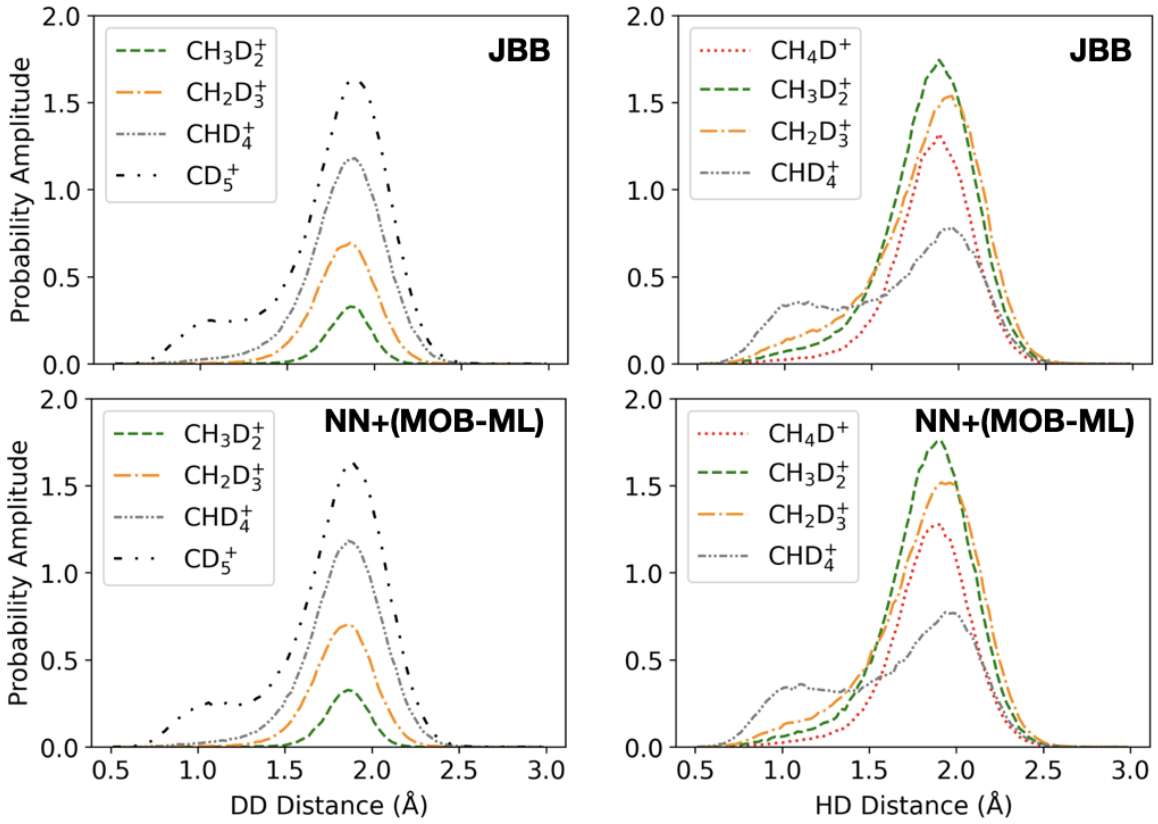

Figure S6: The DMC probability amplitude projected onto DD (left) and HD (right) distances using wave functions from the **JBB** surface (upper panels) and the **NN+(MOB-ML)** surface (lower panels).

Table S1: Harmonic Frequencies for H_2O from Underlying Electronic Structure Calculations (cm^{-1})

Mode	$\omega_{H_2O}^{MRCI,a}$	$\omega_{H_2O}^{CCSD(T)}$
1	1653.1	1646.0
2	3830.7	3810.7
3	3940.5	3919.8

[a] Ref. 38

Table S2: Mean Absolute Error of the NN+(MOB-ML) Training, Validation and Test Sets of H_2O, CH_5^+, and $C_2H_5^+$ (cm^{-1}).

System	Training Error	Validation Error Modified DMC	Test Error Ground State DMC
H_2O	18	24	4
CH_5^+	115	153	68
$C_2H_5^{+a}$	48	64	33
$C_2H_5^{+b}$	39	55	28

[a] Calculated based on the **MOB-ML** model that was trained with 2500 geometries from the 350 K AIMD simulation.

[b] Calculated based on the **MOB-ML** model that was trained with 2000 geometries from the 350 K AIMD simulation and 500 stretched geometries.

Table S3: Learning Curve Data, Plotted in Figure 8.

	MAE[a]			
n_{training}	H_2O	CH_5^+	$C_2H_5^{+b}$	$C_2H_5^{+c}$
50	33.5	13.1	27.0	43.0
100	20.1	8.0	15.5	26.6
200	6.8	4.0	10.1	19.1
300	5.8	3.2	7.6	15.2
500	1.8	2.0	5.2	10.6
800	1.1	1.6	3.9	7.6
1000	1.0	1.4	3.5	6.9
1500			2.7	5.7
2000			2.2	5.1
2500			1.9	5.1

[a] Mean absolute error in cm^{-1}.

[b] MOB-ML model for $C_2H_5^+$, which was trained using only structures from the AIMD trajectory.

[c] MOB-ML model for $C_2H_5^+$, which was trained using structures from the AIMD trajectory and stretched structures in a 4:1 ratio.

References

(1) Anderson, J. B. A Random-Walk Simulation of the Schrödinger Equation: H_3^+. *J. Chem. Phys.* **1975**, *63*, 1499–1503.

(2) Anderson, J. B. Quantum Chemistry by Random Walk. H ^2P, H_3^+ D_{3h} $^1A_1'$, H_2 $^3\Sigma_u^+$, H_4 $^1\Sigma_g^+$, Be ^1S. *J. Chem. Phys.* **1976**, *65*, 4121–4127.

(3) Suhm, M. A.; Watts, R. O. Quantum Monte Carlo Studies of Vibrational States in Molecules and Clusters. *Phys. Rep* **1991**, *204*, 293 – 329.

(4) 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.

(5) Lee, V. G. M.; McCoy, A. B. An Efficient Approach for Studies of Water Clusters Using Diffusion Monte Carlo. *J. Phys. Chem. A* **2019**, *123*, 8063–8070.

(6) Finney, J. M.; DiRisio, R. J.; McCoy, A. B. Guided Diffusion Monte Carlo: A Method for Studying Molecules and Ions that Display Large Amplitude Vibrational Motions. *J. Phys. Chem. A* **2020**, *124*, 9567–9577.

(7) DiRisio, R. J.; Finney, J. M.; McCoy, A. B. Diffusion Monte Carlo Approaches for Studying Nuclear Quantum Effects in Fluxional Molecules. *WIREs Computational Molecular Science* **2022**,

(8) DiRisio, R. J.; Lu, F.; McCoy, A. B. GPU-accelerated Neural Network Potential Energy Surfaces for Diffusion Monte Carlo. *J. Phys. Chem. A* **2021**, *125*, 5849–5859.

(9) McCoy, A. B.; Dzugan, L. C.; DiRisio, R. J.; Madison, L. R. Spectral Signatures of Proton Delocalization in $H^+(H_2O)_{n=1-4}$ Ions. *Faraday Discuss.* **2018**, *212*, 443–466.

(10) Barnett, R.; Reynolds, P.; W.A Lester, J. Monte Carlo Algorithms for Expectation Values of Coordinate Operators. *J. Comput. Phys.* **1991**, *96*, 258 – 276.

(11) DiRisio, R. J.; McCoy, A. B. rjdirisio/pyvibdmc:1.1.8. 2021; https://doi.org/10.5281/zenodo.4695231.

(12) Boyer, M. A.; DiRisio, R. J.; Finney, J. M.; McCoy, A. B. McCoyGroup/PyHPCDMC. 2021; https://doi.org/10.5281/zenodo.4739301.

(13) Shao, Y.; Gan, Z.; Epifanovsky, E.; Gilbert, A. T.; Wormit, M.; Kussmann, J.; Lange, A. W.; Behn, A.; Deng, J.; Feng, X. et al. Advances in Molecular Quantum Chemistry Contained in the Q-Chem 4 Program Package. *Mol. Phys.* **2015**, *113*, 184.

(14) Vosko, S. H.; Wilk, L.; Nusair, M. Accurate Spin-dependent Electron Liquid Correlation Energies for Local Spin Density Calculations: a Critical Analysis. *Can. J. Phys.* **1980**, *58*, 1200.

(15) 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.

(16) Becke, A. D. Density-functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648.

(17) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* **1994**, *98*, 11623.

(18) Hariharan, P. C.; Pople, J. A. The Influence of Polarization Functions on Molecular Orbital Hydrogenation Energies. *Theor. Chim. Acta* **1973**, *28*, 213.

(19) Bussi, G.; Parrinello, M. Accurate Sampling Using Langevin Dynamics. *Phys. Rev. E* **2007**, *75*, 056707.

(20) Johnson, L. M.; McCoy, A. B. Evolution of Structure in CH_5^+ and Its Deuterated Analogs. *J. Phys. Chem. A* **2006**, *110*, 8213–8220.

(21) Welborn, M.; Cheng, L.; Miller III, T. F. Transferability in Machine Learning for Electronic Structure via the Molecular Orbital Basis. *J. Chem. Theory Comput.* **2018**, *14*, 4772–4779.

(22) Cheng, L.; Welborn, M.; Christensen, A. S.; Miller III, T. F. A Universal Density Matrix Functional from Molecular Orbital-based Machine Learning: Transferability across Organic Molecules. *J. Chem. Phys.* **2019**, *150*, 131103.

(23) Husch, T.; Sun, J.; Cheng, L.; Lee, S. J.; Miller III, T. F. Improved Accuracy and Transferability of Molecular-orbital-based Machine Learning: Organics, Transition-metal Complexes, Non-covalent Interactions, and Transition States. *J. Chem. Phys.* **2021**, *154*, 064108.

(24) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. Electron Affinities of the First-row Atoms Revisited. Systematic Basis Sets and Wave Functions. *J. Chem. Phys.* **1992**, *96*, 6796–6806.

(25) Weigend, F. A Fully Direct RI-HF Algorithm: Implementation, Optimised Auxiliary Basis Sets, Demonstration of Accuracy and Efficiency. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4285–4291.

(26) Manby, F. R.; Miller III, T. F.; Bygrave, P.; Ding, F.; Dresselhaus, T.; Batista-Romero, F.; Buccheri, A.; Bungey, C.; Lee, S. J. R.; Meli, R. et al. entos: A Quantum Molecular Simulation Package. **2019**,

(27) Knizia, G. Intrinsic Atomic Orbitals: an Unbiased Bridge Between Quantum Theory and Chemical Concepts. *J. Chem. Theory Comput.* **2013**, *9*, 4834.

(28) Boys, S. F. Construction of Some Molecular Orbitals to be Approximately Invariant for Changes from One Molecule to Another. *Rev. Mod. Phys.* **1960**, *32*, 296–299.

(29) Foster, J. M.; Boys, S. F. Canonical Configurational Interaction Procedure. *Rev. Mod. Phys.* **1960**, *32*, 300–302.

(30) Bartlett, R. J.; Watts, J. D.; Kucharski, S. A.; Noga, J. Non-iterative Fifth-order Triple and

Quadruple Excitation Energy Corrections in Correlated Methods. *Chem. Phys. Lett.* **1990**, *165*, 513–522.

(31) Schütz, M. Low-order Scaling Local Electron Correlation Methods. III. Linear Scaling Local Perturbative Triples Correction (T). *J. Chem. Phys.* **2000**, *113*, 9986–10001.

(32) Rasmussen, C. E.; Williams, C. K. I. *Gaussian Processes for Machine Learning*; MIT Press: Cambridge, MA, 2006.

(33) GPy, GPy: A Gaussian Process Framework in Python. `http://github.com/SheffieldML/GPy`, since 2012.

(34) Wilson, E. B.; Decius, J. C.; Cross, P. C. *Molecular Vibrations*; Dover: New York, 1955.

(35) Colbert, D. T.; Miller, W. H. A Novel Discrete Variable Representation for Quantum Mechanical Reactive Scattering via the S-matrix Kohn Method. *J. Chem. Phys.* **1992**, *96*, 1982–1991.

(36) Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G. S.; Davis, A.; Dean, J.; Devin, M. et al. TensorFlow: Large-scale Machine Learning on Heterogeneous Systems. 2015; `https://www.tensorflow.org/`, Software available from tensorflow.org.

(37) Lill, J.; Parker, G.; Light, J. Discrete Variable Representations and Sudden Models in Quantum Scattering Theory. *Chem. Phys. Lett* **1982**, *89*, 483–489.

(38) 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*, 4618–4639.