GPU-Accelerated Neural Network Potential Energy Surfaces for Diffusion Monte Carlo

Ryan J. DiRisio, Fenris Lu, and Anne B. McCoy *

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

E-mail: abmccoy@uw.edu

Phone: 206-543-7464

June 7, 2021

Abstract

Diffusion Monte Carlo (DMC) provides a powerful method for understanding the vibrational landscape of molecules that are not well-described by conventional methods. The most computationally demanding step of these calculations is the evaluation of the potential energy. In this work, a general approach is developed in which a neural network potential energy surface is trained using data generated from a small-scale DMC calculation. Once trained, the neural network can be evaluated using highly-parallelizable calls to a graphics processing unit (GPU). The power of this approach is demonstrated for DMC simulations on H_2O, CH_5^+, and $(H_2O)_2$. The need to include permutation symmetry in the neural network potentials is explored and incorporated into the molecular descriptors of CH_5^+ and $(H_2O)_2$. It is shown that the zero-point energies and wave functions obtained using the neural network potentials are nearly identical to the results obtained when using the potential energy surfaces that were used to train the neural networks at a substantial savings in the computational requirements of the simulations. Use of the neural network potentials in other types of calculations is also discussed.

Introduction

Solving the vibrational Schrödinger equation is a cornerstone of theoretical vibrational spectroscopy. Methods such as harmonic normal mode analysis, vibrational self-consistent field, and vibrational perturbation theory are often sufficient to address many chemical problems.[1–3] However, molecules that have strong couplings, such as the astronomically-relevant CH_5^+ and molecular clusters that are held together by intermolecular interactions, require more sophisticated approaches due to their complicated potential energy landscapes.[4–6]

Often, one can obtain significant insights into the structure and spectroscopy of molecules and clusters that undergo large amplitude vibrational motions from the ground state wave function and how the wave function changes with isotopic substitution. One effective and general approach for obtaining the ground state solutions to vibrational problems is provided by diffusion Monte Carlo (DMC).[7,8] DMC is stochastic method that is used to solve for the ground state energy and vibrational wave function by propagating an ensemble of localized functions, referred to as walkers, in imaginary time. Often the most computationally demanding aspect of these calculations comes in the evaluation of the potential energy at the coordinates of each walker for each time step in the simulation. A DMC simulation based on 10 000 walkers that is propagated over 10 000 time steps requires roughly 100 million evaluations of the potential energy. Clearly, it is computationally intractable to perform this many calculations at a high level of theory of electronic structure theory for even the very smallest molecules. As such, DMC simulations often employ potential energy surfaces that have been fit to energies obtained from electronic structure calculations, the spectroscopy of the molecule, or both. Even when fitted potentials are used, the evaluations of these functions can become expensive as the system size is increased. For example, DMC studies of $(H_2O)_6$ required ensembles of 10^6 walkers run for 10^5 to 10^6 time steps to obtain converged results.[9,10]

Recently, we have demonstrated that we can achieve a significant reduction in the ensemble size required for a DMC simulation by introducing a guiding function that describes the

dependence of the wave function on the high-frequency intramolecular motions, and have applied the approach to neutral water clusters,[10,11] protonated water clusters, and CH_5^+.[12] This development has led to an order of magnitude speed-up in these calculations, which has opened the door to DMC studies of larger molecules and clusters. Even with these advances, the computational requirements for evaluating the potential will quickly limit the size of the systems that can be studied. Additionally, there are systems of interest for which high-quality potentials are not available, and it would be desirable to be able to perform DMC on-the-fly in such situations.

One solution to this problem is to take advantage of high performance computing resources and introduce multi-node and multi-core parallelism into the DMC code to perform potential calls at each time step. This can be accomplished using packages such as MPI and openMP. We have recently implemented such a strategy for the 10^6 walker calculation of the ground state of the water hexamer.[10] Even with these modifications, this calculation required a wall time of 2 days, and the calculations use 1088 cores spread among 17 KNL compute nodes. To obtain statistically meaningful results, the results of several (typically 5-10) independent DMC simulations are combined to obtain the final wave function and zero-point energy. This further increases the total computational requirements for a DMC study. Such computational demands severely limit the scope of studies. In particular, this high cost makes studies involving isotopic substitutions, which allow us to gain insights into nuclear quantum effects, intractable for all but the smallest molecular systems.

One strategy, which has been gaining increasing traction, is using machine-learning techniques to generate potential surfaces. In these studies, potential energy surfaces are developed from the results of a small number of electronic structure calculations using methods such as Gaussian Process Regression (GPR) and Artificial Neural Networks (NN). For example, Jiang and Guo used *ab initio* data to generate permutationally invariant neural network potential energy surfaces for reactive scattering calculations.[13] Bačič, Tuckermann, and coworkers used a neural network potential to perform enhanced sampling path integral

molecular dynamics on clathrate hydrates.[14] Additionally, Paesani and coworkers compared permutationally invariant polynomials, neural networks, and GPR in calculating many-body energies of water clusters.[15] Miller and co-workers used GPR to fit CCSD(T) quality correlation energies based on molecular orbitals obtained from a Hartree-Fock calculation,[16] and Bowman and Vargas-Hernandez have compared the accuracy and evaluation time of machine learning models using permutationally invariant polynomials and GPR for representing the potential energy surfaces of small molecules from *ab initio* data.[17] In many of these earlier studies, the goal was to develop a global or reactive potential surface with good accuracy across all relevant configurations.

A global potential energy surface is often not necessary. If the process of interest samples a specific region of the potential, describing that region well using neural networks will be sufficient. This is the case for DMC calculations, which focus on the vibrational ground state of the specific system of interest. Clearly, for this approach to be effective, we also need a way to determine the relevant region of the potential before training the neural network.

Based on these ideas, in this study we develop a generic algorithm that trains a neural network to learn the region of a potential energy surface for the system of interest over the range of configurations that are sampled in a DMC simulation. The geometries and corresponding energies that are collected as training data are obtained by running a small DMC simulation that uses the potential energy surface that we want to describe with the neural network. Once trained, this NN-potential can be used to perform large scale DMC simulations. This procedure is illustrated in Figure 1. To train the neural network, we use the Keras API implemented in the TensorFlow library.[18] Using TensorFlow makes training and evaluating the neural network model compatible with graphics processing units (GPUs). Evaluating the neural network using GPUs drastically reduces the computation time and resources required for the evaluation of the potential energy. To distinguish this approach from standard DMC methods, we will refer to it as NN-DMC in the following discussion. While the NN-DMC approach can be applied to any potential surface, in the present study we

will focus on potential energy surfaces that have been previously fit to electronic structure and where converged DMC simulations are achievable. This will allow us to explore the accuracy of the NN-DMC approach and benchmark the results against traditional DMC calculations.

In this study, we focus on three molecular systems, H_2O, $(H_2O)_2$, and CH_5^+. These were chosen because they represent examples of the types of systems that are often studied by DMC. In the case of H_2O, the three high-frequency vibrations make it surprisingly challenging to describe accurately using DMC, and the calculation of the ground state of H_2O requires smaller time steps to converge the results compared to CH_5^+.[9,10] Its small size also makes H_2O amenable to converged variational calculations of the vibrational energies, allowing us to explore the accuracy of the NN-potential for the evaluation of vibrational excited states. As we move to more fluctional molecules, one needs to consider the extent to which it is important to ensure full permutational symmetry in the potential. When full permutational symmetry is required, we must consider how to incorporate the permutational invariance in the NN-potential.

The water dimer presents a good example of a system that has sufficiently high barriers between equivalent minima on the potential such that consideration of full permutational symmetry of the potential will not be necessary for DMC studies that focus on the ground state. In principle, the potential for $(H_2O)_2$ should be invariant under exchange of the two water molecules and exchange of the two hydrogen atoms within either of these water molecules. This leads to eight equivalent minima on the potential. Operationally, a DMC ground state simulation that is initially localized in one of these minima does not sample all eight minima. Based on analysis of the ground state wave function, only the two hydrogen atoms shown in green in Figure 2 are found to exchange.[11] Such situations of reduced permutation symmetry sampled by the ground state wave function come up in a variety of molecular clusters, and $(H_2O)_2$ allows us to explore the efficacy of the NN-DMC approach for such situations.

We also consider CH_5^+. This is a molecular ion for which the barriers for exchange of any pair of hydrogen atoms are less than 350 cm^{-1}.[19] This means that the NN-potential must account for the full permutational symmetry of this ion. This is reflected by all five hydrogen atoms in the structure of CH_5^+ being displayed in green in Figure 2 . Upon partial deuteration, the sizes of the effective barriers that separate the minima increase. This leads to localization of the ground state of the ion in a subset of the minima,[20] and the zero point energies of the isotopologues of CH_5^+ will be sensitive to the the extent to which the ground state is localized. As such, studies of CH_5^+ and its deuterated analogues provide a stringent test of the NN-DMC approach.

In addition to providing representative model systems to explore the efficacy of the NN-DMC approach, potential surfaces have been developed for each of these three systems, and each of these systems have been previously sttudied using DMC.[11,19,20] For H_2O, we use the potential surface generated by Partridge and Schwenke (PS);[21] for CH_5^+ we use the surface generated by Jin, Braams, and Bowman (JBB);[19] and for $(H_2O)_2$ we use the generalized water potential MB-pol developed by Paesani and coworkers (MB-pol).[22–24] We will add NN to the beginning of each of these potentials (e.g. NN-PS) to refer to the NN-potential that was trained based on the indicated surface.

Neural Networks for Fitting Potential Energy Surfaces

Neural networks have been used to generate molecular potential energy surfaces for a variety of systems and in an assortment of contexts. Typically, the training data consists of a set of molecular configurations and the corresponding energies obtained using a specified level of electronic structure theory. For example, Carrington and coworkers generated a neural network potential energy surface for H_2CO using electronic energies evaluated at the CCSD(T) level of theory, which was then used to calculate vibrational energies.[25] Guo and coworkers developed a many-body potential based on CASSCF/CASPT2 data to perform scattering

calculations for N_4.[26] Behler and coworkers developed a scalable (2-8mer) protonated water cluster potential energy surface based on DFT energies and used it in *ab initio* molecular dynamics calculations.[27] As mentioned above, the NN-DMC approach can be used with a variety of sources for the potential energies. Since the focus of the present study is to demonstrate the efficacy of the approach, we will use the energies obtained using potential energy surfaces that were previously fit to electronic structure data. Once the training set is determined there are several other considerations that need to be addressed. These include the structure of the neural network and how it is trained as well as the choice of descriptor, and will be described in the following discussion.

Structure and Training of the Neural Network

The details of generating a potential energy surface using neural networks are described thoroughly by Behler, Manzhos and Carrington, and Jiang and Guo.[13,28,29] In the present study, we employ regression neural networks that will be used to learn the region of the potential surface that is relevant to DMC ground state simulations.

Hyperparameter optimization is an important, yet empirical, component of reducing the prediction error of a neural network. The selection of the number of hidden layers, number of nodes per layer, learning rate, activation function, and optimization algorithm each have a significant impact on the accuracy of the resulting neural network. Numerous studies have addressed on these choices for chemical problems. For example, Manzhos and Carrington have noted that one hidden layer of nodes is typically all that is necessary to approximate a potential energy surface when training with evenly-distributed, unbiased electronic structure training data,[29] while Juang and Guo and Bačič and coworkers have elected to employ deep neural networks (i.e. multiple hidden layers) to obtain potential energy surfaces.[13,30] We have elected to use a deep neural network in this work, since the training data obtained from a DMC simulation will be unevenly distributed and biased based on how the ground state wave function samples the potential surface.

Overfitting is generally a concern when training neural networks. Small numerical instabilities can occur when there is a bias towards the accuracy of the potential at geometries included in the training set when compared to geometries that are not included in the training set.[29] Due to the long wavelength of the vibrational ground state wave function, DMC calculations have been shown to be relatively insensitive to small numerical instabilities in the potential.[31] Nonetheless, we will take steps to avoid overfitting, for example by confirming that our training and test sets show similar errors in the calculated energies.

One other ingredient in generating the NN-potential is the choice of activation functions. Activation functions perform a nonlinear transformation of the input data at a given node. It is typical in the computational chemistry community to use sigmoid or hyperbolic tangent functions.[13,15,25,30] These activation functions are well-suited for single, and in some cases, multiple hidden layer networks, but will suffer from the vanishing gradient problem[32,33] in deep learning contexts. To minimize the effect of the vanishing gradient problem, we use the Swish activation function,

$$f(x) = \frac{x}{1 + e^{-x}} \tag{1}$$

for the nodes in the hidden layers. This function has empirically been shown to perform well in deep learning training.[34] Additionally, we use the the rectified linear unit function (RELU)[35]

$$f(x) = \begin{cases} x & x > 0 \\ 0 & x \leq 0 \end{cases} \tag{2}$$

in the output layer. This function forces the predicted energy of the NN-potential to be either positive or 0, even in regions of configuration space that are poorly described by the training data.

The mean squared error was chosen as the loss function, and the Nadam algorithm is used

to optimize the weights of the neural network.[36] Since neural networks have trouble learning and predicting a large range of energies, it is typical to transform and scale the training energies before using them to train the neural network.[29] To maintain the generality of our algorithm and to compress the energy range, we uniformly shift the energies so that the training data are based on a potential with its global minimum at 100 cm^{-1}. We then take the natural logarithm of the calculated energies and use the resulting quantities as the input for the neural network. The shift is introduced because the natural logarithm transformation spreads out the energies that are close to zero, increasing the likelihood of overfitting the low-energy region of the potential. With this combination of hyperparameters, we have found that using a neural network with three hidden layers and $10 \times (3N - 6)$ nodes per layer, where N is the number of atoms for the system of interest, yields the training and validation accuracy needed for the present application.

Choice of Descriptor

One of the challenges currently facing the chemistry machine learning community is how to best represent molecular coordinates as rotationally, translationally, and permutationally invariant vectors. In addition to this list, for the purposes of interfacing a neural network potential energy surface with DMC, we aim to find a generic and transferable descriptor whose transformation from the Cartesian coordinates of the atoms can be efficiently evaluated using a GPU. If the calculation of the descriptor takes longer than the original potential energy call, our primary goal of developing an efficient approach for evaluating the potential is not met. To this end, we considered using permutationally invariant polynomials of interatomic distances (PIPs), the Coulomb matrix (CM), and the Behler-Parinello neural network structure that uses atom-centered symmetry functions. Using PIPs is a promising option, however the number of polynomials required to adequately describe molecular systems scales poorly with system size. While there have been recent efforts to contract the number of terms required to describe larger symmetrical systems using fundamental invariants,[37,38] this work

has not been extended beyond 10 atoms. The Behler-Parinello neural network formalism is inherently permutationally, rotationally, and translationally invariant.[28] However, finding the sufficient number and types of symmetry functions to use is not obvious and requires empirical testing. Additionally, since all atom-atom distances as well as all atom-atom-atom angles must be calculated for the symmetry functions, this becomes computationally unfavorable.

The CM[39] provides an efficient and effective descriptor due to its generality, its rotational and translational invariance, and its low cost to evaluate. The elements of the CM are given by

$$\mathrm{CM}_{ij} = \begin{cases} 0.5 Z_i^{2.4} & i = j \\ \frac{Z_i Z_j}{r_{ij}} & i \neq j \end{cases} \tag{3}$$

where Z_i is the nuclear charge of the ith atom, and r_{ij} is the interatomic distance between atoms i and j. As such, evaluating the elements of the symmetric CM only requires the evaluation of the interatomic distances. The size of the final descriptor, which is the upper or lower triangle of the matrix, scales as N^2, where N is the number of atoms in the system of interest. To obtain permutation invariance from this descriptor, the rows and columns of the CM are reordered based on the norm of the values in the columns.[40] A limitation of using this sorted CM as a descriptor is that it may lead to discontinuities in the NN-potential.[41] In the discussion that follows, we will use the sorted CM as the descriptor for CH_5^+ to ensure permutational invariance for the NN-potential for this ion. We initially elected not to sort the CM for either H_2O or $(H_2O)_2$ and use the unsorted CM as the descriptor for these systems. We will return to this decision for $(H_2O)_2$ in our discussion of the DMC results. Finally, for efficiency, we implement the unsorted and sorted CM using CuPy, a software package used to execute numerical Python code on GPUs.[42]

Diffusion Monte Carlo

DMC has been described in detail elsewhere,[43,44] and we provide a brief overview of the relevant aspects of the algorithm below. As mentioned above, for DMC simulations, an ensemble of localized functions in 3N-dimensional configuration space, referred to as walkers, is propagated in imaginary time ($\tau = it/\hbar$). Once the average value of

$$V_{\mathrm{ref}}(\tau) = \bar{V}(\tau) - \alpha \frac{N_w(\tau) - N_w(\tau = 0)}{N_w(\tau = 0)} \tag{4}$$

has stabilized, we collect the value of V_{ref} at each time step and the coordinates of the walkers periodically throughout the simulation. The zero-point energy is evaluated by taking the time-averaged value of $V_{\mathrm{ref}}(\tau)$ in Eq. 4. In the expression for V_{ref}, $\bar{V}$ is the average potential energy of the ensemble, α is $0.5/\Delta\tau$,[45,46] and $N_w(\tau)$ is the number of walkers at imaginary time τ. Once equilibrated, the density of the walkers near the molecular configuration represented by the coordinates, $\mathbf{x}$, is proportional to the value of the ground state wave function at that geometry.

At each imaginary time step in the DMC propagation, each of the coordinates of each of the walkers is randomly displaced based on a Gaussian distribution, the width of which depends on the size of the time step ($\Delta\tau$) and the mass of the atom that is being displaced.[7] The potential energies associated with the coordinates of each of the walkers are evaluated after this displacement. Based on the potential energy at the new configuration, the walker may be replicated or removed from the simulation. Typically, ensemble sizes on the order of 10 000 to 100 000 are required to achieve adequate sampling of the ground state wave function for molecules with approximately 2-10 atoms, and these ensembles must be propagated on the order of 10 000 to 100 000 time steps.[12] All DMC calculations were performed using PyVibDMC,[47] an open source, general purpose DMC code developed by our group.

Results and Discussion

Collection of Training Data

In developing a strategy for collecting training data, our goal was to construct a general approach that is based on the collection of wave functions that are obtained from a DMC simulation without any additional pruning or enhancement of this data. In order for this to be successful, we must ensure that the geometries included in the training data fully sample the regions of the potential that are accessed by the ground state wave function in a DMC simulation. The red curve in Figure 3 illustrates the distributions of the energies of walkers obtained from a ground state DMC simulation for CH_5^+. As can be seen, the highest energies that are sampled exceed twice the value of the zero-point energy (dark blue line) of 10 918 cm^{-1}.[12] Additional plots showing the distribution of walker energies obtained from DMC simulations for H_2O and $(H_2O)_2$ are provided in Figure S1 in the Supporting Information.

Above the zero-point energy, the number of walkers with a specified energy decreases with energy. For an $M-$dimensional isotropic harmonic oscillator, the amount of configuration space that corresponds to a particular energy will be the surface area of a $M-$dimensional hypersphere with a radius that is proportional to $\sqrt{E}$. Based on this, the amount of configuration space that corresponds to an energy of E will scale as roughly $E^{(M-1)/2}$. This means that the density of sampled configurations in the DMC ensemble falls off rapidly as a function of E. If we use only the configurations that are captured by the distribution plotted in red in Figure 3, we will not have sufficient training data at energies above the zero-point energy for the neural network to provide an accurate description of the potential energy surface. When we performed DMC on a neural network trained only using data obtained from ground state sampling, we found that the NN-potential had large numerical instabilities in the surface, or holes, which caused DMC simulations based on this surface to fail.

One way to ensure that geometries included in the training set sample these higher energy regions of configuration space is to run DMC simulations in which the masses of the atoms

have been decreased. For example, the light green distribution in Figure 3, which is labeled as 0.5, was obtained by performing a DMC simulation in which the masses of all of the atoms were multiplied by 0.5. This reduction of the atomic mass increases the zero-point energy of the system and expands the region of configuration space that the walkers can access. It also leads to poorer sampling at low-energy regions of the potential.

Based on these observations, we have developed a hybrid approach. The training data is obtained from a DMC simulation in which we propagate a set of walkers with their most abundant masses for the first half of the simulation. During the the second half of the simulation, we decrease the masses of all of the atoms until their masses have been reduced by a factor of ten. This is achieved by multiplying the masses by a constant factor at each time step. In the case of this study, the masses are decreased over the course of 4000 time steps. To decrease the masses of the atoms by a factor of 10 over this time period, at each time step the masses of the atoms are each multiplied by $\sqrt[4000]{0.1}$. The coordinates of the walkers along with their energies are collected periodically throughout this simulation, and this forms the training data for the NN-potential. The resulting energy distribution is shown as the shaded distribution in Figure 3. The shaded distribution follows the red curve at low energies while also providing a larger number of geometries at higher energies.

For all systems, we collect training data from a DMC simulation with 20 000 walkers, which is propagated for 8000 time steps with $\Delta\tau = 1$ a.u. All of the walkers are placed at the minimum in the potential at $\tau = 0$ a.u. The training data consists of the ensembles of walkers and their energies, which are collected every 100 time steps throughout the simulation. This results in a training set containing roughly 1.6×10^6 walkers for each system that is considered. By starting all of the walkers at the potential minimum and, in the case of CH_5^+ and $(H_2O)_2$, collecting the walkers after one time step, we ensure that the region near the potential minimum is well-sampled. This additional step was not required for H_2O, as it lacks the low frequency vibrations and multiple minima exhibited by the other systems. More details on the training procedure are provided in the Supporting Information.

Validation of Neural Network

A validation set was also collected from the training DMC simulation, where 100 000 randomly chosen walkers are collected using snapshots of the wave function between time step 3000 and 4000, which were not used in the construction of the training set. As such, the validation set only contains walkers that have unscaled masses and reflect geometries sampled by a ground state DMC simulation. The mean absolute error (mae) of the validation set was used as a convergence metric throughout the training process. To generate a test data set that consists of walkers and energies that were not used in the training process, we performed an independent DMC simulation using the same procedure as was used to obtain the training set. We obtained the test set by randomly selecting 80 000 of the generated molecular configurations from the wave functions obtained from this second DMC simulation.

After training the neural networks for the three systems of interest, we calculated the error associated with each model. The results are reported in Table 1. We report the mae for each neural network based on the training, validation, and test sets. Since the validation set includes configurations from a typical ground state DMC simulation, the energies of the sampled configurations are expected to be smaller on average than the energies of the configurations that make up the training and test sets. In fact, for all three systems, the validation error is roughly half of the training error. The difference between the training and test error is less than 1 cm^{-1} for H_2O and CH_5^+ and smaller than 4 cm^{-1} for $(H_2O)_2$. This small difference between the training and test errors gives us confidence that the model predicts energies of configurations that are not included in the training set with similar accuracy to configurations that are in the training set.

To further explore how the error is distributed, in Figure 4 we plot the density of configurations in the validation set for H_2O as a function of the energy evaluated using the PS potential (E^{PS}) and the difference between values of the energy obtained using the PS potential and the NN-PS potential, $\Delta E = E^{\mathrm{PS}} - E^{\mathrm{NN-PS}}$. Overlaid on this plot, we show the average and standard deviation of ΔE for E^{PS} ranges of 750 cm^{-1}, shown with black dots

and error bars, respectively, and a white line at $\Delta E = 0$. As is seen, the average value of ΔE remains close to zero up to 15 000 cm^{-1}, which is nearly four times the zero-point energy for water. The standard deviation of ΔE increases with E^{PS}. In the bottom panel, we plot the correlation between $E^{\text{NN-PS}}$ and E^{PS} over the full energy range sampled by the validation set. The dashed line shows $E^{\text{NN-PS}} = E^{\text{PS}}$. As is seen, the data lies close to the dashed line. Taken together, these results show that while there are small, and growing, deviations between $E^{\text{NN-PS}}$ and E^{PS}, the error is centered at 0 and uniformly distributed about that value (see also Figure S2, where we provide one-dimensional cuts through the distribution shown in the top panel of Figure 4). The analogous figures for CH_5^+ and $(H_2O)_2$ can be found in Figures S3 and S4 in the Supporting Information. Consistent with the reported errors in Table 1, the distributions shown in Figures S3 and S4 are similar to those for H_2O, albeit with slightly larger ranges of ΔE. The ensemble that is propagated in DMC randomly samples the potential, and the predicted value for the ground state energy is based on an ensemble average of the potential energy, $\bar{V}(\tau)$. Therefore, the uniformly distributed errors descrribe above should lead to errors in the ground state energy that are much smaller than the mae values reported in Table 1.

Diffusion Monte Carlo Performance on the Neural Network Surfaces

The goal of this work is to develop a NN-based approach that allows for the efficient and accurate determination of the potential energies for a DMC calculation. Therefore, the most important test of the success of this approach is through comparisons of the energies and wave functions obtained using the NN-potential and the potential on which the NN-potential is based.

In Table 2, we compare the zero-point energies for H_2O, CH_5^+, and $(H_2O)_2$ obtained using the NN-potentials with a small and large ensemble of walkers along with previously reported results obtained using the PS, JBB, and MB-pol potentials, respectively. The smaller NN-

DMC calculations were performed using the same ensemble size and propagation time as were used to obtain previously reported results on these systems. We performed the larger calculations to confirm that the results are converged. As can be seen in the results reported in Table 2, the energies obtained from the NN-DMC calculations are in excellent agreement with previously reported results for all three systems, with differences for these three systems of 1 cm^{-1} or less. As anticipated, these differences are much smaller than the mae of the energies calculated using the NN-potentials reported in Table 1. In addition, the differences between the zero-point energies obtained from DMC and NN-DMC simulations are generally much smaller than the reported uncertainties of the DMC simulations.

While validation through comparisons of zero-point energies is important, the power of DMC comes from the ability to obtain the ground state vibrational wave function that describes the structure of the system of interest. To this end, we compare the wave functions generated from DMC using the NN-potentials and the potentials on which the NN-potential was based. We first consider the system that is most difficult to describe quantum mechanically, CH_5^+. As discussed above, CH_5^+ has 120 equivalent minima, and the energetic barriers to isomerize between these minima are lower than the zero-point energy in the associated vibrations. This leads to full permutation equivalence of the hydrogen atoms in the vibrational ground state. Additionally, when one performs isotopic substitution, the wave function becomes localized in a subset of these minima. This phenomenon has been described in previous work.[20,48] This partial localization of the ground state wave function makes the calculation of the ground state energies and wave functions for partially deuterated CH_5^+ a particularly stringent test of the NN-potential, as slight deviations in potential can impact the localization of the isotopic variants of this ion.

We first examine the zero-point energies associated with each of the isotopologues, reported Table 2. There is good agreement between the previously reported values based on DMC calculations that used the JBB surface, the NN-DMC calculation, the large-scale NN-DMC calculation. To compare the ground state probability amplitude obtained from DMC

calculations using the NN-JBB surface to those obtained using the JBB surface, in Figure 5 we project the Ψ_0^2 obtained from these two calculations onto the HH distances in each of the isotopologues of CH_5^+ that contains two or more hydrogen atoms. While there are subtle differences between the distributions obtained when the DMC calculations were performed using the JBB and NN-JBB potentials, these differences are generally smaller than the uncertainty in the values as indicated by the error bars. The corresponding projections onto the HD and DD distances show similar trends (see Figure S6 in the Supporting Information). As noted above, one concern about using the sorted CM as the descriptor in generating the NN-potential comes in the possibility of discontinuities. In fact, if we plot cuts through the NN-JBB surface along each of the four unique CH bond lengths in CH_5^+, constraining all other coordinates to their equilibrium values, we note such discontinuities (see Figure S7). While this is a possible limitation of the use of the sorted CM in thee NN-DMC approach, it does not appear to be affecting the accuracy of the zero-point energies or the probability amplitude obtained when we perform a DMC simulation using the NN-JBB potential.

As discussed above, the water dimer is a system that has eight equivalent minima on the potential surface. However, when the DMC simulation is initialized in one of the eight minima, only two of these minima are sampled by the ground state wave function. The water dimer also exhibits highly anharmonic, low frequency vibrational motions in the intermolecular degrees of freedom. To examine how these attributes of $(H_2O)_2$ are captured by the NN-potential, we project the probability amplitude onto pairs of OH distances in Figure 6. In the top panel, we show the projection of Ψ_0^2 onto both of the intramolecular OH distances in the donor the water molecule (the two OH bonds of the water molecule with red, orange, and yellow atoms in Figure 2). There is excellent agreement between the wave functions collected from DMC simulations using the MB-pol surface and the NN-MB-pol surfaces. In the bottom panel of Figure 6, we show the projection of Ψ_0^2 onto the intermolecular OH distances between the oxygen in the donor water and the two hydrogen atoms in the acceptor water (the red oxygen atom and the two green hydrogen atoms in the structure in Figure

2). We find generally good agreement between these projections, which are plotted in black and red, although there is a small peak around 2 Å that is only partially captured in the DMC calculation that is based on the NN-MB-pol potential. If we train a new NN-potential using the sorted CM as the molecular descriptor and perform a DMC calculation, we recover this feature (gold dotted line). The mae for this new NN-potential is reported in Table 1, the DMC zero-point energies are reported in Table 2, and the distributions of the errors are provided in Figures S2 and S5. The training, test, and validation mae of the sorted CM NN-MB-pol surface are all comparable to the mae of the CH_5^+ model, which contains the same number of atoms and also uses the sorted CM as the molecular descriptor. Additionally, the zero-point energy calculated using the sorted CM NN-potential is in agreement with the NN-DMC simulations that used the unsorted CM NN-potential for $(H_2O)_2$.

Upon closer inspection of the walkers that correspond to the peak near 2 Å in the projections of the ground state probability amplitude plotted in the bottom panel of Figure 6, we find that they are structurally similar to the hydrogen bond interchange transition state,[49,50] which is shown in the inset in the bottom panel of Figure 6. The energy of this transition state has been reported as approximately 235 cm^{-1} based on the MB-pol surface,[49] while the harmonic normal mode frequency that corresponds to the motion that connects the minimum energy structure and this transition state is below 200 cm^{-1}. While the ground state DMC calculation may sample geometries that are near this transition state, the walkers are not able to tunnel through this barrier. When we examined the structures that are sampled by the DMC simulation used to generate the training and test sets, we found they do not show evidence of tunneling through this barrier. The enforced symmetry introduced by the sorted CM provides an improved description of this region of the potential and more accurate ground state wave function. Further evidence of this improvement can be seen in the projections of Ψ_0^2 onto the intermolecular HH distances shown in Figure S8 in the Supporting Information. Finally, we have considered the projections of the probability amplitude onto the OH distance and HOH bend in H_2O, and the results are provided in Figure S9 in the

Supporting Information. The projections obtained using the NN-PS surface show excellent agreement with those obtained using the PS surface.

Efficiency Gain from Neural Network Potentials

Next we consider the relative efficiency of the NN-potentials when compared to the potential energy surfaces on which the NN-potentials are based. To explore this, we calculate the time it takes to evaluate the energies using the MB-pol potential and the GPU accelerated NN-MB-pol potential for $(H_2O)_2$. The analogous timing plots for H_2O and CH_5^+, as well as detailed timings for all the systems, are found in Figure S10 and Table S1 in the Supporting Information. We call the compiled MB-pol surface within Python using the ctypes foreign function interface, and we parallelize the calls to MB-pol across CPU cores using Python's multiprocessing module. In Figure 7, we compare the average time required to evaluate the potential energy over 400 time steps in a DMC simulation for a series of different ensemble sizes. Continuous weighting DMC[43] is used in these calculations in order to keep the ensemble size constant throughout the 400 time steps. The DMC simulations using the MB-pol potential are performed using a single 28-core Intel Xeon E5-2680 v4 2.40 GHz processor. We perform the analogous calculation on the NN-potential using one of the GPU nodes on the Cori supercomputer at the National Energy Research Scientific Computation Center (NERSC). These GPU nodes have 8 NVIDIA Tesla V100 GPUs per node and two 20-core 2.40 GHz Intel Xeon Gold 6148 Skylake processors. For these NN-DMC calculations, only one of the 8 GPU nodes was used.

The results plotted in Figure 7 illustrate that the GPU-optimized NN-MB-pol evaluations scale significantly better than the calls to the underlying MB-pol potential. The improvement is shown most prominently in the 10^6 walker case, where the GPU calls of thee NN-potential are 3x faster for H_2O, 4x faster for CH_5^+, and 10x faster for $(H_2O)_2$ compared to the CPU calls to the potential on which the NN-potential is trained. In some senses this comparison is not entirely fair as similar speed-ups could be obtained by using a version of the underlying

potentials that has been modified to take advantage of GPU's. This is not always a straightforward procedure, and advantage of the present approach for obtaining the NN-potentials is that it can be applied to a broad range of potential surfaces without modification.

With these gains in speed we can perform extremely large NN-DMC simulations at a cost that is independent of the cost to evaluate the underlying potential. As such, the use of these NN-potentials will enable us to perform DMC calculations that use significantly more expensive approaches for evaluation of the potential energy. For example, we are currently exploring applications in which the energies are evaluated using electronic structure calculations directly.

Variational Calculation of Vibrational States Using the NN-Potential

The discussion above focuses on ground state properties, and the geometries that were used to develop the NN potential for the NN-DMC calculation were chosen for the calculation of the ground state. It is interesting to ask how well this potential describes excited states. In the case of water, the excited state vibrational energies are reasonably straightforward to evaluate, and the details of the calculation are provided in the Supporting Information. The results of these calculations are provided in Table 3. Comparing the results obtained using the PS and NN-PS potentials, we find there are small differences in energies. For example, the zero-point energies differ by 0.1 cm^{-1}, while the energies of all other vibrational states differ by less than 1 cm^{-1}. The calculated zero-point energy is also in excellent agreement with the DMC values reported in Table 2. This shows that the NN-potential obtained for DMC calculations can also be used in other types of large scale calculations that require many calls to a potential energy surface.

Conclusions

In this work, we have developed a generic algorithm to train a neural network to learn a potential energy surface using data obtained from a single, small-scale DMC simulation. After training this neural network, we are able to perform large-scale DMC simulations by taking advantage of the efficiency of parallel evaluations of the NN-potential on a GPU. These NN-potentials are system-dependent, meaning that one has to train a new neural network for each system of interest and for different isomers of a particular molecular cluster. These surfaces focus on describing low energy vibrational states, although modification of the training protocols could extend the region covered by the neural network. There are other methods, such as path integral ground state (PIGS) methods[51,52] and Path Integral Monte Carlo[53,54] that require dozens to hundreds of potential energy evaluations per step. These methods can also take advantage of parallel evaluation of these potential energy values on a GPU.

Acknowledgement

Support from the Chemistry Division of the National Science Foundation (CHE-1856125) is gratefully acknowledged. Parts of this work were performed using the Ilahie cluster, which was purchased using funds from a MRI grant from the National Science Foundation (CHE-1624430). This work was also facilitated through the use of advanced computational, storage, and networking infrastructure provided by the Hyak supercomputer system and funded by the STF at the University of Washington as well as resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-05CH11231. RJD was supported by a fellowship from The Molecular Sciences Software Institute under NSF grant OAC-1547580.

Supporting Information Available

Numerical details; number of walkers as a function of energy for H_2O and $(H_2O)_2$; one-dimensional histograms of the number of walkers as a function of the energy difference between the reference surface and the NN-potential; heat maps and correlation plots of the JBB and NN-JBB potentials for CH_5^+ and MB-pol and NN-MBpol $(H_2O)_2$ when using the unsorted CM and the sorted CM as the molecular descriptor; comparisons of the projections of the DMC probability amplitude onto the HD and DD distances of the isotologues of CH_5^+ using the NN-JBB and JBB potentials; one-dimensional cuts through the NN-JBB potential energy surface along the four unique CH stretches in CH_5^+; comparisons of the projections of the DMC probability amplitude onto HH distances in $(H_2O)_2$ for MB-pol, the unsorted CM NN-MB-pol, and the sorted CM NN-MB-pol; comparisons of the projections of the DMC probability amplitude onto the OH stretch and HOH bend in H_2O for the NN-PS and PS potential; potential energy call time for MB-pol and NN-MB-pol for $(H_2O)_2$ and JBB and NN-JBB for CH_5^+; table of potential energy call times for the NN-potentials and underlying potentials for all three systems; TensorFlow Model files for the H_2O, CH_5^+, and $(H_2O)_2$ (both using the sorted and unsorted CM), neural network potential energy surfaces (zip).

References

(1) Dunn, M. E.; Evans, T. M.; Kirschner, K. N.; Shields, G. C. Prediction of Accurate Anharmonic Experimental Vibrational Frequencies for Water Clusters, $(H_2O)_n$, n = 2-5. *J. Phys. Chem. A* **2006**, *110*, 303–309.

(2) Monteiro, J. G.; Barbosa, A. G. VSCF Calculations for the Intra- and Intermolecular Vibrational Modes of the Water Dimer and its Isotopologs. *Chem. Phys.* **2016**, *479*, 81–90.

(3) Watanabe, Y.; Maeda, S.; Ohno, K. Intramolecular Vibrational Frequencies of Water Clusters $(H_2O)_n$ (n=2–5): Anharmonic Analyses Using Potential Functions Based on the Scaled Hypersphere Search Method. *J. Chem. Phys.* **2008**, *129*, 074315.

(4) Wang, X.-G.; Tucker Carrington, J. Vibrational energy levels of CH_5^+. *J. Chem. Phys.* **2008**, *129*, 234102.

(5) Leforestier, C.; Gatti, F.; Fellers, R. S.; Saykally, R. J. Determination of a Flexible (12D) Water Dimer Potential Via Direct Inversion of Spectroscopic Data. *J. Chem. Phys.* **2002**, *117*, 8710–8722.

(6) Leforestier, C.; Szalewicz, K.; van der Avoird, A. Spectra of Water Dimer From a New ab initio Potential with Flexible Monomers. *J. Chem. Phys.* **2012**, *137*, 014305.

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

(8) 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.

(9) Mallory, J. D.; Brown, S. E.; Mandelshtam, V. A. Assessing the Performance of the Diffusion Monte Carlo Method As Applied to the Water Monomer, Dimer, and Hexamer. *J. Phys. Chem. A* **2015**, *119*, 6504–6515.

(10) Lee, V. G. M.; Vetterli, N. J.; Boyer, M. A.; McCoy, A. B. Diffusion Monte Carlo Studies on the Detection of Structural Changes in the Water Hexamer upon Isotopic Substitution. *J. Phys. Chem. A* **2020**, *124*, 6903–6912.

(11) 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.

(12) 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.

(13) Jiang, B.; Guo, H. Permutation Invariant Polynomial Neural Network Approach to Fitting Potential Energy Surfaces. *J. Chem. Phys.* **2013**, *139*, 54112.

(14) Cendagorta, J. R.; Shen, H.; Bačić, Z.; Tuckerman, M. E. Enhanced Sampling Path Integral Methods Using Neural Network Potential Energy Surfaces with Application to Diffusion in Hydrogen Hydrates. *Adv. Theory Simul.* **2021**, *4*, 2000258.

(15) Nguyen, T. T.; Székely, E.; Imbalzano, G.; Behler, J.; Csányi, G.; Ceriotti, M.; Götz, A. W.; Paesani, F. Comparison of Permutationally Invariant Polynomials, Neural Networks, and Gaussian Approximation Potentials in Representing Water Interactions Through Many-Body Expansions. *J. Chem. Phys.* **2018**, *148*, 241725.

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

(17) Qu, C.; Yu, Q.; Van Hoozen, B. L.; Bowman, J. M.; Vargas-Hernández, R. A. Assessing Gaussian Process Regression and Permutationally Invariant Polynomial Approaches To Represent High-Dimensional Potential Energy Surfaces. *J. Chem. Theory. Comput.* **2018**, *14*, 3381–3396.

(18) 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.

(19) Jin, Z.; Braams, B. J.; Bowman, J. M. An ab Initio Based Global Potential Energy Surface Describing $CH_5^+ \rightarrow CH_3^+ + H_2$. *J. Phys. Chem. A* **2006**, *110*, 1569–1574.

(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) 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.

(22) Babin, V.; Medders, G. R.; Paesani, F. Toward a Universal Water Model: First Principles Simulations from the Dimer to the Liquid Phase. *J. Phys. Chem. Lett.* **2012**, *3*, 3765–3769.

(23) Babin, V.; Medders, G. R.; Paesani, F. Development of a "First Principles" Water Potential with Flexible Monomers. II: Trimer Potential Energy Surface, Third Virial Coefficient, and Small Clusters. *J. Chem. Theory. Comput.* **2014**, *10*, 1599–1607.

(24) Paesani, F. Getting the Right Answers for the Right Reasons: Toward Predictive Molecular Simulations of Water with Many-Body Potential Energy Functions. *Acc. Chem. Res.* **2016**, *49*, 1844–1851.

(25) Manzhos, S.; Wang, X.; Dawes, R.; Carrington, T. A Nested Molecule-Independent Neural Network Approach for High-Quality Potential Fits. *J. Phys. Chem. A* **2006**, *110*, 5295–5304.

(26) Li, J.; Varga, Z.; Truhlar, D. G.; Guo, H. Many-Body Permutationally Invariant Polynomial Neural Network Potential Energy Surface for N4. *J. Chem. Theory. Comput.* **2020**, *16*, 4822–4832.

(27) Kondati Natarajan, S.; Morawietz, T.; Behler, J. Representing the Potential-Energy Surface of Protonated Water Clusters by High-Dimensional Neural Network Potentials. *Phys. Chem. Chem. Phys.* **2015**, *17*, 8356–8371.

(28) Behler, J. Atom-Centered Symmetry Functions for Constructing High-Dimensional Neural Network Potentials. *J. Chem. Phys.* **2011**, *134*, 74106.

(29) Manzhos, S.; Carrington, T. Neural Network Potential Energy Surfaces for Small Molecules and Reactions. *Chem. Rev.* **2020**, in press, https://doi.org/10.1021/acs.chemrev.0c00665.

(30) Liu, Y.; Li, J.; Felker, P. M.; Bačič, Z. HCl-H_2O Dimer: An Accurate Full-Dimensional Potential Energy Surface and Fully Coupled Quantum Calculations of Intra- and Intermolecular Vibrational States and Frequency Shifts. *Phys. Chem. Chem. Phys.* **2021**, *23*, 7101–7114.

(31) Crittenden, D. L.; Jordan, M. J. T. Interpolated Potential Energy Surfaces: How Accurate do the Second Derivatives Have to be? *J. Chem. Phys.* **2005**, *122*, 044102.

(32) Glorot, X.; Bordes, A.; Bengio, Y. Deep Sparse Rectifier Neural Networks. Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics. Fort Lauderdale, FL, USA, 2011; pp 315–323.

(33) Xie, X.; Persson, K. A.; Small, D. W. Incorporating Electronic Information into Machine Learning Potential Energy Surfaces via Approaching the Ground-State Electronic Energy as a Function of Atom-Based Electronic Populations. *J. Chem. Theory. Comput.* **2020**, *16*, 4256–4270.

(34) Ramachandran, P.; Zoph, B.; Le, Q. V. Searching for Activation Functions. *arXiv* **2017**, 1710.05941.

(35) Nair, V.; Hinton, G. E. Rectified Linear Units Improve Restricted Boltzmann Machines. Proceedings of the 27th International Conference on International Conference on Machine Learning. Madison, WI, USA, 2010; p 807–814.

(36) Dozat, T. Incorporating Nesterov Momentum into Adam. International Conference on Learning Representations 2016 Workshop. Stanford University: Stanford, CA, USA, 2016; pp 437–453.

(37) Shao, K.; Chen, J.; Zhao, Z.; Zhang, D. H. Communication: Fitting Potential Energy Surfaces with Fundamental Invariant Neural Network. *J. Chem. Phys.* **2016**, *145*, 71101.

(38) Chen, R.; Shao, K.; Fu, B.; Zhang, D. H. Fitting Potential Energy Surfaces with Fundamental Invariant Neural Network. II. Generating Fundamental Invariants for Molecular Systems with up to Ten Atoms. *J. Chem. Phys.* **2020**, *152*, 204307.

(39) Rupp, M.; Tkatchenko, A.; Müller, K.-R.; von Lilienfeld, O. A. Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning. *Phys. Rev. Lett.* **2012**, *108*, 058301.

(40) Montavon, G.; Hansen, K.; Fazli, S.; Rupp, M.; Biegler, F.; Ziehe, A.; Tkatchenko, A.; Lilienfeld, A.; Müller, K.-R. Learning Invariant Representations of Molecules for Atomization Energy Prediction. Advances in Neural Information Processing Systems. 2012; pp 440–448.

(41) Hansen, K.; Montavon, G.; Biegler, F.; Fazli, S.; Rupp, M.; Scheffler, M.; von Lilienfeld, O. A.; Tkatchenko, A.; Müller, K.-R. Assessment and Validation of Machine Learning Methods for Predicting Molecular Atomization Energies. *J. Chem. Theory. Comput.* **2013**, *9*, 3404–3419.

(42) Okuta, R.; Unno, Y.; Nishino, D.; Hido, S.; Loomis, C. CuPy: A NumPy-Compatible Library for NVIDIA GPU Calculations. Proceedings of Workshop on Machine Learning Systems (LearningSys) in The Thirty-first Annual Conference on Neural Information Processing Systems (NIPS). 2017.

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

(44) 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.

(45) Petit, A. S.; Ford, J. E.; McCoy, A. B. Simultaneous Evaluation of Multiple Rotationally Excited States of H_3^+, H_3O^+, and CH_5^+ Using Diffusion Monte Carlo. *J. Phys. Chem. A* **2014**, *118*, 7206–7220.

(46) Petit, A.; McCoy, A. B. Diffusion Monte Carlo Approaches for Evaluating Rotationally Excited States of Symmetric Top Molecules: Application to H_3O^+ and D_3O^+. *J. Phys. Chem. A* **2009**, *113*, 12706–12714.

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

(48) Fore, M. E.; McCoy, A. B. Statistical Analysis of the Effect of Deuteration on Quantum Delocalization in CH_5^+. *J. Phys. Chem. A* **2019**, *123*, 4623–4631.

(49) Babin, V.; Leforestier, C.; Paesani, F. Development of a "First Principles" Water Potential with Flexible Monomers: Dimer Potential Energy Surface, VRT Spectrum, and Second Virial Coefficient. *J. Chem. Theory. Comput.* **2013**, *9*, 5395–5403.

(50) Tschumper, G. S.; Leininger, M. L.; Hoffman, B. C.; Valeev, E. F.; Schaefer, H. F.; Quack, M. Anchoring the Water Dimer Potential Energy Surface with Explicitly Corre-

lated Computations and Focal Point Analyses. *The Journal of Chemical Physics* **2002**, *116*, 690–701.

(51) Schmidt, M.; Roy, P.-N. Path Integral Molecular Dynamic Simulation of Flexible Molecular Systems in Their Ground State: Application to the Water Dimer. *J. Chem. Phys.* **2018**, *148*, 124116.

(52) Sahoo, T.; Iouchtchenko, D.; Herdman, C. M.; Roy, P.-N. A Path Integral Ground State Replica Trick Approach for the Computation of Entanglement Entropy of Dipolar Linear Rotors. *J. Chem. Phys.* **2020**, *152*, 184113.

(53) Ceperley, D. M. Path Integrals in the Theory of Condensed Helium. *Rev. Mod. Phys.* **1995**, *67*, 279–355.

(54) Yan, Y.; Blume, D. Path integral Monte Carlo Ground State Approach: Formalism, Implementation, and Applications. *J. Phys. B: At. Mol. Opt. Phys.* **2017**, *50*, 223001.

Table 1: Mean Absolute Errors for the Training, Validation and Test Set Obtained for H_2O, CH_5^+, and $(H_2O)_2$ (cm^{-1}).

System	Training Error	Validation Error[a]	Test Error[b]
H_2O	5	2	5
CH_5^+	60	33	60
$(H_2O)_2$[c]	28	12	32
$(H_2O)_2$[d]	58	28	61

[a] The validation set consists of 100 000 geometries based on a ground state DMC simulation, as described in the text.

[b] Consists of 80 000 geometries collected during a DMC simulation with scaled masses.

[c] The NN was trained and evaluated using the unsorteed CM descriptor.

[d] The NN was trained and evaluated using the sorted CM descriptor.

Table 2: Calculated DMC Zero-Point Vibrational Energies for H_2O, CH_5^+, and $(H_2O)_2$ (cm^{-1}).

System	Literature Value	NN-DMC	Large NN-DMC[a]
H_2O	4637[b]	4637 (0.4)	
$(H_2O)_2$ [c]	9910 (2)[e]	9910 (2)	9909 (0.3)
$(H_2O)_2$ [d]		9909 (3)	9908 (1)
CH_5^+	10 918 (7)[f]	10917 (4)	10 920 (2)
CH_4D^+	10 307 (11)[f]	10 312 (3)	10 310 (2)
$CH_3D_2^+$	9703 (8)[f]	9701 (6)	9703 (1)
$CH_2D_3^+$	9101 (8)[f]	9105 (2)	9106 (1)
CHD_4^+	8563 (6)[f]	8573 (7)	8570 (1)
CD_5^+	8045 (8)[f]	8048 (6)	8048 (2)

[a] Zero-point energies calculated based on a 200 000 walker DMC simulation described in the Supporting Information.
[b] Variational calculation reported in the manuscript.
[c] The NN was trained and evaluated using the unsorted CM descriptor.
[d] The NN was trained and evaluated using the sorted CM descriptor.
[e] Ref. 11.
[f] Ref. 12.

Table 3: Ground and Excited State Vibrational Energies for H_2O Using the NN-PS and the PS Potentials[21] (cm^{-1}).

v_s^a	v_b	v_a	PS[b]	NN-PS
0	0	0	4636.8	4636.9
0	1	0	1594.4	1594.5
0	2	0	3150.8	3150.8
1	0	0	3656.2	3656.7
0	0	1	3755.1	3755.5
0	3	0	4665.7	4666.0
1	1	0	5233.8	5234.4
0	1	1	5330.0	5330.5

[a] v_s, v_b, v_a correspond to number of quanta in the symmetric OH stretch, the HOH bend, and the antisymmetric OH stretch, respectively.

[b] The first row reports the zero-point energy E_0, and all other values are E-E_0.

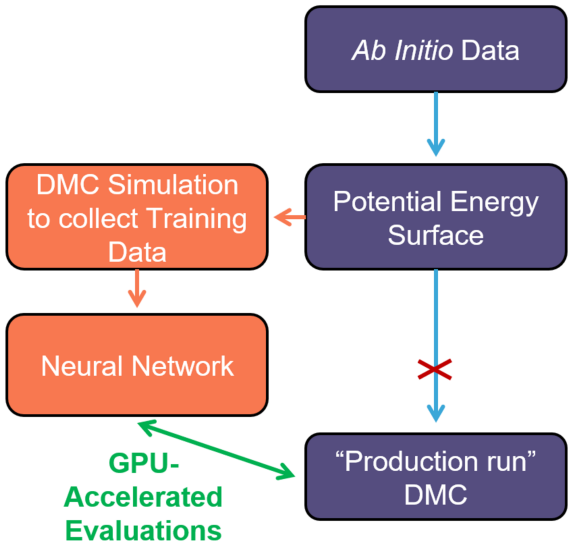

Figure 1: The workflow (purple) of a standard DMC simulation. The NN-DMC workflow (orange). In NN-DMC, rather than evaluating the potential energy directly, we first collect training data and train a neural network potential. Then, we use this neural network for parallel, GPU-accelerated potential energy evaluations.

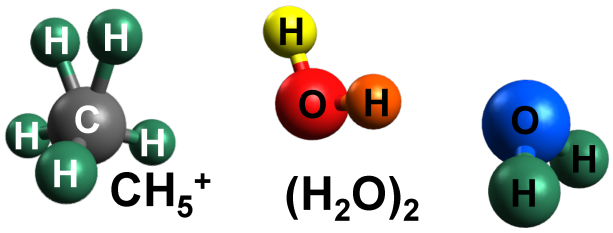

Figure 2: Equilibrium structures of CH_5^+ and $(H_2O)_2$. Permutationally equivalent atoms are shown in green, and permutationally distinct atoms are shown in other colors.

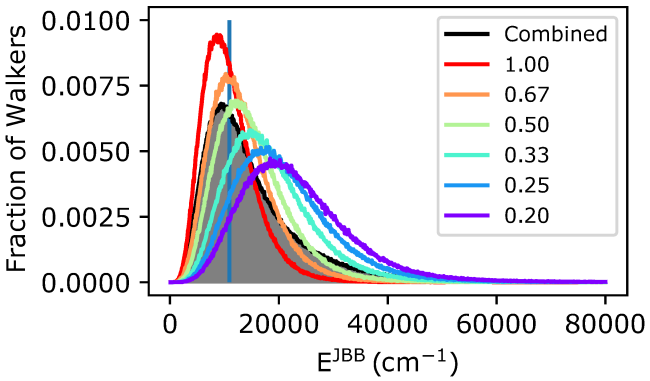

Figure 3: The fraction of the number of walkers obtained from DMC simulations plotted a functions of energy obtained using the JBB potential for CH_5^+.[19] The different colors correspond to walker distributions obtained when the masses of each of the atoms in CH_5^+ are multiplied by the indicated value. The shaded black curve provides the distribution of energies of the training data set used to obtain the NN-JBB potential as described in the text. The dark blue vertical line indicates the value of the calculated zero-point energy of CH_5^+.

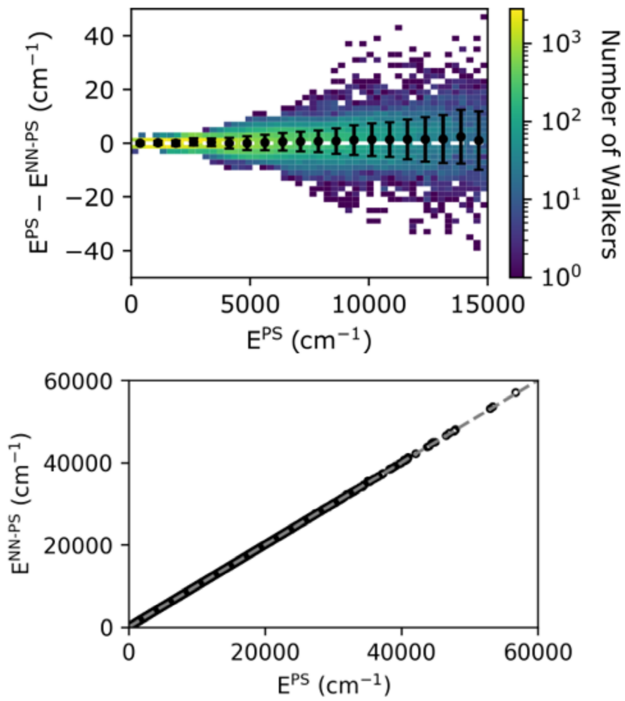

Figure 4: Comparisons between the NN-PS and PS[21] potentials for H_2O. (Top) Plot of the number of walkers as a function of the difference between the energies obtained using the PS and NN-PS surfaces and the energy evaluated using the PS surface. Superimposed on the heat map are the average (black circles) and standard deviation (black error bars) of the energy difference over energy ranges of 750 cm^{-1} centered at the position of the black circle. (Bottom) The predicted NN-PS energy (open circles) as a function of the PS energy over the full energy range sampled by the validation set.

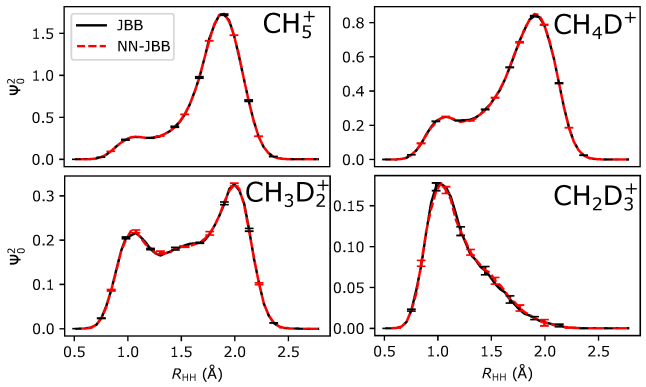

Figure 5: Plots of projections of the DMC probability amplitude onto all HH distances of the indicated system based on the JBB[19] (black) and NN-JBB (red) potential. The error bars indicate the standard deviation of the amplitude of Ψ_0^2 among five independent DMC simulations. The projections are normalized based on the number of HH distances in the ion.

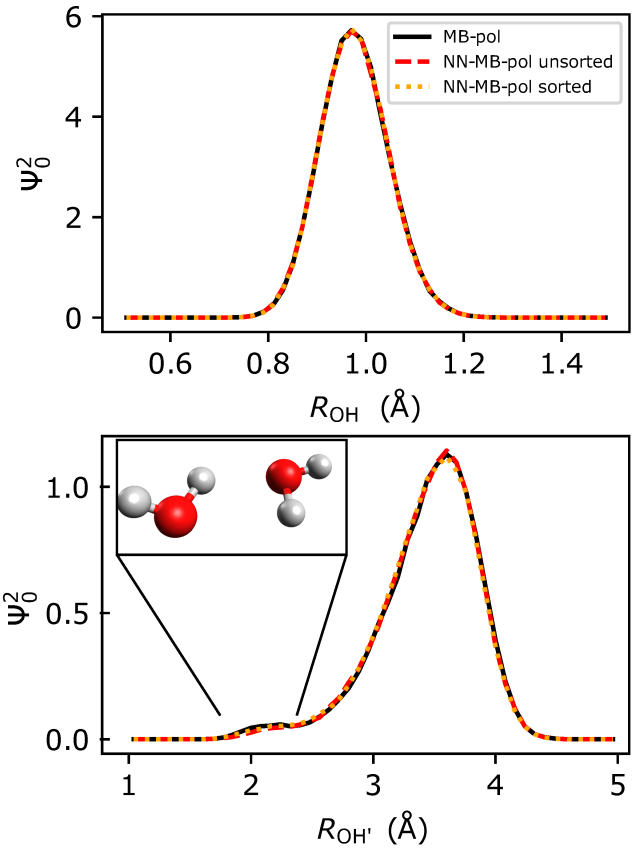

Figure 6: The DMC ground-state probability amplitude obtained from simulations using the NN-MB-pol with an unsorted CM descriptor (red dashed lines) and a sorted CM descriptor (gold dotted lines) and the MB-pol[22–24] potentials (black solid lines) projected onto pairs of OH distances in $(H_2O)_2$. (Top) The DMC probability amplitude projected onto the intramolecular OH distance in the donor water molecule (the two OH bonds between the red oxygen and orange and yellow hydrogen atoms in Figure 2). (Bottom) The DMC probability amplitude projected onto an intermolecular OH distance between the hydrogen atoms in the donor water molecule and the oxygen atom in the acceptor water molecule (the distance between the red oxygen and the two green hydrogen atoms in Figure 2). The inset in the bottom panel shows the geometry of a walker that contributes to the feature near 2 Å, where the walkers sample geometries near the transition state for the exchange of the identities of the donor and acceptor water molecules.

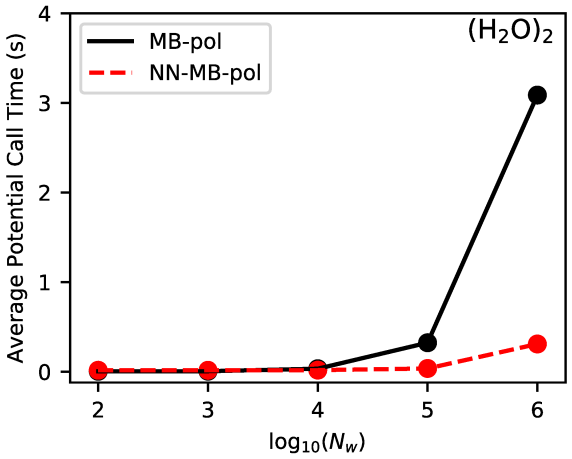

Figure 7: The average potential energy call time in a sample DMC simulation for varying numbers of $(H_2O)_2$ configurations using the MB-pol surface[22–24] (black solid) compared to the NN-MB-pol surface (red dashed).

TOC graphic: