OUKA
The University of Osaka
Institutional Knowledge Archive

Title	Self-learning path integral hybrid Monte Carlo with mixed ab initio and machine learning potentials for modeling nuclear quantum effects in water
Author(s)	Thomsen, Bo; Nagai, Yuki; Kobayashi, Keita et al.
Citation	Journal of Chemical Physics. 2024, 161(20), p. 204109
Version Type	AM
URL	https://hdl.handle.net/11094/99630
rights	This article may be downloaded for personal use only. Any other use requires prior permission of the author and AIP Publishing. This article appeared in Thomsen Bo et al., J. Chem. Phys. 28 November 2024; 161(20): 204109 and may be found at https://doi.org/10.1103/PhysRevMaterials.8.114004.
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

https://ir.library.osaka-u.ac.jp/

The University of Osaka

ACCEPTED MANUSCRIPT

The Journal of Chemical Physics

AIP Publishing

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0230464

Self-Learning Path Integral Hybrid Monte Carlo with Mixed Ab Initio and Machine Learning Potentials for Modelling Nuclear Quantum Effects in Water

Bo Thomsen,[1, a)] Yuki Nagai,[2] Keita Kobayashi,[1] Ikutaro Hamada,[3] and Motoyuki Shiga[1]

[1)] *CCSE, Japan Atomic Energy Agency, 178-4-4, Wakashiba, Kashiwa, Chiba, 277-0871, Japan*

[2)] *Information Technology Center, The University of Tokyo, 6-2-3 Kashiwanoha, Kashiwa, Chiba 277-0882, Japan*

[3)] *Department of Precision Engineering, Graduate School of Engineering, Osaka University, 2-1, Yamadaoka, Suita, Osaka 565-0871, Japan*

(Dated: 18 October 2024)

The introduction of machine learned potentials (MLPs) has greatly expanded the space available for studying Nuclear Quantum Effects (NQEs) computationally with *ab initio* path integral (PI) accuracy, with the MLPs promise of an accuracy comparable to that of *ab initio* at a fraction of the cost. One of the challenges in development of MLPs is the need for a large and diverse training set calculated by *ab initio* methods. This data set should ideally cover the entire phase space, while not searching this space using *ab initio* methods, as this would be counterproductive and generally intractable with respect to computational time. In this paper, we present the self-learning path integral hybrid Monte Carlo Method using a mixed *ab initio* and ML potential (SL-PIHMC-MIX), where the mixed potential allows for the study of larger systems and the extension of the original SL-HMC method [Phys. Rev. B 102, 041124 (2020)] to PI methods and to larger systems. While the MLPs generated by this method can be directly applied to run long-time ML-PIMD simulations, we demonstrate that using PIHMC-MIX with the trained MLPs allows for an exact reproduction of the structure obtained from *ab initio* PIMD. Specifically, we find that the PIHMC-MIX simulations require only 5,000 evaluations of the 32-bead structure, compared to the 100,000 evaluations needed for the *ab initio* PIMD result.

[a)] Electronic mail: thomsen.bo@jaea.go.jp

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0230464

I. INTRODUCTION

Nuclear Quantum Effects (NQEs) play a large role in determining the properties of matter containing light atoms and by extension the isotope effects seen when hydrogen (H) is exchanged for deuterium (D) or tritium (T). One example of this is the observed differences between light (H_2O) and heavy (D_2O) water[1] which has recently been investigated by a series of experiments[2–5]. We have also previously reported some structural and reactive differences between the two liquids and other isotopologues of water[6,7] from *ab initio* or first principles (FP) simulations. Modelling of NQEs in bulk systems relies on path integral (PI) methods based on the Feynman path formulation of quantum mechanics[8–10]. Implementations of these methods[11,12] typically require the simultaneous evaluation of energies and gradients of P copies of the system in each time-step. P is generally considered in tens or in the low hundreds for simulations at room temperature, and thus adds significantly to the cost of performing FP simulations required for accurate description of NQEs in materials.

In the 1990s, methods were suggested for generating machine learned potentials (MLPs)[13–15], with accuracy close to those of FP calculations, but at a much-reduced computational cost. However, MLPs were initially limited to the study of small gas phase clusters. It was only with the introduction of high-dimensional neural network potentials[16–19] by Behler and Parrinello that the MLPs were extended to the study of bulk-phase systems. The development of these MLPs is continuing, with later generations including more physics informed terms such as machine learned atomic charges[20,21] and global charge equilibration[22] for accurate description of charge separation.

From the first MLPs used for the simulation of liquid water[23] the study of bulk phase water using MLPs have undergone a rapid development[24]. With the low cost of evaluation of the MLP allowing for the molecular dynamics simulations of very large systems both with[25] and without[26] NQEs. Generally, fewer FP calculations are needed when training an MLP, and one can thus explore more expensive FP methods for describing the electronic potential in water simulations. MB-Pol[27–29] presents one physics based model for water, which has recently[30,31] been adjusted to fit CCSD(T), *i.e.* the gold standard of quantum chemistry, data for the interaction potentials in water. This model along with other recent fitted MLPs based on FP data from CCSD(T)[32–34] have been shown to accurately reproduce both equilibrium and dynamic properties of water when NQEs are considered. The investigation

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0230464

of NQEs has also been undertaken by a number of studies due to the reduced cost of PI simulations when a MLP is employed[35–40], including comparisons of isotopologues of water[41,42] and the effect of NQEs on the behavior of the hydroxide and hydronium ions[43] in the liquid phase.

Shared by all MLP models are the need for a training set made up of FP data, which should ideally cover the entire phase space while not stemming from an exhaustive search using FP methods. To efficiently carry out the search, one can use on the fly learning[44–48] to train a cheap potential representation, which can be used to accelerate the search. Several of the authors recently suggested the self-learning hybrid Monte Carlo (SL-HMC) method[49,50] based on the hybrid Monte Carlo (HMC) method[51–56]. In the SL-HMC method, a short ML-MD simulation is run between each HMC step to allow efficient sampling of phase space, while training an MLP for the system being studied. Extension of this method to larger systems and to the PI domain is however hindered by the limitations of HMC, as the acceptance ratio scales inversely with the size of the system. Here we introduce the self-learning path integral hybrid Monte Carlo Method using a mixed FP and ML potential (SL-PIHMC-MIX), to overcome this limitation. In brief, this method allows for larger discrepancies between the FP and ML potential energies through the potential mixing, thus enabling larger acceptance ratios and faster sampling of the phase space of the mixed potential Hamiltonian. Thus, reweighting[57] and longer trajectories are necessary to sample the phase space of the FP Hamiltonian. The savings enabled by the larger acceptance rate of the potential mixing scheme are, however, great enough, that the effective length of the trajectory using potential mixing exceeds those using the pure FP potentials. The SL-PIHMC-MIX method is furthermore, as the SL-HMC method, fully general with respect to the FP model used and the MLP model used.

In this study, we will use SL-PIHMC-MIX to train an MLP to model room temperature water. After training the MLP it will be used in a production run using the PIHMC-MIX method, which allows us to rapidly converge the radial distribution functions (RDFs), and thus predict the structure of water and thus predict the structure of water using only 5000 FP calculations along the bead chain, compared to the 100,000 needed in our previous FP-PIMD studies[6,7] to converge the water RDFs. The structure of water has long been a topic of discussion[58], and FP based studies of water using density functional theory (DFT) has since the first report[59], and until the emergence of coupled cluster based MLPs, been the state

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0230464

of the art for studying water, with several studies comparing the accuracy of functionals for this purpose[60,61]. Recent advances in algorithms for PI propagation have allowed for the study of dynamics including NQEs using hybrid functionals[62], and FP-based molecular dynamics (FP-MD) studies have also been conducted at the MP2[63] and quantum Monte Carlo[64] levels of theory. DFT and other FP based studies remain relevant in the context of solvated systems where no general high quality MLP or model is currently available.

This paper is organized as follows. First, we will extend the SL-HMC and HMC methods to the PI formalism and introduce the SL-PIHMC-MIX and PIHMC-MIX methods that allow the study of systems containing many particles. Reweighting of the results from PIHMC-MIX to get the structural properties of the DFT ensemble will also be described in this section. In the following section, the computational details of the simulations used in this work are given. In the result section, the results from the PIHMC-MIX method using an MLP that was fitted using SL-PIHMC-MIX will be compared to the results of FP-PIMD for the RPBE-D3 functional. The effects of the mixed potential method and the accuracy of the MLPs produced by the SL-PIHMC-MIX method will then be discussed. We will briefly discuss the description of heavy water (D_2O) using the PIHMC-MIX method and the MLPs produced by the SL-PIHMC-MIX method. We will then go on to compare the results of PIHMC-MIX for SCAN, rev-vdW-DF2 and optB88-vdW functionals with both experimental data and those from the RPBE-D3 functional. For each of the SCAN, rev-vdW-DF2 and optB88-vdW functionals a unique MLP has been fitted using the SL-PIHMC-MIX method. Finally, we will provide a summary of the findings of this study in the conclusion.

II. THEORY

A. Self-Learning Path Integral Hybrid Monte Carlo

The SL-HMC method has previously been reported by some authors[49,50]. In this section this method will be extended to the PI domain, to the so-called SL-PIHMC method, and then to larger system sizes in the SL-PIHMC-MIX method. In this study, DFT with various functionals will be used as the FP method. However, the approach is fully general and could accommodate a wave function-based method, provided the computational time allows for thousands of evaluations across the entire bead chain. The same holds for the low level

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0230464

method, which is here an MLP denoted by ML, for the SL-PIHMC-MIX method any model which can be updated based on data from the FP data could be used. In PIHMC we use the path integral formulation of quantum mechanics, and the atomic positions, $(\mathbf{R})$, are therefore expanded into P imaginary time-slices or so-called beads, *i.e.* $(\mathbf{R}) = \left(\mathbf{R}^{(1)}, \ldots, \mathbf{R}^{(P)}\right)$. The j^{th} bead contains all the coordinates of the N atoms in the bead, $(\mathbf{R}^j) = \left(\mathbf{R}_1^{(j)}, \ldots, \mathbf{R}_N^{(j)}\right)$. The equations of motion for PIMD and related methods are commonly derived in normal mode space. The reason for this is to better allow energy transfer between the modes at high temperatures, and to ease the derivations of the equations of motion for the system[54,65,66]. Here, the coordinates for all the beads of the I^{th} atom $(\mathbf{R}_I) = \left(\mathbf{R}_I^{(1)}, \ldots, \mathbf{R}_I^{(P)}\right)$ in the system are transformed to the normal mode space $\mathbf{Q}_I = \left(\mathbf{Q}_I^{(1)}, \ldots, \mathbf{Q}_I^{(P)}\right)$. $\mathbf{Q}_I$ and the corresponding momenta $\mathbf{P}_I$ will in the following be assumed to be expanded as vectors.

The heart of the PIHMC method is accepting or rejecting a Monte Carlo move from the point in phase space $\{\mathbf{P}, \mathbf{Q}\}$ to $\{\mathbf{P}', \mathbf{Q}'\}$ with probability of accepting a step given as

$$P_{\text{acc}}\left(\{\mathbf{P}, \mathbf{Q}\} \rightarrow \{\mathbf{P}', \mathbf{Q}'\}\right) = \min\left(1, \exp\left(-\beta\left(H_{\text{FP}}\left(\{\mathbf{P}', \mathbf{Q}'\}\right) - H_{\text{FP}}\left(\{\mathbf{P}, \mathbf{Q}\}\right)\right)\right)\right) \quad (1)$$

where $\beta = \frac{1}{Tk_{\text{B}}}$, T is the temperature and k_{B} is the Boltzmann constant. The Hamiltonian for the whole system, where the potential energy is evaluated within a given model (mod), is

$$H_{\text{mod}}(\{\mathbf{P}, \mathbf{Q}\}) = \frac{1}{2}\sum_{I=1}^{N}\left(\mathbf{P}_I^T\boldsymbol{\mu}_I^{-1}\mathbf{P}_I + M_I\omega_P^2\mathbf{Q}_I^T\boldsymbol{\lambda}\mathbf{Q}_I\right) + V_{\text{av}}^{\text{mod}}\left(\{\mathbf{Q}\}\right). \quad (2)$$

Here $\boldsymbol{\mu}_I$ is a diagonal matrix containing the normal mode masses, M_I is the mass of the Ith particle in the system, $\boldsymbol{\lambda}$ is a diagonal matrix composed of the eigenvalues stemming from the normal mode transform, and $\omega_P = \frac{\sqrt{P}}{\beta\hbar}$. The bead average potential in the model, mod, is given as

$$V_{\text{av}}^{\text{mod}}(\{\mathbf{Q}\}) = \frac{1}{P}\sum_{s=1}^{P} V^{\text{mod}}\left(\mathbf{R}_1^{(s)}\left(\mathbf{Q}_1\right), \ldots, \mathbf{R}_N^{(s)}\left(\mathbf{Q}_N\right)\right) = \frac{1}{P}\sum_{s=1}^{P} V^{\text{mod}}\left(\mathbf{R}^{(s)}\right) \quad (3)$$

Here we introduce a shorthand for the potential energy for the s^{th} bead in the system $\left(V^{\text{mod}}\left(\mathbf{R}^{(s)}\right)\right)$ to avoid direct reference to the normal to Cartesian coordinate transform $\left(\mathbf{R}_i^{(s)}\left(\mathbf{Q}_i\right)\right)$ and to ease the notation later in this manuscript.

The diagrammatic form of SL-PIHMC and PIHMC is shown in Figure 1(a). It is assumed that one has a primitive initial guess for the MLP. As shown in Figure 1(b) before each Monte Carlo step the system is propagated according to the MLP, mod = ML, for n_{ML} steps with

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0230464

the time step Δt_{ML}. The initial momenta for the propagation of the ML trajectory, $\mathbf{P}$, are generated after each Monte Carlo step from a random sample of the Maxwell-Boltzmann distribution with the temperature T. The equations of motion and details of this propagation are widely available in literature[11,54,67–69] describing the PIMD methodology. After n_{test} Monte Carlo steps n_{ML} can be updated, depending on the average acceptance rate from the previous n_{test} steps, $A_{\mathrm{acc}} = \frac{n_{\mathrm{acc}}}{n_{\mathrm{test}}}$, where n_{acc} is the number of accepted Monte Carlo steps out of the last n_{test} steps. This is in our implementation done by either doubling n_{ML}, if $A_{\mathrm{acc}} > P_{\mathrm{upper}}$, to a maximum of $n_{\mathrm{ML}}^{\mathrm{max}}$; or halving n_{ML}, if $A_{\mathrm{acc}} < P_{\mathrm{lower}}$, to a minimum of $n_{\mathrm{ML}}^{\mathrm{min}}$. All of these values, Δt_{ML}, n_{test}, P_{upper}, P_{lower} , $n_{\mathrm{ML}}^{\mathrm{max}}$, and $n_{\mathrm{ML}}^{\mathrm{min}}$ can be provided by the user on input.

A crucial feature of the SL-PIHMC method is the retraining of the MLP during the simulation at every n_{FP} Monte Carlo steps. The MLP will thus implicitly depend on time in the SL-PIHMC method, and the potential in Eq. (3) is formally given as,

$$V_n^{\mathrm{ML}}\left(\mathbf{R}^{(s)}\right), \qquad t_n < t < t_{n+1}. \tag{4}$$

Where t_n and t_{n+1} indicate the simulated timespan according to the collected times propagated in the ML-PIMD trajectories. While the practical benefits of this time dependence cannot be neglected, it does not change fundamentally the working equations of the SL-PIHMC method. We therefore opt to exclude the subscript n of the MLP to simplify the notation. As shown in the supplementary material of Ref. 49 the HMC method fulfills the detailed balance requirement. This also holds for the SL-HMC method, as the time dependence of the MLP does not change the derivation given there.

In the supplementary material (SM) Section SI we have derived the following form of the acceptance probability

$$P_{\mathrm{acc}}\left(\{\mathbf{P},\mathbf{Q}\} \to \{\mathbf{P}',\mathbf{Q}'\}\right) \sim \min\left(1, \exp\left(-\beta \Delta\Delta V\right)\right), \tag{5}$$

under the assumption that the ML-PIMD trajectories conserve the energy of the system. Here the difference between the FP and MLP energies are introduced as

$$\Delta\Delta V \equiv \frac{1}{P}\sum_{s=1}^{P} \Delta V\left(\mathbf{R}^{(s)}\right) - \Delta V\left(\mathbf{R}'^{(s)}\right) \tag{6}$$

with

$$\Delta V\left(\mathbf{R}^{(s)}\right) \equiv V^{\mathrm{FP}}\left(\mathbf{R}^{(s)}\right) - V^{\mathrm{ML}}\left(\mathbf{R}^{(s)}\right). \tag{7}$$

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0230464

It should here be stressed that while the relation in Eq. (5) is very likely to hold, only Eq. (1) is used to calculate P_{acc} in the PIHMC method. The relation in Eq. (5) is only introduced to illustrate what governs the size of P_{acc} in the PIHMC method below.

Given the relationship in Eq. (5) we see that if the ML-PIMD propagates to a region where the difference between the MLPs and FP are smaller, *i.e.* $\left(\Delta V\left(\mathbf{R}'^{(s)}\right) \leq \Delta V\left(\mathbf{R}^{(s)}\right)\right)$ for all beads, the step is always accepted, *i.e.* $P_{\text{acc}} = 1$. On the other hand, if the initial position the MLPs and FP are equal or very close to, *i.e.* $\Delta V\left(\mathbf{R}^{(s)}\right) \approx 0$ for all beads, P_{acc} of the step will only depend on the difference between the MLPs and FP at the end point of the ML-PIMD propagation $\{\mathbf{P}', \mathbf{Q}'\}$. An example that is useful to think of here is going from a region where the MLP very accurately reproduces the FP energy, to a place where extrapolation error creates an unphysical hole in the MLP. In this case, P_{acc} would be greatly reduced, since $\Delta\Delta V \approx -\frac{1}{P}\sum_{s=1}^{P} \Delta V(\mathbf{R}'^{(s)}) \ll 0$. Since the MLP used in the start of the SL-PIHMC procedure may not be well-trained across the phase space, it is important to avoid stepping too far into the untrained regions. This is essential for maintaining the high efficiency of the underlying PIHMC method in accurately sampling the phase space of the FP method.

B. Self-Learning Path Integral Hybrid Monte Carlo using a Mixed FP and ML Potentials

While well-trained MLPs are generally believed to give a good approximation of the FP potential energy surface, they will inevitably differ from the true FP potential. In literature, a mean absolute error (MAE) for energy per atom (σ_{at}) in the system of around 1 meV per atom is generally considered a threshold for a satisfyingly converged MLP. Naturally, the MAE for the whole system $\left(\sigma_{\text{sys}} = N\sigma_{\text{at}}\right)$ will grow with the number of atoms N in the system. This is also expected to be the case if one were to train the MLP for a small system and then use it on a larger system. In terms of the classical HMC, *i.e.,* $P = 1$ in PIHMC, this means that P_{acc} will naturally decrease with an increasing system size. Since, under the assumption that the two points in phase space compared in the MC step are independent, the error will be dominated by the error in the MLP, which grows larger with the system size. For PIHMC the picture is a little more complicated, since the atomic positions in the beads are coupled. We do however expect that this will lead to increasing errors when the

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0230464

number of beads P increases, given that the region in the FP and ML potentials where each bead is located will likely have a similar error. This will then further decrease the acceptance ratio of the PIHMC method over the HMC method, especially in the case where both N and P are large.

The decrease of P_{acc} directly affects the speed with which the phase space is sampled. By extension, this also slows down the training of the MLP, which in turn does not allow us to reduce the errors in the MLP by a more sampled training set. In order to increase P_{acc} for larger systems and PI simulations to maintain a reasonable acceptance, we suggest to modify P_{acc} in the following way,

$$P_{\text{acc}}\left(\{\mathbf{P},\mathbf{Q}\}\rightarrow\{\mathbf{P}',\mathbf{Q}'\}\right)=\min\left(1,\exp\left(-\beta\left(H_{\text{MIX}}\left(\{\mathbf{P}',\mathbf{Q}'\}\right)-H_{\text{MIX}}\left(\{\mathbf{P},\mathbf{Q}\}\right)\right)\right)\right) \tag{8}$$

where the mixed Hamiltonian is given as

$$H_{\text{MIX}}\left(\{\mathbf{P},\mathbf{Q}\}\right)=H_{\text{FP}}\left(\{\mathbf{P},\mathbf{Q}\}\right)-(1-\alpha)\left(V_{\text{av}}^{\text{FP}}\left(\{\mathbf{Q}\}\right)-V_{\text{av}}^{\text{ML}}\left(\{\mathbf{Q}\}\right)\right), \tag{9}$$

and similarly for the phase space point $\{\mathbf{P}',\mathbf{Q}'\}$. α is a tunable parameter between 0 and 1 that effectively allows a bigger discrepancy between the FP and MLP. The mixed Hamiltonian can also be seen as a special case of the Hamiltonian given in Eq. (2), where the potential is given as

$$V_{\text{av}}^{\text{MIX}}\left(\{\mathbf{Q}\}\right)=\alpha V_{\text{av}}^{\text{FP}}\left(\{\mathbf{Q}\}\right)+(1-\alpha)\,V_{\text{av}}^{\text{ML}}\left(\{\mathbf{Q}\}\right). \tag{10}$$

We denote this method path integral hybrid Monte Carlo with potential mixing (PIHCM-MIX). Correspondingly, if we allow for MLP retraining during the propagation, we denote the method as self-learning PIHMC-MIX (SL-PIHMC-MIX). As shown in Figure 1(c) this method does not sample the phase space of the FP functional, but rather the phase space of the mixed potential energy surface, V^{MIX}. Besides the change in the potential term, the steps in the algorithm are the same as for the SL-PIHMC method, shown in Figure 1(a).

Choosing the value of α is a matter of compromise. On one hand, a large α value ensures the relevancy of the points sampled in the context of the phase space of the FP ensemble. On the other hand, a small α value allows for a faster sampling, although it is less likely that the points sampled are relevant in exploring the phase space for the FP method. The efficiency gain also depends on how computationally cheap the evaluation of the MLP is. The cheaper the evaluation, the longer one would wish to propagate in ML-PIMD before doing a costly

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0230464

FP calculation. Finally, it should be mentioned that $\alpha \approx 0$ can have the effect that the ML-PIMD trajectory steps too far into regions with large extrapolation errors, which can lead to instability in the FP calculations due to sampling of physically irrelevant structures of the system. For this study, we have chosen $\alpha = 0.25$ as a compromise between efficiency of the MLP and compatibility to the FP phase space. The effects of this choice on the increase in n_{ML} are shown for the SL-PIHMC-MIX method in Figure 2. Where all training sessions lead to running with $n_{\mathrm{ML}} = 128$ and $\langle A_{\mathrm{acc}} \rangle > 0.33$ relatively fast. We do see some drops in n_{ML} between segments, but the A_{acc} quickly recovers and n_{ML} is increased again. We also observe that as the training set grows beyond the initial 1,000 structures, both the acceptance rate and n_{ML} increases. This indicates the importance of longer trajectories to collect training data that represent the entire phase space of the studied system.

C. Reweighting to obtain the FP ensemble distributions of equilibrium properties

The PIHMC-MIX method allows us to accurately predict the distribution $\rho^{\mathrm{MIX}}(A)$ of a structural parameter (A) in the phase space of $H_{\mathrm{MIX}}(\{\mathbf{P}, \mathbf{Q}\})$. We do however wish to generate the distributions in the phase space of $H_{\mathrm{FP}}(\{\mathbf{P}, \mathbf{Q}\})$, which is guarantied by the PIHMC method. To that end, we employ the reweighting scheme suggested by Miao *et al.* in Ref. 57. In the exact limit the trajectory can be divided into M equally sized bins, and the distribution of the structural parameter A of each bin can be reweighted in the following way to obtain the distribution in the FP ensemble

$$\rho^{\mathrm{FP}}\left(A_j\right) = \rho^{\mathrm{MIX}}\left(A_j\right) \frac{\left\langle \exp\left(\beta\left(\alpha-1\right)\Delta V^{\mathrm{MIX}}\right)\right\rangle_j}{\sum_{j=1}^{M}\left\langle \exp\left(\beta\left(\alpha-1\right)\Delta V^{\mathrm{MIX}}\right)\right\rangle_j}, \tag{11}$$

where the counter is the ensemble-averaged Boltzmann factor for the simulation frames found in the j^{th} bin, and the potential difference is defined as

$$\Delta V^{\mathrm{MIX}} = \frac{1}{P}\sum_{s=1}^{P} V^{\mathrm{ML}}\left(\mathbf{R}^{(s)}\right) - V^{\mathrm{FP}}\left(\mathbf{R}^{(s)}\right) \tag{12}$$

The exact reweighting is however difficult to converge due to the exponentiation of the potential differences. To avoid this the cumulant expansion of the average the exponential is introduced

$$\left\langle \exp\left(\beta\left(\alpha-1\right)\Delta V^{\mathrm{MIX}}\right)\right\rangle = \exp\left\{\sum_{k=1}^{\infty}\frac{\beta^k}{k!}C_k\right\}, \tag{13}$$

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0230464

where the first cumulant is given as

$$C_1 = \left\langle (\alpha - 1)\, \Delta V^{\mathrm{MIX}} \right\rangle = (\alpha - 1) \left\langle \Delta V^{\mathrm{MIX}} \right\rangle . \tag{14}$$

The study of Miao *et al.* established that considering only the first cumulant in this expansion was sufficiently accurate to reweight the results, and we will follow that procedure here. The ensemble-averaged Boltzmann factor does in this case reduce to,

$$\left\langle \exp\left(\beta\, (\alpha - 1)\, \Delta V^{\mathrm{MIX}}\right) \right\rangle \approx \exp\left(\beta C_1\right) = \exp\left(\beta \left\langle (\alpha - 1)\, \Delta V^{\mathrm{MIX}} \right\rangle\right) . \tag{15}$$

This is then inserted into Eq. (11) and forms the following expression

$$\rho^{\mathrm{FP}}\left(A_j\right) \approx \rho^{\mathrm{MIX}}\left(A_j\right) \frac{\exp\left(\beta\, (\alpha - 1) \left\langle \Delta V^{\mathrm{MIX}} \right\rangle_j\right)}{\sum_{j=1}^{M} \exp\left(\beta\, (\alpha - 1) \left\langle \Delta V^{\mathrm{MIX}} \right\rangle_j\right)} \tag{16}$$

This expression has been used to do reweighting the RDFs calculated from the PIHMC-MIX trajectories with a bin size of $M = 20$. In the weighting expression, only structures from accepted HMC steps are considered. This is done in order to avoid adding artificial weight to structures where several trial ML-PIMD trajectories are needed before the MC step is accepted. In section SII, we discuss the addition of higher-order terms in the cumulant expansion and find that the resulting RDFs using the first- and second-order expansion for reweighting PIHMC-MIX data overlap. Furthermore, the "anharmonicity"[57,70] observed in the binned data suggests that binning and expansion to second order should be sufficient for estimating the exponential reweighting in Eq. (11).

III. COMPUTATIONAL DETAILS

All the simulations were undertaken using the `PIMD` software package[71], which is capable of conducting PIMD, PIHMC-MIX and SL-PIHMC-MIX simulations. Through an interface to the quantum chemistry software package `CP2K`[72], potential energy and forces at the FP level within the Born-Oppenheimer approximation can be used for HMC steps and PIMD propagation. The `ELPA`[73] and `FFTW`[74] libraries were used by `CP2K` to speed up the solution of the electronic structure eigenvalue equations and to carry out fast Fourier transform, respectively. The MLPs were trained and evaluated using `AENET`[75]. The parameters of the Behler-Parrinello structural fingerprint parameters[16] used in here are given in Table SI.

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0230464

The neural networks were all prepared with 2 layers with hyperbolic tangential activation functions and 15 nodes per layer, and a single linear combination output layer, resulting in a total of 1290 free parameters for both the O and H atomic potentials.

The RPBE[76], SCAN[77], rev-vdW-DF2[78], and optB88-vdW[79] functionals were used from their implementations in the `libxc` library[80,81]. Grimme's D3 dispersion correction[82,83] was employed to model the van der Waals interactions in the RPBE functional. The electronic structure calculations in the periodic boundary condition (PBC) were performed using the Gaussian and plane-wave (GWP) method[84] with the plane wave cutoffs of 500 Ry for the RPBE functional and 800 Ry for the other functionals to expand the charge density. Only the Γ-point was used for the Brillouin zone sampling. The plane-wave basis set was combined with the TZV2P basis set[85] associated with the Goedecker-Teter-Hutter (GTH) pseudopotentials[86] to describe the electron-ion interactions.

All simulations were carried out in the NVT ensemble with 64 or 256 water molecules in a cubic box with PBC. The volume of the cubic box was chosen to match the experimental density at 298.15 K (1.00 g/mL), *i.e.,* the side lengths of the box at ambient conditions were set to 12.41 Å and 19.71 Å for the systems with 64 and 256 water molecules respectively. The temperature was controlled with the massive Nosé-Hoover chain (MNHC) thermostats[87–89] in all PIMD and MD simulations. The number of imaginary time slices (the number of beads) were $P = 1$ and $P = 32$ for the classical and quantum simulations, respectively. All simulations were conducted with a time step of $\Delta t = 0.25$ fs. The ML-MD, FP-PIMD and ML-PIMD were each propagated for 100,000 steps, corresponding to a trajectory length of 25 ps for each of those trajectories. While the AI-MD simulation was propagated for 200,000 steps, 50 ps, to ensure convergence of the RDFs. The error bars for the RDFs from MD and PIMD simulations were calculated by dividing the trajectory into 4 blocks and calculating the standard deviation of the RDFs from the blocks. The central bold lines of the RDF plots were calculated as the average of the RDFs from these blocks.

The SL-HMC-MIX and SL-PIHMC-MIX simulations were initialized with an MLP trained from around 1000 structures from short FP-MD and FP-PIMD trajectories. These trajectories were initialized using the final structure from the previously reported RPBE-D3 FP-PIMD and FP-MD simulations for all functionals. The self-learning process was run for 5000 steps with retraining every $n_{\mathrm{FP}} = 500$ MC steps. The other parameters governing the process were set to, $n_{\mathrm{test}} = 50$, $n_{\mathrm{max}}^{\mathrm{ML}} = 128$, $n_{\mathrm{min}}^{\mathrm{ML}} = 2$, $P_{\mathrm{upper}} = 40\%$, $P_{\mathrm{lower}} = 10\%$

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0230464

respectively. The initial number of ML steps (n_{ML}) were set to 2. For the SL-PIHMC-MIX trajectories the 32 structures and energies of the beads were saved in every 20th PIHMC step for use in training, resulting in training sets containing around 9000 structures at the end of the SL-PIHMC-MIX simulation. While for the SL-HMC-MIX trajectory for RPBE-D3, every structure from the HMC steps was used for training, resulting in a training set containing around 6000 structures.

The average acceptance rate ($\langle A_{\mathrm{acc}} \rangle$) and effective trajectory length (t_{eff}) of SL-HMC-MIX and SL-PIHMC-MIX trajectories used in this study are given in Table SII. The definition of t_{eff} relies on dividing the PIHMC trajectory of length n_{PIHMC} into $O = \frac{n_{\mathrm{PIHMC}}}{n_{\mathrm{test}}}$ sub-trajectories. t_{eff} of the full PIHMC trajectory can then be calculated as

$$t_{\mathrm{eff}} = \sum_{i}^{O} n_{\mathrm{acc}}^{i} n_{\mathrm{ML}}^{i} \Delta t \tag{17}$$

where n_{acc}^{i} and n_{ML}^{i} is the number of accepted steps and the number of ML-PIMD steps taken respectively in the ith sub-trajectory. ML-MD and ML-PIMD trajectories were run as described for the FP-PIMD and FP-MD simulations previously using the trained MLPs trained by SL-HMC-MIX and SL-PIHMC-MIX. The HMC-MIX and PIHMC-MIX trajectories using these trained MLPs were all run for 5000 steps with the initial setting $n_{\mathrm{ML}} = 128$. The resulting t_{eff} are given in Table I and SIII. The SL-PIHMC-MIX trajectories served as equilibration for both the PIHMC-MIX and ML-PIMD trajectories. For the ML-MD and HMC-MIX trajectories, the final structure from the RPBE-D3 FP-MD simulation from our previous work were used as an equilibrated structure.

IV. RESULTS

A. RPBE-D3 PIHMC-MIX

The RPBE-D3 functional has previously been used to model both room temperature and sub- and supercritical water in FP-MD studies by Schienbein and Marx[90,91]. We have also used the functionals in FP-PIMD studies of both liquid water at room temperature, and under sub- and supercritical conditions[92], and its isotopologues at room temperature[7]. For Sub- and supercritical water, there are a number of differences between the experimentally recorded structures and those found even when including NQEs. For room temperature,

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0230464

our previous works show that the RPBE-D3 gives a good agreement with the experimental RDFs when NQEs are included, which can be seen in the comparisons of FP-PIMD to the experimental RDFs[4,93] in parts (a) of Figures 3-5. The quantitative agreement is also very good as shown by the peak positions and heights in Table II-IV, where the largest discrepancies are found at the interstitial region and second peak of the O-O RDF, *i.e.* in the second hydration shell. The height of the second peak is comparable between FP-PIMD and the experiment at 1.19 and 1.12 respectively, while the position of the second peak of FP-PIMD is at 4.35 Å, while for the experiment it is at 4.53 Å. This indicates that the second hydration shell and disordered water around the first hydration shell are not well described by the RPBE-D3 functional. This might however also be a finite size effect, as the box size is limited to 12.41 Å in those studies, due to the cost of FP-PIMD for larger systems. This claim will later be addressed by ML-PIMD studies of larger system sizes in Subsection IV C. We will in the following use the trajectory data from our previous studies to confirm the ability of the PIHMC-MIX method to reproduce the results of FP-PIMD simulations.

The PIHMC-MIX results for the RPBE-D3 functional were based on using an MLP trained from an SL-PIHMC-MIX trajectory with $t_{\text{eff}} = 69.3$ ps, see Table SII. One of the features of the PIHMC-MIX method is that it with the reweighting scheme will reproduce the results of the FP-PIMD as described in Section II C. The only difference between the two methods being that the PIHMC-MIX method should be able to explore phase space more efficiently and thus require fewer FP calculations. The production run of PIHMC-MIX after the training was completed had $t_{\text{eff}} = 99.9$ ps, and maintained a high acceptance ratio of 55.5 % while running with 128 steps for the entire run of 5000 HMC steps. The resulting RDFs are given in parts (b) of Figures 3-5 with the peak positions and heights given in Table II, III and IV for O-O, O-H and H-H pairs respectively. For the O-H and H-H pairs the FP-PIMD and PIHMC-MIX results the peak positions and heights as well as the other points on the curves match within ± 0.03 Å on R_{XY} and ± 0.03 on $G\left(R_{\text{XY}}\right)$, which we estimate to be within the error bar of the FP-PIMD simulation due to the length of the trajectory. The first peak of the O-O RDF matches similarly to the FP-PIMD result, but for the second O-O RDF peak the maximum for the FP-PIMD at 4.35 Å shifts to 4.24 Å for the PIHMC-MIX trajectory. Part of the reason for this is believed to be the difference in sampling of the two trajectories, given that the PIHMC-MIX trajectory is effectively

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0230464

almost four times longer than the FP-PIMD trajectory. We conclude that the PIHMC-MIX method reproduces the structure observed from FP-PIMD simulations, while using an order of magnitude fewer FP calculations, 5,000 versus 100,000 for the PIHMC-MIX and FP-PIMD simulations respectively and possibly giving a more complete sampling of the phase space. Thus, accepting the description of the second hydration shell calculated from the PIHMC-MIX trajectory as the correct description within the simulations run with the RPBE-D3 functional.

The MLP trained using the SL-PIHMC-MIX trajectory can also be employed to conduct an HMC-MIX simulation, i.e., a simulation without NQEs. The RDFs plotted in Figures 3-5 (d) show agreement between the HMC-MIX and FP-MD, similar to that found for PIHMC-MIX and FP-PIMD. This is further confirmed by comparing the peak heights and positions in Tables SIV-VI. Once again, the HMC-MIX method samples more efficiently than its FP-MD counterpart, achieving an effective trajectory length of 103.7 ps compared to 50 ps for the AI-MD simulation. Additionally, only 10,000 FP calculations were required for HMC-MIX, versus 200,000 for FP-MD.

B. Influence of α on the performance of PIHMC-MIX

Table I includes the average acceptance rates ($\langle A_{\text{acc}} \rangle$), number of ML steps ($n_{\text{ML}}$) and t_{eff} of the PIHMC-MIX trajectories run using the same MLP but with $\alpha \in \{0.25, 0.5, 0.75, 1.0\}$, with $\alpha = 1.0$ corresponding to the unmodified PIHMC method. We find that by increasing α we both lower the acceptance ratio, and more critically n_{ML}, resulting in shorter t_{eff} for even a significantly larger number of HMC steps. In section SIV, the resulting RDFs for water for the different values for α are compared. A good qualitative and quantitative agreement for the O-H and H-H RDFs are found for the PIHMC-MIX simulations with $\alpha \in \{0.25, 0.5, 0.75, 1.0\}$ with the FP-PIMD results. As is the case for the comparison FP-PIMD and PIHMC-MIX with $\alpha = 0.25$, however, the interstitial region and the second peak of the O-O RDFs are not sufficiently converged for $\alpha \in \{0.5, 0.75, 1.0\}$. This is likely due to the fact that the structure of the second hydration shell is intrinsically harder to sample than the first hydration shell. Given that the t_{eff} of the PIHMC-MIX trajectory with $\alpha = 0.25$ is 99.9 ps, it is assumed that this represents the most converged result reported here. Giving enough computational time, the O-O RDFs for $\alpha \in \{0.5, 0.75, 1.0\}$ would converge to the

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0230464

same result, but the low acceptance rate might make it prohibitory expensive to extend these trajectories.

C. Accuracy of trained MLPs for water

The accuracy of the MLPs generated by the SL-PIHMC-MIX method is generally found to be comparable to those trained in other studies, which bodes well for their use in ML-MD and ML-PIMD studies. It is however important to stress that no matter how poorly trained the MLP is, PIHMC-MIX will still be able to reproduce the FP-PIMD result through the reweighting of the property distributions, give that t_{eff} is long enough. A simple way of checking the quality of the MLP the instantaneous acceptance rate and n_{ML} in PIHMC-MIX, which both in the case of a well-trained MLP should be high. We will in this section look more carefully at the trained MLPs and the accuracy of ML-PIMD based on the trained MLPs compared to FP-PIMD and PIHMC-MIX.

In Figures 3-5 parts (c) the O-O, O-H and H-H RDFs respectively are displayed for ML-PIMD simulations with water systems containing 64 and 256 water molecules using the MLP fitted during the SL-PIHMC-MIX training process with the RPBE-D3 functional. The quantitative agreement with FP-PIMD is found to be slightly worse than the case for PIHMC-MIX, as seen from the peak positions and heights are given in Table SIV-SVI. The larger water systems are included in an effort to examine finite size effects on the RFDs and test the behavior of the MLP under NPT like conditions for the first and second hydration shell. In the comparison between the systems containing 64 and 256 water molecules, we find no significant finite size effects, and thus conclude that the fitted MLP is extendable to larger water system sizes. Furthermore, the size of the systems studied using FP-PIMD and PIHMC-MIX are sufficiently large for studies of the first and second hydration shell structure of water.

Validation of the MLPs themselves are done in Section SVII, where the energies and forces obtained from FP and ML calculations of the same structures are compared. The results are in line with those of previous studies training MLPs for water systems. The MAE for energy per atom ($\sigma_{\text{E}}^{\text{at}}$) is 0.36 meV/atom and the MEA for force ($\sigma_{\text{F}}^{\text{at}}$) is 79.0 meV/Å, these quantities are described in Eq. (S12) and (S14). We also tested the transferability of an MLP trained using SL-HMC-MIX, *i.e.* the MLP is constructed without considering

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0230464

NQEs when creating the FP training data. Here we find that both $\sigma_{\mathrm{E}}^{\mathrm{at}}$ and $\sigma_{\mathrm{F}}^{\mathrm{at}}$ are more significant at 3.58 meV/atom and 199.9 meV/Å respectively. This indicates that the MLP trained without including data reflecting NQEs in their training sets, while transferable, will not give accurate modelling of the NQEs of the system studied. In Figures 3-5 (d), we have plotted the RDFs of ML-MD simulations using the MLP trained by the SL-PIHMC-MIX method, which reproduces the FP-MD results with good qualitative agreement. The quantitative agreement of the peak positions, as shown in Tables SIV-VI, is also found to be acceptable. In Section SVIII we find that the transferability of the MLP trained by SL-PIHMC-MIX is generally greater than that trained using SL-HMC-MIX when considering the RDFs calculated by either method using ML-MD or ML-PIMD.

The transfer of the MLP trained using SL-PIHMC-MIX to be used in an HMC-MIX production run is however found to be smoother. Here we find t_{eff} to be around 88 ps, and $\sigma_{\mathrm{E}}^{\mathrm{at}}$ and $\sigma_{\mathrm{F}}^{\mathrm{at}}$ at 0.70 meV/atom and 0.20 eV/Å respectively. This agreement can however stem from the selection of training data in the SL-PIHMC-MIX method. Where the proximity of the 32 beads from each HMC step could be argued to form a training set similar to that suggested by Cooper *et al.* in Ref. 94 to approximate the inclusion of gradients in the fitting of the MLP. In that study the FP data set were augmented with slightly distorted structures where the energy was calculated by Taylor expansion using the FP energy and gradients of a known structure. Here we do not extrapolate, rather we calculate the FP energies of several distorted points directly, but this might lead to an increase in the accuracy of the MLP as were shown in Ref 94.

D. Simulations of heavy water (D_2O)

Another way of examining the transfer-ability of the MLPs and FP models are the comparison of NQEs in both H_2O and D_2O. As the structure of both liquids at room temperature are known experimentally[4,93,95] and show significant differences, these differences being large enough to not have overlapping error bars in theoretical studies. It is however rarely done due to the added cost of running two separate PIMD simulations in place of one. When two simulations are run, they might reveal problems with the underlying potential model. *E.g.* the GGA functional BLYP-D2 were found to not reproduce the correct order of the O-O peak heights of H_2O and D_2O at room temperature[96], this difference was ascribed to the

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0230464

description of dispersion in the functional and the D2 correction. This failure to reproduce the correct ordering is however not present for RPBE-D3[7]. Furthermore, previous studies using MLPs trained on FP data from the PBE0-TS hybrid functional[41] and SCAN meta-GGA functional[42], also show the correct isotopic ordering while being overall over structured compared to the experimental results for both H_2O and D_2O. This exemplifies the delicate balance in the description of the intermolecular potentials needed to model NQEs correctly in both isotopologues of water.

In Figure 6 the RDFs for all pairs in D_2O are presented for FP-PIMD[7], PIHMC-MIX, and ML-PIMD based on the MLP fitted by SL-PIHMC-MIX for H_2O. The peak heights and positions of these RDFs are given in Table SXI. For the PIHMC-MIX trajectory, we obtain a result which agrees with the FP-PIMD reference data for RPBE-D3. For the ML-PIMD trajectory, minor deviations from the FP-PIMD results are found for the O-D and D-D RDFs. The second peak of the O-O RDF show similar deviations as those discussed for H_2O between the three models.

The first peak and the interstitial region of the O-O RDFs does occur at similar distances for all trajectories, the heights for these two extrema (h_1^{OO}, h_{min}^{OO}) are however quite different. These heights are for FP-PIMD found to have the values (2.65, 0.87) for D_2O, whereas they are (2.47,0.83) in the case of H_2O. These results are in line with those we have calculated here by PIHMC-MIX, (2.61, 0.73) and (2.53, 0.77) for D_2O and H_2O respectively. Both FP-PIMD and PIHMC-MIX results match well with the experimental values for D_2O, (2.62, 0.79)[95], and H_2O, (2.50, 0.78)[4]. For ML-PIMD these heights are (2.55, 0.78) and (2.42,0.87) for D_2O and H_2O respectively. This gives the impression that the MLP on its own does not fully reproduce the FP-PIMD results, especially in the case of D_2O.

In order to improve the agreement between ML-PIMD and FP-PIMD some D_2O structures and energies were added to the FP training set of the MLP by running an addition 2000 step SL-PIHMC-MIX simulation for D_2O after the initial 5000 steps for H_2O. The resulting MLP is then used for ML-PIMD and the RDFs are given in Figure S5 and the peak positions and heights are given in Tables SXI and SXII for D_2O and H_2O respectively. A better agreement for D_2O is observed with the heights of the two first O-O RDF extrema given as (2.62, 0.75), while for H_2O these are found to be (2.40, 0.87) which is slightly worse than before. It is therefore not certain if it is possible to make a balanced MLP able to reproduce the O-O RDFs calculated by FP-PIMD result for both H_2O and D_2O simultaneously. The

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0230464

cheap cost of running the PIHMC method does however make it feasible to simply run a simulation for both H_2O and D_2O, to confirm the values of equilibrium properties. If one needed the MLPs for studying the dynamics of the liquid, it would be recommended to run a separate SL-PIHMC-MIX training for D_2O, where the trained MLP for H_2O could be used to speed up the sampling of phase space significantly.

E. SCAN, rev-vdW-DF2 and optB88-vdW results

Given the efficiency gains demonstrated for the RPBE-D3 functional we are able to extend the study of the effect of NQEs in DFT functionals to the SCAN, rev-vdW-DF2 and optB88-vdW functionals, using a limited computational effort compared to that required to run FP-PIMD simulations for each functional. While there are no FP-PIMD data available for all of these functionals, the PIHMC-MIX method has been shown to reproduce the FP-PIMD results in 5000 HMC steps, given a high n_{ML} and $\langle A_{\mathrm{acc}} \rangle$. $\langle A_{\mathrm{acc}} \rangle$ and t_{eff} for the three functionals are given in Table II and in Table SII for the PIHMC-MIX and HMC-MIX trajectories respectively. We find that while the acceptance rates are smaller than they were for RPBE-D3, they are still high enough for the SCAN and rev-vdW-DF2 functionals to be able to run PIHMC-MIX with t_{eff} of 96.8 and 88.7 ps respectively. The performance for the optB88-vdW functional is however less promising, with an average acceptance rate of 36.8 % and t_{eff} drops to 59.4 ps.

The accuracy of the MLPs from the SL-PIHMC-MIX trajectories are analyzed in Figure S3 (b) for the SCAN functional, and (c) for the rev-vdW-DF2 functional. In those figures, the energies and forces calculated by FP and ML from the same structure taken from the PIHMC-MIX trajectories are compared. $\sigma_{\mathrm{E}}^{\mathrm{at}}$ and $\sigma_{\mathrm{F}}^{\mathrm{at}}$ for the SCAN and rev-vdW-DF2 functionals are 0.44 meV/atom, 61.3 meV/Å and 0.59 meV/atom, 60.2 meV/Å respectively. The MLP trained using the optB88-vdW functional has larger errors when comparing to FP results at 2.51 meV/atom, 109.2 meV/Å for $\sigma_{\mathrm{E}}^{\mathrm{at}}$ and $\sigma_{\mathrm{F}}^{\mathrm{at}}$ respectively, and the distributions of energies and forces in Figure S3 (d) are also more spread out than for MLPs constructed using FP data from the other functionals. The quality of the underlying MLP for a given functional should not change the results of PIHMC-MIX, it should only affect t_{eff} through low $\langle A_{\mathrm{acc}} \rangle$ and n_{ML}. The 59.4 ps trajectory for optB88-vdW should in this context still be sufficient to converge the RDFs of the two first hydration shells of water.

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0230464

For the RDFs calculated using ML-PIMD based on FP data from SCAN and rev-vdW-DF2, the agreement to the PIHMC-MIX results from the same functional are found to be similar to that of the RPBE-D3 functional discussed above for the O-H and H-H RDFs. For the O-O RDFs, the MLPs for SCAN and rev-vdW-DF2 even seems better at reproducing the PIHMC-MIX results than for the RPBE-D3 functional. This might be due to the descriptor chosen to be better at describing the more structured O-O RDFs found for SCAN and rev-vdW-DF2. In the case of optB88-vdW, we find significant discrepancies in all RDFs. Especially the first O-O peak, and the secondary O-H and H-H peaks. This indicates that the description of the H-bond is not the same in the FP and ML potentials. The results from PIHMC-MIX should however correct for this, and should be indicative of the true performance of the optB88-vdW functional for modelling water.

In the Sec. IV E 1, IV E 2 and IV E 3 below we will analyze the calculated RDFs of the SCAN, rev-vdW-DF2 and optB88-vdW functionals. The effects of NQEs on the RDFs will also be discussed by comparing the PIHMC-MIX results to those obtained from HMC-MIX, in both cases using the MLPs trained by SL-PIHMC-MIX. The RDFs including NQEs calculated using PIHMC-MIX are give in Figures 7-9, with the peak positions and heights are given in Table II, III, and IV for O-O, O-H and H-H RDFs respectively. For the RDFs calculated using HMC-MIX see Figure S6 of the SM and the peak positions and heights are given in Table SXIII.

The inclusion of NQEs does naturally soften the intramolecular O-H bonds and H-O-H angles the most due to the low mass of the hydrogen atoms and high zero point energies of the intramolecular degrees of freedom. We thus find that for all functionals studied here that the O-H and H-H RDFs in general and in particular the first peaks of these, are softened from the values obtained by HMC-MIX by the inclusion of NQEs in the PIHMC-MIX simulations. We will therefore focus on the softening of the O-O RDFs when comparing classical and quantum results in the following sections, as these are more sensitive to the intermolecular interactions and thus are more challenging to reproduce accurately.

1. The Scan Functional

With the SCAN functional, the O-O RDF from PIHMC-MIX in figure 7(b) is over-structured compared to the experimental RDFs. Looking to the second peaks of the O-H

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0230464

and H-H RDFs, we find that the hydrogen positions for the hydrogens participating in the hydrogen bond are more localized than in the experimental RDFs. The inclusion of NQEs in the PIHMC-MIX simulation softens the liquid structure somewhat, with the O-O RDFs first peak height changing from 3.41 to 3.24 in the HMC-MIX and PIHMC-MIX simulations respectively. Furthermore, softening is also observed for the second peaks of the O-H and H-H RDF with the inclusion of NQEs. The changes in heights are comparable to those observed for the RPBE-D3 functional, but given that the O-O RDF without NQEs is much more structured for the SCAN functional, this softening with the inclusion of NQEs is not enough to reproduce the experimental structure.

The SCAN functional has been studied using both FP-MD[97], FP-PIMD[39,98], and MLP based methods[38] using MLPs trained on FP data from the SCAN functional. These studies have also shown a tendency to over structure the liquid in the NVT ensemble, even when including NQEs at room temperature. The over-structuring when using classical MD simulations have been attributed to the lack of NQEs, which lead to the practice of simulating water at 330 K in an effort to emulate the effects of NQEs[97]. However, a study by Yao and Kanai[38] found this practice problematic due to a fortuitous cancellation of errors in the underlying potential energy surface, which allowed for an accurate reproduction of the O-O RDFs and other properties of the room temperature liquid. In this study, we similarly examined the local structure of the hydrogen bond and compared to the experimental work of Modig, Pfrommer and Halle[99] in section SXII. Our results show that the inclusion of NQEs widens the hydrogen bond angle ($\beta\left(\mathrm{O}\cdots\mathrm{O}-\mathrm{H}\right)$) and contracts the hydrogen bond donor distance ($R_{\mathrm{H}\cdots\mathrm{O}}$), consistent with Yao and Kanai's conclusions for the SCAN functional at room temperature. We note that Li, Peasani and Voth[39] also explored this issue, finding that the dynamical properties from classical simulations at 330 K do not match the effect of NQEs at room temperature across several model potentials and MLPs including one based on the SCAN functional. Given that the PIHMC-MIX result presented here stems from a simulation with a long t_{eff}, we can also conclude that the room temperature structure of SCAN water is significantly different from the experimental structure even when NQEs are included.

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0230464

2. *The rev-vdW-DF2 Functional*

The rev-vdW-DF2 functional shares a similar accuracy with that of the SCAN functional, while being less expensive to execute and at the same time including van der Waals forces directly in the functional. Compared to the experiment we find that using rev-vdW-DF2 the first hydration shell is over-structured when considering all pair RDFs as seen in figures 7(c), 8(c) and 9(c). This over structuring is similar in size to that of the SCAN functional, giving a much more structured liquid phase than for the RPBE-D3 functional. The intramolecular peaks of the O-H and H-H RDFs are also significantly different from the experimental ones, indicating that the differences in the hydrogen bond structure stems from a small difference in molecular structure. Removing the NQEs by using HMC-MIX as shown in the SM leads to a less structured liquid in terms of the O-O RDFs. This indicates that the rev-vdW-DF2 functional is not capable of reproducing the delicate balance in the hydrogen bonds which generally soften the liquid structure as the NQEs are introduced. In figure S9 (e) and (f) it is observed that $\beta\left(\mathrm{O}\cdots\mathrm{O}-\mathrm{H}\right)$ does not widen to the same degree as were the case for the SCAN functional. Additionally, $R_{\mathrm{H}\cdots\mathrm{O}}$ contracts more significantly, suggesting a much stronger hydrogen bond when NQEs are combined with the rev-vdW-DF2 functional, which might explain the larger degree of structure found in PIHMC-MIX compared to HMC-MIX. The description of room temperature liquid water using the rev-vdW-DF2 functional is thus considered to be worse than that of both the SCAN and the RPBE-D3 functionals.

3. *The optB88-vdW Functional*

The MLP constructed by the SL-PIHMC-MIX method with optB88-vdW is the least accurate in reproducing the results from FP calculations among the four functionals studied here. The results from PIHMC-MIX are not improved upon the poor performance of ML-PIMD with respect to the experimental RDFs. It leads to further over structuring of the RDFs as shown in figures 7(d), 8(d) and 9(d). As in the case for rev-vdW-DF2, RDFs are not only over structured, but also the inclusion of NQEs do not have the effect of softening the O-O RDFs. The hydrogen bond structures reported in figure S9 (g) and (h) are much tighter than in any of the other functionals studied here, this along with a trend similar to that observed for rev-vdW-DF2 are likely the cause for the poor performance of the

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0230464

optB88-vdW functional in this study. It should be noted that previous studies which report the structure of water using the optB88-vdW functional[63,98] find a better agreement with the experimental RDFs. The improvements do however not change the fact that the water is over-structured when described by this functional, to an extent that suggests that the inclusion of NQEs should not lead to a better agreement than what is found for the RPBE-D3 functional. However, we cannot rule out the possibility that the current computational setup is a part of the reason for the poor performance of optB88-vdW shown here.

V. CONCLUSIONS

The PIHMC-MIX method has been shown to reproduce the accuracy of FP-PIMD simulations, while requiring an order of magnitude fewer FP calculations. This speedup does however require training of an MLP, which we have shown can be done on the fly through the SL-PIHMC-MIX method. The cost of fitting the MLP is however not prohibitory expensive, and the computational cost of the method is much smaller than that of the FP-PIMD while allowing for the study of longer t_{eff} and thus more efficient sampling of the phase space. The mixing of FP and MLPs through the α parameter in the PIHMC-MIX method is essential in the context of both PI methods and larger systems, such as the case for the water systems studied here. This is shown for the RPBE-D3 functional, where setting $\alpha = 1$, *i.e.,* using the PIHMC method, results in low acceptance rates in the HMC step. The PIHMC method would thus require prohibitively long trajectories and extensive number of FP calculations for convergence of the RDFs. We have also tested the extend-ability of the method to other states of water, namely ice I_h in Section SXI of the supplementary information. Here it was found that the PIHMC-MIX method using the trained MLP for liquid water were able to converge the RDF within 5,000 steps with acceptance ratio and effective trajectory length slightly smaller and shorter than were the case for the PIHMC-MIX simulations of water. The PIHMC-MIX model thus shows promise for extending the study of water across its complex phase diagram. This along with studies of more complex systems will be the subject of future studies.

The MLPs trained by the SL-PIHMC-MIX method were also found to reproduce the FP-PIMD results for all cases studied, except the MLP trained on optB88-vdW data. This gives the perspective of further computational savings by running ML-PIMD simulations instead

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0230464

of the more expensive PIHMC-MIX simulations. Furthermore, it would be possible to study dynamical properties using the MLPs in methods such as ring polymer molecular dynamics (RPMD)[100,101], centroid molecular dynamics (CMD)[102,103], or the recently proposed Brownian chain molecular dynamics (BCMD) method[66]. From our results we do however find reasons to caution direct transfer of an MLP from H_2O to other isotopologues of water, *i.e.* D_2O, and more extremely an MLP trained on only data without NQEs being transferred to a system where NQEs are considered. The MLP is not guarantied to accurately model the differences caused by NQEs, unless they are specifically trained for them, or that PIHMC-MIX is used to guarantee convergence to the FP-PIMD results. It should be noted that for pure water using either MB-Pol[31] another MLPs[32–34] trained on CCSD(T) data would produce more accurate results than what is found here. DFT based FP-PIMD or ML-PIMD trained on DFT data will however still be necessary to study more complex systems, leaving a wide field of applications of SL-PIHMC-MIX for training MLPs and PIHMC-MIX for studying static properties at the DFT level of theory.

Finally, we have been able to provide a survey of the effects of NQEs in the simulations of H_2O with the RPBE-D3, SCAN, rev-vdW-DF2 and optB88-vdW functionals. We find an increased structuring of O-O RDFs for the rev-vdW-DF2 and optB88-vdW functions when NQEs are considered. From the analysis of the shift in hydrogen bond parameters as NQEs are included, this behavior can be explained as the NQEs for these two functionals are found to tighten the hydrogen bonds. For the SCAN functional a slight softening, especially in the hydrogen bond angle is found, leading to a loosening of the structure with the inclusion of NQEs. However, for RPBE-D3 the softening of the hydrogen bond parameters are more subtle and the averages are further from the experiments[99] than the other functionals as seen in figure S9 (a) and (b). However, it seems that the trend of having longer hydrogen bond donor distance ($R_{\mathrm{H\cdots O}}$) in the distribution are key to the good performance of the functional. The conclusion is that among the four functionals studied here, the RPBE-D3 performs the best for studying the structure of water at room temperature in the NVT ensemble. The situation might change for the NPT ensemble, and for higher temperatures and pressures, where non hydrogen bonded contacts between the water molecules become more important.

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0230464

VI. SUPPLEMENTARY MATERIAL

The supplementary material (SM) contains the derivation of Eq. (5), details on the Behler-Parrinello structure fingerprint used in the MLPs, and additional analysis and data on the simulations and RDFs presented in the main text.

VII. ACKNOWLEDGEMENTS

We thank the JSPS Grant-in-Aid for Scientific Research (24K01145, 23K04670, 21H01603, 23H01273, 18H05519) for financial support. The calculations were conducted using the supercomputer facilities at Japan Atomic Energy Agency (JAEA).

VIII. DATA AVAILABILITY

The data that support the findings of this study are available within the article.

REFERENCES

[1]M. Ceriotti, W. Fang, P. G. Kusalik, R. H. McKenzie, A. Michaelides, M. A. Morales, and T. E. Markland, Chem. Rev. **116**, 7529 (2016).

[2]A. Zeidler, P. S. Salmon, H. E. Fischer, J. C. Neuefeind, J. M. Simonson, H. Lemmel, H. Rauch, and T. E. Markland, Phys. Rev. Lett. **107**, 145501 (2011).

[3]A. Zeidler, P. S. Salmon, H. E. Fischer, J. C. Neuefeind, J. M. Simonson, and T. E. Markland, J. Phys.: Condens. Matter **24**, 284126 (2012).

[4]A. K. Soper, ISRN Physical Chemistry **2013**, e279463 (2013).

[5]Y. Kameda, Y. Amo, T. Usuki, Y. Umebayashi, K. Ikeda, and T. Otomo, Bull. Chem. Soc. Jpn. **91**, 1586 (2018).

[6]B. Thomsen and M. Shiga, J. Chem. Phys. **154**, 084117 (2021).

[7]B. Thomsen and M. Shiga, Phys. Chem. Chem. Phys. **24**, 10851 (2022).

[8]R. P. Feynman, *Statistical Mechanics, A Set of Lectures, California, Institute of Technology* (WA Benjamin, Inc. Advanced Book Program Reading, Massachusetts, 1972).

[9]R. P. Feynman, A. R. Hibbs, and D. F. Styer, *Quantum mechanics and path integrals* (Courier Corporation, 2010).

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0230464

[10]L. S. Schulman, *Techniques and appplications of path integration* (Courier Corporation, 2012).

[11]M. Shiga, in *Reference Module in Chemistry, Molecular Sciences and Chemical Engineering* (Elsevier, 2018).

[12]V. Kapil, M. Rossi, O. Marsalek, R. Petraglia, Y. Litman, T. Spura, B. Cheng, A. Cuzzocrea, R. H. Meißner, D. M. Wilkins, B. A. Helfrecht, P. Juda, S. P. Bienvenue, W. Fang, J. Kessler, I. Poltavsky, S. Vandenbrande, J. Wieme, C. Corminboeuf, T. D. Kühne, D. E. Manolopoulos, T. E. Markland, J. O. Richardson, A. Tkatchenko, G. A. Tribello, V. V. Speybroeck, and M. Ceriotti, Comput. Phys. Commun. **236**, 214 (2019).

[13]T. B. Blank, S. D. Brown, A. W. Calhoun, and D. J. Doren, J. Chem. Phys. **103**, 4129 (1995).

[14]D. F. R. Brown, M. N. Gibbs, and D. C. Clary, J. Chem. Phys. **105**, 7597 (1996).

[15]S. Lorenz, A. Groß, and M. Scheffler, Chem. Phys. Lett. **395**, 210 (2004).

[16]J. Behler and M. Parrinello, Phys. Rev. Lett. **98**, 146401 (2007).

[17]J. Behler, J. Phys. Condens. Matter **26**, 183001 (2014).

[18]J. Behler, Int. J. Quantum Chem. **115**, 1032 (2015).

[19]J. Behler, Angew. Chem. **56**, 12828 (2017).

[20]N. Artrith, T. Morawietz, and J. Behler, Phys. Rev. B **83**, 153101 (2011).

[21]T. Morawietz, V. Sharma, and J. Behler, J. Chem. Phys. **136**, 064103 (2012).

[22]J. Behler, Chem. Rev. **121**, 10037 (2021).

[23]T. Morawietz, A. Singraber, C. Dellago, and J. Behler, Proc. Natl. Acad. Sci. U.S.A. **113**, 8368 (2016).

[24]A. Omranpour, P. Montero De Hijes, J. Behler, and C. Dellago, J. Chem. Phys. **160**, 170901 (2024).

[25]A. Krishnamoorthy, K.-i. Nomura, N. Baradwaj, K. Shimamura, P. Rajak, A. Mishra, S. Fukushima, F. Shimojo, R. Kalia, A. Nakano, and P. Vashishta, Phys. Rev. Lett. **126**, 216403 (2021).

[26]D. Lu, H. Wang, M. Chen, L. Lin, R. Car, W. E, W. Jia, and L. Zhang, Comput. Phys. Commun. **259**, 107624 (2021).

[27]V. Babin, C. Leforestier, and F. Paesani, J. Chem. Theory Comput. **9**, 5395 (2013).

[28]V. Babin, G. R. Medders, and F. Paesani, J. Chem. Theory Comput. **10**, 1599 (2014).

[29]G. R. Medders, V. Babin, and F. Paesani, J. Chem. Theory Comput. **10**, 2906 (2014).

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0230464

[30]G. R. Medders, A. W. Götz, M. A. Morales, P. Bajaj, and F. Paesani, J. Chem. Phys. **143**, 104102 (2015).

[31]X. Zhu, M. Riera, E. F. Bull-Vulpe, and F. Paesani, J. Chem. Theory Comput. **19**, 3551 (2023).

[32]Q. Yu, C. Qu, P. L. Houston, R. Conte, A. Nandi, and J. M. Bowman, J. Phys. Chem. Lett. **13**, 5068 (2022).

[33]J. Daru, H. Forbert, J. Behler, and D. Marx, Phys. Rev. Lett. **129**, 226001 (2022).

[34]M. S. Chen, J. Lee, H.-Z. Ye, T. C. Berkelbach, D. R. Reichman, and T. E. Markland, J. Chem. Theory Comput. **19**, 4510 (2023).

[35]B. Cheng, J. Behler, and M. Ceriotti, J. Phys. Chem. Lett. **7**, 2210 (2016).

[36]V. Kapil, J. Behler, and M. Ceriotti, J. Chem. Phys. **145**, 234103 (2016).

[37]V. Kapil, D. M. Wilkins, J. Lan, and M. Ceriotti, J. Chem. Phys. **152**, 124104 (2020).

[38]Y. Yao and Y. Kanai, J. Chem. Phys. **153**, 044114 (2020).

[39]C. Li, F. Paesani, and G. A. Voth, J. Chem. Theory Comput. **18**, 2124 (2022).

[40]Y. Yao and Y. Kanai, J. Phys. Chem. Lett. **12**, 6354 (2021).

[41]H.-Y. Ko, L. Zhang, B. Santra, H. Wang, W. E, R. A. DiStasio Jr, and R. Car, Mol. Phys. **117**, 3269 (2019).

[42]J. Xu, C. Zhang, L. Zhang, M. Chen, B. Santra, and X. Wu, Phys. Rev. B **102**, 214113 (2020).

[43]A. O. Atsango, T. Morawietz, O. Marsalek, and T. E. Markland, J. Chem. Phys. **159**, 074101 (2023).

[44]Z. Li, J. R. Kermode, and A. De Vita, Phys. Rev. Lett. **114**, 096405 (2015).

[45]R. Jinnouchi, J. Lahnsteiner, F. Karsai, G. Kresse, and M. Bokdam, Phys. Rev. Lett. **122**, 225701 (2019).

[46]Z. Cheng, D. Zhao, J. Ma, W. Li, and S. Li, J. Phys. Chem. A **124**, 5007 (2020).

[47]T. A. Young, T. Johnston-Wood, V. L. Deringer, and F. Duarte, Chem. Sci. **12**, 10944 (2021).

[48]P. Montero de Hijes, C. Dellago, R. Jinnouchi, B. Schmiedmayer, and G. Kresse, J. Chem. Phys. **160**, 114107 (2024).

[49]Y. Nagai, M. Okumura, K. Kobayashi, and M. Shiga, Phys. Rev. B **102**, 041124 (2020).

[50]K. Kobayashi, Y. Nagai, M. Itakura, and M. Shiga, J. Chem. Phys. **155**, 034106 (2021).

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0230464

[51]S. Gottlieb, W. Liu, D. Toussaint, R. L. Renken, and R. L. Sugar, Phys. Rev. D **35**, 2531 (1987).

[52]S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, Phys. Lett. B **195**, 216 (1987).

[53]B. Mehlig, D. W. Heermann, and B. M. Forrest, Phys. Rev. B **45**, 679 (1992).

[54]M. E. Tuckerman, B. J. Berne, G. J. Martyna, and M. L. Klein, J. Chem. Phys. **99**, 2796 (1993).

[55]W. Shinoda, M. Shiga, and M. Mikami, Phys. Rev. B **69**, 134103 (2004).

[56]A. Nakayama, T. Taketsugu, and M. Shiga, Chem. Lett. **38**, 976 (2009).

[57]Y. Miao, W. Sinko, L. Pierce, D. Bucher, R. C. Walker, and J. A. McCammon, J. Chem. Theory Comput. **10**, 2677 (2014).

[58]J. L. Finney, J. Chem. Phys. **160**, 060901 (2024).

[59]B. Chen, I. Ivanov, M. L. Klein, and M. Parrinello, Phys. Rev. Lett. **91**, 215503 (2003).

[60]M. J. Gillan, D. Alfè, and A. Michaelides, J. Chem. Phys. **144**, 130901 (2016).

[61]J. Villard, M. P. Bircher, and U. Rothlisberger, Chem. Sci. **15**, 4434 (2024).

[62]O. Marsalek and T. E. Markland, J. Phys. Chem. Lett. **8**, 1545 (2017).

[63]M. Del Ben, J. Hutter, and J. VandeVondele, J. Chem. Phys. **143**, 054506 (2015).

[64]A. Zen, Y. Luo, G. Mazzola, L. Guidoni, and S. Sorella, J. Chem. Phys. **142**, 144111 (2015).

[65]A. Witt, S. D. Ivanov, M. Shiga, H. Forbert, and D. Marx, J. Chem. Phys. **130**, 194510 (2009).

[66]M. Shiga, J. Comput. Chem. **43**, 1864 (2022).

[67]M. Parrinello and A. Rahman, J. Chem. Phys. **80**, 860 (1984).

[68]R. W. Hall and B. J. Berne, J. Chem. Phys. **81**, 3641 (1984).

[69]M. Ceriotti, M. Parrinello, T. E. Markland, and D. E. Manolopoulos, J. Chem. Phys. **133**, 124104 (2010).

[70]O. F. Lange and H. Grubmüller, Proteins **70**, 1294 (2008).

[71]M. Shiga, "PIMD," (2020).

[72]T. D. Kühne, M. Iannuzzi, M. Del Ben, V. V. Rybkin, P. Seewald, F. Stein, T. Laino, R. Z. Khaliullin, O. Schütt, F. Schiffmann, D. Golze, J. Wilhelm, S. Chulkov, M. H. Bani-Hashemian, V. Weber, U. Borštnik, M. Taillefumier, A. S. Jakobovits, A. Lazzaro, H. Pabst, T. Müller, R. Schade, M. Guidon, S. Andermatt, N. Holmberg, G. K. Schenter,

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0230464

A. Hehn, A. Bussy, F. Belleflamme, G. Tabacchi, A. Glöß, M. Lass, I. Bethune, C. J. Mundy, C. Plessl, M. Watkins, J. VandeVondele, M. Krack, and J. Hutter, J. Chem. Phys. **152**, 194103 (2020).

[73]A. Marek, V. Blum, R. Johanni, V. Havu, B. Lang, T. Auckenthaler, A. Heinecke, H.-J. Bungartz, and H. Lederer, J. Phys. Condens. Matter **26**, 213201 (2014).

[74]M. Frigo and S. G. Johnson, Proc. IEEE **93**, 216 (2005).

[75]N. Artrith and A. Urban, Comput. Mater. Sci. **114**, 135 (2016).

[76]B. Hammer, L. B. Hansen, and J. K. Nørskov, Phys. Rev. B **59**, 7413 (1999).

[77]J. Sun, A. Ruzsinszky, and J. P. Perdew, Phys. Rev. Lett. **115**, 036402 (2015).

[78]I. Hamada, Phys. Rev. B **89**, 121103 (2014).

[79]J. Klimeš, D. R. Bowler, and A. Michaelides, J. Phys. Condens. Matter **22**, 022201 (2009).

[80]M. A. L. Marques, M. J. T. Oliveira, and T. Burnus, Comput. Phys. Commun. **183**, 2272 (2012).

[81]S. Lehtola, C. Steigemann, M. J. T. Oliveira, and M. A. L. Marques, SoftwareX **7**, 1 (2018).

[82]S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, J. Chem. Phys. **132**, 154104 (2010).

[83]S. Grimme, S. Ehrlich, and L. Goerigk, J. Comput. Chem. **32**, 1456 (2011).

[84]G. Lippert, J. Hutter, and M. Parrinello, Mol. Phys. **92**, 477 (1997).

[85]J. VandeVondele and J. Hutter, J. Chem. Phys. **127**, 114105 (2007).

[86]S. Goedecker, M. Teter, and J. Hutter, Phys. Rev. B **54**, 1703 (1996).

[87]W. G. Hoover, Phys. Rev. A **31**, 1695 (1985).

[88]S. Nosé, J. Chem. Phys. **81**, 511 (1984).

[89]G. J. Martyna, M. L. Klein, and M. Tuckerman, J. Chem. Phys. **97**, 2635 (1992).

[90]P. Schienbein and D. Marx, J. Phys. Chem. B **122**, 3318 (2018).

[91]P. Schienbein and D. Marx, Angew. Chem. **59**, 18578 (2020).

[92]B. Thomsen and M. Shiga, J. Chem. Phys. **155**, 194107 (2021).

[93]A. K. Soper, Chem. Phys. **258**, 121 (2000).

[94]A. M. Cooper, J. Kaestner, A. Urban, and N. Artrith, Npj Comput. Mater. **6** (2020).

[95]A. K. Soper and C. J. Benmore, Phys. Rev. Lett. **101**, 065502 (2008).

[96]M. Machida, K. Kato, and M. Shiga, J. Chem. Phys. **148**, 102324 (2017).

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0230464

[97]M. Chen, H.-Y. Ko, R. C. Remsing, M. F. C. Andrade, B. Santra, Z. Sun, A. Selloni, R. Car, M. L. Klein, J. P. Perdew, and X. Wu, Proc. Natl. Acad. Sci. U.S.A. **114**, 10846 (2017).

[98]C. Herrero, M. Pauletti, G. Tocci, M. Iannuzzi, and L. Joly, Proc. Natl. Acad. Sci. U.S.A. **119**, e2121641119 (2022).

[99]K. Modig, B. G. Pfrommer, and B. Halle, Phys. Rev. Lett. **90**, 075502 (2003).

[100]I. R. Craig and D. E. Manolopoulos, J. Chem. Phys. **121**, 3368 (2004).

[101]B. J. Braams and D. E. Manolopoulos, J. Chem. Phys. **125**, 124105 (2006).

[102]J. Cao and G. A. Voth, J. Chem. Phys. **100**, 5106 (1994).

[103]J. Liu and Z. Zhang, J. Chem. Phys. **144**, 034307 (2016).

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0230464

FIGURE CAPTIONS

Figure 1: (a) Diagram describing the flow of an SL-PIHMC simulation. The parts shown in black are the core of the PIHMC method. The orange part updates n_{ML} every n_{test} steps and is not strictly required in the PIHMC method. The green part denotes the SL part of the SL-PIHMC method, and is not active during a pure PIHMC simulation. (b) Diagram of a single step in the PIHMC method. (c) Diagram of a single step of PIHMC-MIX method. Here the acceptance is judged based on the Hamiltonian using the V^{MIX}, rather than V^{DFT} (see the text). The numbers on parts (b) and (c) refer to the step in the diagram given in (a).

Figure 2: (Top) The evolution of the number of ML steps (n_{ML}) between HMC steps during the training process for the RPBE-D3 functional using SL-PIHMC-MIX (blue), the RPBE-D3 functional using SL-HMC-MIX (lightblue), the SCAN functional using SL-PIHMC-MIX (green), the rev-vdW-DF2 functional using SL-PIHMC-MIX (orange), and the optB88-vdW functional using SL-PIHMC-MIX (red). (Center) The evolution of the instantaneous acceptance rate (A_{acc}) for every n_{test} steps. (Bottom) The evolution of the accumulated average acceptance rate ($\langle A_{\mathrm{acc}} \rangle$) over the SL-PIHMC-MIX simulations.

Figure 3: O-O RDFs calculated using the RPBE-D3 functional and MLPs trained on FP data from the same functional using SL-PIHMC-MIX. (a) Comparison of FP-MD (green) and FP-PIMD (blue) from our previous works[7,92], with that of experiment[4] (black). (b) Comparison of the FP-PIMD (blue) and PIHMC-MIX (purple) with that of experiment[4] (black). (c) Comparison of FP-PIMD (blue) with the results of ML-PIMD for a system containing 64 water molecules (yellow) or 256 water molecules (orange). (d) Comparison of FP-MD (green) with HMC-MIX (light blue) and ML-MD for systems containing 64 water molecules (red) and 256 water molecules (pink). Note that these results are different from the results trained using SL-HMC-MIX given in Section SVII.

Figure 4: Same as Figure 3 (a-d) for the O-H RDFs. The first O-H RDF peaks are shown in the insets. The experimental data for the first RDF peak is taken from Ref. 93

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0230464

and is given as a dashed black line.

Figure 5: Same as Figure 3 (a-d) for the H-H RDFs. The first H-H RDF peaks for the simulations without NQEs in figure (a) and (d) are shown in the insets. The experimental data for the first RDF peak is taken from Ref. 93, and is given as a dashed black line.

Figure 6: (a) O-O, (b) O-D and (c) D-D RDFs for D_2O. In all figures, the experimental data[95] (black) are given as a reference. The result from our previously published FP-PIMD[7] result are given in red, in green are the results from ML-PIMD, and in blue are the results from PIHMC-MIX. The MLP used in ML-PIMD and PIHMC-MIX stems from the RPBE-D3 data from the SL-PIHMC-MIX trajectory of H_2O. The peak heights and positions are given in Table SXII. In Figure S5, the ML-PIMD results are compared with results for an MLP partially trained using data from SL-PIHMC-MIX for D_2O, the heights and positions of which are also given in Table SXII.

Figure 7: O-O RDFs calculated using various functions and MLPs trained on FP data from said functions. In all figures, the experimental data[4] (black) are given as a reference. (a) Comparison of the results from PIHMC-MIX (blue) and ML-PIMD (light blue) using the RPBE-D3 functional. (b) Comparison of the results from PIHMC-MIX (green) and ML-PIMD (light green) using the SCAN functional. (c) Comparison of the results from PIHMC-MIX (orange) and ML-PIMD (yellow) using the rev-vdW-DF2 functional. (d) Comparison of the results from PIHMC-MIX (red) and ML-PIMD (pink) using the optB88-vdW functional.

Figure 8: Same as for figure 7 (a-d) for the O-H RDFs. Note that all figures contain a subplot of the first O-H RDF peaks, as this goes out of scale when compared to the secondary and tertiary peaks. Furthermore, the experimental data for the first RDF peak is taken from Ref. 93, and is given as a dashed black line.

Figure 9: Same as for figure 7 (a-d) for the H-H RDFs. Note that the experimental data for the first RDF peak is taken from Ref. 93, and is given as a dashed black line.

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0230464

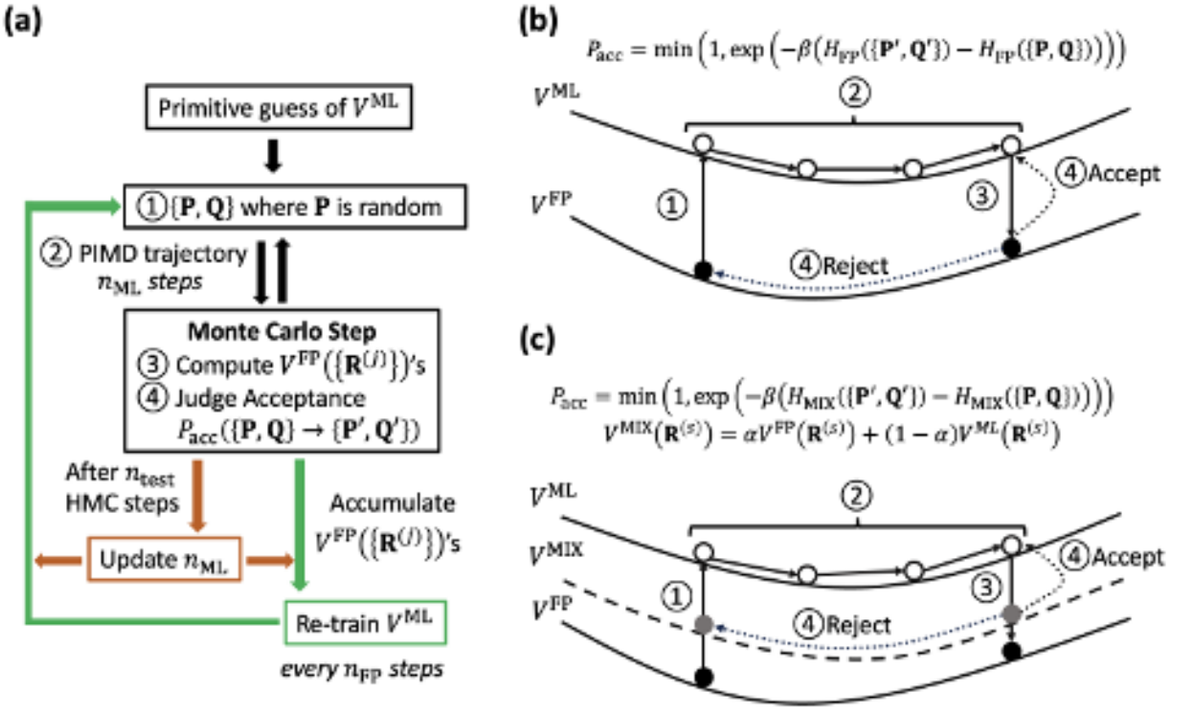

Figure 1, B. Thomsen, Y. Nagai, K. Kobayashi, I. Hamada and M. Shiga, submitted to J. Chem. Phys.

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0230464

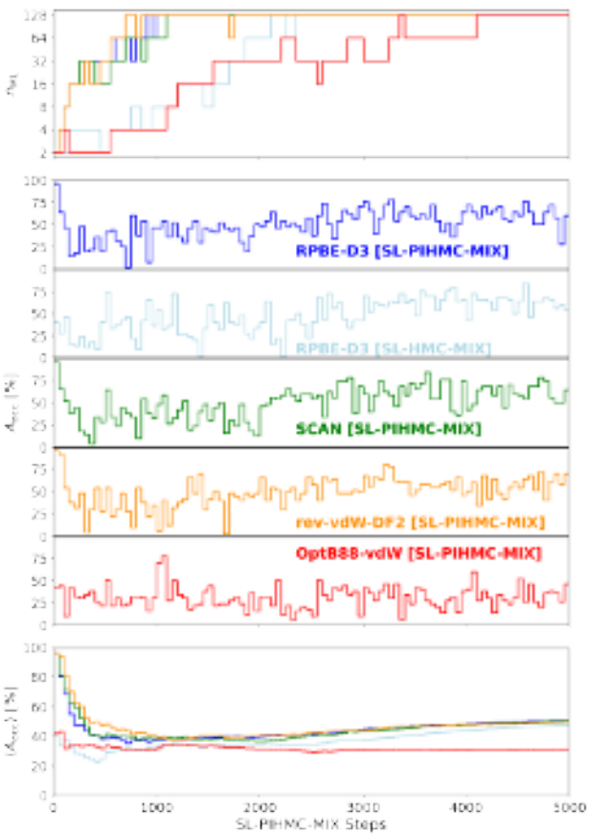

Figure 2, B. Thomsen, Y. Nagai, K. Kobayashi, I. Hamada and M. Shiga, submitted to J. Chem. Phys.

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0230464

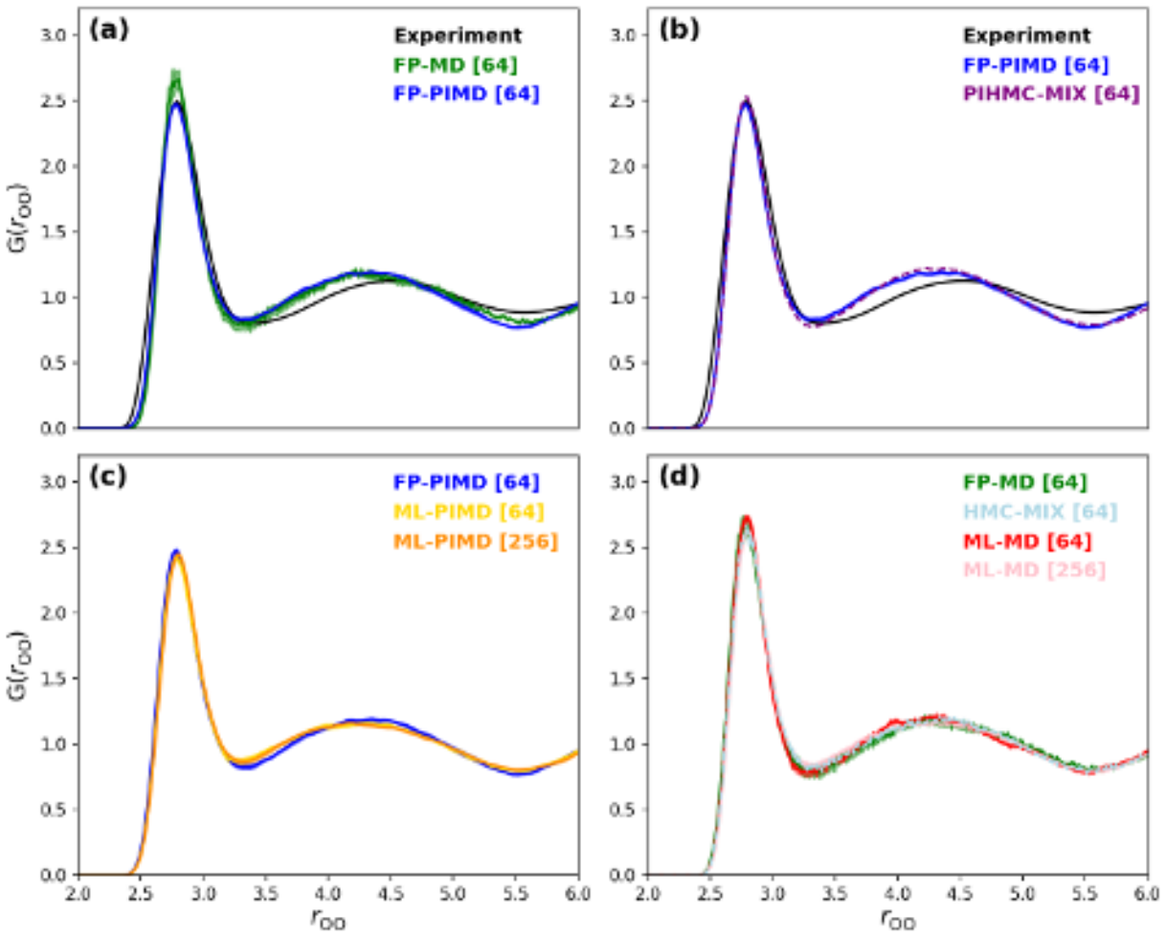

Figure 3, B. Thomsen, Y. Nagai, K. Kobayashi, I. Hamada and M. Shiga, submitted to J. Chem. Phys.

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0230464

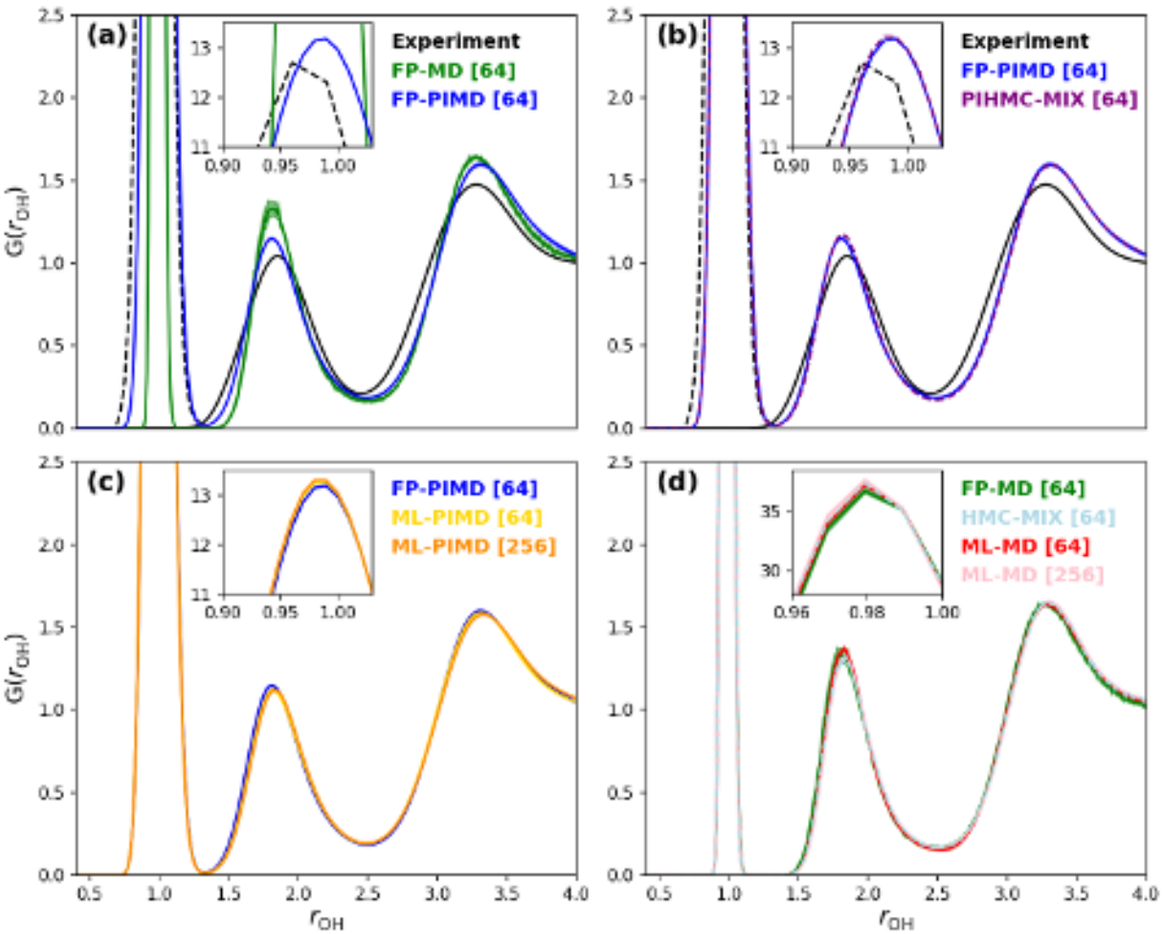

Figure 4, B. Thomsen, Y. Nagai, K. Kobayashi, I. Hamada and M. Shiga, submitted to J. Chem. Phys.

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0230464

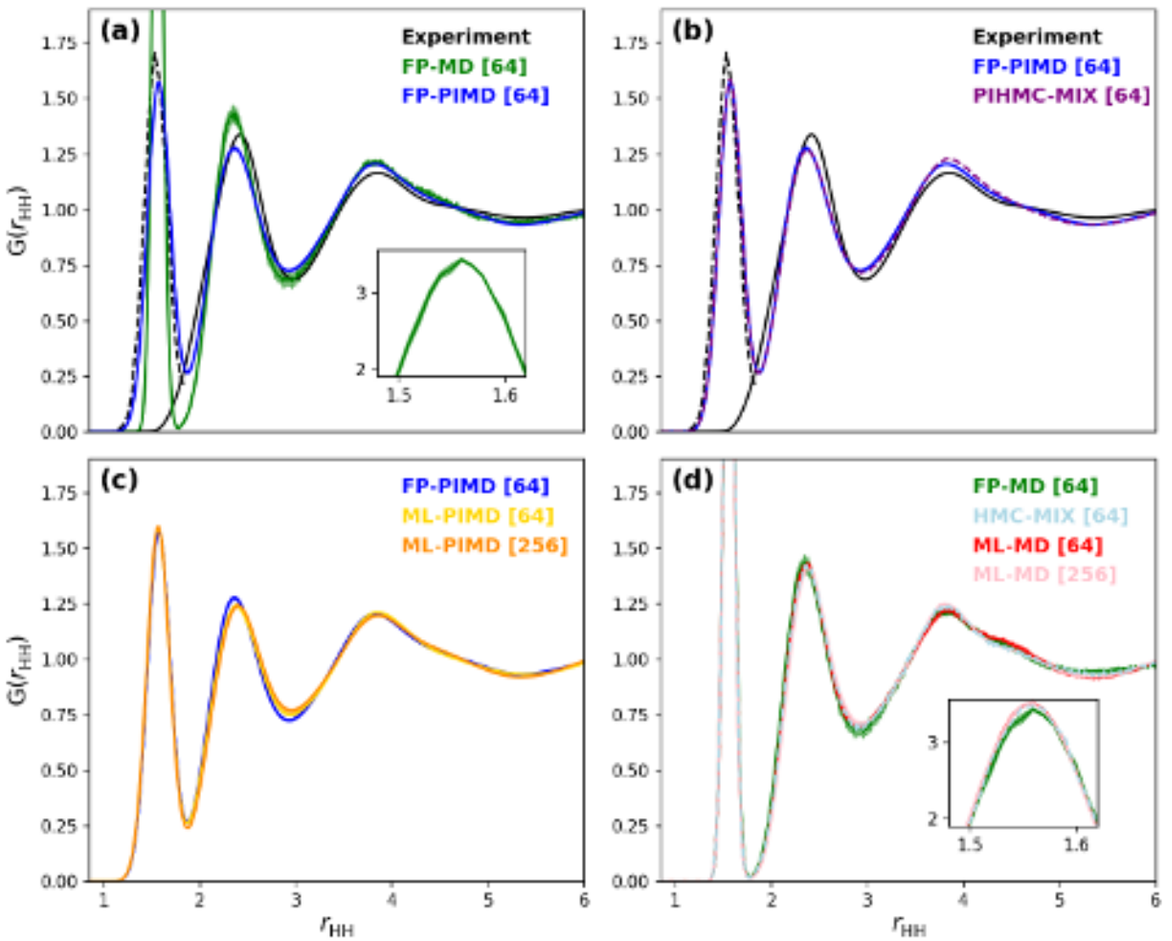

Figure 5, B. Thomsen, Y. Nagai, K. Kobayashi, I. Hamada and M. Shiga, submitted to J. Chem. Phys.

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0230464

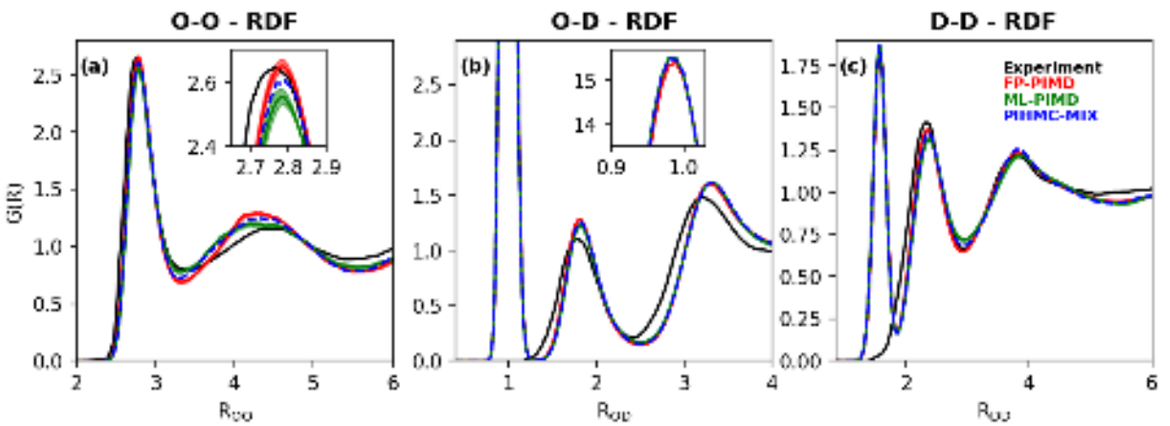

Figure 6, B. Thomsen, Y. Nagai, K. Kobayashi, I. Hamada and M. Shiga, submitted to J. Chem. Phys.

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0230464

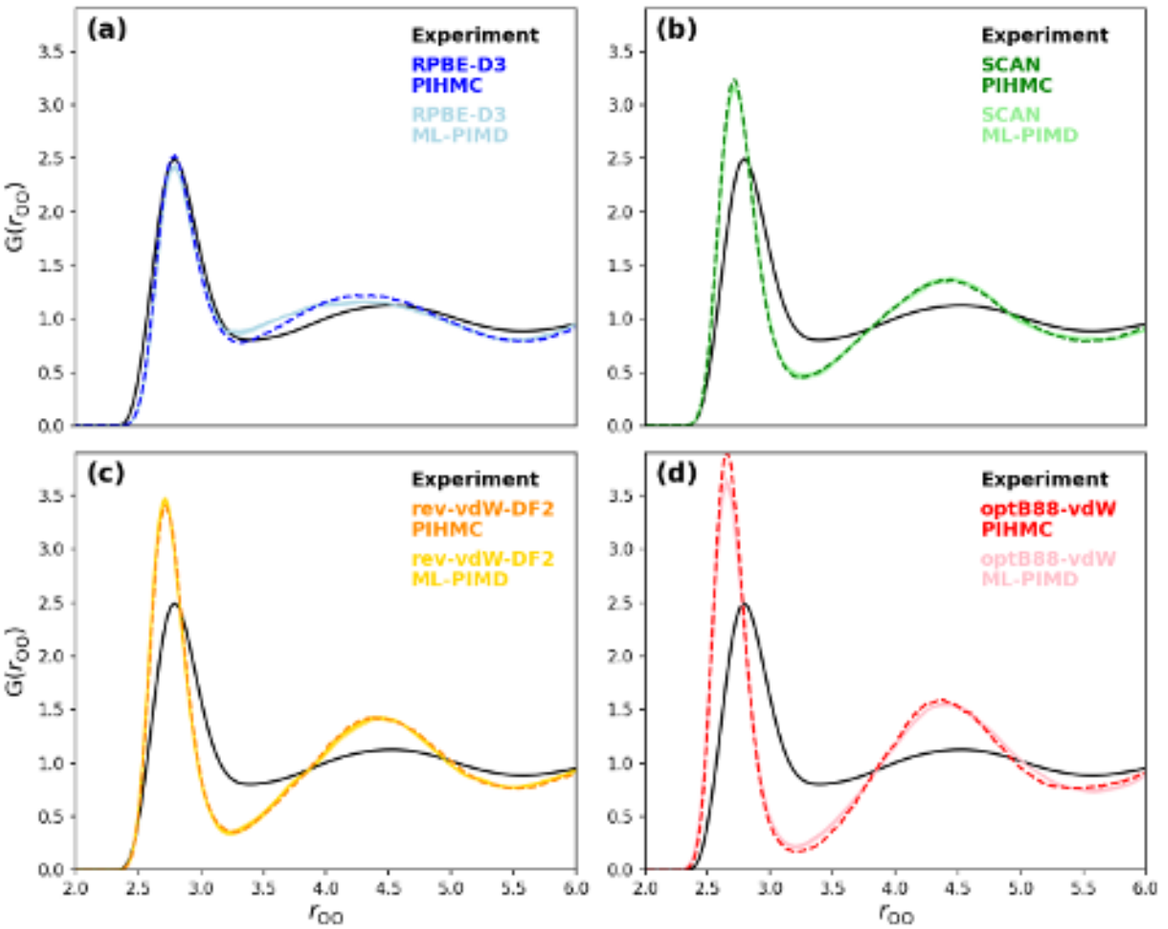

Figure 7, B. Thomsen, Y. Nagai, K. Kobayashi, I. Hamada and M. Shiga, submitted to J. Chem. Phys.

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0230464

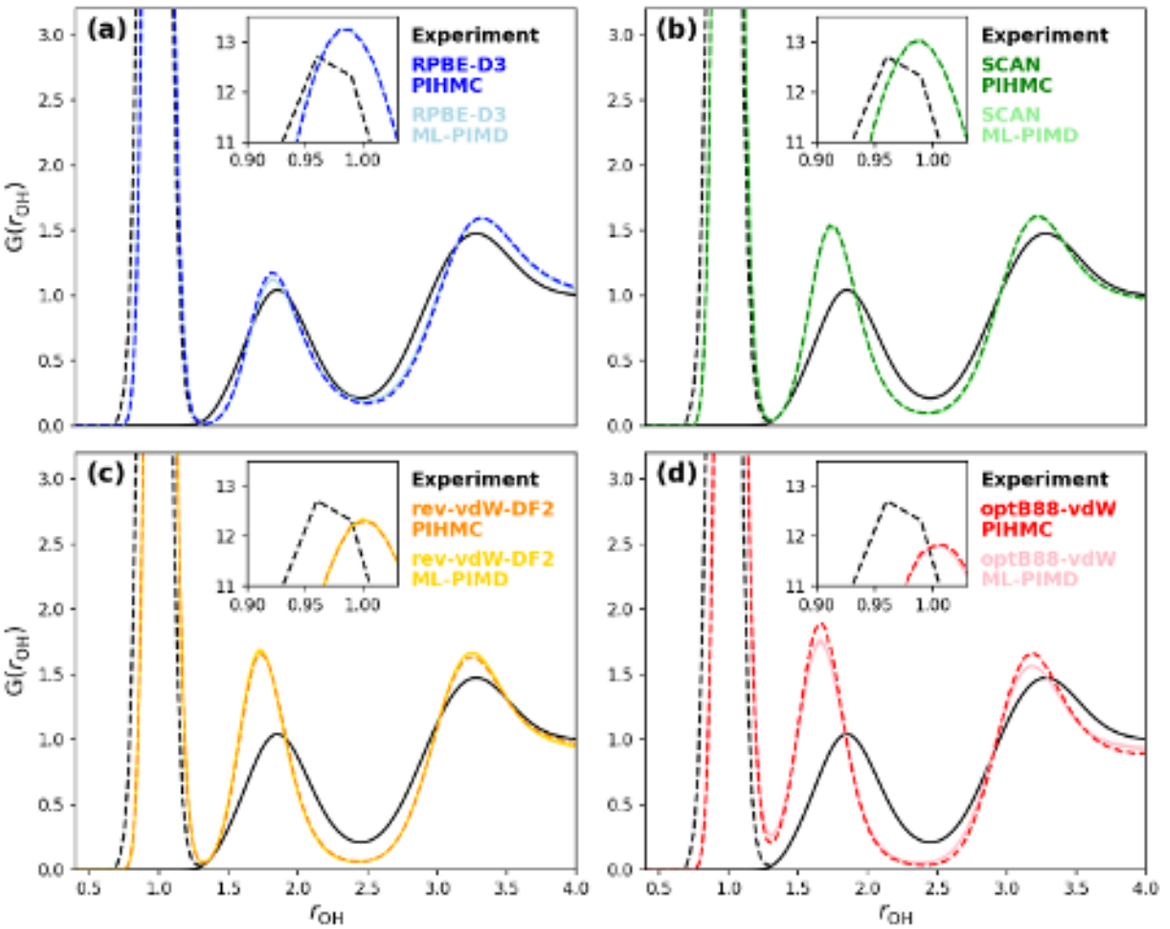

Figure 8, B. Thomsen, Y. Nagai, K. Kobayashi, I. Hamada and M. Shiga, submitted to J. Chem. Phys.

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0230464

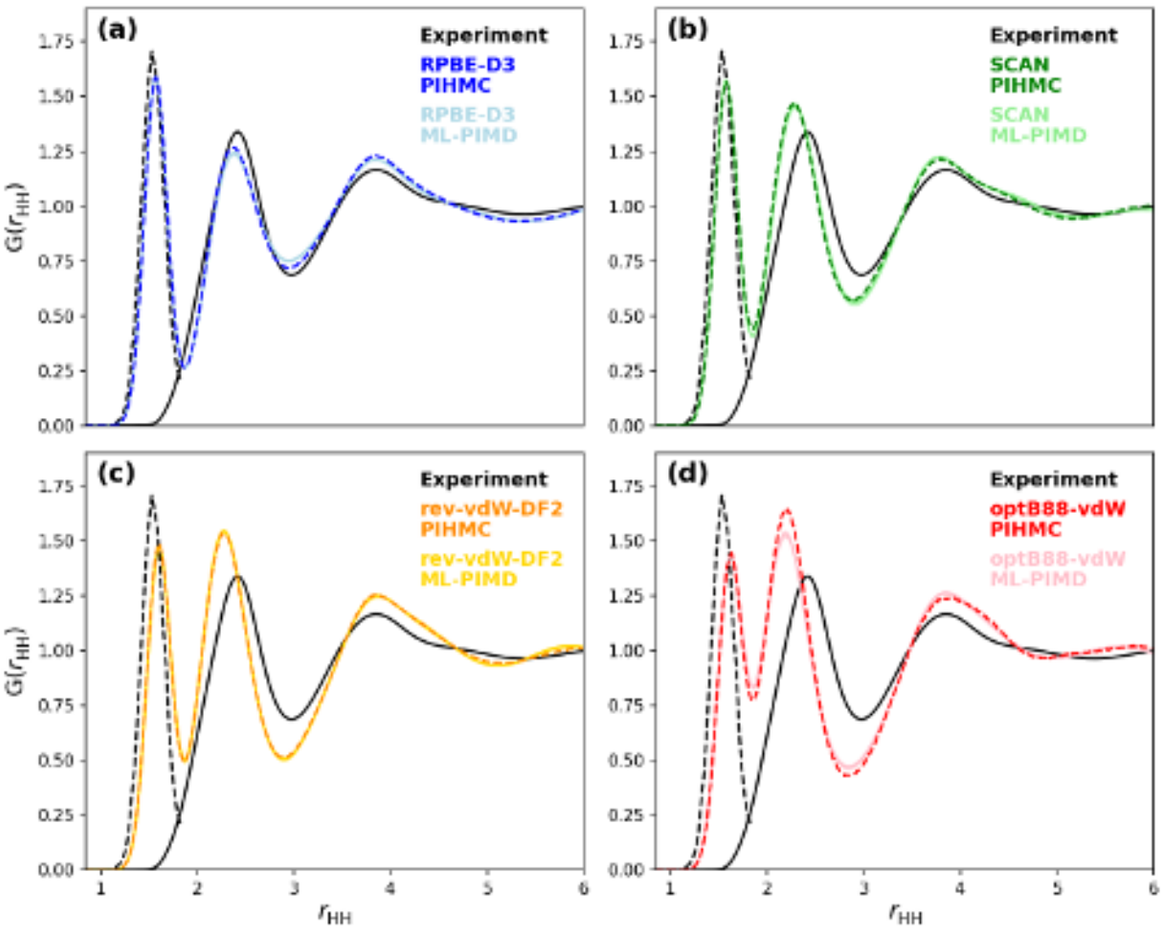

Figure 9, B. Thomsen, Y. Nagai, K. Kobayashi, I. Hamada and M. Shiga, submitted to J. Chem. Phys.

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0230464

TABLE I. The functionals used for the FP calculations, the method used for trajectory propagation, the number of steps, HMC or MD depending on the method, (N_{steps}), the α value used for PIHMC-MIX, the average acceptance ratio ($\langle A_{\text{test}} \rangle$), the number of steps in ML-PIMD for the PIHMC-MIX method (n_{ML}), and the effective trajectory length (t_{eff}) in picoseconds for all simulations presented in the main text. See Table SII in the SM for the SL-PIHMC-MIX trajectories run to train the MLPs, and Table SIII for the additional PIHMC-MIX trajectories only used in the SM.

Functional	Method	N_{steps}	α	$\langle A_{\text{acc}} \rangle$ (%)	n_{ML}	t_{eff} (ps)
RPBE-D3[a]	FP-MD	200,000	-	-	-	50.0
RPBE-D3	HMC-MIX	10,000	0.25	55.3	128	103.7
RPBE-D3[a]	FP-PIMD	100,000	-	-	-	25.0
RPBE-D3	PIHMC	15,000	1.0	24.8	8-128	17.9
RPBE-D3	PIHMC-MIX	10,000	0.75	25.9	8-128	22.3
RPBE-D3	PIHMC-MIX	7,000	0.5	31.5	16-128	62.4
RPBE-D3	PIHMC-MIX	5,000	0.25	55.5	128	99.9
SCAN	PIHMC-MIX	5,000	0.25	54.4	128	96.8
rev-vdW-DF2	PIHMC-MIX	5,000	0.25	51.7	64-128	88.7
optB88-vdW	PIHMC-MIX	5,000	0.25	36.8	128	59.0

[a] These trajectories are from Refs. 7 and 92

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0230464

TABLE II. The positions and heights of the peaks in the O-O RDFs presented in Figure 7. The data is denoted either with a r_i^{OO} or h_i^{OO} referring to the peak position and heights respectively. $r_{\mathrm{min}}^{\mathrm{OO}}$ and $h_{\mathrm{min}}^{\mathrm{OO}}$ refer to the height and position of the minimum of RDF found in the first interstitial region. The experimental results stem from Ref. 4.

DFT Functional	Model	r_1^{OO} (Å)	h_1^{OO}	$r_{\mathrm{min}}^{\mathrm{OO}}$ (Å)	$h_{\mathrm{min}}^{\mathrm{OO}}$	r_2^{OO} (Å)	h_2^{OO}
RPBE-D3	FP-PIMD	2.78	2.47	3.33	0.83	4.35	1.19
RPBE-D3	PIHMC	2.79	2.53	3.33	0.77	4.24	1.22
SCAN	PIHMC	2.72	3.24	3.23	0.44	4.36	1.36
rev-vdw-DF2	PIHMC	2.72	3.43	3.23	0.36	4.46	1.43
optB88-vdW	PIHMC	2.65	3.88	3.20	0.17	4.36	1.58
	Experiment	2.79	2.50	3.36	0.78	4.53	1.12

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0230464

TABLE III. The positions and heights of the peaks in the O-H RDFs presented in figure 8. All peak positions are given in Å. The data is denoted either with a r_i^{OH} or h_i^{OH} referring to the peak position and heights respectively. The experimental results stem from Ref. 4, except those marked by * which are from Ref. 93.

DFT Functional	Model	r_1^{OH} (Å)	h_1^{OH}	r_2^{OH} (Å)	h_2^{OH}	r_3^{OH} (Å)	h_3^{OH}
RPBE-D3	FP-PIMD	0.99	13.19	1.81	1.15	3.32	1.60
RPBE-D3	PIHMC	0.99	13.22	1.81	1.18	3.32	1.59
SCAN	PIHMC	0.99	13.04	1.75	1.54	3.24	1.61
rev-vdw-DF2	PIHMC	1.00	12.33	1.74	1.66	3.26	1.63
optB88-vdW	PIHMC	1.01	11.82	1.66	1.89	3.19	1.66
	Experiment	0.96*	12.71*	1.86	1.04	3.27	1.48

ACCEPTED MANUSCRIPT
The Journal of Chemical Physics
AIP Publishing
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0230464

TABLE IV. The positions and heights of the peaks in the H-H RDFs presented in figure 9. All peak positions are given in Å. The data is denoted either with a r_i^{HH} or h_i^{HH} referring to the peak position and heights respectively. The experimental results stem from Ref. 4, except those marked by * which are from Ref. 93.

DFT Functional	Model	r_1^{HH} (Å)	h_1^{HH}	r_2^{HH} (Å)	h_2^{HH}	r_3^{HH} (Å)	h_3^{HH}
RPBE-D3	FP-PIMD	1.57	1.57	2.36	1.28	3.83	1.21
RPBE-D3	PIHMC	1.57	1.58	2.36	1.27	3.84	1.23
SCAN	PIHMC	1.57	1.56	2.28	1.47	3.80	1.21
rev-vdW-DF2	PIHMC	1.60	1.49	2.25	1.53	3.85	1.25
optB88-vdW	PIHMC	1.63	1.44	2.21	1.64	3.89	1.24
	Experiment	1.53*	1.71*	2.43	1.34	3.84	1.17