JUSTC

Article

http://justc.ustc.edu.cn

Received: February 27, 2024; Accepted: April 01, 2024

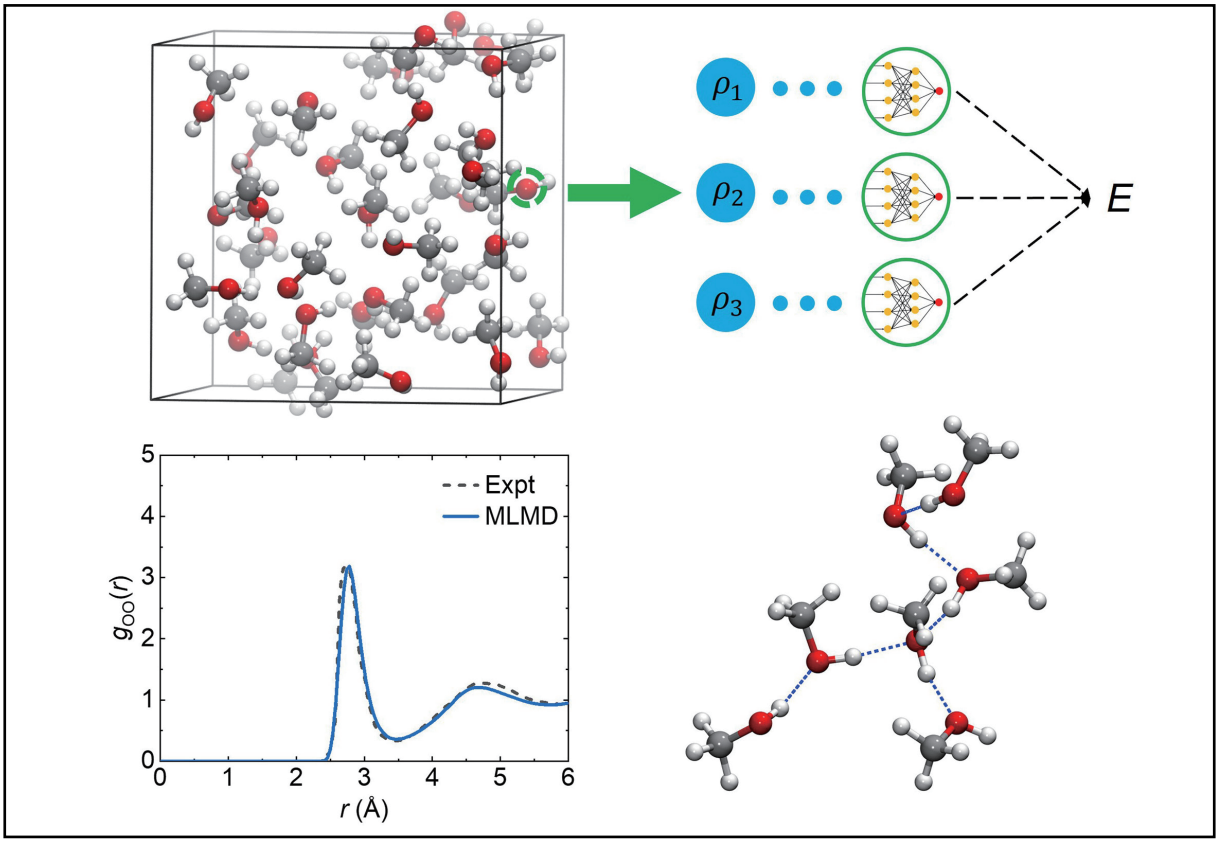
Machine learning molecular dynamics simulations of liquid methanol

Jie Qian, Junfan Xia, and Bin Jiang ✉

Hefei National Research Center for Physical Sciences at the Microscale, Department of Chemical Physics, University of Science and Technology of China, Hefei 230026, China

✉Correspondence: Bin Jiang, E-mail: bjiangch@ustc.edu.cn

Graphical abstract



Simulated liquid methanol by machine learning force field.

Public summary

- We develop a machine learning force field for liquid methanol at the level of hybrid functional revPBE0 plus dispersion correction.
- Machine learning molecular dynamics simulations are orders of magnitude faster than ab initio molecular dynamics simulations.
- Our machine learning force field predicts the radial distribution functions, self-diffusion coefficients and hydrogen bonding features in reasonably good agreement with the experimental data.

Citation: Qian J, Xia J F, Jiang B. Machine learning molecular dynamics simulations of liquid methanol. *JUSTC*, **2024**, 54(6): 0603. DOI: 10.52396/JUSTC-2024-0031

JUSTC

Article

http://justc.ustc.edu.cn

Received: February 27, 2024; Accepted: April 01, 2024

Machine learning molecular dynamics simulations of liquid methanol

Jie Qian, Junfan Xia, and Bin Jiang ✉

Hefei National Research Center for Physical Sciences at the Microscale, Department of Chemical Physics, University of Science and Technology of China, Hefei 230026, China

✉Correspondence: Bin Jiang, E-mail: bjiangch@ustc.edu.cn

Cite This: *JUSTC*, **2024**, 54(6): 0603 (10pp)

Read Online

SI Supporting Information

Abstract: As the simplest hydrogen-bonded alcohol, liquid methanol has attracted intensive experimental and theoretical interest. However, theoretical investigations on this system have primarily relied on empirical intermolecular force fields or ab initio molecular dynamics with semilocal density functionals. Inspired by recent studies on bulk water using increasingly accurate machine learning force fields, we report a new machine learning force field for liquid methanol with a hybrid functional revPBE0 plus dispersion correction. Molecular dynamics simulations on this machine learning force field are orders of magnitude faster than ab initio molecular dynamics simulations, yielding the radial distribution functions, self-diffusion coefficients, and hydrogen bond network properties with very small statistical errors. The resulting structural and dynamical properties are compared well with the experimental data, demonstrating the superior accuracy of this machine learning force field. This work represents a successful step toward a first-principles description of this benchmark system and showcases the general applicability of the machine learning force field in studying liquid systems.

Keywords: liquid methanol; molecular dynamics; machine learning; hydrogen bond; force field

CLC number: O645.16; TP181 **Document code:** A

1 Introduction

Methanol (CH_3OH), the simplest alcohol, is widely employed as a reagent, solvent, and chemical raw material. Due to its high energy density and ease of transportation, methanol has also emerged in recent years as a promising next-generation energy storage material[1,2]. An in-depth understanding of the structure and dynamics of liquid methanol is therefore highly desirable. Intermolecular hydrogen bonds are pivotal features in water and liquid alcohol such as methanol and largely determine the corresponding molecular chain structure. In contrast to bulk water, where each water molecule features four hydrogen bonds (H-bonds) with neighboring molecules, the presence of a hydrophobic methyl (CH_3) group in methanol reduces the number of available hydrogen bonding sites and thus the average number of H-bonds on each methanol molecule[3,4]. Furthermore, methanol molecules are known to be less polar than water molecules with generally weaker H-bonds. The asymmetry of the molecular structure, amphiphilicity, and reduced number of H-bonds contribute to the distinct properties and behavior of liquid methanol compared to those of liquid water.

Over the past few decades, extensive experiments (such as neutron diffraction[3–5], X-ray diffraction[6], IR[7], Raman[8] and ultrafast time-resolved spectroscopy[9–11]) have been conducted to unravel the diverse structural, dynamical, and spectroscopic properties of liquid methanol. Theoretically, molecular dynamics (MD) simulations have also been widely applied to understand these experimental observations. Such MD simulations have often been performed with classical force fields (CFFs) parameterized with physically inspired empirical functions for universal liquid systems, including the widely used OPLS[12] and its variants[13–17] or some universal force fields designed for organic molecules[18]. The pioneering OPLS force field, which was originally proposed by Jorgensen[12], was subsequently adjusted by Haughney et al.[19] to reproduce experimental data for various properties of methanol. More recently, Gonzalez-Salgado and Vega[20] incorporated solid-phase properties and solid–liquid equilibrium to refine the OPLS/2016 force field, which outperformed earlier versions. These CFFs were applied to study pure liquid methanol, methanol solutions[21], and methanol-water mixed solutions[22], achieving satisfactory agreement with some experiments. Despite their super efficiency, CFFs often assume molecular rigidity and adopt the harmonic approximation, thus lose reliability in heavily distorted molecular structures and bond forming/breaking processes.

On the other hand, ab initio molecular dynamics (AIMD) simulations based on on-the-fly electronic structure calculations at the density functional theory (DFT) level are in principle more suitable for studying complex systems from first principles. Since the pioneering investigation by Tsuchida et al.[23], a great deal of AIMD studies on methanol and related systems have been reported[24–33]. For example, Pagliai et al.[24] employed the Car–Parrinello molecular dynamics (CPMD) approach to investigate the hydrogen bond dynamics, molecular dipole moment and infrared spectra of deuterated methanol.

DOI: 10.52396/JUSTC-2024-0031

JUSTC, **2024**, 54(6): 0603

Their results showed good agreement with the experimental data. Sieffert et al.[28] compared three density functionals with or without dispersion corrections in Born–Oppenheimer MD (BOMD) for predicting multiple properties. It was argued that the B97[34] functional outperforms the widely used BLYP[35,36] functional for methanol. Additionally, Saija and coworkers[30,33] investigated the changes in the structural and spectroscopic properties of liquid methanol under an external electric field. Although the AIMD method is powerful, its huge computational cost limits the simulation length and the system size, hindering its applicability for predicting certain properties. For instance, an accurate calculation of self-diffusion coefficients typically requires a simulation time of at least several hundred picoseconds[37]. Moreover, these AIMD studies were limited to using generalized gradient approximation (GGA)-based density functionals.

In recent years, the rapid development of machine learning force fields (MLFFs) has revolutionized the field of molecular modeling, allowing accurate and efficient calculations of thermodynamics[38], spectroscopy[39–41] and chemical reactions[42]. By learning the relationship between structural descriptors and energies (sometimes including other properties) with highly flexible analytical functions, for example, neural networks (NNs), MLFFs can be used to faithfully replace electronic structure calculations and speed up AIMD simulations by several orders of magnitude while retaining first-principles accuracy. In this respect, many MLFFs have been applied to study liquid water[43–50], which has been regarded as a prototypical system that showcases the power of MLFFs. Surprisingly, there are very few MLFF applications to liquid methanol[51–54]. In fact, methanol is a common system for validating many universal CFFs[17,55,56]. Given the similarity and dissimilarity between methanol and water, as well as the abundant experimental data available, the former should serve as another important benchmark for evaluating MLFFs in condensed-phase systems. Although some MLFFs have been attempted for methanol clusters[51–53], Maldonado et al.[54] reported a many-body gradient-domain machine learning (mbGDML) potential from methanol clusters to bulk methanol at the MP2 level. However, the simulated radial distribution functions of liquid methanol deviate largely from the experimental data.

In the present work, we report a new MLFF for liquid methanol. Inspired by successful studies on the ab initio thermodynamics of water[57,58], we carried out electronic structure calculations with a hybrid functional revPBE0 plus dispersion correction instead of the GGA-based functional that was commonly used in previous work for liquid methanol. MD simulations of this MLFF (hereafter referred to as machine learning MD or MLMD) yield converged radial distribution functions, self-diffusion coefficients, and H-bond properties. The remainder of this article is organized as follows. The next section describes the methodology, including the computational details of DFT, the neural network approach, data sampling, and MD simulations. Subsequently, the structural and dynamic properties are compared with the experimental data and previous theoretical results in Section 3. The final section concludes.

2 Method

2.1 Density functional theory calculations

The simulated methanol system consists of 32 CH_3OH molecules in a periodically repeated cubic box with a length of L = 12.93 Å, as shown in Fig. 1a, which corresponds to the experimental density of 0.787 g·cm^{-3} at room temperature. All DFT calculations were conducted using the Quickstep[59] module in the CP2K 7.1[60] package, which employs a mixed Gaussian and plane wave (GPW)[61] basis set. Following recent studies for water, we used a hybrid revPBE0[62] functional that contains 25% exact exchange and included the Grimme D3 correction[63]. This revPBE0 functional was applied in AIMD and MLMD simulations for water that reproduce well the experimental RDFs[64], self-diffusion coefficient[64], and thermodynamics[57]. Goedecker−Teter−Hutter (GTH) pseudopotentials[65–67] were used to describe the core electrons. A molecularly optimized triple-ζ Gaussian basis set augmented with two polarization functions (TZV2P-MOLOPT)[68] was employed to expand the Kohn−Sham orbitals. The plane wave expansion was truncated at 450 Ry. To accelerate the computations of hybrid functionals in periodic systems, we utilized the auxiliary density matrix method (ADMM)[69]. Additionally, we chose the orbital transformation (OT)[70] method to optimize the wave function at each

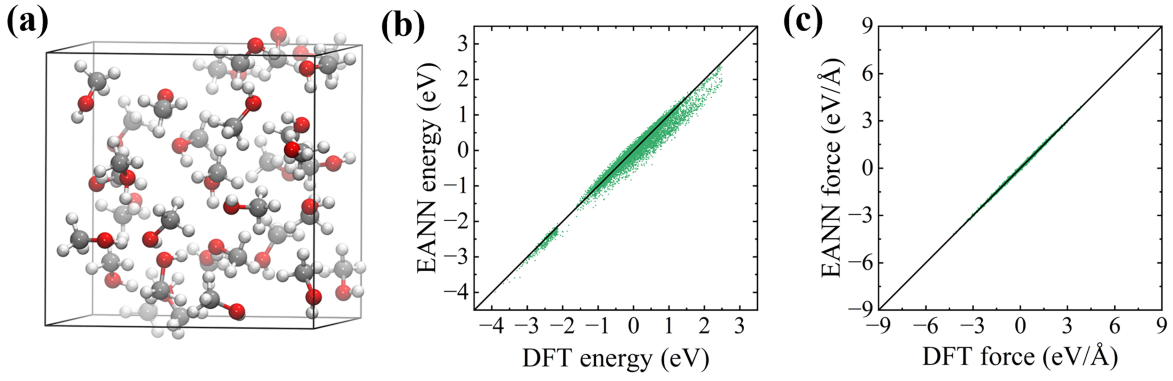


Fig. 1. (a) A snapshot of the simulation cell of liquid methanol containing 32 CH_3OH molecules. (b) Potential energies and (c) forces obtained from the comparison of the EANN potential and DFT results. The energy zero is defined as the mean potential energy of the AIMD trajectory.

DOI: 10.52396/JUSTC-2024-0031
JUSTC, **2024**, 54(6): 0603

step using a self-consistent field (SCF) convergence criterion of 1×10^{-7} a.u.

2.2 Embedded atom neural network approach

The calculated DFT energies and forces were trained to an analytical MLFF using the embedded atom neural network (EANN) approach[48,71] implemented in the open-source REANN package[72]. This approach is based on the atomistic neural network architecture and is generally applicable to molecular, condensed-phase and interfacial systems[73]. It decomposes the total energy of a system into the sum of individual atomic contributions, and each atomic energy (E_i) is dependent on its local environment, which is determined by a series of embedded atom density (EAD) features (ρ_i). The EAD feature is formally expressed as the square of a linear combination of atomic orbitals. For efficiency, these atomic orbitals are practically contracted Gaussian functions expanded in terms of Gaussian primitives ($\varphi^m_{l_x,l_y,l_z}(\hat{r}_{ij})$), leading to

$$\rho_i = \sum\nolimits_{l_x+l_y+l_z=L} \frac{L!}{l_x!l_y!l_z!}\left[\sum_{j\neq i}^{N_c} c_j \sum_{m=1}^{n_{\text{wave}}} d^m_n \varphi^m_{l_x,l_y,l_z}(\hat{r}_{ij}) f_c(r_{ij})\right]^2, \quad (1)$$

where $\hat{r}_{ij}$ represents the vector pointing from the central atom i to the neighboring atom j, r_{ij} is its norm, N_c is the number of neighboring atoms within the cutoff radius (r_c), c_j is an element-dependent orbital coefficient, d^m_n is the contraction coefficient of a Gaussian primitive, and n_{wave} is the total number of contracted Gaussian functions. The EAD feature continuously decays to 0 at r_c, which is ensured by the cosine cutoff function $f_c(r_{ij})$. The Gaussian primitive is defined by its center (r_s), width (σ) and angular momentum ($L = l_x + l_y + l_z$),

$$\varphi^m_{l_x,l_y,l_z}(\hat{r}_{ij}) = (x_{ij})^{l_x}(y_{ij})^{l_y}(z_{ij})^{l_z}\exp\left(-\frac{|r_{ij}-r_s|^2}{2\sigma^2}\right), \quad (2)$$

where L= 0, 1, or 2, resembling the s, p, or d orbitals, respectively. In practice, we chose r_c = 6 Å, n_{wave} = 8, L =0, 1, and 2, yielding 84 EAD features in total. Each atomic neural network contains two hidden layers with 32 neurons in each layer. Once the EANN architecture is determined, all the hyperparameters are optimized during the training process along with the NN parameters, leading to an end-to-end representation.

2.3 Details of the MLFF construction

The dataset for the MLFF construction was generated as follows. Most of the data (15000 configurations) were sampled from an AIMD trajectory of 30 ps (detailed in the next section) every 2 fs and thermostated at 300 K. Five hundred points from these AIMD data were randomly selected as seeds to quickly train an initial EANN potential. This potential was then used in an efficient and iterative sampling procedure, aiming to find possible "holes" in the configuration space of dynamical interest and add new points to fix them. To this end, MLMD simulations on this EANN potential were run with LAMMPS[74] to produce an extensive set of candidate points. We trained two extra EANN potentials (with different initial parameters) to estimate the uncertainty of these candidates. Because any MLFF tends to provide a reliable prediction at the position where training data exist, a large energy difference (ΔE) given by the two EANN potentials for a candidate point indicates that the prediction in this region is not good enough. This point should then be added into the training dataset. Based on this concept, in each iteration, the ten points with the largest energy differences were selected for additional DFT calculations to augment the dataset. New EANN potentials were trained with the updated dataset to repeat this procedure until no new data could be selected from existing MLMD trajectories within our criterion (ΔE > 0.1 eV). New MLMD simulations were then performed on the latest EANN potential to keep the data selection running. This treatment minimizes the effort to run MLMD simulations in the iterative data selection. We collected ~3700 points in this way to be merged into the preselected AIMD dataset, yielding a full set of 18751 points. To train the final MLFF for property predictions, the full dataset was randomly divided into a training set and a validation set at a ratio of 9∶1. The weighting ratio of the energy and force was initially set to 1∶100 and dynamically adjusted during the training process to 1∶1.

2.4 AIMD and MLMD setups

All MD simulations were conducted under the canonical ensemble (NVT) under ambient conditions (T = 300 K, P = 1 bar) with periodic boundary conditions. The equations of motion were integrated with a time step of 0.5 fs. A Nosé-Hoover chain[75,76] thermostat with a chain length of 3 and a time constant of 500 fs was employed to control the temperature. AIMD was performed using CP2K, in which the equilibration run was imposed by ~30 ps, followed by a 20 ps production run for trajectory analysis. MLMD simulations were performed using the LAMMPS package. The system was first equilibrated for 50 ps, followed by subsequent 400 ps for the evaluation of the time correlation functions. To reduce the statistical uncertainties and explore the phase space more thoroughly, five independent simulations were carried out, resulting in a cumulative trajectory duration of 2 ns. This enables much better statistics of MLMD results compared to AIMD results. Postprocessing of the RDFs and mean square displacement (MSD) were performed by Trajectory Analyzer and Visualizer (TRAVIS)[77,78] software. The hydrogen bond behavior was analyzed by an in-house Fortran code. Molecular visualization was implemented by Visual Molecular Dynamics (VMD)[79] program.

3 Results and discussion

3.1 Validation of the MLFF

Let us first check the accuracy of the EANN potential. Fig. 1b and c compare the predictions of the final EANN potential for energies and atomic forces over the entire dataset with DFT values. Here, the energy zero is defined by the mean potential energy of the AIMD trajectory. The root-mean-square-errors (RMSEs) of the EANN potential for the training/validation set are 0.77/0.78 meV per atom for energies and 32.8/32.9 meV·Å^{-1} for forces. In addition, we selected another 1000 configurations by further propagating the AIMD tra-

DOI: 10.52396/JUSTC-2024-0031
JUSTC, **2024**, 54(6): 0603

jectory for an additional 2 ps (every 2 fs), which was not used in the training process, as an independent test set. The test RMSEs for energies and atomic forces are 0.39 meV per atom and 33.2 meV·Å^{-1}, respectively, which are comparable to the training errors. This finding verifies the accuracy of the EANN potential in subsequent MLMD simulations.

3.2 Radial distribution functions

Radial distribution functions (RDFs) are a statistical characterization of the complex structure of condensed phase materials. Each peak in the RDF corresponds to an average distance between two elements of a collection of solvent shell structures. For liquid methanol, there are six RDFs for O–O (g_{OO}), O–H (g_{OH}), H–H (g_{HH}), C–C (g_{CC}), C–O (g_{CO}), and C–H (g_{CH}), which were calculated from AIMD and MLMD simulations and are compared in Fig. 2. In comparison, the AIMD and MLMD results are in excellent agreement with one another, again validating the reliability of the EANN potential. The slight deviations are most likely due to the relatively short simulation time of AIMD compared to that of MLMD simulations. The experimental RDFs derived from neutron diffraction data are also compared in Fig. 2. Specifically, the agreement between our MLMD and experimental results is excellent for the O–O RDF and reasonably good for the C–O RDF with respect to the peak position and intensity. However, in theoretical RDFs involving hydrogen atoms, especially for g_{OH} and g_{HH}, the first peaks appear higher than the experimental values. This disagreement has been previously observed in classical MD simulations for liquid water, which was largely ascribed to nuclear quantum effects (NQEs) that make the hydrogen bond network less compact[80]. The absence of NQEs in current MLMD simulations is likely responsible for the overstructure of the nearest neighbor shell for methanol. Improvements can be made by performing path-integral molecular dynamics (PIMD) simulations on the MLFF, which will be discussed in a forthcoming work.

Notably, to diminish the influence of NQEs, previous AIMD studies have often been conducted on fully deuterated methanol (CD_3OD)[24,28]. Even so, as shown in Fig. 2, our MLMD results for g_{OH} and g_{HH} exhibit a prominent decrease in the first peaks compared to the corresponding AIMD results for g_{OD} and g_{DD} at the BLYP-D3 and B97-D2 levels[28]. This suggests that the GGA density functional does not accurately describe the structure of liquid methanol. On the other hand, the present EANN potential at the revPBE0-D3 level also significantly outperforms the mbGDML potential at the MP2 level with respect to O–O, O–H, and H–H RDFs. The latter results obviously underestimate the structures of the first hydrogen bonding shell, yielding much lower peaks than those of the experiment even without the inclusion of NQEs. Noticeably, the calculated C–C RDF shows the largest disagreement with the experimental RDF, which is narrower and sharper than the MLFF prediction. The same failure also appeared in previous MD results [26, 81], which underscores the great challenge in accurately describing the weak interaction between methyl groups in bulk methanol with DFT. A higher wave function-based level of ab initio calculations is probably needed. Moreover, the experimental site-site RDFs shown here were indirectly derived through an empirical potential structure refinement (EPSR) computer simulation of neutron diffraction data[3,4], which cannot avoid some arbitrariness. Further investigations are necessary to resolve this disagreement.

3.3 Self-diffusion coefficient

The self-diffusion coefficient is another important property of liquid systems. In theory, this quantity can be estimated by the mean square displacement (MSD) of the molecular center, which is related to the molecular diffusion constant (D) via the Einstein equation:

$$6D = \lim_{t\to\infty} \frac{d}{dt} \left\langle [r_i(t_0+t) - r_i(t_0)]^2 \right\rangle, \tag{3}$$

where $r_i(t)$ denotes the position of the centroid of a molecule i relative to the system center of mass at a correlation time t, while the brackets average over all possible time origins (t_0) and molecules.

To obtain a reliable estimate of D, we must choose an appropriate time interval where the slope of the MSD with respect to time becomes stable. Fig. 3 shows the MSDs of both the oxygen atom and the centroid of the methanol molecule as a function of the correlation time t along a long MLMD trajectory. After 1 ps, the two MSD curves are almost indistinguishable. Consequently, the oxygen atom in methanol is employed as the observing particle instead of the molecular centroid when evaluating self-diffusion coefficients, as in some previous studies[25]. Moreover, the diffusion of methanol molecules starts to follow the Einstein relationship, and the MSD becomes linearly proportional to the correlation time after 25 ps. A longer correlation time will enable better statistical analysis. Specifically, a time interval between 100 and 200 ps was chosen to evaluate the slope. The self-diffusion coefficient $D = (2.20 \pm 0.20) \times 10^{-9}\,\mathrm{m^2/s}$ was obtained by fitting Eq. (3) ($R^2 > 0.999$) and taking the average over five MLMD trajectories.

The calculated self-diffusion coefficient is compared with previous theoretical and experimental results in Table 1. As mentioned above, the calculation of the self-diffusion coefficient is more demanding than that of the RDF. Existing theoretical results were mainly based on MD simulations with CFFs. The current MLMD result agrees best with the experimental value at room temperature, outperforming MD results based on CFFs and the CPMD result at the BLYP level[25]. This demonstrates the high accuracy of the newly constructed MLFF.

3.4 Hydrogen bond analysis

The hydrogen bond network is a unique characteristic of condensed phase systems consisting of polar molecules containing hydrogen atoms. To quantify the formation of H-bonds, following Ref. [24], we employed three geometric criteria: (i) $r_{HO\cdots H} < 2.6$Å; (ii) $r_{O\cdots O} < 3.5$Å; and (iii) $\angle HO\cdots O < 30°$, as illustrated in Fig. 4a. Using these criteria, H-bond analysis was performed for each frame along all trajectories. Fig. 4b shows the spatial distribution of H-bonds as a function of $\angle HO\cdots O$ and $r_{HO\cdots H}$. Clearly, the formation of hydrogen bonds is focused in the range of 1.5 Å < $r_{HO\cdots H}$ < 2.2 Å and $\angle HO\cdots O$ < 30°. The distribution peaks are at $r_{HO\cdots H}$ = 1.82 Å and

DOI: 10.52396/JUSTC-2024-0031
JUSTC, 2024, 54(6): 0603

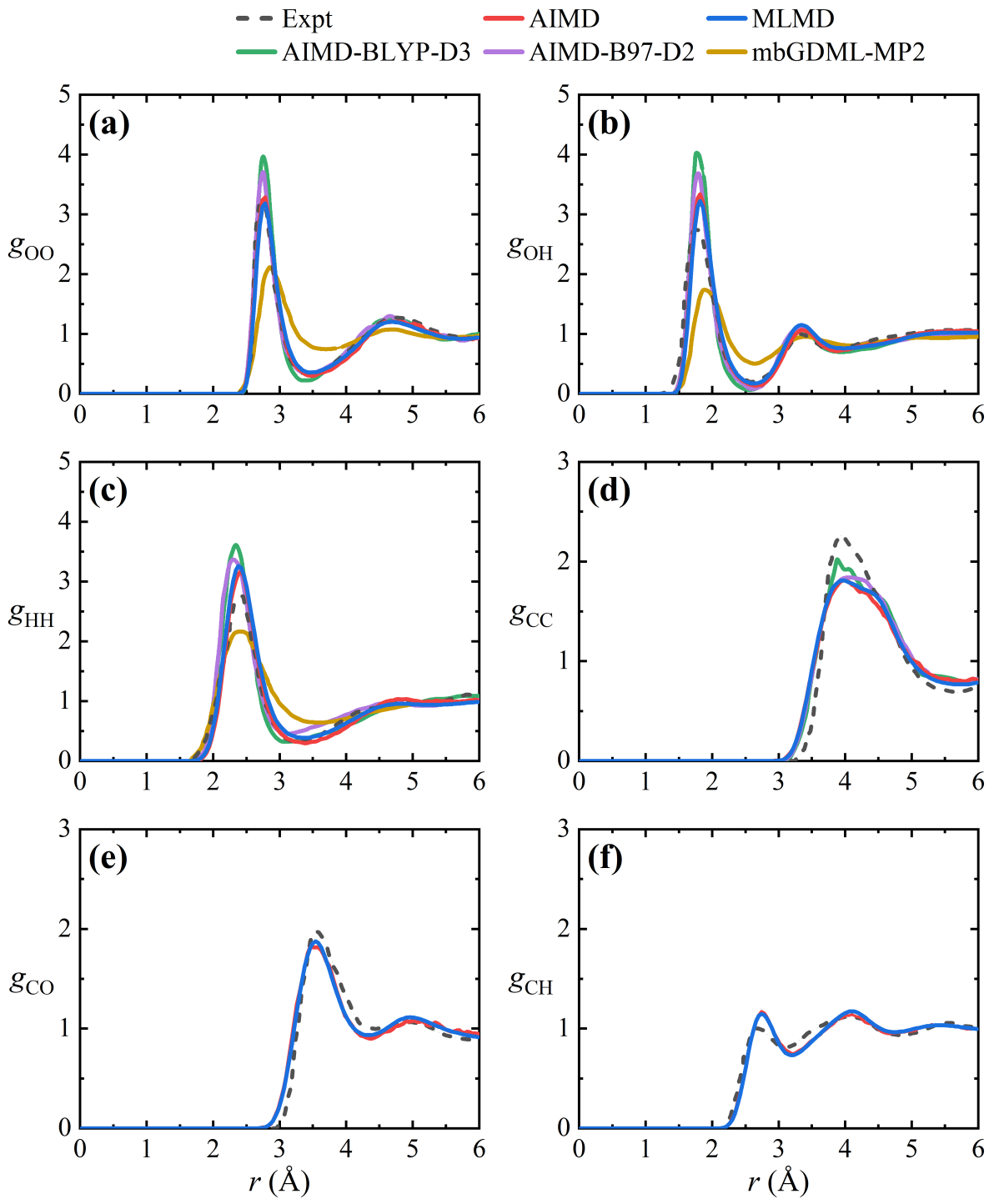


Fig. 2. Comparison of the calculated intermolecular radial distribution functions of (a) O–O, (b) O–H, (c) H–H, (d) C–C, (e) C–O, and (f) C–H in bulk methanol from current AIMD and MLMD simulations at the revPBE0-D3 level with the experimental data[3,4] at room temperature, which are taken from neutron diffraction results fitted by the empirical potential structure refinement (EPSR) computer simulation. Additionally, previous AIMD results at the BLYP-D3 and B97-D2 levels for deuterated methanol and mbGDML results at the MP2 level whenever available are shown.

∠HO···O= 8.8°, which is consistent with the physical picture that the hydrogen bond favors a nearly linear geometry and a more extended bond length than the covalent bond. This peak position also coincides with the first peak position of g_{OH} (~1.85 Å), corresponding to the closest neighbor shell in the H-bond network.

The average number of H-bonds per molecule $\langle n_{HB} \rangle$ is a metric of the average strength of hydrogen bonding interactions, which can be either estimated by the geometric criteria above or calculated by integrating the O–H RDF up to the

DOI: 10.52396/JUSTC-2024-0031
JUSTC, **2024**, 54(6): 0603

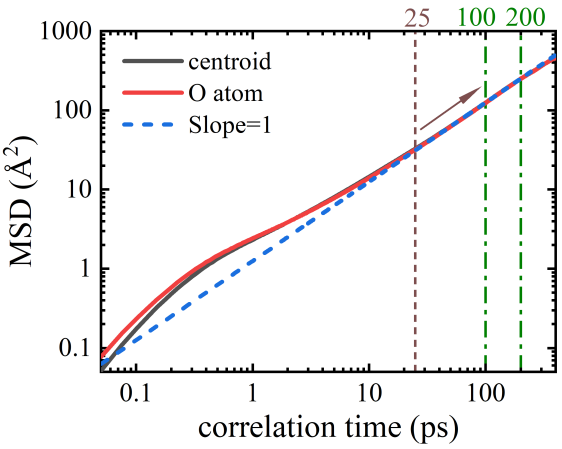


Fig. 3. A log-log plot of the MSD with respect to the correlation time of the O atom and the centroid of methanol was used to identify the "middle" region. The blue dashed line indicates a slope of 1 (target diffusive regime).

Table 1. Comparison of liquid methanol diffusion coefficients obtained from experiments and various theoretical studies.

Method	Diffusion coefficients (10^{-9} m^2/s)
Expt[82]	2.42 ± 0.05
CPMD-BLYP[25]	2.0 ± 0.6
AMOEBA[56]	1.9
OPLS/2016[20]	2.72
This work: MLMD-revPBE0+D3	2.20 ± 0.20

minimum between the first and second peaks. The two kinds of results are summarized in Table 2. Our results agree well with both the experimental and previous simulation results. Similar to previous findings[3,4,24], the geometry-defined $\langle n_{HB} \rangle$ falls in the range of 1.8–1.9, while the RDF-integrated $\langle n_{HB} \rangle$ is slightly larger within 1.9–2.0. This value does not seem to be very sensitive to the force field, indicating that the methanol molecule forms, on average, two hydrogen bonds in the liquid state.

In more detail, we count the fraction of molecules involving n hydrogen bonds f_n, n=0, 1, 2, 3. We find that the $f_0 : f_1 : f_2 : f_3$ ratio is 1.0 ∶ 7.8 ∶ 18.2 ∶ 5.0. The relative proportion of the number of H-bonds provides an overview of the intermolecular interactions. Although the majority of molecules form two H-bonds, a smaller portion of molecules form only one or three H-bonds. Fig. 5a shows an example of a hydrogen bond network from an MD frame, in which methanol molecules are connected with one, two, or three H-bonds. In particular, molecules with two H-bonds constitute a linear H-bond chain, while the third H-bond becomes a branch of the linear chain. Assuming that molecules at the end of the linear H-bond chain contain only one H-bond, given the ratio of $f_2 : f_1 \approx 18.2 : 7.8 \approx 4.7 : 2$, one can estimate that each linear chain contains ~6.7 methanol molecules on average. This value is close to the experimental estimate (5.5 ± 1.0) derived from neutron diffraction data[3,4].

Next, we turn to the dynamics of the H-bond network in liquid methanol. It is convenient to estimate the lifetime (τ_{HB}) of the H-bond by the corresponding H-bond autocorrelation function,

$$C_x(t) = \left\langle \frac{h_{ij}(t_0) h_{ij}(t_0 + t)}{h_{ij}(t_0)^2} \right\rangle, \quad (4)$$

where $h_{ij}(t_0) = 1$ when a donor molecule i forms a hydrogen bond with an acceptor molecule j at moment t_0 and $h_{ij}(t_0) = 0$ otherwise, $h_{ij}(t_0 + t) = 1$ indicates that these atoms remain hydrogen bonded throughout the period t_0 to $t_0 + t$. The angle bracket indicates the average over multiple possible time origins and all molecules in the MD trajectories. By fitting a bi-exponential to the time autocorrelation curve, as shown in Fig. 5b, we can obtain the lifetime of the hydrogen bond dynamics,

$$C_x(t) = A_1 \exp(-t/\tau_1) + A_2 \exp(-t/\tau_2), \quad (5)$$

where τ_1 and τ_2 represent the lifetimes of fast and slow H-bond evolution processes, respectively. The sum of A_1 and A_2 is 1, and their relative ratio indicates the percentage of the

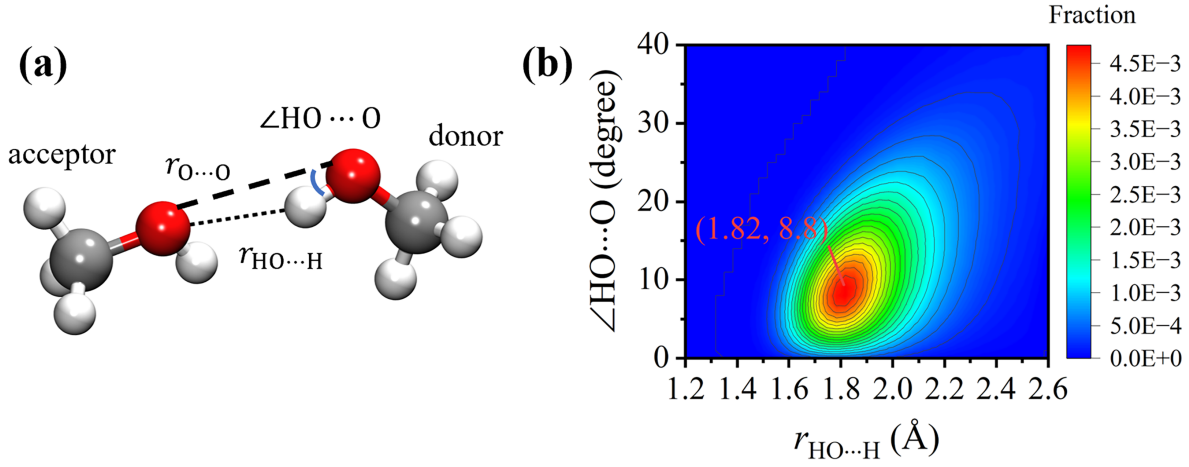


Fig. 4. (a) Definition of the geometric criteria for H-bonds in a methanol dimer. (b) H-bond density probability distribution as a function of ∠HO⋯O and $r_{HO\cdots H}$ obtained from all trajectories. The peak position is marked in red.

DOI: 10.52396/JUSTC-2024-0031
JUSTC, **2024**, 54(6): 0603

Table 2. Comparison of the average number of H-bonds per molecule ($\langle n_{HB} \rangle$) measured with X-ray and neutron diffraction and different theoretical predictions.

Method	$\langle n_{HB} \rangle^a$	$\langle n_{HB} \rangle^b$
X-ray[6]	—	1.80 ± 0.1
Neutron Diffraction[3,4]	1.77 ± 0.07	1.98
CPMD-BLYP[24]	1.89[c]	1.9[c]
CPMD-BLYP-D[83]	—	1.9[c]
CPMD-BLYP-D2[32]	1.8[d]	—
This work: MLMD-revPBE0-D3	1.84 ± 0.005	1.99

$\langle n_{HB} \rangle$ is obtained by either [a]geometric criteria or [b]integrating the g_{OH} curve up to the first minimum.

[c] Results are for deuterium methanol CD_3OD instead of CH_3OH.

[d] Geometric criteria imposed in Ref. [32] are slightly different.

corresponding dynamic process in the overall autocorrelation curve.

Our MLMD results are listed in Table 3. Our prediction gives an estimate of τ_1= 0.23 ps, which agrees well with the experimental results (~0.2 ps). This short lifetime corresponds to a fast process, which was attributed to the libration of the OH group in previous theoretical studies[24] and experiments[11]. This libration process was found to be independent of the H-bond chain length by Jukić et al.[84], who compared the dynamics of hydrogen bonds for water and alcohols with alkane chains of different lengths. τ_2, on the other hand, is a longer lifetime that corresponds to the breaking of the H-bond itself. Our prediction of τ_2 is somewhat smaller than that of the experiments but still on the order of the picosecond scale. The difference may be because methanol is not in a pure liquid phase in experiments but is diluted and dissolved in solvents[9–11].

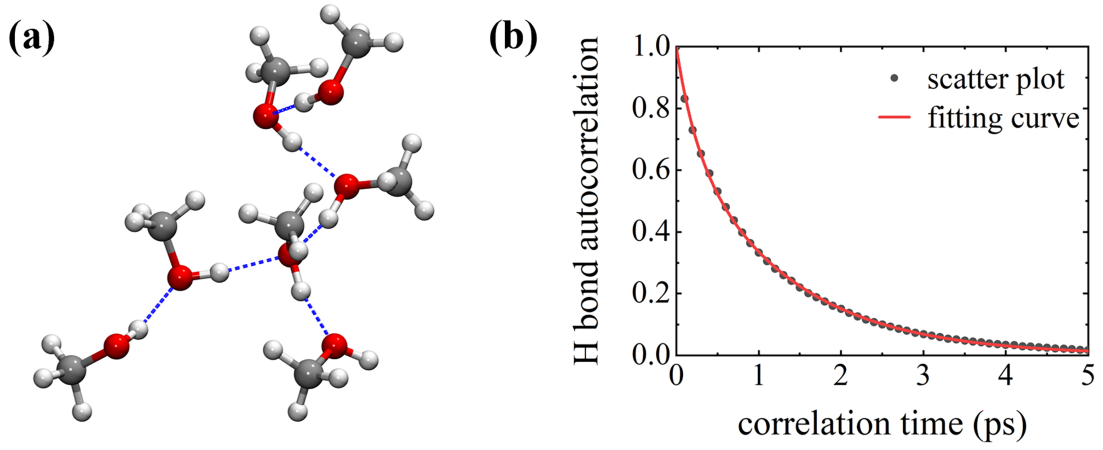


Fig. 5. (a) Snapshot of one chain of methanol molecules illustrating multiple types of H-bonds (blue dashed lines). (b) Scatter plot and fitted biexponential function curve of the H-bond autocorrelation obtained from MLMD.

4 Conclusions

To summarize, we construct an accurate MLFF for liquid methanol by fitting more than 18 thousand data points at the revPBE0-D3 level to an atomistic neural network representation. This MLFF enables sufficiently long molecular dynamics simulations that are necessary to predict the structural and dynamical properties of liquid methanol with good statistics. Specifically, a variety of radial distribution functions, self-diffusion coefficients, and hydrogen bond networks are obtained in comparison with experimental and previous theoretical results. This MLFF generally exhibited better agreement with the experimental data than previous AIMD studies at the GGA density functional level. This work validates the ability of the MLFF to model liquid systems other than water, representing the first step toward a first-principles description of this benchmark system. However, there are still some discrepancies in comparison with the experimental results, especially for the radial distribution functions. These discrepancies are mainly due to two factors. First, the DFT training data at the revPBE0-D3 level used here may not be sufficiently accurate for describing all interatomic interactions. Second, nuclear quantum effects (NQEs), which likely affect most of the properties relevant to the motion of hydrogen atoms, are neglected in our classical MD simulations. The density functional dependence of the MLMD results and the influence of NQEs will be investigated in the future.

Table 3. Comparison of the average lifetimes of H-bonds obtained from the present MLMD results with previous experimental and theoretical results.

Method	τ_1 (ps)	τ_2 (ps)
Expt[9]	~ 0.2	~ 2
Expt[11]	~ 0.2	~ 4
CPMD-BLYP[24]	0.5±0.1	1.9±0.1
This work: MLMD-revPBE0-D3	0.23±0.02	1.27±0.06

Acknowledgements

This work was supported by the CAS Project for Young Scientists in Basic Research (YSBR-005) and the National Natural Science Foundation of China (22325304, 22221003

DOI: 10.52396/JUSTC-2024-0031
JUSTC, **2024**, 54(6): 0603

and 22033007). We acknowledge the Supercomputing Center of USTC, Hefei Advanced Computing Center, Beijing PARATERA Tech Co., Ltd., for providing high-performance computing services.

Conflict of interest

The authors declare that they have no conflict of interest.

Biographies

Jie Qian received his B.S. degree in Chemistry from the University of Science and Technology of China (USTC) in 2020 and is working toward a master's degree at USTC. His research interests include machine learning force fields and relevant applications in condensed phase systems.

Bin Jiang is a Professor of Chemistry at the University of Science and Technology of China (USTC). He received his B.S. and Ph.D. degrees from Nanjing University and performed postdoctoral research at the University of New Mexico. He joined the USTC in 2015. His research interests focus on machine learning method development and applications to potential energy surfaces across gas-phase and gas-surface interfaces, as well as quantum/classical dynamics of gas-surface reactions.

References

[1] Alias M S, Kamarudin S K, Zainoodin A M, et al. Active direct methanol fuel cell: An overview. *Int. J. Hydrog. Energy*, **2020**, *45* (38): 19620–19641.
[2] Olabi A G, Onumaegbu C, Wilberforce T, et al. Critical review of energy storage systems. *Energy*, **2021**, *214*: 118987.
[3] Yamaguchi T, Hidaka K, Soper A K. The structure of liquid methanol revisited: a neutron diffraction experiment at −80 °C and +25 °C. *Mol. Phys*, **1999**, *96* (8): 1159–1168.
[4] Yamaguchi T, Hidaka K, Soper A K. ERRATUM: The structure of liquid methanol revisited: a neutron diffraction experiment at −80 °C and +25 °C. *Mol. Phys*, **1999**, *97* (4): 603–605.
[5] Adya A K, Bianchi L, Wormald C J. The structure of liquid methanol by H/D substitution technique of neutron diffraction. *J. Chem. Phys.*, **2000**, *112* (9): 4231–4241.
[6] Narten A H, Habenschuss A. Hydrogen bonding in liquid methanol and ethanol determined by x-ray diffraction. *J. Chem. Phys.*, **1984**, *80* (7): 3387–3391.
[7] Falk M, Whalley E. Infrared spectra of methanol and deuterated methanols in gas, liquid, and solid phases. *J. Chem. Phys.*, **1961**, *34* (5): 1554–1568.
[8] Lin K, Zhou X G, Luo Y, et al. The microscopic structure of liquid methanol from Raman spectroscopy. *J. Phys. Chem. B*, **2010**, *114* (10): 3567–3573.
[9] Gaffney K J, Davis P H, Piletic I R, et al. Hydrogen bond dissociation and reformation in methanol oligomers following hydroxyl stretch relaxation. *J. Phys. Chem. A*, **2002**, *106* (50): 12012–12023.
[10] Shinokita K, Cunha A V, Jansen T L C, et al. Hydrogen bond dynamics in bulk alcohols. *J. Chem. Phys.*, **2015**, *142* (21): 212450.
[11] Salamatova E, Cunha A V, Shinokita K, et al. Hydrogen bond and lifetime dynamics in diluted alcohols. *Phys. Chem. Chem. Phys.*, **2017**, *19* (41): 27960–27967.
[12] Jorgensen W L. Optimized intermolecular potential functions for liquid alcohols. *J. Phys. Chem.*, **1986**, *90* (7): 1276–1284.
[13] Haughney M, Ferrario M, McDonald I R. Pair interactions and hydrogen-bond networks in models of liquid methanol. *Mol. Phys.*, **1986**, *58* (4): 849–853.
[14] van Leeuwen M E, Smit B. Molecular simulation of the vapor-liquid coexistence curve of methanol. *J. Phys. Chem.*, **1995**, *99* (7): 1831–1833.
[15] Schnabel T, Srivastava A, Vrabec J, et al. Hydrogen bonding of methanol in supercritical CO_2: comparison between ^{1}H NMR spectroscopic data and molecular simulation results. *J. Phys. Chem. B*, **2007**, *111* (33): 9871–9878.
[16] Guevara-Carrion G, Nieto-Draghi C, Vrabec J, et al. Prediction of transport properties by molecular simulation: methanol and ethanol and their mixture. *J. Phys. Chem. B*, **2008**, *112* (51): 16664–16674.
[17] Jorgensen W L, Maxwell D S, Tirado-Rives J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.*, **1996**, *118* (45): 11225–11236.
[18] Wang J, Wolf R M, Caldwell J W, et al. Development and testing of a general amber force field. *J. Comput. Chem.*, **2004**, *25* (9): 1157–1174.
[19] Haughney M, Ferrario M, McDonald I R. Molecular-dynamics simulation of liquid methanol. *J. Phys. Chem.*, **1987**, *91* (19): 4934–4940.
[20] Gonzalez-Salgado D, Vega C. A new intermolecular potential for simulations of methanol: The OPLS/2016 model. *J. Chem. Phys.*, **2016**, *145* (3): 034508.
[21] Watanabe T, Ohashi K. Similarity and dissimilarity between water and methanol in solvent effects on the spectroscopic properties of aniline: Molecular dynamics and time-dependent DFT studies. *Comput. Theor. Chem.*, **2022**, *1215*: 113850.
[22] Galicia-Andrés E, Dominguez H, Pusztai L, et al. Composition dependence of thermodynamic, dynamic and dielectric properties of water–methanol model mixtures. Molecular dynamics simulation results with the OPLS-AA model for methanol. *J. Mol. Liq.*, **2015**, *212*: 70–78.
[23] Tsuchida E, Kanada Y, Tsukada M. Density-functional study of liquid methanol. *Chem. Phys. Lett.*, **1999**, *311* (3/4): 236–240.
[24] Pagliai M, Cardini G, Righini R, et al. Hydrogen bond dynamics in liquid methanol. *J. Chem. Phys.*, **2003**, *119* (13): 6655–6662.
[25] Handgraaf J W, van Erp T S, Meijer E J. Ab initio molecular dynamics study of liquid methanol. *Chem. Phys. Lett.*, **2003**, *367* (5/6): 617–624.
[26] Handgraaf J W, Meijer E J, Gaigeot M P. Density-functional theory-based molecular simulation study of liquid methanol. *J. Chem. Phys.*, **2004**, *121* (20): 10111–10119.
[27] McGrath M J, Kuo I F W, Siepmann J I. Liquid structures of water, methanol, and hydrogen fluoride at ambient conditions from first principles molecular dynamics simulations with a dispersion corrected density functional. *Phys. Chem. Chem. Phys.*, **2011**, *13* (44): 19943–19950.
[28] Sieffert N, Bühl M, Gaigeot M P, et al. Liquid methanol from DFT and DFT/MM molecular dynamics simulations. *J. Chem. Theory Comput.*, **2013**, *9* (1): 106–118.
[29] He J, Noto V D, Paddison S J. The structure of water–methanol mixtures under an electric field: Ab initio molecular dynamics simulations. *Chem. Phys. Lett.*, **2015**, *635*: 99–106.
[30] Cassone G, Giaquinta P V, Saija F, et al. Liquid methanol under a static electric field. *J. Chem. Phys.*, **2015**, *142* (5): 054502.
[31] Jindal A, Vasudevan S. Hydrogen bonding in the liquid state of linear alcohols: molecular dynamics and thermodynamics. *J. Phys. Chem. B*, **2020**, *124* (17): 3548–3555.
[32] Jindal A, Vasudevan S. Geometry of OH···O interactions in the liquid state of linear alcohols from *ab initio* molecular dynamics simulations. *Phys. Chem. Chem. Phys.*, **2020**, *22* (12): 6690–6697.
[33] Cassone G, Trusso S, Sponer J, et al. Electric field and temperature effects on the ab initio spectroscopy of liquid methanol. *Appl. Sci.*, **2021**, *11* (12): 5457.
[34] Grimme S. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.*, **2006**, *27* (15): 1787–1799.

DOI: 10.52396/JUSTC-2024-0031
JUSTC, **2024**, 54(6): 0603

[35] Becke A D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A*, **1988**, *38*: 3098.
[36] Lee C, Yang W, Parr R G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B*, **1988**, *37*: 785–789.
[37] Maginn E J, Messerly R A, Carlson D J, et al. Best practices for computing transport properties 1. Self-diffusivity and viscosity from equilibrium molecular dynamics [article v1.0]. *LiveCoMS*, **2019**, *1* (1): 6324.
[38] Konrad M, Wenzel W. CONI-Net: Machine learning of separable intermolecular force fields. *J. Chem. Theory Comput.*, **2021**, *17* (8): 4996–5006.
[39] Gastegger M, Behler J, Marquetand P. Machine learning molecular dynamics for the simulation of infrared spectra. *Chem. Sci.*, **2017**, *8* (10): 6924–6935.
[40] Zhang Y L, Jiang B. Universal machine learning for the response of atomistic systems to external fields. *Nat. Commun.*, **2023**, *14* (1): 6424.
[41] Gastegger M, Schütt K T, Müller K R. Machine learning of solvent effects on molecular spectra and reactions. *Chem. Sci.*, **2021**, *12* (34): 11473–11483.
[42] Manzhos S, Carrington Jr T. Neural network potential energy surfaces for small molecules and reactions. *Chem. Rev.*, **2021**, *121* (16): 10187–10217.
[43] Medders G R, Babin V, Paesani F. Development of a "first-principles" water potential with flexible monomers. III. Liquid phase properties. *J. Chem. Theory Comput.*, **2014**, *10* (8): 2906–2910.
[44] Grisafi A, Wilkins D M, Csányi G, et al. Symmetry-adapted machine learning for tensorial properties of atomistic systems. *Phys. Rev. Lett.*, **2018**, *120* (3): 036002.
[45] Zhang L, Han J, Wang H, et al. Deep potential molecular dynamics: A scalable model with the accuracy of quantum mechanics. *Phys. Rev. Lett.*, **2018**, *120* (14): 143001.
[46] Nguyen T T, Székely E, Imbalzano G, et al. Comparison of permutationally invariant polynomials, neural networks, and Gaussian approximation potentials in representing water interactions through many-body expansions. *J. Chem. Phys.*, **2018**, *148* (24): 241725.
[47] Zhang L F, Wang H, Car R, et al. Phase diagram of a deep potential water model. *Phys. Rev. Lett.*, **2021**, *126* (23): 236001.
[48] Zhang Y L, Xia J F, Jiang B. Physically motivated recursively embedded atom neural networks: incorporating local completeness and nonlocality. *Phys. Rev. Lett.*, **2021**, *127* (15): 156002.
[49] Gartner T E, Piaggi P M, Car R, et al. Liquid-liquid transition in water from first principles. *Phys. Rev. Lett.*, **2022**, *129* (25): 255702.
[50] Bore S L, Paesani F. Realistic phase diagram of water from "first principles" data-driven quantum simulations. *Nat. Commun.*, **2023**, *14* (1): 3349.
[51] Yao K, Herr J E, Parkhill J. The many-body expansion combined with neural networks. *J. Chem. Phys.*, **2017**, *146* (1): 014106.
[52] Li Y, Li H, Pickard F C, et al. Machine learning force field parameters from ab initio data. *J. Chem. Theory Comput.*, **2017**, *13* (9): 4492–4503.
[53] Jindal S, Hsu P J, Phan H T, et al. Capturing the potential energy landscape of large size molecular clusters from atomic interactions up to a 4-body system using deep learning. *Phys. Chem. Chem. Phys.*, **2022**, *24* (44): 27263–27276.
[54] Maldonado A M, Poltavsky I, Vassilev-Galindo V, et al. Modeling molecular ensembles with gradient-domain machine learning force fields. *Digit. Discov.*, **2023**, *2* (3): 871–880.
[55] Chen B, Potoff J J, Siepmann J I. Monte Carlo calculations for alcohols and their mixtures with alkanes. Transferable potentials for phase equilibria. 5. United-atom description of primary, secondary, and tertiary alcohols. *J. Phys. Chem. B*, **2001**, *105* (15): 3093–3104.
[56] Ren P, Wu C, Ponder J W. Polarizable atomic multipole-based molecular mechanics for organic molecules. *J. Chem. Theory Comput.*, **2011**, *7* (10): 3143–3161.
[57] Cheng B, Engel E A, Behler J, et al. Ab initio thermodynamics of liquid and solid water. *Proc. Natl. Acad. Sci. U. S. A.*, **2019**, *116* (4): 1110–1115.
[58] Ko H Y, Zhang L, Santra B, et al. Isotope effects in liquid water via deep potential molecular dynamics. *Mol. Phys.*, **2019**, *117* (22): 3269–3281.
[59] VandeVondele J, Krack M, Mohamed F, et al. Quickstep: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Comput. Phys. Commun.*, **2005**, *167* (2): 103–128.
[60] Kühne T D, Iannuzzi M, Ben M D, et al. CP2K: An electronic structure and molecular dynamics software package - Quickstep: Efficient and accurate electronic structure calculations. *J. Chem. Phys.*, **2020**, *152* (19): 194103.
[61] Lippert G, Hutter J, Parrinello M. A hybrid Gaussian and plane wave density functional scheme. *Mol. Phys.*, **1997**, *92* (3): 477–487.
[62] Adamo C, Barone V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys.*, **1999**, *110* (13): 6158–6170.
[63] Grimme S, Antony J, Ehrlich S, et al. A consistent and accurate *ab initio* parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.*, **2010**, *132* (15): 154104.
[64] Marsalek O, Markland T E. Quantum dynamics and spectroscopy of ab initio liquid water: the interplay of nuclear and electronic quantum effects. *J. Phys. Chem. Lett.*, **2017**, *8* (7): 1545–1551.
[65] Goedecker S, Teter M, Hutter J. Separable dual-space Gaussian pseudopotentials. *Phys. Rev. B*, **1996**, *54* (3): 1703–1710.
[66] Hartwigsen C, Goedecker S, Hutter J. Relativistic separable dual-space Gaussian pseudopotentials from H to Rn. *Phys. Rev. B*, **1998**, *58* (7): 3641–3662.
[67] Krack M. Pseudopotentials for H to Kr optimized for gradient-corrected exchange-correlation functionals. *Theor. Chem. Acc.*, **2005**, *114*: 145–152.
[68] VandeVondele J, Hutter J. Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases. *J. Chem. Phys.*, **2007**, *127* (11): 114105.
[69] Guidon M, Hutter J, VandeVondele J. Auxiliary density matrix methods for Hartree-Fock exchange calculations. *J. Chem. Theory Comput.*, **2010**, *6* (8): 2348–2364.
[70] VandeVondele J, Hutter J. An efficient orbital transformation method for electronic structure calculations. *J. Chem. Phys.*, **2003**, *118* (10): 4365–4369.
[71] Zhang Y L, Hu C, Jiang B. Embedded atom neural network potentials: efficient and accurate machine learning with a physically inspired representation. *J. Phys. Chem. Lett.*, **2019**, *10* (17): 4962–4967.
[72] Zhang Y L, Xia J F, Jiang B. REANN: A PyTorch-based end-to-end multi-functional deep neural network package for molecular, reactive, and periodic systems. *J. Chem. Phys.*, **2022**, *156* (11): 114801.
[73] Zhang Y L, Lin Q D, Jiang B. Atomistic neural network representations for chemical dynamics simulations of molecular, condensed phase, and interfacial systems: Efficiency, representability, and generalization. *WIREs Comput. Mol. Sci.*, **2023**, *13* (3): e1645.
[74] Thompson A P, Aktulga H M, Berger R, et al. LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Comput. Phys. Commun.*, **2022**, *271*: 108171
[75] Nosé S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.*, **1984**, *81* (1): 511–519.
[76] Hoover W G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A*, **1985**, *31* (3): 1695–1697.
[77] Brehm M, Kirchner B. TRAVIS - a free analyzer and visualizer for

DOI: 10.52396/JUSTC-2024-0031
JUSTC, **2024**, 54(6): 0603

Monte Carlo and molecular dynamics trajectories. *J. Chem. Inf. Model.*, **2011**, *51* (8): 2007–2023.

[78] Brehm M, Thomas M, Gehrke S, et al. TRAVIS—A free analyzer for trajectories from molecular simulation. *J. Chem. Phys.*, **2020**, *152* (16): 164105.

[79] Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. *J. Mol. Graph.*, **1996**, *14* (1): 33–38.

[80] Ceriotti M, Fang W, Kusalik P G, et al. Nuclear quantum effects in water and aqueous systems: Experiment, theory, and current challenges. *Chem. Rev.*, **2016**, *116* (13): 7529–7550.

[81] Wang C C, Tan J Y, Liu L H. *Ab initio* molecular dynamics study of temperature and pressure-dependent infrared dielectric functions of liquid methanol. *AIP Adv.*, **2017**, *7* (3): 035115.

[82] Hurle R L, Woolf L A. The effect of isotopic substitution on self-diffusion in methanol under pressure. *Aust. J. Chem.*, **1980**, *33* (9): 1947–1952.

[83] Yadav V K, Chandra A. Dynamics of hydrogen bonds and vibrational spectral diffusion in liquid methanol from first principles simulations with dispersion corrected density functional. *Chem. Phys.*, **2013**, *415*: 1–7.

[84] Jukić I, Požar M, Lovrinčević B, et al. Universal features in the lifetime distribution of clusters in hydrogen-bonding liquids. *Phys. Chem. Chem. Phys.*, **2021**, *23* (35): 19537–19546.

DOI: 10.52396/JUSTC-2024-0031
JUSTC, **2024**, 54(6): 0603