



Published in final edited form as:

*J Phys Chem Lett.* 2018 June 21; 9(12): 3232–3240. doi:10.1021/acs.jpcclett.8b01131.

## Force Field for Water Based on Neural Network

Hao Wang and

Department of Chemistry, Duke University, Durham, North Carolina 27708, United States

Weitao Yang\*

Department of Chemistry and Department of Physics, Duke University, Durham, North Carolina 27708, United States

Key Laboratory of Theoretical Chemistry of Environment, Ministry of Education, School of Chemistry and Environment, South China Normal University, Guangzhou 510006, China

### Abstract

We developed a novel neural network-based force field for water based on training with high-level ab initio theory. The force field was built based on an electrostatically embedded many-body expansion method truncated at binary interactions. The many-body expansion method is a common strategy to partition the total Hamiltonian of large systems into a hierarchy of few-body terms. Neural networks were trained to represent electrostatically embedded one-body and two-body interactions, which require as input only one and two water molecule calculations at the level of ab initio electronic structure method CCSD/aug-cc-pVDZ embedded in the molecular mechanics water environment, making it efficient as a general force field construction approach. Structural and dynamic properties of liquid water calculated with our force field show good agreement with experimental results. We constructed two sets of neural network based force fields: nonpolarizable and polarizable force fields. Simulation results show that the nonpolarizable force field using fixed TIP3P charges has already behaved well, since polarization effects and many-body effects are implicitly included due to the electrostatic embedding scheme. Our results demonstrate that the electrostatically embedded many-body expansion combined with neural network provides a promising and systematic way to build next-generation force fields at high accuracy and low computational costs, especially for large systems.

### Graphical Abstract

---

\*Corresponding Author : [weitao.yang@duke.edu](mailto:weitao.yang@duke.edu).

ASSOCIATED CONTENT

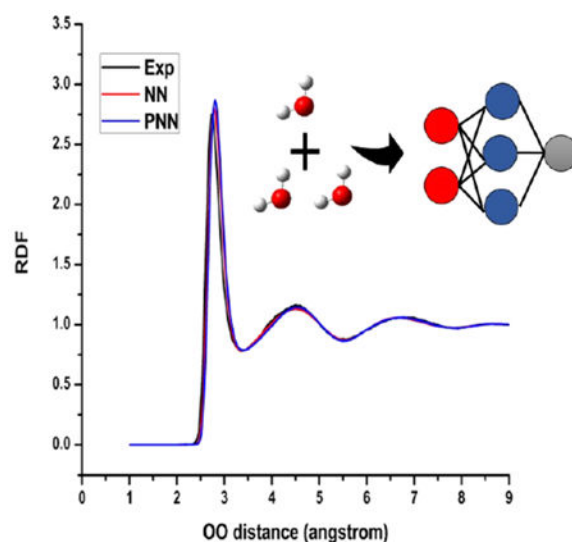
Supporting Information

The Supporting Information is available free of charge on the [ACS Publications website](https://pubs.acs.org) at DOI: [10.1021/acs.jpcclett.8b01131](https://doi.org/10.1021/acs.jpcclett.8b01131).

Details of NN structure, details of MD simulations, dipole moment distribution, hydrogen bond analysis, and dielectric constant calculations ([PDF](#))

NN parameters files ([ZIP](#))

The authors declare no competing financial interest.



Water is one of the most important solvents for many chemical and biological processes. Although water molecules have simple geometric structures, they show various interesting and peculiar phenomena, such as the density maximum at 4° C. In the past few decades, much effort from the molecular dynamics (MD) simulation community has been devoted to simulating water accurately and efficiently. The earliest water models are SPC<sup>14</sup> and TIP3P.<sup>40</sup> Both SPC and TIP3P adopt three-site structures and fixed atomic charges. Though simple, these two models achieve great success in recovering many bulk properties of water,<sup>69</sup> and they are still widely used today. However, they fail to predict intermolecular interactions well. Going beyond these two models, various classical water models have been developed, including TIP4P,<sup>41</sup> TIP4P-FQ,<sup>61</sup> AMOEBA,<sup>60,71</sup> COS/D2,<sup>8,45</sup> and SWM4-NDP.<sup>46</sup> These models add either more interaction sites or explicit polarization effects to further improve the accuracy, and show a significant improvement over SPC and TIP3P not only for bulk properties but for molecular interactions. However, conventional molecular mechanics (MM) force fields fail to give the predictive quantitative accuracy needed for some of the questions of interest.<sup>9,57</sup> To a large extent, the inaccuracy is due to the limited form of the classical MM force fields.

Direct ab initio MD (AIMD)<sup>15,37</sup> simulations can give highly accurate results but at high computational costs. Thus, AIMD is limited to relatively small systems, and the majority of AIMD simulations of water in the literature is at the level of density functional approximations (DFAs),<sup>28,31,42,51,52,56,75,76</sup> which currently have certain limitations.<sup>17</sup> Various approximations have been developed to alleviate the highly demanding quantum mechanical (QM) computational costs without significant loss of accuracy. Fragment-based methods have been designed for large systems, such as the divide-and-conquer approach,<sup>73</sup> molecular fractionation with conjugate caps,<sup>77</sup> systematic molecular fragmentation,<sup>2,18,23</sup> and the generalized energy-based fragmentation approach.<sup>36,47</sup> Another frequently used approximation is to expand the system energy as a sum of many-body expansions and truncate it after the first several terms.<sup>20,32,58</sup> The simplest expansion is to use only two-body interactions. The scheme of many-body expansion can systematically improve the

accuracy by including more and more terms. However, the computational costs increase quickly as more terms are added. In order to improve the accuracy and convergence of many-body expansions, electrostatic embedding from background atoms can be added into the Hamiltonian. Electrostatic interactions with background molecular charges may be added in different forms, such as the pair interaction molecular orbital method<sup>44</sup> and the fragment molecular orbital method of Kitaura and coworkers.<sup>25,26,43,55</sup> However, all these require direct QM calculations, and none of these schemes are in force field form.

Author: Please verify the changes made to the above reference. Author: Please verify the changes made to the above reference. The conventional force field has its limited forms and associated limitation in accuracy. It is possible to go beyond the conventional force field with a neural network (NN), because of its flexibility and low computational costs. It has been shown that NNs are able to approximate unknown functions to, in principle, any desired accuracy.<sup>12,19,49,50</sup> NNs have many applications in chemistry, such as the analysis of spectra,<sup>68</sup> quantitative structure-activity relationship studies<sup>66</sup> and free energy corrections of chemical reactions.<sup>65</sup> The development of NN-based potential energy surface has seen tremendous progress in the past few years. In early applications, NN-based potentials are restricted to low-dimensional systems due to its rapidly increasing complexity with system size. Recently, high-dimensional NN-based potentials was developed by Behler and Parrinello.<sup>13</sup> They expressed the molecular energy as the sum of atomic energies, and each atomic energy is represented by a set of NNs. It turns out that this strategy simplifies the NN structure of molecular systems studied and can achieve very high accuracy.<sup>12,13,53</sup> However, their procedure of building NN-based potentials for water can be expensive, since it requires QM calculations of large water clusters.<sup>53</sup> Parkhill and co-workers recently combined NNs with many-body expansion to predict energy of methanol clusters at the level of second-order Møller-Plesset perturbation theory (MP2).<sup>74</sup> But again, this was not developed for the force field purpose. Paesani and co-workers built a high-precision water force field based on many-body expansion using carefully designed polynomial fitting without electrostatic embedding.<sup>6,7,59</sup> On top of classical intermolecular interactions (electrostatic and dispersion energy), they added short-range two-body and three-body corrections at the level of CCSD(T)/CBS. Their model contains higher-order many-body effects due to polarization.<sup>7,59</sup> Some studies have shown that electrostatic embedding can accelerate the convergence of fragment-based method, especially for polarizable systems, such as water.<sup>20,18,29,30,47</sup> Electrostatically embedded two-body expansion includes many-body effects beyond two-body interactions due to the electrostatic embedding. It has been demonstrated that electrostatically embedded two-body expansion show similar accuracy with three-body expansion in gas phase for water clusters.<sup>20,21</sup> Since Paesani and co-workers successfully developed a high-precision water force field by many-body expansion in the gas phase, we are particularly interested in modeling electrostatically embedded energy of water monomers and dimers with a high-level ab initio method as a reference to predict bulk properties of water. The idea of modeling electrostatically embedded two-body expansion is appealing since it significantly reduces the computational costs compared with higher-order many-body expansion.

In the present work, our aim is to develop a general NN force field for water with high level QM accuracy and low computational costs. Inspired by quantum mechanics/molecular

mechanics (QM/MM) approaches,<sup>27,35,48,78</sup> our NN force field uses the many-body expansion up to binary interactions and NN representation of atomic energies, employing an electrostatic embedding scheme. For each electrostatically embedded one-body or two-body term, only water molecules in the QM regions use NN representation, and water molecules in the MM regions only serve as background charges to provide the electrostatic embedding. To properly represent the atomic environment, many different descriptors have been developed including Coulomb matrix,<sup>24,33,62</sup> symmetry functions,<sup>12,13</sup> bispectrum,<sup>10,11</sup> permutation invariant polynomial,<sup>38,39</sup> metric finger-print<sup>63,64,79</sup> and Fourier series of atomic radial distribution functions.<sup>70</sup> In this work, the atomic environment feature we used is the power spectrum, which is a subset of the bispectrum developed by Csányi and co-workers.<sup>10,11</sup> Although power spectrum is not a one-to-one mapping between geometric configurations of molecules and output value, we find it sufficient to represent an atomic environment for our purpose. Molecular dynamics simulations were performed with our in-house program QM4D.<sup>34</sup>

We use the following form of potential energy for the whole system:

$$E = \sum_{i=1}^N E_1[i; Q_i] + \sum_{i>j}^{R_{ij} < R_{QM}} (E_2[i, j; Q_i \cup Q_j] - E_1[i; j \cup Q_i \cup Q_j] - E_1[j; i \cup Q_i \cup Q_j]) - \sum_{i>j}^{R_{QM} < R_{ij} < R_{SR}} \frac{q_i q_j}{R_{ij}} + \sum_{i>j}^{R_{SR} < R_{ij} < R_{LR}} \frac{q_i q_j}{R_{ij}}$$

(1)

Two-body interactions within  $R_{QM}$  are evaluated with NNs at QM level.  $R_{SR}$  is the short-range electrostatic interactions cutoff and  $R_{LR}$  is the long-range electrostatic interactions cutoff.  $Q_i$  represents background charges within  $R_{SR}$  of QM molecule  $i$ .  $q_i$  in the last two terms is either TIP3P charges (for nonpolarizable NN force field) or electrostatic potential (ESP) derived atomic charges (for polarizable NN force field).  $E_1$  and  $E_2$  are electrostatically embedded one-body and two-body energies, respectively. This expansion scheme was used by Hirata and coworkers.<sup>72</sup> In their case, they ran AIMD for small water clusters at the level of MP2. In our work, instead of carrying out repeated ab initio electronic structure calculations for the small water clusters, we develop NNs to describe the electrostatically embedded one-body and two-body energy terms, using feed-forward NNs for atomic energies in the QM region. For both one-body and two-body terms, each chemical element shares one common set of NNs (i.e., oxygen atoms use one set of NNs and hydrogen atoms use another set of NNs.). The atomic energy  $e_a$  represented by two hidden-layer NNs is given as

$$e_a = \sum_{k=1}^{N_2} w_k^{23} \tanh \left[ \sum_{j=1}^{N_1} w_{jk}^{12} \tanh \left( \sum_{i=1}^{N_0} w_{ij}^{01} G_i + b_j^1 \right) + b_k^2 \right] + b_0 \quad (2)$$

where  $G_i$  is the input variable for node  $i$ ,  $w_{ij}^{01}$  ( $w_{jk}^{12}$ ) is the weight parameter that connects node  $i$  ( $j$ ) in the previous layer and node  $j$  ( $k$ ) in the current layer,  $w_k^{23}$  is the weight parameter that connects node  $k$  in the second hidden layer and output layer,  $b_j^1$  and  $b_k^2$  are bias weight parameters of two hidden layers,  $b_0$  is the bias weight parameter of output layer,  $N_0$  is the number of nodes in input layer, and  $N_1$  and  $N_2$  are the number of nodes in two hidden layers. Energy of electrostatically embedded one-body and two-body terms is the sum of atomic energies in the QM region (Figure 1),

$$E_{1/2} = \sum_a e_a \quad (3)$$

To construct a power spectrum, we project the discrete charge density  $\rho(\mathbf{r}) = \sum_i \delta(\mathbf{r} - \mathbf{r}_i) q_i$ , including both QM charges and MM charges, onto selected radial functions and spherical harmonic functions:

$$\rho(\mathbf{r}) = \sum_{l=0}^l \sum_{m=-l}^l c_{nlm} g_n(r) Y_{lm}(\theta, \phi) \quad (4)$$

After obtaining the coefficients  $c_{nlm}$ , the corresponding power spectrum can then be calculated as

$$\mathbf{c}_{nl}^\dagger \mathbf{c}_{nl} = \sum_i \sum_j \sum_{m=-l}^l g_n(r_i) g_n(r_j) q_i q_j Y_{lm}^*(\theta_i, \phi_i) Y_{lm}(\theta_j, \phi_j) \quad (5)$$

Details of the structures of our NNs can be found in the Supporting Information.

Due to large system size and high computational costs of ab initio methods, the majority of AIMD simulations of water in the literature is at the level of DFAs,<sup>28,31,42,51,52,56,75,76</sup> However, most DFAs suffer from delocalization and strong correlation error and lack accurate van der Waals interactions.<sup>16</sup> This leads to inaccurate simulations of water, in which van der Waals interactions are believed to play an important role. Although some van der Waals density functionals were adopted for AIMD simulations of water, they still cannot fully represent the van der Waals interactions in water.<sup>42,76</sup> Therefore, we developed NNs for electrostatically embedded water monomers and dimers representing coupled cluster singles doubles (CCSD) results.

The reference CCSD was calculated with the basis set aug-ccpVDZ. Compared with using NNs to predict the whole system directly,<sup>53</sup> NNs combined with many-body expansions provide a possible way of simulating water with higher-level QM methods. This is because a smaller size of water clusters is needed for training NNs in the many-body scheme.

The NN optimization object function, in weighted mean squared deviation (MSD), for one-body ( $E_1[i; Q_i]$ ) and two-body terms ( $E_2[i; j; Q_i Q_j] - E_1[i; \mathcal{J} Q_i \mathcal{J} Q_j] - E_1[j; \mathcal{J} Q_i \mathcal{J} Q_j]$ ) is defined as

$$L = \sum_i \left( |E_{\text{ref},i} - E_i|^2 + \frac{\alpha}{3N_{\text{qm}}} |\mathbf{F}_{\text{ref},i} - \mathbf{F}_i|^2 \right) + \beta \sum_i w_i^2 \quad (6)$$

where  $\alpha$  is the relative weight of force with respect to energy and is set to be 0.1 in the current study. The last term is a regularization penalty added to prevent overfitting, and  $\{w_i\}$  are parameters of the NN.  $\beta$  is a small number and is set to be  $1 \times 10^{-6}$ .  $N_{\text{qm}}$  is the number of atoms in the QM region. Only force components of atoms in the QM region are accounted for in the object function. NNs are constructed in the following iterative way:<sup>4</sup> First, several rough NN potentials were built based on snapshots taken from MD simulations. Then, one NN potential was used to run MD simulations, and other NN potentials monitored configurations sampled in this process. When two sets of NN potentials deviated from each other by more than a predetermined value for some configuration, this configuration was then selected out and added into the training set. This process was iterated until one set of satisfying NN was generated. We used 1904 configurations for training and 445 configurations for testing for water monomer model, and 1646 configurations for training and 948 configurations for testing for water dimer model. The embedding cutoff is 12 Å for both training and testing set. There are more than 200 embedding MM water molecules for a typical configuration, which makes it very challenging to model for virial-related calculations. This will be discussed in more details later. The accuracy of our NNs is summarized in Table 1. The energy root mean squared error (RMSE) of the monomer is 0.16 kcal/mol for the training set and 0.22 kcal/mol for the testing set, and the force RMSE of the monomer is 0.77 kcal/mol/Å for the training set and 0.86 kcal/mol/Å for the testing set. The energy RMSE of the dimer is 0.06 kcal/mol/ $H_2O$  for the training set and 0.05 kcal/mol/ $H_2O$  for the testing set, and the force RMSE of the dimer is 0.66 kcal/mol/Å for the training set and 0.54 kcal/mol/Å. As we can see from Table 1, the error of the testing sets is of comparable accuracy with the training sets.

We developed two sets of NN force fields: one using nonpolarizable TIP3P charges as the electrostatic embedding charges, and the other using polarizable ESP charges as the electrostatic embedding charges. Accurate representation of electrostatic embedding is important for reliable energy prediction. As demonstrated in Truhlar's work,<sup>20</sup> TIP3P background charges with electrostatically embedded two-body expansion show very low energy error for water clusters among different point-charge embedding schemes. Thus, we used TIP3P embedding charges for our nonpolarizable force fields. In the polarizable NN force field, a different set of NN was used to predict atomic ESP charges. Before the many-body expansion calculations, atomic ESP charges converge in a self-consistent way. In order to satisfy neutral charge condition for each single water molecule, only the hydrogen atomic charge was predicted by the NN and the oxygen atomic charge was calculated by  $-q(H_1) - q(H_2)$ . The accuracy of NN charge prediction for a hydrogen atom is 0.03 electron. Artrith et al. applied the idea of environment-dependent charges to the zinc oxide system.<sup>5</sup>

However, there are two major differences between their work and ours. First, in their method, environment-dependent charges are predicted using a cutoff of 6 Å, which only partially reflect the contribution of environment. This is fine since their total energy of the bulk system is the sum of atomic energy, and the energy remainder is absorbed into NN. However, in our case, our total energy is the sum of electrostatically embedded energy. To represent electrostatic interactions well, we have to take into account all the contribution from the MM embedding charges within 12 Å. The chemical complexity in our case is higher. Second, direct application of symmetry functions to represent chemical environment of embedding charges within 12 Å (more than 200 water molecules for a typical configuration) is very expensive. We developed the use of embedding charge density to construct a power spectrum to efficiently represent electrostatically embedded energy. To verify the accuracy of our force fields, we first tested it for water clusters of different size (Figure 2). Geometric coordinates of these water clusters were obtained from the Benchmark Energy and Geometry Database.<sup>1</sup> The corresponding absolute error of binding energy is shown in Figure 3, where we compared electrostatically embedded two-body expansion using TIP3P charges with NN force fields, electrostatically embedded two-body expansion using TIP3P charges at the level of CCSD/aug-cc-pVDZ, and full QM calculations at the level of CCSD/aug-cc-pVDZ. The absolute error of binding energy for our NN force field is stable as the size of water clusters increases. It is larger than our training accuracy for monomer and dimer but is still of chemical accuracy. The red bar confirms that electrostatically embedded two-body expansion provides an accurate reference for our force fields.

We now apply the NN force fields to determine bulk water properties using MD simulations. The radial distribution function (RDF) is one of the most important properties reflecting the local structure of water. Experimental data are available for comparison.<sup>67</sup> We calculated the oxygen–oxygen (OO), oxygen–hydrogen (OH) and hydrogen–hydrogen (HH) RDF, which are shown in Figure 4. As we can see, our predicted OO RDF shows excellent agreement with experimental data for the whole range. Here we emphasize that our NN force fields are constructed only from QM calculations of electrostatically embedded one and two water molecules, and are not fitted to any experimental results. We therefore conclude that our NN force fields provide an accurate description of structural features of liquid water. OO RDF obtained with MP2 using electrostatically embedded two-body expansion shows a much higher first peak.<sup>72</sup> We believe this is because van der Waals interactions are not correctly accounted for in MP2. Some studies pointed out that nuclear quantum effects are important for hydrogen atoms in water.<sup>54</sup> The subtle mismatch of the first peak in OO RDF may be due to the lack of consideration of nuclear quantum effects in the current NN force fields or the error of NN representation. We observed that OH and HH RDF do not match experimental results as well as OO RDF. This may affect the prediction of other water properties. Our OH and HH RDF show comparable accuracy with those of TIP3P (not shown here). Compared with TIP3P, our force fields do not improve OH and HH RDF much.

The self-diffusion coefficient ( $D$ ) measures the dynamical properties of liquids. The self-diffusion coefficient was calculated by Einstein's formula:<sup>22</sup>



$$D_{\text{PBC}} = \lim_{t \rightarrow \infty} \frac{|\mathbf{r}(t) - \mathbf{r}(0)|^2}{6t} \quad (7)$$

The self-diffusion coefficient is related to the MSD of atomic positions. MSD as a function of time  $t$  is plotted in Figure 5, and our calculated self-diffusion coefficient is compared with experimental data and TIP3P results in Table 2. From Table 2 we can conclude that NN force fields compares favorably with experimental data and describe the dynamical property of liquid water properly.

The shear viscosity was calculated by the integration over the autocorrelation function (ACF) of stress tensor elements,<sup>3</sup>

$$\eta = \frac{V}{k_{\text{B}}T} \int_0^{\infty} \left\langle \delta P_{\alpha\beta}(t) \delta P_{\alpha\beta}(0) \right\rangle dt \quad (8)$$

where  $\delta P_{\alpha\beta}(t) = P_{\alpha\beta}(t) - \langle P_{\alpha\beta}(t) \rangle$ . There are in total five independent components  $P_{xy}$ ,  $P_{xz}$ ,  $P_{yz}$ ,  $\frac{1}{2}(P_{xx} - P_{yy})$ , and  $\frac{1}{2}(P_{yy} - P_{zz})$ . Simulation results are shown in Table 3 and Figure 6.

Compared with ACF of TIP3P, both nonpolarizable and polarizable NN force fields show longer correlation time and do not show negative correlation. These two factors together make the shear viscosity of NN force fields relatively large. The inaccurate prediction of shear viscosity, or more general virial-related properties, are likely due to two reasons. First, virial-related properties put a higher demand on force accuracy between largely separated molecules. Second, in many-body expansion with electrostatic embedding, forces on one molecule consist of contributions from more terms, i.e., each MM embedding charge contributes to the forces on QM atoms. These two reasons together will amplify small force errors in virial-related calculations in the current scheme. Modeling electrostatically embedded many-body expansion is challenging especially for virial-related properties. Virial calculations are directly related to pressure estimation in MD simulations. Poor virial calculations will limit the ability of the current force field to predict density, heat of vaporization, heat capacity, thermal expansion coefficient, and isothermal compressibility. One possible solution to overcome this difficulty is to use shorter cutoff of embedding charges to construct input features for NN while keeping QM calculations with original long embedding cutoff. For example, suppose all the QM calculations of snapshots in the training set are with 12 Å cutoff of MM charges. When we construct a power spectrum to represent an MM environment, embedding charges are cut at 5 Å. This means embedding charges between 5 Å and 12 Å are treated in mean field. We call this method partial mean field for NN representation of an MM environment. The partial mean field method will reduce both the number of pairwise interactions and the need to calculate forces between largely separated molecules. Thus, it may improve the virial calculations.

In summary, a novel NN force field for water was built based on an electrostatically embedded two-body expansion scheme. The new NN force field is not limited by the



Hamiltonian form as classical force fields and is efficient for applications. It was developed at low cost using high-level electronic structure calculations only for one and two water molecules embedded in MM water molecules and showed high accuracy compared with high-level QM results. Our NN force fields show good structural and dynamical properties of water. Compared with the force field of Paesani and co-workers,<sup>6,7,59</sup> we provide an alternative way to build a water force field. The electrostatically embedded two-body expansion includes many-body effects due to the electrostatic embedding scheme, while in the force field of Paesani and co-workers many-body effects are accounted by long-range classical polarization. The idea of electrostatically embedded two-body expansion is appealing since it significantly reduces the computational costs compared with higher-order many-body expansion. We made progress in modeling electro-statically embedded two-body expansion for bulk systems with high-level QM reference. However, challenges still exist for virial-related calculations in modeling electrostatic embedding. Our work provides a general way of building high accuracy force fields, based on only the high-level QM calculations for small electrostatically embedded clusters, which can be easily extended to the force field construction of other systems.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGMENTS

H.W. thanks Dr. Lin Shen for helpful discussions and Dr. Neil Qiang Su for discussions on parallel programming. Financial support from the National Institutes of Health (Grant No. R01 GM061870-13) is gratefully acknowledged.

## REFERENCES

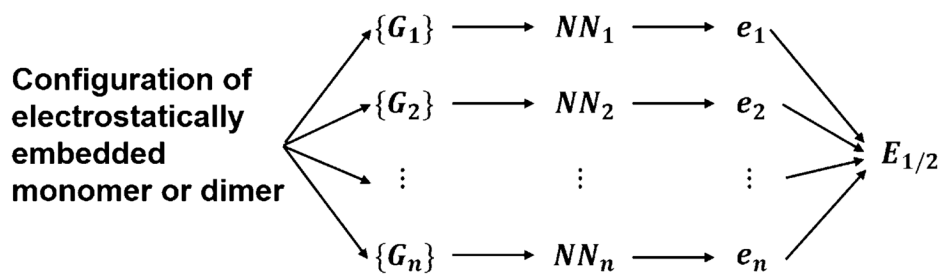
- (1). The benchmark energy and geometry database. <http://www.begdb.com/index.php> (accessed February 23, 2017).
- (2). Addicoat MA; Collins MA Accurate treatment of nonbonded interactions within systematic molecular fragmentation. *J. Chem. Phys* 2009, 131, 104103.
- (3). Allen MP; Tildesley DJ *Computer Simulation of Liquids*; Clarendon Press: Oxford, U.K, 1987.
- (4). Artrith N; Behler J High-dimensional neural network potentials for metal surfaces: A prototype study for copper. *Phys. Rev. B: Condens. Matter Mater. Phys* 2012, 85, 045439.
- (5). Artrith N; Morawietz T; Behler J High-dimensional neural-network potentials for multicomponent systems: Applications to zinc oxide. *Phys. Rev. B: Condens. Matter Mater. Phys* 2011, 83, 153101.
- (6). Babin V; Leforestier C; Paesani F Development of a “first principles” water potential with flexible monomers: Dimer potential energy surface, vrt spectrum, and second virial coefficient. *J. Chem. Theory Comput* 2013, 9 (12), 5395–5403. [PubMed: 26592277]
- (7). Babin V; Medders GR; Paesani F Development of a “first principles” water potential with flexible monomers. ii: Trimer potential energy surface, third virial coefficient, and small clusters. *J. Chem. Theory Comput* 2014, 10 (4), 1599–1607. [PubMed: 26580372]
- (8). Bachmann SJ; van Gunsteren WF An improved simple polarisable water model for use in biomolecular simulation. *J. Chem. Phys* 2014, 141 (22), 22D515.
- (9). Banáš P; Hollas D; Zgarbová M; Jurec ka P; Orozco M; Cheatham TE; Šponer J; Otyepka M Performance of molecular mechanics force fields for rna simulations: Stability of uucg and gnra hairpins. *J. Chem. Theory Comput* 2010, 6 (12), 3836–3849.

- (10). Bartók AP; Kondor R; Csányi G On representing chemical environments. *Phys. Rev. B: Condens. Matter Mater. Phys* 2013, 87, 184115.
- (11). Bartók AP; Payne MC; Kondor R; Csányi G Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons. *Phys. Rev. Lett* 2010, 104, 136403. [PubMed: 20481899]
- (12). Behler J Representing potential energy surfaces by high-dimensional neural network potentials. *J. Phys.: Condens. Matter* 2014, 26 (18), 183001. [PubMed: 24758952]
- (13). Behler J; Parrinello M Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett* 2007, 98, 146401. [PubMed: 17501293]
- (14). Berendsen HJC; Postma JPM; van Gunsteren WF; Hermans J *Intermolecular Forces*; Springer-Science+Business Media, B.V.: Dordrecht, Netherlands, 1981; pp 331–342.
- (15). Car R; Parrinello M Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett* 1985, 55, 2471–2474. [PubMed: 10032153]
- (16). Cohen AJ; Mori-Sánchez P; Yang W Insights into current limitations of density functional theory. *Science* 2008, 321 (5890), 792–794. [PubMed: 18687952]
- (17). Cohen AJ; Mori-Sánchez P; Yang W Challenges for density functional theory. *Chem. Rev* 2012, 112 (1), 289–320. [PubMed: 22191548]
- (18). Collins MA Molecular forces, geometries, and frequencies by systematic molecular fragmentation including embedded charges. *J. Chem. Phys* 2014, 141, 094108. [PubMed: 25194365]
- (19). Ripley BD *Pattern Recognition and Neural Networks*; Cambridge University Press: Cambridge, U.K., 2009.
- (20). Dahlke EE; Truhlar DG Electrostatically embedded many-body expansion for large systems, with applications to water clusters. *J. Chem. Theory Comput* 2007, 3 (1), 46–53. [PubMed: 26627150]
- (21). Dahlke EE; Truhlar DG Electrostatically embedded many-body expansion for simulations. *J. Chem. Theory Comput* 2008, 4 (1), 1–6. [PubMed: 26619974]
- (22). Rapaport DC *The art of molecular dynamics simulation*; Cambridge University Press: Cambridge, U.K., 2004.
- (23). Deev V; Collins MA Approximate ab initio energies by systematic molecular fragmentation. *J. Chem. Phys* 2005, 122, 154102. [PubMed: 15945620]
- (24). Faber F; Lindmaa A; von Lilienfeld OA; Armiento R Crystal structure representations for machine learning models of formation energies. *Int. J. Quantum Chem* 2015, 115, 1094–1101.
- (25). Fedorov DG; Kitaura K The importance of three-body terms in the fragment molecular orbital method. *J. Chem. Phys* 2004, 120, 6832–6840. [PubMed: 15267582]
- (26). Fedorov DG; Kitaura K On the accuracy of the 3-body fragment molecular orbital method (fmo) applied to density functional theory. *Chem. Phys. Lett* 2004, 389, 129–134.
- (27). Friesner RA; Guallar V Ab initio quantum chemical and mixed quantum mechanics/molecular mechanics (qm/mm) methods for studying enzymatic catalysis. *Annu. Rev. Phys. Chem* 2005, 56 (1), 389–427. [PubMed: 15796706]
- (28). Gaiduk AP; Gygi F; Galli G Density and compressibility of liquid water and ice from first-principles simulations with hybrid functionals. *J. Phys. Chem. Lett* 2015, 6 (15), 2902–2908. [PubMed: 26267178]
- (29). Gao J Toward a molecular orbital derived empirical potential for liquid simulations. *J. Phys. Chem. B* 1997, 101 (4), 657–663.
- (30). Gao J A molecular-orbital derived polarization potential for liquid water. *J. Chem. Phys* 1998, 109 (6), 2346–2354.
- (31). Gillan MJ; Alfe D; Michaelides A Perspective: How good is dft for water? *J. Chem. Phys* 2016, 144 (13), 130901. [PubMed: 27059554]
- (32). Gordon MS; Fedorov DG; Pruitt SR; Slipchenko LV Fragmentation methods: A route to accurate calculations on large systems. *Chem. Rev* 2012, 112 (1), 632–672. [PubMed: 21866983]
- (33). Hansen K; Montavon G; Biegler F; Fazli S; Rupp M; Scheffler M; von Lilienfeld OA; Tkatchenko A; Müller K-R Assessment and validation of machine learning methods for

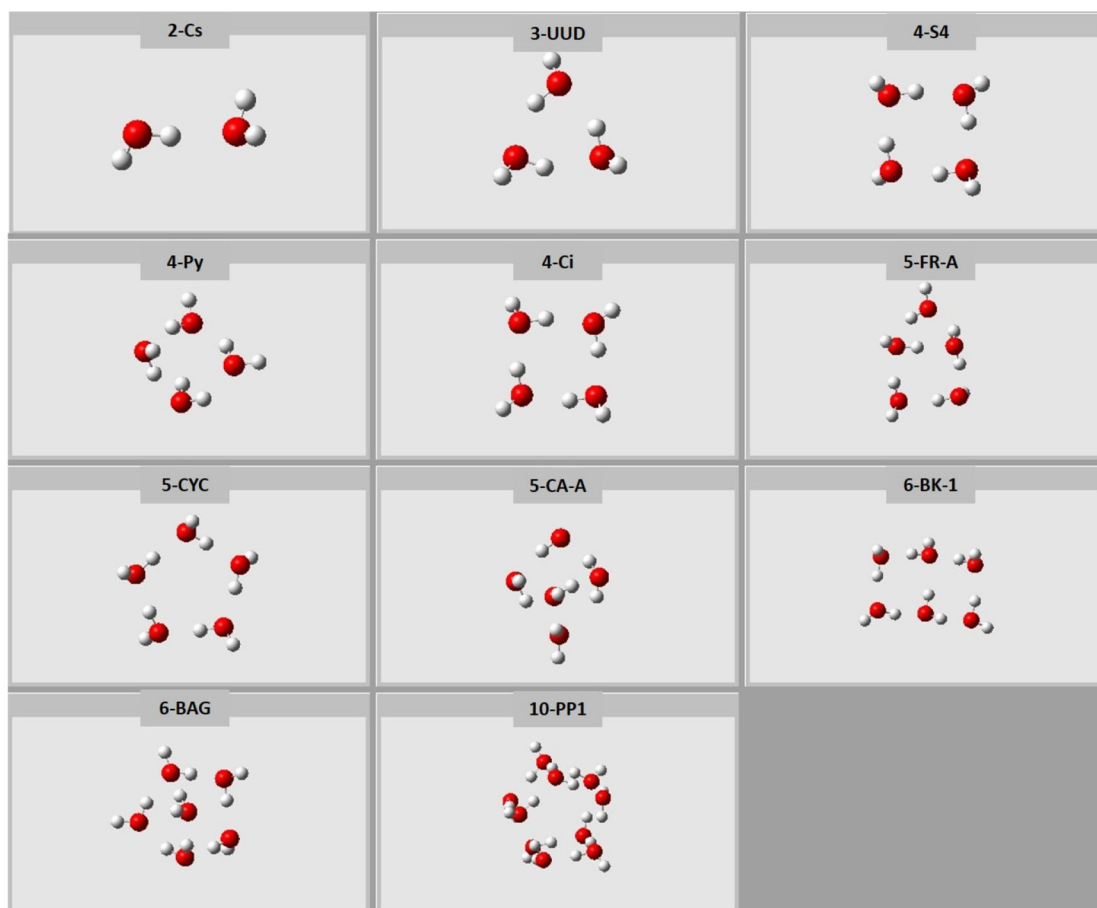
- predicting molecular atomization energies. *J. Chem. Theory Comput* 2013, 9, 3404–3419. [PubMed: 26584096]
- (34). Hu H; Hu X; Yang W An integrated and versatile quantum mechanical/molecular mechanical simulation package <http://www.qm4d.info/>, 2016.
- (35). Hu H; Yang W Free energies of chemical reactions in solution and in enzymes with ab initio quantum mechanics/molecular mechanics methods. *Annu. Rev. Phys. Chem* 2008, 59 (1), 573–601. [PubMed: 18393679]
- (36). Hua S; Li W; Li S The generalized energy-based fragmentation approach with an improved fragmentation scheme: Benchmark results and illustrative applications. *ChemPhysChem* 2013, 14, 108–115. [PubMed: 23239545]
- (37). Ifimie R; Minary P; Tuckerman ME Ab initio molecular dynamics: Concepts, recent developments, and future trends. *Proc. Natl. Acad. Sci. U. S. A* 2005, 102 (19), 6654–6659. [PubMed: 15870204]
- (38). Jiang B; Guo H Permutation invariant polynomial neural network approach to fitting potential energy surfaces. *J. Chem. Phys* 2013, 139, 054112. [PubMed: 23927248]
- (39). Jiang B; Guo H Permutation invariant polynomial neural network approach to fitting potential energy surfaces. iii. molecule-surface interactions. *J. Chem. Phys* 2014, 141, 034109. [PubMed: 25053303]
- (40). Jorgensen WL; Chandrasekhar J; Madura JD; Impey RW; Klein ML Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys* 1983, 79 (2), 926–935.
- (41). Jorgensen WL; Madura JD Temperature and size dependence for monte carlo simulations of tip4p water. *Mol. Phys* 1985, 56 (6), 1381–1392.
- (42). DiStasio RA; Santra B; Li Z; Wu X; Car R The individual and collective effects of exact exchange and dispersion interactions on the ab initio structure of liquid water. *J. Chem. Phys* 2014, 141 (8), 084502. [PubMed: 25173016]
- (43). Kitaura K; Ikeo E; Asada T; Nakano T; Uebayasi M Fragment molecular orbital method: an approximate computational method for large molecules. *Chem. Phys. Lett* 1999, 313, 701–706.
- (44). Kitaura K; Sawai T; Asada T; Nakano T; Uebayasi M Pair interaction molecular orbital method: an approximate computational method for molecular interactions. *Chem. Phys. Lett* 1999, 312, 319–324.
- (45). Kunz A-PE; van Gunsteren WF Development of a nonlinear classical polarization model for liquid water and aqueous solutions: Cos/d. *J. Phys. Chem. A* 2009, 113 (43), 11570–11579. [PubMed: 19663490]
- (46). Lamoureux G; Harder E; Vorobyov IV; Roux B; MacKerell AD, Jr A polarizable model of water for molecular dynamics simulations of biomolecules. *Chem. Phys. Lett* 2006, 418, 245–249.
- (47). Li W; Li S; Jiang Y Generalized energy-based fragmentation approach for computing the ground-state energies and properties of large molecules. *J. Phys. Chem. A* 2007, 111, 2193–2199. [PubMed: 17388268]
- (48). Lu Z; Yang W Reaction path potential for complex systems derived from combined ab initio quantum mechanical and molecular mechanical calculations. *J. Chem. Phys* 2004, 121 (1), 89–100. [PubMed: 15260525]
- (49). Bishop CM. *Neural Networks for Pattern Recognition*; Oxford University Press: Oxford, 1996.
- (50). Mitchell TM. *Machine Learning*; McGraw-Hill: New York, 1997.
- (51). Ma Z; Tuckerman M Constant pressure ab initio molecular dynamics with discrete variable representation basis sets. *J. Chem. Phys* 2010, 133 (18), 184110. [PubMed: 21073216]
- (52). Ma Z; Zhang Y; Tuckerman ME Ab initio molecular dynamics study of water at constant pressure using converged basis sets and empirical dispersion corrections. *J. Chem. Phys* 2012, 137 (4), 044506. [PubMed: 22852630]
- (53). Morawietz T; Singraber A; Dellago C; Behler J How van der waals interactions determine the unique properties of water. *Proc. Natl. Acad. Sci. U. S. A* 2016, 113 (30), 8368–8373. [PubMed: 27402761]
- (54). Morrone JA; Car R Nuclear quantum effects in water. *Phys. Rev. Lett* 2008, 101, 017801. [PubMed: 18764152]

- (55). Nakano T; Kaminuma T; Sato T; Fukuzawa K; Akiyama Y; Uebayasi M; Kitaura K Fragment molecular orbital method: use of approximate electrostatic potential. *Chem. Phys. Lett* 2002, 351, 475–480.
- (56). Prendergast D; Grossman JC; Galli G The electronic structure of liquid water within density-functional theory. *J. Chem. Phys* 2005, 123, 014501. [PubMed: 16035849]
- (57). Price DJ; Brooks CL Modern protein force fields behave comparably in molecular dynamics simulations. *J. Comput. Chem* 2002, 23 (11), 1045–1057. [PubMed: 12116391]
- (58). Raghavachari K; Saha A Accurate composite and fragment-based quantum chemical models for large molecules. *Chem. Rev* 2015, 115 (12), 5643–5677. [PubMed: 25849163]
- (59). Reddy SK; Straight SC; Bajaj P; HuyPham C; Riera M; Moberg DR; Morales MA; Knight C; Gotz AW; Paesani F On the accuracy of the mb-pol many-body potential for water: Interaction energies, vibrational frequencies, and classical thermodynamic and dynamical properties from clusters to liquid water and ice. *J. Chem. Phys* 2016, 145 (19), 194504. [PubMed: 27875875]
- (60). Ren P; Ponder JW Polarizable atomic multipole water model for molecular mechanics simulation. *J. Phys. Chem. B* 2003, 107 (24), 5933–5947.
- (61). Rick SW Simulations of ice and liquid water over a range of temperatures using the fluctuating charge model. *J. Chem. Phys* 2001, 114 (5), 2276–2283.
- (62). Rupp M; Tkatchenko A; Müller K-R; von Lilienfeld OA Fast and accurate modeling of molecular atomization energies with machine learning. *Phys. Rev. Lett* 2012, 108, 058301. [PubMed: 22400967]
- (63). Sadeghi A; Ghasemi SA; Schaefer B; Mohr S; Lill MA; Goedecker S Metrics for measuring distances in configuration spaces. *J. Chem. Phys* 2013, 139, 184118. [PubMed: 24320265]
- (64). Schaefer B; Goedecker S Computationally efficient characterization of potential energy surfaces based on fingerprint distances. *J. Chem. Phys* 2016, 145, 034101. [PubMed: 27448868]
- (65). Shen L; Wu J; Yang W Multiscale quantum mechanics/molecular mechanics simulations with neural networks. *J. Chem. Theory Comput* 2016, 12, 4934–4946. [PubMed: 27552235]
- (66). So S-S; Karplus M Evolutionary optimization in quantitative structure-activity relationship: An application of genetic neural networks. *J. Med. Chem* 1996, 39, 1521–1530. [PubMed: 8691483]
- (67). Soper AK The radial distribution functions of water and ice from 220 to 673 K and at pressures up to 400 {MPa}. *Chem. Phys* 2000, 258, 121–137.
- (68). Thomsen JU; Meyer B Pattern recognition of the 1h {NMR} spectra of sugar alditols using a neural network. *J. Magn. Reson* 1989, 84 (1), 212–217.
- (69). van der Spoel D; van Maaren PJ; Berendsen HJC A systematic study of water models for molecular simulation: Derivation of water models optimized for use with a reaction field. *J. Chem. Phys* 1998, 108 (24), 10220–10230.
- (70). von Lilienfeld OA; Ramakrishnan R; Rupp M; Knoll A Fourier series of atomic radial distribution functions: A molecular fingerprint for machine learning models of quantum chemical properties. *Int. J. Quantum Chem* 2015, 115 (16), 1084–1093.
- (71). Wang L-P; Head-Gordon T; Ponder JW; Ren P; Chodera JD; Eastman PK; Martinez TJ; Pande VS Systematic improvement of a classical molecular model of water. *J. Phys. Chem. B* 2013, 117 (34), 9956–9972. [PubMed: 23750713]
- (72). Willow SY; Salim MA; Kim KS; Hirata S Ab initio molecular dynamics of liquid water using embedded-fragment second-order many-body perturbation theory towards its accurate property prediction. *Sci. Rep* 2015, 5, 14358. [PubMed: 26400690]
- (73). Yang W Direct calculation of electron density in density-functional theory. *Phys. Rev. Lett* 1991, 66, 1438–1441. [PubMed: 10043209]
- (74). Yao K; Herr JE; Parkhill J The many-body expansion combined with neural networks. *J. Chem. Phys* 2017, 146 (1), 014106. [PubMed: 28063436]
- (75). Zhang C; Spanu L; Galli G Entropy of liquid water from ab initio molecular dynamics. *J. Phys. Chem. B* 2011, 115 (48), 14190–14195. [PubMed: 21961845]
- (76). Zhang C; Wu J; Galli G; Gygi F Structural and vibrational properties of liquid water from van der waals density functionals. *J. Chem. Theory Comput* 2011, 7 (10), 3054–3061. [PubMed: 26598149]

- (77). Zhang Da W.; Zhang JZH Molecular fractionation with conjugate caps for full quantum mechanical calculation of proteinscmo-lecule interaction energy. *J. Chem. Phys* 2003, 119 (7), 3599–3605.
- (78). Zhang Y; Lee T-S; Yang W A pseudobond approach to combining quantum mechanical and molecular mechanical methods. *J. Chem. Phys* 1999, 110 (1), 46–54.
- (79). Zhu L; Amsler M; Fuhrer T; Schaefer B; Faraji S; Rostami S; Ghasemi SA; Sadeghi A; Grauzinyte M; Wolverson C; Goedecker S A fingerprint based metric for measuring similarities of crystalline structures. *J. Chem. Phys* 2016, 144, 034203. [PubMed: 26801027]

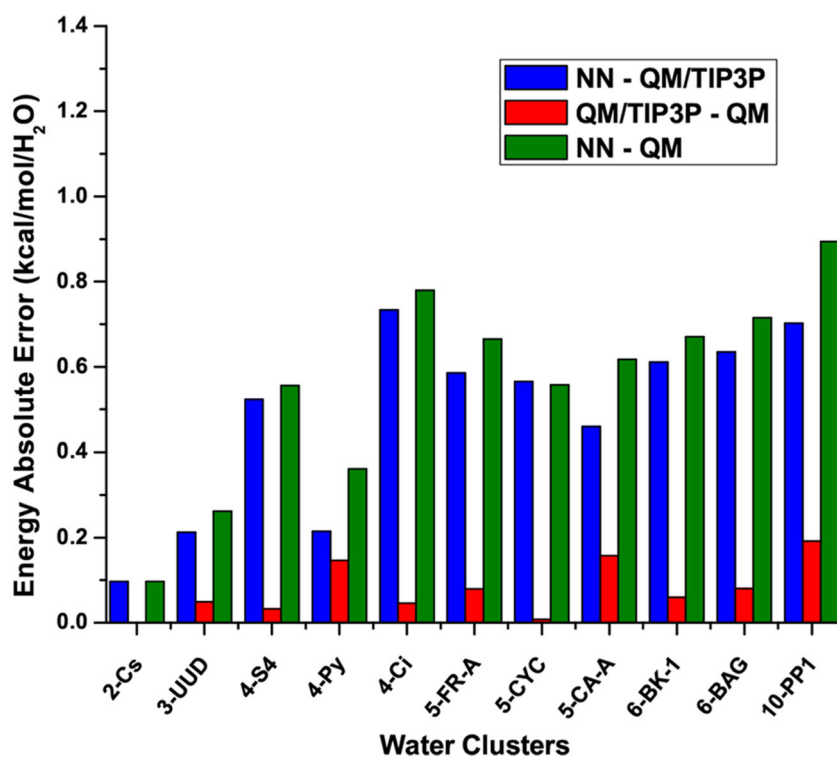


**Figure 1.** Schematic representation of neural network structure for electrostatically embedded water monomer and dimer.  $\{G_i\}$  represents power spectrum of atom  $i$  including both QM and MM environment.  $NN_i$  represents atomic neural network for atom  $i$ .  $e_i$  represents the atomic energy of atom  $i$ .  $E_{1/2}$  is the energy of electrostatically embedded water monomer or dimer.

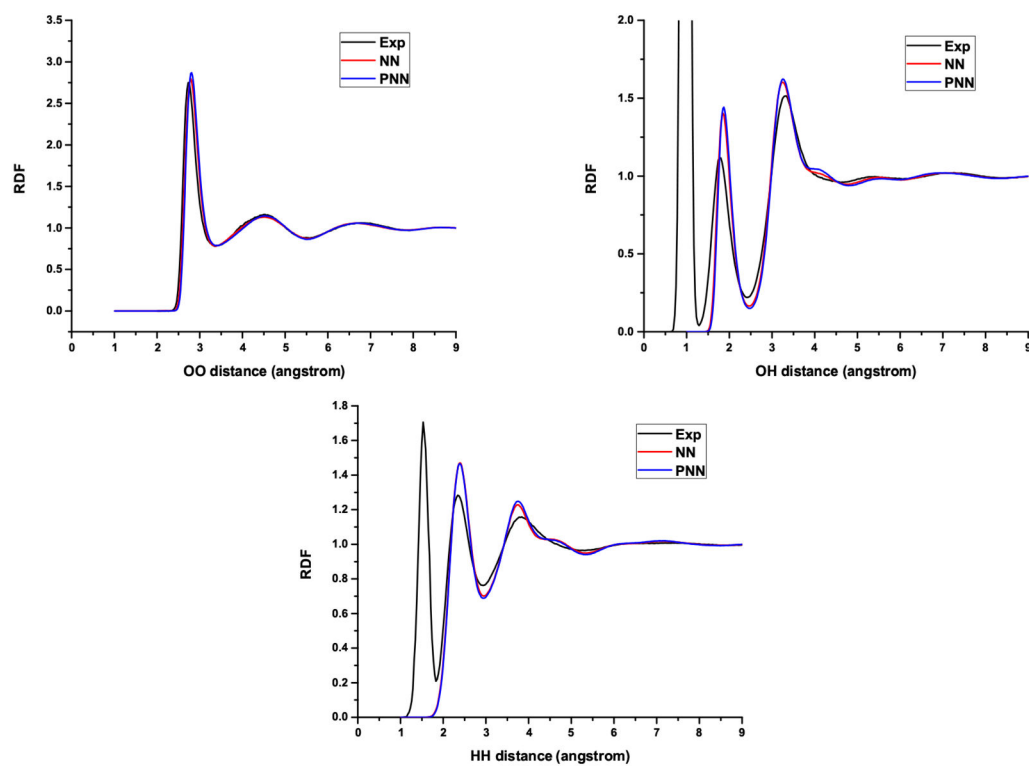


**Figure 2.** Geometric structure of water clusters used to test an NN force field with TIP3P embedding charges.

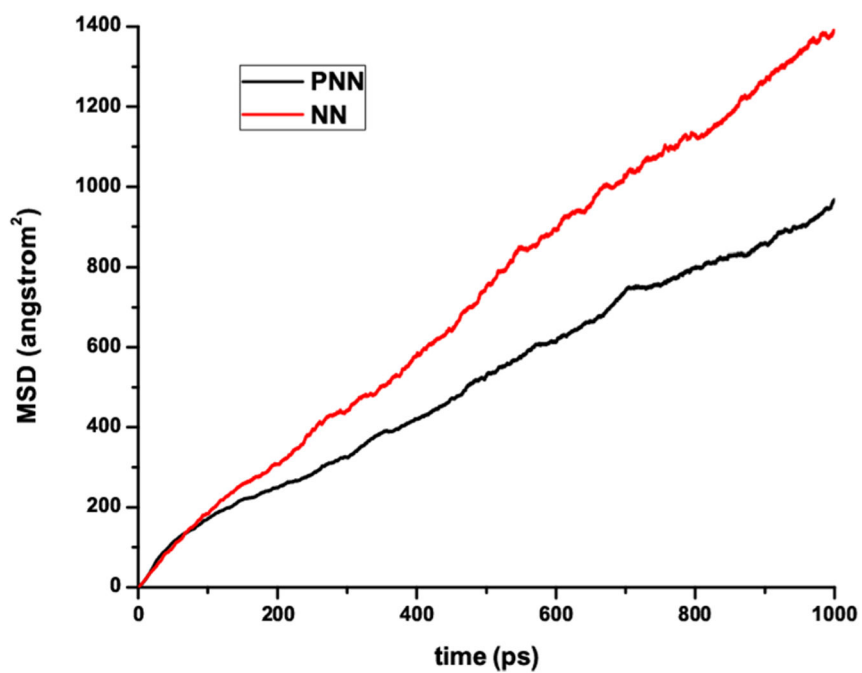




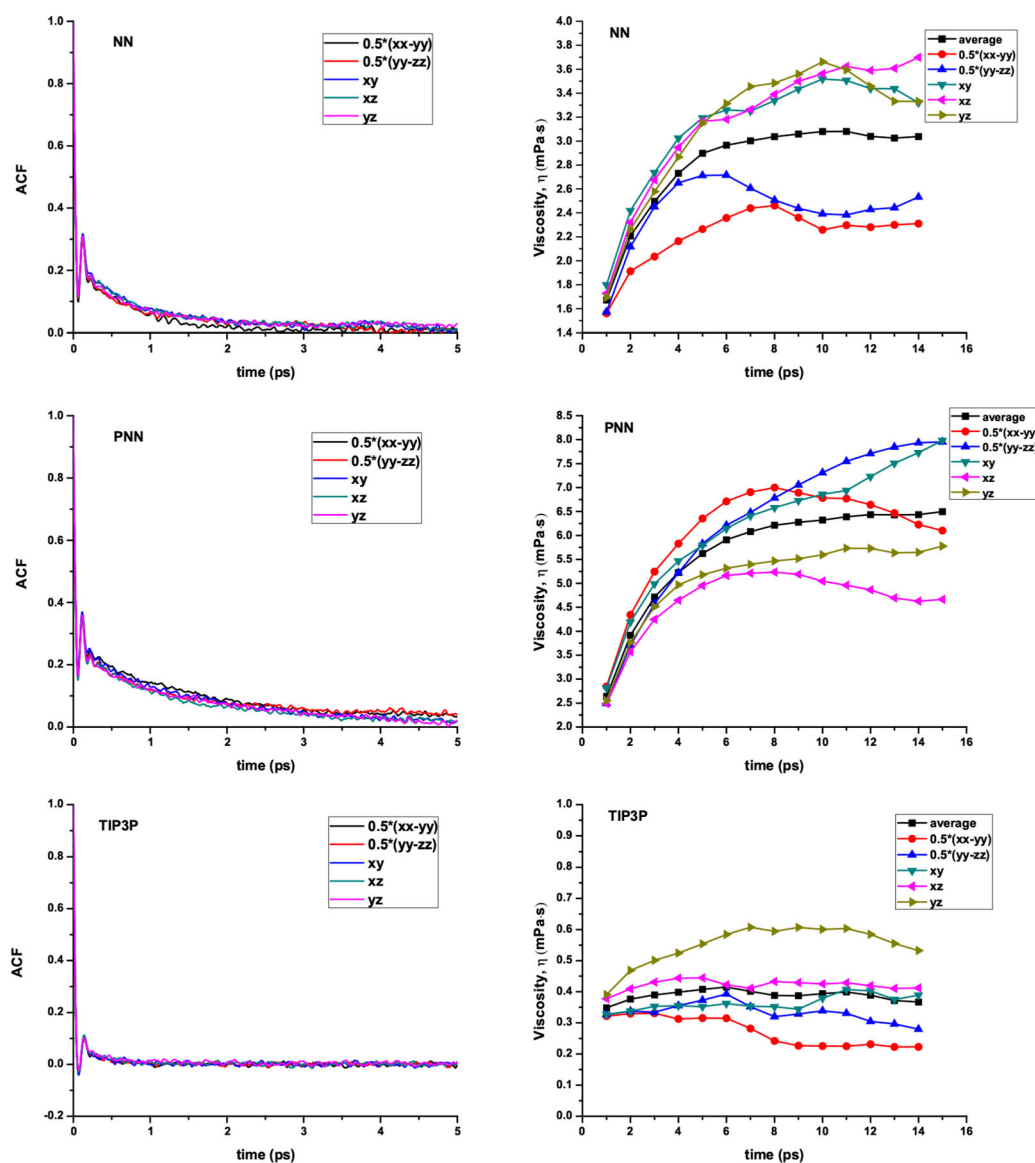
**Figure 3.** Absolute error of binding energy for water clusters of different size. NN represents an NN force field with TIP3P embedding charges. QM/TIP3P represents electrostatically embedded two-body expansion at the level of CCSD/aug-cc-pVDZ using TIP3P charges. QM represents full quantum mechanical calculations at the level of CCSD/aug-cc-pVDZ.



**Figure 4.** Radial distribution function of oxygen–oxygen, oxygen–hydrogen, and hydrogen–hydrogen (from left to right) with nonpolarizable (NN) and polarizable NN (PNN) force fields are compared with experimental results.



**Figure 5.** Average mean squared deviation of atomic positions with nonpolarizable (NN) and polarizable (PNN) neural network force fields.



**Figure 6.** Stress tensor autocorrelation function and shear viscosity. Top: the left one is the autocorrelation function of the stress tensor element for nonpolarizable NN force field; the right one is the shear viscosity changing with time for nonpolarizable NN force field. Middle: the left one is the autocorrelation function of the stress tensor element for polarizable NN force field; the right one is the shear viscosity changing with time for polarizable NN force field. Bottom: the left one is the autocorrelation function of the stress tensor element for TIP3P; the right one is the shear viscosity changing with time for TIP3P.

**Table 1.**Accuracy of Neural Networks for Electrostatically Embedded Water Monomers and Dimers<sup>a</sup>

NN	energy (kcal/mol/H <sub>2</sub> O)		force (kcal/mol/Å)	
	training set	testing set	training set	testing set
monomer	0.16	0.22	0.77	0.86
dimer	0.06	0.05	0.66	0.54

<sup>a</sup>Root mean squared deviation of energy and force for monomer ( $E_1[i;Q_j]$ ) and dimer ( $E_2[i,j;Q_1Q_2]-E_1[i;Q_1Q_2]-E_1[j;Q_1Q_2]$ ).

**Table 2.**

## Self-Diffusion Coefficient

	experiment	TIP3P	NN <sup>a</sup>	PNN <sup>b</sup>
$D(\text{\AA}^2/\text{ps})$	0.23	0.52	0.23	0.16

<sup>a</sup>Nonpolarizable neural network force field.

<sup>b</sup>Polarizable neural network force field.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 3.**Shear Viscosity of TIP3P, Nonpolarizable NN, and Polarizable NN Force Fields<sup>a</sup>

shear viscosity (mPa·s)	TIP3P	NN <sup>b</sup>	PNN <sup>c</sup>	experiment
				0.85
average	0.39	3.05	6.41	
0.5*(xx - yy)	0.23	2.30	6.10	
0.5*(yy - zz)	0.30	2.40	7.95	
xy	0.39	3.48	7.95	
xz	0.41	3.60	4.60	
yz	0.54	3.31	5.65	

<sup>a</sup>Five independent components and average results are shown.<sup>b</sup>Nonpolarizable NN force fields.<sup>c</sup>Polarizable NN force fields.