Liquid water simulation using hydrogen bond corrected SCAN and neural network potentials.

Timothy T. Duignan*

School of Chemical Engineering, The University of Queensland, St Lucia, Brisbane 4072, Australia

E-mail: t.duignan@uq.edu.au

Abstract

Accurately reproducing the structure of liquid water with *ab initio* molecular dynamics (AIMD) simulation is a crucial first step on the path towards accurately predicting the properties of liquid solutions without relying on experiment. Density functional theory (DFT) is normally used to approximate the forces in these simulations. However, no DFT functional has been shown to give an entirely satisfactory description of the structure of liquid water. Here, I propose a simple correction to the strongly constrained and appropriately normalised (SCAN) DFT functional, that corrects the strength of the hydrogen bonding interaction with a simple exponential potential fitted to dimer energy calculations. The resulting SCAN-CH functional provides an excellent description of the structure of liquid water. Long time scale NPT simulations are enabled by the use of neural network potentials, which demonstrate that the simulations are well converged and that the density of water is also more accurately reproduced with this method.

1

Introduction

The accurate simulation of aqueous systems is a fundamental and central problem in physical chemistry that is relevant to a vast range of important applications. *Ab initio* molecular dynamics (AIMD) using density functional theory is an extremely promising tool for this purpose due to its generalisability and freedom from empirical parameterisation.

However, current density functional theory (DFT) functionals are not accurate enough to reliably solve this problem.[1] For example, the most widely used dispersion corrected BLYP-D3 functional has significant errors in its oxygen-oxygen radial distribution function compared with experiment.[2] Other functionals such as revPBE-D3 have significant errors in the forces computed on clusters.[3] Riera et al.[4] have shown that there are significant errors in the cluster energies and many body interactions with these functionals. These problems urgently need to be addressed before this approach can be used to simulate a wider class of aqueous systems such as electrolyte solutions.

The strongly constrained and appropriately normalised (SCAN) DFT functional has shown promise in its ability to reproduce the relative energies changes for a range of water clusters very well, as well as many other systems.[5] One issue is that SCAN has relatively high computational cost to run. However, recent work has nicely demonstrated that deep learning can be used to overcome this problem, by developing neural network potentials (NNPs) that are trained on the forces from a short run. Much larger and longer trajectories can then be obtained.[6–8] I also use this approach here.

However, an additional issue with this functional is that it slightly over structures the radial distribution function (RDF) of pure water. This is why it is often simulated at 330 K instead of 300K to correct for this effect. This is sometimes justified with attribution to the neglect of nuclear quantum effects. (NQEs) This is not a good way to account for these effects however.[9] Additionally, reproducing the sensitive experimental properties of aqueous solutions is obviously going to be difficult if there is a 30 degree discrepancy in temperature. The density of SCAN is also 5 % too high. A better method of correcting for this problem

is therefore needed before using it more widely.

One option to correct these issues is to go to higher levels of theory such as the random phase approximation, density corrected or hybrid SCAN or MP2.[8,10–13] But these can be prohibitively expensive even to obtain the training data to run simulations and inaccuracies can remain.

A simpler and more straightforward method of correcting for this error in the SCAN functional is therefore desirable. It is known that the source of the error is the over stabilisation of hydrogen bonding between water molecules in the two body interaction.[14] Therefore, one solution is to simply add a pair-wise repulsive potential that acts between an oxygen atom and the hydrogen atom on the nearest water molecule it is hydrogen bonded to. The size and shape of this correction can be fitted to remove the error in the dimer interaction energy. This correction should be rigorous as the error in the SCAN functional is contained in the two body term. Here, I demonstrate that this method works well, showing that the SCAN functional with a simple hydrogen bonding correction (SCAN-CH) very accurately reproduces the structure and density of pure water. I demonstrate this with both short AIMD runs and using a neural network potential to perform much longer NPT simulations.

Figure 1 shows the errors in the oxygen-oxygen radial distribution function (RDF) with SCAN compared to experimental X-Ray diffraction (XRD) measurements.[15] The cause of this error is clear from Figure 2, which shows the error in the interaction energy of water dimers with SCAN as a function of distance between the oxygen atom and the nearest neighbouring hydrogen atom. The reference level of theory is CCSD(T) with aug-cc-pVTZ level basis sets. The dimer structures were extracted from AIMD simulations by taking pairs of water molecules from the closest eight water molecules to a central water molecule. It clearly shows that SCAN has an overly strong binding, which explains the poor structural performance of SCAN.

The red line in this Figure 2 is a simple exponential curve fitted to match this error given
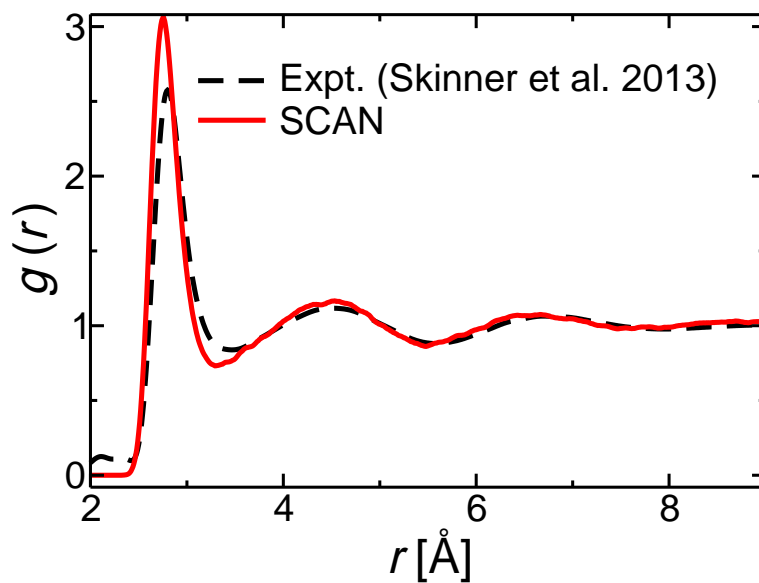
Figure 1: Oxygen-oxygen RDF of water with SCAN compared with experimental XRD data.[15]
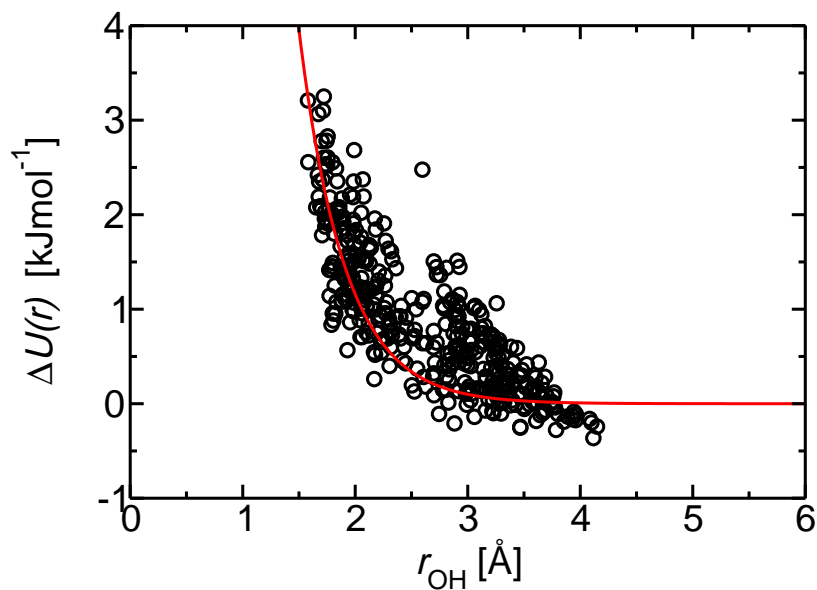

Figure 2: Error of SCAN compared with CCSD(T) in the water-water dimer interaction energy as a function of distance between the oxygen and nearest neighbouring hydrogen atom. (Black circles) Correction potential used to remove this error is shown in red.

by:

$$U(r) = Ae^{-br} \qquad (1)$$

where A= 160 and b= 2.5 Å^{-1}.

Least squares fitting was used to determine this correction potential. Only distances below 1.85 Å were used for the fitting in order to derive a potential that primarily acts only on the closest hydrogen atom. Fitting to the full data results in an overly repulsive potential longer range potential that acts on both hydrogens of the neighbouring water molecule. Including this correction potential strongly reduces the error in the pairwise interaction energy as shown in Figure 3. I use an oxygen-hydrogen correction potential as this will give the smallest alteration to the overall potential energy landscape, whereas an oxygen-oxygen repulsion would act even on water molecules that are not hydrogen bonded. The sampling extracted from the simulation does not include enough cases of water molecules close to each other that are not hydrogen bonded therefore including an oxygen oxygen correction could lead to errors in the overall potential energy surface.

I refer to this SCAN functional including this hydrogen bond correction as SCAN-CH. AIMD simulations can be performed with SCAN-CH resulting in the RDF shown in Figure 4 showing much improved agreement with experiment.

A small difference remains in the first peak position. But this can be attributed to the neglect of NQEs in this simulation. To test this I compare with simulations using the many body polarisable potential (MB-Pol) performed without NQEs.[16] MB-Pol is known to excellently reproduce many of water's properties.[17] Figure 5 shows excellent agreement for the first peak. Figure 6 also shows excellent agreement between the oxygen-hydrogen RDF and the hydrogen-hydrogen RDF for SCANCH in comparison with MBPOL.

Finally, these simulations are only 15 ps long. This is quite short and I cannot rule out the possibility of equilibration issues. Additionally, they use fixed volume not fixed pressure ensembles so the density is fixed to the correct density. To explore longer time scales and constant pressure I use DeepMD-kit software package[18] to develop two NNPs, which learn
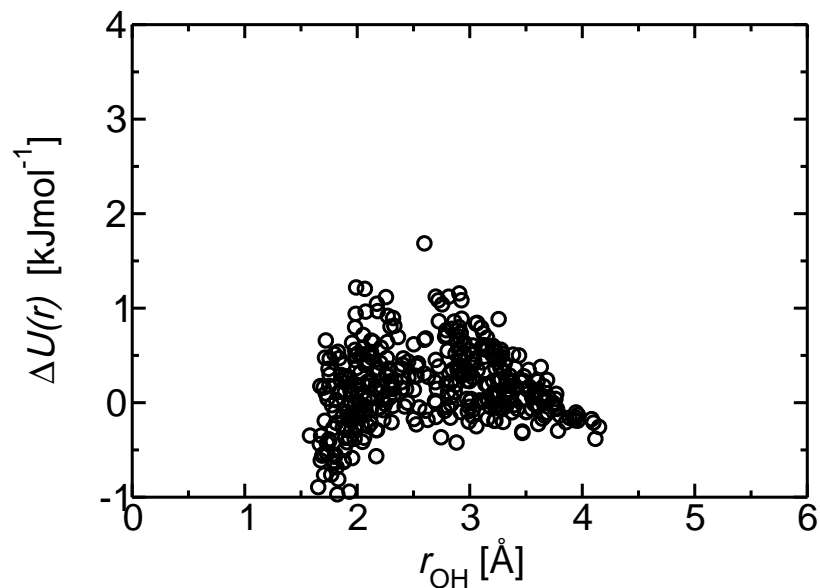
Figure 3: Error of SCAN-CH compared with CCSD(T) in the water-water dimer interaction energy as a function of distance between the oxygen and nearest neighbouring hydrogen atom after correction with exponential repulsion potential. (Black circles)
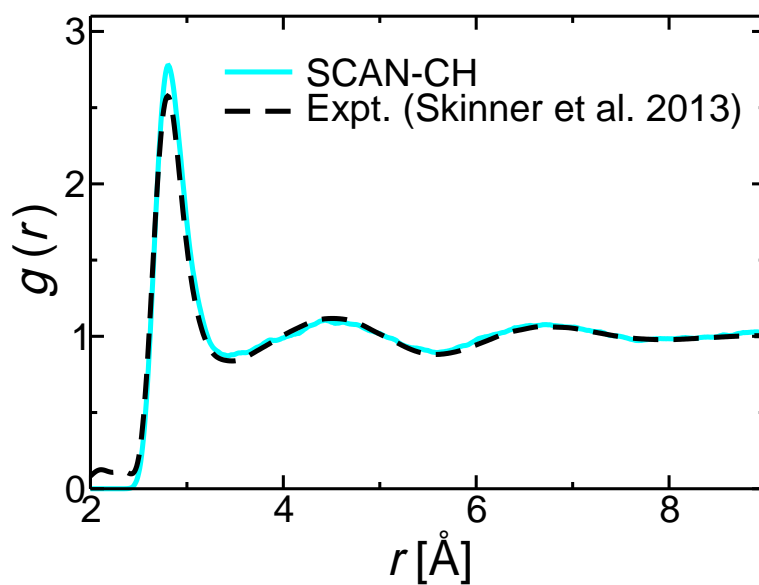

Figure 4: Oxygen-oxygen RDF of water with the hydrogen bond corrected SCAN functional (SCAN-CH) compared with experimental XRD data.[15]
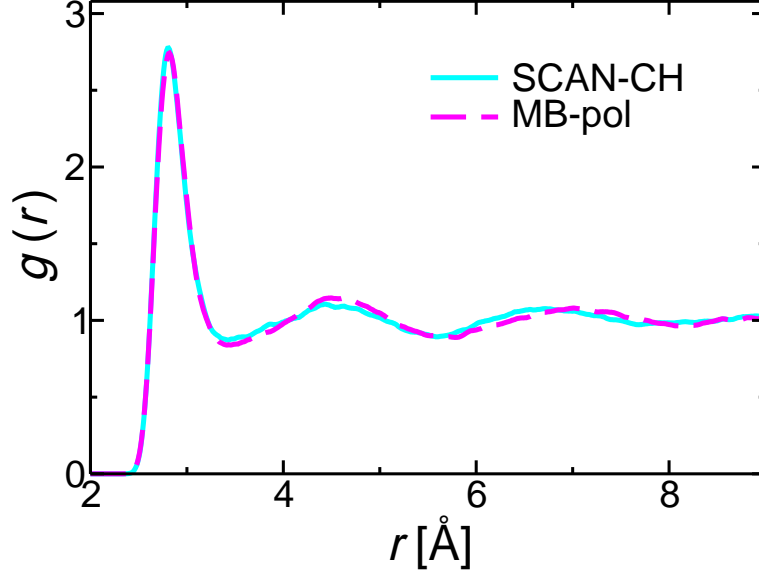
Figure 5: Oxygen-oxygen RDF of water with the hydrogen bond corrected SCAN functional (SCAN-CH) compared with MB-Pol simulation on an identical system.
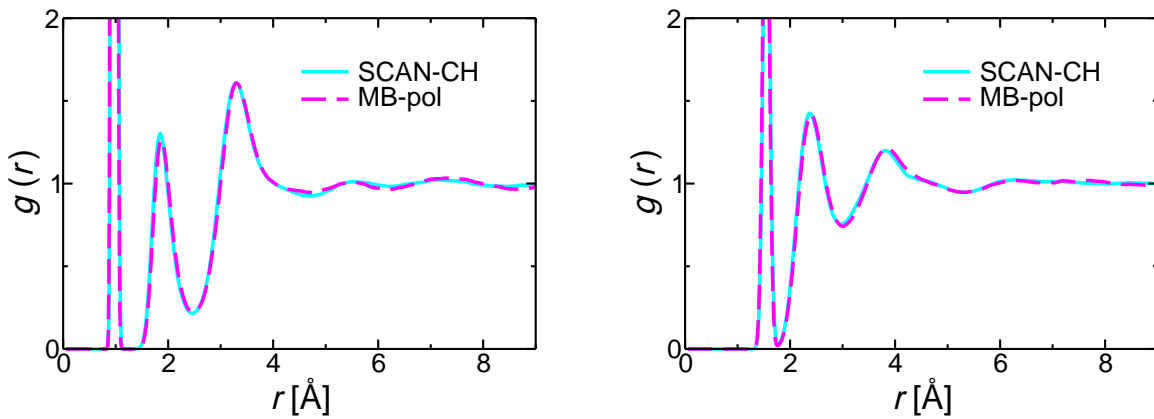

Figure 6: Comparison of the oxygen-hydrogen RDF (left) and hydrogen-hydrogen RDF (right) of water with the hydrogen bond corrected SCAN functional (SCAN-CH) compared with MB-Pol simulation on an identical system.

the forces on the water systems with SCAN-CH. The resulting NNPs are then used to run much longer time scale NPT simulations. The RDFs agree excellently as shown in Figure 7. The resulting density is 1.026 ± 0.008 g/ml very close to the experimental density of 1.00 g/ml and significantly better than the uncorrected SCAN density of approximately 1.05. Most importantly this small difference in density does not appear to lead to a significant difference in the RDF.
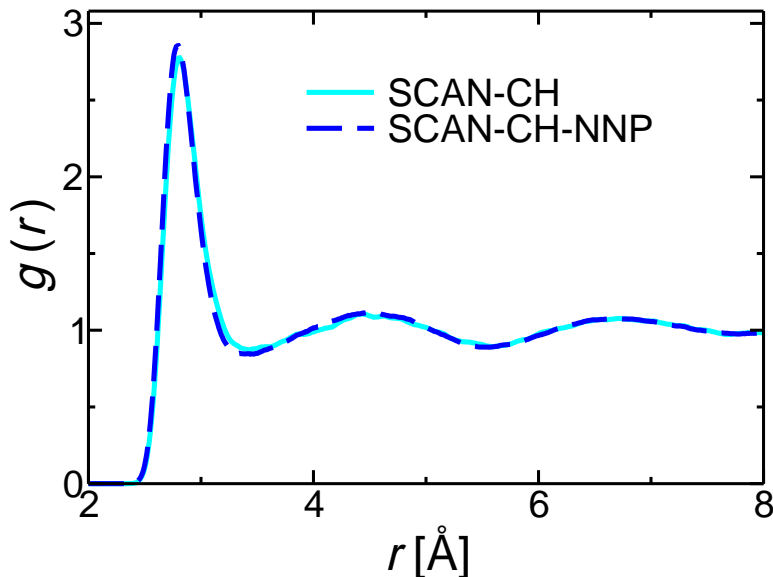

Figure 7: Oxygen-oxygen RDF of water with the hydrogen bond corrected SCAN functional (SCAN-CH) with NVT compared with a NNP fitted to (SCAN-CH) with much longer statistics run with NPT showing good agreement.

In summary, it is clear that a simple correction to remove the error in the short range two body hydrogen bonding interaction of SCAN results in an excellent description of liquid water structure. The next goal is to use this method for the simulation of more complex aqueous solutions.

Computational Details

Ab initio Simulations

Born-Oppenheimer *ab initio* molecular dynamics simulations within the NVT (at 300 K) ensemble using periodic boundary conditions were performed within the `CP2K` simulation suite (http:www.cp2k.org) containing the *QuickStep* module for the DFT calculations.[19] A 0.5 fs time step was used. 97 water molecules in a 14.3^3 Å^3 sized cell was used. A double ζ basis set that has been optimized for the condensed phase[20] was used in conjunction with GTH pseudopotentials[21] using a 1200 Ry cutoff.[22,23] Pseudopotentials optimised for the SCAN functional were used.[24] A Nosé-Hoover thermostat was attached to every degree of freedom to ensure equilibration.[25] The structures were accumulated for $\approx$ 15 ps after 3 ps of equilibration.

MB-Pol Simulations

Simulations of MB-Pol were performed under identical conditions to the AIMD simulations using the MBX program interfaced[26] with i-PI.[27] The strucutres were accumulated for $\approx$ 40 ps after 3 ps of equilibration.

Cluster calculations

The geometries of the 400 water dimers were extracted from a revPBE-D3 simulation of water by extracting dimers from the 9 water molecule cluster structures used in a previous work.[3] ORCA[28] was used to calculate the dimer energies at the CCSD(T) level of theory. The aug-cc-pVTZ basis set was used for the oxygen and hydrogen atoms.[29] Frozen cores were used for the CCSD(T) calculations. For the SCAN cluster energy calculations `CP2K` was used with the periodicity none option and a larger cell size to remove any box size dependence. Otherwise, the same parameters, basis sets *etc.* as the simulation were used.

NNP simulations

The DeepMD-kit package[18] software package was used to train a Deep Potential - Smooth Edition[30] neural network potential fitted to the output of the AIMD SCAN-CH simulations. It was trained on forces and energies extracted from the AIMD simulation every 10 fs to give 1866 frames. Two separate NNPs were trained. One used on an angular descriptor with a 6 Å cutoff. The other used both an angular and radial descriptors with 6 Å cutoff for the radial descriptor and 4 Å for the angular. A three layer fitting net with 240 nodes in each layer was used for both potentials.

The DeepMD interface with LAMMPS[31] was then used to run constant pressure (NPT) molecular dynamics simulations for 150 ps with 6ps of equilibration discarded at 300 K with Nose-Hoover thermo/barostat. The difference between the predictions of the two NNPs were monitored to ensure a good fit. The deviation in energies between the two NNPs remained stable at 1.5 eV and the maximum deviation in forces remained below 0.15 eV/Å.

Acknowledgements

References

(1) Gillan, M. J.; Alfè, D.; Michaelides, A. *J. Chem. Phys.* **2016**, *144*, 130901.

(2) Galib, M.; Duignan, T. T.; Misteli, Y.; Baer, M. M. D.; Schenter, G. K. G.; Hutter, J.; Mundy, C. J. C. *J. Chem. Phys.* **2017**, *146*, 244501.

(3) Duignan, T. T.; Schenter, G.; Mundy, C. J.; Zhao, X. S. *J. Chem. Theory Comput.* **2020**, *16*, 5401–5409.

(4) Riera, M.; Lambros, E.; Nguyen, T. T.; Götz, A. W.; Paesani, F. *Chem. Sci.* **2019**, *10*, 8211–8218.

(5) Sun, J.; Remsing, R. C.; Zhang, Y.; Sun, Z.; Ruzsinszky, A.; Peng, H.; Yang, Z.; Paul, A.; Waghmare, U.; Wu, X.; Klein, M. L.; Perdew, J. P. *Nat. Chem.* **2016**, *8*, 831–836.

(6) Piaggi, P. M.; Panagiotopoulos, A. Z.; Debenedetti, P. G.; Car, R. *J. Chem. Theory Comput.* **2021**, *17*, 3065–3077.

(7) Torres, A.; Pedroza, L. S.; Fernandez-Serra, M.; Rocha, A. R. *J. Phys. Chem. B* **2021**, 10.1021/acs.jpcb.1c04372.

(8) Zhang, C.; Tang, F.; Chen, M.; Xu, J.; Zhang, L.; Qiu, D. Y.; Perdew, J. P.; Klein, M. L.; Wu, X. *J. Phys. Chem. B* **2021**, 10.1021/acs.jpcb.1c03884.

(9) Li, C.; Paesani, F.; Voth, G. A. *ChemRxiv* **2021**, 10.33774/chemrxiv–2021–c603x.

(10) Duignan, T.; Schenter, G. K.; Fulton, J.; Huthwelker, T.; Balasubramanian, M.; Galib, M.; Baer, M. D.; Wilhelm, J.; Hutter, J.; Ben, M. D.; Zhao, X. S.; Mundy, C. J. *Phys. Chem. Chem. Phys.* **2020**, *22*, 10641–10652.

(11) Yao, Y.; Kanai, Y. *J. Phys. Chem. Lett.* **2021**, *12*, 63546362.

(12) Lan, J.; Wilkins, D. M.; Rybkin, V. V.; Iannuzzi, M.; Hutter, J. *ChemRxiv* **2021**, 10.33774/chemrxiv–2021–n32q8.

(13) Dasgupta, S.; Lambros, E.; Perdew, J. P.; Paesani, F. *ChemRxiv* 10.33774/chemrxiv–2021–hstgf.

(14) Lambros, E.; Hu, J.; Paesani, F. *J. Chem. Theory Comput.* **2021**, 10.1021/acs.jctc.1c00141.

(15) Skinner, L. B.; Huang, C.; Schlesinger, D.; Pettersson, L. G. M.; Nilsson, A.; Benmore, C. J. *J. Chem. Phys.* **2013**, *138*, 074506.

(16) Medders, G. R.; Babin, V.; Paesani, F. *J. Chem. Theory Comput.* **2014**, *10*, 2906–2910.

(17) Paesani, F. *Handb. Mater. Model.* **2018**, 1–25.

(18) Wang, H.; Zhang, L.; Han, J.; E, W. *Comput. Phys. Commun.* **2018**, *228*, 178–184.

(19) VandeVondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J. *Comput. Phys. Commun.* **2005**, *167*, 103–128.

(20) VandeVondele, J.; Hutter, J. *J. Chem. Phys.* **2007**, *127*, 114105.

(21) Goedecker, S.; Teter, M.; Hutter, J. *Phys. Rev. B* **1996**, *54*, 1703–1710.

(22) Miceli, G.; Hutter, J.; Pasquarello, A. *J. Chem. Theory Comput.* **2016**, *12*, 3456–3462.

(23) Yao, Y.; Kanai, Y. *J. Chem. Theory Comput.* **2018**, *14*, 884–893.

(24) Hutter, J. https://github.com/juerghutter/GTH/blob/master/SCAN/POTENTIAL. 2021.

(25) Martyna, G. J.; Klein, M. L.; Tuckerman, M. *J. Chem. Phys.* **1992**, *97*, 2635–2643.

(26) Paesani, F.; Riera, M. MBX. 2021.

(27) Kapil, V. et al. *Comput. Phys. Commun.* **2019**, *236*, 214–223.

(28) Neese, F. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2*, 73–78.

(29) Dunning, Jr., T. H. *J. Chem. Phys.* **1989**, *90*, 1007–1023.

(30) Zhang, L.; Han, J.; Wang, H.; Saidi, W. A.; Car, R.; Weinan, E. *Adv. Neural Inf. Process. Syst.* **2018**, *2018-Decem*, 4436–4446.

(31) Plimpton, S. *J. Comput. Phys.* **1995**, *117*, 1–19.