

# Deep Potential: A General Representation of a Many-Body Potential Energy Surface

Jiequn Han<sup>1,†</sup>, Linfeng Zhang<sup>1,†</sup>, Roberto Car<sup>1,2</sup> and Weinan E<sup>1,3,4,\*</sup>

<sup>1</sup> Program in Applied and Computational Mathematics, Princeton University, Princeton, NJ 08544, USA.

<sup>2</sup> Department of Chemistry, Department of Physics, and Princeton Institute for the Science and Technology of Materials, Princeton University, Princeton, NJ 08544, USA.

<sup>3</sup> Department of Mathematics, Princeton University, Princeton, NJ 08544, USA.

<sup>4</sup> Center for Data Science and Beijing International Center for Mathematical Research, Peking University, and Beijing Institute of Big Data Research, Beijing, 100871, China.

Received 5 October 2017; Accepted 17 October 2017

---

**Abstract.** We present a simple, yet general, deep neural network representation of the potential energy surface for atomic and molecular systems. It is “first-principle” based, in the sense that no *ad hoc* approximations or empirical fitting functions are required. When tested on a wide variety of examples, it reproduces the original model within chemical accuracy. This brings us one step closer to carrying out molecular simulations with quantum mechanics accuracy at empirical potential computational cost.

**AMS subject classifications:** 81V55, 92E10, 68T99

**Key words:** Potential energy surface, deep learning, molecular simulation.

---

## 1 Introduction

A representation of the potential energy surface (PES) for general systems of atoms or molecules is the basic building block for molecular dynamics (MD) and/or Monte Carlo (MC) simulations, which are common tools in many disciplines, including chemistry, physics, biology, and materials science. Until now, this problem has been addressed using two very different approaches. At one extreme, empirical potentials have been constructed by fitting limited experimental and/or numerical data from accurate quantum

---

<sup>†</sup>These authors contributed equally to this work.

\*Corresponding author. *Email addresses:* weinan@math.princeton.edu (W. E), jiequnh@princeton.edu (J. Han), linfengz@princeton.edu (L. Zhang), rcar@princeton.edu (R. Car)

mechanical calculations. Well-known examples include the Lennard-Jones potential [1], the Stillinger-Weber potential [2], the embedded-atom method (EAM) potential [3], the CHARMM [4]/AMBER [5] force fields, the reactive force fields [6], etc. These potentials are computationally efficient, allowing large-scale simulations (up to millions of atoms), but their construction is very much an art and their accuracy and transferability are limited. At the other extreme, methods based on first-principle quantum theory such as density functional theory (DFT) [7] have been proposed, the most well-known example being the *ab initio* molecular dynamics (AIMD) [8] scheme. These methods promise to be much more accurate but they are also computationally expensive, limiting our ability to handling systems of hundreds to thousands of atoms only. Until recently, the drastic disparity between these two approaches in terms of accuracy and computational cost has been a major dilemma to be confronted with in molecular simulation.

Recent advances in machine learning, particularly deep learning, have ushered some new hope in addressing this dilemma [9–20]. Several promising new ideas have been suggested, in which deep neural networks are used to represent the potential energy surface. Of particular interest are the Behler-Parrinello neural network (BPNN) [10] and the deep tensor neural network (DTNN) [9]. BPNN uses the so-called symmetry functions as input and a standard neural network as the fitting function; DTNN, on the other hand, uses as input a vector of nuclear charges and an inter-atomic distance matrix, and introduces a sequence of interaction passes where “the atom representations influence each other in a pair-wise fashion” [9]. Both methods are able to predict with great accuracy the potential energy surface of materials in condensed phase, in the case of BPNN, and of small organic molecules, in the case of DTNN. However, the construction of the local symmetry functions for BPNN contains an *ad hoc* and often tedious component where hand-crafted fitting functions and human intervention are required. On the other hand, methods like DTNN may face difficulties when one tries to transfer them directly to large systems.

In this work, we develop a new method, called Deep Potential (DP), that successfully addresses the inadequacies of the existing models. Deep Potential is a simple, yet general, end-to-end deep neural network representation of a many-atom potential energy surface. The network uses as input the raw coordinates of the atoms in a proper frame of reference, and naturally respects the symmetries of the system. Promising results are obtained in a variety of test cases, including small molecular isomers and condensed-phase systems. The Deep Potential method brings us closer to performing molecular modeling with the secure accuracy of first-principle based methods at a computational cost comparable to that of empirical potentials.

## 2 Results

### 2.1 Deep Potential framework

Our goal is to formulate a general and direct end-to-end representation of the many-body potential energy surface that uses the atomic configurations directly as the only

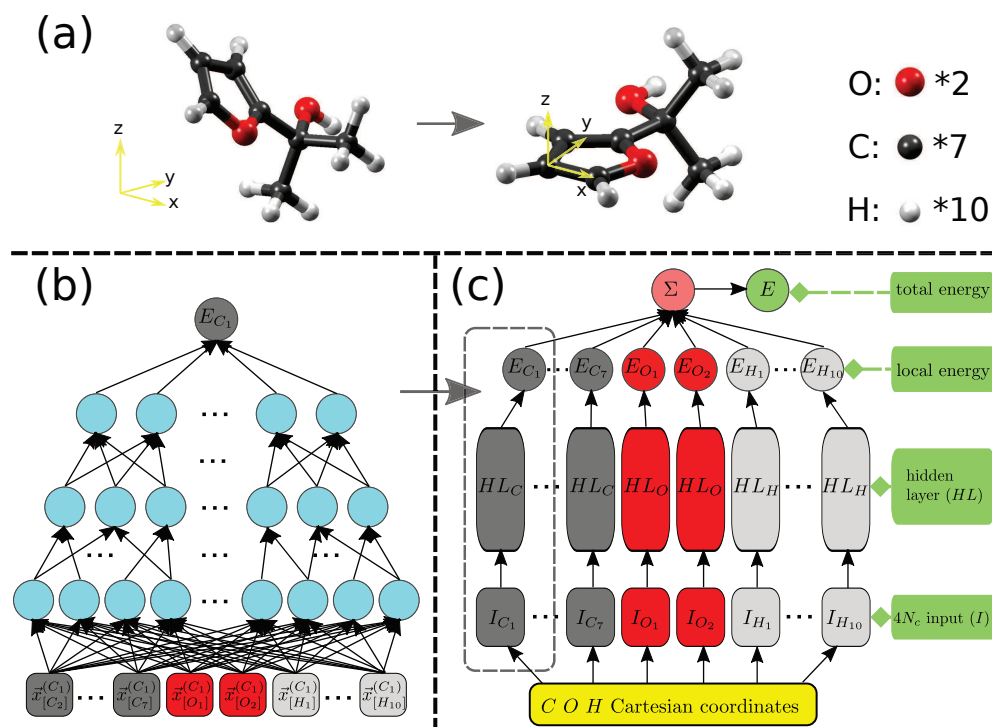


Figure 1: (color online). Schematic diagram of Deep Potential, using C<sub>7</sub>O<sub>2</sub>H<sub>10</sub> as an illustrative example. (a) Transformation from global to local reference frame for a Carbon atom. (b) Sub-network structure for atom C<sub>1</sub>. We use the notion  $\vec{x}_{[A_i]}^{(C_1)}$  to represent the coordinates of the atom A<sub>i</sub> in the list of neighbors of atom C<sub>1</sub>. The atoms A<sub>i</sub> are assigned to different groups corresponding to the different atomic species. Within each one of these groups the atoms are listed in order of increasing distance from C<sub>1</sub>. (c) Full structure of the Deep Potential network.

input data, without any explicit decomposition of two-body or three-body contributions. The main challenge to achieve this goal is to design a deep neural network that obeys the important symmetries of the system like it was achieved, for instance, with the convolutional neural network in computer vision problems [21]. Besides the usual translational and rotational symmetries, we also have the permutational symmetry with respect to the chemically equivalent atoms. We represent the potential energy surface by following the steps that are schematically indicated in Fig. 1.

For a system of  $N$  atoms, our neural network consists of  $N$  small, almost independent, copies. Each copy is a sub-network with a scalar output, corresponding to a different atom in the system. The size of the input data for a sub-network is at most  $4N_c$  (see the detailed description below), where  $N_c$  is the number of atoms within the adopted cut-off radius  $R_c$  (Fig. 1 (b) and (c)). If the number of atoms within  $R_c$  fluctuates,  $N_c$  is the largest fluctuating number. We find that adopting a finite  $R_c$  is sufficient in all the extended material systems considered here.

Next, we introduce a local Cartesian coordinate frame for each atom. In this frame,

the atom under consideration is taken to be at the origin and is labelled as “0”. We fix the  $x$  and  $y$  axes in terms of the atom 0 and its two non-collinear nearest neighbors that we label “1” and “2”, respectively, in order of increasing distance. For organic molecules, we exclude hydrogen atoms in the definition of “1” and “2”. The  $x$ -axis is defined by the 0-1 direction, the  $z$ -axis is along the direction of the cross product of 0-1 and 0-2, and the  $y$ -axis is the cross product of  $z$  and  $x$ . In this Cartesian frame, the coordinates of all the atoms falling inside the cut-off radius centered at the origin, excluding atom 0, define the input data for the corresponding sub-network. In practice, we found that the combination  $\{1/r, \cos\theta, \cos\phi, \sin\phi\}$ , where  $(r, \theta, \phi)$  are the polar coordinates, is a much better representation than the Cartesian coordinates  $(x, y, z)$ . This is so because  $1/r$  automatically differentiates the atoms according to their inverse distances from the tagged atom at the origin. Since the number of inputs has been fixed for each sub-network, we feed 0 to the remaining input entries of the sub-network when the number of atoms falling inside the cut-off radius centered at atom 0 is less than  $N_c$ . This means putting some additional atoms at infinity and hence has negligible influence on the results upon training. In most cases (see below) we find that it is sufficient to use the radial and angular coordinates of a smaller subset of atoms closer to the origin while keeping only the radial coordinates of all the other atoms within the cut-off radius.

The sub-networks are only coupled through summation in the last step of the scheme, when we compute the total energy. From a qualitative point of view, one can think about the sub-networks as providing different local energy contributions to the potential energy surface. To preserve the permutational symmetry of the input, in each sub-network the atoms are first assigned to different groups corresponding to the different atomic species, and then within each one of these groups the atoms are sorted in order of increasing distance to the origin. Global permutational symmetry is preserved by assigning the same parameters to all the sub-networks corresponding to atoms of the same species.

The rest of the formulation of the deep neural network as well as the training procedure is fairly standard. In this work, we use a fully-connected feedforward neural network [22], sometimes combined with Batch Normalization [23], in the architecture of the sub-networks. To train the network we employ stochastic gradient descent with the Adam optimizer [24]. We find that using hidden layers with the number of their nodes in decreasing order gives better results. Such phenomenon is similar to the coarse-graining process in the convolutional neural network [21]. The compactness of the input data has the effect of reducing the number of parameters and thereby the complexity of the training process. For the systems considered here, it generally takes only a few hours on a NERSC Cori CPU node to train a small neural network capable of predicting the energy within chemical accuracy ( $\sim 1\text{kcal/mol}$ )<sup>§</sup>. See Table 2 for the architecture of the sub-networks and details of the training process.

---

<sup>§</sup>This doesn't mean that the Deep Potential model is chemically accurate. Only if the electronic structure data is chemically accurate can the Deep Potential model be chemically accurate.

## 2.2 Molecular system

As a representative molecular system, we consider  $C_7O_2H_{10}$ , which comprises the largest ensemble of stable isomers in the QM9 database [25,26]. Our goal is to predict the energies of all the MD trajectories of 113 randomly selected  $C_7O_2H_{10}$  isomers in this database. The same collection of data was also used to benchmark the DTNN scheme. Following the convention of the DTNN benchmark, we measure the accuracy of the energy predictions using Deep Potential in terms of the mean absolute error (MAE) (Table 1).

## 2.3 Condensed-phase systems

To test the performance of Deep Potential in condensed phase, we use data from EAM and AIMD simulations. In particular, we use 80000 MD snapshots in a 256 Cu atoms trajectory generated with the EAM potential. Consecutive snapshots are separated by 0.1 ps and we consider a trajectory in which the temperature is increased by 100 K every 200 ps under zero pressure. After reaching 2400 K, the temperature is decreased with the same schedule until the system reaches again a temperature of 500 K. We also consider 14000 snapshots along several AIMD trajectories for a solid Zr sample in which the atoms are randomly displaced to model radiation damage. Finally, we use 100000 snapshots of a path-integral (PI) AIMD trajectory modeling liquid water at room temperature (300 K) and standard pressure (1 atm). In this trajectory, almost all the time steps are used for the training snapshots, so that a significant amount of correlation is present in the data. For the condensed-phase systems, we follow the convention of BPNN and measure the accuracy of the energy predictions with Deep Potential in terms of the root mean square error (RMSE) (Table 1).

# 3 Discussion

## 3.1 Generality

By design Deep Potential is a very general framework. We obtain satisfactory results for finite and extended systems, using data generated either by empirical potentials or by DFT simulations. Based on the experience gained from the test cases, we expect that the method should also work well for more complex systems, such as biological molecules. In addition, we could use for training more accurate data than DFT, if available, such as, *e.g.*, data from Coupled-Cluster calculations with Single and Double and Perturbative Triple excitations [27] (CCSD(T)), or Quantum Monte Carlo simulations [28].

## 3.2 Accuracy

In all the test cases, the Deep Potential predictions reproduce the original data within chemical accuracy and compare well with either DTNN or BPNN (Table 1). For  $C_7O_2H_{10}$ ,

Table 1: Comparison of energy prediction on the test set obtained by Deep Potential and DTNN/BPNN.

System	MAE of DP (DTNN*) (eV)
C <sub>7</sub> O <sub>2</sub> H <sub>10</sub>	0.04 (0.07)
System	RMSE of DP (BPNN**) (meV/atom***)
Cu	0.20 (0.20)
Zr	0.09 (0.21)
H <sub>2</sub> O	1.8 (2)

\* The MAE of DTNN is from Ref. [9];

\*\* The results for Cu and Zr come from our implementation of BPNN on the same training and test set, with the symmetry functions for Cu in Ref. [29]; the result for H<sub>2</sub>O is from Ref. [30], which is tested on a different set of configurations.

\*\*\* The unit of RMSE for H<sub>2</sub>O is meV/molecule.

using the same training/testing protocols, we obtain the up-to-date best result (0.04 eV) compared with the existing benchmark (0.07 eV). In the case of EAM Cu, we only use radial information to train the network, obtaining an accuracy comparable to BPNN. In the case of Zr with DFT data, Deep Potential with only radial information gives a better result (0.09 meV/atom) than BPNN (0.21 meV/atom), with the caveat that the symmetry functions we use for Zr are the same of those of Cu, *i.e.*, they are not redesigned for Zr. In the case of liquid water, including the radial and the first-shell angular information gives an error (1.8 meV/H<sub>2</sub>O) that is one fourth of the error obtained with distance information only.

In our scheme, rotational and permutational symmetries are imposed by fixing operations. The importance of the symmetry constraints can be assessed by selectively removing corresponding operations. For instance, removing rotational symmetry fixing in the input data of C<sub>7</sub>O<sub>2</sub>H<sub>10</sub> gives an MAE of 0.05 eV, slightly larger than the optimal value for this molecule (0.04 eV). On the other hand, removing permutational symmetry fixing in the input data of bulk Cu increases substantially the RMSE, from 0.2 meV/atom to 15.6 meV/atom.

### 3.3 Computational cost

For a system with  $N$  atoms and at most  $N_c$  atoms inside  $R_c$ , the computational cost of Deep Potential is linear with  $N$ , similar to EAM, but requires a larger number of size independent operations. In all systems,  $N_c$  is size independent. Using Cu as an example, both EAM and Deep Potential require  $O(N)N_c$  operations to locate the atoms inside  $R_c$ . Within EAM,  $\sim 10NN_c$  multiplications are needed to fit the potential. Within Deep Potential,  $\sim 15NN_c \log N_c$  operations are needed for the sorting, and  $\sim 100NN_c$  multiplications

are needed for the calculation of the energy by the neural network. Both EAM and Deep Potential are highly parallelizable. Thus, large-scale MD simulations with Deep Potential are computationally feasible.

Within our framework we have the freedom to tune the size of the sub-networks in order to achieve a good balance of computational cost and accuracy of the predictions. In the examples that we report, we use the sub-network sizes of Table 2, but we could use significantly smaller sub-network sizes with a minor reduction of accuracy. For instance, using a sub-network with 80-40-20-10 nodes in the hidden layers for  $C_7O_2H_{10}$ , we obtain a still acceptable MAE of 0.07 eV, to be compared with the MAE of 0.04 eV with the considerably larger 600-400-200-100-80-40-20 sub-network. The smaller sub-network reduces the computational cost by almost two orders of magnitude.

### 3.3.1 Scalability

The scalability (and extensivity) of the energy of Deep Potential is intimately related to the fact that the “local energies”  $E_i$  depend only on a finite environment of site  $i$ . This is consistent with physical intuition because most of the interactions can be captured with a large enough  $R_c$ . Typically including the first two shells of neighbors would allow us to describe covalent bonding interactions, such as bond stretching and bending, and dihedral angle forces. Van der Waals effects have longer range but we find that including up to the third shell of neighbors is usually sufficient. A small number of neighboring shells is also sufficient in metals due to screening effects. Longer range effects are present in ionic and/or dipolar systems. The long-range part of the Coulombic interactions in such systems is typically treated exactly with techniques such as the Ewald summation [31]. We have not included explicitly these effects in the current implementation of Deep Potential, although in the cases in which these effects are important they are included in the training data. A possible way of explicitly including these effects in neural network potentials was presented in Ref. [32]. We leave the issue of how to include long-range effects into Deep Potential to future studies. As a limited test of scalability, we use the Deep Potential model trained from the sample of 256 Cu atoms, mentioned earlier, to predict the potential energy of systems with 864, 2048, and 4000 atoms in periodic simulation cells. Comparison with EAM calculations for these systems gives an RMSE per atom of 0.10, 0.07, and 0.06 meV/atom, respectively.

## 4 Summary

Deep Potential is a simple, yet general, end-to-end deep neural network representation of the potential energy function for atomic and molecular systems. It should enable us to evaluate the many-atom potential energy surface with the accuracy of quantum mechanics models at a computational cost comparable to that of empirical models.

## 5 Methods

### 5.1 Reference data sets

The AIMD data set and the training/testing protocols for  $C_7O_2H_{10}$  are available at <http://quantum-machine.org/datasets/>. The simulation of Cu uses the MD package LAMMPS 2016 [33]. We adopt the EAM potential for Cu [34] with an MD integration time step of 1 fs. The training system contains 256 atoms in a cubic cell with periodic boundary conditions. We adopt a Nosé-Hoover thermostat and barostat following Ref. [35]. 80000 configurations at intervals of 0.1 ps along the EAM trajectory for Cu are stored. Of these, 72000 are used for training and 8000 to test the predictions. The AIMD data for Zr are generated with VASP [36] and consist of 14000 snapshots from 30 short trajectories that are used to determine the threshold displacement energies (TDE) along symmetric directions. These simulations use large supercells with 8000 atoms. Each trajectory runs for 1 ps with a variable time step ensuring that the atom with the largest velocity moves by less than 0.1 Å in each time step. 90% of the corresponding snapshots are used for training and 10% to test the energy predictions. The PI-AIMD trajectory for liquid water is obtained with Quantum Espresso [37] interfaced with i-PI [38]. The system contains 64 water molecules in a cubic box under periodic boundary conditions. The exchange-correlation functional PBE0 [39] is adopted, and the long-range dispersion interactions are approximated self-consistently with the Tkatchenko-Scheffler model [40]. The generalized Langevin equation [41] in i-PI requires 8 beads for a converged representation of the Feynman paths. The trajectory is approximately 10 ps long with a time step of 0.48 fs. 100000 snapshots are randomly selected along the PI-AIMD trajectory. Of these, 90000 are used for training and 10000 to test the energy predictions.

### 5.2 Details on the Deep Potential method

After generating and standardizing the input data from the atomic configurations, the implementation of the method follows standard deep learning procedures [22]. We use a fully connected neural network with the rectifier (ReLU) as the activation function. We use the Adam optimizer [24] to train the neural network with a batch size of 128. See Table 2 for the detailed architecture of the sub-networks, the total training epochs, the learning rate scheme, and the decay rate of the moving-average parameter for the Batch Normalization procedure. It is widely observed that Batch Normalization can effectively reduce the training time and improve the predictions of a deep and large neural network, like in the case of the  $C_7O_2H_{10}$  isomers, but for neural networks that are not deep and large, Batch Normalization is less effective. Thus, we use it for  $C_7O_2H_{10}$  and  $H_2O$ , but, after testing it, we decide not to use it for Cu or Zr. The adopted cutoff radii are 6.0 Å, 7.0 Å, and 5.8 Å for Cu, Zr, and  $H_2O$ , respectively.



Table 2: Details of training process.

System	Architecture*	Training epochs	LR scheme**	Batch Normalization***
C <sub>7</sub> O <sub>2</sub> H <sub>10</sub>	600-400-200-100-80-40-20	300	(0.01, 0.96, 1.5)	used
Cu	80-40-20-10	600	(0.005, 0.96, 3.6)	not used
Zr	80-20-5-5	1200	(0.005, 0.96, 7.2)	not used
H <sub>2</sub> O	160-40-10-10	300	(0.01, 0.96, 1.6)	used

\* The architecture is given by the number of nodes in each hidden layer, from input to output.

\*\* The parameters ( $a$ ,  $b$ ,  $c$ ) in the learning rate (LR) scheme are the starting learning rate, the decay rate, and the decay epoch, respectively. If the current training epoch is  $x$ , then the learning rate will be  $a * b^{x/c}$ .

\*\*\* If Batch Normalization is used, the parameter for moving average will be  $a * b^{0.6x/c}$ .

## Acknowledgments

We thank H. Wang for providing the trajectory data of Zr and helpful discussions, and H.-Y. Ko and B. Santra for sharing the AIMD data on water and ice. The work of J. Han and W. E is supported in part by Major Program of NNSFC under grant 91130005, ONR grant N00014-13-1-0338, DOE grants DE-SC0008626 and DE-SC0009248. The work of R. Car is supported in part by DOE-SciDAC grant DE-SC0008626.

## References

- [1] J. E. Jones, "On the determination of molecular fields," in *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 106, pp. 463–477, The Royal Society, 1924.
- [2] F. H. Stillinger and T. A. Weber, "Computer simulation of local order in condensed phases of silicon," *Physical Review B*, vol. 31, no. 8, p. 5262, 1985.
- [3] M. S. Daw and M. I. Baskes, "Embedded-atom method: Derivation and application to impurities, surfaces, and other defects in metals," *Physical Review B*, vol. 29, no. 12, p. 6443, 1984.
- [4] A. D. MacKerell Jr, D. Bashford, M. Bellott, R. L. Dunbrack Jr, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, *et al.*, "All-atom empirical potential for molecular modeling and dynamics studies of proteins," *The Journal of Physical Chemistry B*, vol. 102, no. 18, pp. 3586–3616, 1998.
- [5] J. Wang, P. Cieplak, and P. A. Kollman, "How well does a restrained electrostatic potential (resp) model perform in calculating conformational energies of organic and biological molecules?," *Journal of Computational Chemistry*, vol. 21, no. 12, pp. 1049–1074, 2000.
- [6] A. C. Van Duin, S. Dasgupta, F. Lorant, and W. A. Goddard, "Reaxff: a reactive force field for hydrocarbons," *The Journal of Physical Chemistry A*, vol. 105, no. 41, pp. 9396–9409, 2001.
- [7] W. Kohn and L. J. Sham, "Self-consistent equations including exchange and correlation effects," *Physical Review*, vol. 140, no. 4A, p. A1133, 1965.

- [8] R. Car and M. Parrinello, "Unified approach for molecular dynamics and density-functional theory," *Physical Review Letters*, vol. 55, no. 22, p. 2471, 1985.
- [9] K. T. Schütt, F. Arbabzadah, S. Chmiela, K. R. Müller, and A. Tkatchenko, "Quantum-chemical insights from deep tensor neural networks," *Nature Communications*, vol. 8, p. 13890, 2017.
- [10] J. Behler and M. Parrinello, "Generalized neural-network representation of high-dimensional potential-energy surfaces," *Physical Review Letters*, vol. 98, no. 14, p. 146401, 2007.
- [11] F. A. Faber, L. Hutchison, B. Huang, J. Gilmer, S. S. Schoenholz, G. E. Dahl, O. Vinyals, S. Kearnes, P. F. Riley, and O. A. von Lilienfeld, "Fast machine learning models of electronic and energetic properties consistently reach approximation errors better than dft accuracy," *arXiv preprint arXiv:1702.05532*, 2017.
- [12] S. Chmiela, A. Tkatchenko, H. E. Sauceda, I. Poltavsky, K. T. Schütt, and K.-R. Müller, "Machine learning of accurate energy-conserving molecular force fields," *Science Advances*, vol. 3, no. 5, p. e1603015, 2017.
- [13] C. M. Handley, J. Behler, *et al.*, "Next generation of interatomic potentials for condensed systems," *European Physical Journal B*, vol. 87, p. 152, 2014.
- [14] J. Behler, "Constructing high-dimensional neural network potentials: A tutorial review," *International Journal of Quantum Chemistry*, vol. 115, no. 16, pp. 1032–1050, 2015.
- [15] M. Rupp, A. Tkatchenko, K.-R. Müller, and O. A. Von Lilienfeld, "Fast and accurate modeling of molecular atomization energies with machine learning," *Physical Review Letters*, vol. 108, no. 5, p. 058301, 2012.
- [16] R. Ramakrishnan and O. A. von Lilienfeld, "Many molecular properties from one kernel in chemical space," *CHIMIA International Journal for Chemistry*, vol. 69, no. 4, pp. 182–186, 2015.
- [17] B. Huang and O. A. von Lilienfeld, "Communication: Understanding molecular representations in machine learning: The role of uniqueness and target similarity," *The Journal of Chemical Physics*, vol. 145, no. 16, p. 161102, 2016.
- [18] K. Hansen, F. Biegler, R. Ramakrishnan, W. Pronobis, O. A. Von Lilienfeld, K.-R. Müller, and A. Tkatchenko, "Machine learning predictions of molecular properties: Accurate many-body potentials and nonlocality in chemical space," *The Journal of Physical Chemistry Letters*, vol. 6, no. 12, p. 2326, 2015.
- [19] A. P. Bartók, R. Kondor, and G. Csányi, "On representing chemical environments," *Physical Review B*, vol. 87, no. 18, p. 184115, 2013.
- [20] J. S. Smith, O. Isayev, and A. E. Roitberg, "Ani-1: an extensible neural network potential with dft accuracy at force field computational cost," *Chemical Science*, vol. 8, no. 4, pp. 3192–3203, 2017.
- [21] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, pp. 1097–1105, 2012.
- [22] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT Press, 2016.
- [23] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *Proceedings of The 32nd International Conference on Machine Learning (ICML)*, 2015.
- [24] D. Kingma and J. Ba, "Adam: a method for stochastic optimization," in *Proceedings of the International Conference on Learning Representations (ICLR)*, 2015.
- [25] L. Ruddigkeit, R. Van Deursen, L. C. Blum, and J.-L. Reymond, "Enumeration of 166 billion organic small molecules in the chemical universe database gdb-17," *Journal of Chemical Information and Modeling*, vol. 52, no. 11, pp. 2864–2875, 2012.

- [26] R. Ramakrishnan, P. O. Dral, M. Rupp, and O. A. Von Lilienfeld, "Quantum chemistry structures and properties of 134 kilo molecules," *Scientific Data*, vol. 1, 2014.
- [27] J. D. Watts, J. Gauss, and R. J. Bartlett, "Coupled-cluster methods with noniterative triple excitations for restricted openshell hartreefock and other general single determinant reference functions. energies and analytical gradients," *The Journal of Chemical Physics*, vol. 98, no. 11, pp. 8718–8733, 1993.
- [28] D. Ceperley and B. Alder, "Quantum monte carlo," *Science*, vol. 231, 1986.
- [29] N. Artrith and J. Behler, "High-dimensional neural network potentials for metal surfaces: A prototype study for copper," *Physical Review B*, vol. 85, no. 4, p. 045439, 2012.
- [30] T. Morawietz, A. Singraber, C. Dellago, and J. Behler, "How van der waals interactions determine the unique properties of water," *Proceedings of the National Academy of Sciences*, p. 201602375, 2016.
- [31] P. P. Ewald, "Die berechnung optischer und elektrostatischer gitterpotentiale," *Annalen Der Physik*, vol. 369, no. 3, pp. 253–287, 1921.
- [32] N. Artrith, T. Morawietz, and J. Behler, "High-dimensional neural-network potentials for multicomponent systems: Applications to zinc oxide," *Physical Review B*, vol. 83, no. 15, p. 153101, 2011.
- [33] S. Plimpton, *Fast parallel algorithms for short-range molecular dynamics*. Academic Press Professional, Inc., 1995.
- [34] S. Foiles, M. Baskes, and M. Daw, "Embedded-atom-method functions for the fcc metals cu, ag, au, ni, pd, pt, and their alloys," *Physical Review B*, vol. 33, no. 12, p. 7983, 1986.
- [35] M. E. Tuckerman, J. Alejandre, R. López-Rendón, A. L. Jochim, and G. J. Martyna, "A liouville-operator derived measure-preserving integrator for molecular dynamics simulations in the isothermal-isobaric ensemble," *Journal of Physics A: Mathematical and General*, vol. 39, no. 19, p. 5629, 2006.
- [36] G. Kresse and J. Furthmüller, "Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set," *Physical Review B*, vol. 54, pp. 11169–11186, 1996.
- [37] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. Dal Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari, and R. M. Wentzcovitch, "Quantum espresso: a modular and open-source software project for quantum simulations of materials," *Journal of Physics: Condensed Matter*, vol. 21, no. 39, p. 395502 (19pp), 2009.
- [38] M. Ceriotti, J. More, and D. E. Manolopoulos, "i-pi: A python interface for ab initio path integral molecular dynamics simulations," *Computer Physics Communications*, vol. 185, pp. 1019–1026, 2014.
- [39] C. Adamo and V. Barone, "Toward reliable density functional methods without adjustable parameters: The pbe0 model," *The Journal of Chemical Physics*, vol. 110, no. 13, pp. 6158–6170, 1999.
- [40] A. Tkatchenko and M. Scheffler, "Accurate molecular van der waals interactions from ground-state electron density and free-atom reference data," *Physical Review Letters*, vol. 102, p. 073005, 2009.
- [41] M. Ceriotti, D. E. Manolopoulos, and M. Parrinello, "Accelerating the convergence of path integral dynamics with a generalized langevin equation," *The Journal of Chemical Physics*, vol. 134, no. 8, p. 084104, 2011.