PCCP

Prediction of Optoelectronic Properties of Cu_2O Using Neural Network Potentials

Journal:	*Physical Chemistry Chemical Physics*
Manuscript ID	CP-ART-02-2020-001112.R1
Article Type:	Paper
Date Submitted by the Author:	20-May-2020
Complete List of Authors:	Selvaratnam, Balaranjan; University of South Dakota Koodali, Ranjit; University of South Dakota Miro Ramirez, Pere; University of South Dakota

Journal Name

ARTICLE TYPE

Cite this: DOI: 00.0000/xxxxxxxxxx

Prediction of Optoelectronic Properties of Cu_2O Using Neural Network Potentials†

Balaranjan Selvaratnam, Ranjit T. Koodali, and Pere Miró*

Received Date
Accepted Date

DOI: 00.0000/xxxxxxxxxx

Neural Network Potentials (NNPs) trained against density functional theory (DFT) are capable of reproducing the potential energy surface at a fraction of the computational cost. However, most NNP implementations focus on energy and forces. In this work, we modified NNP model introduced by Behler and Parrinello to predict Fermi energy, band edges, and partial density of states of Cu_2O. Our NNPs can reproduce the DFT potential energy surface and properties at a fraction of the computational cost. We used our NNP to perform molecular dynamics (MD) simulations and validated them against DFT calculations. Our model achieved a root mean squared error of 16 meV for the energy prediction. Furthermore, we addressed the uncertainty in the predicted potential energies during MD simulations by using the standard deviation of the snapshot ensemble energies to estimate the uncertainty. This allows us to switch from the NNP to DFT on-the-fly during the MD simulation to evaluate the forces when the uncertainty is high.

1 Introduction

The predictive power of Machine Learning (ML) has been demonstrated to be useful in several avenues of chemical research and the application of this data-driven approach is nowadays considered as the fourth paradigm in materials science.[1,2] In general, machine learning algorithms improve their performance at a given task by learning from experience. In supervised machine learning, this improved performance is achieved by learning from data to minimize the error between the predicted and expected outcome for the given inputs. This process is known as training and once completed, the model acts as a surrogate to the original theory/experiment and can predict outcomes for new inputs.[3] Machine learning has been used in a broad range of applications in chemistry such as the accelerated prediction of various experimental and *ab initio* properties[4,5], inverse design of molecules and materials with desired physico-chemical properties[6], optimizing synthesis parameters toward desired properties[7,8], synthesis planning[9] and, catalyst design[10–12], machine learning potentials[13,14], etc.

Neural Network Potentials (NNP) are machine leaning potentials that model the Potential Energy Surface (PES) of a system using features derived from positions and identities of all atoms in the system as input. Although first principles methods such as Density Functional Theory (DFT) can provide accurate PES for many systems, using DFT for large systems or long simulation times are computationally expensive. Alternatively, PES can be described using empirical potentials which are relatively inexpensive in terms of computational cost, but the development of such potentials is usually a lengthy and difficult process. In addition, their accuracy is rather limited compared to first principles methods. In 2007, Behler and Parrinello presented a NN model to describe the PES of silicon using Atom Centered Symmetry Functions (ACSF) as descriptors.[15] This NN-PES approach is orders of magnitude faster than DFT with comparable accuracy.[16] An important aspect of NN potentials compared to classical potentials is that NNP does not assume a specific functional form and fits the PES from the training data rather than fitting the parameters of an empirical function. NN-PES have been used in several applications such as modeling Phase Changing Materials[17], studying the atomic structure of nanoparticles[18,19], Molecular Dynamics (MD) simulations metal nanoparticles on support[20], MD simulations of amorphous materials[21,22], atom diffusion[23], surface phonons[24], etc. However, most NNPs do not include information beyond energy and forces such as band gap, band structure or density of states (DOS). Availability of these properties can expand the applications of neural network potentials to study the dynamic evolution of these properties during MD simulations. There are two recent works focused on the prediction of DOS using ML. Yeo *et al.* used a ML model to predict the principal components (PCs) of the DOS vector using d-orbital occupation ratio, coordination number, mixing factor, and the inverse of Miller indices as input features.[25] Later, the DOS probability matrix for the DOS vector was calculated and then converted to DOS pattern for the given system. They tested this approach to predict DOS for multicomponent alloy system and obtained an accuracy greater than

Department of Chemistry, University of South Dakota, Vermillion, SD 57069, USA. E-mail: pere.miro@usd.edu

† Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 00.0000/00000000.

90% compared to DFT. Chandrasekaran *et al.* used NN to predict the charge density and local DOS around a given grid point using a rotation invariant atomic environment descriptor as input representation.[3] The DOS predicted using this approach for polyethylene and aluminum systems were in excellent agreement with the DFT values. Although these approaches predict DOS, neither energy nor forces were included in their models. To authors' best knowledge, there are no works using Behler-Parrinello's ASCFs to predict optoelectronic properties such as band edges or DOS. In this work, we show that a multitask neural network model can be used to predict DOS, Fermi energy (EF), Valence Band Maximum (VBM), and Conduction Band Minimum (CBM), and that can be used to predict these properties during molecular dynamic simulations. Furthermore, we also show that the standard deviation of the energies predicted by ensemble of training snapshots can be used to identify when the NNP ventures into PES regions beyond the ones included in the training. By using forces from DFT at these points in the trajectory, the molecular dynamics can be continued.

2 Methods

In the NN-PES approach reported by Behler and Parrinello, Atom Centered Symmetry Functions (ASCF) are used as descriptors, from which a NN learns to predict atomic energies. These atomic energies are added to give the total energies and the forces are calculated from the analytic gradients. This NN-PES model is available in several codes such as RuNNer[26], Atomistic Machine learning Package (AMP)[27], and ænet[28]. In this work, we extend the NN-PES approach to predict optoelectronic properties by modifying the NN model implemented in the AMP code. The optoelectronic properties are predicted by modifying the output layer of the NN model from 1 to 104 where energy, Fermi energy, VBM, and CBM were predicted by the first four nodes and the remaining hundred nodes were used to predict the partial density of states (PDOS). Since the DOS can be calculated as summation of PDOS, in this work, we decided to focus on PDOS. Since the region near the Fermi level will be relevant for the optoelectronic properties, the PDOS near the Fermi level was modeled with 100 points however, the PDOS region of interest can be changed if needed. This gives the flexibility of focusing on different region of PDOS when needed at the expense of retraining of the NNP.

Prior to training, the PDOS data was preprocessed. Initially, the Fermi energy was subtracted from PDOS in order to focus PDOS region near band edges. After, the PDOS was interpolated using SciPy's 1-D interpolate function with third order and resampled at 0.1 eV intervals from -4.9 to +5.0 eV to give 100 data points.[29] However, we note that one can avoid the interpolation scheme by sampling the PDOS at fine intervals when acquiring the training data. The raw PDOS data are presented in the supporting information (Figure S1). The DFT and NNP PDOS data were smoothed by Savitzky-Golay (as implemented in SciPy) filter using third order polynomial with a window length of 5. The energy and PDOS are calculated as sum of atomic contributions, however, EF, VBM and CBM are calculated by summation of the atomic contributions followed by division by the number of atoms in the system. The atom centered symmetry functions used in this work are given equations 1, and 2, and the Gaussian parameters are given in supporting information.

$$f_i^I = \sum_{j\neq i}^{N} e^{-\eta R_{ij}^2/R_c^2} f_c(R_{ij}) \tag{1}$$

Where N is the number of atoms within cutoff radius, i, j atom indices, R_{ij} is the distance between atom i and j, R_c is the cutoff radius, η is Gaussian parameter, and f_c is the cutoff function.

$$f_i^{II} = 2^{1-\zeta} \sum_{j,k\neq i, j\neq k}^{N} (1+\lambda\cos\theta_{ijk})^{\zeta} e^{-\eta(R_{ij}^2+R_{ik}^2+R_{jk}^2)/R_c^2} \times f_c(R_{ij}) f_c(R_{ik}) f_c(R_{jk}) \tag{2}$$

Where ζ and λ are Gaussian parameters, and θ_{ijk} is the angle between atoms indicated by the indices. The f_c is the cutoff function, given below.

$$f_c(r) = \begin{cases} 0.5(1+\cos(\pi r/R_c)) & \text{if } r \leq R_c, \\ 0 & \text{if } r > R_c. \end{cases} \tag{3}$$

The loss function also modified to optimize the parameters for the additional properties being predicted. The modified loss function is given in equation 4. This modified model was implemented in Python (version 3.6) using TensorFlow[30], (version 1.13) framework in AMP code.

$$L = C_E L_E + C_F L_F + C_{EF} L_{EF} + C_{VBM} L_{VBM} + C_{CBM} L_{CBM} + C_{PDOS} L_{PDOS} \tag{4}$$

$$L_i = \frac{1}{n} \sum_{i=0}^{n} (y_i - \hat{y}_i)^2 \tag{5}$$

Where C_i and L_i are coefficient and loss for property i. The subscripts E, F, EF, VBM, CBM, $PDOS$ stands for energy, forces, Fermi energy, valence band maximum, conduction band minimum, and partial density of states, respectively. y_i and $\hat{y}_i$ stands for true and predicted properties, respectively.

2.1 Training Data

We tested this modified model to predict electronic properties of cuprous oxide, Cu_2O. Since we are interested in evaluating the feasibility of using NNP to predict optoelectronic properties, we performed all calculations and MD simulations on the unit cell of Cu_2O containing four copper and two oxygen atoms. The training data was generated by following the dataset aggregation method. Initially, a MD simulation was performed at 1550 K and all snapshots from this trajectory were added into the training set. The starting temperature was selected to be above the melting point of Cu_2O in order obtain configurations that are far from the equilibrium structure. Then a coarse NNP was trained using unmodified AMP and then it was validated by performing MD at random

temperatures between 50 K and 1550 K (in 25 K intervals) using random snapshots from the training set as starting points. See supporting information for details. Then the energies predicted by the coarse potential were compared with DFT values and if the absolute difference between the NNP and DFT energy was greater than 5 kcal, then the structure was added to training set. Finally, the NNP was retrained and the cycle was continued until no structures had energy difference greater than 5 kcal when compared to DFT for three consecutive MD validations. This cycle was implemented in Python using Atomic Simulation Environment (ASE) [31]. This data generation cycle was performed with a lower k-point mesh in order to reduce the computational cost. After the training data was generated, single point calculations were performed with the optimized k-points. See computational methods for details.

2.2 Hyperparameter Optimization

The hyperparameters, number of layers, number of neurons in each layer, and activation function were optimized using the training set and a separate DFT-MD trajectory at 300 K as validation set. Initially, hyperbolic tangent was used as activation function to optimize the layers and number of neurons. For each hyperparameter configuration, three separate models were trained and performance metrics on the evaluation set were averaged. The hyperparameters with lowest root mean squared error (RMSE) for energy and forces were selected for further studies. All models used in this work were trained using L-BFGS-B optimizer implemented in the SciPy library.

2.3 Computational Methods

All reference calculations were performed using the periodic density functional theory implementation in the Vienna Ab-Initio Simulation Package (VASP 5.1). The exchange-correlation functional is approximated by the generalized gradient approximation according to Perdew, Burke and Ernzerhof (PBE) functional with a Hubbard U parameter of 6 eV [32]. The cut-off energy for the plane wave basis set was set to 700 eV. Gaussian smearing with sigma of 0.02 was used for all calculations. For the starting geometry, all atoms in the unit cell were allowed to relax until the forces were less than 0.02 eV/Å. For populating the training database using the dataset aggregation method, we used (1, 1, 1) k-points and for single point calculations (4, 4, 4) k-points were used.

3 Results and Discussion

3.1 NNP Training and Hyperparameter Optimization

The dataset aggregation method generated a training set with 2712 entries. A total of 22 iterations were performed by the dataset aggregation method. From this, structures with forces greater than 5 eV/Å were removed, which resulted in a training set with 2492 entries. Then the NNP was trained with different loss coefficients to get the optimum values for the coefficients in equation 4. However, in all different coefficients we tried, the agreement between the true and predicted PDOS was not satisfactory. Therefore, to obtain a better fit, we used a step-wise training scheme. Initially, we optimized the entire network with the coefficients $C_E = 0.91$, $C_F = 0.05$, $C_{EF} = 0.01$, $C_{VBM} = 0.01$, $C_{CBM} = 0.01$, and $C_{PDOS} = 0.01$. Then the output layer parameters corresponding to EF, VBM, and CBM were fine tuned by retraining the model with coefficient of 1.0 for EF, VBM, and CBM, and 0.0 for others ($C_{EF} = 1.0$, $C_{VBM} = 1.0$, $C_{CBM} = 1.0$, and $C_i = 0.0$ for others). Finally, the parameters for the PDOS part of the output layer was retrained with $C_{PODS} = 1.0$ and all other $C_i = 0.0$. During the fine tuning of the output layer parameters corresponding to EF, VBM, CBM, and PDOS, the parameters of the other parts of the network were kept frozen. Then the hyperparameters (number of layers, number of neurons, and activation function) were optimized using the step-wise training scheme and the results are given in Table 1.

The optimization data in Table 1 shows that the Root Mean Squared Error (RMSE) on the evaluation set decreases sharply as the number of nodes in the hidden layers of NN increases. However, when increasing the number of hidden layers, the improvement at higher node counts is minimal. For entry 8, we tested use of Rectified Linear Unit (ReLU) activation function and the performance with ReLU is poor compared to hyperbolic tangent activation. Hence, we selected entry 8 (two hidden layers each with 512 nodes, hyperbolic tangent activation) as optimum hyperparameters. The RMSE error on energy and forces at this configuration is similar to metrics previously reported and the error on energy is below chemical accuracy (0.016 eV) [20,23]. In this configuration, the network has 326,676 parameters per element. The comparison between the DFT and NNP potential energies, band edges and Fermi energy are given in Figure 1 and the forces are compared in Figure 2. The energies and forces predicted by the NNP shows a good agreement with DFT. The predicted PDOS also aligns well with the DFT data and the histogram of Root Mean Squared Deviation (RMSD) for the PDOS predictions and sample PDOS of Cu and O are given in the supporting information.

3.2 NNP Molecular Dynamics

To further validate the NNP, we performed constant energy (NVE) molecular dynamics simulations using the trained potential. These MD were performed using the Velocity Verlet algorithm implemented in ASE. Initially we performed a molecular dynamics at 300 K for 500 steps using 1 fs timestep. Then the potential energy and optoelectronic properties predicted by NNP were compared with DFT values by performing single point calculations for each snapshot along the MD trajectory. The results are given in Figure 3 and values predicted by the NNP are in good agreement with the DFT values. The error between the DFT and NNP potential energies are on the order of 1 kcal for most of the structures and the predicted band edges are also close to the DFT ones. However, the Fermi energy predicted by the NNP differs from the DFT as high as 147 meV. The Fermi energy calculated by DFT during this dynamics drops significantly between steps 70-80 and this is due to the use of Gaussian smearing in the DFT calculations. The RMSD between the DFT and NNP PDOS of Cu varies from 0.1 to 2.5, for PDOS of O, the RMSD varies between 0.15 and 0.40. From this we randomly selected two snapshots each for Cu and O as representative to assess the quality of PDOS predictions.

For Cu, two snapshots with RMSD-Cu between $1.2 < x < 1.3$ and $2.2 < x < 2.3$ were selected and the DFT and NNP PDOS are compared in (Figures 4b and 4c, respectively). Similarly, we randomly selected two snapshots where the RMSD-O is $0.29 < x < 0.31$ for the first snapshot and is $0.35 < x < 0.36$ for the second snapshot (Figures 4e and 4f, respectively). The density predicted by the NNP shows a general agreement with density calculated by the DFT.

3.3 Uncertainty in NNP Molecular Dynamics

In order to further test the NNP, we performed additional MD simulations under same conditions and found that during some simulations, the system would eventually reach unphysical structures due to high forces. This is a clear indication that the system reached points in the PES that are outside of the training points of our potential. To address this issue, one needs to identify the point where the predictions by the NNP are out of its training data. Once identified, by using the forces from DFT at those points on-the-fly, the MD can be continued.[33] To identify these uncertain regions, several approaches have been proposed in the literature. For instance, Janet *et al.* showed that the distance between training data and test data at latent space can be used as a measure of uncertainty.[34] Peterson *et al.* used an ensemble of NNPs and showed that the width of the ensemble can be used to quantify the uncertainty.[35] This method requires the training of several NNPs. Instead of using an ensemble of several NN models, Cortés-Ciriano and Bender used weights (snapshots) taken during the training of one NN model as ensemble.[36] In this work, we use the standard deviation of the snapshot-ensemble energies as measure of uncertainty. By running a few MD simulations using the trained NNP, we found that when the standard deviation of the ensemble of the predicted energies is greater than 5 meV, the dynamics eventually leads to unpysical structures. Hence, when the ensemble standard deviation is above 5 meV, we calculated the forces using DFT single point calculation during the MD and updated the forces to continue the molecular dynamics. The results of the dynamics are given in Figure 5 and the steps where DFT calculations were performed during MD are shaded in Figure 5b. The DFT and NNP potential energies, Fermi energy, and band edges of this simulation are consistent with the DFT values. See supporting information for details (Figure S3 and S4). The force norm plot shows that DFT is used in four different spans during the dynamics and that DFT is continuously used in each span. During this MD, DFT was used for a total of 58 steps.

For further validation of this approach we performed NVT MD simulations using this scheme at 300 K, 600 K, and 900 K with a timestep of 0.5 fs for 500 steps. These simulations were performed using NVTAndersen dynamics implemented in the TSASE code by Henkelman and coworkers[37]. Collision strength of 0.8 was used for the dynamics. Then DFT single point calculations were performed for the trajectory. The comparison of DFT and NNP potential energies are presented in Figure S5 for all three temperatures. During these runs, the number of on-the-fly DFT calculations increased with increasing temperature with 0, 59, and 278 DFT force calculations for 300 K, 600 K, and 900 K respectively. The increasing number of DFT force calculations with temperature shows the need for an extensive sampling to generate the training dataset. Furthermore, the changes in the ensemble standard deviation also plotted in the respective potential energy comparison plots which shows that the ensemble standard deviation method can be used as uncertainty measure for most cases. However, for some configurations, this method yields poor results where we saw high difference between DFT and NNP energy eventhough the ensemble standard deviation is lower than 5.0 threshold. Upon closer inspection, we noticed that the ensemble sandard deviation is above 3.5 meV and below 5.0 meV for most of these configurations. Hence, we decided to set the ensemble standard deviation threshold to 3.5 meV for future MD simulations.

Finally, we performed NVT MD using Andersen thermostat for 1 ps with 0.5 fs steps at 300 K. Based on the uncertainty changes and energy differences in the high temperature simulations, we set the threshold for ensemble standard deviation to 3.5 meV. The change in the potential energy along with the ensemble standard deviation are given in Figure 6. A total of 951 DFT force calculations were made during this 1 ps simulation consisting 2000 steps. This scheme can be used for extended time simulations and when the number of calculations by reference calculator (DFT) is above a certain threshold number, one can retrain the NNP with the data points calculated using the reference calculator.

The extensive MD validations shows the need for more through sampling of the PES to build the training dataset. Since the main objective of this work is to evaluate the feasibility of using NNP to predict the optoelectronic properties using a single NNP, we will test the effectiveness of a well sampled training dataset against an automatic retraining scheme when the number of DFT force calculation exceeds a certain number in a future work.

Conclusion

In summary, we implemented, and trained a multitask neural network model to predict optoelectronic properties using atom centered symmetry functions as descriptor. Using this neural network potential model, we predict energy, forces, Fermi energy, valence band maximum, conduction band minimum, and partial density of states for molecular dynamics trajectories for a representative metal oxide system Cu_2O. The potential energies, forces, band edges, and partial density of states predicted by the model are in good agreement with the DFT calculated data. In addition, we also show that the standard deviation of energies predicted by snapshot ensemble can be used as measure of uncertainty.

Conflicts of interest

There are no conflicts to declare.

Acknowledgments

Computations supporting this project were performed on High Performance Computing systems at the University of South Dakota funded by NSF (ACI-1626516). We are thankful to the Department of Chemistry of the University of South Dakota (USD) for Graduate Assistantship and to the USD Neuroscience, Nanotechnology, and Networks program (USD-N3) funded by the Na-

tional Science Foundation (DOE-1633213). P.M. is thankful to the Department of Chemistry of the University of South Dakota (USD) for the start-up funds.

Notes and references

1 L. Himanen, A. Geurts, A. S. Foster and P. Rinke, *Advanced Science*, 2019, **6**, 1900808.

2 A. Agrawal and A. Choudhary, *APL Materials*, 2016, **4**, 053208.

3 A. Chandrasekaran, D. Kamal, R. Batra, C. Kim, L. Chen and R. Ramprasad, *npj Computational Materials*, 2019, **5**, 1–7.

4 A. Jain and T. Bligaard, *Physical Review B*, 2018, **98**, 214112.

5 G. Pilania, A. Mannodi-Kanakkithodi, B. P. Uberuaga, R. Ramprasad, J. E. Gubernatis and T. Lookman, *Scientific Reports*, 2016, **6**, srep19375.

6 B. Sánchez-Lengeling and A. Aspuru-Guzik, *Science*, 2018, **361**, 360–365.

7 J. M. Granda, L. Donina, V. Dragone, D.-L. Long and L. Cronin, *Nature*, 2018, **559**, 377–381.

8 E. Kim, K. Huang, S. Jegelka and E. Olivetti, *npj Computational Materials*, 2017, **3**, 53.

9 M. H. S. Segler, M. Preuss and M. P. Waller, *Nature*, 2018, **555**, 604–610.

10 Y. Huang, Y. Chen, T. Cheng, L.-W. Wang and W. A. Goddard III, *ACS Energy Letters*, 2018, 2983–2988.

11 K. Tran and Z. W. Ulissi, *Nature Catalysis*, 2018, **1**, 696–703.

12 E. L. Kolsbjerg, A. A. Peterson and B. Hammer, *Physical Review B*, 2018, **97**, 195424.

13 V. L. Deringer, M. A. Caro and G. Csányi, *Advanced Materials*, 2019, **31**, 1902765.

14 J. Behler, *The Journal of Chemical Physics*, 2016, **145**, 170901.

15 J. Behler and M. Parrinello, *Physical Review Letters*, 2007, **98**, 583.

16 N. Artrith, T. Morawietz and J. Behler, *Physical Review B*, 2011, **83**, 153101.

17 G. Sosso and M. Bernasconi, *MRS Bulletin*, 2019, **44**, 705–709.

18 R. Jinnouchi, H. Hirata and R. Asahi, *The Journal of Physical Chemistry C*, 2017, **121**, 26397–26405.

19 J. S. Elias, N. Artrith, M. Bugnet, L. Giordano, G. A. Botton, A. M. Kolpak and Y. Shao-Horn, *ACS Catalysis*, 2016, **6**, 1675–1679.

20 N. Artrith, B. Hiller and J. Behler, *physica status solidi (b)*, 2012, **250**, 1191–1203.

21 G. C. Sosso, V. L. Deringer, S. R. Elliott and G. Csányi, *Molecular Simulation*, 2018, **44**, 866–880.

22 V. L. Deringer, N. Bernstein, A. P. Bartók, M. J. Cliffe, R. N. Kerber, L. E. Marbella, C. P. Grey, S. R. Elliott and G. Csányi, *The Journal of Physical Chemistry Letters*, 2018, **9**, 2879–2885.

23 W. Li, Y. Ando, E. Minamitani and S. Watanabe, *The Journal of Chemical Physics*, 2017, **147**, 214106.

24 K. Shakouri, J. Behler, J. Meyer and G.-J. Kroes, *The Journal of Physical Chemistry Letters*, 2017, **8**, 2131–2136.

25 B. C. Yeo, D. Kim, C. Kim and S. S. Han, *Scientific Reports*, 2019, **9**, 1–10.

26 J. B. L. f. T. Chemie, Ruhr-Universität and 2018, *RuNNer–A neural network code for high-dimensional potential-energy surfaces*.

27 A. Khorshidi and A. A. Peterson, *Computer Physics Communications*, 2016, **207**, 310–324.

28 N. Artrith and A. Urban, *Computational Materials Science*, 2016, **114**, 135–150.

29 P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. Jarrod Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. Carey, İ. Polat, Y. Feng, E. W. Moore, J. Vand erPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt and S. . . Contributors, *Nature Methods*, 2020.

30 M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu and X. Zheng, *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, 2015, `https://www.tensorflow.org/`, Software available from tensorflow.org.

31 A. H. Larsen, J. J. Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, E. D. Hermes, P. C. Jennings, P. B. Jensen, J. Kermode, J. R. Kitchin, E. L. Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. B. Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiøtz, O. Schütt, M. Strange, K. S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng and K. W. Jacobsen, *Journal of Physics: Condensed Matter*, 2017, **29**, 273002.

32 L. I. Bendavid and E. A. Carter, *The Journal of Physical Chemistry C*, 2013, **117**, 26048–26059.

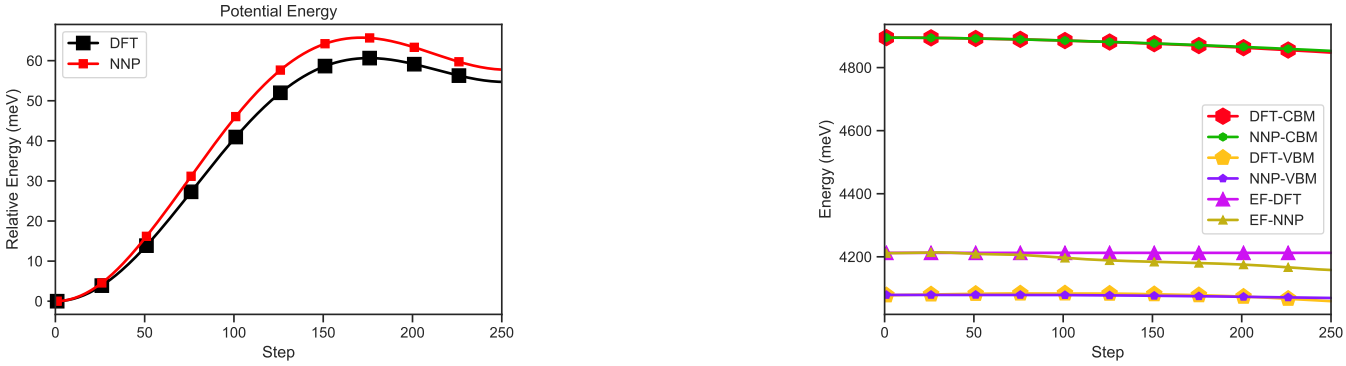
33 Z. Li, J. R. Kermode and A. De Vita, *Physical Review Letters*, 2015, **114**, 096405.

34 J. P. Janet, C. Duan, T. Yang, A. Nandy and H. J. Kulik, *Chemical Science*, 2019, **10**, 7913–7922.

35 A. A. Peterson, R. Christensen and A. Khorshidi, *Physical Chemistry Chemical Physics*, 2017, **19**, 10978–10985.

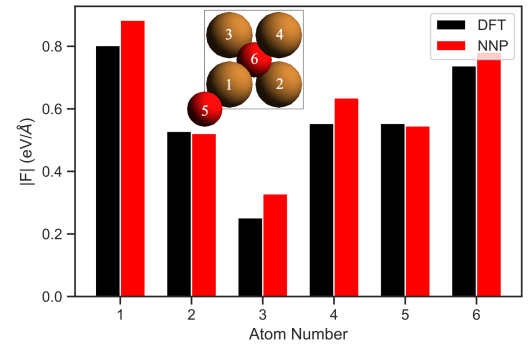
36 I. Cortés-Ciriano and A. Bender, *Journal of Chemical Information and Modeling*, 2019, **59**, 1269–1281.

37 P. Xiao, D. Sheppard, J. Rogal and G. Henkelman, *The Journal of Chemical Physics*, 2014, **140**, 174104.

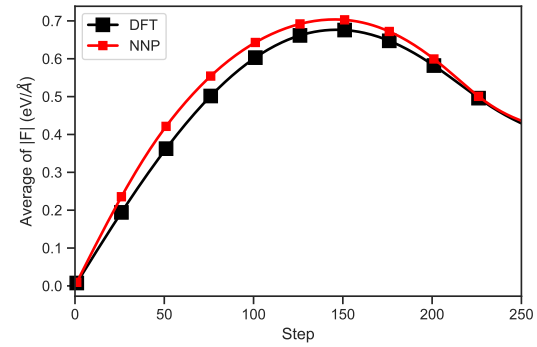


(a) (b)

Fig. 1 Comparison between DFT and NNP (a) Energy (b) Fermi Energy (EF), Valence Band Maximum (VBM), and Conduction Band Minimum (CBM) along the evaluation molecular dynamics trajectory.



(a)



(b)

Fig. 2 Comparison between DFT and NNP forces for (a) all atoms in a randomly selected MD snapshot (inset shows the unit cell structure of Cu_2O bulk used for MD), and (b) average $|F|$ of all atoms along the evaluation molecular dynamics trajectory.

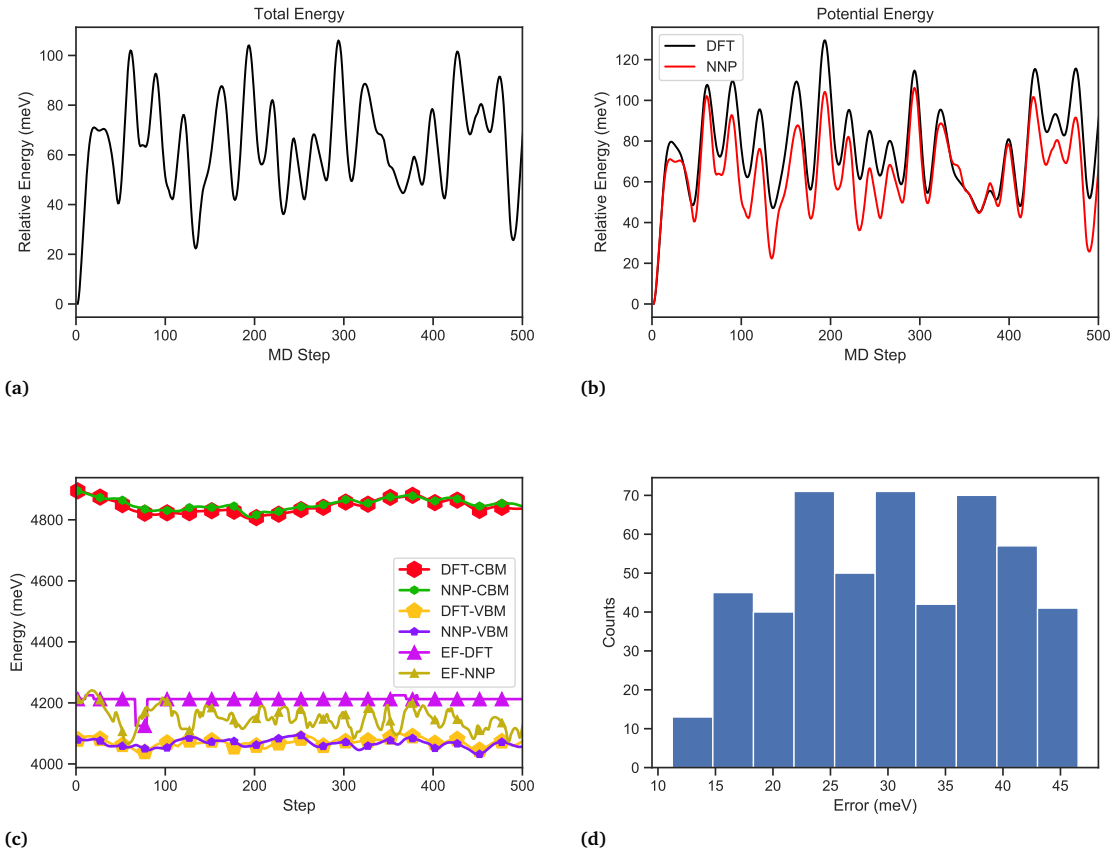


Fig. 3 (a) Total energy (b) comparison of potential energies, (c) comparison between Fermi level, and band edges, and (d) error between DFT and NNP potential energies.

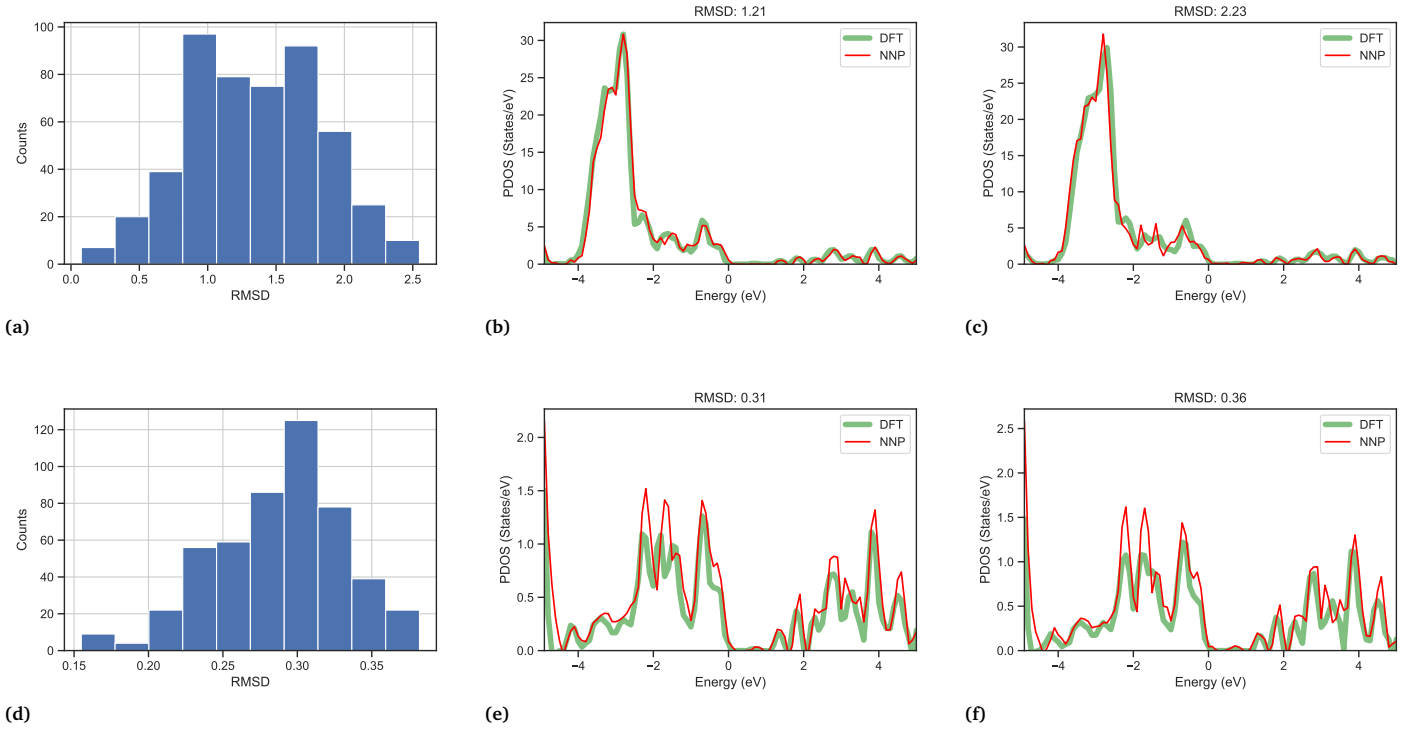
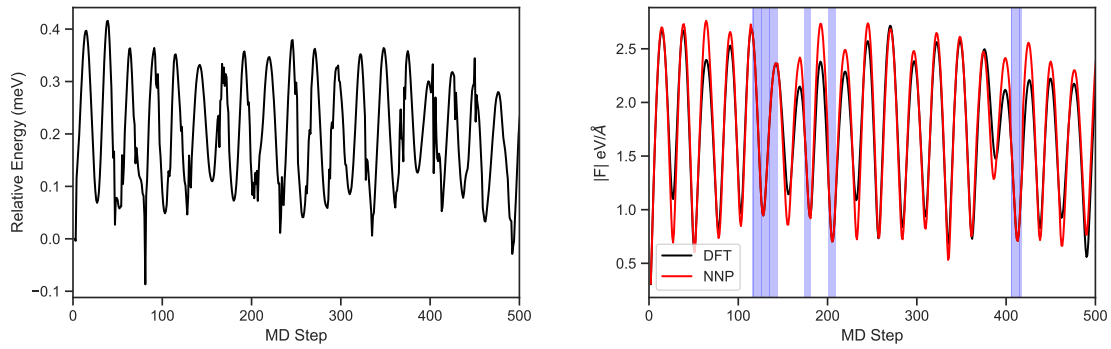
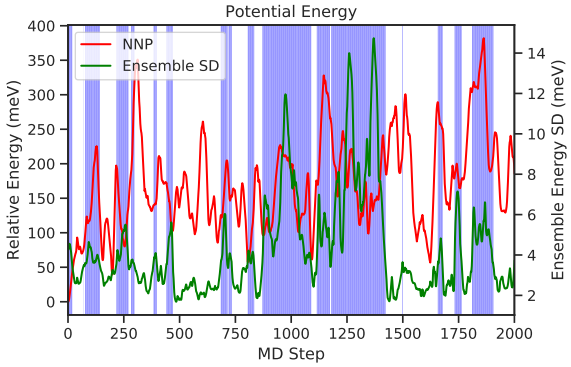


Fig. 4 Comparison between DFT and NNP Partial Density of States. (a) histogram RMSD of PDOS-Cu (b) and (c) PDOS of Cu at two different snapshots, (d) histogram of RMSD of PDOS-O, and (e) and (f) PDOS of O at two different snapshots.



(a) (b)

Fig. 5 (a) Total energy, and (b) comparison of norm of forces on all atoms by DFT and NNP (highlighted regions indicates the use of DFT forces during molecular dynamics).



(a)

Fig. 6 Change of predicted energy and standard deviation of the ensemble energies during 1 ps MD simulation at 300 K (highlighted regions indicate the use of DFT forces during molecular dynamics).

Table 1 Results of the hyperparameter optimization.

Entry	Hidden Layers	RMSE on Evaluation Set						
		E	F	EF	VBM	CBM	PDOS Cu	PDOS O
		(meV)	(meV/Å)	(meV)	(meV)	(meV)	(a.u.)	(a.u.)
1	4-4	345 ± 68	297	14	5	2	1.10	0.25
2	8-8	70 ± 53	258	13	4	3	0.65	0.19
3	16-16	73 ± 45	199	27	6	3	0.53	0.20
4	32-32	15 ± 7	194	43	5	4	0.81	0.20
5	64-64	23 ± 3	108	33	6	2	0.66	0.21
6	128-128	23 ± 13	116	67	4	2	0.89	0.21
7	256-256	26 ± 11	72	86	5	2	1.34	0.21
8	512-512	16 ± 5	57	45	5	1	0.74	0.20
9	4-4-4	234 ± 58	348	19	5	4	0.99	0.25
10	8-8-8	105 ± 50	214	6	10	5	0.58	0.19
11	16-16-16	37 ± 40	170	7	8	6	0.47	0.21
12	32-32-32	33 ± 17	174	18	11	4	0.57	0.20
13	64-64-64	9 ± 2	132	20	6	2	0.59	0.20
14	128-128-128	22 ± 4	97	21	5	3	0.63	0.20
15	256-256-256	17 ± 4	91	50	6	1	0.76	0.21
16	512-512-512	15 ± 5	63	57	6	2	0.80	0.20
17	512-512[a]	21 ± 10	233	50	4	1	0.91	0.20

[a] Rectified Linear Unit was used as activation function.