Modeling H_2O/Rutile-TiO_2(110) Potential Energy Surfaces with Deep Networks

1st Stefan Oehmcke
Machine Learning Group
Copenhagen University
Denmark
ORCID: 0000-0002-0240-1559
Email: stefan.oehmcke@di.ku.dk

2nd Thomas Teusch
Theoretical Chemistry Group
University of Oldenburg
Germany
ORCID: 0000-0002-8250-6018
Email: thomas.teusch@uol.de

3rd Thorben Petersen
Theoretical Chemistry Group
University of Oldenburg
Germany
ORCID: 0000-0003-2052-0567
Email: thorben.petersen@uol.de

4th Thorsten Klüner
Theoretical Chemistry Group
University of Oldenburg
Germany
ORCID: 0000-0003-1389-6013
Email: thorsten.kluener@uol.de

5th Oliver Kramer
Computational Intelligence Group
University of Oldenburg
Germany
ORCID: 0000-0001-7607-1700
Email: oliver.kramer@uol.de

***Abstract*—Successful and cost-effective water splitting could be one of the most interesting energy sources of the future. The modification of the transition metal oxide rutile-TiO_2 as photocatalyst can lead to improved performance in the water splitting process. For that purpose, an accurate description of the interaction potential of a water molecule and the rutile-TiO_2(110) surface in the ground and electronically excited state after photoexcitation is crucial. The electronic Schrödinger equation for the states involved is solved pointwise for different nuclear configurations within the Born-Oppenheimer approximation, and accurate fits to these energy points are required to obtain an analytic expression for the potential energy surface. This is too computationally expensive for fine-grained surface calculations of quantum chemical models. In this paper, we propose to use state-of-the-art deep learning techniques to provide accurate fits for this problem. Namely, we employ a fully connected variant of ResNet and DenseNet with heavy regularization ($L2$, RReLU, Dropout, and BatchNormalization). Previous literature applied neural network approaches before, but with unsatisfactory accuracy. In an experimental evaluation we show that the root mean squared error (RMSE) can be 6.8 times lower for the exited state and 12.7 times lower for the ground state compared to former approaches.**

***Index Terms*—Neural Networks, Water Splitting, ResNet, DenseNet**

I. Introduction

Water splitting represents a promising way to produce molecular hydrogen, which stores chemical energy that can be used in fuel cells. Nowadays, the main source of hydrogen is crude oil (with CO_2 as a by-product), but the former alternative represents an auspicious alternative in the context of green chemistry, because the only byproduct is oxygen. Unfortunately, there is no known method to operate this reaction efficiently. Therefore, not only the search for appropriate catalysts is a cutting-edge research topic, but also the modification of presently applied materials to increase their performance. For this purpose, an understanding of the elementary reaction steps is substantial.

One promising class in this context are transition metal oxides (TMOs). In general, these class of materials exhibits a band gap, which allows for an electronic response due to incident light irradiation. These electronically excited states may open up new reaction pathways for adsorbed molecules. In addition, most TMOs are cheap, non-toxic, and earth abundant. The most prominent representative TMO is rutile-TiO_2. Fujishima and Honda have shown in their now famous experiment, that water can be successfully split up into hydrogen and oxygen using TiO_2 as a photocatalyst [1].

To understand the process of water splitting, a combination of quantum chemical and quantum dynamical studies is necessary. The general workflow sheet for quantum dynamical studies is depicted in Figure 1. The first part contains the calculation of the high-dimensional potential energy surface (PES) in the ground state (GS) and the excited state (ES). In particular, the ES is characterized by a complex topology that can no longer be mapped using classical analytic fit methods. The calculated points therefore represent crucial points on the overall PES. In the second part we utilize these points as a basis for the fitting procedure to acquire a dense grid of points. The task here is to fit and extrapolate the energetic behavior of different configurations of the water molecule

We are very grateful to the DFG for funding this project within the framework of the Research Training Group QM3 (GRK 2247). Stefan Oehmcke acknowledges support from the Danish Industry Foundation through the Industrial Data Analysis Service (IDAS).

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

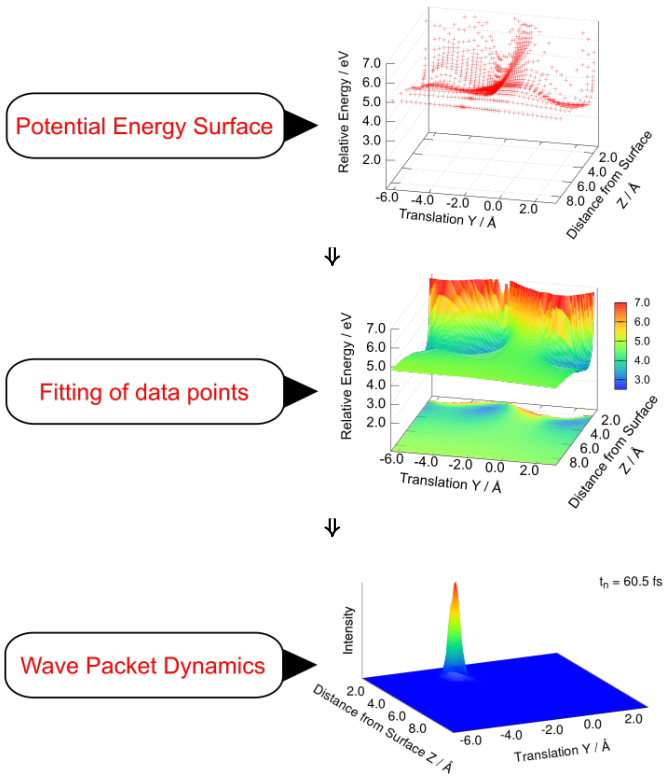


Fig. 1. Working scheme towards quantum dynamical studies. This work focuses on the second task of fitting data points.

on the PES. Finally, we use these fitted PESs for quantum dynamical studies to gain time-resolved insights into the photodissociation process. The present study contributes to the second part of this scheme, to the fitting of data.

To fit these PESs, actual physical experiments are not feasible, as they require intensive preparations and fast measurements that are challenging to acquire. Thus, quantum chemical simulations are employed for this task. However, these are computationally very demanding, i.e. it might take several hours to calculate a single data point. A typical high-dimensional study contains from 10 000 to several 100 000 single energy points depending on the degrees of freedom considered in the chosen adsorbate-substrate system. Hence, the need for an fitting procedure is evident. Previous work has applied a multi-layer-perceptron (MLP) as fitting procedure with the quantum chemical results as training data [2]. Since the given data points exhibit a significantly small simulation error, the fit should not only represent the topology of the energy surface very precisely, but should also be as accurate as possible. Although these MLPs were successfully deployed in the quantum dynamical calculation for the motion of the nuclei, there is still an arguably great need for improvement with respect to their accuracy [3].

In this paper, we apply a fully connected variant of the ResNet [4] and DenseNet [5] architectures to fit these PESs. Our hypothesis is that we can improve on the performance of the existing MLPs by making intensive use of state-of-the-art methods such as batch normalization (BN) [6], dropout [7], and residual connections [4]. We also ensemble the ResNet and DenseNet models to improve the prediction. These assumptions are confirmed by our empirical evaluation, e.g. we achieve a substantially lower RMSE that is 6.8 times lower for the ES and 12.7 for the GS, respectively.

This work is structured as follows: To put our work into context, related work is described in Section II. We explain the used quantum chemical model that defines our data in Section III. Then, the applied architectures are introduced in Section IV. In Section V we present our experimental setup and findings. Finally, we draw our conclusions in Section VI.

II. Related Work

Previous studies approach the fitting of arbitrary PES by representing the total energy as a sum of atomic contributions [8]. For this purpose, a symmetric atomic local environment must be defined to account for the invariance of rotation and translation of the whole system. This is accomplished by introducing a cutoff function for certain interatomic distance thresholds. Consequently, the atomic energies are fitted independently and summed up afterwards.

For instance, [9] fit a two-dimensional PES for the GS and ES of CH_2NH with neural networks. Their chosen network architecture is four parallel running seven layer MLPs for three atoms of the studied molecule, that are eventually summed up to output the energy level. They train on 90 000 *ab initio* data points in a three-dimensional input space, of which 10% are used for testing. For validation an independent data set of 26 522 points was chosen. Their loss function is specific to the model system and incorporates neighboring data points which are identified through a cutoff radius of 6 Å [10]. The final fit of the energy exhibits a RMSE of 0.011 eV for GS and 0.003 eV.

To tackle the problem of generalization [11], [12] introduced an approach that defines permutation invariant polynomial neural networks. It relies on adapting analytic functions reflecting the symmetry of the input data such as the translational periodicity of molecule-surface interactions. Albeit the interaction of the diatomic hydrogen molecule on a rigid Ag(111) surface was proven to be described accurately (RMSE 0.003 eV) [12], working out the symmetric functions for constructing the input data highly depends on the studied surface and might lead to neglecting certain potentially relevant data points.

On a broader scope, Blaschke *et al.* [13] use variational autoencoder to generate completely new molecule structures. They represent a molecular structure as time series. The encoder uses convolutional layers, while the decoder employs recurrent network layers. This technique is not applicable here though, since the molecules are fixed in

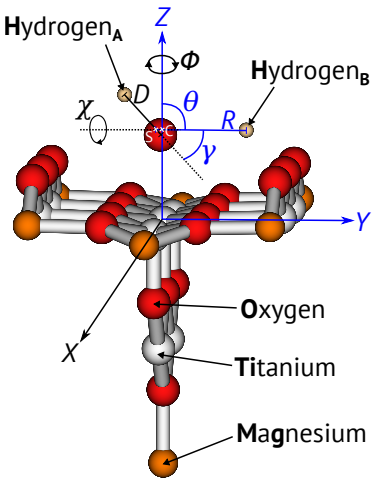


Fig. 2. Illustration of the $H_2O/Ti_9O_{18}Mg_7{}^{14+}$ cluster model with 9D Jacobi coordinates (without point charge field). The labels S and C denote the center of mass for the terminal OH group and the whole water molecule, respectively. Note, that the relevant Jacobi coordinates for our study are depicted in blue.

their study, but the position of the atoms has to be variable for our scope.

There has also been research to fit PESs with other machine learning methods. [14] apply kernel methods called gradient-domain machine learning to find accurate fits, while [15] employ Gaussian process regression. Further, [16] have been incorporating kernel ridge regression to generate new data points, if possible.

III. Quantum Chemical Model System

To investigate the interaction of surfaces with a variety of molecules still represents one of the major topics in condensed matter physics as the knowledge of the underlying steps aids in the design and application of heterogeneous catalysts in industrial processes.

In our study, we used the quantum chemical model system proposed by [2] in order to describe the mechanism of the photocatalytic water splitting reaction with high-level quantum chemistry and quantum dynamical methods. The applied model is depicted in Figure 2. This cluster model is embedded into a cubic field of 4421 point charges to model the solid adequately by incorporating long-range Coulomb effects. Jacobi coordinates are employed to simulate the motion of the water molecule on the rutile-$TiO_2(110)$ surface, as well as the dissociation process [17]. Since it is experimentally proven that dissociation is preceded by electronic excitation, we calculated a five-dimensional PES in θ, γ, Y, Z, and R for the electronic GS and ES, resulting in a total of 171 239 data points for both states, respectively.

In addition to the 171 239 *ab initio* data points, further quantum chemical simulations in the repulsive region of the adsorbate (e.g. for low values in the desorption coordinate Z) were necessary to describe this chemically "forbidden" region. Using a minimal basis set for these coordinates, 103 689 data points for the GS and 64 569 for the ES were additionally calculated.

IV. Network Architectures

In contrast to most deep learning scenarios, in our application, the popular convolutional layers are hard to apply. Although the data represents spatial coordinates with additional information about the rotation of the axes, we cannot create simple 3D-images, as the distance between points is non-equidistant unlike, e.g., in images. We need to be able to define areas with either finer or coarser meshed points to focus on areas of higher interest. Consequentially, we have applied fully connected layers, but also other techniques, to improve the performance of the networks.

A. Baseline MLP

We initially used two MLPs to fit the PES of GS and ES, respectively. These present our baseline. While the ES-model has three hidden layer with $(10 - 15 - 20)$ neurons and a total amount of 566 parameters, the GS-model utilizes two hidden layers with $(10 - 30)$ neurons, which totals to 421 parameters. As non-linear activation function, $tanH$ is employed. For initialization and training of the MLPs, the [18] program package was used. Details of the overall procedure as well as representative two-dimensional PES cuts were presented by [2] and due to low accuracy extended in [3]. Backpropagation according to Bayesian Regularization algorithm was employed to optimize the mean squared error during training [19]. This algorithm was observed to outperform the more simple Levenberg-Marquardt model due to automatic improvement of regularization parameters.

B. ResNet and DenseNet

The ResNet architecture proposed by [4], [20] is highly successful due to its residual connections, which allow the information of previous layers to be used throughout the network. These connections sum up the output of the previous and the current layer. The network learns which residuals f_{l+1} to add to the previous outputs $\boldsymbol{x}_l$ to acquire the next output $\boldsymbol{x}_{l+1}$:

$$\boldsymbol{x}_{l+1} := \boldsymbol{x}_l + f_{l+1} \ . \tag{1}$$

This architecture is well suited for our purposes, as it has been shown that residual connections also help networks with only fully connected layers [21].

ResNet is organized in building blocks which is illustrated in the upper part of Figure 3. We use the pre-activation variant with BN [6], randomized leaky ReLU (RReLU) [22], and fully connected weight layers. BN is normalizing the layer output with mini-batch statistics and helps the training process by beneficially reparameterizing the optimization problem [23]. We choose RReLU as non-linearity function, since it has the advantages of leaky ReLU, but also helps to avoid overfitting. Our ResNet

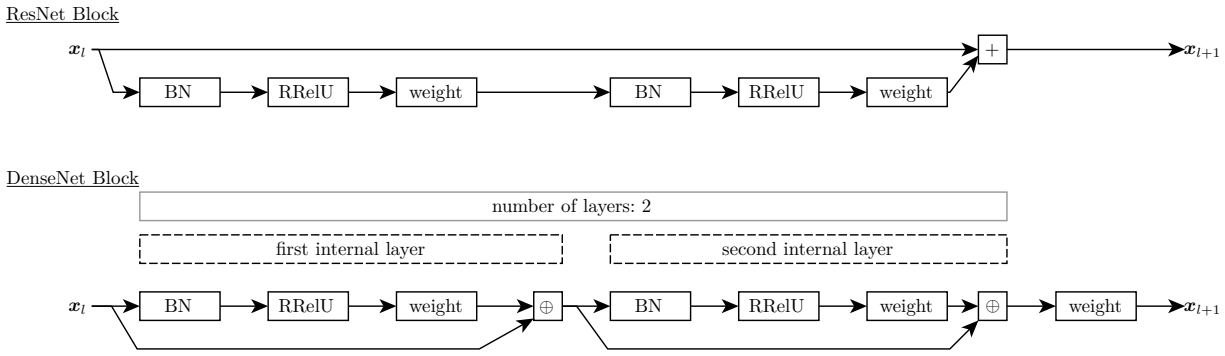


Fig. 3. Building blocks of ResNet and DenseNet with summation operator + and concatenation operator ⊕ employed in this paper. The shown DenseNet block has a layer depth of 2, more internal layers are possible.

model has 8 of these blocks with an increasing number of neurons $(256-512-768-1024)$ every 2 blocks, an initial fully connected layer with 256 neurons, and an output layer with 1024 neurons that are averaged to predict the energy values E. This totals to 18 layers and 8 901 120 parameters. We conducted further tests with more layers observing no improvements.

An architecture that builds upon similar principles is DenseNet proposed by [5]. One of the main differences is that instead of adding up the output of previous layers, here they are concatenated throughout a block, and such a block can contain more than two weight layers. In the lower part of Figure 3 such a block is illustrated. Each internal layer receives the input of all previous layers of this block, leading to a linear increase of the input size for these layers. Our DenseNet model consists of 4 blocks; the numbers of neurons are $(32-64-192-256)$ and the numbers of layers per block are $(6-12-24-16)$. The initial layer contains 16 neurons and the output layer is the same as in the ResNet model. In total, our DenseNet model consists of 64 layers and 7 888 272 parameters.

To train our ResNet and DenseNet models, we employed Adam [24] with AMSGrad [25] as optimizer and a batch size of 1024. We train each model for 10 000 epochs and use the checkpoint with the best performance on the validation set. To reduce overfitting, we apply dropout [7] with a rate of $p = 0.5$ before the output layer, which results in half the neurons being randomly set to zero during one training iteration. Additionally, a $L2$ weight decay of 1×10^{-5} is utilized for all weights. Our experiments are based on implementations in PYTHON using the deep learning framework PYTORCH [26].

V. Experimental Evaluation

Our hypothesis that state-of-the-art deep learning methods create better PES fits than the classical MLP is empirically tested in a quantitative and qualitative way for the H_2O/rutile-TiO_2(110) system.

A. Settings

We train models on the aforementioned data for the GS and ES calculated by a quantum chemical model and compare the baseline MLP with the ResNet, DenseNet, and an ensemble of both models. In addition, for greater comparability, we also include experiments with selected other machine learning algorithms: RandomForest (RF) in its extremely randomized trees variant [27], [28] and a scalable machine learning system for tree boosting (XGBoost) [29]. For these algorithms, we also conducted an extensive hyperparameter search.

Before training, all input data are shuffled randomly, normalized to fall in the range $[-1:1]$ and divided into 70% training, 15% validation, and 15% test data sets. To better quantify our model accuracy, we applied 10-fold cross-validation (CV).

The loss function used by our neural network models is the weighted mean squared error:

$$L := \frac{1}{n}\sum_{i=0}^{n}\frac{1}{w_i^q}\cdot\left(f(\boldsymbol{x}_i)-y_i\right)^2 \ , \tag{2}$$

with the batch size n, input Jacobi-coordinates $\boldsymbol{x}$, model f, and energy level y from the quantum chemical model. Moreover, the parameter w^q (with $0 < w^q \leq 1$) was added to manually account for the quality of the *ab initio* data ($w^q = 1$ for high-precision data points).

We select the RMSE and the coefficient of determination (R^2) as evaluation metrics. For our subsequent quantum dynamical studies, especially the data points around the local minima and the pathways in between (transition states) are important. Thus, a weighted root mean squared error (wRMSE) and weighted coefficient of determination (wR^2) is employed that neglects the aforementioned "forbidden" data points [2]: For the GS, we remove the chemically irrelevant energy values above 7 eV. Further, for the ES, all values above 20 eV are ignored.

B. Results

In Table I we present our results for the ES. We observe that ResNet achieves the best of the single models per-

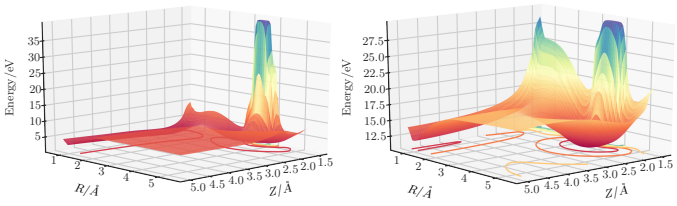
TABLE I
PERFORMANCE ON THE ES-DATA FOR RMSE, WRMSE, R^2, AND WR^2 SCORE OF THE AVERAGED CV-RUNS. THE BEST RESULTS ARE HIGHLIGHTED IN BOLD FONT.

model	RMSE	wRMSE	R^2	wR^2
MLP (base)	0.835 ± 0.065	0.321 ± 0.014	$0.9811 \pm 2.838 \times 10^{-3}$	$0.9788 \pm 1.853 \times 10^{-3}$
RF	0.250 ± 0.009	0.212 ± 0.011	$0.9983 \pm 0.124 \times 10^{-3}$	$0.9907 \pm 1.002 \times 10^{-3}$
XGBoost	0.212 ± 0.011	0.166 ± 0.008	$0.9907 \pm 0.222 \times 10^{-3}$	$0.9943 \pm 0.530 \times 10^{-3}$
ResNet	0.133 ± 0.004	0.126 ± 0.004	$0.9995 \pm 0.032 \times 10^{-3}$	$0.9967 \pm 0.210 \times 10^{-3}$
DenseNet	0.138 ± 0.010	0.131 ± 0.009	$0.9995 \pm 0.077 \times 10^{-3}$	$0.9964 \pm 0.510 \times 10^{-3}$
Ensemble	$\mathbf{0.123} \pm 0.006$	$\mathbf{0.119} \pm 0.006$	$\mathbf{0.9996} \pm 0.043 \times 10^{-3}$	$\mathbf{0.9970} \pm 0.295 \times 10^{-3}$

formances, but the ensemble can further improve on these results. Nevertheless, it is noticeable that the ensemble model has a higher standard deviation than ResNet. To ignore high energy values improves the performance for wRMSE, especially for the MLP approach, but the error is still more than two times as high as compared to our new models. Interestingly, the results for R^2 generally show higher values than for wR^2, most likely due to the smaller overall variance of values when neglecting high energy values. The RF and XGBoost models performed notably better than the MLP, but cannot achieve the same accuracy as deeper NNs. In short, the best model compared to the MLP has better evaluation metrics by a factor of ∼6.8 for RMSE, ∼2.7 for wRMSE, and ∼1.02 for R^2 as well as wR^2.

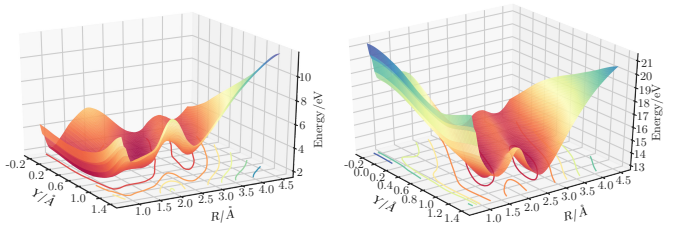
Table II shows the results of our models on the GS-data. Here, DenseNet achieves the best results without the weighting of the metrics. With the metric weighting, DenseNet offers a slightly worse performance than XGBoost, but DenseNet would still be preferable as it has a lower variance over the runs. The ensemble approach does not help here. Note, that the wRMSE of the MLP is nearly as high as the ResNet error. Again, wR^2 results are scaled differently and thus show lower score values, except for the MLP, where the score actually is better than the unweighted one. Here, the relative improvements compared to the MLP are better by a factor of ∼12.7 for RMSE, ∼1.8 for the wRMSE, ∼1.1 for R^2, and ∼1.03 for wR^2. Note that the different states cannot be compared since they have different target distributions, which explains why the error appears to be worse on the GS although the ES just has smaller value differences in general.

In Figures 4, 5, 6, and 7 we display projections of the PES fits chosen by experts, where three input dimensions are fixed. These are the same projections that are used in the work of [2], since they have distinct physical properties and show complex topologies that help to qualify a fit. To calculate these projections, we applied an ensemble of models from all CV-runs, where we combined ResNet and DenseNet for the ES and only the DenseNet models for the GS. As expected the GS-PESs show lower energy values and the topology of the ES-PES is more complex. Only Figure 7 shows a nearly identical topology, but still the energy level is significantly higher. All the PESs are smooth and show no sudden jumps, which would be implausible from a physical perspective. The experts use these projections to evaluate the topology for physical plausibility and confirm that our models are well fitted.



(a) ground state (b) excited state

Fig. 4. Projections with $R/\mathring{A}$ and $Z/\mathring{A}$ against energy/eV with $Y/\mathring{A} = -0.3$, $\theta = 88.0°$, $\gamma = 50.0°$.



(a) ground state (b) excited state

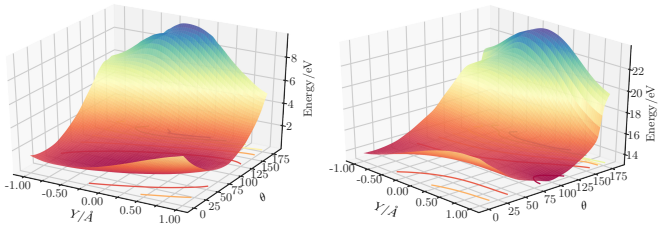
Fig. 5. Projections with $Y/\mathring{A}$ and $R/\mathring{A}$ against energy/eV with $Z/\mathring{A} = 1.82$, $\theta = 82.0°$, $\gamma = 40.0°$.

VI. CONCLUSION

In this paper, we apply state-of-the-art neural network techniques to fit PESs for the GS and ES of the H_2O/rutile-TiO_2(110) system. We employ ResNet and DenseNet architectures that are completely fully connected and benefit from various regularization techniques ($L2$, Dropout, RReLU, BatchNormalization). The results presented here offer new insights into a highly complex photochemical process and can contribute significantly to

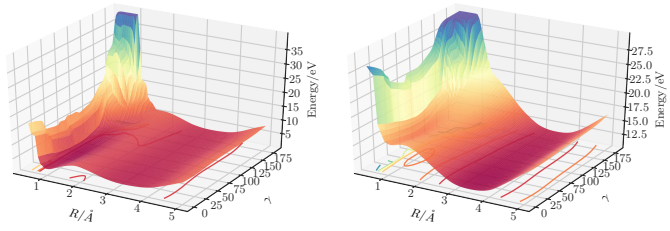
TABLE II
PERFORMANCE ON THE GS-DATA FOR RMSE, WRMSE, R^2, AND WR^2 SCORE OF THE AVERAGED CV-RUNS. THE BEST RESULTS ARE HIGHLIGHTED IN BOLD FONT.

model	RMSE	wRMSE	R^2	wR^2
MLP (base)	3.605 ± 0.570	0.362 ± 0.025	$0.9107 \pm 2.999 \times 10^{-2}$	$0.9589 \pm 5.619 \times 10^{-3}$
RF	0.419 ± 0.017	0.237 ± 0.022	$0.9988 \pm 0.001 \times 10^{-2}$	$0.9822 \pm 3.349 \times 10^{-3}$
XGBoost	0.440 ± 0.025	$\mathbf{0.202 \pm 0.019}$	$0.9987 \pm 0.015 \times 10^{-2}$	$\mathbf{0.9870 \pm 2.562 \times 10^{-3}}$
ResNet	0.568 ± 0.026	0.318 ± 0.021	$0.9978 \pm 0.021 \times 10^{-2}$	$0.9683 \pm 4.166 \times 10^{-3}$
DenseNet	$\mathbf{0.284} \pm 0.018$	$\mathbf{0.204 \pm 0.012}$	$\mathbf{0.9995} \pm 0.007 \times 10^{-2}$	$\mathbf{0.9869 \pm 1.503 \times 10^{-3}}$
Ensemble	0.362 ± 0.018	0.234 ± 0.013	$0.9991 \pm 0.009 \times 10^{-2}$	$0.9828 \pm 1.829 \times 10^{-3}$



(a) ground state (b) excited state

Fig. 6. Projection with $Y/\mathring{A}$ and $\theta/°$ against energy/eV with $Z/\mathring{A} = 2.2$, $R/\mathring{A} = 0.98$, $\gamma = 65.0°$.



(a) ground state (b) excited state

Fig. 7. Projection with $R/\mathring{A}$ and $\gamma/°$ against energy/eV with $Y/\mathring{A} = 0.0$, $Z/\mathring{A} = 2.2$, $\theta = 88.0°$.

the mechanistic understanding. They represent an important part in the optimization of photocatalysts without the need for experiments.

In future work, it might be interesting to include more information about the PES topology into the networks *a priori*, e.g. by introducing an educated guess upon initialization of the neural network weights. Another interesting line of research is the viability of transfer-learning techniques of neural network models to fit PES other, comparable adsorbate-substrate systems.

REFERENCES

[1] A. Fujishima and K. Honda, "Electrochemical photolysis of water at a semiconductor electrode," *Nature*, vol. 238, pp. 1476–4687, 1972. [Online]. Available: https://doi.org/10.1038/238037a0

[2] J. Mitschker and T. Klüner, "Photodesorption of water from rutile(110): ab initio calculation of five-dimensional potential energy surfaces of ground and excited electronic states and wave packet studies," *Physical Chemistry Chemical Physics (PCCP)*, vol. 17, pp. 268–275, 2015. [Online]. Available: http://dx.doi.org/10.1039/C4CP04593A

[3] T. Petersen, J. Mitschker, and T. Klüner, "High-dimensional wave packet dynamics from first principles: Photodissociation of water on tio2-rutile (110)," *Journal of Photochemistry and Photobiology A*, vol. 366, pp. 3–11, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1010603017316696

[4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2016, pp. 770–778.

[5] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2017, pp. 2261–2269.

[6] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International Conference on Machine Learning (ICML)*. International Machine Learning Society, 2015, pp. 448–456.

[7] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research (JMLR)*, vol. 15, no. 1, pp. 1929–1958, 2014.

[8] J. Behler and M. Parrinello, "Generalized neural-network representation of high-dimensional potential-energy surfaces," *Physical Review Letters (PRL)*, vol. 98, p. 146401, 2007. [Online]. Available: http://dx.doi.org/10.1103/PhysRevLett.98.146401

[9] W.-K. Chen, X.-Y. Liu, W.-H. Fang, P. O. Dral, and G. Cui, "Deep learning for nonadiabatic excited-state dynamics," *Journal of Physical Chemistry Letters (JPCL)*, vol. 9, no. 23, pp. 6702–6708, 2018.

[10] L. Zhang, J. Han, H. Wang, R. Car, and E. Weinan, "Deep potential molecular dynamics: a scalable model with the accuracy of quantum mechanics," *Physical Review Letter (PRL)*, vol. 120, no. 14, p. 143001, 2018.

[11] B. Jiang and H. Guo, "Permutation invariant polynomial neural network approach to fitting potential energy surfaces," *Journal of Chemical Physics (JCP)*, vol. 139, no. 5, p. 054112, 2013. [Online]. Available: https://doi.org/10.1063/1.4817187

[12] ——, "Permutation invariant polynomial neural network approach to fitting potential energy surfaces. iii. molecule-surface interactions," *Journal of Chemical Physics (JCP)*, vol. 141, no. 3, p. 034109, 2014. [Online]. Available: https://doi.org/10.1063/1.4887363

[13] T. Blaschke, M. Olivecrona, O. Engkvist, J. Bajorath, and H. Chen, "Application of generative autoencoder in de novo molecular design," *Molecular Informatics*, vol. 37, no. 1-2, p. 1700123, 2018. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/minf.201700123

[14] 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.

[15] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, "Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons," *Physical Review Letters (PRL)*, vol. 104, no. 13, p. 136403, 2010.

[16] V. Botu and R. Ramprasad, "Adaptive machine learning framework to accelerate ab initio molecular dynamics," *International*

Journal of Quantum Chemistry (IJQC), vol. 115, no. 16, pp. 1074–1083, 2015.

[17] R. Schinke, *Photodissociation dynamics – spectroscopy and fragmentation of small polyatomic molecules*. Cambridge Univ. Press, Cambridge, 1995.

[18] "Matlab, *version 8.2.0.701 (R2013b)*," The MathWorks Inc., 2013.

[19] D. J. C. MacKay, "Bayesian interpolation," *Neural Computation*, vol. 4, no. 3, pp. 415–447, 1992. [Online]. Available: https://doi.org/10.1162/neco.1992.4.3.415

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *European Conference on Computer Vision (ECCV)*, 2016, pp. 630–645.

[21] J. Martinez, R. Hossain, J. Romero, and J. J. Little, "A simple yet effective baseline for 3d human pose estimation," in *IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2640–2649.

[22] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical evaluation of rectified activations in convolutional network," *Computing Research Repository (CoRR)*, vol. abs/1505.00853, 2015. [Online]. Available: http://arxiv.org/abs/1505.00853

[23] S. Santurkar, D. Tsipras, A. Ilyas, and A. Madry, "How does batch normalization help optimization?" in *Advances in Neural Information Processing Systems (NeurIPS)*, 2018, pp. 2488–2498.

[24] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *Computing Research Repository (CoRR)*, vol. abs/1412.6980, 2014. [Online]. Available: http://arxiv.org/abs/1412.6980

[25] S. J. Reddi, S. Kale, and S. Kumar, "On the convergence of adam and beyond," in *International Conference on Learning Representations (ICLR)*, 2018. [Online]. Available: https://openreview.net/forum?id=ryQu7f-RZ

[26] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in pytorch," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

[27] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001. [Online]. Available: https://doi.org/10.1023/A:1010933404324

[28] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Machine Learning*, vol. 63, no. 1, pp. 3–42, 2006. [Online]. Available: https://doi.org/10.1007/s10994-006-6226-1

[29] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *International Conference on Knowledge Discovery and Data Mining (SIGKDD)*. ACM, 2016, pp. 785–794.