PCCP

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal

Cite this: DOI: 10.1039/d3cp06312g

Accessing the usefulness of atomic adsorption configurations in predicting the adsorption properties of molecules with machine learning†

Walter Malone, *[a] Johnathan von der Heyde[b] and Abdelkader Kara[b]

Received 28th December 2023,
Accepted 18th March 2024

DOI: 10.1039/d3cp06312g

rsc.li/pccp

We present a systematic study into the effect of adding atomic adsorption configurations into the training and validation dataset for a neural network's predictions of the adsorption energies of small molecules on single metal and bimetallic, single crystal surfaces. Specifically, we examine the efficacy of models trained with and without H and X atomic adsorption configurations, where X is C, N, or O, to predict XH_n adsorption energies. In addition, we compare our machine learning models to traditional simple scaling relationships. We find that models trained with the atomic adsorption configurations outperform models trained with only molecular adsorption configurations, with as much as a 0.37 eV decrease in the MAE. We find that models trained with the atomic adsorption configurations slightly outperform traditional scaling relationships. In general, these results suggest it may be possible to vastly reduce the number of adsorption configurations one needs for training and validation datasets by supplementing said data with the adsorption configurations of composite atoms or smaller molecular fragments.

1. Introduction

Machine learning (ML) emerged recently as a tool to accelerate the study of interfaces, especially the adsorption of small molecules on surfaces. Studies have successfully predicted properties of the adsorbate/substrate system such as adsorption energy,[1–15] reaction barrier,[16] adsorption height,[17] and buckling of the surface.[17] The main application in many of these studies has been heterogenous catalysis where one can employ adsorption energy or another adsorption property as a chemical descriptor,[18–21] or an indicator of catalytic activity. The authors of these studies hope to rapidly predict adsorption properties, such as adsorption energy, and map them to the efficacy of the catalysts they are studying. Thus, by the quick and efficient calculation of adsorption properties using ML algorithms, one can rapidly screen for novel, more efficient catalysts. However, potential applications of ML methods to surface adsorption problems are not limited to heterogenous catalysis. These methods may be easily applied to any system where molecules interact with an interface such as energy storage systems[22] or light-harvesting systems.[23]

Despite their successes, ML methods are often held back from widespread deployment to surface science problems by the methods' requirement of a large database of experimental observations or high-level theoretical calculations to serve as a training dataset. The emergence of large datasets of high-level theoretical calculations and online data repositories to store and distribute those datasets, has somewhat mitigated this obstacle to the widespread application of ML methods to material discovery,[24–27] However, due to the relatively large computational time associated with surface calculations, it remains imperative to be as efficient as possible with the data available.

Being able to predict the adsorption properties of larger, more complex adsorbates from smaller, simpler adsorbates would vastly improve data efficiency. For example, using the adsorption properties of C, H, and CH to predict the adsorption properties of CH_3 or CH_4 would greatly decrease computational cost, improving efficiency. This relationship between the adsorption properties of smaller adsorbates and similar, larger adsorbates is known as a scaling relationship[28–38] and has been successfully deployed many times without the inclusion of ML. With a traditional scaling relationship, one assumes the adsorption energy of the larger adsorbate scales linearly with the adsorption energy of the smaller adsorbate. Scaling relationships have been known to fail with certain adsorbates and

[a] Department of Physics, Tuskegee University, 1200 W. Montgomery Rd., Tuskegee, AL 36088, USA. E-mail: wmalone@tuskegee.edu
[b] Department of Physics, University of Central Florida, 4000 Central Florida Blvd., Orlando, Florida, 32816, USA

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d3cp06312g

larger molecules,[37,38] making ML an enticing option for resolving these nuances.

In this manuscript, we report that ML methods can be greatly aided in the prediction of the adsorption properties of larger, more complex adsorbates from smaller, simpler related adsorbates. Specifically, we utilize the adsorption energies of N and H, O and H, and C and H over various single metal and bimetallic single crystal surfaces to improve the predictions of the adsorption energies of NH, OH and H_2O, and CH, CH_2, and CH_3, respectively. We demonstrate that including the atomic adsorption configurations can drastically reduce the number of data points we need in the training datasets for the more complex molecules to make successful predictions. Importantly, we also directly compare our ML results to traditional scaling relationships. In Section 2 we present the computational details of our calculations, in Section 3 we present our results, and in Section 4 we present our conclusions.

2. Computational details

For our training dataset, we utilized density functional theory (DFT) calculations on H, C, O, N, CH, CH_2, CH_3, NH, OH, and H_2O adsorbed on various single crystal, single metal, and bimetallic fcc (111) surfaces. DFT, which models the total energy as a function of electron density, provides a good tradeoff between computational cost and accuracy. The reference calculations were taken from a high-throughput screening study[39] conducted by Mamun *et al.*, for which the data is available for free on Cataylsis-Hub.org.[24] The dataset was generated with the BEEF-vdW functional,[40] which contains a correction for the van der Waals (vdW) interaction. The vdW interaction is known to be important for the description of many systems including small molecules adsorbed on transition metal surfaces.[41–47]

As the details of the calculations can be found in ref. 39, here we will only give a brief overview of the methods utilized to create the dataset. To begin, each surface constructed was 3 layers thick. They fixed the bottom two layers of the slab at their bulk value and allowed the top layer and adsorbate to relax. To limit the interaction between slabs, at least 17 Å of vacuum were placed in between slabs. Each layer is a 2×2 structure. To sample the Brillouin zone the authors utilized a $(12 \times 12 \times 12)$ and a $(6 \times 6 \times 1)$ Monkhorst–Pack *k*-point grid for bulk and slab calculations respectively. Moreover, they used a 500 eV planewave cutoff and 5000 eV density cutoff for all the calculations. The authors tried all stoichiometries 4 : 0, 3 : 1, 2 : 2, 1 : 3, 0 : 4 for elements A : B, where A and B were elements with possible atomic numbers 13, 21–31,39–50, 57, or 72–83. For elements Fe, Ni, Co, and Mn the authors performed spin-polarized calculations with starting magnetic moments of 3, 3, 2, and $1\mu_B$ respectively. Overall, the dataset of calculations consisted of 9278, 8078, 8333, 8430, 2749, 2193, 1916, 2663, 2887, and 2430 adsorption calculations for H, C, O, N, CH, CH_2, CH_3, NH, OH, and H_2O respectively. The H, C, O, and N datasets were pruned according to ref. 17, removing configurations that led to more than 0.79 Å of buckling of the first layer of the surface. Adsorption energy was calculated as:

$$E_{\text{ads}} = -E_{\text{molecule/surface}} + E_{\text{clean surface}} + \sum_i \alpha_i E_{i,\text{molecule}} \quad (1)$$

where $E_{\text{molecule/surface}}$ is the total energy of the molecule on the surface, $E_{\text{clean surface}}$ is the total energy of the clean surface, and E_{molecule} are the total energies of the gas phase species chosen from H_2, C_2H_2, C_2H_4, C_2H_6, O_2, H_2O, and NH_3. Finally, Fig. 1 illustrates the adsorption sites studied in ref. 39 to build the datasets. In Fig. 1 atop sites are denoted with squares, bridge sites are denoted with triangles, and hollow sites are denoted with circles. For the 2 : 2 surfaces the authors sampled 10 different adsorption sites, see Fig. 1(a), for the 3 : 1 And 1 : 3 surfaces the authors sampled 9 different adsorption sites, see Fig. 1(b), and for the single element surfaces the authors sampled four different adsorption sites, see Fig. 1(c). All adsorption sites and surfaces were attempted for H, C, N, and O adsorption while a limited, due to computational resource constraints, number of surfaces and adsorption sites were sampled for the molecular adsorption.

For our machine learning needs, we employed the Hierarchically Interacting Particle Neural Network[48] (HIP-NN). HIP-NN, a deep neural network, has been successfully utilized to predict the adsorption energies, adsorption heights, and buckling of the first layer of the surface for small adsorbates on

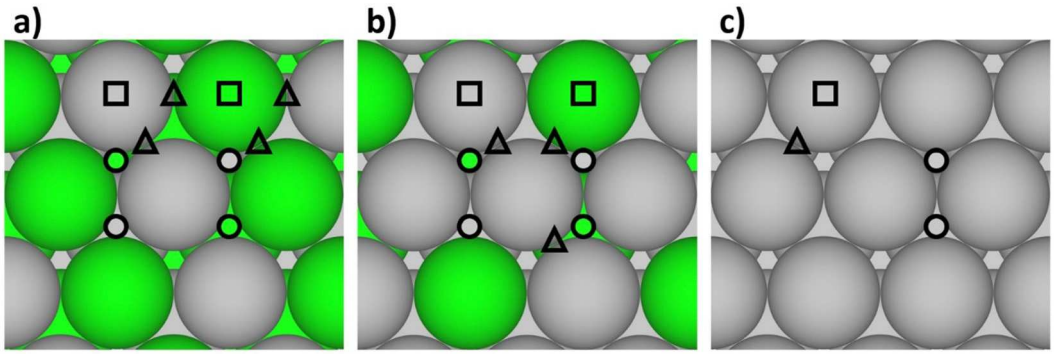

Fig. 1 Adsorption sites sampled for the (a) 2 : 2 surfaces, (b) 3 : 1 and 1 : 3 surfaces, and (c) single element surfaces. Squares denote atop sites, triangles denote bridge sites, and circles hollow sites.

transition metal surfaces.[17] HIP-NN makes predictions with both the local atomic density, similar to a continuous convolutional network, and information about atom pairs, similar to a message-passing neural network. Specifically, HIP-NN builds predictions of each quantum observable one is interested in hierarchically, meaning HIP-NN takes some global property like adsorption energy and breaks it into learnable local contributions of order n. Within this framework, HIP-NN contains both onsite and interaction layers. Onsite layers gather information about individual atoms up to some cutoff while interaction layers transmit data between atoms within the same cutoff. Given HIP-NN's ease of use, prior success, and explicit inclusion of periodic boundary conditions, we successfully utilized it to predict the adsorption energies of NH, OH, H_2O, CH, CH_2, and CH_3 on various transition metal surfaces. The hyperparameters were taken from ref. 17, which contained optimized HIP-NN hyperparameters for the adsorption of atoms on transition metal surfaces. For brevity, the number of features was set at 20, the number of sensitivity functions at 60, the number of interaction layers at 3, the number of atoms per layer at 4, the soft cutoff at 5.3 Å, and the hard cutoff at 9.0 Å.

To begin with, we trained several HIP-NN models, with periodic boundary conditions, to predict the adsorption energy of CH, CH_2, CH_3, NH, OH, and H_2O. We began with 60% of the configurations for each molecule in the training dataset, 20% in a validation dataset, and the remaining 20% in a held-out testing dataset. One purpose of the validation dataset is to help with the overtraining problem. We kept a new model if it performed better on the validation dataset. We then periodically reduced the number of configurations in the training/validation dataset to 1250 configurations, 750 configurations, 500 configurations, and then to 250 configurations, training new models after each reduction and placing the removed configurations in the held-out testing dataset. We took a similar approach for models trained with H, C, N, or O adsorption configurations. We began with a model trained only to C and H, N and H, or O and H adsorption configurations. We then added 10, 50, 250, or 1250 CH, CH_2, CH_3, NH, OH, or H_2O adsorption configurations, training a new model after each new addition of configurations to our training and validation dataset. We would test the model on the remaining CH, CH_2, CH_3, NH, OH, or H_2O configurations not in the training or validation dataset. For the models trained with H and C, N, or O adsorption configurations along with 250 or 1250 CH, CH_2, CH_3, NH, OH, or H_2O adsorption configurations, we also tried increasing the weight of the molecular adsorption configurations two or five-fold in our training and validation dataset.

3. Results and discussion

(A) Machine learning results

The results of each model's MAE for molecular adsorption configurations in the testing dataset are displayed in Fig. 2. The numerical values are listed in Table S1 in the ESI.† In Fig. 2 we exclude the models only trained to C and H, N and H, or O and H adsorption configurations as these models, as perhaps expected, performed poorly on molecular adsorption configurations. From Fig. 2 we see that models trained on only the molecules, CH, CH_2, CH_3, NH, OH, or H_2O, performed, generally, more poorly than models that contained the atomic adsorption configurations as well as an equal amount of molecular adsorption configurations. On average, we find a 0.16 eV decrease in MAE for the 250 configuration models, a 0.13 eV decrease in MAE for the 500 configuration models, a 0.14 eV decrease in MAE for the 750 configuration models, and a 0.13 eV decrease in MAE for the 1250 configuration models when introducing the atomic adsorption configurations into the training and validation dataset. However, the decrease in MAE is not uniform among the different models.

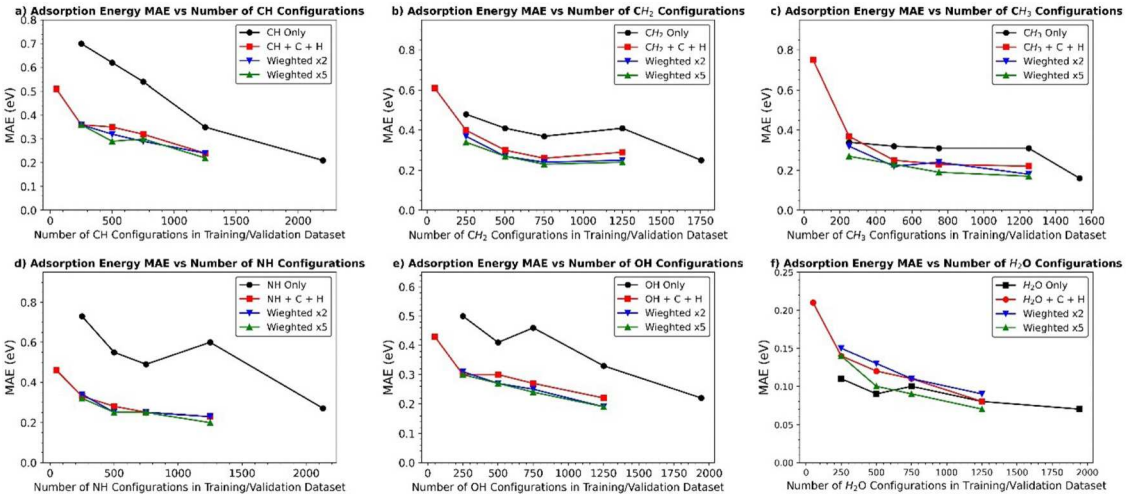

Fig. 2 Adsorption energy MAE *versus* number of XH_n configurations in the training and validation dataset for (a) CH, (b) CH_2, (c) CH_3, (d) NH, (e) OH, and (f) H_2O. Black points correspond to only XH_n configurations in the training and validation dataset, red points correspond to X, H, and XH_n configurations in the training and validation dataset, and blue and green points correspond to X, H, and XH_n configurations in the training and validation dataset with XH_n configurations weighted two-fold and five-fold respectively. Lines are provided to guide the reader's eye.

For the 250 configuration models, the difference is largest for NH, where the difference in MAE between the model trained with only NH configurations and the model trained with NH, N, and H configurations is 0.40 eV. Moreover, the difference in MAE is smallest for H_2O and CH_3, where the difference in MAE between the model trained with only H_2O configurations and the model trained with OH, N, and H configurations is 0.03 eV, which is smaller than the accuracy of the DFT reference data. For the 500, 750, and 1250 configuration models, we note a similar story. The difference in MAE is largest for NH, where the model trained with the atomic adsorption configurations outperforms the model trained without the atomic adsorption configurations by 0.27 eV, 0.24 eV, and 0.37 eV for the 500, 750, and 1250 configuration models. Moving further, the difference in MAE is smallest for H_2O where both models, those trained with atomic adsorption configurations and those trained without atomic adsorption configurations, perform about the same, less than 0.03 eV, for the 500, 750, and 1250 configuration models. In fact, most models, except those models trained with less than 250 H_2O configurations in the training and validation dataset, perform roughly the same, within 0.08 eV MAE, which is again less than the accuracy of DFT. We believe this to be the case as, on most of the surfaces, H_2O physisorbs (weak adsorption). With many adsorption configurations close to 0.0 eV in adsorption energy, one would expect a lower overall MAE for all models. The case of H_2O will be discussed in further detail later. In general, these results demonstrate that including atomic adsorption configurations in one's training and validation dataset will noticeably improve their model's predictions on the molecules' adsorption energies.

Moreover from Fig. 2, we further note that the models trained with X, H, and 1250 XH_n configurations generally approach or exceed the accuracy of the models trained with more molecular adsorption configurations without the atomic adsorption configurations. To sum up the results, for CH, Fig. 2(a), the model trained with atomic adsorption configurations and 1250 molecular adsorption configurations, produces a MAE of 0.22 eV while a model trained with the full 2200 molecular adsorption configurations gives a similar MAE of 0.21 eV. For CH_2, Fig. 2(b), the model trained with atomic adsorption configurations and 1250 molecular adsorption configurations, weighted five-fold, produces a MAE of 0.24 eV while a model trained with the full 1755 molecular adsorption configurations gives a similar MAE of 0.25 eV. For CH_3, Fig. 2(c), the model trained with atomic adsorption configurations and 1250 molecular adsorption configurations, weighted five-fold, produces a MAE of 0.17 eV while a model trained with the full 1533 molecular adsorption configurations gives a similar MAE of 0.16 eV. For NH, Fig. 2(d), the model trained with atomic adsorption configurations and 1250 molecular adsorption configurations, weighted five-fold, produces a MAE of 0.20 eV while a model trained with the full 2131 molecular adsorption configurations gives inferior predictions with a MAE of 0.27 eV. For OH, Fig. 2(e), the model trained with atomic adsorption configurations and 1250 molecular adsorption configurations, weighted five-fold, produces a MAE of 0.19 eV while a model trained with the full 2310 molecular adsorption configurations gives a similar MAE of 0.22 eV. Finally, for H_2O, Fig. 2(f), the model trained with atomic adsorption configurations and 1250 molecular adsorption configurations, weighted five-fold, produces a MAE of 0.07 eV while a model trained with the full 1944 molecular adsorption configurations gives the same MAE of 0.07 eV. This data serves as further evidence of the importance of including the atomic adsorption configurations in the training and validation dataset. It also suggests that one can reduce the size of one's training and validation dataset if one includes atomic adsorption configurations, providing one a means to greatly reduce the computational time one expends building one's training and validation dataset if the atomic adsorption configurations are readily available.

As hinted at above, weighting the molecular configurations either two or five-fold within the dataset often led to small improvements or no improvements, that is MAEs of the weighted models were within 0.01 eV of the unweighted models. Overall, we witnessed the largest change in MAE for the CH_3 250 configuration models. In the model where we weighted the CH_3 molecular configurations five-fold, the MAE decreased by 0.10 eV compared to the unweighted model. While there is not a huge decrease in MAE, and sometimes we do not see any change in MAE, the fact that the error would never appreciably increase for predictions on the testing dataset illustrates that weighting unrepresented configurations a bit more may be a viable strategy to improve results. It can lead to one's model better capturing the chemistry one is interested in when data is limited.

Moving on, in general, we note a modest decrease in MAE by increasing the number of molecular adsorption configurations in our training dataset from 250 to 500 to 750 to 1250 molecular configurations with or without the atomic configurations present. There are exceptions to this general rule, though mostly in the models trained with no atomic adsorption configurations. For example, we see a 0.11 eV and 0.05 eV spike in MAE for NH and OH going from smaller to larger training datasets, demonstrating that purely molecular adsorption configuration training datasets are likely insufficient to train these models. Another indicator that these models fall short are their MAEs, which rise above 0.33 eV for OH and 0.49 for NH. The large overall MAEs and sporadic increases in MAE demonstrates that the purely molecular configuration datasets are insufficient to train a robust model to make adsorption energy predictions, especially for NH and OH. For the models that include atomic adsorption configurations, we note a smoother more consistent decrease in MAE as the training dataset increases. For NH we record a decrease from a MAE of 0.33 eV to 0.23 eV going from a model trained with 250 molecular configurations to 1250 molecular configurations. For the same size datasets for OH, CH, and CH_3 we witness a decrease of MAE from 0.30 eV, 0.36 eV, and 0.37 eV to 0.22 eV, 0.24 eV, and 0.22 eV respectively. For H_2O, we note only a 0.06 eV decrease in MAE likely as HIP-NN has nearly reached its limit of accuracy. Finaly for CH_2, the MAE, like H_2O, also seems to reach a limit, oscillating

±0.04 eV, from the 500 configuration model to the 1250 configuration model. Weighting the molecular configurations appears to help break this limit, lowing the MAEs closer to that of NH or CH. CH_2 will be discussed further in detail a little later. Taken together with the generally larger decrease in MAE when including atomic adsorption configurations, these results indicate that the presence of the atomic configurations is more important to improve training than more molecular configurations. This demonstrates that once the network possesses enough data to predict on the atomic species, it does not require much more (250 to 1250 configurations) to make accurate predictions on molecular species that contain the relevant atomic species. This suggests that it may be easier to learn how the X–H bond affects adsorption than the adsorption characteristics of the X atomic species or X–H molecular species outright. Further support for this conclusion comes from the fact that the 1250 configurations weighted models, trained with the atomic species, perform better than or on par with, within 0.01 eV, models trained with more molecular configurations and no atomic configurations.

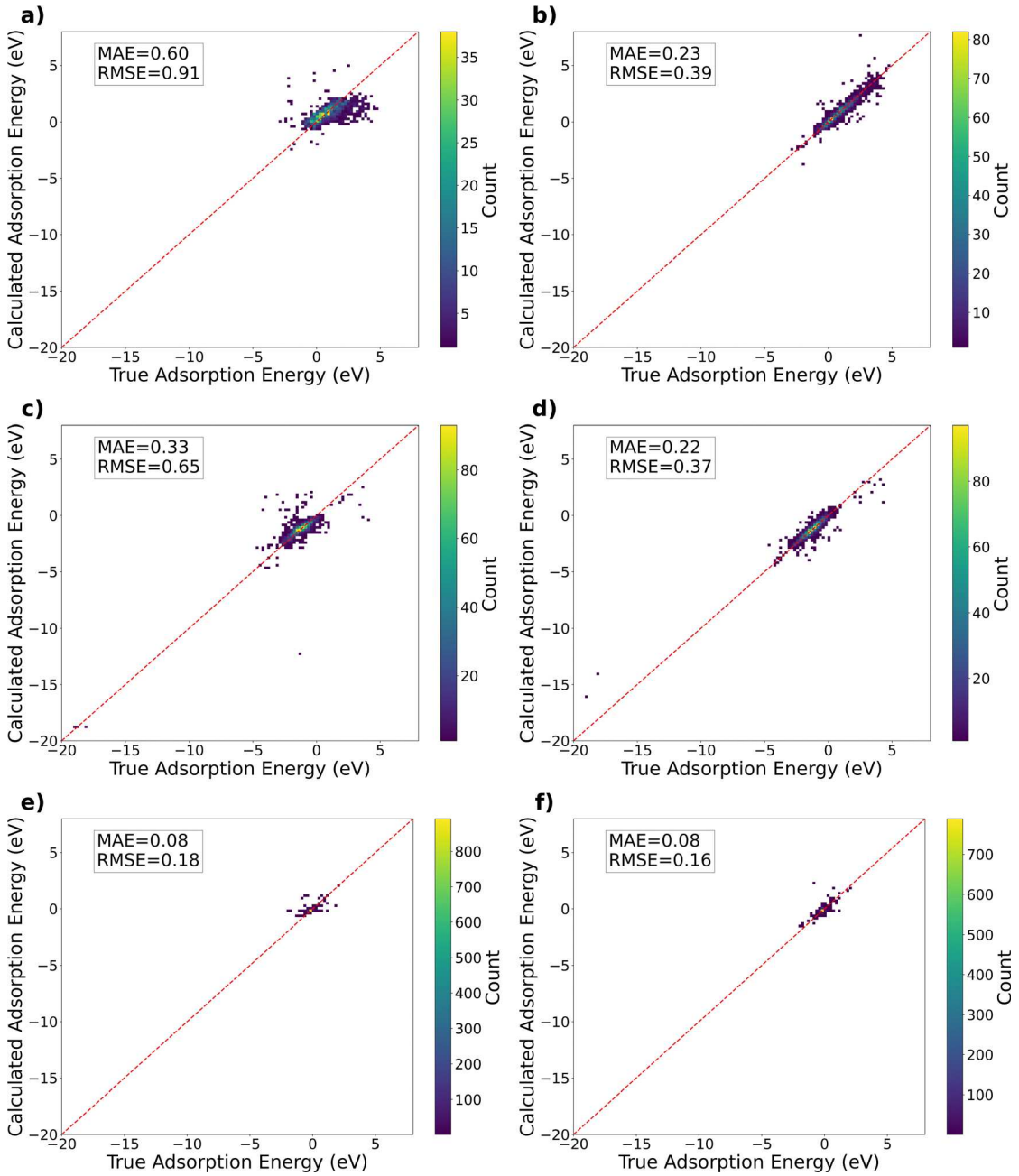

Fig. 3 Calculated adsorption energy *vs.* true adsorption energy with 1250 XH_n configurations in the training and validation dataset for (a) NH without and (b) with N and H adsorption configurations in the training and validation dataset, (c) OH without and (d) with O and H adsorption configurations in the training and validation dataset, and (e) H_2O without and (f) with O and H adsorption configurations in the training and validation dataset.

To further analyze our results, we now focus on the models trained with 1250 molecular adsorption configurations in the training and validation dataset. We display in Fig. 3 the predicted *versus* the true adsorption energies for NH, OH, and H_2O for models trained with 1250 molecular configurations without and with the atomic configurations in the training and validation dataset. All panels are plotted with the same scales for comparison. A zoomed in version of Fig. 3 is available in the ESI† as Fig. S1. From Fig. 3 we note many outliers for the models trained without the atomic adsorption data, reflected in a larger RMSE value. This is true for every molecular species we tried to predict on, including H_2O, demonstrating the benefit of including atomic species adsorption data into your training and validation dataset. The decrease in MAE and RMSE is largest again for NH where we note a 0.37 eV and a 0.52 eV decrease respectively. Fig. 3 also illustrates for H_2O, the molecule bonds relatively weakly to the substrates, giving a relatively large cluster of configurations with adsorption energy around 0.0 eV. Given this pattern, one would expect the neural network to perform better with a large clump of values around zero. Indeed, the MAE for the H_2O model's predictions was 0.08 eV while the MAE was 0.23 eV and 0.22 eV for the NH and OH models, respectively. Moreover, as previously stated, the fact that all models perform very similarly for H_2O suggests we are nearly at the limit of accuracy of HIP-NN, which appears to be 0.05 eV to 0.10 eV, in agreement with our previously published report.[17] Overall, we see that adding the atomic adsorption sites lowers the MAE and/or the RMSE for HIP-NN's predictions on NH, OH, and H_2O.

We note the exact same trend for CH, CH_2, and CH_3. Fig. 4 plots the HIP-NN calculated adsorption energy *versus* the true adsorption energy for CH, CH_2, and CH_3 adsorbed on the various metallic surfaces in our test dataset for both models with and without the atomic adsoption configurations in the training and validation datasets. The results in Fig. 4 are for the models trained with 1250 molecular configurations. From Fig. 4 we see that adding the atomic adsorption configurations decreases both the MAE and RMSE of our models' predictions. This improvement is largest for CH_2 where we note a 0.12 eV decrease in the MAE and a 0.17 eV decrease in RMSE going from a model without atomic adsorption configurations in the training and validation dataset to a model with atomic adsorption configurations in the training and validation dataset. The results for CH_x molecules reinforces this insight that one should always include the atomic adsorption configurations whenever they are available. Including these configurations drastically improves the fully trained model's performance, as made evident by lower errors and fewer outliers, while adding a small amount of computational time to the training procedure. Finally, looking specifically by molecular species, we see HIP-NN performs roughly about equally well for CH, CH_2, and CH_3 with the MAE for all three models within 0.07 eV. Despite all being relatively close, CH_2 performs the worst of any of the 1250 configuration models trained with atomic adsorption site configurations. For this subset of models, neglecting H_2O which we have already discussed, CH_2 predictions yield a MAE of 0.29 eV compared to 0.22 eV to 0.25 eV for the other four adsorbates. There are a variety of reasons this can occur, but it is likely that it is simply harder for the network to predict on CH_2. HIP-NN gathers information about local atom pairs. The specific CH_2 geometry and symmetry may make it harder for HIP-NN to learn the adsorption energy. Weighting the CH_2 configurations in the training dataset partially solves this problem reducing the MAE by 0.05 eV, making the prediction accuracy closer to that of CH. It also reinforces the idea that there is something particular about CH_2 adsorption that the network needs to see more of during training. It is not surprising though that HIP-NN performs slightly worse on one molecule than some of the others. Many other ML reports have found a larger than 0.05 eV spread in error when they apply their specific ML algorithm to predict the adsorption energy of different adsoprbates.[2,4,15,17]

Taking the results from Fig. 3 and 4 together we see that including molecular adsorption configurations into your training and validation dataset can greatly improve ML predictions at a negligible computational cost. Excluding H_2O, we see MAE and RMSE values ranging from 0.31 eV to 0.60 eV and 0.46 eV to 0.91 eV, respectively, for models trained with a 1250 molecular adsorption configuration dataset without atomic adsorption configurations. This drops to a MAE and RMSE ranging from 0.22 eV to 0.29 eV and 0.30 eV to 0.43 eV, respectively, after adding the atomic adsorption configurations into the training dataset. Besides H_2O, which HIP-NN makes better predictions for, NH is also an outlier, where HIP-NN makes worse predictions when only trained on the molecular adsorption configurations. This large MAE and RMSE of 0.60 eV and 0.91 eV are most likely caused by the small, randomly selected training dataset, missing important configurations. To examine this possibility, we looked at the standard deviation (SD) of the adsorption energy of the 1250 molecular adsorption configuration training and testing dataset. For NH we note a SD of the training dataset of 2.00 eV while for OH, CH, CH_2, and CH_3 we note a SD in the training dataset of 3.36 eV, 2.74 eV, 2.75 eV, and 2.76 eV. This relatively low SD of NH likely means our random sampling is missing configurations of a particular bonding strength. This explanation looks more likely as we calculate a SD of 1.89 eV, 1.71 eV, 2.13 eV, 1.39 eV, and 0.85 eV for the testing dataset of NH, OH, CH, CH_2, and CH_3. NH, by far, has the lowest SD for the training dataset, but for the testing dataset NH possesses the second largest SD. This inadequacy in sampling is overcome by adding the atomic configurations into the training dataset, reducing the MAE to 0.23 eV and RMSE to 0.39 eV, again further illustrating the power of adding atomic adsorption configurations to one's training and validation dataset.

In addition to demonstrating the importance of including the atomic adsorption configurations, our neural network predictions share a similar accuracy to other state of the art ML methods[2,4,5,15–17] which usually have a MAE of adsorption energy predictions around 0.1 eV to 0.3 eV, depending on the species studied and the methods employed. For example, Li *et al.* predicted OH adsorption energies with a RMSE of

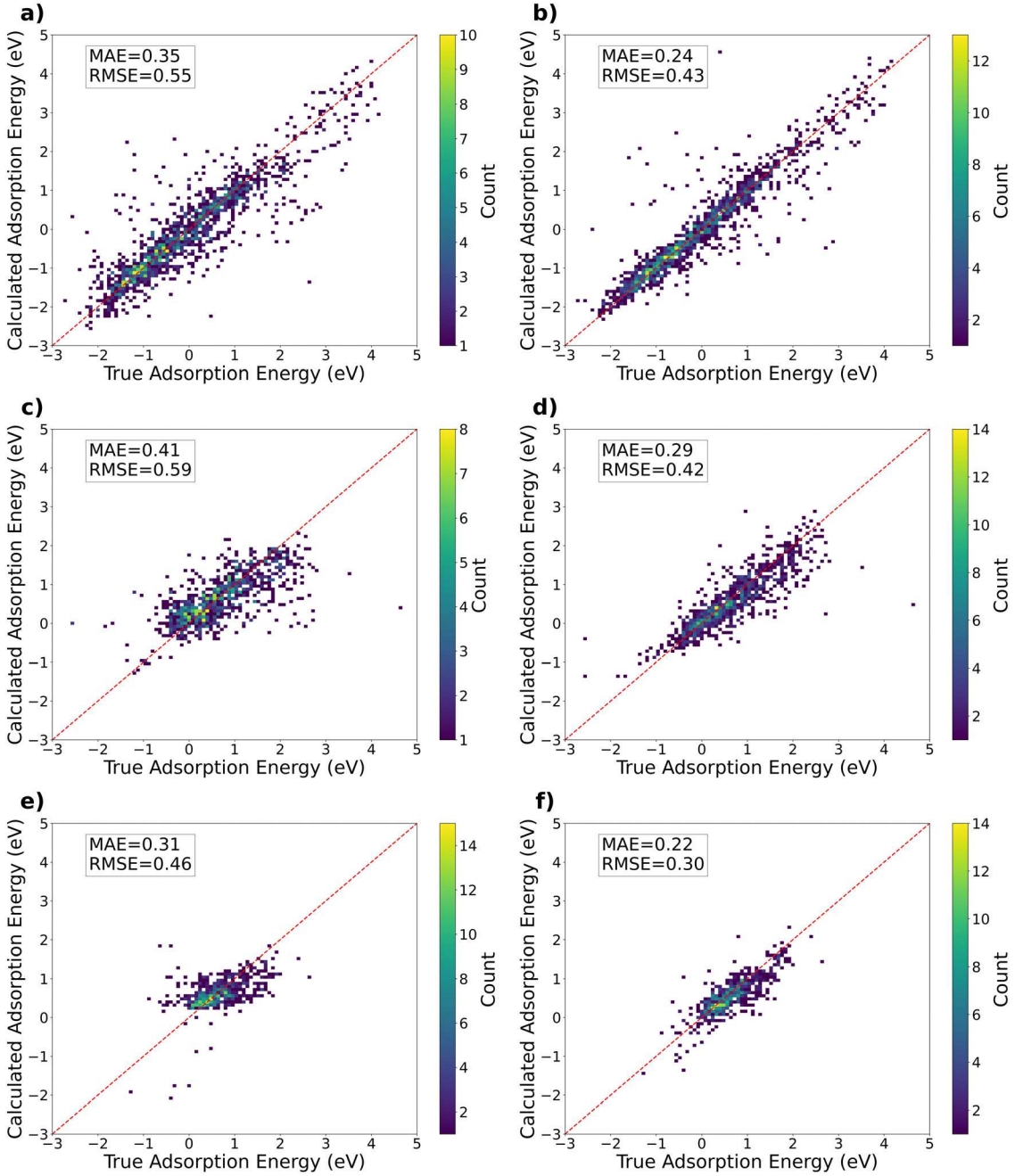

Fig. 4 Calculated adsorption energy *vs.* true adsorption energy with 1250 CH_n configurations in the training and validation dataset for (a) CH without and (b) with C and H adsorption configurations in the training and validation dataset, (c) CH_2 without and (d) with C and H adsorption configurations in the training and validation dataset, and (e) CH_3 without and (f) with C and H adsorption configurations in the training and validation dataset.

0.24 eV. However, they utilized many inputs for their neural network including information about the d-band of the adsorption site and local electronegativity, among others.[16] In this report, we simply use atomic species and the atomic positions of the starting adsorption configurations. In another study, Nayak *et al.* predicted atomic adsorption energies, H, C, and N, and molecular adsorption energies, OH, CO, and NO, over twenty-five transition metal surfaces using ML with a RMSE of 0.41 eV and 0.26 eV respectively. They utilized a much smaller dataset, about 200 calculations, but many more descriptors to train their model: 7 features for the atomic adsorption configurations and 8 for the molecular adsorption configurations. They also only studied single metal surfaces.[15] As in this study, Tran *et al.* studied surfaces with up to two elements and recorded a MAE of 0.29 eV for CO adsorption energies and 0.24 eV for H adsorption energies for their ML predictions.[2] In another ML study, Back *et al.* predicted the adsorption energies of OH and H both to 0.15 eV on materials consisting up to 4 elements chosen from a total of 37 elements.[4] In a different study, Fung *et al.* predicted H, CH, CH_2, CH_3, NH, and OH, among other adsorbates, adsorption

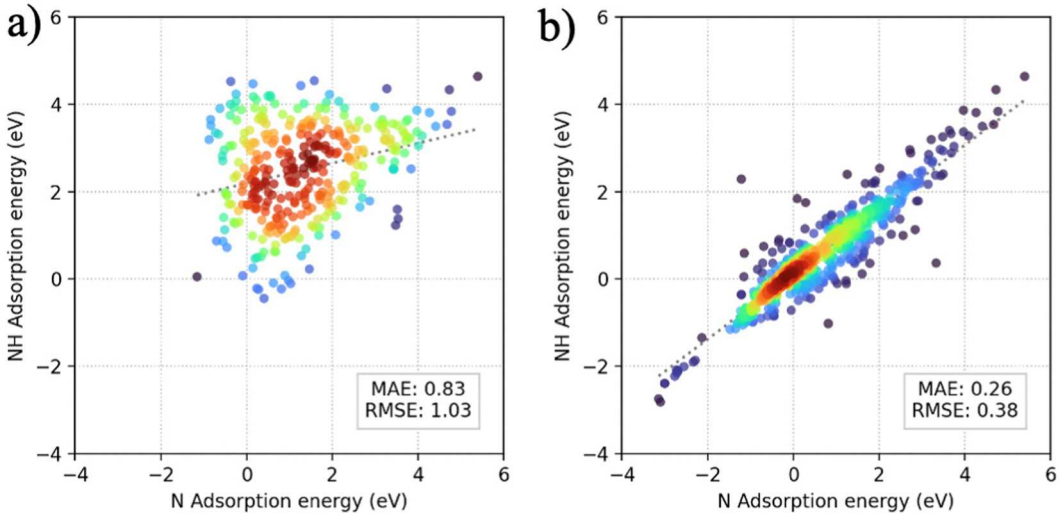

Fig. 5 (a) NH adsorption energy *vs.* N adsorption energy for the most stable adsorption configuration on each surface, and (b) NH adsorption energy *vs.* N adsorption energy for each adsorption site on each surface. A line of best fit is provided for each plot which would represent the scaling relationship.

energies on bimetallic surfaces. They found adsorption energies with a MAE of 0.07 eV, 0.15 eV, 0.13 eV, 0.11 eV, 0.16 eV and 0.16 eV for H, CH, CH_2, CH_3, NH, and OH, respectively.[5] While their predictions slightly outclass HIP-NN's predictions some considerations must be made. They trained using features from the density of states. HIP-NN needs only an initial geometry to make a prediction, saving valuable computational time. Fung also noted that they needed about ~2700 calculations in their H training dataset to achieve convergence with respect to testing error.[5] Our largest dataset, NH, had 2310 datapoints in the testing dataset, indicating with more data our results would likely improve too. They also noted, like this report, a small decrease in MAE when including other adsorbates in their H training dataset.[5] They did not investigate this further with other adsorbates. Taken together we see that our results' accuracy is within the accuracies reported by others, MAEs anywhere from 0.07 eV to 0.29 eV and RMSEs anywhere from 0.24 eV to 0.41 eV, using ML to predict the adsorption energies of small molecules and atoms on metallic surfaces. HIP-NN performs quite well considering the relatively small sizes of the datasets. Moreover, while HIP-NN has not achieved the accuracy of DFT, approximately 0.13 eV or roughly 3 kcal mol^{-1}, except for H_2O, adsorption energies predicted with a MAE of 0.22 eV to 0.29 eV is suitable for search space reduction, saving valuable computational time and achieving the end goal of using ML for material discovery.[49]

(B) Traditional scaling relationships

To assess the relative effectiveness of the neural network, we compare our ML results with traditional scaling relationships. Fig. 5(a) plots the adsorption energy of NH *versus* the adsorption energy of N for the highest, most stable, adsorption energy configuration on each surface, and the best fit linear line for this relationship. From Fig. 5(a) we note that a scaling relationship between the highest NH and H adsorption energies generally fails with a MAE of 0.83 eV from the best fit line. This failure is a result of the most stable adsorption site sometimes being different for N and NH, and different adsorption sites possess different coordination numbers. However, scaling relationships rely on the coordination number of the surface being the same from one adsorbate to another, highlighting a weakness of the method. Successful attempts have been made to build a more complicated scaling relationship for a specific adsorption site from a different adsorption site.[30] Considering adsorption site, in Fig. 5(b) we plot the adsorption energy of NH *vs.* N for identical adsorption sites and a linear best fit line. From Fig. 5(b) we note a MAE of 0.26 eV, a huge improvement from plotting the most stable adsorption sites and utilizing the best fit line. A MAE of 0.26 eV for NH sits in between the MAE for the ML without the N adsorption sites in the training and validation dataset, 0.60 eV, and the MAE for the ML predictions with the N adsorption sites in the training and validation dataset, 0.23 eV. This illustrates that ML may have the edge over traditional scaling relationships if sufficent data is available. Indeed, in a recent study, Vijay *et al.* pointed out several weaknesses of scaling relationships. Scaling relationships may perform poorly for surfaces of noble metals and for surfaces with metals that do not belong to the same row in the periodic table as these different metals may lead to qualitatively different normalized adsorbate valence energies. Other problems with scaling relationships include issues with larger molecules where the greater degree of freedom of the adsorbate may break the relationship, and higher coverages where adsorbate–substrate interactions could also alter or break simple scaling relationships.[38] Several of the aftermentioned issues likely are contributing to the scatter in Fig. 5(b).

4. Conclusions

Overall, we notice that ML can compete and even surpass simple scaling relationships for small molecules and atoms adsorbed on transition metal surfaces. ML, perhaps unsurprisingly,

performs better as we increase the amount of data we utilize to train the network. With the addition of atomic adsorption configurations, our methods reach a similar accuracy to other ML methods when predicting the adsorption energy of small molecules and atoms on transition metal surfaces. More importantly, we have demonstrated that a neural network's predictions of the adsorption energies of small molecules can be greatly improved by including the atomic adsorption configurations in the training and validation dataset. This improvement manifests itself in smaller MAEs, RMSEs, and fewer outliers for models that have been trained with atomic adsorption configurations in addition to the molecular adsorption configurations. Specifically, we saw that models trained with 250, 500, 750, and 1250 molecular adsorption configurations generally saw an overall reduction in MAE and RMSE when introducing the atomic adsorption configurations into the training and validation dataset. Not only did models trained with atomic adsorption configurations in the training and validation dataset outperform similar models lacking atomic adsorption configurations in the training and validation dataset, but these models, which included the atomic adsorption configurations during training, could reach or even exceed the accuracy of models trained with many more molecular adsorption configurations.

In fact, it appears that including the atomic adsorption configurations in the training and validation dataset is more important than more molecular configurations, as demonstrated by a generally larger decrease in error when including atomic adsorption configurations in the training procedure, compared to increasing the number of molecular configurations in the training procedure from 250 to 1250 adsorption configurations. This suggests that it may be more difficult for the neural network to initially learn the adsorption energies associated with the X–metal or H–metal bond than how a new X–H bond will change the adsorption dynamics of the metal-non-metal bonds.

As a consequence of including atomic adsorption configurations, one will need fewer of the more computationally expensive molecular adsorption configurations in their dataset. This can potentially improve computational efficiency by allowing one to, instead of building an entirely new dataset when one is interested in larger new molecules, build a training dataset of smaller, already available, molecular, or atomic adsorption sites supplemented by a much smaller amount of larger, novel molecular adsorption sites. In principle, this can eliminate a great deal of computational time by finding or reusing datasets of smaller molecules absorbed on the surfaces one is interested in. By placing these simpler adsorbate configurations in their dataset, one will reduce, as we have illustrated, the number of more computationally demanding adsorption calculations with the larger molecules. Finally, we found that weighting the molecular configurations inside the training and validation dataset can further, although to a smaller degree, increase the quality of predictions. As such, the concept may be worth exploring in future high-throughput ML surface studies. Especially when building a diverse training dataset, it may be valuable to examine underrepresented configurations and modify their weight within the entire dataset.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Author contributions

Walter Malone: conceptualization (lead); methodology (lead); software (lead); writing – original draft (lead); writing – review & editing (equal); visualization (equal). Johnathan van der Hyde: writing – original draft (support); writing – review & editing (equal); visualization (equal). Abdelkader Kara: methodology (supporting); writing – original draft (supporting); writing – review & editing (equal).

Conflicts of interest

The authors have no conflicts to disclose.

Acknowledgements

Walter Malone acknowledges support from the National Science Foundation: National Science Foundation RISE grant #2122985. This work was made possible in part by a grant of high-performance computing resources and technical support from the Alabama Supercomputer Authority.

Notes and references

1 O. Mamun, K. T. Winther, J. R. Boes and T. Bligaard, *npj Comput. Mater.*, 2020, **6**, 177.
2 K. Tran and Z. W. Ulissi, *Nat. Catal.*, 2018, **1**, 696–703.
3 J. Noh, S. Back, J. Kim and Y. Jung, *Chem. Sci.*, 2018, **9**, 5152–5159.
4 S. Back, J. Yoon, N. Tian, W. Zhong, K. Tran and Z. W. Ulissi, *J. Phys. Chem. Lett.*, 2019, **10**, 4401–4408.
5 V. Fung, G. Hu, P. Ganesh and B. G. Sumpter, *Nat. Commun.*, 2021, **12**, 88.
6 D. Roy, S. C. Mandal and B. Pathak, *J. Phys. Chem. Lett.*, 2022, **13**, 5991–6002.
7 E. Hu, C. Liu, W. Zhang and Q. Yan, *J. Phys. Chem. C*, 2023, **127**, 882–893.
8 B. Meyer, B. Sawatlon, S. Heinen, O. A. von Lilienfeld and C. Corminboeuf, *Chem. Sci.*, 2018, **9**, 7069–7077.
9 Z. W. Ulissi, M. T. Tang, J. Xiao, X. Liu, D. A. Torelli, M. Karamad, K. Cummins, C. Hahn, N. S. Lewis, T. F. Jaramillo, K. Chan and J. K. Nørskov, *ACS Catal.*, 2017, **7**, 6600–6608.
10 X. Ma, Z. Li, L. E. K. Achenie and H. Xin, *J. Phys. Chem. Lett.*, 2015, **6**, 3528–3533.
11 G. H. Gu, J. Noh, S. Kim, S. Back, Z. Ulissi and Y. Jung, *J. Phys. Chem. Lett.*, 2020, **11**, 3185–3191.
12 X. Li, W. Paier and J. Paier, *Front. Chem.*, 2020, **8**, 601029.

13 O. Piqué, I. Z. Koleva, A. Bruix, F. Viñes, H. A. Aleksandrov, G. N. Vayssilov and F. Illas, *ACS Catal.*, 2022, **12**, 9256–9269.
14 X. Zong and D. G. Vlachos, *J. Chem. Inf. Model.*, 2022, **62**, 4361–4368.
15 S. Nayak, S. Bhattacharjee, J.-H. Choi and S. C. Lee, *J. Phys. Chem. A*, 2020, **124**, 247–254.
16 Z. Li, S. Wang, W. S. Chin, L. E. Achenie and H. Xin, *J. Mater. Chem. A*, 2017, **5**, 24131–24138.
17 W. Malone and A. Kara, *Surf. Sci.*, 2023, **731**, 122252.
18 A. Mazheika, Y.-G. Wang, R. Valero, F. Viñes, F. Illas, L. M. Ghiringhelli, S. V. Levchenko and M. Scheffler, *Nat. Commun.*, 2022, **13**, 419.
19 M. Andersen, S. V. Levchenko, M. Scheffler and K. Reuter, *ACS Catal.*, 2019, **9**, 2752–2759.
20 L. M. Ghiringhelli, J. Vybiral, S. V. Levchenko, C. Draxl and M. Scheffler, *Phys. Rev. Lett.*, 2015, **114**, 105503.
21 W. Malone, W. E. Kaden and A. Kara, *Catal. Lett.*, 2019, **149**, 2953–2960.
22 Y. Huang, H. Yang, T. Xiong, D. Adekoya, W. Qiu, Z. Wang, S. Zhang and M.-S. Balogun, *Energy Storage Mater.*, 2020, **25**, 41–51.
23 H.-W. Cheng, P. Raghunath, K. Wang, P. Cheng, T. Haung, Q. Wu, J. Yuan, Y.-C. Lin, H.-C. Wang, Y. Zou, Z.-K. Wang, M. C. Lin, K.-H. Wei and Y. Yang, *Nano Lett.*, 2020, **20**, 715–721.
24 K. T. Winther, M. J. Hoffmann, J. R. Boes, O. Mamun, M. Bajdich and T. Bligaard, *Sci. Data*, 2019, **6**, 75.
25 A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder and K. A. Persson, *APL Mater.*, 2013, **1**, 011002.
26 S. Kirklin, J. E. Saal, B. Meredig, A. Thompson, J. W. Doak, M. Aykol, S. Rühl and C. Wolverton, *npj Comput. Mater.*, 2015, **1**, 15010.
27 D. D. Landis, J. S. Hummelshoj, S. Nestorov, J. Greeley, M. Dulak, T. Bligaard, J. K. Norskov and K. W. Jacobsen, *Comput. Sci. Eng.*, 2012, **14**, 51–57.
28 W. A. Saidi, *npj Comput. Mater.*, 2022, **8**, 86.
29 R. Christensen, H. A. Hansen, C. F. Dickens, J. K. Nørskov and T. Vegge, *J. Phys. Chem. C*, 2016, **120**, 24910–24916.
30 F. Calle-Vallejo, D. Loffreda, M. T. M. Koper and P. Sautet, *Nat. Chem.*, 2015, **7**, 403–410.
31 F. Abild-Pedersen, J. Greeley, F. Studt, J. Rossmeisl, T. R. Munter, P. G. Moses, E. Skúlason, T. Bligaard and J. K. Nørskov, *Phys. Rev. Lett.*, 2007, **99**, 016105.
32 M. M. Montemore and J. W. Medlin, *Catal. Sci. Technol.*, 2014, **4**, 3748–3761.
33 J. Greeley, *Annu. Rev. Chem. Biomol. Eng.*, 2016, **7**, 605–635.
34 M. T. M. Koper, *Chem. Sci.*, 2013, **4**, 2710.
35 Z.-J. Zhao, S. Liu, S. Zha, D. Cheng, F. Studt, G. Henkelman and J. Gong, *Nat. Rev. Mater.*, 2019, **4**, 792–804.
36 I. C. Man, H. Su, F. Calle-Vallejo, H. A. Hansen, J. I. Martínez, N. G. Inoglu, J. Kitchin, T. F. Jaramillo, J. K. Nørskov and J. Rossmeisl, *ChemCatChem*, 2011, **3**, 1159–1165.
37 A. J. Medford, A. Vojvodic, J. S. Hummelshøj, J. Voss, F. Abild-Pedersen, F. Studt, T. Bligaard, A. Nilsson and J. K. Nørskov, *J. Catal.*, 2015, **328**, 36–42.
38 S. Vijay, G. Kastlunger, K. Chan and J. K. Nørskov, *J. Chem. Phys.*, 2022, **156**, 231102.
39 O. Mamun, K. T. Winther, J. R. Boes and T. Bligaard, *Sci. Data*, 2019, **6**, 76.
40 J. Wellendorff, K. T. Lundgaard, A. Møgelhøj, V. Petzold, D. D. Landis, J. K. Nørskov, T. Bligaard and K. W. Jacobsen, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2012, **85**, 235149.
41 A. Tkatchenko, L. Romaner, O. T. Hofmann, E. Zojer, C. Ambrosch-Draxl and M. Scheffler, *MRS Bull.*, 2010, **35**, 435–442.
42 J. Klimeš, D. R. Bowler and A. Michaelides, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2011, **83**, 195131.
43 J. Carrasco, B. Santra, J. Klimeš and A. Michaelides, *Phys. Rev. Lett.*, 2011, **106**, 026101.
44 J. Carrasco, J. Klimeš and A. Michaelides, *J. Chem. Phys.*, 2013, **138**, 024708.
45 J. Klimeš and A. Michaelides, *J. Chem. Phys.*, 2012, **137**, 120901.
46 W. Liu, J. Carrasco, B. Santra, A. Michaelides, M. Scheffler and A. Tkatchenko, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2012, **86**, 245405.
47 W. Malone, J. Matos and A. Kara, *Surf. Sci.*, 2018, **669**, 121–129.
48 N. Lubbers, J. S. Smith and K. Barros, *J. Chem. Phys.*, 2018, **148**, 241715.
49 X. Zhang, A. Chen and Z. Zhou, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2019, **9**, e1385.

