Predicting Intermetallic Surface Energies with High-Throughput DFT and Convolutional Neural Networks

Aini Palizhati,^{†,‡} Wen Zhong,^{†,‡} Kevin Tran,[†] and Zachary W. Ulissi*,[†]

[†]*Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, PA 15217*

[‡]*Contributed equally to this work*

E-mail: zulissi@andrew.cmu.edu

Abstract

Surface energy of inorganic crystals is crucial in understanding experimentally-relevant surface properties and thus important in designing materials for many applications including catalysis. Predictive methods and datasets exist for surface energies of monometallic crystals but predicting these properties for bimetallic or more complicated surfaces is an open challenge. Here we present a workflow for predicting surface energies *ab initio* using high-throughput DFT and a machine learning framework. We calculate the surface energy of 3,285 intermetallic alloys with combinations of 36 elements and 47 space groups. We used this high-throughput workflow to seed a database of surface energies, which we used to train a crystal graph convolutional neural network (CGCNN). The CGCNN model was able to predict surface energies with a mean absolute test error of $0.0082 \text{ eV}/\text{\AA}^2$ and can qualitatively reproduce nanoparticle surface distributions (Wulff constructions). Our workflow provides quantitative insights into which surfaces are more stable and therefore more realistic. It allows us to down-select interesting candidates that we can study with robust theoretical and experimental methods for applications such as catalysts screening and nanomaterials synthesis.

Introduction

Surface energy is a fundamental property in understanding and explaining surface properties and effects, such as surface structures, various phenomena including reconstruction, segregation, catalytic activities, and crystal equilibrium shape.¹⁻³ These surface properties and effects are important in studying and designing new materials for many applications, such as catalysts screening and nanomaterials synthesis.² Catalysts are used in many emerging new applications such as fuel cells and bio-refining, in addition to approximately 90% of the chemical processes.⁴⁻⁶ Surface energy describes the stability of a surface, and surface stability is an important consideration in catalyst screening process.^{6,7} Surface energy also affects the formation or mechanical behavior of nanomaterials such as intermetallic nanoparticles,^{1,8} polymer nanocomposites,⁹ and quantum dots.¹⁰

The determination of surface energy can be achieved using first principle computations such as density functional theory (DFT).¹¹ The standard method in DFT calculations is to evaluate the total energy of a slab and subtract the bulk energy obtained from an individual calculation (Equation 1).

$$\sigma = \lim_{N \rightarrow \infty} \frac{1}{2} (E_{slab}^N - N E_{bulk}) \quad (1)$$

where σ is the surface energy, N is the number of atoms in the slab, E_{slab}^N is the total energy of an N -layer slab, E_{bulk} is the total energy of the bulk, and the $\frac{1}{2}$ accounts for both the top and bottom surface of a slab. However, this method becomes computational divergent as N increases.¹² Fiorentini *et al.* addressed this issue using a linear extrapolation method.¹³ They found that as N became large and convergence was approaching, surface energy can be approximated with

$$\sigma \approx (E_{slab} - n_{slab} E_{bulk}) / (2A_{slab}) \quad (2)$$

where σ is the surface energy, E_{slab} is the energy of the slab itself, n_{slab} is the number of atoms in the slab, E_{bulk} is the energy of the optimized bulk structure per atom, and A_{slab} is the cross-sectional area of the slab's unit cell. The quantity of E_{bulk} and σ can be extracted

by fitting a linear line to slab total-energy data versus n_{slab} with as little as three slabs at different thicknesses, where E_{bulk} is the slope and σ is the intercept of the linear fit. The method resolved the divergence problem because no separately calculated bulk energy is needed.^{13,14}

Extensive DFT calculations have been performed to calculate the surface energy of different metals as a starting point to understand surface phenomena. For example, Yu *et al.* calculated orientation-dependent surface energies of over 30 low-index and vicinal Pb surfaces.¹⁵ Zhang *et al.* determined 36 orientation-dependent surface energies of Ni.¹ Vito *et al.* compiled a database of low index surface energies for 60 metals.¹⁶ More recently, Tran *et al.* presented a large database of surface energies of more than 100 polymorphs of about 70 elements.¹ However, these calculations are largely limited to single-element crystals. This is a major limitation for the large and growing community studying bimetallic nanoparticle synthesis, faceting, and catalysis.

Surface energy calculations have three major obstacles. (1) Even though real surfaces terminate on one side of a semi-infinite bulk, slabs used in DFT calculations are thin-films with two surface terminations. Equation 2 assumes both terminations of a slab are identical and contribute equally to the surface energy. However, this is only true for symmetric slabs. For asymmetric slabs, Equation 2 calculates the "cleavage energy".¹⁷ For monometallic systems, methods exist to modify one surface to obtain symmetry. Applying these methods to bimetallic systems is more difficult as they tend to break stoichiometry and thus require a detailed accounting of the elemental free energy. (2) The choice of exchange-correlation functional and other parameters for DFT can affect the calculated surface energy. They affect the accuracy and convergence of the DFT calculations, result in potential surface reconstructions due to poor initial guess, and introduce variation across research studies.^{2,18} (3) The intensive computational cost caused by the calculation requirements and large exploration space. Performing slab relaxations to extrapolate surface energy requires the use of large supercells in a large vacuum region, especially for surfaces with high Miller indices

or low symmetry.¹ The large exploration space comes from the high number of potential surface orientations. For example: in order to characterize one titanium gold (TiAu) crystal structure, with surfaces limited to a maximum Miller index of 2, we enumerated 19 unique facets. We performed 57 DFT slab relaxations to calculate surface energies for these surfaces, which we then used to create a Wulff structure (**Figure 1**). When we extend the same process to many more intermetallic materials of interests, the number of DFT calculations needed increases drastically.

Figure 1: The large number of surface energy calculations required to understand the surface stability of multi-component materials. For example: given a TiAu crystal structure and a maximum Miller index of 2, we attempted surface energy calculations for 19 unique facets and 17 of them converged. All of these converged energies were needed to construct the Wulff structure, which illustrates the relative stabilities of various surfaces.

The limitations of DFT and large exploration spaces demand accelerated discovery methods, such as a machine learning framework.¹⁹ Most machine learning algorithms require the

size of crystal systems represented as a fixed-length vector.²⁰ This vector usually is constructed using simple material properties^{21,22} or symmetry-invariant atoms coordinates.^{23,24} However, these methods either require case-by-case design or generate models that are hard to interpret. Xie and Grossman developed a generalized crystal graph convolutional neural networks (CGCNN) framework to represent periodic crystal systems. It converts atomic structures into a graph representation that entails both atomic information and bond interactions. Then it builds a convolutional neural network on top of the graph representation to provide high accuracy prediction of material property and atomic level knowledge extraction.²⁰ These methods have not yet been applied to various surface phenomena.

In this work, we present a workflow to predict surface energy combining high-throughput DFT and CGCNN. We calculated the surface energy of 3,285 different intermetallic surfaces across 36 elements and 47 space groups. This data served as the foundation of a growing surface energy database. With this database, we trained a CGCNN model to predict surface energies from initial, unrelaxed atomic structures. The CGCNN model is an effective approach to reduce the number of DFT jobs in large exploration spaces required for applications such as high-throughput catalyst screening or equilibrium shape construction for nanoparticles.

Materials and Methods

Slab Generation and DFT Relaxation

A list of stable crystal bulk structures was compiled from the Materials Projects²⁵ and relaxed using DFT. Surfaces were enumerated from relaxed bulk structure using pymatgen.²⁶ Each surface was characterized by three features: Materials Project ID number (MPID), Miller index, and shift (i.e., location of surface termination). These surfaces were limited to a maximum Miller index of 2 and a maximum of three elements. To calculate surface energies, we used slab structures of three different thickness for the linear extrapolation method. We

first generated base slab structures with a minimum thickness of 7 Å, calculated the height of the smallest unit slabs, and added one and two layers of the unit slabs on top of the base slabs to generate two other structures. All slabs were created with a vacuum of at least 20 Å and were not constrained to have symmetric top and bottom surfaces. These slabs were relaxed using DFT. All DFT calculations (of bulk structures and slabs) were performed using our open-source Generalized Adsorption Simulator for Python (*GASpy*).²⁷ DFT calculations were performed in VASP with the PBEsol functional (for full details see supplementary document DFT calculation parameters section).

Surface Energy Calculation

The surface energy of a given surface was obtained using the linear extrapolation method.^{13,14} For each surface, the total-energy data of three slabs versus the number of atoms were fit into a linear correlation using statsmodels²⁸ ordinary least square model. The intercept of the linear correlation was the intrinsic surface energy (**Figure 2.a**). The standard error of the linear fit was also recorded. Energies higher than $0.03 \text{ eV}/\text{\AA}^2$, with standard errors (from linear extrapolation) higher than $0.01 \text{ eV}/\text{\AA}^2$, or surfaces with atomic movements higher than 1 Å were considered ill-converged surface energies and were filtered from the final dataset.

Crystal Graph Convolutional Neural Network

We applied CGCNN on atomic structures to predict surface energies (**Figure 2.b**). Initial atomic configuration of surfaces (before DFT relaxation) was used as input, and the natural logarithm of surface energies was used as prediction labels. The logarithm transformation was used to reduce the impact of percentage error of the small values.²⁹ The mean absolute error (MAE) of the residuals were calculated with surface energies to evaluate prediction accuracy. In addition, we modified CGCNN to visualize the contribution of each atom to the surface energy prediction. We indicated the degree of contribution by the opacity of the atoms (i.e., higher contribution corresponds to higher opacity).³⁰

a) Surface Energy Extrapolation

b)

Figure 2: **Methods for generating a) DFT-calculated surface energies and b) CGCNN predictions.** a) First, three slabs with different thicknesses were generated for each surface pattern. DFT was then used to calculate the energies of each of these slabs, which were used to calculate surface energies via linear extrapolation (Equation 2¹⁴). b) For the CGCNN method,²⁰ a crystal graph was generated using the thinnest slab of each surface pattern. The graph was input to a convolutional neural network, where the surface energy prediction was provided as an output.

Two different train-test split methods were used during training: random assignment and leave-one-out. The random assignment method split 80% of the DFT data into a training set and 20% into a test set. It demonstrated the overall prediction accuracy of the CGCNN model. The leave-one-out method used the DFT results of a composition of interest as the test set and the remaining data as the training set. It tested whether CGCNN can predict surface energies and Wulff structures for intermetallic combinations not in the training data set, and thus demonstrated CGCNN's prediction ability on new, unseen materials.

Hyperparameter Optimization

Deep learning models such as CGCNN are difficult to optimize because they contain large numbers of hyperparameters. Tuning a model using trial and error is manually intensive and time-consuming. We used SigOpt,³¹ which is an end-to-end optimization package that we configured to perform this hyperparameter optimization to minimize MAE between DFT-calculated surface energies and CGCNN-predicted surface energies. The list of hyperparameters and optimized values are summarized in the supplementary document (Table S1).

Wulff Construction

Theoretical simulation based on Wulff construction characterizes the equilibrium shape of the nanoparticle given the surface energies of a material. Wulff constructions are used widely to explain experimental findings and lead to the design of better materials with tailored properties.^{1,3} In this work, Wulff structures were constructed using pymatgen’s²⁶ Wulff module. This module defines a Wulff shape from a lattice, a list of indices, and their corresponding surface energies.² For a material of interest, two Wulff structures were constructed, where one used DFT-calculated data and the other used CGCNN-predicted data, both using the lowest surface energy of each Miller index. For each Miller index, the ratio of its surface area to the total surface area of the Wulff structure (area fraction) was recorded. The MAE of area fractions between DFT-calculated and CGCNN-predicted Wulff structures were calculated to quantify the accuracy of CGCNN predictions.

Results

Surface Energy Database

We created a surface energy database comprising 3,285 different intermetallic surfaces across 36 elements and 47 space groups. Of which, 72% of surfaces are bimetallic, 25% of surfaces

are trimetallic, and 3% are monometallic. The surface energies are between 0.02 and 0.3 eV/Å², with an average standard error of 0.00095 eV/Å². The maximum atomic movement of surfaces is less than 1 Å. The force exerted on each atom of surfaces is lower than 0.05 eV/Å for 3255 (out of 3285) surfaces, and 30 (out of 3285) surfaces have moderate residual forces (i.e. between 0.05 eV/Å and 0.3 eV/Å). In addition, we compared our DFT calculated surface energies with other literature surface energies results on some of the elemental compounds. The comparison is included in the supplementary document (Table S2).

CGCNN Analysis

For the random assignment method (80:20 train-test split), we achieved a training error of 0.0048 eV/Å² (MAE) and 0.008 eV/Å² (RMSE), and a test error of 0.0082 eV/Å² (MAE) and 0.016 eV/Å² (RMSE)(**Figure 3.a**). A learning curve was generated for the CGCNN model learning performance over the number of DFT training data (**Figure 3.b**). For each number of DFT calculations, we obtained an error estimate using a 5-fold cross-validation, and computed the mean and standard deviation of the estimates. The mean absolute errors are represented by the points, and the standard deviations are represented by the shaded region that bounds the curve.

Figure 3.c-d are visualizations of atoms’ contribution to the total surface energy for a nickel aluminum slab. As shown in the figure, aluminum atoms on the top of the slab are most opaque, and therefore have the highest contribution to the surface energy prediction. This result makes physical sense because surface energy is the excess energy of surface atoms due to broken bonds yielding under-coordinated atoms.²

A similar analysis was performed for CGCNN where training data was selected using the leave-one-out method, which demonstrated CGCNN’s prediction ability on new intermetallic combinations not seen in the training data. **Figure 4**. For example, we created one CGCNN model where we used all our DFT results as training, except we left out all nickel gallium

Figure 3: **CGCNN analysis using 80:20 train, test split.** (a) Parity plot of CGCNN predictions. (b) The learning curve of the CGCNN model over the number of DFT training data, where the training error is indicated by orange and the test error is indicated by blue. c-d) Visualizations of atoms' contribution to the total surface energy for a nickel (green) and aluminum (grey) surface. c) is the side view and d) is the top view. Higher opacity corresponds to higher contributions towards surface energy.

(Ni_xGa_y) results, which is a known catalyst for CO₂ reduction.³² We repeated this again using copper aluminum (Cu_xAl_y)²⁷ as the leave-one-out set. Table 1 shows the CGCNN performance of these two models.

For the leave-one-out method, in addition to the parity plots, Wulff structures were constructed to characterize relative surface stabilities and equilibrium crystal shapes. We constructed one set of Wulff structures using only DFT results and one set of structures using only CGCNN's predictions (**Figure 5**). The MAE of the area fraction for each surface is also reported in **Figure 5**. Wulff structures created from CGCNN-predicted data appear to provide reasonable accuracy, and the top facets in the Wulff structures created from DFT data match the top facets created from CGCNN predictions.

Figure 4: **CGCNN analysis using a composition of interest as the test data.** The surface energies of a composition of interest were left out as the test set, and the remaining data was used as the training set for CGCNN. The parity plots shown here were CGCNN prediction vs. DFT data that used (a) Ni_xGa_y as the test set, and (b) Cu_xAl_y as the test set.

Discussion

We presented a framework of predicting surface energy combining high-throughput DFT and a machine learning method. The database we created has an extensive collection of calculated surface energies for intermetallic crystal surfaces that we have not yet seen in other literature work. It contains over 3,000 DFT calculated surface energies of different intermetallic alloys across 36 elements and 47 space groups. The full dataset in easy-to-read

Table 1: **CGCNN prediction results for leave-one-out method.**

	Ni_xGa_y	Cu_xAl_y
Number of Test Data	30	68
Train MAE ($\text{eV}/\text{\AA}^2$)	0.0051	0.0055
Train RMSE ($\text{eV}/\text{\AA}^2$)	0.0008	0.0090
Test MAE ($\text{eV}/\text{\AA}^2$)	0.0070	0.0094
Test RMSE ($\text{eV}/\text{\AA}^2$)	0.0080	0.0120

Figure 5: **Comparison of Wulff structures constructed using DFT-calculated data and CGCNN-predicted data.** For each crystal composition, Wulff structures were constructed using both DFT-calculated data and CGCNN-predicted data for all facets up to Miller index of 2. The top three facets in each structure were labeled (facets in the dashed box means that facet is also one of the top facets, but was not shown in the figure). The MAE of surface area fraction difference of each facet between DFT Wulff structure and CGCNN Wulff structure were also recorded.

and well-documented format is included in the Github repository below. We hope that this open dataset will encourage other researchers to develop even more predictive models for this important surface property.

The CGCNN framework trained on this dataset provides a reasonably accurate estimation of DFT-calculated surface energy and suggests it can be used to down-select candidates for DFT calculations. Combination of high-throughput DFT and machine learning can be automated to develop a dynamic workflow. This dynamic workflow will enable the prediction of all potentially relevant surface energies of a material of interest. These predictions can be

used to reduce the number of DFT calculations needed and provide insights into surface phenomena. This can be an order of magnitude more efficient than calculating surface energies using the traditional trial-and-error DFT calculations, which will have a meaningful impact on nanoparticle synthesis and high-throughput catalysts screening. We have demonstrated similar approaches for catalyst thermochemistry³⁰ and for the prediction of oxide material termination.³³

We highlight several challenges using high throughput workflows for these systems. First, in performing automated calculations some percentage of the calculations will fail due to various DFT convergence issues, reconstructions, or other problems. These failures, although removed in the dataset in this work, could bias the remaining dataset. Second, it is still an open challenge on how to treat surface energies of inherently asymmetric slabs as discussed above. However, the demonstrated approach and models are sufficient to act as a surrogate for detailed theoretical/experimental studies. This framework extracts knowledge from machine learning prediction and down-select candidates in large search spaces so that we can further investigate interesting materials with more robust studies.

Acknowledgement

This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. The Titan Xp used for this research was donated by the NVIDIA Corporation. This work was also supported by Carnegie Mellon University chemical engineering departmental PPG fellowship for 2017-2018, courtesy of the PPG Foundation.

Supporting Information Available

DFT calculation parameters details. The optimized hyperparameters of the crystal graph

convolutional neural networks. A summarized table of comparison of our DFT results and previous DFT calculations.

Code and Data availability

The code and data used to produce the figures in this article are available in the Surface_Energy_Manuscript repository at “https://github.com/ulissigroup/Surface_Energy_Manuscript”.

References

- (1) Zhang, W. B.; Chen, C.; Zhang, S. Y. Equilibrium Crystal Shape of Ni from First Principles. *The Journal of Physical Chemistry* **2013**, *117*, 21274–21280.
- (2) Tran, R.; Xu, Z.; Radhakrishnan, B.; Winston, D.; Sun, W.; Persson, K. A.; Ong, S. P. Surface Energies of Elemental Crystals. *Scientific Data* **2016**, *3*, 160080.
- (3) Barmparis, G. D.; Lodziana, Z.; Lopez, N.; Remediakis, I. N. Nanoparticle Shapes by Using Wulff Constructions and First-Principles Calculations. *Beilstein Journal of Nanotechnology* **2015**, *6*, 361–368.
- (4) Nørskov, J. K.; Studt, F.; Abild-Pedersen, F.; Bligaard, T. *Angewandte Chemie International Edition*; John Wiley & Sons, Inc., 2015.
- (5) Edenhofer, O.; Madrugá, R. P.; Sokona, Y. *Climate Policy*, 1st ed.; Cambridge University Press, 2012; p 1088.
- (6) Seh, Z. W.; Kibsgaard, J.; Dickens, C. F.; Chorkendorff, I.; Nørskov, J. K.; Jaramillo, T. F. Combining Theory and Experiment in Electrocatalysis: Insights into Materials Design. *Science* **2017**, *355*.
- (7) Medford, A. J.; Ross Kunz, M.; Ewing, S. M.; Borders, T.; Fushimi, R. Extracting Knowledge from Data through Catalysis Informatics. *ACS Catal.* **2018**, *8*.
- (8) Yang, T.; Zhao, Y. L.; Tong, Y.; Jiao, Z. B.; Wei, J.; Cai, J. X.; Han, X. D.; Chen, D.; Hu, A.; Kai, J. J.; Lu, K.; Liu, Y.; Liu, C. T. Multicomponent Intermetallic Nanoparticles and Superb Mechanical Behaviors of Complex Alloys. *Science* **2018**, *362*, 933–937.
- (9) Khoshkava, V.; Kamal, M. R. Effect of Surface Energy on Dispersion and Mechanical Properties of Polymer/Nanocrystalline Cellulose Nanocomposites. *Biomacromolecules* **2013**, *14*, 3155–3163.

- (10) Long, F.; Gill, S. P. A.; Cocks, A. C. F. Effect of Surface-Energy Anisotropy on the Kinetics of Quantum Dot Formation. *Phys. Rev. B* **2001**, *64*, 121307.
- (11) Harrison, M. J.; Woodruff, D. P.; Robinson, J. Adsorbate-Induced Surface Reconstruction and Surface-Stress Changes in Cu100/O: Experiment and Theory. *Physical Review B* **2006**, *74*, 165402.
- (12) Boettger, J. C. Nonconvergence of Surface Energies Obtained from Thin-Film Calculations. *Phys. Rev. B* **1994**, *49*, 16798–16800.
- (13) Fiorentini, V.; Methfessel, M. Extracting Convergent Surface Energies from Slab Calculations. *Journal of Physics: Condensed Matter* **1996**, *8*, 6525–6529.
- (14) Tang, M. T.; Ulissi, Z. W.; Chan, K. Theoretical Investigations of Transition Metal Surface Energies under Lattice Strain and CO Environment. *J. Phys. Chem. C* **2018**, *122*, 14481–14487.
- (15) Yu, D. K.; Bonzel, H. P.; Scheffler, M. The Stability of Vicinal Surfaces and the Equilibrium Crystal Shape of Pb by First Principles Theory. *New Journal of Physics* **2006**, *8*, 65.
- (16) Vitos, L.; Ruban, A.; Skriver, H.; Kollár, J. The Surface Energy of Metals. *Surface Science* **1998**, *411*, 186–202.
- (17) Sun, W.; Ceder, G. Efficient Creation and Convergence of Surface Slabs. *Surface Science* **2013**, *617*, 53–59.
- (18) Stekolnikov, A. A.; Furthmüller, J.; Bechstedt, F. Absolute Surface Energies of Group-IV Semiconductors: Dependence on Orientation and Reconstruction. *Physical Review B* **2002**, *65*, 115318.
- (19) Nandy, A.; Duan, C.; Janet, J. P.; Gugler, S.; Kulik, H. J. Strategies and Software for

- Machine Learning Accelerated Discovery in Transition Metal Chemistry. *Industrial Engineering Chemistry Research* **2018**, *57*, 13973–13986.
- (20) Xie, T.; Grossman, J. C. Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties. *Physical Review Letters* **2018**, *120*.
- (21) Ghiringhelli, L. M.; Vybiral, J.; Levchenko, S. V.; Draxl, C.; Scheffler, M. Big Data of Materials Science: Critical Role of the Descriptor. *Phys. Rev. Lett.* **2015**, *114*, 105503.
- (22) Isayev, O.; Fourches, D.; Muratov, E. N.; Oses, C.; Rasch, K.; Tropsha, A.; Curtarolo, S. Materials Cartography: Representing and Mining Materials Space Using Structural and Electronic Fingerprints. *Chem. Mater.* **2015**, *27*, 735–743.
- (23) Schütt, K. T.; Glawe, H.; Brockherde, F.; Sanna, A.; Müller, K. R.; Gross, E. K. U. How to Represent Crystal Structures for Machine Learning: Towards Fast Prediction of Electronic Properties. *Phys. Rev. B* **2014**, *89*, 205118.
- (24) Seko, A.; Hayashi, H.; Nakayama, K.; Takahashi, A.; Tanaka, I. Representation of Compounds for Machine-Learning Prediction of Physical Properties. *Phys. Rev. B* **2017**, *95*, 144110.
- (25) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K. A. The Materials Project: A Materials Genome Approach to Accelerating Materials Innovation. *APL Materials* **2013**, *1*, 1–11.
- (26) Ping Ong, S.; Davidson Richards, W.; Jain, A.; Hautier, G.; Kocher, M.; Cholia, S.; Gunter, D.; Chevrier, V. L.; Persson, K. A.; Ceder, G. Python Materials Genomics (pymatgen): A robust, Open-Source Python Library for Materials Analysis. *Computational Materials Science* **2013**, *68*, 314–319.

- (27) Tran, K.; Ulissi, Z. W. Active Learning Across Intermetallics to Guide Discovery of Electrocatalysts for CO₂ Reduction and H₂ Evolution. *Nature Catalysis* **2018**, *1*, 696–703.
- (28) Seabold, S.; Perktold, J. Statsmodels: Econometric and statistical modeling with python. 9th Python in Science Conference. 2010.
- (29) Keene, O. N. The Log Transformation Is Special. *Stat Med.* **1995**, *14*, 811–819.
- (30) Back, S.; Yoon, J.; Tian, N.; Zhong, W.; Tran, K.; Ulissi, Z. Convolutional Neural Network of Atomic Surface Structures to Predict Binding Energies for High-Throughput Screening of Catalysts. *ChemRxiv* **2019**, DOI: 10.26434/chemrxiv.8150666.
- (31) SIGOPT. <https://sigopt.com/>, Accessed: 2019-05.
- (32) Ulissi, Z. W.; Tang, M. T.; Xiao, J.; Liu, X.; Torelli, D. A.; Karamad, M.; Cummins, K.; Hahn, C.; Lewis, N. S.; Jaramillo, T. F.; Chan, K.; Nørskov, J. K. Machine-Learning Methods Enable Exhaustive Searches for Active Bimetallic Facets and Reveal Active Site Motifs for CO₂ Reduction. *ACS Catal.* **2017**, *7*, 6600–6608.
- (33) Back, S.; Tran, K.; Ulissi, Z. Towards a design of active oxygen evolution catalysts: Insights from automated density functional theory calculations and machine learning. *ChemRxiv* **2019**, DOI: 10.26434/chemrxiv.7926869.