Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Combining Artificial Intelligence and Physics-based Modeling to Directly Assess Atomic Site Stabilities: From Sub-nanometer Clusters to Extended Surfaces

- Supporting Information -

Philomena Schlexer Lamoureux,[†,‡] Tej S. Choksi,[†,‡] Verena Streibel,[†,‡] and Frank Abild-Pedersen[*,‡]

[†]*Department of Chemical Engineering, Stanford University, 443 Via Ortega, Stanford, CA 94305, USA*

[‡]*SLAC National Accelerator Laboratory, SUNCAT Center for Interface Science and Catalysis, 2575 Sand Hill Road, Menlo Park, California 94025, USA*

E-mail: abild@slac.stanford.edu

Keywords

Machine learning, scaling relations, atomic site stabilities, density functional theory

Computational details

DFT details

Periodic density functional theory (**DFT**) calculations were performed using Quantum ESPRESSO[1], the BEEF-vdw exchange-correlation functional[2] in combinations with ultra-soft Vanderbilt pseudo-potentials[3], and the atomic simulation environment (ASE)[4]. Plane waves and charge densities were expanded up to a kinetic energy of 500 eV and 5000 eV, respectively.

Site stabilities of individual atoms were computed by single-point DFT calculations. The site stability of site A $\mathrm{E_{site}(A)}$ is defined in equation (1), where $\mathrm{E_{defect-free}}$ is the total energy of the original, defect-free cluster, $\mathrm{E_A}$ is the total energy of the free-standing atom A and $\mathrm{E_{defective}}$ is the original structure without atom A:

$$\mathrm{E_{site}(A) = E_{defect-free} - E_A - E_{defective}} \tag{1}$$

The spin states of free-standing atoms were thereby:

Table SI-1: Spin states of atoms in vacuum (box of (15 × 15 × 15) Å^3).

Atom	mag/cell (Bohr)
Ag	1.0
Au	1.0
Cu	1.0
Co	3.0
Rh	3.0
Ir	3.0
Fe	4.0
Os	4.0
Mn	5.0
Re	5.0
Ni	2.0
Pd	0
Pt	0
Ru	0

The cohesive energy $\mathrm{E_{coh}}$ of a cluster $\mathrm{A_xB_y}$ is computed as shown in equation (2), where

$E_{\mathrm{defect-free}}$ is the total energy of the defect-free cluster, $E_{A/B}$ is the total energy of the free-standing atom A/B, and x and y are stoichiometric coefficients. As the cohesive energy is given per atom, it can be interpreted as mean site stability of all atoms in the cluster.

$$E_{\mathrm{coh}}(A_xB_y) = (E_{\mathrm{defect-free}} - x \times E_A - y \times E_B)/(x + y) \tag{2}$$

Cluster details

Mono- and bimetallic sub-nanometer clusters with 3-13 atoms were generated by randomly positioning atoms in a box of $(15 \times 15 \times 15)$ Å^3. The box was located in a $(30 \times 30 \times 30)$ Å^3 unit cell. A genetic algorithm[5] combined with effective medium theory (EMT) potentials was then used to search for stable particle structures. This search was performed using a starting population of 40 candidates of random atomic positions per particle stoichiometry. The lowest energy candidate in the final population of the EMT-based search was then further optimized with DFT. The DFT calculations for clusters and nano-particles were performed with a $(1 \times 1 \times 1)$ k-point set using the Monkhorst-Pack scheme for Brillouin-zone integration.[6] All possible unique combinations of A_xB_y with A, B $\in$ {Ni, Cu, Pd, Ag, Pt, Au} and (x+y) $\in$ {3, 4, 5 . . . 13} were computed.

Nanoparticle details

55-atom nanoparticles were generated using the layer specification approach within ASE, considering only (111) and (100) surfaces. Specifying 2 layers in each direction resulted in a cuboctahedral, 55-atom nanoparticle. From this template, we generated monometallic nanoparticles by setting the interatomic distances to those found in the respective bulk material. For bimetallic particles, we used the lattice constant estimated via Vegard's law[7] to adjust interatomic distances.

Extended surfaces details

Metal surfaces were represented by (2 × 2) surface unit cells of slabs consisting of 7 atomic layers for monometallic systems. For slabs, we used a (7 × 7 × 1) k-point mesh and slab replicas were separated by more than 12 Å. For the slabs and 55 atom nanoparticles, all atoms were constrained to their bulk geometries.

We optimized the metal lattice constants with the computational set up specified in the main manuscript. With this set up, we obtained the following lattice parameters, which we used to set up nanoparticles and slabs.

Table SI-2: Lattice constants as calculated with our computational set up, see computational details.

System	a (Å)
Ag	4.22
Au	4.20
Cu	3.66
Pd	3.99
Pt	3.99

Alloy stability model

One objective of this work is to compare the multi-feature machine-learning (ML) based models with the two feature physics-based alloy-stability model (ASM).[8–13] The ASM was parametrized using the RPBE functional.[14] We do not expect the choice of the XC functional to influence model accuracy, since model accuracies are represented by differences between model predictions and DFT calculations. Nevertheless, we computed selected site stabilities on Ag, Au, and Cu systems using both RPBE and BEEF-vdW. Parity plots comparing predicted and DFT calculated values for both XC functionals are shown in Figure SI-1. Clearly, both functionals show a mean averaged errors (MAEs) of approximately 0.03 eV.

We performed additional DFT calculations on size-selected Pt and Au cubo-octahedral nanoparticles (CUBO) ranging from 13 atoms, 55 atoms, 147 atoms, 175 atoms (truncated 309 atom clusters) to 309 atoms. Our space of materials for extended surfaces consisted of

(111), (100), and (211) crystal planes of Ir, Pd, Pt, and Rh. The calculation details are similar to our earlier papers introducing the alloy stability model[8,9] with the salient details summarized below. We computed atomic site stabilities from first principles following equation (1) for atoms having diverse coordination numbers ranging from 3 to 12. Total energies were computed using the RPBE functional in Quantum ESPRESSO[1] within the Atomic Simulation Environment (ASE).[4] Total energies were determined with kinetic energy cutoffs of 500 eV and density cutoffs of 5000 eV using Vanderbilt ultrasoft pseudopotentials.[3] Spin polarized calculations were used for 13 and 55 atom clusters since these clusters have significant magnetic moments. All other calculations were spin paired. To facilitate convergence, we used fermi dirac smearing with a width of 0.1 eV. Kohn Sham equations were iteratively solved with total energies converged to a threshold of 10^{-5} eV and forces to 0.02 eV/Å. For all nanoparticles, we ensured a vacuum of at least 10 Å. Total energies of nanoparticles were determined using a (1x1x1) Monkhorst Pack k-point grid.[6] An (8x8x1) k-point grid was used for (111) and (100) surfaces while a (6x6x1) k-point grid was used for (211) surfaces. For extended surfaces, a dipole correction was employed along the z direction to reduce spurious electrostatic interactions between periodic images.

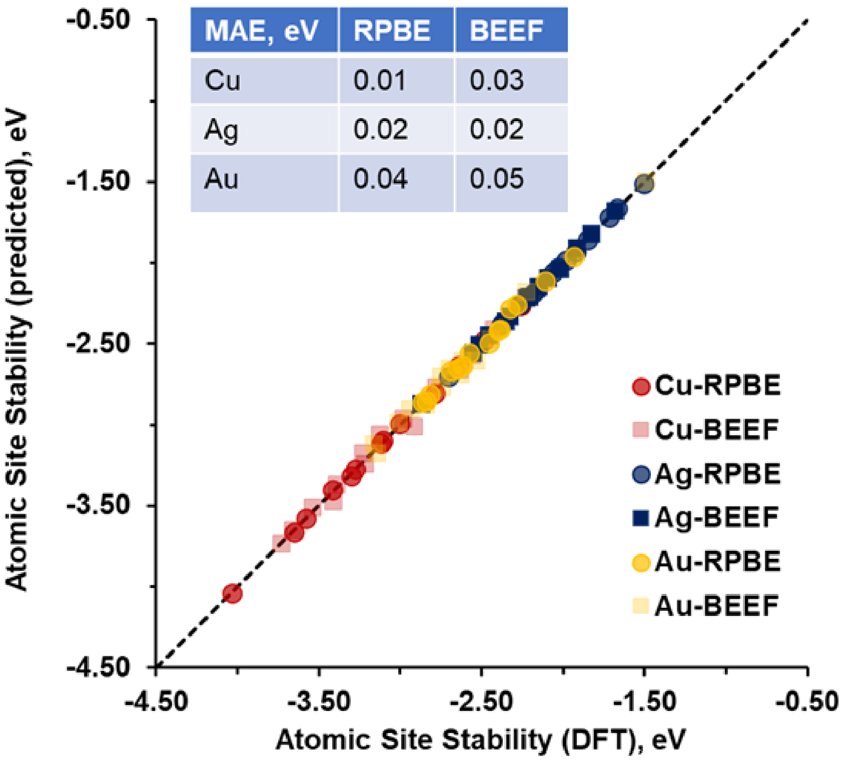

Figure SI-1: Parity plots of ASM-predicted site stabilities vs true DFT-based site stabilities computed using RPBE and BEEF. The mean absolute error of predictions are comparable for both DFT functionals.

We predict atomic site stabilities for metal atoms in nanoparticles and crystal planes using the alloy stability model. The alloy stability model calculates site stabilities based on the site composition and coordination numbers of the first coordination shell. The interested reader is referred to the following publications for additional details.[8–11] Parity plots comparing DFT calculated site stabilities with model predictions are shown in Figure SI-2. Model predictions on nanoparticles determined with the alloy stability model parametrized using extended surfaces are shown in the blue series in Figure SI-2. Significant deviations from the parity line, especially for the 309 atom nanoparticles indicate that there are residual quantum and finite size effects. To account for these effects, we re-parametrized the alloy stability model for a given nanoparticle size (e.g. 147 atoms) by now fitting the model to DFT derived atomic site stabilities corresponding to the same nanoparticle size. We evaluated the accuracy of the re-parametrized alloy stability models using a leave one out cross validation.

The leave one out predictions are shown by the red series in Figure SI-2.

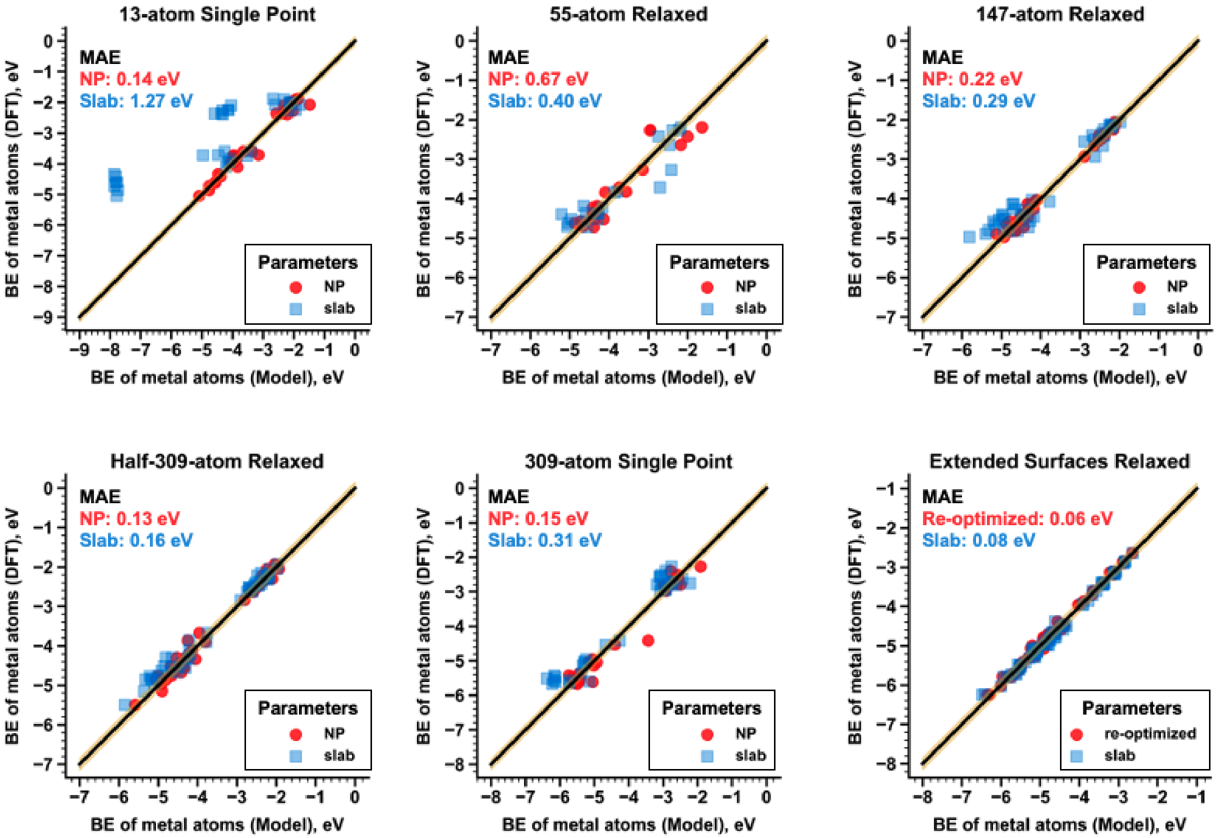

Figure SI-2: Parity plots comparing atomic site stabilities predicted using the Alloy Stability Model (ASM) with DFT derived values. The blue series represents predictions using model parameters fitted to extended surfaces. The red series depicts model predictions when the ASM model is re-parametrized for a specific size range. Due to quantum and finite size effects, the as parametrized ASM does poorly on 13 atom CUBO clusters (MAE of 1.27 eV). These 13 atom clusters are single point calculations because the ASM does not include structral features like the machine learning model. For 55, 147, and 309 atom nanoparticles, the re-parametrized ASM is predictive within 0.15 to 0.20 eV, which are errors typical to screening paradigms. Calculations with 309 atom nanoparticles are also single point to minimize computational cost. Selected geometry optimizations reveal that 309 atom nanoparticles have minimal reconstructions, validating this assumption.

Machine learning details

We deployed various machine learning (**ML**) models, including ordinary linear regression (**ORL**), Gaussian process regression[15] (**GPR**), neural networks[16] (**NN**), random forests

(**RF**), and extreme gradient boost[17] (**XGB**) decision trees. We used the scikit-learn packages for ORL/GPR/RF and keras[16]/ternsorflow for NN, as well as the XGBoost package for XGB. No regularization was used for the ORL. The Gaussian process regression was set up using a linear and a radial basis kernel of length scales between $10^{-3} - 10^{3}$. We applied hyper-parameter optimization using the randomized cross-validation (**cv**) scheme to optimize the XGB and the random forest. Details on the neural network, including the architecture and learning curve are shown in the supporting information, section 3.3. We applied early stopping with a patience of 50 epochs in order to prevent overfitting.

A genetic algorithm (**GA**) for feature selection was implemented, the module is available on Github [https://github.com/schlexer/CatLearn]. The algorithm uses a selection of p features (=genes) from the total feature pool ($\mathrm{p_{tot}}$=28 features) to form a chromosome. A random population of n chromosomes is initialized by randomly selecting chromosomes from the pool of all unique chromosomes; the number of unique chromosomes being given by $\frac{\mathrm{p_{tot}}!}{\mathrm{p}!(\mathrm{p_{tot}}-\mathrm{p})!}$. For each chromosome, it's fitness (mean of 4f-cv R^2 on the training set) is calculated. The chromosomes are ranked according to their fitness of which m parent chromosomes are selected from the top to create m-1 offspring chromosomes via combination of 2 parent chromosomes, respectively. As the gene order in the chromosome doesn't matter, the parent chromosomes are shuffled before crossover. Then, the first half (or the integer rounding up) of parent 1 is combined with the second half (or the integer rounding down) of parent 2. If any, duplicated genes in the offspring are replaced by a random choice of genes that aren't already in the chromosome (random mutation). Then the fitness of the offspring is calculated and the least well performing m-1 members of the population are replaced by the offspring. The population is ranked again, and the procedure is repeated for g generations.

In order to balance exploration vs exploitation, we made sure that all features were available in the population at all times. We created a sufficiently large population for the algorithm to run smoothly (less noise during evolution). Additionally, we use an aggressive mutation rate where in each evolution, 5-20% of all genes are mutated. The exact percentage is chosen via a uniform random sampling within the bounds of 5-20%. This percentage range enhances the algorithm's exploration. We ran the genetic algorithm over 50+ evolution steps to further ensure balance between exploration and exploitation.

Especially for small population sizes $n \rightarrow \mathrm{p}_{\mathrm{tot}}$, the occurrence probability of weak genes can drop to 0 and therefore the gene can become absent in the evolution process. In order to prevent this, we monitor the homogeneity index, defined as $1 - \mathrm{p}_{\mathrm{pop}}/\mathrm{p}_{\mathrm{tot}}$, where $\mathrm{p}_{\mathrm{pop}}$ is the number of unique features in the population. Furthermore, at each generation, after the crossover, a random percentage of chromosomes between 0-20% are chosen for random mutation. Here, one random gene position is selected and replaced by a random choice of the residual $\mathrm{p}_{\mathrm{tot}} - \mathrm{p}$ genes which were absent in the original chromosome.

Featurization

The coordination numbers of atomic sites and the coordination number of their neighbors, as well as relevant statistics thereof where automatically gathered. A neighbor is defined as an atom in a vicinity of 3 Å. Note that this may lead to more neighbors for smaller atoms. We assume that this discrepancy does not affect the performance of our models, however we abstained from investigating the effect of different neighbor definitions, as this would exceed the scope of this study. We computed not only the coordination number of the site itself, but also that of its neighbors. Chemical intuition suggests that if the neighbors have many neighbors themselves, they would be forming less strong bonds to the site of interest, although of course there can be odd-even effects related to the valence of each respective neighbor, Figure SI-SI-3.

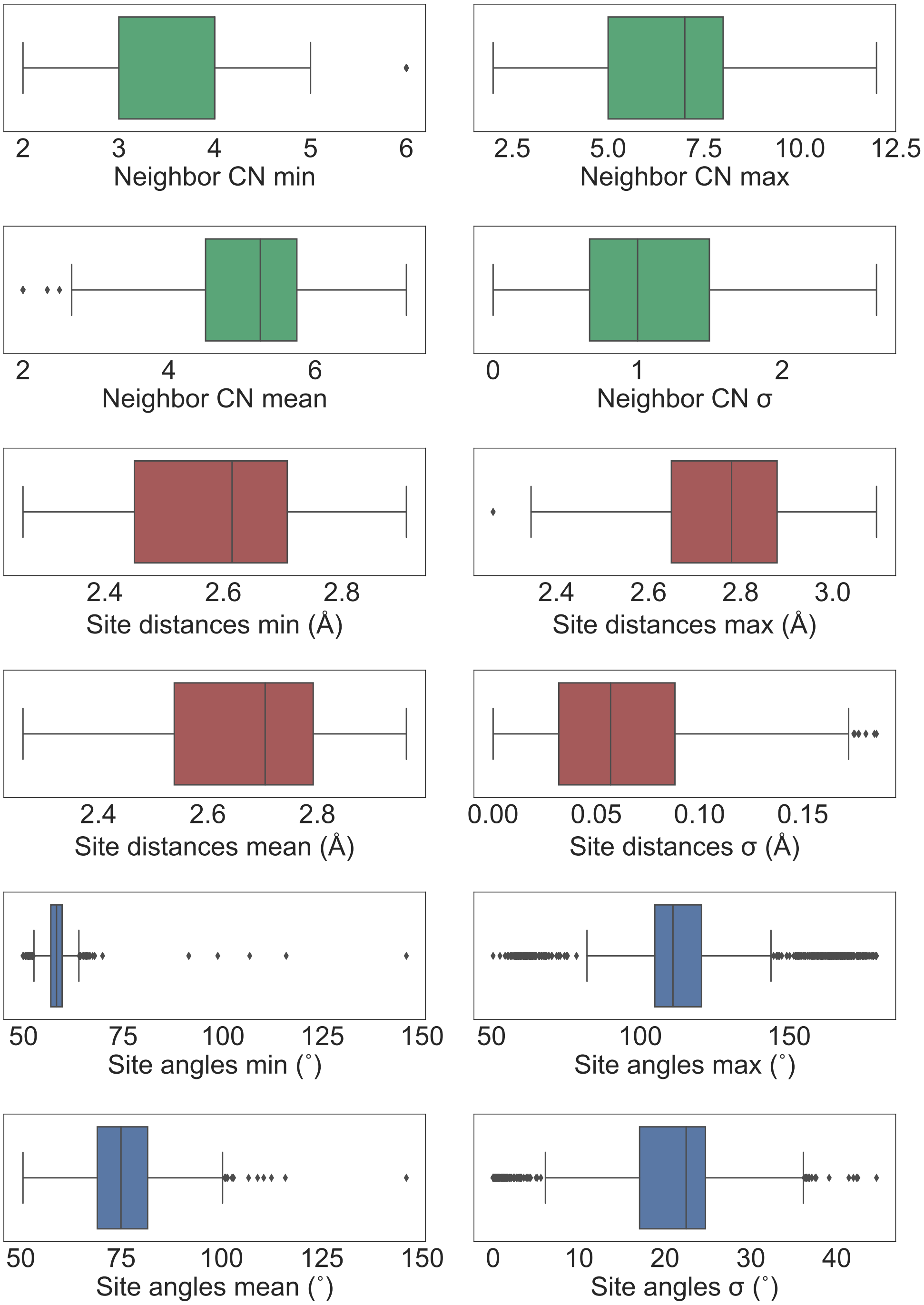

Figure SI-3: Distributions of site features. Green: Site neighbor coordination number. Red: Site distances. Blue: Site angles.

Site distances are distances between the atomic position of the site and its neighbors, respectively. As each atomic site has several neighbors, statistics like the mean and standard deviation of distances represent the nature of the structure in terms of order/disorder (amorphicity in broader terms) and bond strength.

Another interesting structural metric is the angle between 3 atoms. Site angles are defined so that the site constitutes the central atom. That is, a site angle is the smallest possible angles measures between $\mathrm{N_i - S - Nj}$, where $\mathrm{N_{i,j}}$ denote the neighbors with $\mathrm{i \neq j}$ and S denotes the site.

As many of the features are derived from the same set of values, such as distance mean, max, standard deviation etc. are measures of the same set of numbers, we can assume some linear correlation between the features. In order to obtain a rough idea on feature correlations, we computed the Pearson feature correlation matrix, shown in Figure SI-4.

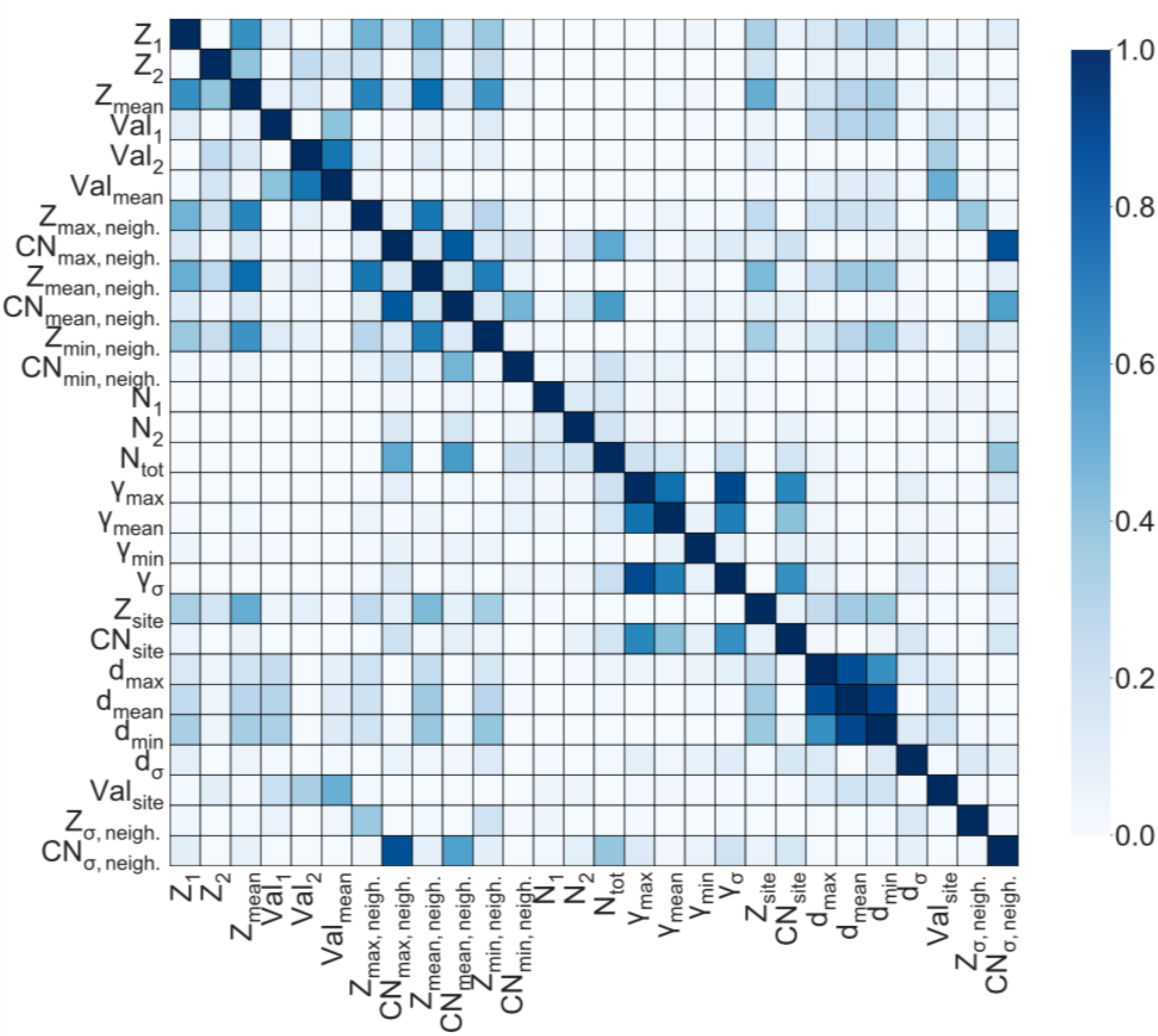

Figure SI-4: Pearson correlation matrix of features. In bimetallic clusters, element 1 is denoted by the number **1** and element 2 by the number **2**. Z is the atomic number, Val is the valence electrons in the elementary state, CN is the coordination number and γ is the angle between the atomic site and two neighboring atoms, and d denote inter-atomic distances.

Model training and feature selection

Training set size convergence

We investigated the convergence of our performance metrics which is the mean R^2 in 4-fold cross-validation with the training set size, including all features. Note that the number of the cross-validation folds is informed by the training set size. A smaller set will need more folds to increase the training split to a reasonable size, so the model can learn from sufficient examples.We consistently employed a leave one out analysis for the alloy stability model since the training sets for nanoparticles contained between 15 to 35 adsorption events. Earlier studies[8,9] showed that a minimum of 14 adsorption events is required to fit the 10 parameters per metal identity (e.g. Au) Interestingly, the data-greedy neural network algorithm already performs well with a smaller training set size. All algorithms reach a plateau at a training set size of 300-400 data points, Figure SI-5.

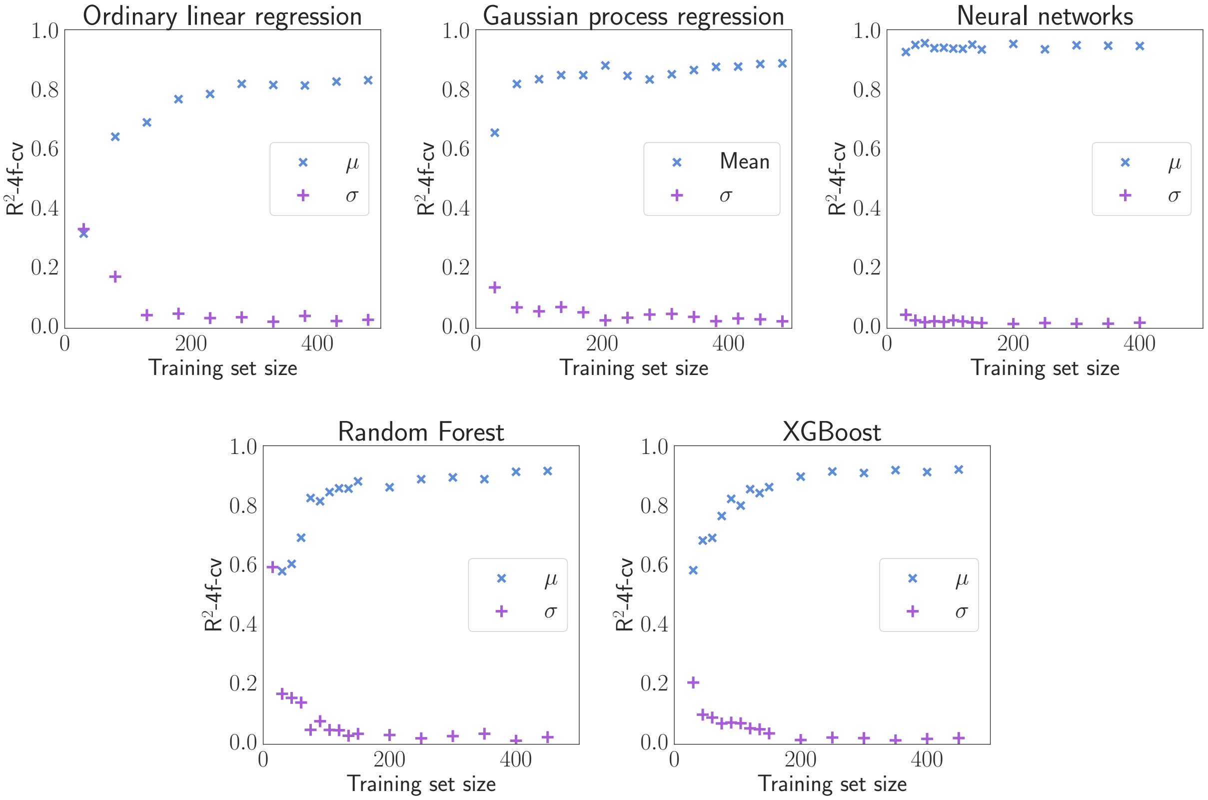

Figure SI-5: Convergence of mean R^2 on 4-fold validation set with training set size.

Ordinary linear regression

The ordinary linear regression model using all features reaches a R^2 value on the test set of 0.81. As some features show collinearity (see Figure SI-4), we tested, if we could achieve similar performance using less features. To quantify the degree of linear correlations in models, we use the sum of pair-wise Pearson feature correlations, $\Sigma_{i<j}\ P_{ij} < 1$. Indeed, using 5 features, the R^2 of 0.79 is close to that using all features (0.81), and a few of the solutions showing an R^2 close to 0.8 even have small feature correlation $\Sigma_{i<j}\ P_{ij} < 1$, Figure SI-6. However, for models with 5 features, the majority of the models show significant feature correlation $\Sigma_{i<j}\ P_{ij} > 1$. We therefore conclude that a feature space of 5 is sufficient to for a subsequent analysis of the importance of individual features.

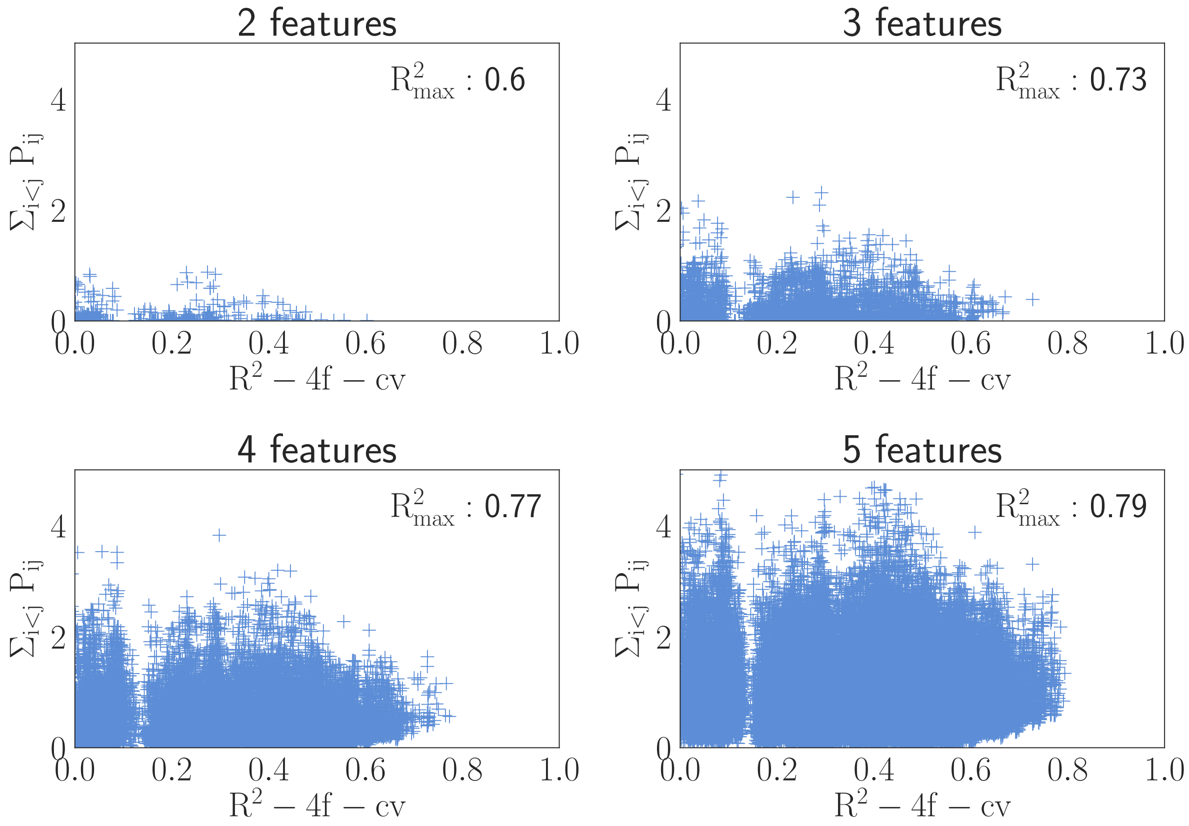

Figure SI-6: Sum of pair-wise Pearson feature correlations, $\Sigma_{i<j} P_{ij} < 1$ and performance metrics mean R^2 on 4-fold cross validation set for linear models using 2, 3, 4 and 5 unique features respectively.

Neural network

We designed the architecture of the neural network manually by comparing cross-validation performances on the cluster site stability training set. As we're interested in a proof of concept that the featurization scheme works for neural networks, we abstain from performing an in-depth model selection. In other words, we are aware that there are more sophisticated hyper-parameter optimizations, but as this network performed sufficiently well, we abstained from further systematic investigations of this matter. The architecture consists of 6 fully connected layers and one dropout layer as the second layer. We closely monitored the learning curve of the NN to avoid over-fitting. During an epoch, the neural net trains on each training point (or batches of training points). After having gone over all training points, the neural net updates its parameters. Then, it can train again on the training set with the new parameters, which would correspond to the second epoch. After using early stopping after 150 epochs, we still slightly over-fit to the data. The learning curve is shown in Figure SI-7. For high-throughput studies, we recommend trying out neural networks if your training data set is large.

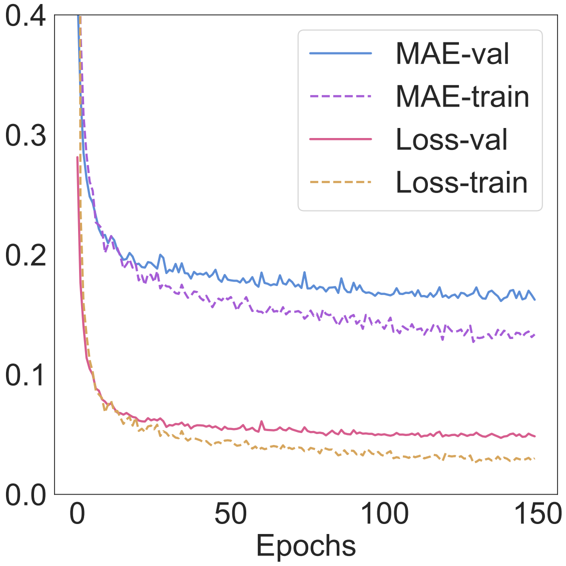

Figure SI-7: Learning curve of the neural network. The mean absolute error and the loss decrease with the number of epochs the neural net goes over the training data.

Feature importance

Clusters

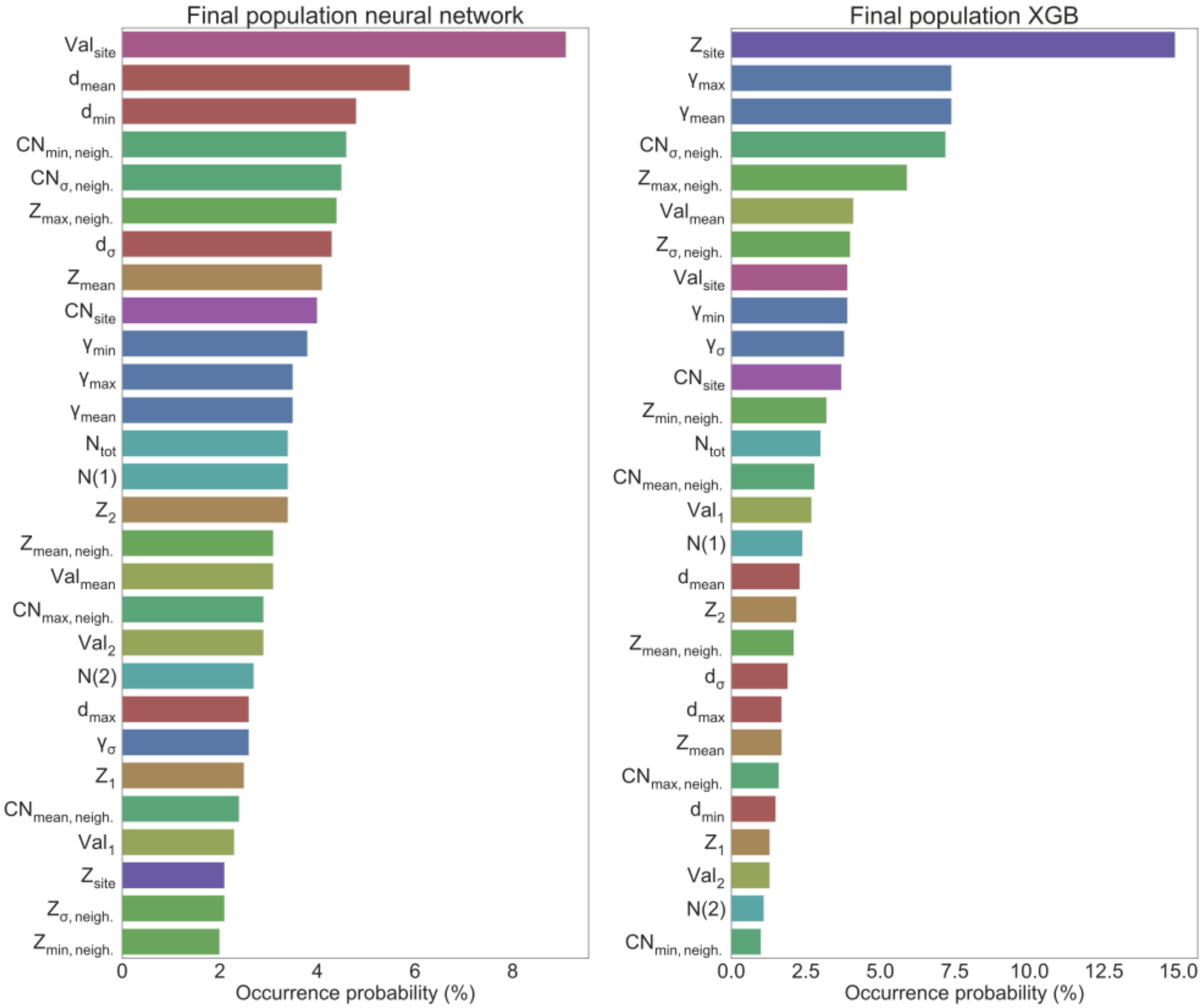

Figure SI-8: Feature distributions of populations (size 200) after 50 generations of evolution for the neural network (offspring size 10) and the XGB (offspring size 20). Please note that the feature occurrence probability from the GA analysis is not generally transferable to other materials or other models. However, we make the observation that for the metal clusters and nano-particles we investigated, an atomic identifier such as the atomic number or valence electrons, followed by structural features is the consistent trend.

Surfaces

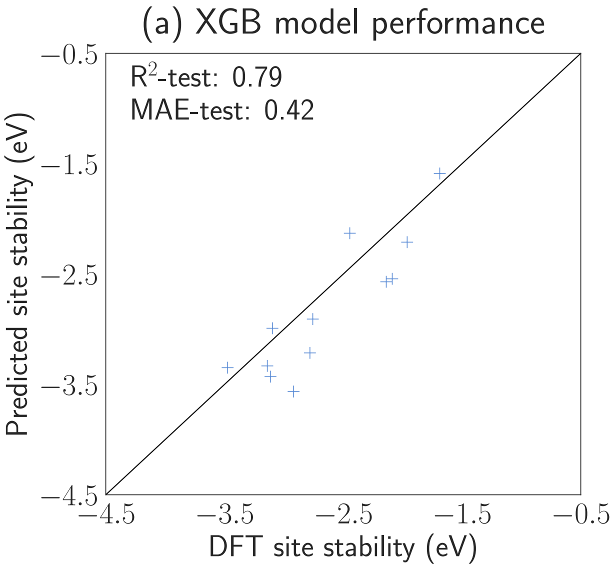

Figure SI-9: Model performances for surface site stabilities on full feature data set. MAE given in eV. Important features are mainly distance-based.

Table SI-3: Comparing prediction errors in binding energies of metal atoms and catalytic descriptors across different feature sets. All feature sets for metallic systems listed below yield mean averaged errors between 0.1 to 0.2 eV on the test sets considered by the respective studies.

Scheme	Mean average error, eV	Comments
SOAP[18]	0.08 eV	Adsorption energies of O*, N*, NO* on RhAu alloys
Coordination-based[19]	0.15 eV	Adsorption energies of OH*, CO*, CH_3* on monometallic Ag, Au, and Cu
Coordination-based ([8–11], alloy stability model)	0.1 to 0.2 eV based on composition, and whether the system is an extended surface or a nanoparticle	Adsorption energies of metal atoms on extended surfaces and nanoparticles of monometallic/bimetallic Ag, Au, Cu, Ir, Pd, Pt, and Rh
Moments of projected d-states[20]	0.15 eV	OH*, O*, CH*, C* adsorption on d-block alloys
Moments of projected d-states[21]	0.13 eV	CO* adsorption on (100) and (111) Cu-alloys
Graph-based neural networks[22]	0.2 eV	CO* and H* adsorption on transition metal alloys

References

(1) Giannozzi, P.; Baroni, S.; Bonini, N.; Calandra, M.; Car, R.; Cavazzoni, C.; Ceresoli, D.; Chiarotti, G. L.; Cococcioni, M.; Dabo, I., et al. QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *Journal of physics: Condensed matter* **2009**, *21*, 395502.

(2) Wellendorff, J.; Lundgaard, K. T.; Møgelhøj, A.; Petzold, V.; Landis, D. D.; Nørskov, J. K.; Bligaard, T.; Jacobsen, K. W. Density functionals for surface science: Exchange-correlation model development with Bayesian error estimation. *Physical Review B* **2012**, *85*, 235149.

(3) Vanderbilt, D. Soft self-consistent pseudopotentials in a generalized eigenvalue formalism. *Physical review B* **1990**, *41*, 7892.

(4) Larsen, A. H.; Mortensen, J. J.; Blomqvist, J.; Castelli, I. E.; Christensen, R.; Dułak, M.; Friis, J.; Groves, M. N.; Hammer, B.; Hargus, C., et al. The atomic simulation environment—a Python library for working with atoms. *Journal of Physics: Condensed Matter* **2017**, *29*, 273002.

(5) Vilhelmsen, L. B.; Hammer, B. A genetic algorithm for first principles global structure optimization of supported nano structures. *The Journal of chemical physics* **2014**, *141*, 044711.

(6) Monkhorst, H. J.; Pack, J. D. Special points for Brillouin-zone integrations. *Physical review B* **1976**, *13*, 5188.

(7) Vegard, L. Die konstitution der mischkristalle und die raumfüllung der atome. *Zeitschrift für Physik A Hadrons and Nuclei* **1921**, *5*, 17–26.

(8) Roling, L. T.; Li, L.; Abild-Pedersen, F. Configurational Energies of Nanoparticles

Based on Metal–Metal Coordination. *The Journal of Physical Chemistry C* **2017**, *121*, 23002–23010.

(9) Roling, L. T.; Choksi, T. S.; Abild-Pedersen, F. A coordination-based model for transition metal alloy nanoparticles. *Nanoscale* **2019**, *11*, 4438–4452.

(10) Streibel, V.; Choksi, T. S.; Abild-Pedersen, F. Predicting metal-metal interactions. I. The influence of strain on nanoparticle and metal adlayer stabilities. *J. Chem. Phys.* **2020**, *152*.

(11) Choksi, T. S.; Streibel, V.; Abild-Pedersen, F. Predicting metal–metal interactions. II. Accelerating generalized schemes through physical insights. *J. Chem. Phys.* **2020**, *152*.

(12) Roling, L. T.; Abild-Pedersen, F. Structure-Sensitive Scaling Relations: Adsorption Energies from Surface Site Stability. *ChemCatChem* **2018**, *10*, 1643–1650.

(13) Choksi, T. S.; Roling, L. T.; Streibel, V.; Abild-Pedersen, F. Predicting Adsorption Properties of Catalytic Descriptors on Bimetallic Nanoalloys with Site-Specific Precision. *The journal of physical chemistry letters* **2019**, *10*, 1852–1859.

(14) Hammer, B.; Hansen, L. B.; Nørskov, J. K. Improved adsorption energetics within density-functional theory using revised Perdew-Burke-Ernzerhof functionals. *Physical review B* **1999**, *59*, 7413.

(15) Rasmussen, C. CKI Williams Gaussian processes for machine learning. 2006.

(16) Chollet, F. keras. `https://github.com/fchollet/keras`, 2015.

(17) Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. 2016; pp 785–794.

(18) Jinnouchi, R.; Asahi, R. Predicting catalytic activity of nanoparticles by a DFT-aided machine-learning algorithm. *The journal of physical chemistry letters* **2017**, *8*, 4279–4283.

(19) Dean, J.; Taylor, M. G.; Mpourmpakis, G. Unfolding adsorption on metal nanoparticles: Connecting stability with catalysis. *Science advances* **2019**, *5*, eaax5101.

(20) Andersen, M.; Levchenko, S. V.; Scheffler, M.; Reuter, K. Beyond scaling relations for the description of catalytic materials. *ACS Catalysis* **2019**, *9*, 2752–2759.

(21) Ma, X.; Li, Z.; Achenie, L. E.; Xin, H. Machine-learning-augmented chemisorption model for CO2 electroreduction catalyst screening. *The journal of physical chemistry letters* **2015**, *6*, 3528–3533.

(22) Back, S.; Yoon, J.; Tian, N.; Zhong, W.; Tran, K.; Ulissi, Z. W. Convolutional neural network of atomic surface structures to predict binding energies for high-throughput screening of catalysts. *The journal of physical chemistry letters* **2019**, *10*, 4401–4408.