



PCCP

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

Journal:	<i>Physical Chemistry Chemical Physics</i>
Manuscript ID	CP-ART-05-2021-002198.R2
Article Type:	Paper
Date Submitted by the Author:	26-Aug-2021
Complete List of Authors:	schlexer, philomena; Stanford University, Chemical Engineering Choksi, Tej; Nanyang Technological University, Streibel, Verena; Stanford University, Chemical Engineering; SLAC National Accelerator Laboratory, SUNCAT Center for Interface Science and Technology Abild-Pedersen, Frank; Stanford Linear Accelerator Center, SUNCAT

SCHOLARONE™  
Manuscripts

Cite this: DOI: 00.0000/xxxxxxxxxx

# Combining Artificial Intelligence and Physics-based Modeling to Directly Assess Atomic Site Stabilities: From Sub-nanometer Clusters to Extended Surfaces<sup>†</sup>

Philomena Schlexer Lamoureux,<sup>a,b</sup> Tej S. Choksi,<sup>a,b</sup> Verena Streibel<sup>a,b</sup> and Frank Abild-Pedersen<sup>\*b</sup>Received Date  
Accepted Date

DOI: 00.0000/xxxxxxxxxx

The performance of functional materials is dictated by chemical and structural properties of individual atomic sites. In catalysts, for example, the thermodynamic stability of constituting atomic sites is a key descriptor from which more complex properties, such as molecular adsorption energies and reaction rates, can be derived. In this study, we present a widely applicable machine learning (ML) approach to compute the stability of individual atomic sites in structurally and electronically complex nano-materials on the fly. Conventionally, we determine such site stabilities using computationally intensive first-principles calculations. With our approach, we predict the stability of atomic sites in sub-nanometer metal clusters of 3-55 atoms with mean absolute errors in the range of 0.11-0.14 eV. To extract physical insights from the ML model, we introduce a genetic algorithm (GA) for feature selection. This algorithm distills the key structural and chemical properties governing the stability of atomic sites in size-selected nanoparticles, allowing for physical interpretability of the models and revealing structure-property relationships. The results of the GA are generally model and materials specific. In the limit of large nanoparticles, the GA identifies features consistent with physics-based models for metal-metal interactions. By combining the ML model with the physics-based model, we predict atomic site stabilities in real time for structures ranging from from sub-nanometer metal clusters (3-55 atom) to larger nanoparticles (147 to 309 atoms) to extended surfaces using a physically interpretable framework. Finally, we present a proof of principle showcasing how our approach can determine stable and active nanocatalysts across a generic materials space of structure and composition.

## 1 Introduction

The working principle of a vast variety of material classes is governed by atomic-scale features, such as low-coordinated sites, point defects and other complex atomic arrangements at the sub-nanometer scale. Prominent examples of such materials are heterogeneous catalysts<sup>1-3</sup>, molecular catalysts<sup>4-6</sup>, semiconductors<sup>7</sup>, molecular electronics<sup>8</sup>, bio-engineered materials<sup>9</sup>, and enzymes<sup>10</sup>. To date, meaningful computational studies of such materials require complex atomic-scale models combined with high-accuracy first-principle calculations.<sup>1,4</sup> The knowledge generated by such studies paves the way for the rational design of high-performance materials.

An enduring challenge in the rational design of materials is the computationally guided design of nanostructures which are synthesizable in practice. These nanostructures include small clusters, nanoparticles, nanorods, and ultra-thin layers. The first step towards gauging the synthesis feasibility of any *in silico*-designed material is to evaluate its thermodynamic stability. Prevailing computational paradigms to assess the thermodynamic stability of nanomaterials are generally limited by at least three key requirements. First, these paradigms should enable predictions of thermodynamic stability with atomic-site level resolution. Second, these paradigms should be generalizable across the vast structural (sub-nm to 10 nm), morphological, and compositional (p-, d-block elements) space typical to most nanomaterials. Third, for improved transferability and realistic speed of high-throughput efforts, these paradigms should use inputs that are obtained on the fly. Such inputs can, for example, include basic chemical information of constituting elements, coordination numbers, and other structural properties.

An efficient and accurate evaluation of atomic site stabilities

<sup>a</sup> Department of Chemical Engineering, Stanford University, 443 Via Ortega, Stanford, CA 94305, USA.

<sup>b</sup> 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.

<sup>†</sup> Electronic Supplementary Information (ESI) available. See DOI: 00.0000/00000000.

has implications beyond simply finding stable nanoparticles. Recent experimental<sup>11–13</sup> and theoretical studies<sup>14–17</sup> have shown that the stability of atomic sites is a robust descriptor for adsorption energies of reaction intermediates. These studies introduce generalizable correlations between the stability of adsorption sites and adsorbate binding energies. Since these adsorption energies are, in turn, linked to reaction rates through the Sabatier volcano,<sup>18–20</sup> computing atomic site stabilities permits us to estimate reaction kinetics. These emerging correlations between stability and reactivity complement existing volcano-based screening of reaction kinetics, such as those based on adsorption energies<sup>1,21–24</sup> and Brønsted-Evans-Polanyi<sup>25,26</sup> relations. Using site stabilities as descriptors therefore opens up the possibility to screen catalytic properties across a large set of materials.

In the last decade, machine learning (ML)-aided high throughput screening has accelerated materials discovery.<sup>27–29</sup> These methods have, however, predominantly been applied to periodic bulk materials.<sup>20,27,28,30–34</sup> To predict properties of materials using machine learning, the atomic structure information should be transformed into a format that is digestible by an ML algorithm. This process is called *featurization*. Featurization methods such as Voronoi tessellation<sup>30,35</sup>, are well developed for perfect bulk materials. These methods, however, have not been widely employed for complex sub-nanometer structures or non-crystalline/amorphous structures.

Features for individual sites should reflect the physical and chemical properties of the local environment. For this reason, several site-specific features, broadly categorized into (1) *electronic* and (2) *geometric* features have been previously used. Prominent examples for electronic structure features are the d-band center<sup>36,37</sup> and moments of the d-band distribution<sup>38</sup>. These features can be computed via density functional theory (DFT) or the tight-binding model.<sup>39</sup> Computing electronic structure features requires careful selection of the electronic structure method (e.g. tight-binding vs. DFT vs. DFT + Hubbard U<sup>40</sup> vs. higher-level first principles methods). Once established, these electronic structure-based featurization schemes have limited computational efficiency and limited generalization potential in materials space.

In terms of geometric features, notable featurization schemes include the coordination-based alloy stability model (ASM)<sup>41–44</sup>, generalized coordination numbers (GCN)<sup>45–48</sup>, and orbital-wise coordination numbers<sup>49</sup>. The alloy stability model partitions energies of metal atoms into contributions arising from individual metal-metal bonds formed. The parameters reflecting these contributions are fitted using a limited set of DFT calculations (as few as 6 DFT calculations per metal). The model explicitly considers the coordination numbers of nearest neighboring atoms while composition effects are treated in a mean-field approach. Streibel et al.<sup>43</sup> have shown how the effects of strain can be incorporated into the alloy stability model thus making the model more accurate in the finite-size regime. This simplistic deconstruction of metal-metal interactions is valid because long-range interactions are effectively screened out by d-electrons.<sup>50,51</sup> Even with this simple linear model, mean absolute errors (MAE) of 0.10 eV (extended surfaces) and 0.21 eV (147 atom nanoparticles) are

achieved.

The coordination-based models described above generally work best for ordered systems, such as nanoparticles > 2nm in diameter and extended crystal planes. Since the coordination-based models have been derived from bulk and surface structures, they generally lose accuracy for disordered structures and sub-nanometer particles, which are affected by finite- and quantum-size effects.<sup>52–56</sup> Such effects can, amongst others, result in the discretization of electronic states, giving rise to the well-known odd-even effects<sup>52,57</sup>.

Hence, encoding local atomic geometries of nanoparticles requires more sophisticated schemes such as the smooth overlap atomic position (SOAP) approach<sup>39</sup>. This approach compares materials sections using a similarity kernel. Jinnouchi et al.<sup>58,59</sup> successfully used this approach to predict adsorption energies on RhAu nano- and sub-nanometer particles. The authors, however, needed to include small nanoparticle data in their training set and could not use calculations on extended surfaces to accurately predict energetics of nanoparticles. The accuracy for RhAu sub-nanometer particles remained limited with MAEs between 0.2–0.4 eV. For kernel methods, such as the SOAP method, the prediction performance systematically improves by increasing the number of training data to cover all possible local structures. However, since kernel methods scale unfavorably with training set size, the SOAP approach becomes computationally demanding when higher accuracy or better generalizability is required. Furthermore, the model needs to be re-parameterized for every new bimetallic composition in consideration.

Non-linear machine learning models have successfully predicted chemical adsorption energies based on geometric and electronic structure features.<sup>28,60–65</sup> For example, Alexandrova and coworkers developed a featurization scheme based on electronic structure features.<sup>66–69</sup> As mentioned before, however, using features derived from electronic structure is computationally demanding and reduces the efficiency of re-training the model for new structures. For an in-depth overview on featurization schemes and machine learning techniques in computational heterogeneous catalysis, the reader is referred to our recent review on this topic.<sup>70</sup>

The key challenges in predicting the site stabilities, and therefore materials stabilities generally, are: First, to use "cheap and fast" features based on basic chemical and structural properties. Second, to develop approaches that are applicable to non-crystalline, non-periodic complex structures commonly found in materials with nano-scale functionality. Third, to ensure model transferability across the vast space of materials in terms of size scale (sub-nm to extended surfaces) and composition (metals, oxides, etc.). Finally, to formulate physically interpretable machine learning models, which can be integrated with physics-based approaches.

In this study, we present machine learning algorithms to predict the site stability of metal structures ranging from atomic clusters, to sub-nm nanoparticles, to extended surfaces with site-by-site precision. We begin by discussing a site-specific featurization scheme relying only on basic chemical and geometric information. For a given site, this scheme generates a set of fea-

tures including both the local chemical environment and non-local, system-wide features. The generated features, when employed in conventional linear and non-linear ML models, achieve high-accuracy predictions for atomic site stabilities. We further present a genetic algorithm for feature selection to obtain physical insights from the ML algorithms. We successfully predict atomic site stabilities of atoms in sub-nm nanoparticles, achieving MAEs as low as 0.1 eV using only the 5 most important features. By examining the relative importance of features with increasing nanoparticle size, we show that the physical insights revealed by the machine learning models are consistent with the physics-based ansatz of the alloy stability model. Taken together, both the machine learning and physics-based models present a unified framework for determining catalyst stability for structures ranging from sub-nm clusters to extended surfaces. We conclude with a proof of principle demonstrating the utility of our model in screening stable and active catalysts with atomic level precision across different nanoparticle size regimes. All data handling, featurization, and feature selection functionalities used in this work are provided to the community in open-source python modules. ([github.com/schlexer/CatLearn](https://github.com/schlexer/CatLearn))

## 2 Results

The Results section is structured as follows: We will first show how to featurize atomic sites in 3-13 atom clusters, use machine learning to predict atomic site stabilities, and present a genetic algorithm for feature selection in the section *3-13 atom clusters*. The genetic algorithm furnishes an in-depth understanding of the importance of different features. In the sections *Nanoparticles* and *Surfaces*, we apply the scheme to 55-atom nanoparticles and extended surfaces, respectively. We then demonstrate in the section *Determining Catalyst Stability from sub-nm Nanoparticles to Extended Surfaces* that a unified approach combining machine learning models and the physics-based alloy stability model accurately predicts atomic site stabilities across different nanoparticle size regimes ranging from atomic clusters to extended surfaces. Finally, in the section *Applications of Site Stabilities in Screening Stable and Active Catalysts*, we discuss how efficient predictions of site stabilities accelerate the screening of stable and active nanoparticles having diverse sizes, shapes, and compositions.

### 2.1 3-13 atom clusters

#### 2.1.1 Featurization of atomic sites

We generated mono- and bimetallic sub-nanometer clusters as described in the Computational Details. All possible unique combinations of  $A_xB_y$  with  $A, B \in \{\text{Ni}, \text{Cu}, \text{Pd}, \text{Ag}, \text{Pt}, \text{Au}\}$  and  $(x+y) \in \{3, 4, 5 \dots 13\}$  were computed. Briefly, atomic positions were pre-optimized using the EMT potential in combination with the genetic algorithm from ASE.<sup>71</sup> We emphasize that the ASE algorithm is different from the GA we developed for feature selection; [github.com/schlexer/CatLearn](https://github.com/schlexer/CatLearn). The best candidates were then further relaxed using DFT.

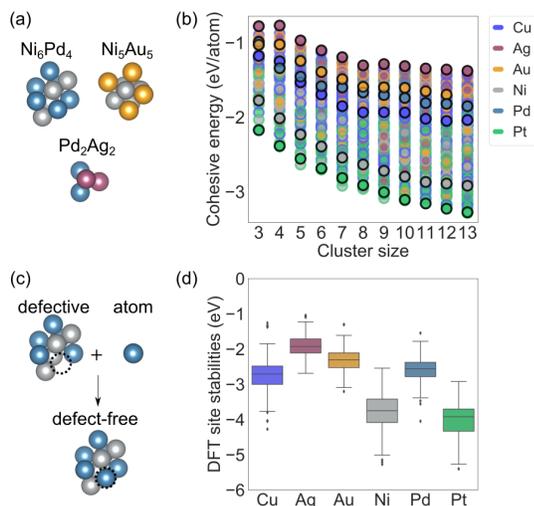
This relaxation procedure yields stable minima of nanoclusters. Notwithstanding the development of global optimization methods like genetic algorithms and simulated annealing, determin-

ing global minima for sub-nm nanoclusters remains a formidable challenge. Computationally predicted global minima should ideally be verified with experimental synthesis. This verification is difficult because of the sensitivity of the nanocluster structures to the synthesis method employed. The objective of this machine learning study is to map site-specific structural features to the stabilities of individual atoms composing the nanoclusters. Sub-nm nanoclusters are known to be fluxional under reaction conditions, i.e., these nanoclusters can alter their structure during an elementary step of a catalytic cycle<sup>72,73</sup>. Furthermore, recent studies have unequivocally shown that the reaction rates on sub-nm nanoparticles arise from a pool of metastable structure, and not one single global minimum<sup>74</sup>. Determining the stability of atoms in sub-nm nanoclusters will enable direct considerations of the dynamic nature of nanocatalysts under reaction conditions. Hence, we employ structure-stability relations to predict the atomic site stability across a variety of nanoclusters, regardless of them being metastable or global minima. Our goal is not to use these structure-stability relations to find the global minimum of a given nanocluster.

Examples of selected nanoclusters are shown in Figure 1 (a). We computed the formation energies, i.e. cohesive energies, of all clusters, as defined in the supporting information, section *DFT details*. The results are shown in Figure 1 (b), where the monometallic clusters have a black circle and bimetallic clusters are bi-colored according to their composition. As expected, we observe a steady decrease in cohesive energy with increasing cluster size.

We then randomly selected atomic sites from the set of all possible atoms of all monometallic and bimetallic sub-nanometer particles, and computed their site stability, according to the reaction illustrated in Figure 1 (c), and defined in the supporting information, section *DFT details*. We plotted the distribution of the site stabilities based on the chemical identity of the site in Figure 1 (d). As expected, the coinage metals Ag and Au bind weaker than the other elements, and also show a tighter distribution. Ni and Pt show the strongest site stabilities and also show a larger variance of their site stabilities depending on their chemical environment.

In order to predict the stability of these atomic sites, we analyze their features, i.e. properties. We can conceptually distinguish between *system-specific* features (e.g. stoichiometry of the nanoparticle, particle size (# atoms)), and *site-specific* features (e.g. the coordination number or chemical environment of the site). Furthermore, we can distinguish between *physical/chemical* and *structural* features. Physical/chemical features are for instance atomic numbers, valence electrons, or electronegativity, whereas structural features entail metrics of inter-atomic distances, angles, or coordination numbers.



**Fig. 1** Cohesive energies and site stabilities of 3-13 atom clusters. (a) A selection of optimized clusters. (b) Cohesive energies of all defect-free clusters. For example, a green dot surrounded by a red circle denotes a Pt-Ag bimetallic cluster. The cluster size is  $x+y$ , where  $x$  ( $y$ ) is the number of atoms of element A (B), and black circles denote monometallic clusters. (c) Computing atomic site stabilities: Negative energies denote exothermic site stabilities. (d) Boxplots of site stabilities by element of site.

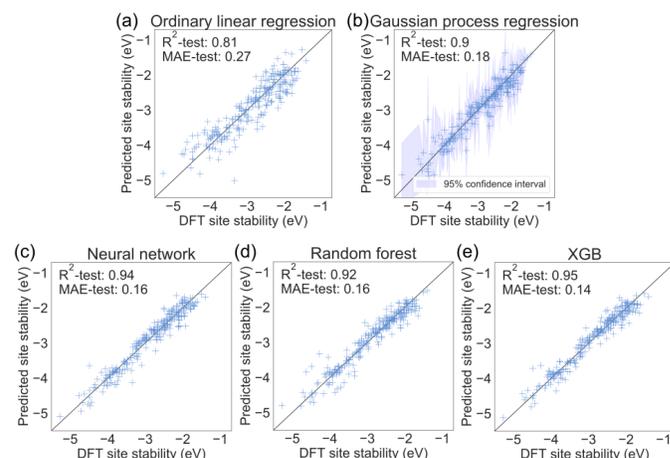
Our goal is to predict site stabilities from features that are determined in a computationally efficient way. We therefore manually chose a set of features that consists only of structural (distances, angles, coordination numbers, etc.) and basic chemical information (atomic number, valence electrons, stoichiometric coefficients, etc.). These features are selected to reflect the local structural and compositional environments of the first and second coordination shells. The well-established near-sightedness of d-electrons implies that perturbations beyond the third-coordination shell will be effectively screened out.<sup>53,55</sup> This choice results in **28 unique features**. We abstain from using the EMT site stability as a feature since the EMT potentials are available for only a limited set of elements. The features are summarized in Figure SI-4 and Table 1. There are several methods to featurize active sites of transition metals. These methods include SOAP<sup>39,58</sup>, coordination-based models<sup>17,41–44</sup>, moments of the projected d-states<sup>19,28</sup>, and converting active site environments into graph based representations<sup>35</sup>. While these models describe active sites in different ways, they generally demonstrate accuracies on the order of 0.1 to 0.2 eV in predicting adsorption energies of small molecules and metal atoms. Two distinctive aspects of our featurization method are: (1) we employ features which can be computed on-the-fly without additional inputs from first principles, (2) we aim to unravel how the most important features listed in Table 1 evolve with increasing nanoparticle size. This discussion is presented in sections *Feature importance analysis*, *Nanoparticles*, and *Determining Catalyst Stability from sub-nm Nanoparticles to Extended Surfaces*.

### 2.1.2 Model selection: Algorithms and features

Having introduced our featurization schemes, we now investigate the performance of various machine learning models to pre-

dict site stabilities. The models include ordinary linear regression (also referred to as ordinary least squares, OLS), Gaussian process regression<sup>75</sup> (GPR), neural networks (NN), random forests (RF), and extreme gradient boost<sup>76</sup> (XGB) decision trees. Further details about these models are in the supplementary information. To ensure that our training set is sufficiently large for all of these algorithms, we tested the convergence of our performance metrics with training set size, see supporting information Figure SI-5. All models reach a plateau in performance after 300-400 data points. We split our data in a training/test ratio of 958/240, therewith achieving a sufficiently large training set to produce meaningful results for our problem of interest. We optimized the model using 4-fold cross validation (**4f-cv**) on the training set, in combination with hyper-parameter optimization using python.

Considering all 28 features and fitting the most promising models on the full training set, we predict atomic site stabilities with MAEs between 0.14 and 0.27 eV. The results are shown in Figure 2. Clearly, the neural network and the extreme gradient boost decision trees are the best performing models with an  $R^2$  on the test set of 0.94 and 0.95, respectively.



**Fig. 2** Performance of various models on the test set (training/test = 958/240) using all 28 features, all MAEs are in eV. (a) Ordinary linear regression model (OLS). (b) Gaussian process regression (GPR) with 95% confidence interval based on posterior distribution. (c) Neural network (NN). (d) Random forest (RF). (e) Extreme gradient boost (XGB).

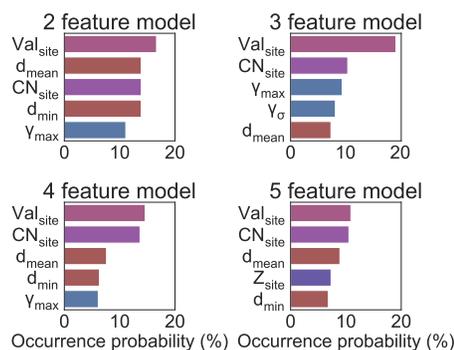
### 2.1.3 Feature importance analysis

Besides evaluating different ML models, our objective is to investigate structure-stability relationships for bimetallic nanoparticles. By determining the most important features, we can work towards a physical interpretation of the model and investigate recurring trends in terms of structure-stability relations. The well-established role of screening effects by d-electrons in transition metals<sup>50,51</sup> implies that minimalist models containing a small feature set will suffice. We aim to identify the most relevant features from the pool of 28 features in Table 1. Hence, instead of rank ordering all 28 features in terms of their importance, we will identify the 2-5 most important features which describe nanoparticle stability. Feature importances can be model dependent, especially if the features show correlation, as depicted in Fig. SI-4. That be-

**Table 1** Features of  $A_{N_1}B_{N_2}$  binary systems. **n** denotes the number of features in this category, **t** denotes the category type: **s** denotes site-specific, **g** denotes "global", i.e. system-specific. \*x can be mean, min, max or the standard deviation ( $\sigma$ )

Feature(s)	Description	n	t
$N_{\text{tot}}$	Total number of atoms in the system	1	g
$N_1, N_2$	Stoichiometric coefficients of A or B	2	g
$Z_1, Z_2$	Atomic number of A or B	2	g
$Z_{\text{mean}}$	Mean of atomic numbers of all atoms	1	g
$\text{Val}_1, \text{Val}_2$	Valence electrons of A or B	2	g
$\text{Val}_{\text{mean}}$	Mean of valence electrons of all atoms	1	g
$Z_{\text{site}}$	Site atomic number	1	s
$\text{Val}_{\text{site}}$	Site valence electrons	1	s
$\text{CN}_{\text{site}}$	Site coordination number	1	s
$Z_{x,\text{neigh.}}$	Metrics x of neighbor atomic numbers*	4	s
$\text{CN}_{x,\text{neigh.}}$	Metrics x of neighbor coordination numbers*	4	s
$\gamma_x$	Metrics x of the angles between the atomic site and each unique pair of two neighbors*	4	s
$d_x$	Metrics x of the distances between the atomic site and each neighbor*	4	s

ing said, we can investigate general trends in the features which hold across different models and diverse cluster sizes.



**Fig. 3** Feature importance in ordinary linear regression (ORL) models. Using the genetic algorithm, we analyzed populations of models containing between 2 and 5 features, respectively. Shown are the feature occurrence probability (%) in the sub-set of linear models with a mean 4f-cv  $R^2$  in the top 5 percentile of the population.

In this study, the property of interest is the stability of the site. However, the machine learning + genetic algorithm approach can be applied to any kind of property. The physical interpretability of the machine learning model furthermore enables a direct comparison to the physics-based ansatz of the alloy stability model. We begin the feature analysis with linear models to inform our understanding on individually relevant features, i.e. disregarding synergies between 2 or more features. In order to analyze the importance of features, we generated a model for every possible combination of 2 features and determined each model's mean  $R^2$  in 4-fold cross-validation (4f-cv) on the training set. We then ranked the resulting models according to their 4f-cv performance and subsequently analyzed the features used by the best models. Since every model uses only two features, we can compute the number of times each feature was used by the set of models which perform best. In other words, we analyzed the *occurrence probability* of each individual feature across all models performing in the top 5% of 4f-cv  $R^2$ . The top 5 most occurring features are shown in Figure 3 (upper left panel). We repeated the process for models with additional features, i.e. every possible combination of  $p \in \{2,3,4,5\}$  out of  $p_{\text{tot}} = 28$ . The results are shown in Figure 3.

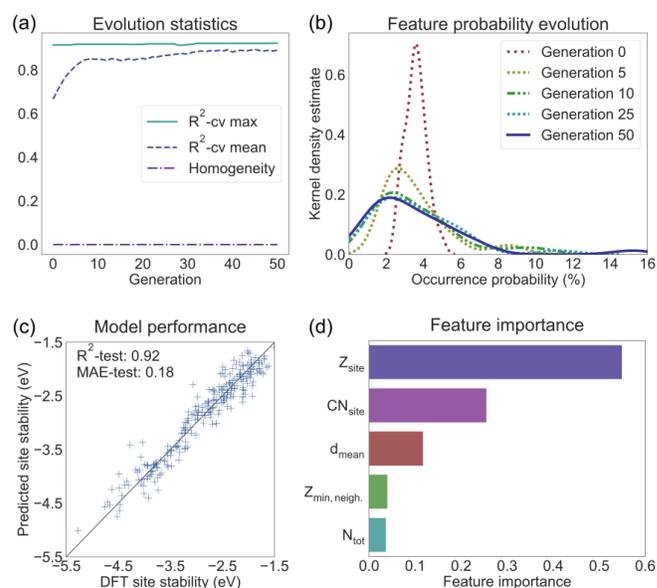
Across all reduced order linear models, the number of valence electrons of the element constituting the site (oxidation state 0),  $\text{Val}_{\text{site}}$ , is the most frequently occurring feature with an occurrence probability of 10-20% in the top 5% of models. The second most frequent feature in all cases is the coordination number of the site,  $\text{CN}_{\text{site}}$ . Thus, accounting for linear correlation of a feature with the target only, the number of valence electrons in combination with the coordination number shows great predictive power. Other important features represent measures of distances and angles. In this feature space, the mean distance of the site to its neighbors  $d_{\text{mean}}$  and the maximum angle between two neighbors and the site  $\gamma_{\text{max}}$  stand out as important features. The ansatz for the linear alloy stability model, inspired by the effective medium theory<sup>77</sup> also uses the site composition and site coordination number as the two main features. Thus, the feature importance obtained through the machine learning model is consistent with physical principles.

Some features are linearly correlated, such as the distance measures ( $d_{\text{min}}/d_{\text{max}}/d_{\sigma}$ ) and angles/coordination numbers ( $\gamma_{\text{mean}}/\gamma_{\text{max}}/\gamma_{\sigma}/\text{CN}_{\text{site}}$ ), see Pearson correlation matrix in the supporting information Figure SI-4. Therefore, the solutions of ordinary linear regression (and those of other models) may not be unique. These linear correlations are not problematic,

though, since our goal is not to identify the best model. Rather, we identify the most critical features determining site stability by quantifying the occurrence probability of a given feature in the top 5% of models. Since we used the linear model in this analysis, we included all possible combinations of  $p$  features, as fitting the model and determining the performance metrics is computationally fast. Non-linear models and especially neural networks are computationally more expensive in training and prediction.

For a more general analysis of feature importance, we need to develop a method that is applicable to problems, for which not all models (feature combinations + algorithm) can be exhaustively computed. We therefore developed a genetic algorithm (GA) for model selection. Briefly, the GA creates a population of chromosomes. A chromosome is a list of  $p$  features, which are called genes in this context. For instance, a 3-gene chromosome consists of  $\{d_{\text{mean}}, \gamma_{\text{max}}, \text{Val}_{\text{site}}\}$ , with the genes being individual features. The GA determines the fitness of all chromosomes given a certain regressor class (e.g. XGB or linear model) and selects the best performing model via evolution over a certain number of generations. The working principle is described in more detail in the supporting information. The goal of the genetic algorithm is to ensure that good feature candidates are enriched in the population. We may already have great feature combinations from the start, but this cannot be guaranteed and may depend on random chance, as we select the initial feature population randomly. The features which are accumulated throughout evolution can be regarded as important features. These features are generally model and material dependent.

We chose to investigate models containing 5 features (genes) to reduce overall feature correlation while allowing higher-order feature interactions in non-linear models. Most feature combinations with  $p=5$  exhibit a sum of pairwise feature correlations  $> 1$ , Figure SI-6. As the NN and the XGB performed best when we used all features (Figure 2), we ran the GA with these two models. We generally used a population size of 200 and an offspring size of 20. In the case of the neural network, we decreased the offspring size to 10 to make the computation faster. The larger the offspring size, the faster the algorithm evolves.



**Fig. 4** Genetic algorithm evolution of feature selection process using the XGB regressor. (a) Solid line: Max fitness (fitness = mean 4f-cv  $R^2$  of best performing chromosome). Dashed line: mean population fitness. Dotted-dashed line: Homogeneity (a value of 0 indicates that all possible 28 features are available in the population). (b) Evolution of feature occurrence probability. At Generation 0, all features show an occurrence probability close to  $1/28 \approx 3.57\%$ . During evolution strong features become more probable, and weak features become less probable. (c) Best performing 5 feature model after evolution, achieving a  $R^2$  of 0.92 (MAE=0.18 eV) on the test set. (d) Feature importances based on the average gain of splits of the XGB model used in (c).

An in-depth analysis of the GA process using the XGB-based process is summarized in Figure 4. In Figure 4 (a), the mean and max 4f-cv  $R^2$  throughout the GA evolution are shown. At generation 0, the population consists of 200 unique random chromosomes, which all show different fitness measured by the mean 4f-cv  $R^2$ . At generation 0, the best performing chromosome (5-feature model) shows a fitness close to 0.9 (solid line) and the population shows a mean fitness of around 0.7 (dashed line). We furthermore follow the homogeneity, which indicates whether all features are still in the population. A value of 0 means that all features are available in the population. The exact definition of homogeneity is given in the supporting information.

During the GA evolution, the population performance increases, consequently improving the mean population fitness shown in Figure 4 (a). At the same time, the occurrence frequency of features changes noticeably in Figure 4 (b). At generation 0, all features show an occurrence probability of around  $1/p = 1/28 \approx 3.57\%$ . Given the randomly created starting population of the chromosome generation, we see a tight distribution around the ideal value of  $1/28$ . In the initial phase of the GA evolution, weak feature combinations are depleted while stronger combinations are enriched, thus broadening the Gaussian distribution. In the later phase of the evolution (generations 25-50) the distribution becomes asymmetric with a broader tail towards larger occurrence probabilities (6-16%). After 50 generations, a small broad peak at around 15% appears. This peak represents the

most relevant features.

After an evolution of 50 generations, the best XGB model (chromosome) using 5 features (genes) achieves an  $R^2$  of 0.92 (MAE=0.18 eV) on the test set, see parity plot in Figure 4 (c). The 5 features used by this model are shown in Figure 4 (d), together with their feature importance. The best neural network model using 5 features achieves a  $R^2$  of 0.89 (MAE=0.20 eV) on the test set. The parity plot shows that reduced order 5 feature models identified through the genetic algorithm retain their predictive accuracy. These 5 feature models also in principle, require smaller training sets as compared to the 28 feature models. The final population feature occurrence probability breakdown is shown in Figure SI-8.

Based on the occurrence probabilities shown in Figure SI-8, both the NN and XGB models suggest a chemical feature to be most important in determining the site stability. For the NN, it is the site valence electrons ( $Val_{site}$ ) and for the XGB the site atomic number ( $Z_{site}$ ). In both cases, the chemical measure is immediately followed by structural features like the coordination number of the site ( $CN_{site}$ ). This trend in occurrence probabilities was also found using linear models (compare Figure 3, Figure 4 and Figure SI-8).

### 3 Nanoparticles

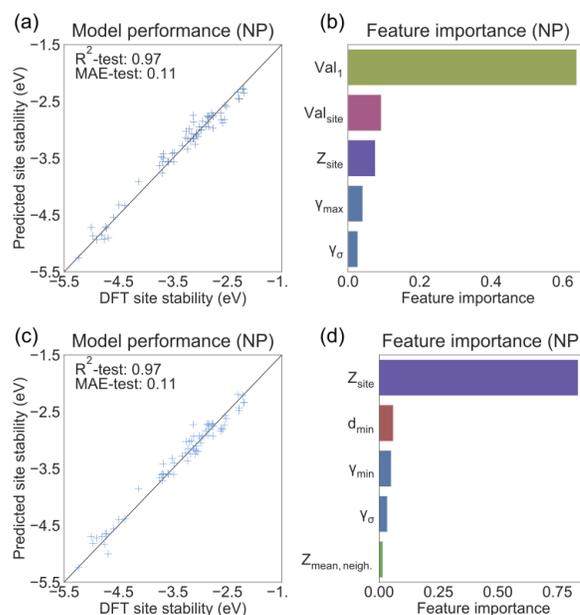
We have seen that the machine learning models predict atomic site stabilities in sub-nanometer clusters with impressive accuracy. We now investigate if models trained on sub-nanometer particle data sets are transferable to larger nanoparticles. To address this question, we created 30 bimetallic, cuboctahedral nanoparticles of stoichiometry  $A_xB_y$  with  $x=28$  and  $y=27$  and  $A \neq B$ . We used the same elements and compositions as for the smaller clusters. For these nanoparticles, we systematically computed the site stability of structurally distinct surface sites, including corners, edges, and terraces. The resulting data set has 327 atomic sites represented by the same set of 28 features discussed above.

To verify model transferability, we predicted site stabilities on 55-atom nanoparticles with the XGB model trained on sub-nanometer clusters. However, we only achieve a  $R^2$  of 0.44 and a MAE of 0.47 eV using the complete 55-atom nanoparticle data set. The other ML models yield similarly high MAEs, indicating that the structure-property dependencies are inherently different for clusters made up of sub-nm clusters vs. nano-particles. These differences in feature importance with nanoparticle size arise because the finite- and quantum-size effects are strongly size-dependent. The large error suggests that the models must be re-trained on the 55-atom nanoparticle data set. This error is not surprising because the influence of quantum- and finite-size effects on the 3-13 atom clusters is too pronounced to capture the properties of the larger, 55-atom nanoparticles. Fortunately, based on our training set convergence tests, we anticipate that training a ML model on a few hundred data points for a respective material class should be sufficient to obtain a performance for practical use. We refitted machine learning models on the 55-atom nanoparticle data set. We split the data set into training/test sets with a ratio of 261/66. The performance metrics may show a larger variance since the data set is small. To quantify the vari-

ance, we computed the mean and standard deviation on the  $R^2$  in a 5-fold cross-validation on the training set. As the random forest (RF) and the extreme gradient boost (XGB) algorithm have proven fast and accurate for the sub-nanometer particles, we will continue our study with these algorithms only and abstain from performing an in-depth model comparison.

Training the random forest (with cross-validation based hyperparameter optimization) using all 28 features gives a  $R^2 = 0.95$  (MAE=0.14 eV) on the test set. Interestingly, the 5 most important features are mainly related to the chemistry of the site and the nanoparticle stoichiometry ( $Val_{site}/Val_1/Z_{site}/Val_{mean}/CN_{site}$ ). We verified this outcome using different random states for the training set.

The XGB algorithm outperforms the RF with a  $R^2$  of 0.97 (MAE=0.11 eV) on the test set. The parity plot and the feature importance are shown in Figure 5 (a-b).



**Fig. 5** Model performance and feature importance of the extreme gradient boost regressor (XGB), training and predicting on the nanoparticle data. (a) Parity plot of hyper-optimized XGB using all features. MAE given in eV. (b) Top 5 features of model from (a). (c) Parity plot of hyper-optimized XGB using only 5 features (best model from genetic algorithm evolution). (d) Feature importances based on the average gain of splits of the XGB model used in (c).

In alignment with the RF model, the XGB model uses various chemical parameters as the most important features. This finding shows how an increase in nanoparticle size makes chemical features more relevant at the expense of complex structural details. This transition in the relative importance of chemical versus structural features converges with the physics-based ansatz the simple coordination-based alloy stability model derived for larger nanoparticles and surfaces.<sup>42</sup> However, we note that for sub-nanometer particles too, chemical information (like atomic number and valence electrons) has always ranked as top feature, followed by structural features.

As the chemical features used by the best models have low cor-

relation, see supporting information Figure SI-4 and SI-6, we anticipate that reducing the feature space could give similarly precise results. We performed an evolutionary selection of 5-feature models using the genetic algorithm introduced in the sections *Computational Details* and *3-13 atom clusters*. After 50 generations of evolution, the best performing model using only 5 features achieves an  $R^2$  of 0.97 (MAE=0.11 eV) on the test set, shown in Figure 5 (c-d). The best 5-feature model performs just as well as the 28-feature model. Hence, by leveraging the genetic algorithm, we can build accurate models for predicting site stabilities with fewer features which are easier to train.

Furthermore, the best 5-feature model for the 55-atom nanoparticles has a similar ranking of feature importance as that of the 3-13 atom clusters, with the site atomic number being the most important feature, followed by structural parameters. However, the relative importance of the structural features changes, compare Figure 4 (d) and Figure 5 (d). In fact, the importance of structural features decreases with increasing system size. The decreasing importance of structural features can be rationalized by the comparatively less structural variability of surface atoms in 55-atom nanoparticles. Thus, the main distinction lies within the chemical identity of the site of interest and its neighbors. We will show in the section *Determining Catalyst Stability from sub-nm Nanoparticles to Extended Surfaces* that nanoparticles larger than 55 atoms are modelled with adequate accuracy through the coordination-based alloy stability model (ASM). This physics-based ansatz employs only two site specific features namely the site composition and the site coordination number.

Overall, the accuracy of the machine learning model on 55 atom nanoparticles is encouraging despite the relatively small data set. Our analysis with the 55-atom nanoparticles reveals that the featurization scheme works for materials that are different from the sub-nanometer particles. Our featurization scheme can, in principle, be transferred to other material classes, such as oxides, carbides, or two-dimensional materials. The software is publicly available for the scientific community to perform such studies ([github.com/schlexer/CatLearn](https://github.com/schlexer/CatLearn)).

## 4 Surfaces

We now use the machine learning approach to predict site stabilities on extended surfaces. Atoms embedded in extended surfaces have less structural variability because of the periodic lattice. As discussed in the introduction, the alloy stability model<sup>15,41,42</sup> works well on extended surfaces. A key strength of the alloy stability model is that this model employs a small training set of only 6 DFT calculations per metal. Moreover, being an analytical model, it is more intuitive than ML-based models. We have nonetheless demonstrated in this study how GA-based feature selection lends physical interpretability to ML models. The main differences in the two approaches are that ML models require increased data availability while not needing DFT-derived features, whereas the CN-based models require limited, DFT-based training sets and are less automated in terms of training and validation.

Here, we construct a ML model for determining site stabilities on extended surfaces. We computed the site stability of atoms at chemically distinct surface sites in (111), (211), and

(100) surfaces of Ag, Au, Cu, Pd, and Pt. As we considered only monometallic systems, we removed features representing binary interactions  $Z_{\text{mean}}$ ,  $\text{Val}_{\text{mean}}$ ,  $Z_{\text{max,neigh}}$ , etc. from the ML models. The resulting data set consists of 70 site stabilities and 16 features. This data set is split into a training/test ratio of 56/14. As this data set is very small, we compared the performance of all models (XGB, RF, NN, ORL) using all features. The tree-based models were trained using hyper-parameter optimization via 10-fold cross-validation on the training set.

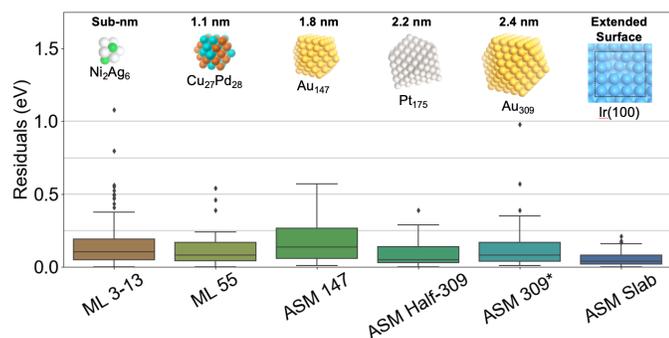
The best performing model is the XGB with a  $R_{\text{XGB}}^2 = 0.79 \pm 0.08$  (MAE = 0.42 eV). We anticipate that performance metrics will improve significantly with larger training set sizes. Importantly, the alloy stability model has a MAE of 0.09 eV on test sets of mono- and bimetallic surfaces.<sup>42</sup> The physics-based functional form of the the alloy stability model results in relatively smaller training sets than needed by machine learning models. On the other hand, the data-heavy machine learning models accurately represent the fluxionality of sub-nm nanoparticles, but require larger data sets for high accuracy. In the next section, we discuss how combining the machine learning and alloy stability models accurately predicts atomic site stabilities across different nanoparticle size regimes ranging from sub-nm nanoparticles to larger nanoparticles and extended surfaces. We also illustrate how the feature importance systematically varies with nanoparticle size, ultimately converging to the physics-based model in the limit of large nanoparticles.

## 5 Determining Catalyst Stability from sub-nm Nanoparticles to Extended Surfaces

This work shows how machine learning can predict atomic site stabilities efficiently in sub-nm nanoparticles. The training sets for the machine learning models contain on the order of  $10^2$  DFT calculations. While such training set sizes are computationally feasible for nanoparticles in the sub-nm size range, such extensive training sets become intractable for larger nanoparticles (e.g. 147 or 309 atoms). We now discuss an approach to predict atomic site stabilities across different nanoparticle size regimes integrating the machine learning framework shown here with the physics-based alloy stability model.<sup>42</sup> In the sub-nanometer regime of 3-55 atoms, the ML models outperform the alloy stability model, since their featurization explicitly considers the non-crystalline structure. In Figure 6 we obtain residuals (absolute errors) of 0.10 to 0.15 eV on the test set consisting of 3-55 atom nanoparticles. Although the residuals in Figure 6 are calculated using the 28-feature model, we show in Figure 4 that the five-feature model obtained using the genetic algorithm yields similar accuracy.

We examine the accuracy of the physics-based alloy stability model on larger nanoparticles (147 to 309 atoms). This model is trained to (211), (111), and (100) crystal planes using two features; namely the site identity and coordination number. We obtain MAEs of 0.29 eV (147), 0.16 eV (Half-309), and 0.31 eV (309) using the alloy stability model fitted to extended surfaces;<sup>41</sup> with the nanoparticle size in atoms in parenthesis. The corresponding parity plots are shown in the Supporting Information. These relatively high errors suggest that the alloy stability

model originally fitted to extended surfaces is less accurate on nanoparticles, possibly because of residual quantum- and finite-size effects in nanoparticles between 147 and 309 atoms. To increase the model accuracy, we can simply re-optimize the parameters of the alloy stability model by re-training the model to specific nanoparticle sizes. A leave one out cross validation yields lower MAEs of 0.22 eV (147), 0.13 eV (Half-309), and 0.15 eV (309) for this re-optimized model. We note that on extended surfaces in the test set, the original alloy stability model predicts atomic site stabilities with a low MAE of 0.08 eV.<sup>41,42</sup> The broadest possible pool of compositions represented in Figure 6 includes monometallic and bimetallic structures of Ni, Cu, Pt, Ag, Pd, and Au. Our previous studies have shown that the mean average errors are independent of the composition for late transition metal systems.<sup>41,42</sup> Further details about the test sets on nanoparticles and extended surfaces used with the alloy stability model are presented in the Supporting Information. Re-parametrizing either the machine learning or the alloy stability model beyond the composition space stated above is a straightforward task. Through the genetic algorithm, we identify reduced order 5 feature models which are, in principle, easier to train as compared to the 28 feature ones. We have also conceptualized an accelerated scheme to parameterize the alloy stability model for monometallic structures<sup>44</sup>. This scheme can be extended to bimetallic alloys.

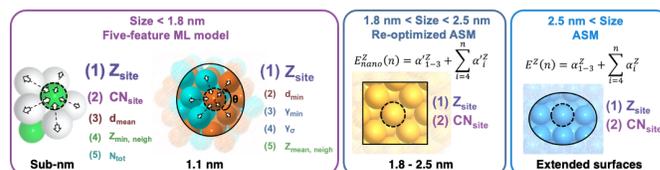


**Fig. 6** Accuracy of real-time predictions using both the machine learning and the alloy stability model (ASM) across different nanoparticle size regimes. Distribution of residuals (absolute errors) in eV for structure-optimized metal systems. The box plots show outliers ( $1.5 \times$  inter-quartile range) as dots. The machine learning (ML) models use the full set of 28 features and achieve MAEs of 0.14 eV and 0.11 eV for 3-13 atom clusters and 55 atom nanoparticles, respectively. For larger nanoparticles, we obtain MAEs between 0.08-0.18 eV using the ASM. \*Calculations with 309 particles are single-point because relaxations are computationally intensive.

To accurately predict atomic site stabilities in different nanoparticle size regimes, we can combine the machine learning model (3-55 atoms), re-optimized alloy stability model (147 - 309 atoms) and the alloy stability model (> 309 atoms). Figure 6 shows that this combined approach yields residuals well within 0.2 eV across these size regimes, which is within the error of high throughput screening models. To the best of our knowledge, this work presents among the few computational frameworks that can predict atomic site stabilities for nanoparticles in different size regimes. In the section *Applications of Site Stabilities in Screening*

*Stable and Active Catalysts*, we show how predictions of atomic site stabilities facilitate designing active and stable nanocatalysts with active site-specific precision. Before discussing how site stabilities enable catalyst design, we first elucidate in Figure 7 how the most important features of these three models systematically evolve with nanoparticle size.

We distill the most important features using a genetic algorithm, thus allowing physical interpretability of the machine learning model. For 3-13 atom clusters (sub-nm in diameter), a combination of features like the site identity ( $Z_{\text{site}}$ ), the coordination number ( $\text{CN}_{\text{site}}$ ), and structural features have high importance. The enhanced accuracy of the ML model is due to the comprehensive feature space considered together with non-linear feature interactions. Moving on to 55 atom nanoparticles (1.1 nm in diameter), we find the relative importance of the site identity ( $Z_{\text{site}}$ ) increases while the importance of structural features decreases. This change is not unexpected because larger nanoparticles have a lower tendency for structural reconstructions. For nanoparticles in the 147 to 309 atom size regime (1.8 - 2.5 nm in diameter), the site identity ( $Z_{\text{site}}$ ) and the coordination number ( $\text{CN}_{\text{site}}$ ) are sufficient to estimate the site stability. Residual quantum- and finite-size effects prevalent in these nanoparticles are accounted for by re-parameterizing the model for a specific size. This analysis demonstrates the progressive changes in feature importance with nanoparticle size ultimately converging to the features used by the physics-based alloy stability model. These results reveal the powerful physical insights that can be extracted by interpreting features of machine learning models in the context of existing physical theories.

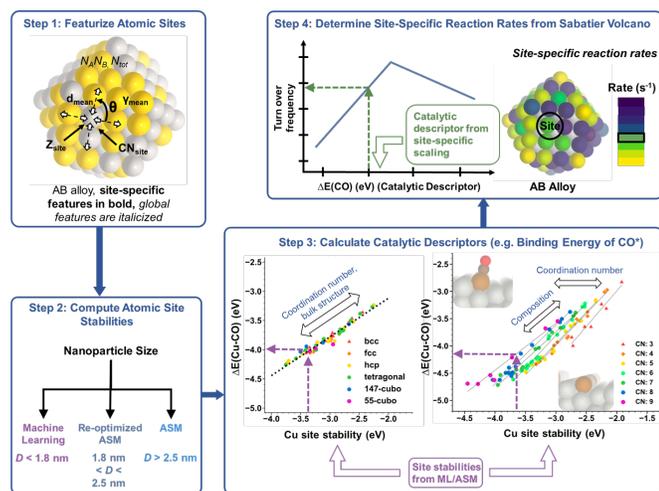


**Fig. 7** Evolution of best-performing models and features with particle size. In the sub-nanometer to nanometer size regime, the best performing models are ML models using information about the site identity, chemical identity of nearest neighbors ( $Z_{\text{site}}$ ,  $Z_{\text{min,neigh}}$ ), the site coordination number  $\text{CN}_{\text{site}}$ , and structural features, such as  $d_{\text{mean}}$ ,  $d_{\text{min}}$ ,  $\gamma_{\chi}$  and  $N_{\text{tot}}$ . With increasing nanoparticle size, the relative importance of structural features diminishes. The features ultimately converge to only the site identity and site coordination number as shown by the alloy stability model (ASM).

## 6 Applications of Site Stabilities in Screening Stable and Active Catalysts

The Sabatier volcano has emerged as the workhorse for *in-silico* catalyst design, especially for bimetallic systems. While volcano plots permit accelerated screening of reactivity and selectivity, similar accelerated approaches to determine catalyst stability remain elusive. Modeling capabilities that can determine catalyst surface energies instantaneously will thus close the gap between candidate catalysts identified through the Sabatier volcano and their inherent synthesizability. In the section *Determining Catalyst Stability from sub-nm Nanoparticles to Extended Surfaces* we

augment machine learning with physical principles to determine atomic site stabilities of bimetallic nanoparticles in different size regimes. We now discuss a proof of principle showcasing how our approach determines stability and reactivity metrics directly. Computations of nanoparticle cohesive energies aid in determining the most stable bimetallic nanoparticle morphology containing a given number of atoms. This unified approach is equally accurate for both sub-nanometer clusters and larger nanoparticles alike. Segregation energies in bimetallic nanoparticles are simply the differences in cohesive energies as atoms are shuffled. Our machine learning and physics-based frameworks can determine the tendency of a bimetallic nanoparticle to segregate in different nanoparticle size regimes. Segregation effects, when evaluated directly, explicitly consider the dynamic nature of bimetallic systems under reaction conditions. By calculating metal atom energies on-the-fly, our machine learning and physics-based models increases the efficiency of workflows evaluating nanoparticle dynamics like sintering. Current sintering models considering both particle migration and Ostwald ripening<sup>78</sup> mechanisms are computationally intensive. The principle hurdle is requiring a priori knowledge of metal diffusion energetics. Because of their lack of generality, such studies have primarily been performed on the most common catalysts like monometallic Pt. Our framework computes energies of metal atoms in a nanoparticle using a set of features determined by the nanoparticle size. These energies, in turn, reveal the thermodynamic barriers of elementary steps in particle diffusion and Ostwald ripening processes. Since our model determines such thermodynamic barriers instantaneously, these barriers can be readily inputted into existing kinetic Monte Carlo and mean field sintering models,<sup>78</sup> enabling sintering evaluations across a broad space of structures and compositions.



**Fig. 8** We first featureize adsorption sites of a given nanoparticle in step 1. In step 2, we compute atomic site stabilities (e.g. Cu) using appropriate models as determined by the nanoparticle size. The atomic site stabilities are then inserted into site-specific scaling relations to calculate binding energies of Cu-CO\* complexes in step 3. From these binding energies, we extract binding energies of CO\*, which are catalytic descriptors. In step 4, catalytic descriptors are inputted into volcano plots (e.g. CO<sub>2</sub> hydrogenation) to evaluate reactivity metrics. Since all steps are evaluated on-the-fly, we can construct heat maps depicting reactivity trends with site-by-site resolution.

Computing atomic site stabilities directly has applications beyond simply evaluating stability metrics. Atomic site stabilities are inherently linked to reactivity metrics as predicated by experiments and illustrated by the alloy stability model. In Figure 8 we present a workflow that connects atomic site stabilities with reaction rates through the Sabatier volcano. Figure 8 contains a family of site-specific scaling relations connecting binding energies of metal-adsorbate complexes (Cu-CO\*) with binding energies of atomic sites (Cu\*). These simple, yet powerful linear trends unify different bulk morphologies, nanoparticles, extended surfaces, and chemical environments. To compute catalytic descriptors of a generic bimetallic nanoparticle, we present an on-the-fly scheme that first computes atomic site stabilities using machine learning or physics-based models and then leverages simple linear correlations to extract catalytic descriptors. In Figure 8, we illustrate a four step process that links active sites of generic bimetallic nanoparticles to reactivity descriptors. We first featurize active sites in terms of global and site-specific properties indicated in Table 1. Next, based on the nanoparticle size, we select the appropriate machine learning or physics-based model to compute active site stabilities. In the third step, these site stabilities are inserted into the family of linear correlations connecting binding energies of atomic sites (e.g. Cu\*) to the binding energies of metal-adsorbate complexes (e.g. Cu-CO\*). We can then extract binding energies of CO\* as  $\Delta E(\text{Cu-CO}) - \Delta E(\text{Cu})$ . CO\* binding energies are popular catalytic descriptors for C<sub>1</sub> conversion reactions like thermal and electrochemical CO<sub>2</sub> reduction.<sup>18,79</sup> These catalytic descriptors can be inputted into existing volcano plots to estimate reactivity metrics. Our scheme directly connects basic structural and chemical properties of generic nanoparticles to reactivity trends using atomic site stabilities as a conduit. Taken together, this comprehensive approach can determine stability and reactivity metrics directly in different nanoparticle size regimes. This work represents an important step towards closing the materials gap between model catalysts and their dynamic structures under reaction conditions.

## 7 Conclusions

In this study, we present a robust scheme to compute stability and reactivity metrics on-the-fly for bimetallic catalysts ranging from sub-nm nanoclusters to extended surfaces. We transform atomic structure data into global and site-specific features that can be fed into any common machine learning (ML) algorithm. Training various ML models on DFT-based site stabilities, this approach predicts site stabilities of sub-nanometer and nanometer bimetallic particles instantaneously and with high accuracy (MAE 0.11-0.14 eV).

We extract physical insights from ML models using a genetic algorithm-based feature selection. Our feature analysis shows how the importance of structural versus chemical features increases as we transition from sub-nm nanoclusters, to nanoparticles, to extended surfaces. Specifically, we establish that a chemical feature (site valence electrons (Val<sub>site</sub>) or the site atomic number (Z<sub>site</sub>)), immediately followed by structural features (such as the site coordination number (CN<sub>site</sub>), distances and angles) represents a powerful feature combination across all models and

nanoparticle sizes. The physical insights interpreted from our feature selection scheme highlight that quantum- and finite-size effects in nanoparticles are likely governed by structural features, since their importance diminishes with increasing nanoparticle size.

Scientific insights obtained from the feature selection scheme are wholly consistent with the physics-based alloy stability model in the limit of large nanoparticles. By integrating both the machine learning and the alloy stability model, we estimate stability metrics in different nanoparticle size regimes with mean average errors of 0.15 eV. We obtain such accuracies only because our model synergistically combines numerous structural features needed to describe fluxional nanoclusters in the sub-nm regime with a conceptually simpler physics-based model in the limit of larger nanoparticles. We use a proof of principle to demonstrate how our unified model computes stability and reactivity metrics of bimetallic nanoparticles directly. Our scheme sets the stage for a new generation of Sabatier volcanos that yield catalytic properties on-the-fly with active site specific precision while concurrently determining stability metrics for candidate catalysts.

Our machine learning workflows including the featurization and feature analysis tools are directly transferable for high-throughout screening of complex, amorphous, and sub-nanometer materials that exhibit diverse structural variety.

## 8 Computational Details

First-principles calculations were performed using Quantum ESPRESSO<sup>80</sup> within the Atomic Simulation Environment (ASE)<sup>81</sup>. We created and optimized atomic structures of mono- and bimetallic sub-nanometer 3-13 atom clusters, cuboctahedral nanoparticles of 55 atoms, and extended surfaces ((100), (111), (211)) using ASE. Further details about structure generation are in the Supporting Information. Total energies were computed using the BEEF-vdW functional<sup>82</sup>. Core states were represented by ultrasoft Vanderbilt pseudopotentials<sup>83</sup>. We computed total energies using a k-point grid generated with the Monkhorst-Pack method<sup>84</sup>. We introduced a dipole correction for extended surfaces that eliminated spurious interactions between periodic images.<sup>85</sup> Additional electronic structure details like the energy cutoffs and numerical convergence criteria are in the Supporting Information. The featurization scheme and the genetic algorithm feature selection are explained in detail in the Supporting Information. The source code for the featurizer and the genetic algorithm based feature selection is open-source available at: [github.com/schlexer/CatLearn](https://github.com/schlexer/CatLearn). The alloy stability model was originally parameterized using the RPBE functional<sup>86</sup> by Roling et al.<sup>41,42</sup>. One objective of this work is to bridge the physical interpretations obtained by the genetic algorithm for sub-nm nanoparticles with the physics-based alloy stability model fitted to extended surfaces. To facilitate this integration across diverse nanoparticle length scales, we performed additional calculations on cuboctahedral nanoparticles containing 13, 55, 147, truncated-309, and 309 atoms, together with (100), (111), (211) crystal planes. In the Supporting Information, we show that the alloy stability model parametrized with either BEEF-vdW or RPBE total energies reveals similar MAEs of under 0.05 eV.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgments

This work was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Chemical Sciences, Geosciences, and Biosciences Division, Catalysis Science Program to the SUNCAT Center for Interface Science and Catalysis. The authors would like to acknowledge the use of computer time for the m2997 allocation at 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. PSL and VS gratefully acknowledge the Alexander von Humboldt Foundation (AvH) for financial support.

## Notes and references

- 1 J. K. Nørskov, T. Bligaard, A. Logadottir, S. Bahn, L. B. Hansen, M. Bollinger, H. Bengaard, B. Hammer, Z. Sljivančanin, M. Mavrikakis *et al.*, *Journal of catalysis*, 2002, **209**, 275–278.
- 2 Z. Ma and F. Zaera, *Surface Science Reports*, 2006, **61**, 229–281.
- 3 W. Tang, L. Zhang and G. Henkelman, *The journal of physical chemistry letters*, 2011, **2**, 1328–1331.
- 4 J. P. Janet and H. J. Kulik, *Chemical science*, 2017, **8**, 5137–5152.
- 5 B. Meyer, B. Sawatlon, S. Heinen, O. A. von Lilienfeld and C. Corminboeuf, *Chemical science*, 2018, **9**, 7069–7077.
- 6 A. Nandy, J. Zhu, J. P. Janet, C. Duan, R. B. Getman and H. J. Kulik, *ACS Catalysis*, 2019, **9**, 8243–8255.
- 7 S. Chen, J.-H. Yang, X.-G. Gong, A. Walsh and S.-H. Wei, *Physical Review B*, 2010, **81**, 245204.
- 8 C. Joachim and M. A. Ratner, *Proceedings of the National Academy of Sciences*, 2005, **102**, 8801–8808.
- 9 B. Kasemo, *Surface science*, 2002, **500**, 656–677.
- 10 M. D. Toscano, K. J. Woycechowsky and D. Hilvert, *Angewandte Chemie International Edition*, 2007, **46**, 3212–3236.
- 11 C. T. Campbell, *Accounts of chemical research*, 2013, **46**, 1712–1719.
- 12 C. T. Campbell and J. R. Sellers, *Faraday discussions*, 2013, **162**, 9–30.
- 13 Z. Mao and C. T. Campbell, *ACS Catalysis*, 2021, **11**, 82842–8291.
- 14 L. T. Roling and F. Abild-Pedersen, *ChemCatChem*, 2018, **10**, 1643–1650.
- 15 T. S. Choksi, L. T. Roling, V. Streibel and F. Abild-Pedersen, *The journal of physical chemistry letters*, 2019, **10**, 1852–1859.
- 16 M. Núñez, J. Lansford and D. Vlachos, *Nature chemistry*, 2019, **11**, 449.
- 17 J. Dean, M. G. Taylor and G. Mpourmpakis, *Science advances*, 2019, **5**, eaax5101.
- 18 J. Schumann, A. J. Medford, J. S. Yoo, Z.-J. Zhao, P. Bothra,

- A. Cao, F. Studt, F. Abild-Pedersen and J. K. Nørskov, *ACS Catalysis*, 2018, **8**, 3447–3453.
- 19 M. Andersen, S. V. Levchenko, M. Scheffler and K. Reuter, *ACS Catalysis*, 2019, **9**, 2752–2759.
- 20 F. Studt, I. Sharafutdinov, F. Abild-Pedersen, C. F. Elkjær, J. S. Hummelshøj, S. Dahl, I. Chorkendorff and J. K. Nørskov, *Nature chemistry*, 2014, **6**, 320.
- 21 F. Abild-Pedersen, J. Greeley, F. Studt, J. Rossmeisl, T. Munter, P. G. Moses, E. Skulason, T. Bligaard and J. K. Nørskov, *Physical review letters*, 2007, **99**, 016105.
- 22 F. Calle-Vallejo, D. Loffreda, M. T. Koper and P. Sautet, *Nature chemistry*, 2015, **7**, 403.
- 23 S. Wang, V. Petzold, V. Tripkovic, J. Kleis, J. G. Howalt, E. Skulason, E. Fernandez, B. Hvolbæk, G. Jones, A. Toftelund *et al.*, *Physical Chemistry Chemical Physics*, 2011, **13**, 20760–20765.
- 24 E. M. Lopato, E. A. Eikey, Z. C. Simon, S. Back, K. Tran, J. Lewis, J. F. Kowalewski, S. Yazdi, J. R. Kitchin, Z. W. Ulissi *et al.*, *ACS Catalysis*, 2020, **10**, 4244–4252.
- 25 J. Brønsted and K. Pedersen, *Zeitschrift für Physikalische Chemie*, 1924, **108**, 185–235.
- 26 M. Evans and M. Polanyi, *Transactions of the Faraday Society*, 1936, **32**, 1333–1360.
- 27 K. Tran and Z. W. Ulissi, *Nature Catalysis*, 2018, **1**, 696–703.
- 28 X. Ma, Z. Li, L. E. Achenie and H. Xin, *The journal of physical chemistry letters*, 2015, **6**, 3528–3533.
- 29 S. Chakraborty, W. Xie, N. Mathews, M. Sherburne, R. Ahuja, M. Asta and S. G. Mhaisalkar, *ACS Energy Letters*, 2017, **2**, 837–845.
- 30 L. Ward, R. Liu, A. Krishna, V. I. Hegde, A. Agrawal, A. Choudhary and C. Wolverton, *Physical Review B*, 2017, **96**, 024104.
- 31 J. H. Montoya and K. A. Persson, *npj Computational Materials*, 2017, **3**, 14.
- 32 J. R. Boes, O. Mamun, K. Winther and T. Bligaard, *The Journal of Physical Chemistry A*, 2019, **123**, 2281–2285.
- 33 J. P. Janet and H. J. Kulik, *J. Phys. Chem. A*, 2017, **121**, 8939–8954.
- 34 S. Curtarolo, G. L. Hart, M. B. Nardelli, N. Mingo, S. Sanvito and O. Levy, *Nature materials*, 2013, **12**, 191.
- 35 S. Back, J. Yoon, N. Tian, W. Zhong, K. Tran and Z. W. Ulissi, *The journal of physical chemistry letters*, 2019, **10**, 4401–4408.
- 36 S. Dahl, A. Logadottir, R. Egeberg, J. Larsen, I. Chorkendorff, E. Törnqvist and J. K. Nørskov, *Physical Review Letters*, 1999, **83**, 1814.
- 37 B. Hammer, Y. Morikawa and J. K. Nørskov, *Physical review letters*, 1996, **76**, 2141.
- 38 H. Xin, A. Vojvodic, J. Voss, J. K. Nørskov and F. Abild-Pedersen, *Phys. Rev. B*, 2014, **89**, 115114.
- 39 A. P. Bartók, R. Kondor and G. Csányi, *Physical Review B*, 2013, **87**, 184115.
- 40 V. I. Anisimov, J. Zaanen and O. K. Andersen, *Physical Review B*, 1991, **44**, 943.
- 41 L. T. Roling, L. Li and F. Abild-Pedersen, *The Journal of Physical Chemistry C*, 2017, **121**, 23002–23010.
- 42 L. T. Roling, T. S. Choksi and F. Abild-Pedersen, *Nanoscale*, 2019, **11**, 4438–4452.
- 43 V. Streibel, T. S. Choksi and F. Abild-Pedersen, *The Journal of Chemical Physics*, 2020, **152**, 094701.
- 44 T. S. Choksi, V. Streibel and F. Abild-Pedersen, *The Journal of Chemical Physics*, 2020, **152**, 094702.
- 45 F. Calle-Vallejo, J. Tymoczko, V. Colic, Q. H. Vu, M. D. Pohl, K. Morgenstern, D. Loffreda, P. Sautet, W. Schuhmann and A. S. Bandarenka, *Science*, 2015, **350**, 185–189.
- 46 F. Calle-Vallejo, J. I. Martínez, J. M. García-Lastra, P. Sautet and D. Loffreda, *Angewandte Chemie International Edition*, 2014, **53**, 8316–8319.
- 47 A. S. Bazhenov, L. Lefferts and K. Honkala, *The Journal of Physical Chemistry C*, 2017, **121**, 4324–4331.
- 48 J. R. Boes, G. Gumuslu, J. B. Miller, A. J. Gellman and J. R. Kitchin, *ACS Catalysis*, 2015, **5**, 1020–1026.
- 49 X. Ma and H. Xin, *Physical review letters*, 2017, **118**, 036101.
- 50 E. Prodan and W. Kohn, *Proceedings of the National Academy of Sciences*, 2005, **102**, 11635–11638.
- 51 J. K. Nørskov, T. Bligaard, B. Hvolbæk, F. Abild-Pedersen, I. Chorkendorff and C. H. Christensen, *Chemical Society Reviews*, 2008, **37**, 2163–2171.
- 52 Y. Volokitin, J. d. Sinzig, L. De Jongh, G. Schmid, M. Vargaftik and I. Moiseevi, *Nature*, 1996, **384**, 621.
- 53 L. Li, A. H. Larsen, N. A. Romero, V. A. Morozov, C. Glinsvad, F. Abild-Pedersen, J. Greeley, K. W. Jacobsen and J. K. Nørskov, *The journal of physical chemistry letters*, 2013, **4**, 222–226.
- 54 I. V. Yudanov, R. Sahnoun, K. M. Neyman and N. Rösch, *The Journal of chemical physics*, 2002, **117**, 9887–9896.
- 55 J. Kleis, J. Greeley, N. Romero, V. Morozov, H. Falsig, A. H. Larsen, J. Lu, J. J. Mortensen, M. Dułak, K. S. Thygesen, J. K. Nørskov and K. W. Jacobsen, *Catalysis Letters*, 2011, **141**, 1067–1071.
- 56 M. M. Kauppinen, V. Korpelin, A. M. Verma, M. M. Melander and K. Honkala, *The Journal of chemical physics*, 2019, **151**, 164302.
- 57 W. Bouwen, F. Vanhoutte, F. Despa, S. Bouckaert, S. Neukermans, L. T. Kuhn, H. Weidele, P. Lievens and R. E. Silverans, *Chemical physics letters*, 1999, **314**, 227–233.
- 58 R. Jinnouchi and R. Asahi, *The journal of physical chemistry letters*, 2017, **8**, 4279–4283.
- 59 R. Jinnouchi, H. Hirata and R. Asahi, *The Journal of Physical Chemistry C*, 2017, **121**, 26397–26405.
- 60 Z. Zhao, P. Schlexer Lamoureux, A. Kulkarni and M. Bajdich, *ChemCatChem*, 2019, **11**, 3423–3431.
- 61 R. Gasper, H. Shi and A. Ramasubramaniam, *The Journal of Physical Chemistry C*, 2017, **121**, 5612–5619.
- 62 Z. Li, X. Ma and H. Xin, *Catalysis Today*, 2017, **280**, 232–238.
- 63 M. O. Jäger, E. V. Morooka, F. F. Canova, L. Himanen and A. S. Foster, *npj Computational Materials*, 2018, **4**, 37.
- 64 Z. Li, S. Wang, W. S. Chin, L. E. Achenie and H. Xin, *Journal of Materials Chemistry A*, 2017, **5**, 24131–24138.
- 65 Z. W. Ulissi, A. J. Medford, T. Bligaard and J. K. Nørskov, *Nature communications*, 2017, **8**, 14621.

- 66 A. N. Alexandrova and A. I. Boldyrev, *Journal of chemical theory and computation*, 2005, **1**, 566–580.
- 67 H. Zhai and A. N. Alexandrova, *Journal of chemical theory and computation*, 2016, **12**, 6213–6226.
- 68 A. N. Alexandrova, *The Journal of Physical Chemistry A*, 2010, **114**, 12591–12599.
- 69 B. G. Sumpter and D. W. Noid, *Chemical physics letters*, 1992, **192**, 455–462.
- 70 P. Schlexer Lamoureux, K. T. Winther, J. A. Garrido Torres, V. Streibel, M. Zhao, M. Bajdich, F. Abild-Pedersen and T. Bligaard, *ChemCatChem*, 2019, **11**, 3581–3601.
- 71 L. B. Vilhelmsen and B. Hammer, *The Journal of chemical physics*, 2014, **141**, 044711.
- 72 H. Zhai and A. N. Alexandrova, *Fluxionality of catalytic clusters: when it matters and how to address it*, 2017.
- 73 G. Sun, P. Sautet and A. Alexandrova, 2019 North American Catalysis Society Meeting, 2019.
- 74 G. Sun and P. Sautet, *Journal of the American Chemical Society*, 2018, **140**, 2812–2820.
- 75 C. Rasmussen, *CKI Williams Gaussian processes for machine learning*, 2006.
- 76 T. Chen and C. Guestrin, Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, 2016, pp. 785–794.
- 77 K. Jacobsen, P. Stoltze and J. Nørskov, *Surface Science*, 1996, **366**, 394 – 402.
- 78 E. M. Dietze, F. Abild-Pedersen and P. N. Plessow, *The Journal of Physical Chemistry C*, 2018, **122**, 26563–26569.
- 79 X. Liu, P. Schlexer, J. Xiao, Y. Ji, L. Wang, R. B. Sandberg, M. Tang, K. S. Brown, H. Peng, S. Ringe *et al.*, *Nature communications*, 2019, **10**, 1–10.
- 80 P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo *et al.*, *Journal of physics: Condensed matter*, 2009, **21**, 395502.
- 81 A. H. Larsen, J. J. Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus *et al.*, *Journal of Physics: Condensed Matter*, 2017, **29**, 273002.
- 82 J. Wellendorff, K. T. Lundgaard, A. Møgelhøj, V. Petzold, D. D. Landis, J. K. Nørskov, T. Bligaard and K. W. Jacobsen, *Physical Review B*, 2012, **85**, 235149.
- 83 D. Vanderbilt, *Physical review B*, 1990, **41**, 7892.
- 84 H. J. Monkhorst and J. D. Pack, *Physical review B*, 1976, **13**, 5188.
- 85 L. Bengtsson, *Phys. Rev. B*, 1999, **59**, 12301–12304.
- 86 B. Hammer, L. B. Hansen and J. K. Nørskov, *Physical review B*, 1999, **59**, 7413.