ACS Combinatorial Science

pubs.acs.org/acscombsci

Research Article

Efficient Machine-Learning-Aided Screening of Hydrogen Adsorption on Bimetallic Nanoclusters

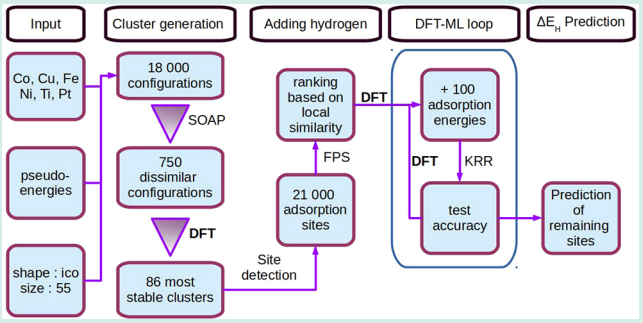
Marc O. J. Jäger,* Yashasvi S. Ranawat, Filippo Federici Canova, Eiaki V. Morooka, and Adam S. Foster

Cite This: *ACS Comb. Sci.* 2020, 22, 768−781

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Nanoclusters add an additional dimension in which to look for promising catalyst candidates, since catalytic activity of materials often changes at the nanoscale. However, the large search space of relevant atomic sites exacerbates the challenge for computational screening methods and requires the development of new techniques for efficient exploration. We present an automated workflow that systematically manages simulations from the generation of nanoclusters through the submission of production jobs, to the prediction of adsorption energies. The presented workflow was designed to screen nanoclusters of arbitrary shapes and size, but in this work the search was restricted to bimetallic icosahedral clusters and the adsorption was exemplified on the hydrogen evolution reaction. We demonstrate the efficient exploration of nanocluster configurations and screening of adsorption energies with the aid of machine learning. The results show that the maximum of the *d*-band Hilbert-transform ϵ_u is correlated strongly with adsorption energies and could be a useful screening property accessible at the nanocluster level.



KEYWORDS: *nanoclusters, adsorption, machine learning, hydrogen evolution reaction, catalysis, workflow automation, computational screening*

INTRODUCTION

Excess electricity in the grid can be stored in a long-term energy carrier such as hydrogen, via electrolysis of water.[1,2] The energy is released either directly back into the electric grid at times of high demand, in fuel cell vehicles, or through the gas grid for heating.[2−4] However, the production of hydrogen through electrolysis is generally more expensive than from natural gas, oil, or coal, that means it is not a competitive energy carrier yet.[2] Since the electricity accounts for 70−90% of the cost, the situation can change in the future when more volatile electricity prices are expected.[5−7] Water splits electrolytically into hydrogen at the cathode via the hydrogen evolution reaction (HER) and oxygen at the anode via the oxygen evolution reaction (OER).[8−10] Whereas OER catalysts are mostly transition metals in oxidized states,[11] HER catalysts normally contain reduced transition metals and platinum is often used commercially.[8] Platinum plays a key role as its price and availability can influence whether a hydrogen-based energy carrier can be deployed at a competitive price. In order to reduce prices, the platinum content has been minimized in the past decade, but catalyst loading still makes up a significant portion of the proton-exchange membrane (PEM) electrolyzer cost.[12] There is a high risk associated with the supply of platinum due to (a) the high volatility of the price and (b) the scarce resources being mined only in few countries worldwide.[13] Hence, its reduction or replacement remains a key component to producing hydrogen via electrolysis competitively.[14]

Several potential new catalysts with reduced or zero content of platinum group metals (PGMs) are being considered, ranging from alloys which are dispersed nanoparticles, highly diluted (single atom alloys) or high-entropy compounds with a nonmetallic component such as phosphides, carbides, nitrides, and chalcogenides.[15−18] When picking a new potential catalyst one is spoilt for choice from a large array of compound classes with varying compositions. While experimentally screening a series of novel materials is time-consuming and expensive, their properties can be approximated by high-throughput simulations.[19−22]

However, in order to account for the kinetics and adsorbate coverage of catalysts, extensive ab initio simulations are necessary to calculate the theoretical catalytic activity. Requiring information about transition states and adsorbate−

Received: May 26, 2020
Revised: September 21, 2020
Published: November 4, 2020

https://dx.doi.org/10.1021/acscombsci.0c00102
ACS Comb. Sci. 2020, 22, 768−781

adsorbate interactions, it becomes prohibitively computationally demanding to screen a large data set. Instead, properties which correlate with catalytic activity but are cheaper to compute are preferable. Exploiting the Sabatier principle, the adsorption energy as a surrogate property allowed scaling relations and *d*-band theory to predict trends in catalytic activity.[23−26] Moreover, properties such as bond lengths, bond angles, and (orbital-wise) coordination numbers were used to describe catalyst structures and correlate them to adsorption energies.[27−29]

However, the correlation trends between adsorption energies and electronic descriptors have not been investigated in the nanocluster regime. An electronic descriptor which requires only DFT calculations of the nanocluster and correlates linearly with the adsorption energy would drastically reduce the computation effort to predict catalytic activity. Trying to find such a descriptor is challenging and in this work we explore the possibilities, starting with existing descriptors such as ϵ_d, ϵ_d^w, and ϵ_u. Alternatively, adsorption energies of nanoclusters can be obtained via machine learning at a fraction of the computational cost. To facilitate this, we have developed a set of open source tools that allow the generation and analysis of large sets of nanoclusters and the prediction of adsorption energies.

In this work, we investigate trends of HER on bimetallic nanoclusters based on simulations of the hydrogen adsorption energy. We generated a data set containing platinum and five different transition metals which form icosahedra. We restricted the search space to icosahedral structures of a fixed size (55 atoms). We begin by describing how the screening process is done efficiently with the aid of machine learning, and how the automated workflow was designed. We then analyze the nanocluster stability and report the machine learning accuracy and computational efficiency. We further compare electronic descriptors with each other and with adsorption energy distributions. We conclude by comparing the descriptors to experimental and other simulated data sets.

METHODS

Adsorption energy. The adsorption energy ΔE_H is readily available through total energy simulations:

$$\Delta E_H = E_{cluster+H} - E_{cluster} - \frac{1}{2}E_{H_2} \quad (1)$$

where $E_{cluster+H}$, $E_{cluster}$, and E_{H_2} are the total energies of the nanocluster with one adsorbed hydrogen, the nanocluster alone and the hydrogen molecule in the gas phase. From the adsorption energy alone the catalytic activity via a microkinetic model[30] is not accessible. It requires additional computational information, in particular: (i) activation energy, (ii) free energy correction, and (iii) equilibrium adsorbate coverage. Since these properties are orders of magnitude more expensive to calculate, methods to circumvent them are necessary.

The activation energy can be mitigated thanks to the Bell-Evans−Polanyi principle which states that activation energies of the same reaction family depend linearly on the reaction energy.[31] As a direct conclusion, according to the Sabatier principle, adsorption sites with a change in Gibbs free energy $\Delta G_H \approx 0$ are expected to have the highest catalytic activity.[23,24] Furthermore, the free energy can be approximated with various degrees of accuracy ranging from a constant shift to molecular dynamics simulations.[32,33] Lastly, the equilibrium adsorbate coverage is difficult to approximate. Closing this gap via experimental data or expensive simulations is paramount in order to get a proper estimate of the catalytic activity. However, given only the adsorption energy, it is at least possible to determine trends in similar systems (under the assumption that coverage remains similar). It is also possible to bridge the gap by machine learning a correlation between adsorption energies and catalytic activity.[34]

Density Functional Theory calculations. Ab initio simulations were performed using the Density Functional Theory (DFT) as implemented in the CP2K package[35] using Gaussian and planewave (GPW) basis sets and the spin-polarized GGA-functional by Perdew−Burke−Ernzerhof (PBE).[36] All atom types had short-ranged double-ζ valence plus polarization molecularly optimized basis sets (MOLOPT-SR-DZVP)[37] and norm-conserving Goedecker-Teter-Hutter (GTH) pseudopotentials.[38−40] The D3 method of Grimme et al. with Becke-Johnson damping (DFT-D3(BJ))[41,42] accounted for Van-der-Waals interactions. The energy cutoff for the auxiliary PW basis and the reference grid was set to 600 and 60 Ry, respectively. Atomic positions were relaxed with the Broyden−Fletcher−Goldfarb−Shanno (BFGS) optimizer until the maximum force component converged to 0.02 eV/Å. The simulation box was 2.5 times the diameter of the nanocluster, resulting in a gap of approximately 15 Å vacuum.

Descriptors. The local atomic and electronic environment of adsorption sites are characterized using various descriptors in this work, providing unique *fingerprints* that are useful for comparison. If the property is derived from the atomic structure of the system, we refer to them as structural descriptors, if the property is derived from the electron density, we refer to them as electronic descriptors.

For structural descriptors, Smooth Overlap of Atomic Positions (SOAP) performed best on nanoclusters in a previous work[43] and here we briefly outline its properties. SOAP is a structural descriptor that overlaps Gaussian-smeared atomic positions in space and maps them to coefficients of orthonormal basis functions.[44,45] Being a local descriptor, SOAP facilitates comparing environments around an adsorption site (within a certain radial cutoff R_c) of different nanoclusters or within the same nanocluster. SOAP can also compare full structures for, for exaample, stability by matching several local environments with each other.[45]

Since the chemical reactivity of transition metals is determined by the *d*-band, electronic descriptors focus on its position and shape. The *d*-band theory links the energy distribution of the transition metal *d*-band with the adsorption strength.[24,46,47] The *d*-band center was introduced as the first catalyst descriptor, since adsorption energies scaled linearly with the filling of the *d*-band. It has exceptions in alloys with almost filled *d*-orbitals (d^9 or d_{10}) where the peak of the hydrogen-surface antibonding state traverses the Fermi-level.[46,48]

Following the conclusion that mostly the shape of the *d*-band at the Fermi-level would determine the position of the hydrogen-surface antibonding state, by adding half of the *d*-bandwidth W_d, the descriptor $\epsilon_d^w = \epsilon_d + \frac{W_d}{2}$ is brought closer to the *d*-band edge near the Fermi-level.[49] The most recent maximum of the Hilbert-transform ϵ_u as a direct measure of the *d*-band edge was reported to correlate best with adsorption energies on late transition metal alloys.[50]

https://dx.doi.org/10.1021/acscombsci.0c00102
ACS Comb. Sci. 2020, 22, 768−781

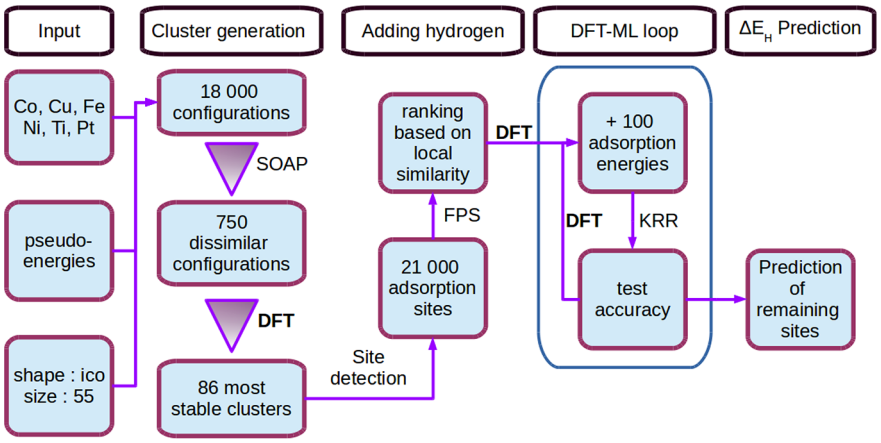


Figure 1. Workflow sketch shows the detailed steps from cluster generation to the prediction of the adsorption energy distributions. FPS stands for farthest point sampling.

The above electronic descriptors are simple and intuitive, yet are prone to information loss due to their one-dimensional nature. In order to retain information for machine learning, an accurate description of the local density of states (LDOS) of the *d*-band of a surface atom is necessary. LDOS alone, however, cannot compete with the accuracy of SOAP,[51] since it is a discrete representation with few features, containing only indirect structural information. A combination of SOAP and LDOS proved to be an effective spatial and electronic descriptor.[51] A way of describing LDOS is through its moments, defined as

$$\mu_n(i) = \int_{E_F-\Delta}^{E_F+\Delta} \mathrm{d}E(E - E_F)^n \mathrm{LDOS}^{(i)}(E) \tag{2}$$

where Δ is the LDOS range from the Fermi energy E_F. The total kernel of SOAP plus LDOS amounts to

$$k(i,j) = k_{\mathrm{SOAP}}(i,j) \prod_{n=0}^{n_{\max}} \exp^{-\gamma_n(\mu_n(i)-\mu_n(j))^2} \tag{3}$$

where k_{SOAP} is the radial-basis-function kernel derived from the SOAP descriptor.

Machine Learning. The relatively cheap evaluation of the adsorption energy enables high-throughput DFT calculations on large data sets. However, to screen even more efficiently, DFT calculations can be supplemented by interpolating data with machine learning.[52] Since the cost of machine learning is negligible compared to DFT calculations, it can reduce the computational cost by an order of magnitude or more, depending on the data set.

Input properties required for machine learning models are ideally available without heavy computation from the relaxed nanoclusters without adsorbates. Given such an inexpensive property we predict more expensive ones, such as adsorption energy distributions or catalytic activities from simulations or experiments. The local environment around the initial guess of the adsorbate was encoded with the SOAP descriptor, calculated with the Dscribe package.[53] The cutoff radius is the only hyperparameter from the descriptor side. Adsorption energies were predicted through Kernel Ridge Regression (KRR) using the radial-basis-function kernel. Adsorption energies of top, bridge and hollow sites were acquired in a ML-DFT loop (see workflow below) whereas the data size increased by 100 DFT calculations in each loop. The data set was randomly split into 80% training and 20% test set. The kernel hyperparameters (regularization parameter $\alpha = 0.01$ and length-scale $\gamma = 10^{-6}$ of the kernel) alongside the cutoff radius $r_c = 7.0$ Å were optimized via 5-fold cross-validation grid search.

We computed LDOS up to the sixth moment $n_{\max} = 6$ (see eq 3). The LDOS range Δ was set to 3 eV. For each adsorption site except for top sites, the LDOS moments were averaged over the transition metal atoms of the site. LDOS and SOAP were combined to a descriptor using the kernel product in eq 3. As above, the same data set differing only by the enhanced descriptor SOAP+LDOS was split with a ratio 80:20. The hyperparameters α (regularization parameter), the kernel length-scales γ_{SOAP}, γ_2 to γ_6 as well as the cutoff radius r_c were optimized via the Nelder–Mead method using the 5-fold cross-validation MAE as its optimization target.

Workflow. Since the search space was large, it was important to develop a repeatable and scalable workflow which requires minimal maintenance. From the choice of elements to the prediction of adsorption energies, the process was automated as much as possible. The steps in the workflow range from cluster generation, adsorption site detection, DFT to machine learning. The full workflow is sketched in Figure 1.

The nanocluster data set is composed of binary combinations of the elements Pt, Ti, Fe, Co, Ni, and Cu with compositions of A_nB_{55-n} ($n \in \{0, 6, 13, 27, 42, 49, 55\}$). The size and shape are kept constant at icosahedra of 55 atoms. The above elements were chosen, since—with the exception of platinum—they form stable icosahedra at 55 atoms[54] and are studied in experiments for their catalytic potential. For icosahedral shapes, 55 is a magic number, resulting in 2 shells around a center atom. Upon testing 13-atom clusters (a single shell around a center atom), we noticed a significant surface reconstruction, which made analyses more difficult. Larger clusters are out of the scope of this study due to computational limitations, but also we chose nanoclusters as small as possible in order to magnify the size effects. With the additional shape constraint, the problem complexity was further reduced.

First, the configuration space of the bimetallic icosahedra was sampled on a grid of different combinations of pseudoenergies in Monte Carlo runs of 1000 steps. The Monte Carlo-threshold was set to $e^{E_1-E_0/0.2}$. The pseudoenergy was defined as

https://dx.doi.org/10.1021/acscombsci.0c00102
ACS Comb. Sci. 2020, 22, 768−781

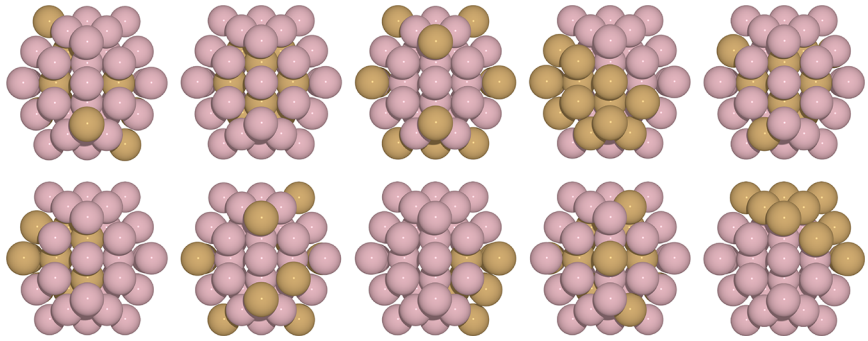


Figure 2. Clusters of a given composition (e.g., the depicted $Cu_{13}Co_{42}$) were generated automatically by Monte Carlo assuming various combinations of interaction and segregation energies. Experimentally observable composites such as core−shell, segregated, ordered, and random as well as structures in-between emerged naturally.

$$E = \sum_{i}^{n} E_{c,i} \times c_i + \sum_{i}^{n} \sum_{j>i}^{n} E_{x,i,j} \tag{4}$$

where E_c characterizes the element-specific attraction to the core and E_x stands for the interaction pseudoenergy between atoms of the same or different types. For a binary nanocluster the pseudoenergies amount to five parameters. Interactions were only counted where atoms were connected through Voronoi tesselation. For each bimetallic composition the 10 most dissimilar structures were selected from 243 gridpoints. The clusters were compared by averaging over the atomic environments expressed by the descriptor SOAP. The process of cluster generation, in particular sampling configurations and compositions of a fixed scaffold, was automated such that a diverse set of clusters emerged ranging from core−shell, to randomly distributed and segregated clusters. This selection resulted in a data set of 750 bimetallic and 6 pure nanoclusters. The selected clusters of the composition $Cu_{13}Co_{42}$ are shown in Figure 2 as an excerpt. Before adding adsorbates to them, the above 756 nanoclusters were relaxed by DFT.

Determining adsorption sites can be especially challenging when a surface is irregular and requires initially determining all the atoms that belong to the surface *class*. Hence, the volume of the clusters were compartmentalized by Delaunay tetrahedralization and, subsequently, surface atoms along the outermost tetrahedral faces were detected. On the surface top, bridge and hollow sites were defined as the atom position (top), middle point (bridge), and center (hollow) of each triangular face. The adsorption vector or the binding direction of the initial guess of the adsorbate was set to the average of outward-pointing normal vectors of triangles containing the site point. This procedure depicted in Figure 3 resulted in well-defined sites of arbitrarily shaped nanoclusters without the requirement of visual inspection.

Even with irregular shapes or after surface reconstruction it was still possible to identify sites consistently. Adsorption sites were classified on the surface of DFT-relaxed clusters. Given the most stable structures of each composition, the workflow identified around 21 000 adsorption sites. Since those contained many redundancies, randomly picking training points to calculate via DFT for machine-learning is inefficient. The sites were ranked based on similarity of their local environment using farthest-point sampling.[55] We chose kernel ridge regression as our machine learning model because it proved to work well in a smooth feature space such as SOAP.[43,56] The adsorption energy was predicted given the local environment of the initial position of the adsorbate.

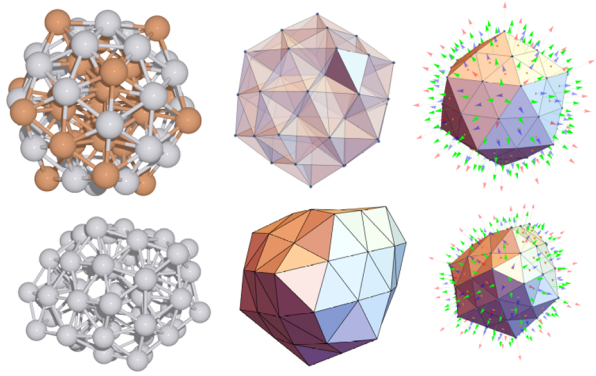


Figure 3. Through Delaunay tetrahedralization the whole surface is triangulated and surface atoms are detected. Adsorption vectors of top, bridge, and hollow sites are defined as the average of outward-pointing normal vectors of surface triangles containing the site point.

Starting from the relaxed nanoclusters, from adsorption site detection via ranking to DFT calculations and machine learning, the whole workflow was automated with the aid of the workflow manager Fireworks.[57] Upon completion of several DFT calculations, episodically, the prediction accuracy of the machine learning model was tested to see if the convergence criterion was met. The workflow along with others is available in the public domain.[58] The tools for generating clusters and detecting adsorption sites were compounded in the python package Cluskit,[59] a library devised to build cluster-adsorbate structures.

RESULTS

Nanocluster Stability and Electronic Descriptors. Except for platinum, the pure nanoclusters form stable icosahedra.[54] The excess energy of a nanocluster A_nB_{55-n} (A,B = Fe, Co, Ni, Cu, Ti, Pt) is defined as in ref:[60]

$$E_{exc} = \frac{E_{A_nB_{55-n}} - \frac{n}{55}E_{A_{55}} - \frac{n-55}{55}E_{B_{55}}}{55} \tag{5}$$

where E_x denotes the total energy of the systems A_nB_{55-n}, A_{55}, and B_{55}. For every composition, there are 10 nanocluster configurations. The excess energies along with an analysis on core−shell and segregation distribution with respect to stability can be found in the Supporting Information (SI).

To summarize, all bimetallic systems form a convex hull except for FeCu and CoCu, where the compositions $Fe_{28}Cu_{27}$ and $Co_{28}Cu_{27}$ were slightly above. The energy gap between the

https://dx.doi.org/10.1021/acscombsci.0c00102
ACS Comb. Sci. 2020, 22, 768−781

Table I. Descriptors ϵ_d, ϵ_d^w and ϵ_u for a Given Bimetallic Composition Depend on the Cluster Configuration and the Environment of the Respective Surface Atom Sites[a]

	σ_{rel}							*R*-value						
ϵ_d	Fe	Co	Ni	Cu	Ti	Pt	all	Fe	Co	Ni	Cu	Ti	Pt	all
Fe		0.59	0.29	0.17	0.11	0.17	0.26		**0.99**	**0.95**	**0.94**	0.84	**0.90**	**0.92**
Co	0.52		0.22	0.14	0.11	0.13	0.22	**-0.99**		**0.97**	**0.91**	0.74	**0.92**	0.51
Ni	0.40	0.31		0.17	0.13	0.26	0.25	**-0.92**	**-0.92**		0.87	−0.69	**0.98**	−0.14
Cu	0.16	0.15	0.15		0.09	0.17	0.15	−0.89	−0.86	**-0.94**		−0.72	0.82	−0.52
Ti	0.36	0.46	0.36	0.20		0.38	0.35	−0.72	nss	nss	0.82		0.86	0.32
Pt	0.34	0.29	0.44	0.29	0.20		0.31	−0.85	−0.79	**-0.96**	−0.89	−0.83		−0.87
ϵ_d^w	Fe	Co	Ni	Cu	Ti	Pt	all	Fe	Co	Ni	Cu	Ti	Pt	all
Fe		0.37	0.37	0.24	0.16	0.17	0.26		**0.98**	**0.94**	0.86	**0.91**	**0.92**	**0.92**
Co	0.36		0.29	0.20	0.11	0.14	0.22	**-0.93**		**0.94**	0.88	0.80	**0.95**	0.53
Ni	0.51	0.42		0.23	0.14	0.25	0.31	**-0.95**	**-0.95**		0.78	nss	**0.97**	−0.04
Cu	0.19	0.20	0.22		0.08	0.19	0.17	−0.86	−0.84	**-0.91**		nss	0.74	−0.47
Ti	0.38	0.37	0.34	0.17		0.33	0.32	−0.82	nss	0.69	0.71		0.85	0.36
Pt	0.32	0.31	0.43	0.31	0.20		0.31	−0.07	0.22	−0.84	−0.81	nss		−0.38
ϵ_u	Fe	Co	Ni	Cu	Ti	Pt	all	Fe	Co	Ni	Cu	Ti	Pt	all
Fe		0.16	0.07	0.04	0.13	0.03	0.08		−0.76	−0.82	**-0.93**	nss	nss	−0.84
Co	0.22		0.27	0.03	0.12	0.03	0.13	0.88		−0.82	**-0.91**	0.65	0.78	0.11
Ni	0.18	0.89		0.04	0.07	0.03	0.24	0.77	nss		nss	0.77	nss	0.77
Cu	0.13	0.16	0.20		0.05	0.14	0.13	−0.76	nss	0.88		nss	−0.83	−0.24
Ti	0.03	0.02	0.02	0.03		0.02	0.02	nss	−0.69	nss	−0.88		0.67	−0.30
Pt	0.29	0.30	0.33	0.30	0.13		0.27	0.69	nss	0.72	nss	0.82		0.74

[a]Relative standard deviations σ_{rel} of these descriptors are computed for each composition and then averaged. The described surface atom type is shown on the left and all bimetallic combinations are averaged on the last column. The correlation coefficients *R* on the right side correlate the above descriptors with the stability of the nanoclusters in the same fashion (bold font indicating strong correlation). The short nss stands for not statistically significant. More detailed correlation ellipses for each composition are shown in the SI.

lowest and second lowest energy structure was lower than 0.05 eV for all compositions. The gap tended to increase toward equiatomic compositions in accordance with the total permutations $\binom{55}{n}$. The accuracy of the convex hull could further be improved by increasing the number of maximally different configurations for each composition (see SI).

The most stable nanoclusters were selected to determine the adsorption energy distributions. Along with 81 lowest-energy structures, 5 structures had a boltzmann-factor $\frac{p_i}{p_0} = e^{-E_i - E_0/kT}$ greater than 0.01 at room temperature with respect to the lowest-energy structure, which amounted to a total of 86 nanoclusters. The observed segregation was in good agreement with the miscibility analysis of Zhang et al.[61] In general, the formation of core−shell structures as the most stable configurations agreed well with the literature.[62]

The most successful electronic descriptors for periodic slabs that provide an intuitive property of surface atoms are the *d*-band center ϵ_d,[24,46,47] the *d*-band center plus half the *d*-bandwidth ϵ_d^w[49] and the maximum of the *d*-band Hilbert-transform ϵ_u.[50] Examining the full set of nanoclusters, for each composition the nanoclusters could exhibit low to high variance in electronic descriptors. The descriptors depend on the cluster configuration and the environment of the respective surface atom sites. The extent of this variance as well as the correlation with nanocluster stability is summarized in Table I.

Relative standard deviations σ_{rel} of ϵ_d, ϵ_d^w, and ϵ_u are computed for each composition and then averaged. The surface atom type (from which the LDOS and descriptor is evaluated) constitutes the row and the other atom type constitutes the column. For bimetallic systems with a high σ_{rel}, for example, NiFe, NiCo, TiFe, or TiPt, a dense nanocluster sampling could be important. To cover the range of clusters with different electronic properties, it is important to select multiple low-energy clusters with dissimilar configurations. The relative standard deviations were lowest for ϵ_u among the descriptors for iron and titanium sites.

The correlation coefficients *R* on the right side reveal a trend between the electronic descriptors and the stability of the nanoclusters. If the correlation coefficient *R* are close to 1 (−1) and σ_{rel} is high, the property is expected to drop (rise) when the number of sampled cluster configurations increases, since the convex hull will lower and the descriptor will change alongside it. In such cases, a dense sampling is important, as too sparse sampling could introduce a systematic under- or overestimation of the electronic descriptor. The correlation coefficient was strong for systems containing iron, cobalt, or nickel. The *d*-band center ϵ_d showed a trend in the sign of *R*. In the series Fe-Co-Ni-Cu-Pt, elements on the left of the LDOS atom would increase its *d*-band center, whereas elements on the right would decrease it upon converging to more stable nanoclusters. All those correlations were moderate to strong. Titanium would fall between cobalt and nickel, but is excluded from the list due to many nonsignificant data points. The same picture emerges for ϵ_d^w although the correlation coefficients of PtFe and PtCo were weak, whereas other combinations were moderate to strong.

A possible explanation for the above trends was that the *d*-band filling decreases when paired with an element toward the left of the periodic table, and opposite for elements on the right. The exception here was titanium as an early transition metal. The consistent positive or negative correlation with stability might stem from a larger overlap of the *d*-orbitals in stable nanoclusters. The maximum of the *d*-band Hilbert-transform ϵ_u did not seem to adhere to the above trend. Since the correlation was not statistically significant for many

https://dx.doi.org/10.1021/acscombsci.0c00102
ACS Comb. Sci. 2020, 22, 768−781

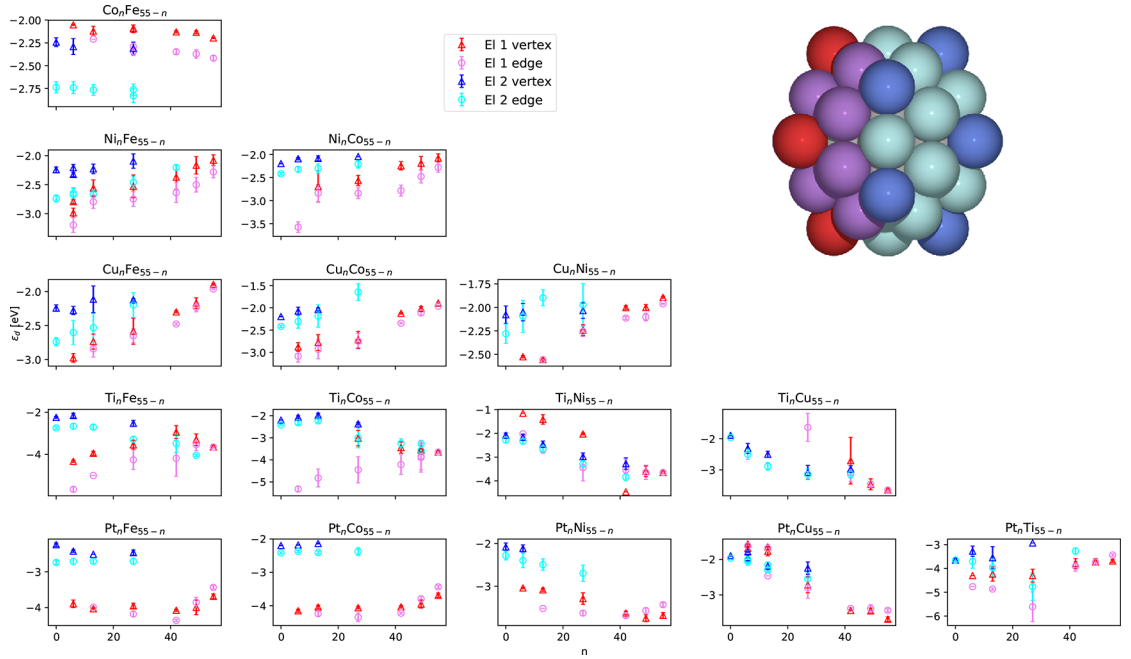


Figure 4. *d*-band center ϵ_d of the 86 most stable nanoclusters. The error bars indicate the standard deviation of the distribution among surface atoms split into edges and vertices.

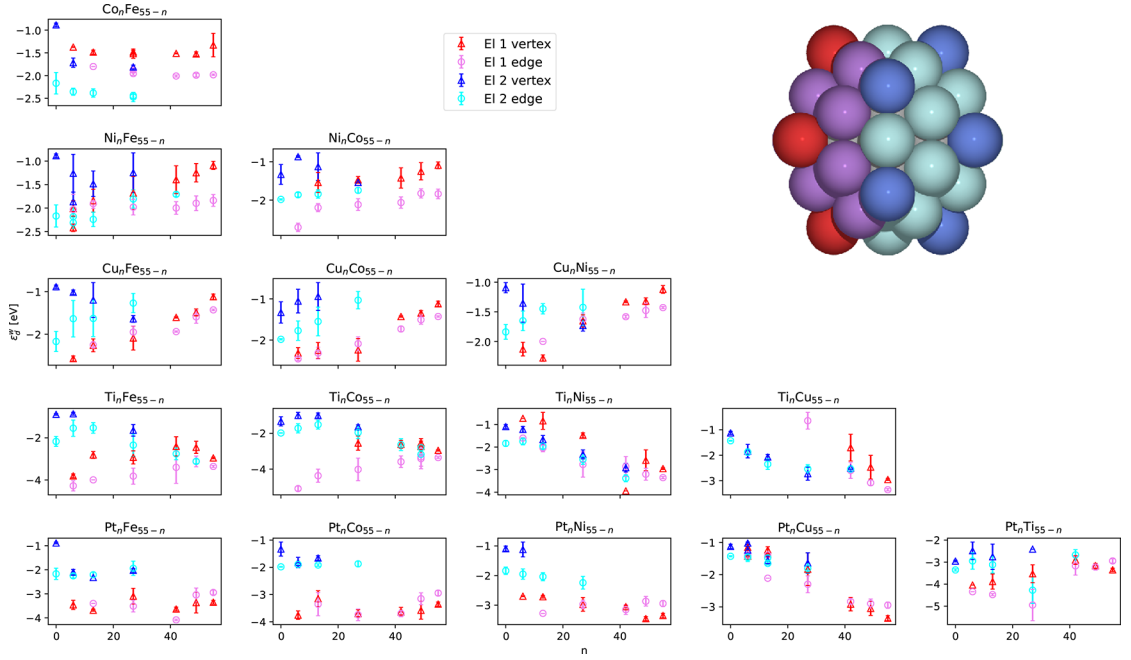


Figure 5. *d*-band center plus half *d*-bandwidth ϵ_d^w of the 86 most stable nanoclusters. The error bars indicate the standard deviation of the distribution among surface atoms split into edges and vertices.

bimetallic systems, a detailed analysis was not possible. However, it can be noted that the shape of the *d*-band near the Fermi-level does not change smoothly with respect to nanocluster stability. A more detailed descriptor distribution with correlation ellipses for each composition can be found in the SI.

Moving to the stable nanoclusters, the above *d*-band properties would still vary among surface atom LDOS within a nanocluster. Due to the small icosahedral shape, there were only two types of surface atoms, edges and vertices. The Figures 4, 5, and 6 depict, given a certain composition of the 86 most stable clusters, the mean and the variance w.r.t surface atoms of ϵ_d, ϵ_d^w, and ϵ_u, respectively.

The error bars show the standard deviation of the descriptor value for one element at a certain distribution. The atoms were further split into vertex and edge atoms. Vertex and edge descriptor values differed significantly with ϵ_d^w and ϵ_d with a relative difference between vertex and edge values of 28% and 19%, respectively. The distribution spread tended to increase toward equi-atomic compositions except for ϵ_u (Figure 6) where the relative difference between vertex and edge values amounted to an average of 4.5%.

https://dx.doi.org/10.1021/acscombsci.0c00102
ACS Comb. Sci. 2020, 22, 768−781

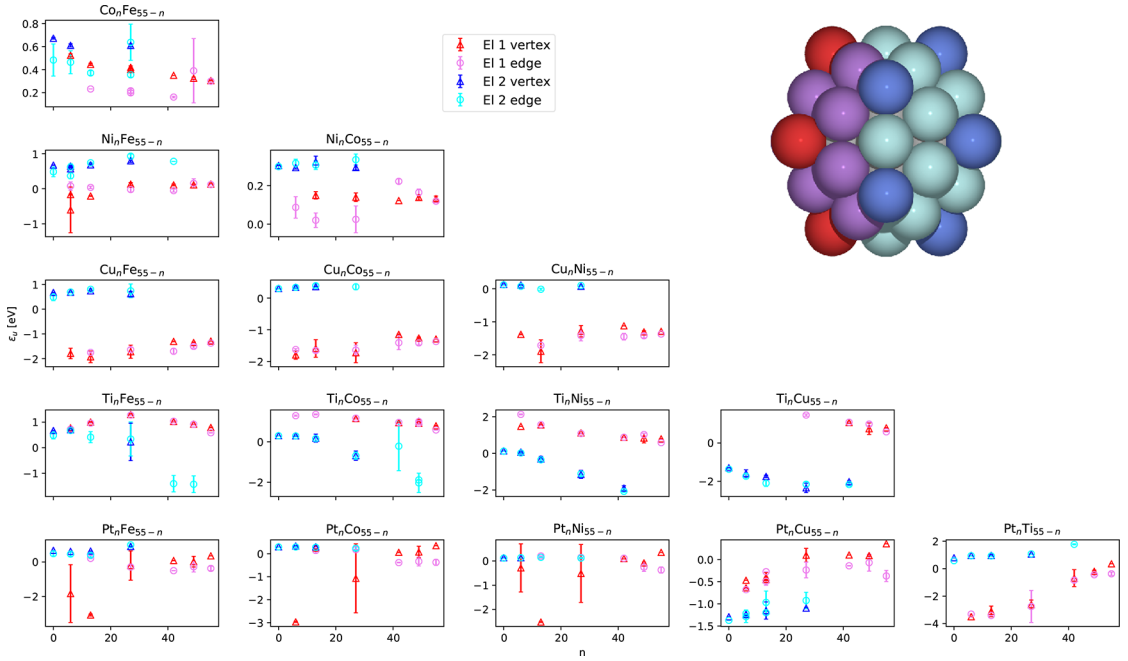


Figure 6. *d*-band maximum of the hilbert-transform ϵ_u of the 86 most stable nanoclusters. The error bars indicate the standard deviation of the distribution among surface atoms split into edges and vertices.

Out of all electronic descriptors ϵ_u had the sharpest distributions for all binary element combinations with a few exceptions at certain compositions. The overall trends for each bimetallic system plot of ϵ_d and ϵ_d^w were the same, but compositions with titanium were shifted downward. The electronic descriptor ϵ_u showed comparable trends to ϵ_d and ϵ_d^w but clearer and with fewer outliers enhancing the difference between elements and damping the difference between edges and vertices. Since the *d*-bandwidth strongly fluctuated among nanoclusters of the same composition, ϵ_d^w turned out not to be as stable as the other descriptor, especially for nanoclusters containing Pt. The SI provides an additional explanation to the broad distribution of ϵ_d, most of it, but not all could be attributed to the atom types of the nearest neighbors around a surface atom.

Machine Learning Precision. In our previous work, we showed that it was possible to efficiently interpolate the potential energy surface of several nanoclusters simultaneously with a limited number of single point calculations.[43] In this work, instead of finding the positions of the adsorbates, we machine learn the adsorption energy of a relaxed cluster-adsorbate structure directly from its initial guess. A similar strategy was applied to adsorption on amorphous carbon.[51] Adsorption energies for different sites on the 86 most stable nanoclusters were predicted with a subsequently increasing training set. Figure 7a shows the learning curve from kernel ridge regression with features from the SOAP descriptor.

Training, validation, and test set are in blue, green, and orange, respectively. By subsequently computing 1767 adsorption energies with DFT, an accuracy of 0.11 eV MAE (mean absolute error) was reached. With almost DFT accuracy we could predict the remaining adsorption energies, hence we required DFT calculations for less than 10% of the sites. The parity plot in Figure 7b shows outliers in the high- and low-energy region. The few data points of low adsorption energies (below −1.3 eV) were predicted worse than average due to under-sampling in the low-energy region.[43] Analyzing further

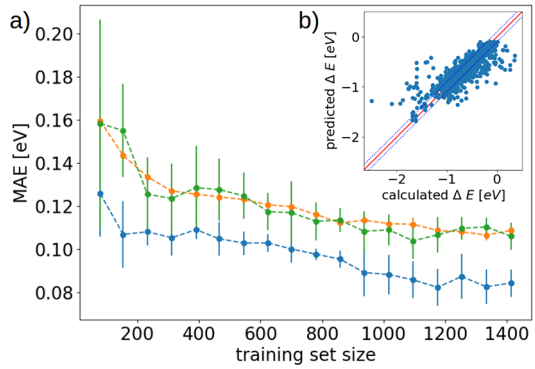


Figure 7. (a) Learning curve of KRR. The errors are averaged over 20 randomized runs and the error bars indicate the standard deviation of those errors. Training, validation and test set are in blue, yellow and green, respectively. (b) Calculated vs predicted hydrogen adsorption energy of 1767 DFT calculations.

where the largest error contributors were, the mean absolute errors in Table II confirmed that there were significant differences in prediction accuracy by element combination.

The errors on copper and titanium adsorption energies were particularly high. The bimetallic alloy CuTi had the largest MAE with 0.28 eV. Different element combinations apparently were machine-learned with different accuracies. Without prior knowledge, the initial ranking of the training points did not reflect that difference. In future studies, that could be solved by an optimization engine, or splitting the workflow into separate runs per bimetallic combination. We also evaluated a combined structural and electronic descriptor of SOAP with moments of the local density of states of a surface atom (LDOS), as described in eq 3.[51] With our labeled data (1767 DFT calculations) from the machine learning with SOAP we were able to achieve an MAE of 0.10 eV on the test set (20%). The validation set MAE amounted to 0.11 eV. The reduction in error was not significant (compare Figure 7).

https://dx.doi.org/10.1021/acscombsci.0c00102
ACS Comb. Sci. 2020, 22, 768−781

Table II. Machine Learning Accuracy by Element[a]

MAE [eV]							
n	Fe	Co	Ni	Cu	Ti	Pt	all
Fe	0.057 4	0.058 133	0.056 194	0.174 133	0.124 148	0.073 104	0.095 716
Co		0.247 2	0.084 113	0.066 140	0.167 145	0.061 99	0.091 632
Ni			0.125 5	0.105 69	0.127 121	0.053 76	0.082 578
Cu				0.148 4	0.276 78	0.093 118	0.134 542
Ti					N.A 0	0.126 79	0.157 571
Pt						0.071 2	0.081 478

[a]The dataset contains the elements Fe, Co, Ni, Cu, Ti, and Pt. The total number of labeled data points *n* was 1767.

As possible contributors to the prediction error, we considered the surface reconstruction, that is, changes in local cluster geometry, when adding the adsorbate and the movement of the adsorbate from its initial guess to a neighboring site during structure relaxation. In Figure 8b the effect of surface reconstruction and adsorption site drift on the machine learning accuracy is visualized. The global SOAP dissimilarity between the nanocluster structure before and after the relaxation with the adsorbate acts as a metric for surface reconstruction. If there was a significant effect of the surface relaxation on the predictive power, one could observe the outliers to have a high SOAP distance metric, which was not uniformly the case. On top of that, the metric average did not increase with increasing error (s. binned mean). Second, the subset of data points in yellow where the adsorption site remained the same after structure relaxation, retained a slightly better accuracy than the full data set, with an MAE of 0.10 eV. The other data points (purple) where the adsorption site changed increased the MAE to 0.12 eV. Hence, surface reconstruction had no apparent effect and adsorption site drift had a small effect on the predictive power of the model. Overall, machine learning systems of different elements at different sampling densities, as well as predicting which sites are stable, could further improve adsorption energy predictions.

Adsorption Energy Distribution. A summary of the predicted adsorption energy distributions is given in Figure 8a. They were divided into 15 bimetallic combinations and 7 compositions.

Each distribution was characterized by its mean and standard deviation. The distribution trends were not trivial. Some curves for a bimetallic combination, for example, CoFe are convex upward, whereas others are convex downward. The distribution is generally broadest at an equi-atomic ratio. The metals platinum and copper retain their narrow distribution as a majority component in CuFe, CuCo, CuNi, and NiPt, CuPt.

Since electronic descriptor data require fewer DFT calculations, it would be advantageous to replace adsorption energy distributions. A perfect linear relation would make the latter obsolete. Figure 9 correlates the adsorption energies of each site with its electronic descriptors ϵ_d, ϵ_u, and ϵ_d^w (from top to bottom). There is a weak correlation between the adsorption energies and the electronic descriptors ϵ_d and ϵ_d^w (a), even after constraining the data to only top sites (b) or sites made up of a single atomic type (c). Only by constraining sites further to pure nanoclusters (d), the correlation increases to moderate.

A strong linear correlation was observed with the descriptor ϵ_u, but platinum and titanium were shifted downward from the linear trend with other descriptors. Titanium as an early transition metal is characterized by strong chemisorption contrary to late transition metals. The unexpected downshift of the platinum *d*-band center could not be explained by the fact that platinum does not prefer icosahedral structure. Compared to the global minimum structure with a reduced core, the

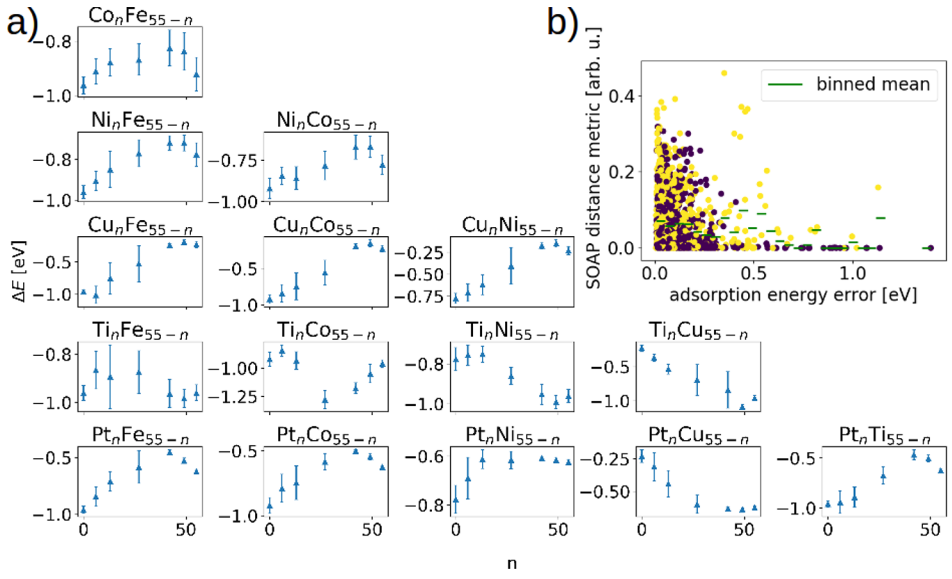


Figure 8. (a) Predicted hydrogen adsorption energy distribution. The mean and standard deviation is given for each composition. (b) Effect of surface reconstruction and adsorption site drift on the machine learning accuracy. Yellow points represent adsorbates retaining their initial adsorption site, purple points represent adsorbates which drifted to neighboring sites. The green bars average the SOAP distance metric over an interval of 0.1 eV adsorption energy error.

https://dx.doi.org/10.1021/acscombsci.0c00102
ACS Comb. Sci. 2020, 22, 768−781

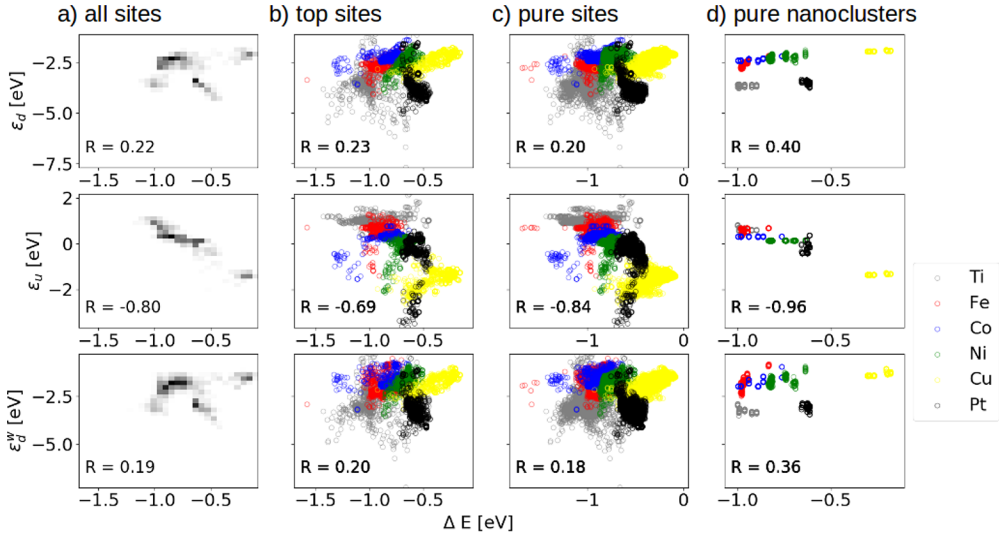


Figure 9. Electronic descriptors of surface atoms which form the adsorption site against the adsorption energy. From left to right, the columns show (a) all adsorption sites, (b) all top sites, (c) only pure adsorption sites (top, bridge, and hollow sites made up of a single atomic type), and (d) only adsorption sites from pure nanoclusters. Each subplot has its own correlation coefficient *R*.

descriptor values do not change significantly (see SI). Apparently, the above electronic descriptors could not replace adsorption energies without loss of information, but the descriptor ϵ_u prevailed as a property, against which nanoclusters could be prescreened.

Mutual Information and Clustering of Nanocluster LDOS. To understand this weak correlation of the adsorption energies with the electronic descriptors, we computed the difference of density of states (DDOS) of the nanocluster, before and after the adsorption process. We used the different adsorption sites on Cu_xPt_{1-x} nanoclusters. The mutual information (MI)[63] between the tensor of DDOS—comprising the *s*, *p*, and *d* states of each atom in the nanocluster, separately—and the adsorption energy was calculated (see SI). On comparison with the MI with the DDOS comprising only the *d* states of each atom in the nanocluster, no significant difference was found (MI of 0.91, in both cases). This implies that no additional information is gained by including the *s* and *p* states, when correlating the LDOS with the adsorption energy, in agreement with earlier studies.[50]

Further, t-distributed stochastic neighbor embedding (t-SNE)[64] plots of the *d*-band of an element in various nanocluster concentrations is shown in Figure 10. t-SNE resolves the distribution of the multidimensional *d*-band tensor into a 2-dimensional space. The clusters in a t-SNE plot indicate similar LDOS shapes, and points farther apart would be dissimilar. A color is assigned to each point, corresponding to the color-map of the element concentration. Across all elements in nanoclusters, a high correlation of the t-SNE clusters with the element concentration is seen, with an average MI of 1.39. It indicates that the LDOS of the atoms is a substantial function of a property global to the nanocluster. This can be attributed to the small size of these nanoclusters, and the delocalized electrons of the transition metals in the nanoclusters. t-SNE plots with *s*, *p*, and *d* states were also plotted, but no significant change in the MI was seen (see SI Figure S9). This MI, between t-SNE plots of the d orbitals and the element concentration, was also compared with other descriptors of the LDOS, ε_d, ε_d^w, and ε_u, as seen in Table III. It confirms the previous discussion of comparison between the descriptors.

Additionally, a similar t-SNE plot of the *d*-band is compared with the site of the atom—core, edge, or vertex—in the nanocluster (see SI Figure S8). A low average MI of 0.68 is seen across all the elements in the nanoclusters. It further lends credence to the substantial global influence on the LDOS. This global influence on LDOS would also explain their weak correlation with the adsorption energies, that are local to a site. This could be remedied with future descriptors, which are a combination of descriptors with longer range, and the ones that describe the entire nanocluster.

Comparison with Experiments. Platinum is known for its unrivalled catalytic activity and acts as the reference catalyst, so we compared its descriptor values to other nanoclusters to indicate good catalyst candidates. We compared it to three references with consistent series of experiments on HER. Table IV lists the catalytic activities from references[10] (A),[65] (B), and[66] (C) as specific current density j_s or exchange current density j_0.

Since none of the above series of experiments were done on nanoparticles of controlled size, the results could deviate from theoretical predictions. Hidden variables in experiments as well as approximations in the theoretical model could also play a role here. Catalytic activity followed the trends (A) NiCo > Ni > Co > NiFe, (B) NiTi > Ni ≈ Co, and (C) Ni ≈ Co > Fe > Cu > Ti. The *d*-band centers in Figure 4 of NiTi, NiFe, and NiCo drop closer to the *d*-band center of platinum compared to pure Ni and Co. This agrees with (B) but disagrees partly with (A). The ordering of catalytic activity of pure elements in (C) was also not reflected by the *d*-band center of nanoclusters with titanium having a similar *d*-band center as platinum. Upon adding half the *d*-bandwidth, ϵ_d^w in Figure 5 the picture changes slightly. Similar to the *d*-band center, NiTi, NiFe, and NiCo ϵ_d^w get closer to platinum. Again, this agrees with (B) but disagrees partly with (A). The ϵ_d^w values are similar for Ni, Co, Fe, and Cu; hence, could not discriminate between the catalytic activities in (C). Only Ti no longer had a value comparable to platinum. The descriptor ϵ_u disagrees at least partly with all observed catalytic activities. This is due to copper and NiTi

https://dx.doi.org/10.1021/acscombsci.0c00102
ACS Comb. Sci. 2020, 22, 768−781

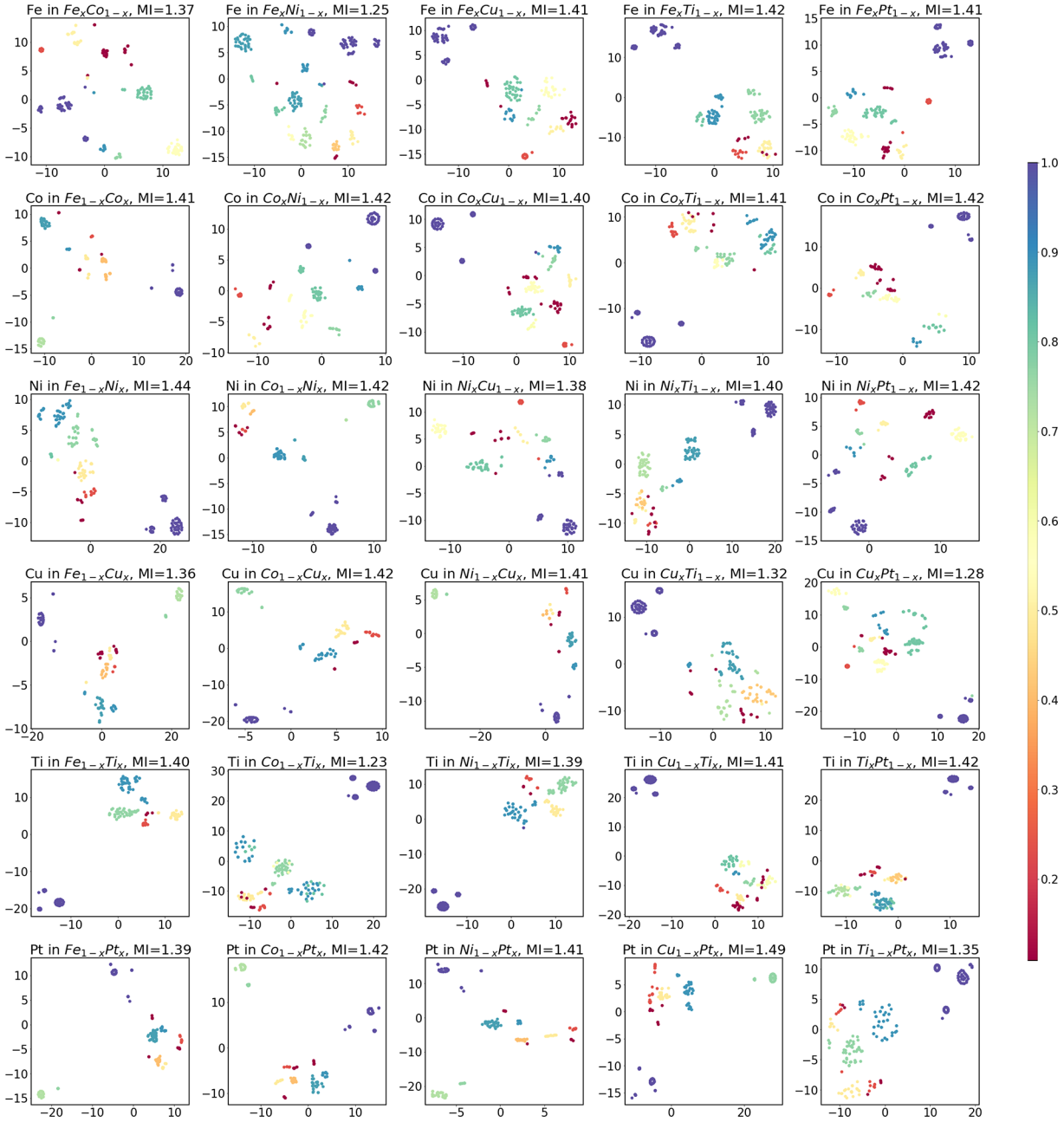


Figure 10. t-SNE plots for LDOS of the *d*-band of an element in various nanoclusters. A colormap of the concentration of an element in a nanocluster is added to correlate the t-SNE clusters with the concentration.

Table III. Average MI of the Descriptors over States of the d Orbitals of an Element in Various Nanoclusters, With the Concentration of the Element[a]

	LDOS	ϵ_d	ϵ_d^w	ϵ_u
MI	**1.39**	0.98	0.93	0.98

[a]LDOS, used as is, has the highest mutual information with the element concentration, followed by the descriptors of ϵ_u and ϵ_d.

having a too low ϵ_u and NiFe having a value similar to platinum. Overall, the agreement is not good, possibly for the reasons of difference in catalytic activity of nanoclusters and/or lower predictive power of electronic descriptors in the nanoregime.

The adsorption energies of NiCo, NiTi, and NiFe (see Figure 8) were shifted upward closer to platinum adsorption, which agrees with (B) but disagrees partly with (A). The ordering of catalytic activity of pure elements in (C) turned out to be the same as the ordering of adsorption energy trends of pure nanoclusters, except for copper which had a higher adsorption energy than platinum. This was expected as copper surfaces are usually less reactive than platinum.[46] The adsorption energy distribution could describe the trends in catalytic activity well for the pure elements. It, however, did not explain all activity trends of bimetallic systems. In particular, NiFe is less active than cobalt or nickel, but according to the adsorption energy distribution, the opposite was expected. A possible reason for that could be that iron has the highest magnetic moment among the studied elements.[60] Apart from that, iron forms bcc crystals as opposed to hcp or fcc for all the other elements.

https://dx.doi.org/10.1021/acscombsci.0c00102
ACS Comb. Sci. 2020, 22, 768−781

Table IV. Experimental Catalytic Activities of Selected Pure and Binary Transition Metals Given As Either Specific Current Density j_s at $\eta = 0.35V$ or Exchange Current Density j_0

	$j_{s,\eta=0.35V}$		j_0	
catalyst	[mAcm^{-2}]	ref	[mAcm^{-2}]	ref
Pt	0.54	A	0.57	B
Pt			1.0	C
Co	0.002	A	0.00094	B
Co	0.002	A	0.005	C
Fe	0.004	A	0.0025	C
Ni	0.021	A	0.0004	B
Ni			0.0056	C
Ti			0.000005	C
Cu			0.000016	C
NiCo	0.062	A		
NiFe	0.002	A		
NiTi			0.003	B

Comparison with Other Computed Data Sets. Adsorption energies of nanoclusters differ from those on periodic slabs. The structure of the slab surface is predetermined by the crystal, since the forces exerted are generally not large enough to induce a bulk distortion (high bulk-to-surface ratio). In nanoclusters on the other hand, the core structure is distorted or under strain, especially in small icosahedral clusters. Apart from that, edges are more prevalent in nanoclusters. In order to view how adsorption energies shift between the two models, our nanocluster data set was compared to two slab databases from the literature.[20,67] Figure 11b shows the average adsorption energy on a pure or bimetallic nanocluster top site with a top site on a periodic slab from ref 67. Figure 11a shows the differences between those models for each pure or bimetallic system.

The barchart shows that adsorption energies on nanoclusters were shifted downward. This was due to the different model geometries but also due to different computational methods. The shift was, however, not constant with respect to elements. Co and Ni, for instance were expected to have similar adsorption energies by the slab model, but had a 0.2 eV difference in the nanocluster model (see Figure 11b). Figure 12 adds a large database from ref 20 of adsorbates on periodic slabs of higher Miller indices into the picture.

The downward shift of adsorption energies of nanoclusters became also evident when looking at the lowest-energy adsorbates on slabs (orange) or their mean (black) in Figure 12. It was most prominent in nickel and cobalt, but also platinum and copper show a slight downward shift. Pure iron seemed to have outliers at very low adsorption energies. All the six outliers had low Miller indices of 100 and 110, small unit cells and significant surface reconstruction. Only the adsorption energy of titanium was close between the two data sets. The periodic slab data set contained most bimetallic data on platinum compounds. There, the curves of the lowest-energy adsorbates on slabs (orange) as well as the mean adsorption energies (black) followed the same trends as of the nanocluster adsorbates, however the PtCo and PtNi curves diverged toward the left.

CONCLUSIONS

We automated the procedure for creating a bimetallic data set, from the generation of nanoclusters through adsorption site detection to machine learning predictions of the hydrogen adsorption onto the clusters. Given a fixed shape, namely icosahedral, the problem of high number of configurations was mitigated by measuring how similar configurations were to one another. This subset was big enough to include experimentally observable composites such as core−shell, segregated, ordered, and random.

The *d*-band properties of nanoclusters correlated consistently with nanocluster stability. This indicated that a dense sampling is required to avoid systematic descriptor errors. In

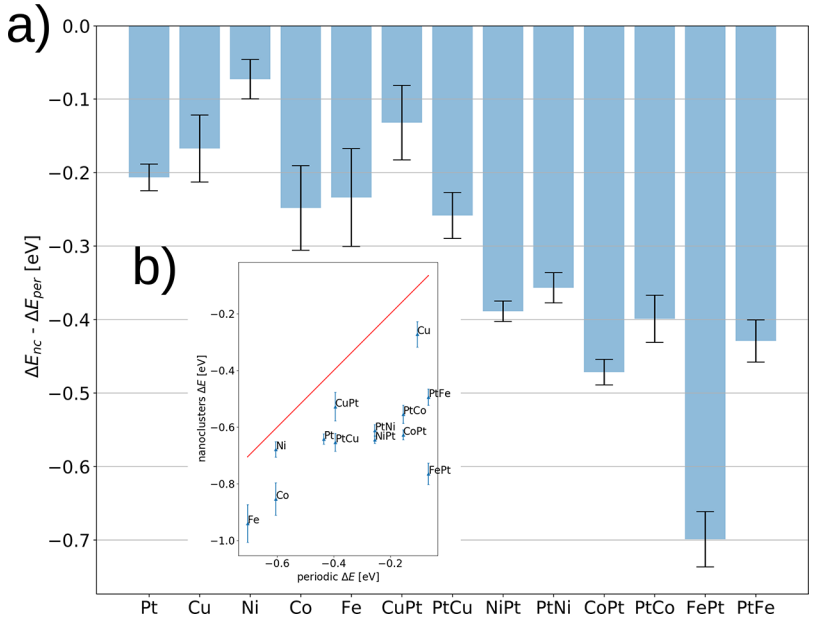


Figure 11. Comparison of adsorption energies of only top nanocluster adsorption sites with a periodic slab data set from ref 67. The error bars display the standard deviation of the distribution of adsorption energies. The first element mention denotes the binding site. (a) A histogram of the difference in adsorption energies. (b) A parity plot with the parity line $y = x$ is shown in red.

https://dx.doi.org/10.1021/acscombsci.0c00102
ACS Comb. Sci. 2020, 22, 768−781

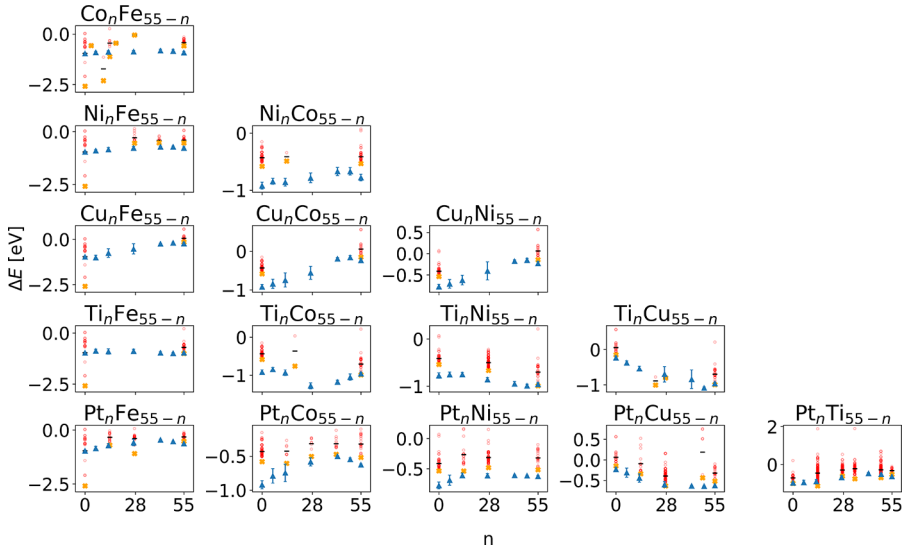


Figure 12. Comparison of adsorption energies of nanocluster adsorption sites (blue) with a data set of slabs with higher Miller indices from ref 20 (red). The minimum (average) of the latter data for a given composition is depicted in orange (black).

the future, descriptor statistics could be used to monitor and actively modulate the extent of cluster configuration sampling.

The adsorption energy distribution on nanoclusters was computed efficiently by simulating the most dissimilar data points (less than 10%) and interpolating the remaining data points with machine learning to a mean absolute error of 0.1 eV. Enhancing the structural descriptor SOAP by the LDOS of surface atoms did not significantly improve the prediction accuracy. That result was in contrast to the work where the method was suggested[51] where a 30% accuracy improvement over SOAP on pure amorphous carbon surfaces was observed. Since our systems contained both multiple elements and *d*-orbitals were involved in binding, the LDOS was composed of more multifaceted states.

Adsorption energies as a descriptor are much closer to the actual catalytic activity than electronic descriptors. The electronic descriptors ϵ_d and ϵ_d^w show only weak correlation with adsorption energies. The maximum of the *d*-band Hilbert transform ϵ_u did correlate moderately to strongly with the adsorption energy, especially strongly on pure nanoclusters. The results indicate that ϵ_u can be used semiquantitatively for small cluster sizes, as a descriptor for prescreening at lower computational cost than adsorption energies. Further analysis of the LDOS showed that the contribution of *s* and *p* states to the adsorption energy do not significantly provide additional information to the descriptor. The similarity clustering of LDOS with respect to nanocluster composition indicated that an improved descriptor should incorporate global or at least long-range information.

A qualitative comparison of electronic descriptors with experiments showed no consistent agreement. This could have two reasons: either the nanocluster structures were not representative of nanoparticles in real conditions, or the electronic descriptors lose their predictive power when applied to the nanocluster model. The adsorption energy distribution could describe the trends in catalytic activity better, at least for the pure elements. It, however, did not explain activity trends of bimetallic systems well. It is possible that the adsorption energy does not suffice to make predictions across the board, therefore it might be necessary to expand to more expensive coverage simulations. Additionally, a future series of experiments with controlled size and composition could help make more substantiated comparisons. When we compared our data to other computational data sets, there were nonconstant shifts, likely due to structural differences between nanoclusters and periodic slabs.

Our statistical approach yields not only descriptors but also their distribution. This indicates how much the local properties can change upon configurational changes. Since we selected only the 10 most dissimilar structures, the convex hull was not accurate. The number of sampled configurations could be increased on-the-fly until electronic descriptors are converged. In the future, the data could be enhanced so that it also reveals shape changes. With the current workflow, different nanocluster shapes can be encompassed in the search. The automation tools support arbitrary shapes and also molecular adsorbates, so that more complex reactions such as the oxygen evolution reaction can be screened.

A broader search with different sizes (up to 2 nm) and shapes (wulff, dodecahedral, and icosahedral) could open the door for quantitative structural effects on electronic descriptors and adsorption energies. Comparing them to periodic slabs could help untangle nanocluster effects and attribute them to core strain, the finite nature of the system and contribution of faces and edges. If we understand those contributions better, we can explore refined descriptors, such as the incorporation of structural effects into *d*-band properties, to better guide nanocatalyst design.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acscombsci.0c00102.

Figures for excess energies of nanoclusters and stability change of nanoclusters with respect to a similarity metric. Additionally, for each descriptor ϵ_d, ϵ_d^w and ϵ_u, figures resolve the correlation with nanocluster stability in Table I per composition. The correlation between ϵ_d and the nearest and next-nearest neighbor environment is also shown. Furthermore, descriptor values of the reduced core Pt nanocluster are compared to the values

https://dx.doi.org/10.1021/acscombsci.0c00102
ACS Comb. Sci. 2020, 22, 768−781

of the icosahedral configuration. Lastly, the mutual information between the LDOS and other properties such as electronic descriptors is shown(PDF)

AUTHOR INFORMATION

Corresponding Author

Marc O. J. Jäger − *Department of Applied Physics, Aalto University, 00076 Aalto, Espoo, Finland;* orcid.org/0000-0001-9045-7431; Email: marc.jager@aalto.fi

Authors

Yashasvi S. Ranawat − *Department of Applied Physics, Aalto University, 00076 Aalto, Espoo, Finland*

Filippo Federici Canova − *Department of Applied Physics, Aalto University, 00076 Aalto, Espoo, Finland; Nanolayers Research Computing Ltd, London N12 0HL, United Kingdom*

Eiaki V. Morooka − *Department of Applied Physics, Aalto University, 00076 Aalto, Espoo, Finland*

Adam S. Foster − *Department of Applied Physics, Aalto University, 00076 Aalto, Espoo, Finland; WPI Nano Life Science Institute (WPI-NanoLSI), Kanazawa University, Kakuma-machi, Kanazawa 920-1192, Japan;* orcid.org/0000-0001-5371-5905

Complete contact information is available at: https://pubs.acs.org/10.1021/acscombsci.0c00102

Author Contributions

M.J created the data, designed the automated workflow and wrote the manuscript. M.J., Y.R., E.M., and F.F. contributed to the automation tools and analyzed the results. All authors reviewed and commented on the manuscript. A.F. supervised the project.

Funding

The work was supported by the European Union's Horizon 2020 research and innovation program under grant agreement no. 676580 NOMAD, a European Center of Excellence and no. 686053 CRITCAT. We thank the Finnish Foundation for Technology Promotion for their funding.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We acknowledge the generous computing resources from CSC - IT Center for Scientific Computing and the computational resources provided by the Aalto Science-IT project. This work was partially supported by the Finnish Foundation for Technology promotion. A.S.F. was supported by the World Premier International Research Center Initiative (WPI), MEXT, Japan.

REFERENCES

(1) Gahleitner, G. Hydrogen from renewable electricity: An international review of power-to-gas pilot plants for stationary applications. *Int. J. Hydrogen Energy* **2013**, *38*, 2039−2061.

(2) Parra, D.; Valverde, L.; Pino, F. J.; Patel, M. K. A review on the role, cost and value of hydrogen energy systems for deep decarbonisation. *Renewable Sustainable Energy Rev.* **2019**, *101*, 279−294.

(3) Zhang, Y.-h.; et al. Development and Application of Hydrogen Storage. *J. Iron Steel Res. Int.* **2015**, *22*, 757−770.

(4) Yu, X.; Tang, Z.; Sun, D.; Ouyang, L.; Zhu, M. Recent advances and remaining challenges of nanostructured materials for hydrogen storage applications. *Prog. Mater. Sci.* **2017**, *88*, 1−48.

(5) Bertuccioli, L. et al. *Development of Water Electrolysis in the European Union*, Tech. Rep., Fuel Cells and Hydrogen Joint Undertaking, 2014.

(6) Graves, C.; Ebbesen, S. D.; Mogensen, M.; Lackner, K. S. Sustainable hydrocarbon fuels by recycling CO2 and H2O with renewable or nuclear energy. *Renewable Sustainable Energy Rev.* **2011**, *15*, 1−23.

(7) Bhandari, R.; Trudewind, C. A.; Zapp, P. Life cycle assessment of hydrogen production via electrolysis − a review. *J. Cleaner Prod.* **2014**, *85*, 151−163.

(8) Roger, I.; Shipman, M. A.; Symes, M. D. Earth-abundant catalysts for electrochemical and photoelectrochemical water splitting. *Nature Reviews Chemistry* **2017**, *1*, 0003.

(9) Jiao, Y.; Zheng, Y.; Jaroniec, M.; Qiao, S. Z. Design of electrocatalysts for oxygen- and hydrogen-involving energy conversion reactions. *Chem. Soc. Rev.* **2015**, *44*, 2060−2086.

(10) McCrory, C. C. L.; et al. Benchmarking Hydrogen Evolving Reaction and Oxygen Evolving Reaction Electrocatalysts for Solar Water Splitting Devices. *J. Am. Chem. Soc.* **2015**, *137*, 4347−4357.

(11) Seh, Z. W.; et al. Combining theory and experiment in electrocatalysis: Insights into materials design. *Science* **2017**, *355*, 4998.

(12) Mayyas, A.; Ruth, M.; Pivovar, B.; Bender, G.; Wipke, K. *Manufacturing Cost Analysis for Proton Exchange Membrane Water Electrolyzers*, Tech. Rep., National Renewable Energy Laboratory, 2019).

(13) Deloitte Sustainability, British Geological Survey, Bureau de Recherches Géologiques et Minières & Netherlands Organisation for Applied Scientific Research. *Study on the review of the list of critical raw materials: Final report*, Tech. Rep, 2017.

(14) Li, A.; Sun, Y.; Yao, T.; Han, H. Earth-Abundant Transition-Metal-Based Electrocatalysts for Water Electrolysis to Produce Renewable Hydrogen. *Chem. - Eur. J.* **2018**, *24*, 18334−18355.

(15) Xie, P.; et al. Highly efficient decomposition of ammonia using high-entropy alloy catalysts. *Nat. Commun.* **2019**, *10*, 1−12.

(16) Zhang, G.; et al. High entropy alloy as a highly active and stable electrocatalyst for hydrogen evolution reaction. *Electrochim. Acta* **2018**, *279*, 19−23.

(17) Zeng, M.; Li, Y. Recent advances in heterogeneous electrocatalysts for the hydrogen evolution reaction. *J. Mater. Chem. A* **2015**, *3*, 14942−14962.

(18) Darby, M. T.; Stamatakis, M.; Michaelides, A.; Charles Sykes, E. H. Lonely Atoms with Special Gifts: Breaking Linear Scaling Relationships in Heterogeneous Catalysis with Single-Atom Alloys. *J. Phys. Chem. Lett.* **2018**, *9*, 5636−5646.

(19) Ulissi, Z. W.; Medford, A. J.; Bligaard, T.; Nørskov, J. K. To address surface reaction network complexity using scaling relations machine learning and DFT calculations. *Nat. Commun.* **2017**, *8*, 14621.

(20) Tran, K.; Ulissi, Z. W. Active learning across intermetallics to guide discovery of electrocatalysts for CO_2 reduction and H_2 evolution. *Nature Catalysis* **2018**, *1*, 696−703.

(21) Li, Z.; Wang, S.; Chin, W. S.; Achenie, L. E.; Xin, H. High-throughput screening of bimetallic catalysts enabled by machine learning. *J. Mater. Chem. A* **2017**, *5*, 24131−24138.

(22) Nørskov, J. K.; Bligaard, T.; Rossmeisl, J.; Christensen, C. H. Towards the computational design of solid catalysts. *Nat. Chem.* **2009**, *1*, 37−46.

(23) Parsons, R. The rate of electrolytic hydrogen evolution and the heat of adsorption of hydrogen. *Trans. Faraday Soc.* **1958**, *54*, 1053.

(24) Nørskov, J. K.; et al. Trends in the Exchange Current for Hydrogen Evolution. *J. Electrochem. Soc.* **2005**, *152*, J23−J26.

(25) Greeley, J. Theoretical Heterogeneous Catalysis: Scaling Relationships and Computational Catalyst Design. *Annu. Rev. Chem. Biomol. Eng.* **2016**, *7*, 605−635.

(26) Zhao, Z.-J.; et al. Theory-guided design of catalytic materials using scaling relationships and reactivity descriptors. *Nature Reviews Materials* **2019**, *4*, 792−804.

(27) Gasper, R.; Shi, H.; Ramasubramaniam, A. Adsorption of CO on Low-Energy, Low-Symmetry Pt Nanoparticles: Energy Decomposition Analysis and Prediction via Machine-Learning Models. *J. Phys. Chem. C* **2017**, *121*, 5612−5619.
(28) Wexler, R. B.; Martirez, J. M. P.; Rappe, A. M. Chemical Pressure-Driven Enhancement of the Hydrogen Evolving Activity of Ni_2P from Nonmetal Surface Doping Interpreted via Machine Learning. *J. Am. Chem. Soc.* **2018**, *140*, 4678−4683.
(29) Ma, X. Orbitalwise Coordination Number for Predicting Adsorption Properties of Metal Nanocatalysts. *Phys. Rev. Lett.* **2017**, *118*, 036101.
(30) Dumesic, J. A.; Rudd, D. F.; Aparicio, M.; Rekoske, J. E.; Trevino, A. A. *The Microkinetics of Heterogeneous Catalysis*; American Chemical Society, 1993.
(31) Singh, A. R.; Rohr, B. A.; Gauthier, J. A.; Nørskov, J. K. Predicting Chemical Reaction Barriers with a Machine Learning Model. *Catal. Lett.* **2019**, *149*, 2347−2354.
(32) Hinnemann, B.; et al. Biomimetic Hydrogen Evolution: MoS_2 Nanoparticles as Catalyst for Hydrogen Evolution. *J. Am. Chem. Soc.* **2005**, *127*, 5308−5309.
(33) Brüssel, M.; di Dio, P. J.; Muñiz, K.; Kirchner, B. Comparison of Free Energy Surfaces Calculations from Ab Initio Molecular Dynamic Simulations at the Example of Two Transition Metal Catalyzed Reactions. *Int. J. Mol. Sci.* **2011**, *12*, 1389−1409.
(34) Smith, A.; Keane, A.; Dumesic, J. A.; Huber, G. W.; Zavala, V. M. A machine learning framework for the analysis and prediction of catalytic activity from experimental data. *Appl. Catal., B* **2020**, *263*, 118257.
(35) Hutter, J.; Iannuzzi, M.; Schiffmann, F.; VandeVondele, J. CP2K: Atomistic simulations of condensed matter systems. *Wiley Interdisciplinary Reviews-Computational Molecular Science* **2014**, *4*, 15−25.
(36) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.
(37) VandeVondele, J.; Hutter, J. Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases. *J. Chem. Phys.* **2007**, *127*, 114105.
(38) Goedecker, S.; Teter, M.; Hutter, J. Separable dual-space Gaussian pseudopotentials. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1996**, *54*, 1703−1710.
(39) Krack, M. Pseudopotentials for H to Kr optimized for gradient-corrected exchange-correlation functionals. *Theor. Chem. Acc.* **2005**, *114*, 145−152.
(40) Hartwigsen, C.; Goedecker, S.; Hutter, J. Relativistic separable dual-space Gaussian Pseudopotentials from H to Rn. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1998**, *58*, 3641−3662.
(41) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys. 132* (**2010**).154104
(42) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Funtion in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.* **2011**, *32*, 1456−1465.
(43) Jäger, M. O. J.; Morooka, E. V.; Canova, F. F.; Himanen, L.; Foster, A. S. Machine learning hydrogen adsorption on nanoclusters through structural descriptors. *npj Computational Materials* **2018**, *4*, 1−8.
(44) Bartók, A. P.; Kondor, R.; Csányi, G. On representing chemical environments. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2013**, *87*, 184115.
(45) De, S.; Bartók, A. P.; Csányi, G.; Ceriotti, M. Comparing molecules and solids across structural and alchemical space. *Phys. Chem. Chem. Phys.* **2016**, *18*, 13754−13769. 1601.04077.
(46) Hammer, B.; Nørskov, J. K. Why gold is the noblest of all the metals. *Nature* **1995**, *376*, 238−240.
(47) Hammer, B.; Nørskov, J. Theoretical Surface Science and Catalysis — Calculations and Concepts. *Adv. Catal.* **2000**, *45*, 71−129.
(48) Xin, H.; Linic, S. Communications: Exceptions to the d-band model of chemisorption on metal surfaces: The dominant role of repulsion between adsorbate states and metal d-states. *J. Chem. Phys.* **2010**, *132*, 221101.
(49) Vojvodic, A.; Nørskov, J. K.; Abild-Pedersen, F. Electronic Structure Effects in Transition Metal Surface Chemistry. *Top. Catal.* **2014**, *57*, 25−32.
(50) Xin, H.; Vojvodic, A.; Voss, J.; Nørskov, J. K.; Abild-Pedersen, F. Effects of d-band shape on the surface reactivity of transition-metal alloys. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2014**, *89*, 115114.
(51) Caro, M. A.; Aarva, A.; Deringer, V. L.; Csányi, G.; Laurila, T. Reactivity of Amorphous Carbon Surfaces: Rationalizing the Role of Structural Motifs in Functionalization Using Machine Learning. *Chem. Mater.* **2018**, *30*, 7446−7455.
(52) Schlexer Lamoureux, P.; et al. Machine Learning for Computational Heterogeneous Catalysis. *ChemCatChem* **2019**, *11*, 3581−3601.
(53) Himanen, L. et al. DScribe: Library of descriptors for machine learning in materials science. *Comput. Phys. Commun. 106949* (**2020**).106949
(54) Piotrowski, M. J.; et al. Theoretical Study of the Structural, Energetic, and Electronic Properties of 55-Atom Metal Nanoclusters: A DFT Investigation within van der Waals Corrections, Spin−Orbit Coupling, and PBE+ U of 42 Metal Systems. *J. Phys. Chem. C* **2016**, *120*, 28844−28856.
(55) Bartók, A. P.; et al. Machine learning unifies the modeling of materials and molecules. *Science Advances* **2017**, *3*, e1701816.
(56) Faber, F. A.; et al. Prediction Errors of Molecular Machine Learning Models Lower than Hybrid DFT Error. *J. Chem. Theory Comput.* **2017**, *13*, 5255−5264.
(57) Jain, A.; et al. FireWorks: A dynamic workflow system designed for high-throughput applications. *Concurrency and Computation: Practice and Experience* **2015**, *27*, 5037−5059.
(58) SIN. Critcatworks. https://github.com/SINGROUP/critcatworks (2020).
(59) SIN. Cluskit. https://github.com/SINGROUP/cluskit (2019).
(60) Guedes-Sobrinho, D.; Nomiyama, R. K.; Chaves, A. S.; Piotrowski, M. J.; Da Silva, J. L. F. Structure, Electronic, and Magnetic Properties of Binary Pt_nTM_{55-n} (TM = Fe, Co, Ni, Cu, Zn) Nanoclusters: A Density Functional Theory Investigation. *J. Phys. Chem. C* **2015**, *119*, 15669−15679.
(61) Zhang, R. F.; et al. An informatics guided classification of miscible and immiscible binary alloy systems. *Sci. Rep.* **2017**, *7*, 1−12.
(62) Wang, L.-L.; Johnson, D. D. Predicted Trends of Core-Shell Preferences for 132 Late Transition-Metal Binary-Alloy Nanoparticles. *J. Am. Chem. Soc.* **2009**, *131*, 14023−14029.
(63) Kraskov, A.; Stögbauer, H.; Grassberger, P. Estimating mutual information. *Phys. Rev. E* **2004**, *69*, 066138.
(64) Maaten, L. v. d.; Hinton, G. Visualizing data using t-sne. *J. Mach. Learn. Res.* **2008**, *9*, 2579−2605.
(65) Lu, S.; Zhuang, Z. Electrocatalysts for hydrogen oxidation and evolution reactions. *Science China Materials* **2016**, *59*, 217−238.
(66) Trasatti, S. Work function, electronegativity, and electrochemical behaviour of metals: III. Electrolytic hydrogen evolution in acid solutions. *J. Electroanal. Chem. Interfacial Electrochem.* **1972**, *39*, 163−184.
(67) Greeley, J.; Mavrikakis, M. Alloy catalysts designed from first principles. *Nat. Mater.* **2004**, *3*, 810−815.

https://dx.doi.org/10.1021/acscombsci.0c00102
ACS Comb. Sci. 2020, 22, 768−781