Journal of
Materials Chemistry A

Materials for energy and sustainability

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: H. Li, S. Xu, M. Wang, Z. Chen, F. Ji, K. Cheng, Z. Gao, Z. Ding and W. Yang, *J. Mater. Chem. A*, 2020, DOI: 10.1039/D0TA04615A.

ROYAL SOCIETY OF CHEMISTRY

rsc.li/materials-a

View Article Online
DOI: 10.1039/D0TA04615A

Computational Design of (100) Alloy Surfaces for Hydrogen Evolution Reaction

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

Hao Li[a,b], Shaopeng Xu[a], Min Wang[a], Ziheng Chen[c], Fengfeng Ji[a], Kewei Cheng[d], Zhengyang Gao[a], Zhao Ding[e], and Weijie Yang[a,*]

With the rapid development of kinetically-controlled techniques, synthesis of cubic bimetalllic catalysts with tunable component and composition becomes possible. In recent years, many of the bimetallic alloy catalysts with (100) as the predominant facet have shown outstanding electrocatalytic activity and stability. However, alloying effects on the activity at (100) were less explored, compared to the well-studied closed-packed surfaces. Here, using density functional theory (DFT) calculations, we explore the catalytic activity of bimetallic alloy (100) surfaces for electrocatalytic hydrogen evolution reaction (HER). The ensemble and compositional effects of (100) facets alloyed by strong- (Pd and Pt) and weak-binding (Ag, Au, and Cu) transition metals were studied. Based on DFT calculations, the catalytic activity of bimetallic alloys for HER were systematically investigated using H binding energy as the reaction descriptor, at all of the typical surface ensembles on (100). Our results suggest that Pd_xAg_{1-x} and Pd_xAu_{1-x}(100) have promising theoretical HER activity in acidic media, due to the presence of highly active four-fold ensembles which reach the peak of the HER volcano activity plot. The electronic structure and stability of the alloys with predicted promising HER activities were studied. Furthermore, based on the DFT-calculated database, we performed feature analysis and developed a robust machine learning model which can help to predict HER activities of those out-of-sample alloys. Most importantly, this study provides helpful guideline for the design of (100) dominated bimetallic alloys towards promising HER activity.

1. Introduction

Bimetallic cubic catalyst is a new type of materials for electrochemistry due to the rapid development of advanced kinetically-controlled synthesis techniques.[1-3] As the predominant facet of cubic catalysts, bimetallic (100) has more complicated heterogeneous surfaces due to the presence of four-fold adsorption sites, compared to the simpler three-fold sites on (111).[4] Although (100) usually has higher surface energy than those close-packed surfaces,[5] kinetically-controlled synthesis is able to form a metastable cubic nanoparticle (NP) structure with the use of proper capping element such as bromine.[6, 7] Meanwhile, some commonly seen low energy structures (*e.g.*, cuboctahedron)[8] also have a large ratio of (100) facets. Once a cubic shape is formed, many of the bimetallic (100) surfaces have shown promising stability and durability in electrocatalytic experiments.[4, 9-19] Meanwhile, with advanced synthesis methods, both classically miscible and immiscible bimetallic alloys with well-defined NP structures have been successfully synthesized and reported recently, such as PdAu,[20] PdAg,[21] PdCu,[22] PtAu,[10] PtAg,[23] PtCu,[24] RhAu,[11] RhAg,[12] IrAg,[13] IrCu,[25] PdRh,[26] RhIr,[15] and PdIr.[27] Based on some recent combined experimental and computational studies, it was found that once an alloyed cubic structure is formed, the geometry of a NP is kinetically trapped, making the surface difficult to collapse into a more close-packed structure. Meanwhile, some of the synthesized bimetallic structures are not easy to segregate under mild temperature, due to the proven high kinetic barriers for these processes.[28-30] These make some of the cubic alloy NPs ideal for electrocatalysis – an electrocatalytic reaction environment is usually with ambient conditions, which does not provide high temperature to overcome these kinetic barriers of segregation or collapsing. Therefore, (100) dominated NPs can be a type of potential candidates for industrially important electrocatalysis.

Recently, there are some outstanding experimental studies which have shown that alloy cubic NPs could be promising catalysts for electrocatalysis. Rodene *et al*.[31] realized the synthesis of cubic $Ni_{1-x}Mo_x$ alloy NPs for water splitting to produce hydrogen. They found that the hydrogen evolution reaction (HER) activity increased with the increasing Mo content until x=6.6%. Among the candidate catalysts, cubic $Ni_{0.93}Mo_{0.07}$ NPs showed the best catalytic performance, with the overpotentials of −64 and −85 mV respectively at −10 and −20 mA/cm^2. Some cubic nanomaterials (*e.g.*, Cu-Ni-$CoSe_x$, Ni-$CoSe_x$, and Cu-$CoSe_x$ nanocubes) were prepared by Qian *et al*.[32] for dye-sensitized solar cells and HER. Their research showed that Cu-Ni-$CoSe_x$ could accelerate HER, with an ultralow overpotential (50.2 mV at 10 mA/cm^2). In the research of Yang *et al.*,[33] using a simple two-step method, cubic iron carbides wrapped in N-doped carbon shell were successfully synthesized. Their results showed that the overpotential of the catalyst was 209 mV at 10 mA/cm^2. Cao *et al*.[34] explored the catalytic activity of face-centred cubic Pd-Ni-P NPs for HER. It was shown that Pd-Ni-P NPs is a promising catalyst, which can promote HER with the overpotential of only 32 mV at 10 mA/cm^2. All of these studies have suggested that the low-index facets (predominantly, the (100) facet) at cubic NPs possess

[a.] *Department of Power Engineering, School of Energy, Power and Mechanical Engineering, North China Electric Power University, Baoding 071003, China .*
[b.] *Department of Physics, Technical University of Denmark, 2800 Kongens Lyngby, Copenhagen, Denmark*
[c.] *Department of Applied Mathematics and Statistics, Stony Brook University , Stony Brook, New York 11794, USA*
[d.] *Department of Computer Science, University of California, Los Angeles, Los Angeles, Los Angeles 90095, USA*
[e.] *Department of Mechanical, Materials and Aerospace Engineering Illinois Institute of Technology, Chicago, Illinois, USA, 60616*
* Corresponding Author: yangwj@ncepu.edu.cn
Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x

highly active sites for electrocatalysis. For acidic HER, it has been shown that strong-binding transition metals (*e.g.*, Pd and Rh) tend to over-bind H, leading to H poisoning which hinders the subsequent evolution step.[4] On the contrary, most of the weak-binding transition metals (*e.g.*, Ag, Au, and Cu) bind H too weakly, leading to difficulty in the proton-electron combination step. Therefore, alloying these strong- and weak-binding elements together could be a promising method which results in a well-tuned adsorption strength that provides an ideal H binding for HER.[4, 17, 35] However, relevant theoretical studies were less-explored on alloyed (100) surfaces, compared to (111). Although there were several studies focusing on the design of alloyed surfaces for HER,[36-38] to the best of our knowledge, very few study has focused on the design of HER catalyst at a more detailed ensemble scale.

Recent studies have shown that atomic ensemble effect (the specific composition of an alloyed adsorption site on surface) is the predominant effect for (111) surfaces alloyed by strong- and weak-binding transition metals.[9, 17, 39] This is because a specific surface ensemble directly determines the adsorption environment of an adsorbate, while the electronic effect[4] (contributed from the elements outside of the site) and strain effect[18] (originated in the lattice difference between the two elements) have less significant influence on the adsorbate binding. This new insight helps to understand why many of the alloy NPs showed outstanding catalytic activities, compared to their monometallic counterparts.[10-16] However, to the best of our knowledge, very few study has focused on the alloying effect at (100) surfaces,[19, 40] due to the more complicated composition of four-fold adsorption sites (**Fig. 1a**) as compared to the simpler three-fold sites on (111). For example, a bimetallic alloy (100) has six different symmetries at a four-fold ensemble, while there are only four different symmetries at the three-fold site. This leads to a much higher computational cost on the studies for bimetallic cubic catalysts.

In this paper, bimetallic (100) surfaces alloyed by strong- (Pd and Pt) and weak-binding (Ag, Au, and Cu) transition metals with different compositions were studied using DFT calculations. The reason of choosing this type of alloy is that compared to those close-packed structures, (100) has lower coordination environment at the surface, leading to over-binding of adsorbate at the strong-binding elements. However, alloying weak-binding elements into strong-binding surface can lead to a tunable adsorbate binding at an appropriate surface ensemble, which could lead to an optimal adsorbate binding for a catalytic reaction due to the *Sabatier* principle.[41] The outline of this paper is shown as follows: first, H binding energies on different binding sites of bimetallic alloys with different compositions were calculated at each type of four-fold ensemble on (100). Second, using HER as a probe reaction, catalytic activities of bimetallic alloys for HER were analysed through the calculated H binding energies, HER volcano model, heat matrix, and free energy variation. Third, the electronic structures of bimetallic alloy with the best predicted activities were investigated to explain their improved catalytic performance. Finally, machine learning (ML) models based on the calculated database were developed to predict other different bimetallic alloys with promising HER activity. This computational screening research based on DFT calculations and ML predictions can effectively guide the design and synthesis of bimetallic alloys for HER.

2. Computational and Modelling Methods

2.1 DFT Calculations

All of the DFT calculations were performed in Vienna *Ab initio* Simulation Package (VASP5.4.4) with Perdew-Burke-Ernzerhof (PBE) functional and projector augmented wave (PAW) methods.[42-44] The kinetic energy cutoff was chosen as 400 eV. To describe the electron orbit occupation, Gaussian smearing with a width of 0.05 eV (σ=0.2 eV) was defined. Structural optimizations were carried out in a (3×3×1) Monkhorst–Pack *k*-point,[45] with a convergence standard of 0.05 eV/Å. The total energy convergence criteria in VASP (EDIFF) was set as 10^{-5}. Convergence tests with spin-polarization and stricter criterion were performed (**Table S1**); no significant change was found in the binding energies and optimized geometries. Since van der Waals correction contributes little to H adsorption on transition metal surfaces, and PBE functional showed relatively small errors in H binding energy calculations as compared to experimental benchmarking database,[46] this correction is not included in our calculations. We also tested our current results with a DFT-D2 method on Pd(100),[47] and only found that the H binding varies less than 0.1 eV without configuration change.

A four-layer slab model with the upper two layers allowed to relax was modelled to represent the catalytic surfaces, which was proven to be reasonable in previous studies.[48] Thicker slabs were tested in our previous studies; no significant change was found in the binding energies and optimized geometries.[49] To eliminate the error caused by periodic mirror images, a vacuum layer of at least 12 Å was built. Bimetallic alloys were composed of Pd, Pt, Ag, Au, and Cu with different ratios of 25%, 50%, and 75%, and each model of bimetallic alloy contains 64 atoms. The schematic pictures of randomly sampled bimetallic alloys are shown in **Fig. 1a**. To obtain abundant alloyed configurations, at least 15 random structures were generated for each composition based on the Atomic Simulation Environment (ASE) library,[50] building at least 270 different structures of bimetallic alloys. All of the bare structures were re-relaxed with our calculation parameters after they were generated. All of the binding sites of H were sampled from these surfaces. Our previous studies have shown that this sampling method is helpful for analysing the uncertainty of various alloyed structures, which is representative for the alloy catalysts synthesized by kinetically-controlled methods.[9-16]

For each ensemble composition at a four-fold site, at least three binding sites were sampled for the binding energy calculations. In total, more than 450 binding energies of H were calculated. The binding energy of H (E_b) was calculated according to $E_b = E_{tot} - E_{sur} - E_{H2}$, where E_{tot}, E_{sur}, and E_{H2} are the total energies of the adsorption system, the bare surface, and a H_2 molecule in vacuum, respectively. All of the structural information are stored in our online database.[51] To acquire the Gibbs free energy of HER, the entropic and ZPE corrections for

View Article Online
DOI: 10.1039/D0TA04615A

transition metal (+0.24 eV) was applied to the calculated H binding energies.[52]

2.2 ML Modelling

ML modelling were carried out using a three-layer artificial neural network with a back-propagation algorithm (BPNN), using *TensorFlow* as the backend (**Fig. 2b**).[53, 54] To describe the characteristics of different bimetallic alloys, atomic and electronic information of the adsorption environment were described using physical indicators, including: electronegativity, *d*-orbital information, *d*-band centre (average energy of *d*-electrons) of the monometallic element, and atomic information. The completed list of these variables are shown in **Table S2**. After feature selection analysis (as discussed later in this paper), these indicators were used as the inputs (independent variables) of the BPNN model. 462 DFT-calculated H binding energy data after normalization were randomly split into training and test sets in different ratios. Adaptive momentum optimizer algorithm with mean squared error (MEA) as the loss function was adopted for ML model training. Rectified linear unit was applied as the transfer function between layers. In addition, to avoid over-fitting, L2 regularization was adopted with a coefficient of 0.001. Cross-validation and sensitivity tests were performed to confirm the robustness of the models.

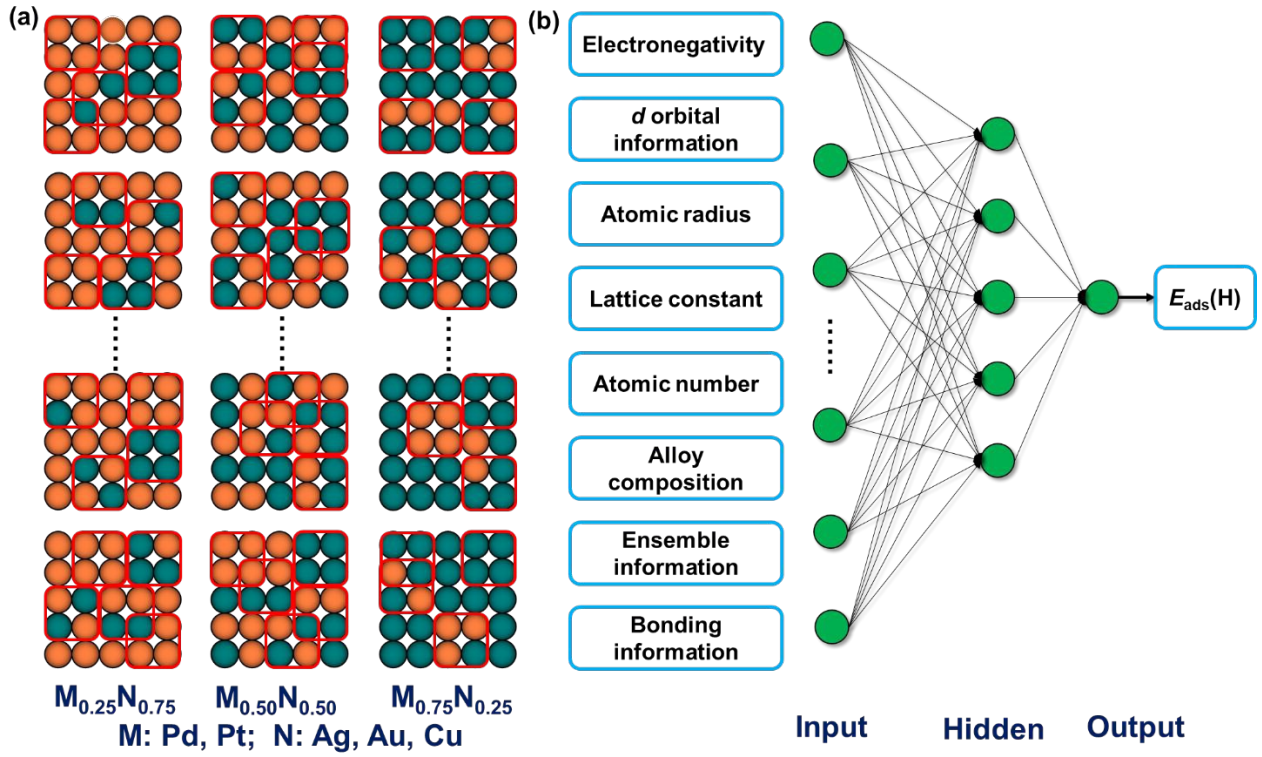


Fig. 1 (a) Schematic pictures of the random sampling method for (100) bimetallic alloys. Red squares represent the four-fold ensemble which provides the specific adsorption environment of H. (b) Algorithmic architecture of the BPNN model employed for this study.

3. Results and Discussion

3.1 H Binding Energies

H binding energies were calculated at the four-fold sites consisting of four neighbouring transition metal atoms, with the reason that a four-fold ensemble is the smallest repeat unit that determines the adsorption environment of H on a (100) surface.[55] According to the composition of these four neighbouring atoms, there are six types of ensembles for H adsorption based on the symmetry of the alloying elements, as shown in **Fig. 2a**. In detail, these six binding sites are defined as: M_4N_0, M_3N_1, M_2N_2-d (diagonal), M_2N_2-p (parallel), M_1N_3, and M_0N_4, respectively (where M represents strong-binding element, Pd and Pt; N represents weak-binding element, Ag, Au, and Cu). The average binding energies of H are plotted in **Fig. 2b-2g** (with the representative binding configurations shown in **Fig. S1**). The calculated error bars (standard deviation) are relatively small at all of the binding sites, indicating that ensemble effect is still the most predominant effect for H adsorption, while the electronic environments outside of the ensemble (from both surface and subsurface) only have slight influence on the H binding. This also reveals that H binding energy is less dependent to the specific configuration of an alloy system. Our previous study also has shown that the binding energies at ordered alloys (*e.g.*, intermetallic alloys) also fall within the error bars which calculated from random sampling.[16] It can be clearly seen from the results that with the increased ratio of weak-binding element in the ensemble, H binding energy at Pd_xN_y becomes monotonically weaker (*i.e.*, more positive) (**Fig 2b-2d**). This shows that as a typical *4d* transition metal, Pd-alloys have tunable H binding energy under various compositions. Interestingly, while the compositional effects are less significant in the binding energies at Pd_xAg_{1-x} and Pd_xAu_{1-x}(100), it has significant influence on the H bindings at Pd_xCu_{1-x}(100) (**Fig. 2d**): lower composition of Pd leads to significantly stronger H bindings at all of the Pd-Cu ensembles. This indicates that compared to Ag and Au, Cu

leads to more significant electronic and strain effects on the adsorbate binding when they are alloyed with Pd. This is also in good agreement with the previous studies on Pd_xCu_{1-x}(111) for nitrogen adoption,[9] suggesting that ensemble effect is not the only significant effect on Pd_xCu_{1-x} alloy. Compared to Pd-based alloys, Pt-based alloys are less tunable for H bindings, with generally similar average H bindings among Pt_4N_0, Pt_3N_1, and Pt_2N_2-p (**Fig. 2e-2g**). A plausible explanation is that Pt is a *5d* transition metal, having larger *d*-orbitals which have slight misalignment when alloying with other transition metals, leading to a less tunable H binding. We also found similar phenomenon in other previously reported *5d* metal alloyed (111) surfaces such as PtAu,[4, 10] IrAg,[13] and IrCu.[14] Meanwhile, based on previous studies, coverage effect is not significant on the H binding energies on most of the systems alloyed with strong-binding metals.[13, 56] Therefore, we expect that H coverage only leads to neglectable influence on the calculated site-specific H binding energy in our studied systems.

In addition, with the same composition of the M_xN_y binding sites, binding energies of H on Pt_xN_y are mostly weaker than those of Pd_xN_y. However, given that Pd_xN_y has more tunable features for H binding, we expect that there should be more Pd-based alloy materials which are promising for HER. Our catalytic modelling, which discussed later, will provide further analysis.

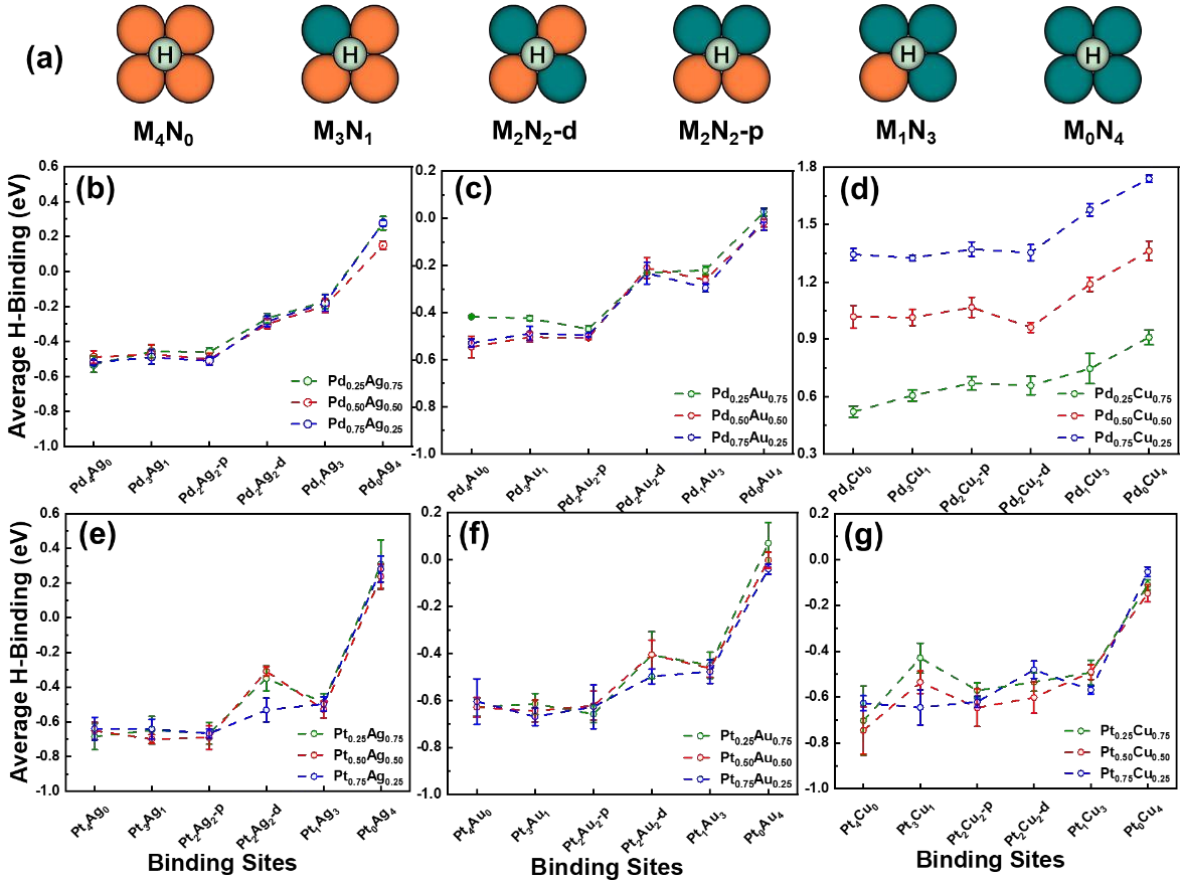


Fig. 2 (a) Binding sites of the four-fold ensembles where H was placed for DFT optimization. (b-g) Calculated average binding energies of H on (b) Pd_xAg_{1-x}(100), (c) Pd_xAu_{1-x}(100), (d) Pd_xCu_{1-x}(100), (e) Pt_xAg_{1-x}(100), (f) Pt_xAu_{1-x}(100), and (g) Pt_xCu_{1-x}(100) (x=0.25, 0.50, and 0.75). For each alloy composition, more than fifteen randomly alloyed geometries were generated. Each error bar is the standard deviation calculated from at least three randomly sampled sites from the generated surfaces.

3.2 HER Activity

Based on the calculated binding energies of H, we used an acidic HER volcano proposed by Nørskov *et al*.[52] to analyse the catalytic activity of HER for different bimetallic alloys (**Fig. 3**). The theoretical current density (in the unit of A/cm^2) is the indicator of the HER activity under the potential of 0 V *vs.* reversible hydrogen electrode with pH=0. This volcano activity plot has shown precise predictive power for HER in acidic media, in good agreement with many previous experiments.[36] It can be clearly seen that, the (100) surfaces of Pd_xAg_{1-x} and Pd_xAu_{1-x} have the surface ensembles close to the peak of the volcano (Pd_1N_3 and Pd_2N_2-d) in all of the three compositions (**Fig. 3a** and **3b**), showing outstanding site-specific HER activities. Given that these two active ensemble geometries are respectively with half- or less-Pd composition, we expect that those Pd_xAg_{1-x} and Pd_xAu_{1-x} cubic alloys with the Pd compositions of 50% or less would lead to the maximized amount of these active sites, leading to outstanding activity and turnover frequency (TOF) in experiments. In contrast, Pd_xCu_{1-x}(100) has most of the ensembles which weakly adsorb H (**Fig. 3c**), indicative of the difficulty in the formation of H*. In addition to Pd_xAg_{1-x} and Pd_xAu_{1-x}(100), Pt_xAg_{1-x}(100) also has an active ensemble (Pd_2Ag_2-d) when x=0.50 (**Fig. 3d**, middle frame). However, for other compositions and components of Pt-based alloys, most of their surface ensembles are at the left-leg of the HER volcano, indicating the poisoning of H* and poor HER activity (**Fig. 3d-3f**).

Based on the volcano model, the optimal H binding occurs when H binding energy reaches the peak of the HER volcano (~ -0.24 eV), where the reaction free energies of the two HER steps become zero. The binding site that drops at the left- and right-legs of the volcano respectively indicate the "strong-" and "weak"-binding regions. Therefore, we further plotted a heat matrix which was quantified by |H binding − target binding|, as shown in **Fig. 4**. In the heat matrix, a smaller value indicates a higher catalytic activity. Based on the results, there are six types of bimetallic ensembles that have the highest catalytic activity of HER among the examined bimetallic (100) surfaces: Pd_2Ag_2-d/$Pd_{0.25}Ag_{0.75}$, Pd_1Au_3/$Pd_{0.25}Au_{0.75}$, Pd_2Au_2-d/$Pd_{0.25}Au_{0.75}$, Pd_1Au_3/$Pd_{0.50}Au_{0.50}$, Pd_2Au_2-d/$Pd_{0.50}Au_{0.50}$, and Pd_2Au_2-d/$Pd_{0.75}Au_{0.25}$. Meanwhile, Pd- and Pt-based bimetallic ensembles containing Cu show the lowest catalytic activities, which suggests that Cu may not be an appropriate component for HER. The original data can be found in the **Supplementary Information**.

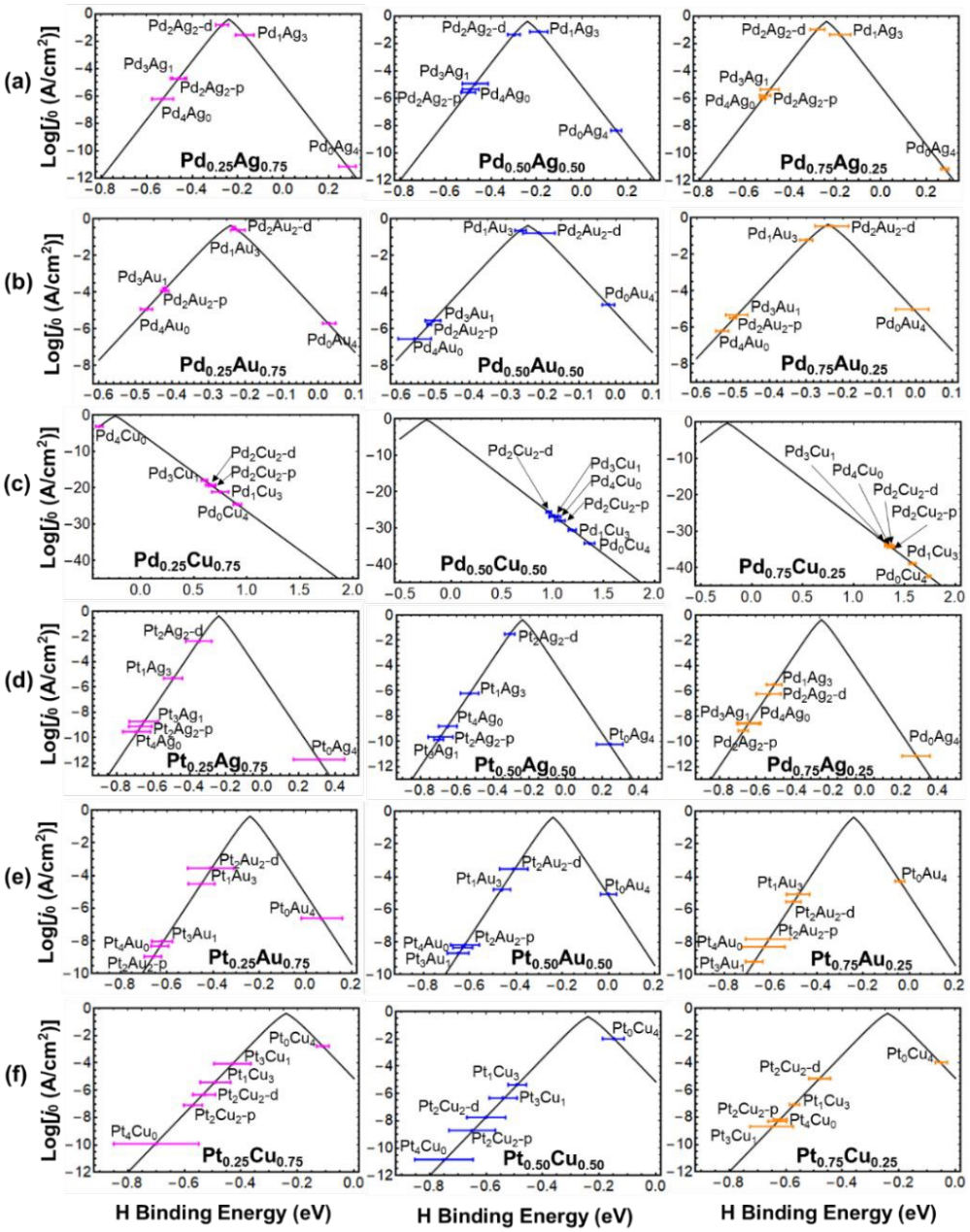


Fig. 3 Predicted acidic HER activity for the binding sites at (a) Pd_xAg_{1-x}(100), (b) Pd_xAu_{1-x}(100), (c) Pd_xCu_{1-x}(100), (d) Pt_xAg_{1-x}(100), (e) Pt_xAu_{1-x}(100), and (f) Pt_xCu_{1-x}(100) (x=0.25, 0.50, and 0.75). The left, middle, and right frames represent the compositions where x=0.25, 0.50, and 0.75, respectively. For each alloy composition, more than fifteen randomly alloyed geometries were generated. Each error bar is the standard deviation calculated from at least three randomly sampled sites from the generated surfaces.

View Article Online
DOI: 10.1039/D0TA04615A

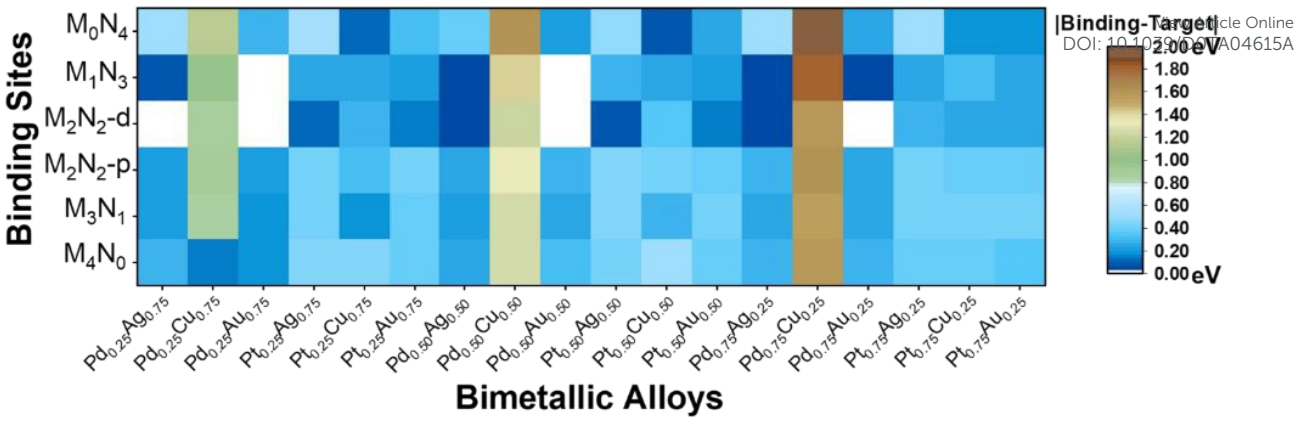


Fig. 4 Heat matrix of the relative binding energy of H. As indicated from the HER volcano model, the target H binding is -0.24 eV.

3.3 Analysis of the Screened Catalysts

Based on the catalytic screening, six bimetallic alloys with high catalytic activities were determined. To provide more understandings to the enhanced activity after alloying at (100), these bimetallics are compared with the catalytic activities of their monometallic counterparts (Pd, Au, and Ag) (**Fig. 5**). Firstly, the free energy diagram of acidic HER at a potential of U = 0 V was plotted, as shown in **Fig. 5a**. The negative reaction free energies at some of the surfaces indicate an exothermic step for the formation of H* after a proton being coupled with an electron and then adsorbed on the surface. Meanwhile, negative free energy in this step results in an endothermic step in the subsequent hydrogen evolution. For Pd(100), this indicates that H poisoning can easily happen because the surface over-binds hydrogen (yellow pathway, **Fig. 5a**), which in turn hinders the hydrogen evolution step. Obviously, the bimetallic alloys of Pd_xAg_{1-x} and Pd_xAu_{1-x}(100) have lower free energy barriers than Pd, Au, and Ag(100), showing that the catalytic activity for HER can be significantly improved through the rational design of bimetallic (100) alloy. Moreover, due to its promising catalytic activity, we picked Pd_2Au_2-d/$Pd_{0.75}Au_{0.25}$(100) as an example to discuss the binding site, catalytic activity of HER, and stability. To identify the binding site of H on Pd_2Au_2-d/$Pd_{0.75}Au_{0.25}$(100), electron transfer in the adsorption process of H and surface electronic structure were investigated, as shown in **Fig. 5b** and **5c**. According to the electron density difference (EDD) of H on Pd_2Au_2-d/$Pd_{0.75}Au_{0.25}$(100), H and the surface respectively act as the electron acceptor and donor. Other similar results on Pd and Pt(100) can be found in **Fig. S2**. In the surface map of electrostatic potential distribution (EPD) (**Fig. 5c**), the red and blue regions represent the nucleophilic and electrophilic sites, respectively. Obviously, the hollow site composed of two Au and two Pd atoms is the nucleophilic site. Considering that H acts as electron acceptor (**Fig. 5b**), it should be preferentially adsorbed on the nucleophilic site, which is consistent with the calculated H adsorption results. To further explain the catalytic activity difference among the (100) surfaces of Pd_xAu_{1-x}, Pd, and Au, their projected density of states (PDOS) are plotted in **Fig. 5c**. Based on the calculated PDOS, the *d*-band centre (ε_d) of the three surfaces were calculated as -1.49, -1.09, and 2.29 eV for Pd_2Au_2-d/$Pd_{0.75}Au_{0.25}$, Pd, and Au(100), respectively. These suggest that the order of H binding strength should be Pd(100) > Pd_2Au_2-d/$Pd_{0.75}Au_{0.25}$(100) > Au(100), which is consistent with their free energy variations (**Fig. 5a**). The H bindings on the surface of Pd and Au(100) are respectively too strong and too weak, which result in lower catalytic activities as compared to their alloyed surfaces. The *d* electron distribution of Pd_xAu_{1-x}(100) was also regulated through the tuning of composition. The H binding strength on Pd_2Au_2-d/$Pd_{0.75}Au_{0.25}$(100) is moderate, leading to a higher catalytic activity of HER which drops close to the peak of the volcano. Furthermore, the stability of Pd_2Au_2-d/$Pd_{0.75}Au_{0.25}$(100) was analysed using *ab initio* molecular dynamics (AIMD) simulation with 10 ps at the working temperature of 300 K (**Fig. 5e**). There was no obvious fluctuation in the key bond lengths and system energies, suggesting that Pd_2Au_2-d/$Pd_{0.75}Au_{0.25}$(100) has good stability with a similar experimental process. In addition, the stability of Pd_2Au_2-d/$Pd_{0.75}Au_{0.25}$(100) at higher temperature (400 and 500 K) were also studied through the AIMD simulations (**Fig. S3**), showing a similar conclusion.

View Article Online
DOI: 10.1039/D0TA04615A

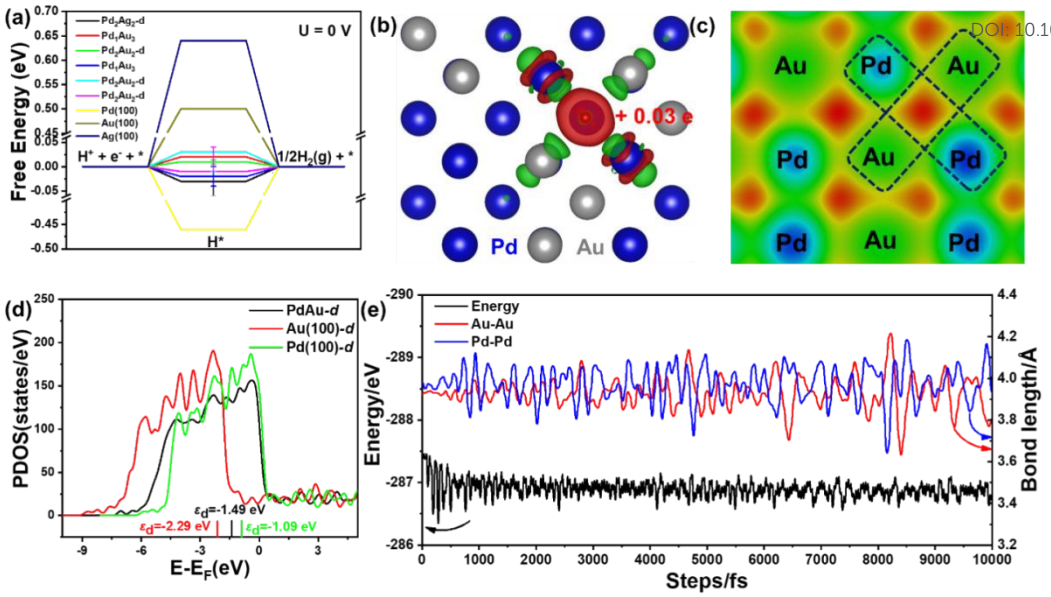


Fig. 5 (a) Free energy diagram of acidic HER at Pd, Au, and Ag(100), and at different ensembles on Pd_xAg_{1-x} and Pd_xAu_{1-x}(100). Each error bar is the standard deviation calculated from at least three randomly sampled sites from the generated surfaces. (b) EDD of H binding on Pd_2Au_2-d/$Pd_{0.75}Au_{0.25}$(100) (isosurface = 0.002 e/Å^3) Red and green regions represent the increase and decrease of electrons, respectively. (c) EPD of Pd_2Au_2-d/$Pd_{0.75}Au_{0.25}$(100) (isosurface = 0.002 e/Bohr^3). (d) PDOS of the binding sites at Pd_2Au_2-d/$Pd_{0.75}Au_{0.25}$, Au, and Pd(100). Red and blue regions represent the nucleophilic and electrophilic sites, respectively. (e) Energy and bond length variation of Pd_2Au_2-d/$Pd_{0.75}Au_{0.25}$ under AIMD simulations at 300 K.

3.4 ML Modelling

As an extensive section of this study, we further developed ML models based on our calculated database, for the rapid screening of out-of-sample (100) alloy catalysts. The modelling details are shown in ***Section 2.2***. Before ML modelling, we performed analysis on the feature selection (**Fig. 6**), which helps to determine the most important input variables for the prediction of H binding energy. As shown in **Fig. 1b**, there are a number of physical features (*e.g.*, input variables) that can be considered for the ML modelling (with the completed list of our input variables shown in **Table S2**). Obviously, there exist some slightly correlated features in the input variables according to the correlation map depicted in **Fig. 6a**. Therefore, it is necessary for us to make an appropriate feature reduction. To minimize the dimension of features, the extremely randomized trees (ERT) method derived by Geurts *et al*. [57] is one of the most commonly used methods to realize feature selection.[58-60] Different from other tree-based ensemble methods, ERT chooses the cut-points to split each nodes fully at random.[61-63] Besides, it uses the whole training samples to build the "tree" rather than bootstrapped samples.[64, 65] From the bias-variance perspective, introducing randomness of cut-point combined with ensemble methods should reduce the variance more significantly.[64, 65] Meanwhile, using the whole training samples can significantly reduce the bias.[65, 66] Considering that it is stable and robust to noisy data both empirically and theoretically,[57] we adopted it as our method for feature selection. Using this method, we ranked all the features and selected the top 26 features out of the total 29 features (**Fig. 6b**). We determined the number of features by gradually increasing the number of features and observed the absolute mean square error (MSE) on the training data. Once we finished feature selection, the reduced data set were fed to our machine learning models. More details of the ERT method can be found in Ref. [57].

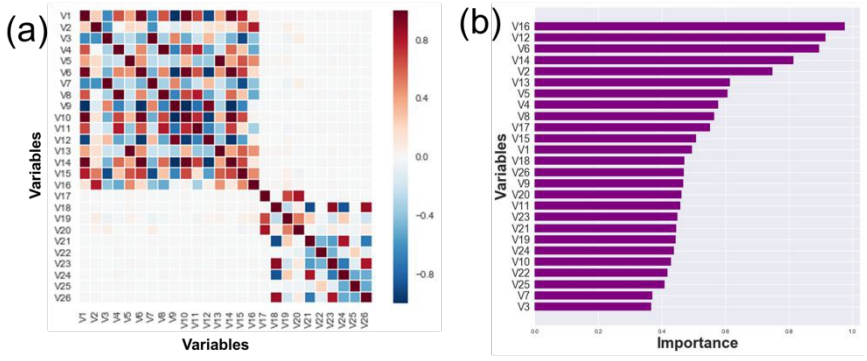


Fig. 6 Feature analysis of the ML model. (a) Correlation analysis of the input variables and (b) calculated feature ranking with the ERT method. The completed list of the input variables can be found in **Table S2**.

View Article Online
DOI: 10.1039/D0TA04615A

Using the selected features as the input variables, we performed BPNN modelling using our DFT database of (100) surfaces. As shown in **Fig. 7**, good accuracy can be found in both the training and test sets. The training and test scores with different combinations are all above 0.98, with the MSEs all less than 0.01. The average errors of ML prediction are all relatively low, with the predictive model approaching the DFT-PBE level. Cross-validation and sensitivity tests were also performed and showed robust modelling results using the current database, input variables, and BPNN hyper-parameters.[54] Noted that we have performed hundreds of repeated training and testing of the model using the current database, with various hyper-parameters (*e.g.*, different initial weights of the network); no significant change was found on the MSEs in the dataset. Based on this ML model, H binding energies on Ir and Ru alloys with different compositions were predicted and examined, as shown in the our online database.[51] Among the ~900 predicted bimetallic alloy structures, the average H binding energies on Ru_2Ag_2-d/$Ru_{0.25}Ag_{0.75}$(100) and Pt_4/$Ir_{0.75}Pt_{0.25}$(100) locate at the HER volcano peak (-0.24 eV), suggesting that these catalysts should also possess high catalytic activities for HER with a zero reaction free energy in the HER steps (**Fig. S4**). Therefore, it is expected that the predicted catalysts can be potential candidates for HER, which is worthy of further experimental research. Our future research will further focus on the ML-assisted design of bimetallic alloy catalysts.

We also note that there was a pioneering study which mainly focused on HER on intermetallic alloy systems using combined DFT and ML modelling.[67] In our study, different from intermetallic alloys, we used a random sampling method to evaluate various randomly generated alloy structures, and analysed their average binding energies and variations (*i.e.*, error bars). Previous combined experimental and theoretical studies showed that the study of random alloy structures can help better understand the uncertainties in an alloy catalyst synthesized by kinetically-controlled synthesis methods.[9-11, 13, 14, 16, 68] Since many of the alloys are meta-stable (especially for the alloys containing (100) facet), we expect that the study on various randomly-mixed structures are more representative. Also, study using intermetallic alloy surfaces sometimes may dismiss some alloy ensembles (*e.g.*, the pure M_4 site in an alloyed M-N system) due to the orderly alloyed structure. However, these specific sites may generally exist in an alloy catalyst synthesized by kinetically-controlled synthesis. But our sampling method, which considers all different ensemble symmetries for all the compositions, can help study the activity of these sites. Meanwhile, some of the previous studies also indicated that the results of intermetallic alloys also fall within the variation of our sampled results,[4, 16] suggesting that our studied systems can provide generalized understandings to the nature of synthesized alloy catalysts.

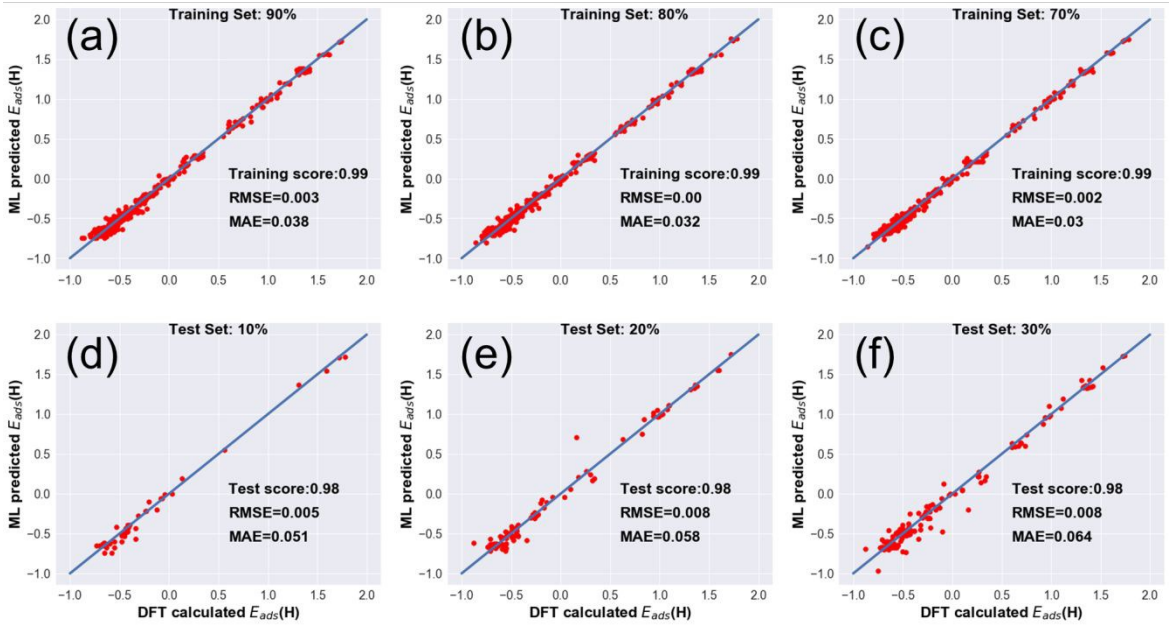


Fig. 7 Selected representative training and testing results of DFT calculated *vs.* predicted H binding energies, with the data composition of (a,d) 90% training + 10% testing, (b,e) 80% training + 20% testing, and (c,f) 70% training + 30% testing. Other ML modelling results can be found in our online database.[51]

4. Conclusion

In this paper, we have shown that some of the (100) surfaces alloyed by strong- (Pd and Pt) and weak-binding (Ag, Au, and Cu) transition metals can be promising HER catalysts in acidic media, using high-throughput DFT calculations and catalytic modelling. Especially, we found that the (100) surfaces of Pd_xAg_{1-x} and Pd_xAu_{1-x} have promising HER activities due to the presence of highly active four-fold ensembles which reach the peak of the HER volcano plot, compared to their monometallic counterparts and other evaluated alloys. Electronic and stability analyses were performed on the surfaces which are predicted

with promising activities. Finally, as an extensive section, ML modelling was performed based on the DFT-calculated database for alloyed (100) structures, showing a promising predictive power which guarantees rapid screening of future bimetallic catalysts. Most importantly, this combined DFT, catalytic analysis, and ML modelling study suggests that the catalytic activity of (100) dominated catalysts for HER can be significantly improved through the rational design of alloy catalysts. We expect that this study provides helpful guideline for future experiments. Our Python codes, ML database, and optimized geometries can be found in the GitHub repository: https://github.com/alloycat/database.

Conflicts of interest

There are no conflicts to declare.

Acknowledgement

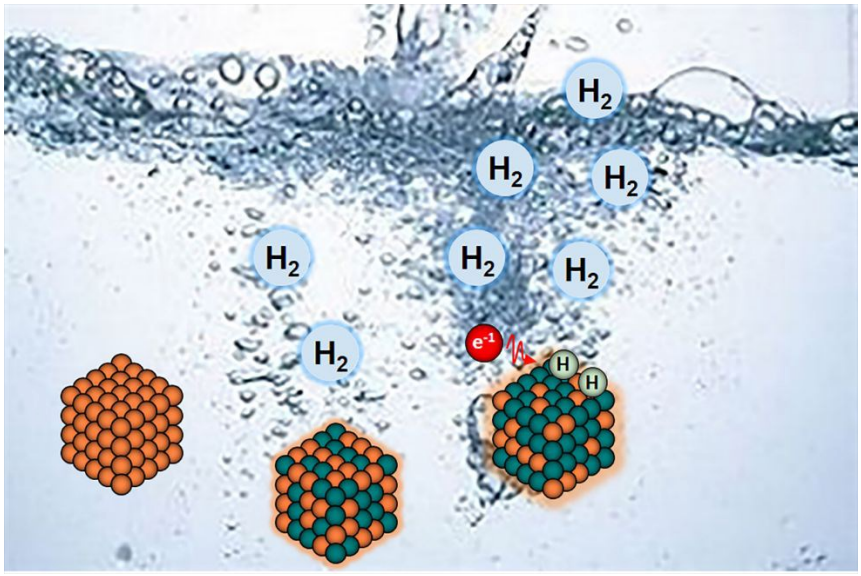
This work was supported by Beijing Natural Science Foundation (No. 2182066), Natural Science Foundation of Hebei Province of China (No. B2018502067 and No. E2020502023), and the Fundamental Research Funds for the Central Universities (2020MS104). All of the computations were performed in The Computational Centre for North China Electric Power University. Hao Li thanks the research funds and computational resources provided by North China Electric Power University. We thank Dr. Ryan Ciufo (The University of Texas at Austin, USA) for improving the writing of our manuscript.

Reference

1. Q. Chen, Y. Yang, Z. Cao, Q. Kuang, G. Du, Y. Jiang, Z. Xie and L. Zheng, *Angew. Chem. Int. Edit.*, 2016, **55**, 9021-9025.
2. Z. Peng, H. You, J. Wu and H. Yang, *Nano Lett.*, 2010, **10**, 1492-1496.
3. Y. Kang and C. B. Murray, *J. Am. Chem. Soc.*, 2010, **132**, 7568-7569.
4. H. Li, K. Shin and G. Henkelman, *J. Chem. Phys.*, 2018, **149**, 174705.
5. Y.-N. Wen and J.-M. Zhang, *Solid State Commun.*, 2007, **144**, 163-167.
6. T. K. Sau and C. J. Murphy, *Langmuir*, 2005, **21**, 2923-2929.
7. Y. Fu, P. Richardson, K. Li, H. Yu, B. Yu, S. Donne, E. Kisi and T. Ma, *Nano-Micro Lett.*, 2020, **12**, 65.
8. S.-W. Chou, C.-L. Zhu, S. Neeleshwar, C.-L. Chen, Y.-Y. Chen and C.-C. Chen, *Chem. Mater.*, 2009, **21**, 4955-4961.
9. H. Li, S. Guo, K. Shin, M. S. Wong and G. Henkelman, *ACS Catal.*, 2019, **9**, 7957-7966.
10. L. Luo, Z. Duan, H. Li, J. Kim, G. Henkelman and R. M. Crooks, *J. Am. Chem. Soc.*, 2017, **139**, 5538-5546.
11. H. Li, L. Luo, P. Kunal, C. S. Bonifacio, Z. Duan, J. C. Yang, S. M. Humphrey, R. M. Crooks and G. Henkelman, *J. Phys. Chem. C*, 2018, **122**, 2712-2716.
12. S. García, L. Zhang, G. W. Piburn, G. Henkelman and S. M. Humphrey, *ACS Nano*, 2014, **8**, 11512-11521.
13. H. Guo, H. Li, K. Jarvis, H. Wan, P. Kunal, S. G. Dunning, Y. Liu, G. Henkelman and S. M. Humphrey, *ACS Catal.*, 2018, **8**, 11386-11397.
14. H. Guo, H. Li, D. Fernandez, S. Willis, K. Jarvis, G. Henkelman and S. M. Humphrey, *Chem. Mater.*, 2019, **31**, 10225-10235.
15. H. Guo, Z. Fang, H. Li, D. Fernandez, G. Henkelman, S. M. Humphrey and G. Yu, *ACS Nano*, 2019, **13**, 13225-13234.
16. H. Guo, J. A. Trindell, H. Li, D. Fernandez, S. M. Humphrey, G. Henkelman and R. M. Crooks, *J. Mater. Chem. A*, 2020, DOI: 10.1039/C9TA13711D.
17. P. Liu and J. K. Nørskov, *Phys. Chem. Chem. Phys.*, 2001, **3**, 3814-3818.
18. M. Mavrikakis, B. Hammer and J. K. Nørskov, *Phys. Rev. Lett.*, 1998, **81**, 2819.
19. Q. Li, L. Song, L. Pan, X. Zhuang, M. Ling and L. Duan, *Phys. Chem. Chem. Phys.*, 2013, **15**, 20345-20353.
20. X. Yuan, L. Zhang, L. Li, H. Dong, S. Chen, W. Zhu, C. Hu, W. Deng, Z.-J. Zhao and J. Gong, *J. Am. Chem. Soc.*, 2019, **141**, 4791-4794.
21. M. Liu, Y. Lu and W. Chen, *Adv. Funct. Mater.*, 2013, **23**, 1289-1296.
22. N. Yang, Z. Zhang, B. Chen, Y. Huang, J. Chen, Z. Lai, Y. Chen, M. Sindoro, A.-L. Wang, H. Cheng, Z. Fan, X. Liu, B. Li, Y. Zong, L. Gu and H. Zhang, *Adv. Mater.*, 2017, **29**, 1700769.
23. Z. Peng, H. You and H. Yang, *Adv. Funct. Mater.*, 2010, **20**, 3734-3741.
24. A. Kloke, C. Köhler, R. Gerwig, R. Zengerle and S. Kerzenmacher, *Adv. Mater.*, 2012, **24**, 2916-2921.
25. T. Kwon, H. Hwang, Y. J. Sa, J. Park, H. Baik, S. H. Joo and K. Lee, *Adv. Funct. Mater.*, 2017, **27**, 1604688.
26. J. Bai, Q. Xue, Y. Zhao, J.-X. Jiang, J.-H. Zeng, S.-B. Yin and Y. Chen, *ACS Sustain. Chem. Eng.*, 2019, **7**, 2830-2836.
27. A. Yang, T. Li, S. Jiang, X. Wang, X. Qiu, W. Lei and Y. Tang, *Nanoscale*, 2019, **11**, 14561-14568.
28. S. D. House, C. S. Bonifacio, J. Timoshenko, P. Kunal, H. Wan, Z. Duan, H. Li, J. C. Yang, A. I. Frenkel, S. M. Humphrey, R. M. Crooks and G. A. Henkelman, *Microsc. Microanal.*, 2017, **23**, 2030-2031.
29. Z. Duan, J. Timoshenko, P. Kunal, S. D. House, H. Wan, K. Jarvis, C. Bonifacio, J. C. Yang, R. M. Crooks, A. I. Frenkel, S. M. Humphrey and G. Henkelman, *Nanoscale*, 2018, **10**, 22520-22532.
30. L. Li, X. Li, Z. Duan, R. J. Meyer, R. Carr, S. Raman, L. Koziol and G. Henkelman, *Nanoscale*, 2019, **11**, 10524-10535.
31. D. D. Rodene, E. H. Eladgham, R. B. Gupta, I. U. Arachchige and V. Tallapally, *ACS Appl. Energ. Mater.*, 2019, **2**, 7112-7120.
32. Y. Niu, X. Qian, C. Xu, H. Liu, W. Wu and L. Hou, *Chem. Eng. J.*, 2019, **357**, 11-20.
33. Z. Ye, Y. Qie, Z. Fan, Y. Liu, Z. Shi and H. Yang, *Dalton Trans.*, 2019, **48**, 4636-4642.
34. M. Zhao, Y. Wu, W. Cai, T. Xia, W.-J. Jiang, W. Ding and J.-P. Cao, *Int. J. Hydrogen Energ.*, 2019, **44**, 31053-31061.
35. N. Takehiro, P. Liu, A. Bergbreiter, J. K. Nørskov and R. J. Behm, *Phys. Chem. Chem. Phys.*, 2014, **16**, 23930-23943.
36. J. Greeley, T. F. Jaramillo, J. Bonde, I. Chorkendorff and J. K. Nørskov, *Nat. Mater.*, 2006, **5**, 909-913.
37. J. Greeley and J. K. Nørskov, *Surf. Sci.*, 2007, **601**, 1590-1598.
38. M. E. Björketun, A. S. Bondarenko, B. L. Abrams, I.

Chorkendorff and J. Rossmeisl, *Phys. Chem. Chem. Phys.*, 2010, **12**, 10536-10541.
39. V. Fung, G. Hu and B. Sumpter, *J. Mater. Chem. A*, 2020, **8**, 6057-6066.
40. I.-C. M. Oğuz, Tzonka ; Guesmi, Hazar, *J. Chem. Phys.*, 2018, **148**, 024701.
41. A. J. Medford, A. Vojvodic, J. S. Hummelshøj, J. Voss, F. Abild-Pedersen, F. Studt, T. Bligaard, A. Nilsson and J. K. Nørskov, *J. Catal.*, 2015, **328**, 36-42.
42. G. Kresse and J. Furthmüller, *Phys. Rev. B*, 1996, **54**, 11169.
43. G. Kresse and J. Furthmüller, *Comp. Mater. Sci.*, 1996, **6**, 15-50.
44. J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865.
45. H. J. Monkhorst and J. D. Pack, *Phys. Rev. B*, 1976, **13**, 5188-5192.
46. J. Wellendorff, T. L. Silbaugh, D. Garcia-Pintos, J. K. Nørskov, T. Bligaard, F. Studt and C. T. Campbell, *Surf. Sci.*, 2015, **640**, 36-44.
47. S. Grimme, *J. Comput. Chem.*, 2006, **27**, 1787-1799.
48. S. M. Foiles, M. I. Baskes and M. S. Daw, *Phys. Rev. B,* 1986, **33**, 7983-7991.
49. H. Li, E. J. Evans, C. B. Mullins and G. Henkelman, *J. Phys. Chem. C*, 2018, **122**, 22024-22032.
50. A. Hjorth Larsen, J. Jorgen Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dulak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, E. D. Hermes, P. C. Jennings, P. Bjerre Jensen, J. Kermode, J. R. Kitchin, E. Leonhard Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. Bergmann Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiotz, O. Schutt, M. Strange, K. S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng and K. W. Jacobsen, *J. Phys-Condens. Mat.*, 2017, **29**, 273002.
51. https://github.com/alloycat/database
52. J. K. Noerskov, T. Bligaard, A. Logadottir, J. R. Kitchin, J. G. Chen, S. Pandelov and U. Stimming, *J. Electrochem. Soc.* 2005, **152,** J23.
53. M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu and X. Zheng, *Proceedings of the 12th USENIX conference on Operating Systems Design and Implementation*, Savannah, GA, USA, 2016.
54. H. Li, Z. Zhang and Z.-Z. Zhao, *Processes,* 2019, **7**, 151.
55. Z. Cao, H. Li, X. Xu and J. Xu, *Chem. Eng. J.*, 2020, **394**, 124876.
56. E. J. Evans, H. Li, W.-Y. Yu, G. M. Mullen, G. Henkelman and C. B. Mullins, *Phys. Chem. Chem. Phys.*, 2017, **19**, 30578-30589.
57. Geurts, P., Damien E., Louis W., *Mach. Learn.*, 2006, **63**: 3-42.
58. Goetz M., Weber C., Bloecher J., Stieltjes B., Meinzer HP., Maier-Hein K., *Proceeding of BRATS challenge-MICCAI,* 2014.
59. Geurts P., Louppe G., *J. Mach. Learn. Res.*, 2011: 49-61.
60. Zou T., Yang W., Dai D., Sun H., *EURASIP J. Adv. Signal. Process.,* 2010, 1.
61. Liaw A, Wiener M. Classification and regression by randomForest. *R. news.* 2002, **2**, 18-22.
62. S. R. Safavian and D. Landgrebe, *IEEE T. Syst. Man. Cy.*, 1991, **21**, 660-674.
63. B. Schölkopf, Z. Luo and V. Vovk, *Empirical Inference: Festschrift in Honor of Vladimir N. Vapnik*, Berlin, Heidelberg, 2013.
64. Geurts P, Louppe G., *Proceedings of Machine Learning Research*, 2011, **14**, 49-61.
65. Friedman J, Hastie T, Tibshirani R., *The elements of statistical learning.* New York, 2001.
66. Bishop C. M, *Pattern recognition and machine learning*. 2006.
67. K. Tran and Z. W. Ulissi, *Nat. Catal.*, 2018, **1**, 696.
68. H. Li, C. Yan, H. Guo, K. Shin, S. M. Humphrey, C. J. Werth and G. Henkelman, *ACS Catal.*, 2020, DOI: 10.1021/acscatal.0c01604.

View Article Online
DOI: 10.1039/D0TA04615A



TOC