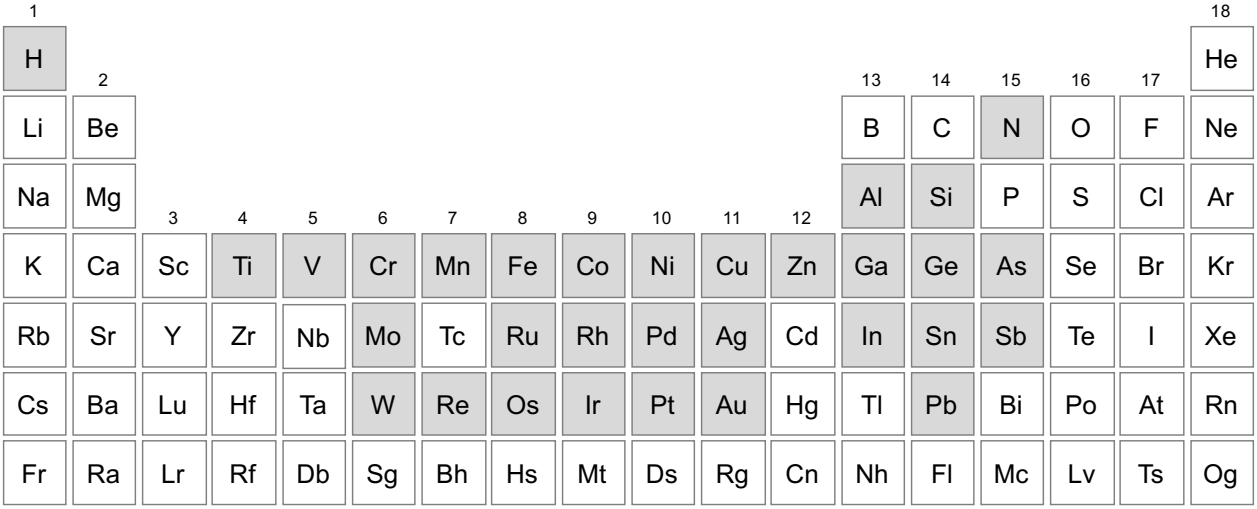
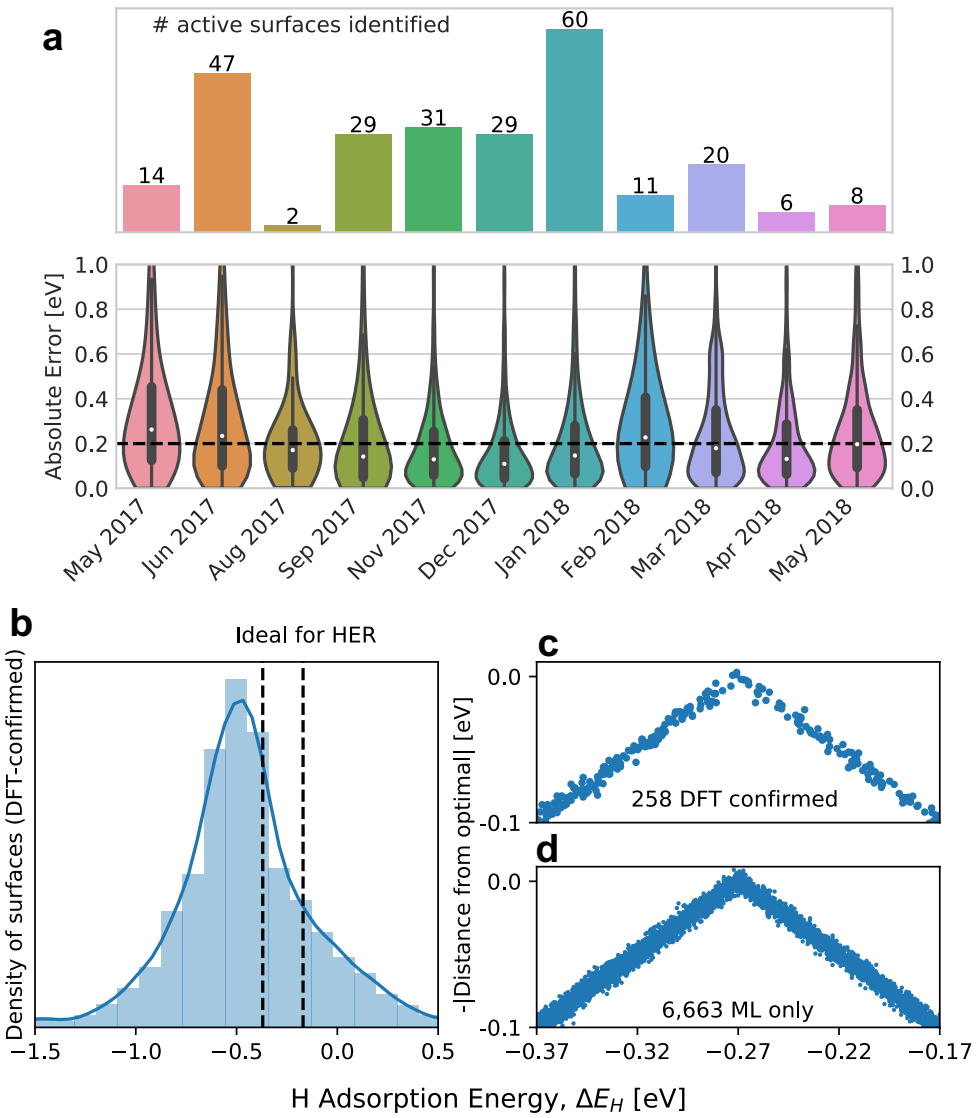
Supplementary Information for: Active learning across intermetallics to guide discovery of electrocatalysts for CO_2 reduction and H_2 evolution

Kevin Tran & Zachary W. Ulissi
Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, PA 15217
zulissi@andrew.cmu.edu

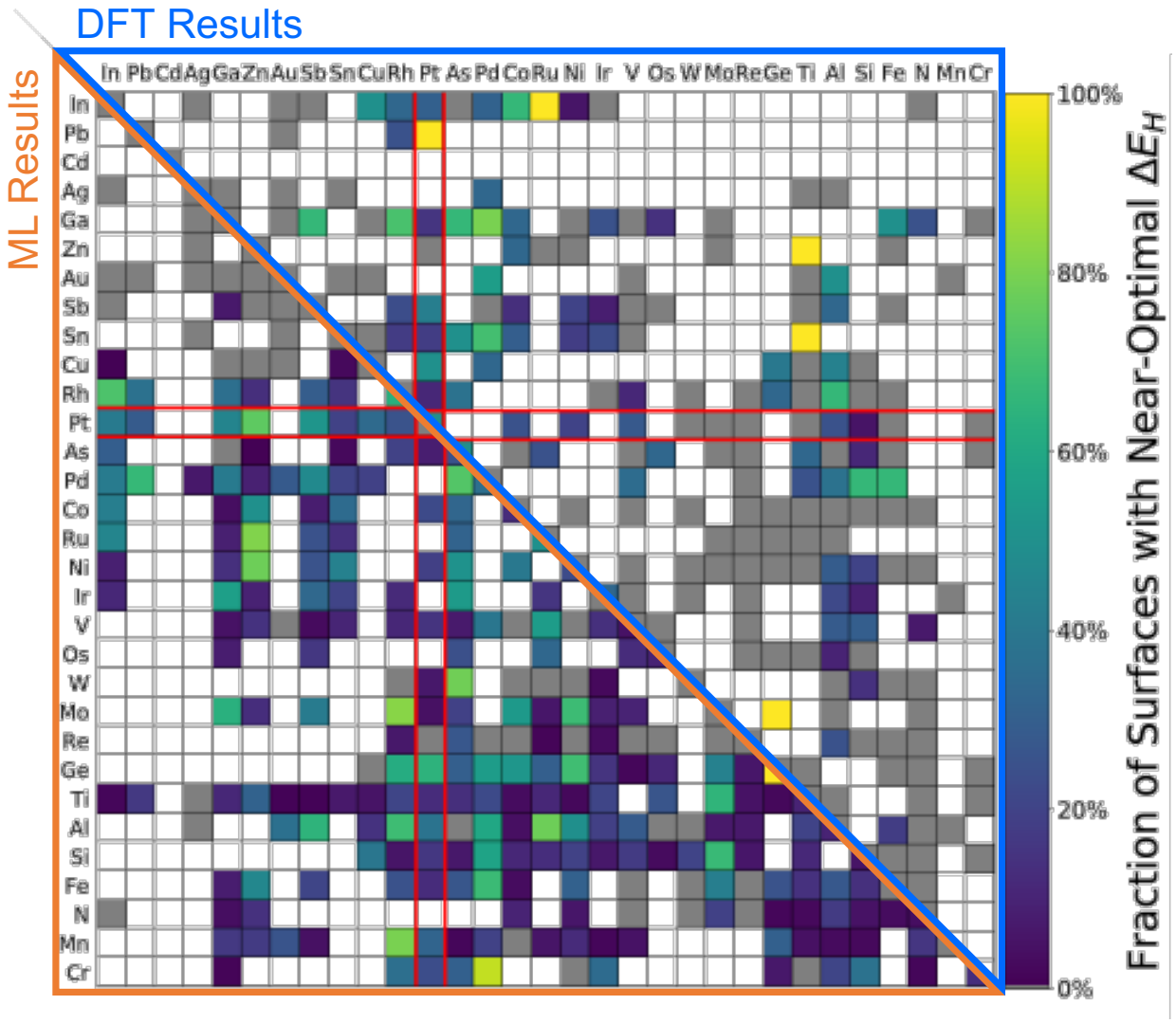
Supplementary Figures



Supplementary Figure 1 | Elements considered in this screening. Shading of an element indicates its inclusion in this screening study.



Supplementary Figure 2 | Identification of surfaces with near-optimal ΔE_{H} values for HER. a, Distribution of the number of near-optimal surfaces identified. **b,** The normalized distribution of the low-coverage H adsorption energies of all of the surfaces enumerated by this study. Dashed lines indicate the $\pm$0.1 eV range around the optimal ΔE_{H} value of -0.27 eV. **c,** Surfaces whose low-coverage H adsorption energies have been calculated and verified with DFT. **d,** Surfaces whose low-coverage H adsorption energies have been calculated only by the machine learning models.



Supplementary Figure 3 | H_2 evolution efficiency map for bimetallics. Visualization of two component intermetallics whose surfaces have low-coverage H adsorption energy (ΔE_H) values inside the range of [-0.37, -0.17] eV. White shading indicates an absence of any enumerated surfaces; grey shading indicates that all ΔE_H values are outside the range of [-0.37, -0.17] eV; and colored shading indicates possible efficiency. The ΔE_{CO} values used to create the upper half of this figure were calculated by DFT, and the values used to create the bottom half were calculated by the surrogate model.

Supplementary Notes

Supplementary Note 1

We performed a coarse study of the TPOT[1] hyperparameters and how they effect model performance. Details can be found on the Github website mentioned in the main text's Code Availability section. This coarse study yielded the following settings for using TPOT: 1 generation, a population size of 16, and an offspring size of 16. A generation size of 1 was used because the coarse study showed that the error did not change as a function of the number of generations. The study also showed that population and offspring sizes of 16 were sufficent and that changes in population/offspring were overshadowed by stochastic variations inherent in TPOT's genetic algorithm. Note that we used a median absolute deviation scoring function to allow the model to fit the majority of the data well at the expense of more severe outliers, and we used TPOT's default setting of 5-fold cross validation.

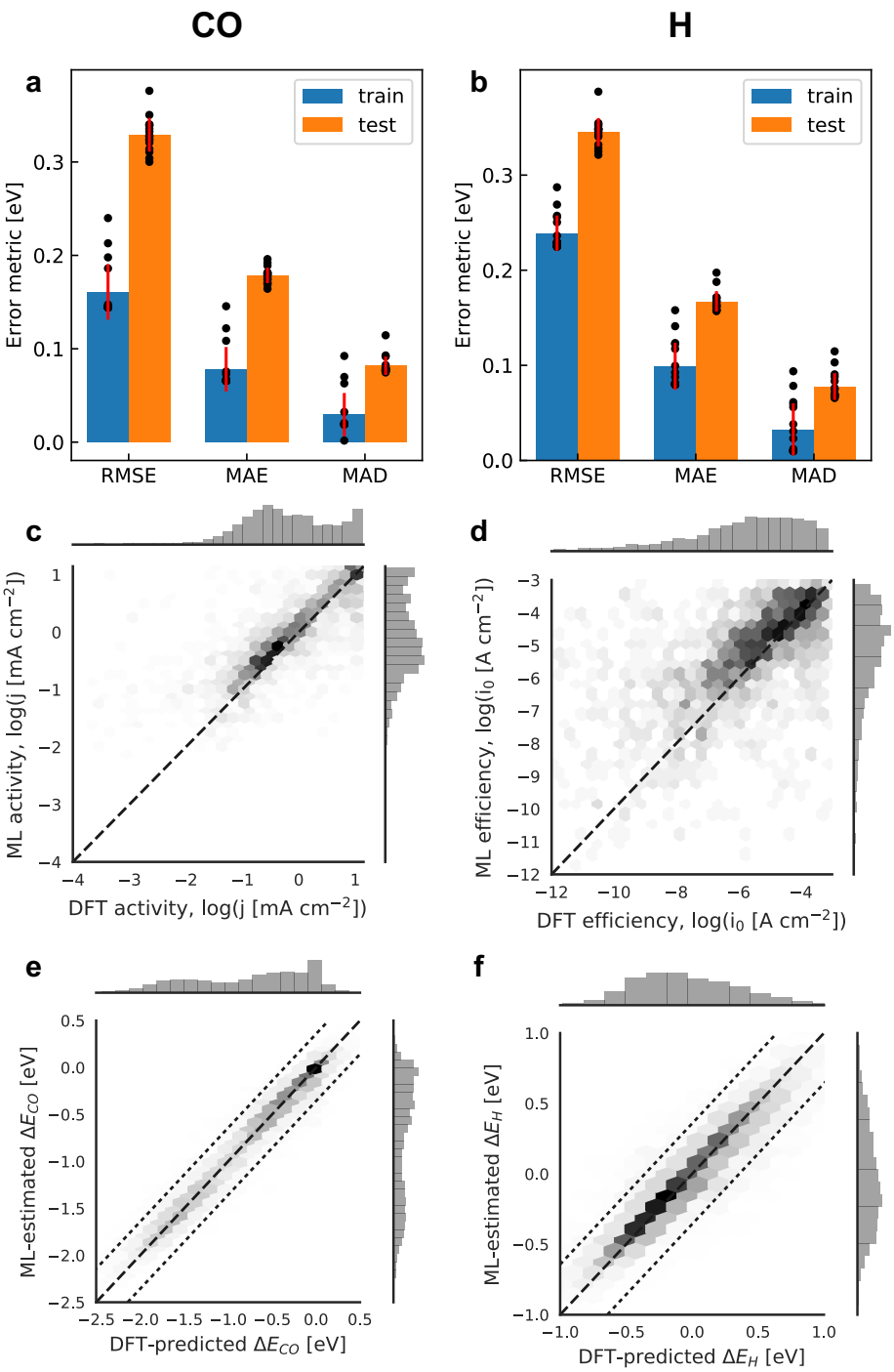
We also performed a cursory hyperparameter optimization test to verify whether or not TPOT's hyperparameter tuning worked well, because we were concerned that the resolution of hyperparameter changes that TPOT examined were too coarse to perform an optimization properly. We performed this test by comparing TPOT's hyperparameter tuning with SigOpt's tuning for one example pipeline that TPOT has chosen—Random Forest Regression. Using a 90/10 train/test split for predicting ΔECO, TPOT's choice of hyperparameters yielded a test RMSE, MAE, and MAD of 0.32, 0.18, and 0.10 eV, respectively. SigOpt yielded a test RMSE, MAE, and MAD of 0.33, 0.19, and 0.10 eV, respectively. These results suggest that TPOT's hyperparameter tuning is sufficient for our use-case.

Supplementary Note 2

Static surrogate models are typically judged via train/test errors and parity plots. These were omitted from the main manuscript in favor of evaluations that focus on dynamic, iterative workflows—i.e., evaluation on a rolling forecast origin.[2] We still show the classical evaluations here for the sake of completeness.

We took the 19,644 DFT-calculated ΔE_{CO} values that we have and performed 16 different stratified 90/10 train/test splits,[3] and then we used TPOT create 16 different models from these splits. We repeated this for our 23,141 ΔE_{H} calculations, resulting in 32 total models. Of the 32 models, 19 of them were K-nearest-neighbor models, 3 were gradient boosting regressors, and 10 were random forest or extra trees regressors. The subsequent distributions of root-mean-squared-error (RMSE), mean absolute error (MAE), and median absolute deviation (MAD) are shown in Supplementary Figures 4a and b for ΔE_{CO} and ΔE_H, respectively.

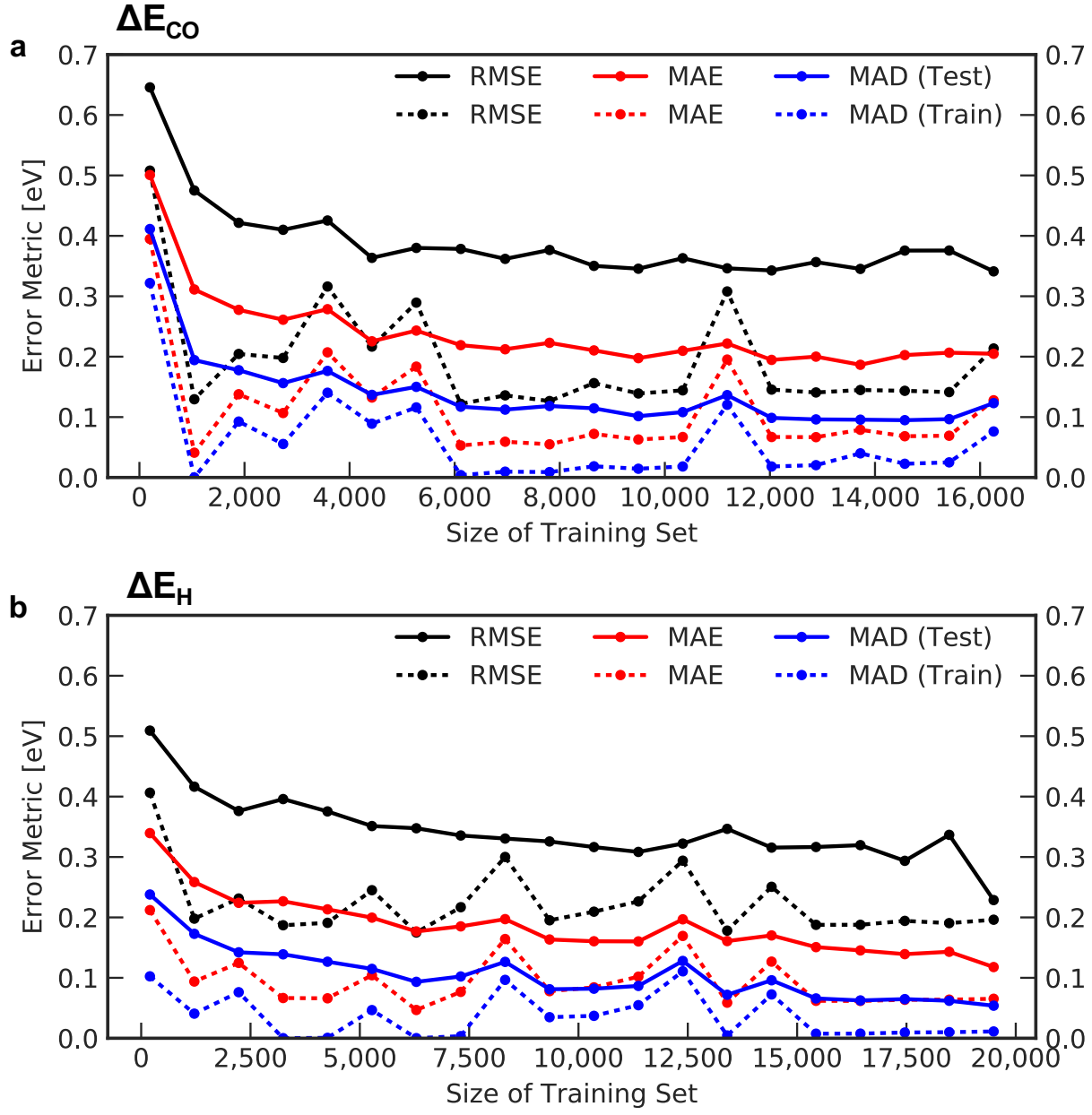
We then used TPOT to train final models on 100% of our DFT data—i.e., 19,644 for ΔE_{CO} and 23,141 for ΔE_{H}. When predicting ΔE_{CO}, TPOT identified an extra trees regressor with a 5-fold cross-validation (CV) MAD of 0.09 eV. This regressor was created with a maximum feature splitting proportion of 0.95, a minimum of 5 samples per leaf, a minimum of 6 samples per split, 100 trees, and is fit to minimize RMSE. When predicting ΔE_{H}, TPOT identified a random forest regressor with a 5-fold CV MAD of 0.07 eV. This regressor was created with a maximum feature splitting proportion of 0.65, a minimum of 11 samples per leaf, a minimum of 18 samples per split, 100 trees, and is fit to minimize RMSE. We then converted the final models' adsorption energy predictions to performance metrics using scaling relations.[4,5] The subsequent parity plots of catalyst performance are shown in Supplementary Figures 4c and d for CO_2RR and HER, respectively. Darker shading indicates a higher density of points. Lastly: We created parity plots for this model's predictions of ΔE_{CO} and ΔE_{H} in Supplementary Figures 4e and f, respectively.



Supplementary Figure 4 | Surrogate modeling performance of final model. Performance metrics for 16 models created from 90/10 train/test splits on all DFT data for predicting **(a)** ΔE_{CO} (n=19,644) and **(b)** ΔE_H (n=23,141). Black points represent results for each of the 16 models, bar heights indicate average values, and red lines indicate average values ± 1 standard deviation. Parity plots when predicting catalyst performance for **(c)** CO_2RR and **(d)** HER. Parity plots when predicting **(e)** ΔE_{CO} and **(f)** ΔE_{H}. Darker shading indicates a higher density of points; dashed lines indicate parity; and dotted lines indicate 95% prediction intervals—i.e., parity$\pm 2 \times$ MAE. The parity plots in **(c-f)** were generated using a model that was trained using a 100/0 train/test split.

It is common with surrogate modeling to calculate the modeling error as a function of training set size—

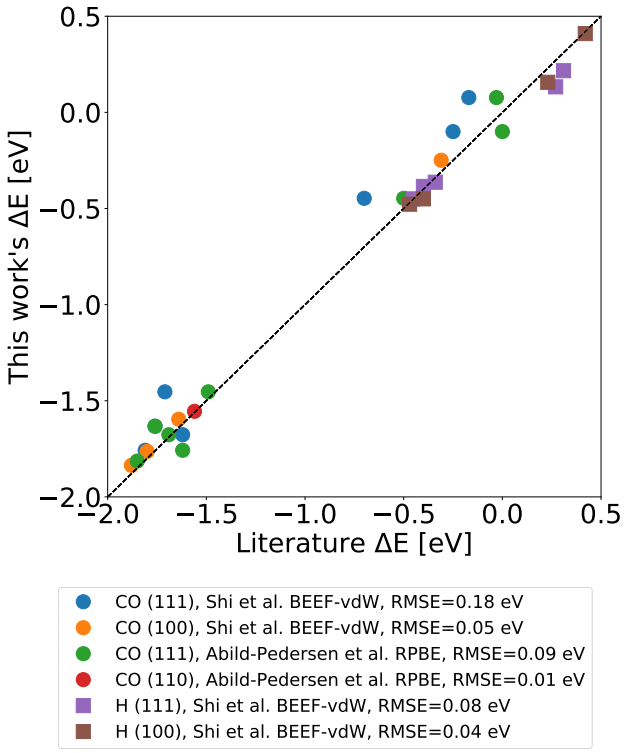
i.e., a learning curve. Learning curves give the modeler a quantitative estimate for how much training data was actually necessary to converge on a final model. We do not have a final model because we are not yet finished exploring (or even defining) our search space, and so any learning curve we make may not illustrate properly the amount of data needed to converge. We can still show retrospective learning curves as if we treated our current DFT data set as if it was our entire search space—Supplementary Figure 5. These figures were created by fitting 20 different models using varying sizes of training data that were randomly selected from our database of DFT results. Test errors were calculated from the remaining data—e.g., given a training size of 2,000 points from a pool of 19,000, the test size would be 17,000 points.



Supplementary Figure 5 | Learning curves. Error metrics vs. size of training set when calculating adsorption energies for **(a)** CO and **(b)** H. RMSE is the root mean squared error; MAE is the mean absolute error; and MAD is the median absolute deviation.

Supplementary Note 3

To verify the integrity of our 42,785 DFT calculations, we compared a subset of our results for pure metal calculations on simple facets with RPBE calculations found in literature,[6] including sources[7] that used the BEEF-vdW functional (Supplementary Figure 6). Some of the differences between our results and literature results may be caused by our frameworks' neglect in accounting for spin magnetism for magnetic elements like Fe and Ni, in contrast to other literature methods.



Supplementary Figure 6 | Benchmarking of DFT results. The adsorption energies (ΔE) of various adsorbate/surface combinations calculated by our framework vs. the energies calculated by various literature sources.[6,7] The diagonal line indicates parity.

Supplementary Methods

Calculation of free energies

Supplementary Equation 1 can be used to calculate the change in chemical potential (ΔG_X) during adsorption of arbitrary adsorbate X,[8] which is the descriptor that this study uses to predict catalyst performance.[4,5] μ_{X*} is the chemical potential of the adsorbate bound on a surface, μ_* is the chemical potential of the surface, and μ_X is the chemical potential of the adsorbate in gas phase.

$$\Delta G_X = \mu_{X*} - \mu_* - \mu_X \tag{1}$$

Chemical potentials can be calculated[5,8] using Supplementary Equation 2, where E is the electronic energy as calculated by density functional theory (DFT), ZPE is the zero-point energy, C_p is the heat capacity, T is temperature, S is entropy, $\Delta\mu_{solv}$ is the change in chemical potential from solvent [de]stabilization, and $\Delta\mu_{corr}$ is any experimental correction required to account for differences between experimental chemical potentials and DFT-based chemical potentials.

$$\mu = E + ZPE + \int C_p dT - TS + \Delta\mu_{solv} + \Delta\mu_{corr} \tag{2}$$

Literature[9] reports that ZPE_{CO} is 0.130 eV, $\int C_{p,\,CO}dT$ is 0.091 eV at 298 K, S_{CO} is 0.002092 eV K^{-1}, and that $\Delta\mu_{exp,\,CO}$ is 0.02 eV for gas phase CO. Thus the chemical potential of CO in the gas phase is:

$$\begin{aligned} \mu_{CO,\,g} &= E_{CO,\,g} + ZPE_{CO,\,g} + \int C_{p,\,CO}dT - TS_{CO,\,g} + \Delta\mu_{solv,\,g} + \Delta\mu_{corr,\,g} \\ \mu_{CO,\,g} &= E_{CO,\,g} + 0.130\ eV + 0.091\ eV - (298 \cdot 0.002092)\ eV + 0\ eV + 0.02\ eV \\ \mu_{CO,\,g} &= E_{CO,\,g} - 0.362\ eV \end{aligned} \tag{3}$$

The same method can be used to calculate the chemical potential of CO in the adsorbed-state (μ_{CO*}) given literature values[9] for ZPE, C_p, and S as well as a solvation correction term[5] (μ_{solv}) and an empirical correction correction term[6] (μ_{corr})—see Supplementary Equation 4. Note that we assumed a CO stretch frequency of 2,000 cm^{-1}, and deviations from this frequency may cause errors.

$$\begin{aligned} \mu_{CO*} &= E_{CO*} + ZPE_{CO*} + \int C_p dT - TS_{CO*} + \Delta\mu_{solv} + \Delta\mu_{corr} \\ \mu_{CO*} &= E_{CO*} + 0.192\ eV + 0.085\ eV - (298 \cdot 0.000452)\ eV - 0.2\ eV + 0.2\ eV \\ \mu_{CO*} &= E_{CO*} + 0.142\ eV \end{aligned} \tag{4}$$

Assuming that μ_* is equal to E_* and by combining Supplementary Equations 1, 3, and 4, we can calculate the change in chemical potential of adsorbing CO:

$$\begin{aligned} \Delta G_{CO} &= \mu_{CO*} - \mu_* - \mu_{CO} \\ \Delta G_{CO} &= [E_{CO*} + 0.142\ eV] - E_* - [E_{CO,\,g} - 0.362\ eV] \\ \Delta G_{CO} &= E_{CO*} - E_* - E_{CO,\,g} + 0.50\ eV \\ \Delta G_{CO} &= \Delta E_{CO} + 0.50\ eV \end{aligned} \tag{5}$$

where the change in electronic energy, ΔE_{CO}, is defined using Supplementary Equation 6 and is calculated with DFT.[8]

$$\Delta E_X = E_{X*} - E_* - E_{X,\,g} \tag{6}$$

We combined Supplementary Equation 5 with the optimal ΔG_{CO} found in literature,[5] -0.17 eV, to calculate the optimal ΔE_{CO}: -0.67 eV. Note that the optimal ΔG_{CO} of -0.17 eV was based on microkinetic modeling of the reaction pathway from CO_2 to methane on single metal surfaces where the rate-limiting step was hydrogenation of CO, which is why the binding energy of CO is used as a descriptor of activity. By using this descriptor, we assume that CO is still involved in the rate-limiting step of CO_2 reduction. This is acceptable because our framework is used only as a screening method. If we want to find a catalyst for a reaction where CO is not a key intermediate in a rate-limiting step, then we can use our framework to model the adsorption of whichever intermediate is appropriate. Note also that the source of the CO_2RR scaling relationship used the BEEF functional[10] while our work used the RPBE functional,[11] which may impart errors in our estimates of activity.

In contrast to the methods we used to find the optimal ΔE_{CO} for CO_2 reduction, the literature sources[4,12] we used for the HER already performed the chemical potential calculations, allowing us to use their energy calculation methods directly:

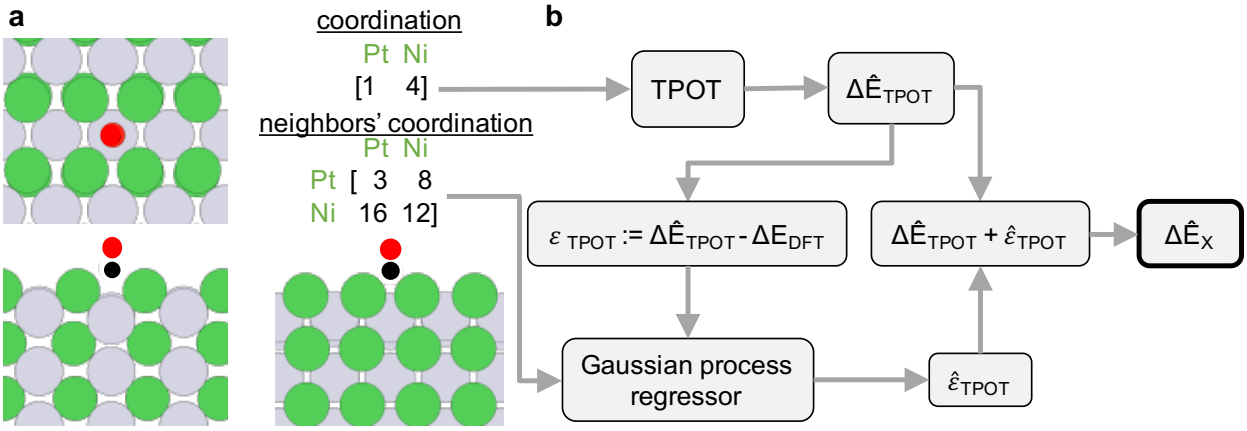
$$\Delta G_H = \Delta E_H + 0.24\ eV \tag{7}$$

where the optimal ΔG_H value is -0.03 eV, yielding an optimal ΔE_H value of -0.27 eV.

Previous surrogate modeling methods

We used various methods for fingerprinting adsorption sites and various regression methods during the development of our workflow. The primary method used during development involved both a different fingerprint and a different regression method than the one described in the main text. This alternative fingerprint was composed of a vector of coordination numbers for each element and an array of neighbors' coordination numbers as determined by Voronoi tesselation performed by pymatgen.[13] Supplementary Figure 7a illustrates a simplified example of the fingerprint. The coordination vector contains one item for each of the 31 elements

considered in the screening, and each item in this vector is the sum of the number of atoms of that element that are coordinated with the adsorbate. The neighbors' coordination vector is a flattened array that contains 31×31 items for each elemental pairing, and each item in this vector is the sum of the number of atoms of one element that are coordinated with the all adsorbate neighbors that belong to another element. Algorithm 1 outlines explicitly how we calculated the neighbors' coordination number.



Supplementary Figure 7 | Alternative method for modeling adsorption energy. a, Alternative fingerprinting method. Each item in the "coordination" vector represents the coordination number for a particular element, e.g., Pt or Ni. Each item in the "neighbors' coordination" array represents this same coordination vector, but for each of the adsorbate's neighbors. Note that this example is for illustrative purposes only. The vectors and arrays actually used contained enough items to represent 31 different elements. **b,** Old method for performing regression. Coordination was used by TPOT to estimate adsorption energy, then the residuals of that model were combined with the neighbors' coordinations and used by a Gaussian process regressor to estimate residuals of the first model. Summation of the results of the first and second models yielded the final estimate.

Algorithm 1 Calculating array of neighbors' coordination numbers for an adsorption site

```
ncn := zeros(n, n)                ▷ where n=31, the total number of elements we are investigating
for all neighbors do
    i := index of the element of neighbor
    for all neighbor's neighbors do
        j := index of the element of neighbor's neighbor
        ncn_{i,j} := ncn_{i,j} + 1
```

The alternative regression method used these coordination and neighbors' coordination fingerprints in a multi-staged regression approach to predict adsorption energies (Supplementary Figure 7b). In the first stage, we used TPOT[1] to regress the DFT-calculated adsorption energies (ΔE_{DFT}) against the vector of coordination numbers. In the second layer, we used SKLearn's[14] Gaussian processor to regress the residuals of the first stage's estimates ($\hat{\epsilon}_{TPOT}$) against the neighbors' coordination numbers. The dimensionality of the neighbors' coordination numbers was reduced by using principal component analysis as implemented in SKLearn. We then added the estimates of the models from both stages to yield a final estimate for adsorption energy ($\Delta\hat{E}_{X*}$).

These alternative fingerprinting and regression methods were eventually forgone in favor of methods that yielded lower test errors. Test set root mean squared error, mean absolute error, and median absolute deviation values were approximately 0.91, 0.23, and 0.14 eV, respectively. A stratified, 90/10 train/test split was used to build the model that yielded these errors.

t-SNE methods

We then used the t-SNE algorithm[15] to visualize our data by reducing all of the DFT-modeled adsorption sites into a 2-dimensional representation. We did so by taking the fingerprints illustrated in Supplementary Figure 7 and then scaling them such that each feature had a mean of zero and a variance of one, and then sending the fingerprints through SKLearn's[14] principle component analyzer where the dimensionality was reduced until only 85% of the variance was accounted for, yielding a vector of 113 items instead of 31×32 items. This reduced vector was processed by the t-SNE algorithm[15] with a perplexity of 50 and a learning rate of (ϵ) of 750 and then stopped after 3,000 iterations, yielding a 2-dimensional reduction of the adsorption sites. Note that the fingerprint in Supplementaly Figure 7a was still used to create the t-SNE diagrams instead of the fingerprint in Figure 2, because Supplementary Figure 7's fingerprint codified adsorption sites only by the identities and the numbers of neighboring atoms. Thus the sites in the reduced space were described only by elemental identities and coordination numbers, which we assumed to be more representative of our controllable materials design space.

Supplementary Tables

When using adsorption energies as descriptors of catalyst performance, it is common to characterize an entire surface by calculating the adsoprtion energies of all the adosorption sites on that surface and then assuming that the lowest energy site is representative of the entire surface, because that site is the most thermodynamically stable and therefore more likely to bind the adsorbate. Thus we define an "adsorption energy of a surface" as the minimum adsorption energy of all sites on that surface. Using this method, we identified dozens of surfaces with near-optimal adsorption energies. These surfaces are listed in Table 1 and Table 2 for CO_2 reduction and H_2 evolution, respectively. Note that we chose to include surfaces whose energies were below the optimal energy targets by 0.2 eV and above the optimal targets by 0.1 eV, because we assumed that over-binding surfaces were more likely to perform well than under-binding surfaces.

SupplementaryTable 1 | List of possibly active surfaces for CO_2 reduction. All of the surfaces whose minimum CO adsorption energies were between -0.87 and -0.57 eV. MPID represents the Materials Project ID. Apparently duplicate entries for MPID/Miller index combinations are caused by different, nonidentical cuts along a Miller plane, which yields multiple surfaces for a single Miller index. All energies reported here were calculated by DFT, not with machine learning.

Formula	Miller index	ΔE_{CO} [eV]	MPID
Ag3Pd	[2, 0, 1]	-0.84	mp-985296
Al2Cu	[1, 1, 0]	-0.66	mp-985806
Al2Cu	[1, 1, 1]	-0.75	mp-985806
Al2Cu	[1, 1, 1]	-0.73	mp-985806
Al2Cu	[2, 1, 0]	-0.75	mp-985806
Al2Cu	[2, 1, 0]	-0.68	mp-985806
Al2Cu	[2, 2, 1]	-0.73	mp-985806
Al2Cu6	[0, 0, 1]	-0.79	mp-12802
Al2Cu6	[0, 0, 1]	-0.64	mp-12802
Al2Cu6	[1, 1, 0]	-0.80	mp-12802
Al2Cu6	[1, 1, 2]	-0.84	mp-12802
Al2Cu6	[1, 2, 0]	-0.65	mp-12802
Al2Cu6	[1, 2, 1]	-0.69	mp-12802
Al2Cu6	[1, 2, 2]	-0.78	mp-12802
Al2Cu6	[2, 1, 0]	-0.70	mp-12802
Al2Cu6	[2, 2, 1]	-0.76	mp-12802
Al3Cu2	[1, 1, 1]	-0.83	mp-10886
Al3Cu2	[1, 1, 1]	-0.82	mp-10886
Al3Cu2	[2, 0, 1]	-0.82	mp-10886
Al5Cu5	[1, 1, -2]	-0.83	mp-2500
Al6Cu2	[1, 0, 1]	-0.58	mp-985825
Al6Cu2	[2, 1, 0]	-0.83	mp-985825
AlCu	[1, 1, 1]	-0.80	mp-1022721
AlCu3	[1, 1, 0]	-0.76	mp-1008555
AlCu3	[1, 1, 0]	-0.72	mp-1008555
AlCu3	[1, 1, 1]	-0.84	mp-12777
AlCu3	[2, 1, 0]	-0.78	mp-1008555
AlCu3	[2, 1, 0]	-0.75	mp-1008555
Al2N2	[2, 1, 2]	-0.66	mp-661
Al2Pd	[2, 2, 1]	-0.85	mp-16522
Al6Pd10	[1, 0, 0]	-0.78	mp-16523
AlPd	[1, 0, 0]	-0.85	mp-829
Al2Pt	[1, 0, 0]	-0.72	mp-1502
Al2Ti2	[0, 0, 1]	-0.60	mp-1953
Al4Ti2	[0, 0, 1]	-0.82	mp-11809
Al12W	[1, 0, 0]	-0.77	mp-11227
AsGa	[1, 0, 0]	-0.81	mp-2534
AsGa	[2, 1, 1]	-0.76	mp-2534
As4Si3	[1, 1, 1]	-0.87	mp-570744
As6Si6	[2, 2, -1]	-0.67	mp-1863
As4Zn4Cu4	[1, 0, 1]	-0.66	mp-676828
As4Zn4Cu4	[1, 2, 1]	-0.75	mp-676828
As4Zn4Cu4	[2, 1, 0]	-0.75	mp-676828
As4Zn4Cu4	[2, 1, 2]	-0.79	mp-676828
Au2Al	[0, 0, 1]	-0.73	mp-1018179
Au2Al	[1, 0, 2]	-0.74	mp-1018179
Au4Al4	[0, 0, 1]	-0.62	mp-1399

Continued on next page...

Table 1 Continued from previous page

Formula	Miller index	ΔE_{CO} [eV]	MPID
Au2AlCu	[1, 1, 1]	-0.73	mp-867306
Au2AlCu	[2, 1, 0]	-0.79	mp-867306
Au2Cu2	[0, 0, 1]	-0.67	mp-522
Au2Cu2	[1, 0, 0]	-0.69	mp-522
Au2Cu2	[2, 1, 0]	-0.84	mp-522
Au3Cu	[2, 1, 0]	-0.76	mp-2103
Au3Cu	[2, 1, 0]	-0.61	mp-2103
Au3Cu	[2, 1, 1]	-0.64	mp-2103
Au3Cu	[2, 2, 1]	-0.71	mp-2103
AuCu3	[1, 1, 0]	-0.68	mp-2258
AuCu3	[1, 1, 0]	-0.63	mp-2258
AuCu3	[2, 1, 0]	-0.80	mp-2258
AuCu3	[2, 1, 0]	-0.79	mp-2258
AuCu3	[2, 1, 1]	-0.75	mp-2258
AuGa2	[1, 0, 0]	-0.57	mp-2776
AuPd2In	[2, 2, 1]	-0.79	mp-863724
Au5Sn	[1, 1, 0]	-0.64	mp-30418
Au4Ti	[1, 1, 0]	-0.58	mp-12635
Co8Al20	[0, 0, 1]	-0.69	mp-196
Cu	[1, 1, 0]	-0.63	mp-30
Cu	[2, 1, 0]	-0.78	mp-30
Cu	[2, 1, 1]	-0.69	mp-30
CuAlPt2	[1, 1, 0]	-0.77	mp-12550
Cu28In12	[0, 0, 1]	-0.84	mp-21985
Fe2Al12	[1, 1, 0]	-0.69	mp-570001
FeAl	[1, 0, 0]	-0.60	mp-2658
Ga2Cu	[0, 0, 1]	-0.78	mp-11359
Ga2Cu	[1, 0, 0]	-0.64	mp-11359
Ga2Cu	[1, 0, 1]	-0.67	mp-11359
Ga2Cu	[1, 1, 1]	-0.66	mp-11359
Ga2Cu	[1, 1, 2]	-0.74	mp-11359
Ga2Cu	[2, 0, 1]	-0.72	mp-11359
Ga2Cu	[2, 1, 0]	-0.82	mp-11359
Ga2Cu6	[1, 0, 0]	-0.79	mp-865798
Ga2Cu6	[1, 0, 0]	-0.69	mp-865798
Ga2Cu6	[1, 0, 1]	-0.75	mp-865798
Ga2Cu6	[1, 1, 0]	-0.80	mp-865798
Ga2Cu6	[1, 1, 1]	-0.72	mp-865798
Ga2Cu6	[2, 2, 1]	-0.84	mp-865798
Ga10Pd2	[1, 1, 2]	-0.66	mp-30660
Ga6Pd10	[1, 0, 0]	-0.79	mp-2408
Ga2Pt	[1, 0, 0]	-0.64	mp-22095
Ga2Pt	[1, 1, 0]	-0.65	mp-22095
Ga2Pt	[2, 1, 0]	-0.70	mp-22095
Ga3Pt2	[1, 1, 0]	-0.65	mp-21400
Ga3Pt2	[2, 1, 1]	-0.68	mp-21400
Ga3Pt2	[2, 2, 1]	-0.81	mp-21400
GaPt2Cu	[1, 1, 0]	-0.74	mp-644280
Ge2	[2, 1, 1]	-0.75	mp-32
Ge4Co4Cu2	[1, 0, 0]	-0.71	mp-19955

Continued on next page...

Table 1 Continued from previous page

Formula	Miller index	ΔE_{CO} [eV]	MPID
Ge2Cu6	[1, 0, 2]	-0.71	mp-19724
Ge2Cu6	[1, 2, 1]	-0.75	mp-19724
Ge2Cu6	[2, 1, 0]	-0.72	mp-19724
Ge4Fe4Cu2	[1, 0, 0]	-0.70	mp-21141
H6Al2	[2, 1, 1]	-0.74	mp-23933
H2Si2	[1, 0, 0]	-0.77	mp-29803
H2Si2	[2, 0, 1]	-0.85	mp-29803
H2Ti	[0, 0, 1]	-0.68	mp-24726
H2Ti	[1, 0, 0]	-0.64	mp-24161
H2Ti	[1, 1, 0]	-0.63	mp-24726
Ir4In12	[0, 2, 1]	-0.86	mp-636498
IrRhGa2	[1, 0, 0]	-0.81	mp-865743
Mn2Al2	[0, 0, 1]	-0.74	mp-771
Mn3GaN	[1, 0, 0]	-0.85	mp-627439
N4	[1, 1, 0]	-0.84	mp-999498
NIn	[1, 0, 0]	-0.76	mp-20411
N2V2	[1, 1, 1]	-0.74	mp-1017532
N2V2	[2, 1, 0]	-0.77	mp-1017532
Ni4Al12	[0, 0, 1]	-0.64	mp-622209
NiAl	[1, 0, 0]	-0.69	mp-1487
NiFeAl2	[1, 0, 0]	-0.73	mp-867330
NiGa	[1, 1, 1]	-0.75	mp-1941
Ni4Si4	[0, 1, 0]	-0.70	mp-351
NiZnCu2	[1, 0, 0]	-0.68	mp-30593
NiZnCu2	[1, 1, 0]	-0.87	mp-30593
OsAl	[1, 0, 0]	-0.73	mp-875
Pd4Cu16	[0, 0, 1]	-0.71	mp-30594
Pd4Cu16	[2, 1, 0]	-0.82	mp-30594
PdCu	[1, 0, 0]	-0.66	mp-1018029
PdCu3	[1, 0, 0]	-0.66	mp-580357
PdCu3	[1, 1, 0]	-0.76	mp-580357
PdCu3	[1, 1, 1]	-0.82	mp-580357
PdCu3	[2, 2, 1]	-0.84	mp-580357
Pd2CuIn	[1, 0, 0]	-0.79	mp-867308
PtCu	[0, 0, 1]	-0.72	mp-644311
PtCu3	[1, 0, 0]	-0.69	mp-12086
PtCu3	[1, 1, 0]	-0.79	mp-12086
PtCu3	[2, 2, 1]	-0.76	mp-12086
PtIn2	[2, 1, 0]	-0.74	mp-22682
Re2Al12	[0, 0, 1]	-0.66	mp-16528
Re2Al12	[1, 1, 0]	-0.70	mp-16528
ReAl12	[1, 0, 0]	-0.80	mp-1648
RuCoGa2	[1, 0, 0]	-0.69	mp-865779
RuIrGa2	[1, 0, 0]	-0.65	mp-866041
Sb2Al2	[1, 0, 2]	-0.82	mp-1018100
SbAl	[1, 1, 1]	-0.82	mp-2624
SbAl	[2, 1, 1]	-0.80	mp-2624
SbCuNi2	[1, 0, 0]	-0.66	mp-30069
Sb2Ga2	[2, 0, 1]	-0.76	mp-1018059
SbGa	[1, 1, 1]	-0.57	mp-1156

Continued on next page...

Table 1 Continued from previous page

Formula	Miller index	ΔE_{CO} [eV]	MPID
Sb2Pd2	[1, 1, 1]	-0.57	mp-1769
Sb2Pd2	[2, 2, 1]	-0.58	mp-1769
Sb8Pd4	[2, 1, 0]	-0.71	mp-1356
Sb2Pt2	[1, 0, 1]	-0.61	mp-2845
Sb2Pt2	[1, 1, 0]	-0.67	mp-2845
Sb2Pt2	[2, -1, 2]	-0.61	mp-2845
Sb2Pt2	[2, 1, 0]	-0.68	mp-2845
Si2Cu6	[1, 0, 0]	-0.60	mp-867317
Si2Cu6	[1, 0, 2]	-0.73	mp-867317
Si2Cu6	[1, 0, 2]	-0.73	mp-867317
Si2Cu6	[1, 1, 0]	-0.78	mp-867317
Si2Cu6	[1, 1, 1]	-0.79	mp-867317
Si2Cu6	[2, 0, 1]	-0.83	mp-867317
Si2Cu6	[2, 1, 1]	-0.71	mp-867317
SiCu3	[0, 0, 1]	-0.65	mp-972828
SiCu3	[1, 0, 0]	-0.69	mp-972828
SiCu3	[2, 0, 1]	-0.82	mp-972828
SiCu3	[2, 1, 0]	-0.77	mp-972828
SiCu3	[2, 1, 1]	-0.81	mp-972828
SiCu3	[2, 1, 1]	-0.79	mp-972828
Si16Fe8	[0, 1, 0]	-0.86	mp-1714
Si4Pd4	[0, 1, 1]	-0.83	mp-389
Si4Pd4	[1, 0, 0]	-0.63	mp-389
Si4Pt4	[1, 0, 1]	-0.80	mp-696
Si4Pt4	[1, 2, 0]	-0.67	mp-696
Si4Rh4	[0, 1, 0]	-0.83	mp-818
Sn10Cu12	[0, 1, 0]	-0.64	mp-1233
Sn10Cu12	[1, 0, 0]	-0.68	mp-1233
Sn10Cu12	[1, 1, -2]	-0.70	mp-1233
Sn16Cu20	[0, 0, 1]	-0.65	mp-845
Sn16Cu20	[1, 0, 0]	-0.63	mp-845
Sn16Cu20	[1, 0, 0]	-0.61	mp-845
Sn2Cu2	[2, 1, 1]	-0.61	mp-10598
Sn12Pd4	[1, 1, 2]	-0.76	mp-1371
Sn12Pd4	[1, 2, 1]	-0.77	mp-1371
Sn12Pd4	[2, 0, 1]	-0.74	mp-1371
Sn16Pd8	[1, 1, 0]	-0.66	mp-1573
Sn4Pd4	[1, 0, 0]	-0.58	mp-2369
Sn2Pt	[1, 1, 0]	-0.69	mp-19962
Sn2Pt2	[1, 0, 1]	-0.75	mp-19856
Sn2Pt2	[2, -1, 2]	-0.74	mp-19856
Sn2Pt2	[2, 1, 0]	-0.85	mp-19856
Zn4Au2Cu2	[0, 1, 2]	-0.62	mp-12759
Zn4Au2Cu2	[1, 1, 1]	-0.68	mp-12759
Zn4Au2Cu2	[1, 2, 1]	-0.63	mp-12759
Zn4Au2Cu2	[1, 2, 2]	-0.79	mp-12759
ZnAu2Cu	[1, 0, 0]	-0.78	mp-864623
ZnAu2Cu	[2, 1, 0]	-0.72	mp-864623
Zn16Cu10	[1, 1, 0]	-0.82	mp-1368
ZnCu	[1, 0, 0]	-0.61	mp-987

Continued on next page...

Table 1 Continued from previous page

Formula	Miller index	ΔE_{CO} [eV]	MPID
ZnCu	[2, 1, 0]	-0.79	mp-987
Zn2Pd2	[2, 0, 1]	-0.58	mp-1652

SupplementaryTable 2 | List of possibly efficient surfaces for H_2 evolution. All of the surfaces whose minimum H adsorption energies were between -0.47 and -0.17 eV. MPID represents the Materials Project ID. Apparently duplicate entries for MPID/Miller index combinations are caused by different, nonidentical cuts along a Miller plane, which yields multiple surfaces for a single Miller index. All energies reported here were calculated by DFT, not with machine learning.

Formula	Miller index	ΔE_H [eV]	MPID
Ag3Pd	[1, 0, 0]	-0.24	mp-985296
Al2Cu	[1, 1, 1]	-0.45	mp-985806
Al2Cu	[1, 1, 1]	-0.21	mp-985806
Al2Cu	[2, 1, 0]	-0.34	mp-985806
Al2Cu	[2, 1, 0]	-0.28	mp-985806
Al2Cu	[2, 1, 1]	-0.37	mp-985806
Al2Cu6	[1, 1, 0]	-0.18	mp-12802
Al2Cu6	[1, 1, 1]	-0.21	mp-12802
Al2Cu6	[1, 2, 2]	-0.21	mp-12802
Al3Cu2	[1, 1, 1]	-0.20	mp-10886
Al4Cu2	[1, 0, 1]	-0.22	mp-998
Al4Cu2	[1, 1, 2]	-0.23	mp-998
Al5Cu5	[1, 1, -1]	-0.17	mp-2500
Al5Cu5	[1, 1, -2]	-0.25	mp-2500
Al6Cu2	[1, 1, 1]	-0.22	mp-985825
Al6Cu2	[1, 1, 2]	-0.20	mp-985825
AlCu3	[1, 1, 1]	-0.30	mp-12777
Al2Pd	[2, 1, 0]	-0.37	mp-16522
Al4Pd4	[1, 0, 0]	-0.22	mp-7189
Al4Pd8	[0, 0, 1]	-0.24	mp-2824
Al2Pt	[2, 2, 1]	-0.37	mp-1502
Al3Pt2	[2, 1, 1]	-0.46	mp-10905
Al4Pt12	[0, 0, 1]	-0.29	mp-607111
Al4Pt12	[1, 0, 1]	-0.40	mp-607111
Al4Pt4	[1, 0, 0]	-0.42	mp-10904
Al4Pt4	[2, 1, 1]	-0.38	mp-10904
Al4Pt4	[2, 1, 1]	-0.34	mp-10904
Al4Pt8	[0, 0, 1]	-0.42	mp-16526
Al4Pt8	[1, 0, 0]	-0.37	mp-16526
Al6Pt10	[1, 0, 0]	-0.34	mp-1501
AlPt3	[1, 0, 0]	-0.38	mp-188
AlPt3	[1, 1, 1]	-0.45	mp-188
Al3V	[1, 1, 0]	-0.46	mp-2554
Al3V	[1, 1, 1]	-0.35	mp-2554
Al3V	[1, 1, 2]	-0.33	mp-2554
Al12W	[1, 1, 1]	-0.47	mp-11227
As2	[1, 1, 1]	-0.21	mp-11
As2	[2, 1, 0]	-0.47	mp-11
As2	[2, 2, 1]	-0.23	mp-11
As4Co4	[1, 0, 0]	-0.43	mp-15679
As4Co4	[1, 1, 0]	-0.45	mp-583

Continued on next page...

Table 2 Continued from previous page

Formula	Miller index	ΔE_H [eV]	MPID
As2Ga2	[2, 2, 1]	-0.19	mp-8883
AsGa	[2, 1, 0]	-0.21	mp-2534
AsIn	[2, 1, 0]	-0.44	mp-20305
As4Os2	[1, 1, 1]	-0.37	mp-2455
AsPd5In	[2, 1, 0]	-0.23	mp-1025293
AsPd5In	[2, 1, 2]	-0.46	mp-1025293
As4Rh4	[0, 0, 1]	-0.46	mp-22079
As4Rh4	[0, 0, 1]	-0.19	mp-22079
As4Rh4	[1, 0, 2]	-0.25	mp-22079
As4Rh4	[2, 2, 1]	-0.45	mp-22079
As4Rh4	[2, 2, 1]	-0.33	mp-22079
As8Rh4	[1, 0, -1]	-0.45	mp-15954
As2Rh10Ga4	[1, 0, 0]	-0.46	mp-18561
As6Si6	[1, 1, 0]	-0.20	mp-1863
AsSn	[1, 1, 1]	-0.23	mp-2182
As6Ti8	[1, 1, 0]	-0.33	mp-567082
As4Zn4Cu4	[1, 1, 1]	-0.36	mp-676828
As4Zn4Cu4	[1, 1, 1]	-0.24	mp-676828
As4Zn4Cu4	[1, 2, 1]	-0.31	mp-676828
As4Zn4Cu4	[2, 1, 0]	-0.30	mp-676828
AsZnPt5	[0, 0, 1]	-0.40	mp-1025356
AsZnPt5	[1, 1, 1]	-0.44	mp-1025356
Au4Al4	[0, 0, 1]	-0.29	mp-1399
Au4Al4	[0, 0, 1]	-0.21	mp-1399
Au4Al4	[2, 2, 1]	-0.30	mp-1399
AuAl2	[1, 1, 1]	-0.46	mp-2647
AuAl2	[2, 1, 0]	-0.36	mp-2647
AuAl2	[2, 2, 1]	-0.37	mp-2647
Au2AlCu	[2, 1, 0]	-0.28	mp-867306
Au3Pd	[2, 1, 1]	-0.19	mp-973834
AuPd3	[1, 0, 0]	-0.30	mp-999298
AuPd3	[1, 0, 0]	-0.18	mp-999298
AuPd3	[2, 1, 1]	-0.40	mp-999298
AuPd3	[2, 1, 1]	-0.36	mp-999298
AuPd3	[2, 2, 1]	-0.29	mp-999298
AuPdCu2	[2, 1, 0]	-0.19	mp-862256
Co4Al18	[0, 0, 1]	-0.41	mp-16488
CoFeAl2	[2, 1, 0]	-0.45	mp-862691
Co4Ga12	[1, 0, 1]	-0.21	mp-20559
Co4Ga12	[1, 1, 0]	-0.18	mp-20559
Co4In12	[1, 0, 0]	-0.33	mp-22236
Co4In12	[2, 1, 0]	-0.31	mp-22236
Co4N2	[1, 0, 1]	-0.41	mp-22631
Co2Pt2	[1, 0, 1]	-0.40	mp-949
CoPt3	[1, 0, 0]	-0.36	mp-922
CoPt3	[2, 1, 0]	-0.36	mp-922
CoPt3	[2, 2, 1]	-0.37	mp-922
Co2Sn4	[1, 1, 1]	-0.24	mp-20155
Co2Sn4	[1, 1, 1]	-0.18	mp-20155
Co2Sn4	[2, 1, 1]	-0.20	mp-20155

Continued on next page...

Table 2 Continued from previous page

Formula	Miller index	ΔE_H [eV]	MPID
Co3Sn3	[2, 1, 1]	-0.40	mp-20536
CrCoPt2	[1, 1, 1]	-0.35	mp-570863
CrCoPt2	[1, 1, 2]	-0.36	mp-570863
CrCoPt2	[2, 1, 0]	-0.45	mp-570863
CrCoPt2	[2, 1, 0]	-0.39	mp-570863
CuAlPt2	[0, 0, 1]	-0.40	mp-12550
CuAlPt2	[1, 0, 0]	-0.27	mp-12550
CuAlPt2	[1, 1, 1]	-0.33	mp-12550
CuAlPt2	[1, 1, 2]	-0.29	mp-12550
CuAlPt2	[2, 1, 0]	-0.43	mp-12550
CuAlPt2	[2, 1, 2]	-0.44	mp-12550
Cu28In12	[0, 1, 0]	-0.29	mp-21985
Fe2Al12	[0, 0, 1]	-0.30	mp-570001
Fe2Al6	[1, 0, 0]	-0.43	mp-984873
Fe2Al6	[1, 1, 0]	-0.24	mp-984873
Fe2Al6	[2, 1, 1]	-0.46	mp-984873
Fe4Ga12	[1, 0, 1]	-0.33	mp-636368
Fe4Ga12	[1, 1, 0]	-0.25	mp-636368
Fe2Pd2	[1, 0, 1]	-0.44	mp-2831
Fe2Pd2	[1, 1, 0]	-0.30	mp-2831
Fe2Pd2	[2, 1, 2]	-0.35	mp-2831
Fe2Pd2	[2, 1, 2]	-0.33	mp-2831
FePd3	[1, 0, 0]	-0.30	mp-21845
FePt	[1, 0, 0]	-0.40	mp-2260
FePt	[1, 1, 1]	-0.41	mp-2260
FePt	[2, 0, 1]	-0.41	mp-2260
FePt	[2, 1, 1]	-0.47	mp-2260
Fe2WAl	[1, 0, 0]	-0.44	mp-862288
Fe2WAl	[2, 1, 0]	-0.44	mp-862288
Ga2N2	[1, 0, 1]	-0.37	mp-804
Ga10Pd2	[1, 1, 0]	-0.27	mp-30660
Ga3Pd7	[0, 0, 1]	-0.27	mp-31467
Ga3Pd7	[1, 0, 0]	-0.38	mp-31467
Ga3Pd7	[1, 0, 0]	-0.35	mp-31467
Ga3Pd7	[1, 1, -1]	-0.31	mp-31467
Ga4Pd8	[0, 0, 1]	-0.27	mp-1869
Ga4Pd8	[0, 1, 0]	-0.24	mp-1869
Ga4Pd8	[1, 0, 1]	-0.24	mp-1869
Ga4Pd8	[1, 0, 1]	-0.23	mp-1869
Ga4Pd8	[1, 1, 0]	-0.34	mp-1869
Ga4Pd8	[1, 1, 0]	-0.31	mp-1869
Ga4Pd8	[2, 1, 0]	-0.32	mp-1869
Ga6Pd10	[0, 0, 1]	-0.34	mp-2408
Ga2Pd2Ti2	[1, 0, 0]	-0.26	mp-1025045
Ga2Pt6	[0, 0, 1]	-0.27	mp-862621
Ga2Pt6	[0, 0, 1]	-0.19	mp-862621
Ga2Pt6	[1, 0, 1]	-0.42	mp-862621
Ga2Pt6	[1, 0, 2]	-0.42	mp-862621
Ga3Pt5	[0, 0, 1]	-0.43	mp-30663
Ga3Pt5	[0, 1, 0]	-0.39	mp-30663

Continued on next page...

Table 2 Continued from previous page

Formula	Miller index	ΔE_H [eV]	MPID
Ga3Pt5	[0, 1, 1]	-0.43	mp-30663
Ga3Pt5	[0, 1, 1]	-0.43	mp-30663
Ga3Pt5	[1, 1, 2]	-0.43	mp-30663
Ga4Pt12	[0, 0, 1]	-0.32	mp-623066
Ga4Pt12	[1, 0, 2]	-0.43	mp-623066
Ga8Pt16	[1, 0, 0]	-0.45	mp-2223
GaPt3	[1, 0, 0]	-0.46	mp-11407
GaPt3	[1, 1, 0]	-0.33	mp-11407
GaPt3	[2, 1, 1]	-0.42	mp-11407
GaPt2Cu	[1, 0, 0]	-0.44	mp-644280
GaPt2Cu	[1, 0, 2]	-0.42	mp-644280
GaPt2Cu	[1, 1, 0]	-0.45	mp-644280
GaPt2Cu	[1, 1, 0]	-0.35	mp-862791
GaPt2Cu	[1, 1, 1]	-0.39	mp-862791
GaPt2Cu	[1, 1, 2]	-0.35	mp-644280
GaPt2Cu	[2, 0, 1]	-0.46	mp-644280
GaPt2Cu	[2, 1, 0]	-0.41	mp-644280
GaPt2Cu	[2, 1, 1]	-0.28	mp-644280
GaPt2Cu	[2, 1, 2]	-0.37	mp-644280
GaPt2Cu	[2, 2, 1]	-0.44	mp-644280
Ge4Co4Cu2	[1, 0, 0]	-0.22	mp-19955
Ge2Cu6	[1, 1, 2]	-0.21	mp-19724
Ge2Cu6	[1, 2, 1]	-0.18	mp-19724
Ge2Mn2Ga2	[2, 0, 1]	-0.47	mp-1018802
Ge2Mo	[1, 1, 2]	-0.35	mp-10201
GePtTi	[2, 1, 0]	-0.45	mp-1008680
Ge4Rh4	[0, 0, 1]	-0.44	mp-22239
Ge4Rh4	[0, 2, 1]	-0.38	mp-22239
Ge4Rh4	[1, 0, 0]	-0.28	mp-20866
Ir	[1, 1, 1]	-0.37	mp-101
Ir2Al6	[0, 0, 1]	-0.44	mp-2294
Ir2Al6	[0, 0, 1]	-0.41	mp-2294
Ir2Al6	[1, 0, 0]	-0.46	mp-2294
Ir2Al6	[1, 0, 0]	-0.26	mp-2294
Ir2Al6	[2, -1, 2]	-0.33	mp-2294
IrFeAl2	[1, 1, 0]	-0.45	mp-866031
Ir4Ga18	[1, 2, 0]	-0.31	mp-31311
Ir6Ga12	[0, 0, 1]	-0.43	mp-31253
Ir6Ga12	[1, 1, 0]	-0.37	mp-31253
IrGa	[1, 1, 1]	-0.46	mp-11388
Ir4In8	[0, 1, 0]	-0.38	mp-22812
IrRhAl2	[1, 1, 1]	-0.33	mp-862694
Ir6Sn14	[1, 0, 0]	-0.21	mp-22040
IrSn2	[1, 0, 0]	-0.31	mp-2083
IrSn2	[2, 1, 1]	-0.22	mp-2083
Ir6V2	[1, 1, 0]	-0.39	mp-865496
MnRh2In	[2, 1, 0]	-0.47	mp-864968
Mn2RhPt	[2, 1, 0]	-0.39	mp-865032
Mn2RhPt	[2, 1, 0]	-0.33	mp-865032
MnRhTi2	[1, 0, 0]	-0.39	mp-866218

Continued on next page...

Table 2 Continued from previous page

Formula	Miller index	ΔE_H [eV]	MPID
MoN	[2, 2, 1]	-0.38	mp-13036
MoZn7	[1, 1, 1]	-0.40	mp-644500
N2V2	[2, 1, 0]	-0.18	mp-1017532
Ni	[1, 1, 0]	-0.41	mp-23
Ni2Al3	[1, 1, 1]	-0.38	mp-1057
Ni2Al3	[2, -1, 2]	-0.45	mp-1057
Ni2Al3	[2, 1, 0]	-0.30	mp-1057
Ni2Al3	[2, 2, 1]	-0.32	mp-1057
Ni4Al12	[0, 0, 1]	-0.35	mp-622209
Ni4Al12	[0, 1, 0]	-0.40	mp-622209
Ni5Al3	[0, 0, 1]	-0.41	mp-16514
Ni5Al3	[0, 2, 1]	-0.31	mp-16514
Ni5Al3	[1, 1, 2]	-0.46	mp-16514
Ni5Al3	[1, 1, 2]	-0.45	mp-16514
Ni5Al3	[2, 0, 1]	-0.32	mp-16514
NiAl	[1, 0, 0]	-0.36	mp-1487
NiAl	[1, 1, 1]	-0.23	mp-1487
NiAl	[2, 1, 0]	-0.43	mp-1487
Ni2AuSn4	[1, 0, 1]	-0.23	mp-568925
NiCo2Ga	[1, 0, 0]	-0.32	mp-20551
NiCo2Ga	[2, 1, 1]	-0.41	mp-1018060
Ni2Fe2	[1, 0, 0]	-0.41	mp-2213
Ni2Fe2	[2, 1, 0]	-0.42	mp-2213
Ni2Fe2	[2, 1, 2]	-0.41	mp-2213
Ni3Fe	[1, 0, 0]	-0.40	mp-1007855
Ni3Fe	[1, 1, 0]	-0.42	mp-1418
NiFeAl2	[2, 1, 0]	-0.40	mp-867330
NiFePt2	[0, 0, 1]	-0.39	mp-13463
NiFePt2	[0, 0, 1]	-0.31	mp-13463
NiFePt2	[1, 0, 0]	-0.39	mp-13463
NiFePt2	[1, 0, 0]	-0.37	mp-13463
NiFePt2	[1, 1, 1]	-0.38	mp-13463
NiFePt2	[2, 1, 1]	-0.45	mp-13463
Ni13Ga9	[1, 0, 0]	-0.42	mp-21589
Ni2Ga3	[2, -1, 2]	-0.43	mp-11397
Ni2Ga3	[2, -1, 2]	-0.43	mp-11397
Ni4Ga2	[1, 1, 1]	-0.39	mp-570904
Ni5Ga3	[0, 0, 1]	-0.44	mp-11398
Ni4Ge4	[1, 0, 2]	-0.37	mp-1099
Ni2In3	[1, 1, 1]	-0.22	mp-21385
Ni6Mo2	[0, 1, 0]	-0.46	mp-11506
Ni2Pt2	[0, 0, 1]	-0.36	mp-945
Ni2Pt2	[1, 0, 1]	-0.39	mp-945
Ni2Pt2	[2, 1, 0]	-0.47	mp-945
Ni2Pt2	[2, 1, 0]	-0.39	mp-945
Ni2Pt2	[2, 2, 1]	-0.45	mp-945
Ni2Pt2	[2, 2, 1]	-0.42	mp-945
Ni3Pt	[1, 0, 0]	-0.43	mp-12798
Ni3Pt	[1, 0, 0]	-0.33	mp-12798
NiRh2Ga	[1, 1, 1]	-0.44	mp-866037

Continued on next page...

Table 2 Continued from previous page

Formula	Miller index	ΔE_H [eV]	MPID
NiRh2Ga	[2, 1, 0]	-0.43	mp-866037
NiRh2Sn	[1, 1, 1]	-0.40	mp-11519
Ni2Si2	[0, 0, 1]	-0.31	mp-999192
Ni2Si2	[1, 2, 1]	-0.38	mp-999192
Ni2Si2	[1, 2, 2]	-0.36	mp-999192
Ni2Si2	[2, 1, 2]	-0.36	mp-999192
Ni4Si4	[1, 0, 2]	-0.39	mp-351
Ni8Si4	[1, 1, 1]	-0.45	mp-1118
Ni3Sn4	[1, 1, -2]	-0.28	mp-20174
Ni6Sn2	[1, 0, 0]	-0.17	mp-20112
Ni3SnN	[2, 1, 1]	-0.40	mp-1017632
Ni4W	[0, 0, 1]	-0.40	mp-30811
OsAl	[1, 0, 0]	-0.29	mp-875
Os4Ga12	[0, 0, 1]	-0.29	mp-570844
OsV3	[1, 1, 1]	-0.41	mp-866121
PbMnRh2	[2, 1, 0]	-0.36	mp-4652
Pb4Pt2	[1, 0, 1]	-0.27	mp-21318
Pb10Rh8	[0, 1, 0]	-0.18	mp-569678
Pd	[1, 1, 0]	-0.41	mp-2
Pd	[1, 1, 1]	-0.45	mp-2
Pd4Cu16	[1, 0, 0]	-0.19	mp-30594
PdCu3	[1, 1, 1]	-0.22	mp-580357
PdCu3	[2, 2, 1]	-0.20	mp-580357
Pd2CuIn	[1, 0, 0]	-0.21	mp-867308
Pd2CuIn	[1, 1, 0]	-0.24	mp-867308
Pd2CuIn	[2, 1, 0]	-0.20	mp-867308
Pd3In	[1, 0, 2]	-0.36	mp-510436
Pd3In	[1, 1, 1]	-0.42	mp-510436
Pd3In	[2, 0, 1]	-0.39	mp-510436
Pd3In	[2, 0, 1]	-0.36	mp-510436
Pd3In	[2, 1, 0]	-0.39	mp-31337
Pd3In	[2, 1, 0]	-0.36	mp-31337
Pd3In	[2, 1, 2]	-0.46	mp-510436
Pd3In	[2, 1, 2]	-0.40	mp-510436
Pd8In4	[1, 0, 2]	-0.36	mp-22646
Pd8In4	[1, 1, 0]	-0.38	mp-22646
Pd8In4	[2, 1, 0]	-0.23	mp-22646
Pd2Ti	[1, 1, 0]	-0.42	mp-1018121
Pd2Ti	[1, 1, 2]	-0.37	mp-1018121
Pd2Ti	[2, 1, 2]	-0.47	mp-1018121
Pd2V	[1, 0, 1]	-0.26	mp-11549
Pd2V	[1, 1, 0]	-0.44	mp-11549
Pd2V	[2, 1, 2]	-0.41	mp-11549
Pd3V	[1, 0, 0]	-0.40	mp-873
Pd3V	[1, 0, 1]	-0.34	mp-873
Pd3V	[1, 1, 0]	-0.32	mp-873
Pd3V	[2, 1, 1]	-0.40	mp-568711
Pd3V	[2, 1, 1]	-0.33	mp-568711
Pd3V	[2, 2, 1]	-0.39	mp-568711
Pd6V2	[1, 0, 0]	-0.46	mp-979980

Continued on next page...

Table 2 Continued from previous page

Formula	Miller index	ΔE_H [eV]	MPID
Pd6V2	[1, 0, 2]	-0.24	mp-979980
Pd6V2	[1, 1, 1]	-0.37	mp-979980
Pt	[1, 1, 1]	-0.36	mp-126
Pt7Cu	[1, 0, 0]	-0.44	mp-12608
Pt7Cu	[1, 0, 0]	-0.36	mp-12608
PtCu3	[1, 1, 0]	-0.34	mp-12086
PtCu3	[1, 1, 0]	-0.28	mp-12086
PtCu3	[1, 1, 1]	-0.32	mp-12086
PtCu3	[2, 1, 1]	-0.27	mp-12086
PtCu3	[2, 2, 1]	-0.25	mp-12086
Pt2CuIn	[1, 0, 0]	-0.35	mp-639659
Pt2CuIn	[2, 0, 1]	-0.44	mp-639659
Pt2CuIn	[2, 1, 1]	-0.46	mp-639659
Pt2FeCu	[0, 0, 1]	-0.31	mp-3702
Pt2FeCu	[1, 0, 1]	-0.42	mp-3702
Pt13In9	[1, 0, 0]	-0.44	mp-571060
Pt13In9	[2, 0, -1]	-0.37	mp-571060
Pt3In	[1, 0, 0]	-0.46	mp-20516
Pt3In	[1, 0, 0]	-0.25	mp-20516
Pt5In5	[0, 0, 1]	-0.36	mp-510438
Pt5In5	[1, 1, -1]	-0.30	mp-510438
Pt8Ti	[2, 1, 1]	-0.38	mp-30852
Pt2V	[0, 0, 1]	-0.35	mp-12108
Pt2V	[0, 1, 2]	-0.27	mp-12108
Pt2V	[1, 1, 0]	-0.45	mp-12108
Pt2V2	[1, 0, 0]	-0.45	mp-1017531
Pt3V	[0, 0, 1]	-0.44	mp-2705
Pt3V	[0, 0, 1]	-0.21	mp-2705
Pt3V	[1, 1, 2]	-0.26	mp-2705
Pt3V	[2, 1, 0]	-0.31	mp-372
Re2Al12	[0, 1, 0]	-0.17	mp-16528
Re2Al12	[1, 1, 0]	-0.46	mp-16528
ReAl12	[1, 0, 0]	-0.22	mp-1648
Re6Pt2	[0, 0, 1]	-0.46	mp-862589
Re4Si4	[1, 0, 0]	-0.44	mp-7948
Rh	[1, 0, 0]	-0.45	mp-74
Rh	[1, 1, 1]	-0.39	mp-74
Rh4Al18	[0, 0, 1]	-0.40	mp-1645
Rh4Al18	[0, 0, 1]	-0.20	mp-1645
Rh4Al18	[0, 1, 1]	-0.37	mp-1645
Rh4Al18	[1, 0, -1]	-0.19	mp-1645
Rh4Al18	[1, 0, 0]	-0.35	mp-1645
Rh4Al18	[1, 0, 0]	-0.30	mp-1645
Rh4Al18	[1, 0, 1]	-0.36	mp-1645
Rh4Al18	[1, 1, 0]	-0.31	mp-1645
Rh4Al18	[1, 1, 0]	-0.30	mp-1645
Rh8Al20	[0, 0, 1]	-0.45	mp-1791
Rh8Al20	[1, 1, 0]	-0.42	mp-1791
RhAl	[1, 0, 0]	-0.37	mp-364
RhAl	[1, 1, 0]	-0.26	mp-364

Continued on next page...

Table 2 Continued from previous page

Formula	Miller index	ΔE_H [eV]	MPID
RhAl	[2, 1, 0]	-0.18	mp-364
RhAl	[2, 1, 1]	-0.17	mp-364
RhAl	[2, 2, 1]	-0.34	mp-364
Rh2AlTi	[2, 1, 0]	-0.45	mp-866153
Rh2AlTi	[2, 1, 0]	-0.41	mp-866153
Rh2CoSn	[1, 0, 0]	-0.42	mp-1018085
Rh2CuGa	[1, 1, 1]	-0.47	mp-862485
Rh2CuV	[1, 0, 0]	-0.43	mp-979910
Rh2CuV	[1, 1, 1]	-0.37	mp-979910
Rh2CuV	[2, 1, 0]	-0.42	mp-979910
Rh2FeAl	[2, 1, 0]	-0.42	mp-861953
Rh6Ga10	[0, 0, 1]	-0.23	mp-30923
Rh6Ga10	[1, 0, -2]	-0.29	mp-30923
Rh6Ga10	[1, 2, -1]	-0.20	mp-30923
RhGa	[1, 0, 0]	-0.37	mp-2444
RhGa	[1, 0, 0]	-0.36	mp-2444
Rh4In12	[1, 0, 0]	-0.28	mp-18614
Rh4In12	[1, 0, 1]	-0.28	mp-18614
Rh4In12	[2, 1, 1]	-0.31	mp-18614
Rh3Pt	[1, 0, 0]	-0.45	mp-974370
Rh3Pt	[1, 0, 0]	-0.39	mp-974370
Rh3Pt	[1, 0, 0]	-0.34	mp-974376
Rh3Pt	[1, 0, 1]	-0.42	mp-974376
Rh3Pt	[1, 0, 1]	-0.39	mp-974376
Rh3Pt	[1, 1, 1]	-0.43	mp-974370
Rh3Pt	[2, 1, 0]	-0.41	mp-974376
RhPt3	[0, 0, 1]	-0.41	mp-974616
RhPt3	[1, 1, 2]	-0.37	mp-974616
Rh3Sn	[1, 1, 2]	-0.42	mp-978974
Rh4Sn4	[1, 0, 0]	-0.23	mp-317
Rh4Sn4	[1, 1, 0]	-0.37	mp-317
Rh4Sn4	[1, 1, 1]	-0.33	mp-317
Rh4Sn4	[2, 1, 0]	-0.25	mp-317
Rh2SnTi	[1, 0, 0]	-0.34	mp-865707
Rh2SnTi	[2, 1, 0]	-0.45	mp-865707
Rh2V2	[0, 0, 1]	-0.34	mp-971751
Rh2V2	[1, 0, 0]	-0.32	mp-1251
Rh2V2	[2, 2, 1]	-0.46	mp-1251
Rh2V2	[2, 2, 1]	-0.41	mp-1251
Rh2V6	[1, 1, 0]	-0.44	mp-1578
Rh3V	[1, 0, 0]	-0.43	mp-1185
Ru2	[1, 0, 0]	-0.44	mp-33
Ru2As4	[1, 2, 0]	-0.19	mp-766
Ru2As4	[2, 1, 0]	-0.43	mp-766
RuCoAl2	[1, 1, 1]	-0.41	mp-862695
Ru4In12	[1, 0, 0]	-0.31	mp-672326
Ru2SiGa	[2, 1, 0]	-0.35	mp-865615
Ru2Zn6	[1, 0, 1]	-0.46	mp-1380
Sb2Al2	[1, 1, 1]	-0.45	mp-1018100
Sb2Al2	[1, 1, 1]	-0.28	mp-1018100

Continued on next page...

Table 2 Continued from previous page

Formula	Miller index	ΔE_H [eV]	MPID
Sb2Al2	[2, 1, 1]	-0.35	mp-1018100
Sb2Al2	[2, 1, 2]	-0.32	mp-1018100
SbAl	[1, 1, 1]	-0.38	mp-2624
SbAl	[2, 2, 1]	-0.34	mp-2624
Sb4Co2	[2, 1, 2]	-0.30	mp-9835
Sb8Co4	[2, 1, 0]	-0.20	mp-755
Sb8Co4	[2, 1, 1]	-0.22	mp-755
Sb2CuNi	[1, 0, 0]	-0.36	mp-11834
Sb2CuNi	[1, 0, 1]	-0.29	mp-11834
Sb2Ga2	[2, 1, 1]	-0.24	mp-1018059
SbGa	[2, 1, 0]	-0.31	mp-1156
Sb8Ir4	[1, 1, 0]	-0.37	mp-1247
SbMnRh	[2, 1, 0]	-0.37	mp-4846
SbMnRh2	[1, 1, 2]	-0.42	mp-571163
SbMnRu2	[2, 1, 0]	-0.41	mp-864957
Sb2Ni2	[1, 0, 0]	-0.43	mp-810
Sb2Ni2	[1, 1, 1]	-0.25	mp-810
Sb2Ni2	[1, 1, 1]	-0.21	mp-810
Sb2Ni2	[2, 1, 1]	-0.23	mp-810
Sb2Ni2	[2, 1, 1]	-0.21	mp-810
Sb4Os2	[0, 1, 2]	-0.43	mp-2695
Sb4Os2	[0, 1, 2]	-0.40	mp-2695
Sb2Pt2	[1, 0, 0]	-0.33	mp-2845
SbPt7	[1, 0, 0]	-0.33	mp-1030
SbPt7	[1, 1, 1]	-0.29	mp-1030
Sb4Rh4	[1, 0, 0]	-0.20	mp-20619
Sb4Rh8	[1, 0, 0]	-0.35	mp-21359
Sb4Rh8	[2, 0, 1]	-0.42	mp-21359
Sb8Rh4	[1, 1, -2]	-0.41	mp-2682
Sb8Rh4	[1, 2, -1]	-0.20	mp-2682
Sb8Rh4	[2, 1, -1]	-0.21	mp-2682
SbRh2Cu	[2, 1, 0]	-0.44	mp-867753
Sb8Si8Pt8	[1, 0, 0]	-0.45	mp-11152
Sb8Si8Pt8	[1, 1, 0]	-0.43	mp-11152
SbSiPt5	[0, 0, 1]	-0.30	mp-1025366
SbSiPt5	[1, 0, 0]	-0.31	mp-1025366
SbSiPt5	[2, 1, 2]	-0.47	mp-1025366
Sb4V2	[2, 1, 1]	-0.41	mp-2851
SiAgPt5	[1, 0, 0]	-0.44	mp-1025220
SiAgPt5	[1, 0, 1]	-0.44	mp-1025220
SiAgPt5	[2, 1, 2]	-0.47	mp-1025220
Si2Co2Al	[1, 1, 0]	-0.41	mp-10010
Si2Co2Al	[1, 1, 1]	-0.42	mp-10010
Si2Co2Al	[2, 1, 0]	-0.42	mp-10010
Si2Co2Al	[2, 1, 0]	-0.42	mp-10010
Si2Fe4	[1, 1, 1]	-0.43	mp-22787
Si8Fe4Al12	[1, 1, 0]	-0.38	mp-505229
Si8Fe6Al4	[0, 0, 1]	-0.40	mp-29111
SiFe2Al	[1, 1, 1]	-0.37	mp-867878
Si2Ir6	[0, 0, 1]	-0.30	mp-1841

Continued on next page...

Table 2 Continued from previous page

Formula	Miller index	ΔE_H [eV]	MPID
Si2Ir6	[2, 1, 1]	-0.47	mp-1841
Si4Ir4	[0, 1, 1]	-0.44	mp-1128
Si2Mn6Al18	[1, 1, 0]	-0.37	mp-15819
Si16Os8	[0, 0, 1]	-0.47	mp-17123
SiOs2V	[2, 1, 0]	-0.46	mp-865506
SiOs2V	[2, 1, 0]	-0.45	mp-865506
Si3Pd6	[1, 0, 1]	-0.19	mp-697068
Si4Pd12	[0, 0, 1]	-0.34	mp-20622
Si4Pd12	[0, 1, 1]	-0.25	mp-20622
Si4Pd12	[1, 0, 0]	-0.31	mp-20622
Si4Pd12	[1, 0, 2]	-0.35	mp-20622
Si4Pd12	[1, 0, 2]	-0.32	mp-20622
Si4Pd12	[1, 2, 0]	-0.33	mp-20622
Si4Pd12	[2, 1, 0]	-0.31	mp-20622
Si2Pt6	[0, 1, 0]	-0.42	mp-13363
Si2Pt6	[2, 0, -1]	-0.33	mp-13363
Si4Pt12	[0, 0, 1]	-0.44	mp-21163
Si4Pt12	[1, 2, 0]	-0.39	mp-21163
Si4Pt12	[1, 2, 1]	-0.40	mp-21163
Si4Pt4	[2, 0, 1]	-0.46	mp-696
SiPt2	[1, 1, 2]	-0.44	mp-1299
SiPt5In	[1, 0, 0]	-0.41	mp-1025370
SiPt5In	[2, 0, 1]	-0.34	mp-1025370
SiPt5In	[2, 1, 0]	-0.40	mp-1025370
Si6Rh10	[1, 0, 0]	-0.38	mp-21012
SiRh2Cu	[1, 1, 0]	-0.47	mp-978532
SiRh2Zn	[1, 0, 0]	-0.46	mp-977384
SiRh2Zn	[2, 1, 0]	-0.42	mp-977384
SiSnPt5	[1, 0, 0]	-0.42	mp-1025324
SiSnPt5	[1, 1, 1]	-0.45	mp-1025324
Si6V3	[1, 0, 0]	-0.34	mp-10711
Si6V3	[1, 1, 0]	-0.21	mp-11190
Si6V3	[2, 1, 1]	-0.36	mp-11190
Si2W	[1, 1, 0]	-0.38	mp-1620
Si6W3	[1, 1, 0]	-0.34	mp-8939
Sn2Cu2	[1, 0, 0]	-0.39	mp-10598
Sn4Pd4	[0, 1, 0]	-0.26	mp-2369
Sn4Pd4	[1, 0, 1]	-0.17	mp-2369
Sn4Pd4	[2, 1, 2]	-0.26	mp-2369
Sn4Pd8	[1, 0, 1]	-0.28	mp-1851
Sn4Pd8	[1, 1, 0]	-0.23	mp-1851
Sn4Pd8	[1, 1, 2]	-0.24	mp-1851
Sn4Pd8	[1, 2, 2]	-0.29	mp-1851
SnPd3	[1, 0, 0]	-0.39	mp-718
SnPdTi	[2, 1, 0]	-0.46	mp-961682
Sn2Pt2	[1, 1, 1]	-0.36	mp-19856
Sn2Pt2	[2, 1, 0]	-0.17	mp-19856
SnPt3	[1, 0, 0]	-0.46	mp-20971
Sn2Ti4	[1, 0, 0]	-0.20	mp-30875
TiPd2In	[1, 0, 0]	-0.32	mp-866168

Continued on next page...

Table 2 Continued from previous page

Formula	Miller index	ΔE_H [eV]	MPID
TiRh2In	[1, 0, 0]	-0.39	mp-866170
TiRh2In	[2, 1, 0]	-0.47	mp-866170
Zn2As4Sn2	[1, 0, 0]	-0.38	mp-5190
Zn3Co	[1, 1, 2]	-0.20	mp-971948
Zn13Fe	[2, 0, 1]	-0.41	mp-1722
ZnPt3	[1, 0, 0]	-0.45	mp-30856
ZnPt3	[2, 1, 0]	-0.39	mp-30856
ZnPt3	[2, 1, 0]	-0.38	mp-30856
Zn2RhPd	[2, 1, 0]	-0.43	mp-864839
ZnRh2V	[1, 0, 0]	-0.35	mp-865487
ZnRh2V	[2, 1, 0]	-0.45	mp-865487
ZnRh2V	[2, 1, 0]	-0.43	mp-865487
Zn3Ti	[2, 2, 1]	-0.34	mp-21289

Supplementary References

[1] Olson, R.S. *et al.* Automating Biomedical Data Science Through Tree-Based Pipeline Optimization. In *European Conference on the Applications of Evolutionary Computation*, 123–137 (Springer International Publishing, Porto, Portugal, 2016).

[2] Hyndman, R.J. & Athanasopoulos, G. *Forecasting: Principles and practice* (otexts.com, 2014), 1 edn. URL `http://otexts.com/fpp`.

[3] Thompson, S.K. Stratified Sampling. In *Sampling*, 139–156 (John Wiley and Sons Inc., 2012). URL `http://dx.doi.org/10.1002/9781118162934.ch11`.

[4] Greeley, J., Jaramillo, T.F., Bonde, J., Chorkendorff, I. & Nørskov, J.K. Computational high-throughput screening of electrocatalytic materials for hydrogen evolution. *Nature Materials* **5**, 909–913 (2006).

[5] Liu, X. *et al.* Understanding trends in electrochemical carbon dioxide reduction rates. *Nature Communications* **8** (2017).

[6] Abild-Pedersen, F. & Andersson, M.P. CO adsorption energies on metals with correction for high coordination adsorption sites - A density functional study. *Surface Science* **601**, 1747–1753 (2007).

[7] Shi, C., Hansen, H.A., Lausche, A.C. & Nørskov, J.K. Trends in electrochemical CO2 reduction activity for open and close-packed metal surfaces. *Physical Chemistry Chemical Physics* **16**, 4720 (2014).

[8] Nørskov, J.K., Studt, F., Abild-Pedersen, F. & Bligaard, T. *Fundamental Concepts in Heterogeneous Catalysis* (John Wiley & Sons, Inc., 2015).

[9] Studt, F., Abild-Pedersen, F., Varley, J.B. & Nørskov, J.K. CO and CO2 hydrogenation to methanol calculated using the BEEF-vdW functional. *Catalysis Letters* **143**, 71–73 (2013).

[10] Wellendorff, J. *et al.* Density functionals for surface science: Exchange-correlation model development with Bayesian error estimation. *Physical Review B - Condensed Matter and Materials Physics* **85**, 32–34 (2012).

[11] Hammer, B., Hansen, L.B. & Nørskov, J. Improved adsorption energetics within density-functional theory using revised Perdew-Burke-Ernzerhof functionals. *Physical Review B* **59**, 7413–7421 (1999).

[12] Nørskov, J.K. *et al.* Trends in the Exchange Current for Hydrogen Evolution. *Journal of The Electrochemical Society* **152**, J23 (2005).

[13] Ong, S.P. *et al.* Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Computational Materials Science* **68**, 314–319 (2013).

[14] Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **12**, 2825–2830 (2012). URL `http://dl.acm.org/citation.cfm?id=2078195%5Cnhttp://arxiv.org/abs/1201.0490. 1201.0490`.

[15] Maaten, L.V.D. Accelerating t-SNE using Tree-Based Algorithms. *Journal of Machine Learning Research* **15**, 1–21 (2014). `1307.1662`.