Machine Learning Aided Discovery of High-Entropy Alloy Catalysts for CO$_2$ Reduction

Jack Kirk Pedersen, Supervisor: Jan Rossmeisl

In short
The many surface site environments that emerge from a composite surface made from a high-entropy alloy (HEA) (an alloy made up of 5 or more metals) facilitate a near-continuum of adsorption energies. Using machine learning on density functional theory data, adsorption energies can be predicted for any surface site. By tuning the composition of the HEA to favor surface sites with the desired adsorption energies, three rational candidates for a CO$_2$ reduction catalyst are suggested.

Introduction

Descriptor based search
Using properties of materials that correlate with the catalytic activity and are readily measured or calculated eases the task of discovering new materials that are likely to be suited for the catalytic reaction in question as it facilitates a screening of many materials without the need for experimental testing.

The adsorption energy of reactants, intermediates and products of a chemical reaction to surfaces are common descriptors that correlate with catalytic activity, as established by the Sabatier principle, which states that the adsorption strength between adsorbates and surface should neither be too weak, nor too strong for optimal catalytic activity.

Catalyst discovery using high-entropy alloys

High-entropy alloys (HEAs)
HEAs are materials made up of 5 or more metals. The driving force for stabilizing such an alloy is the high configurational entropy associated with mixing the elements.

We hypothesize that the many different surface site environments made possible by the many constituent elements of an HEA will provide a near-continuous distribution of adsorption energies. The composition of the HEA can then be optimized to increase the probability of obtaining surface sites with virtually any adsorption energy.

Layout of methodology
The procedure that we have devised for a rational design of new catalysts is based on experimental knowledge of catalytic activity or selectivity for a range of materials. It is an easy and quick way of suggesting new catalyst candidates for experimental testing and proceeds as follows:

1) Select a chemical reaction of interest.
2) Examine experimental data on catalytic activity or selectivity for different materials on the same type of surface.
3) Find one or more adsorption energies that correlate with the activity or selectivity.
4) Select 5 or more elements that make up the HEA.
5) Calculate the lattice constant of an equimolar, solid solution of the HEA.
6) Calculate adsorption energies for various and numerous surface environments of the HEA.
7) Use a machine learning algorithm to predict the adsorption energies of all possible surface environments of the HEA.
8) Optimize the composition of the HEA using the catalogue of all possible surface sites and their adsorption energies.
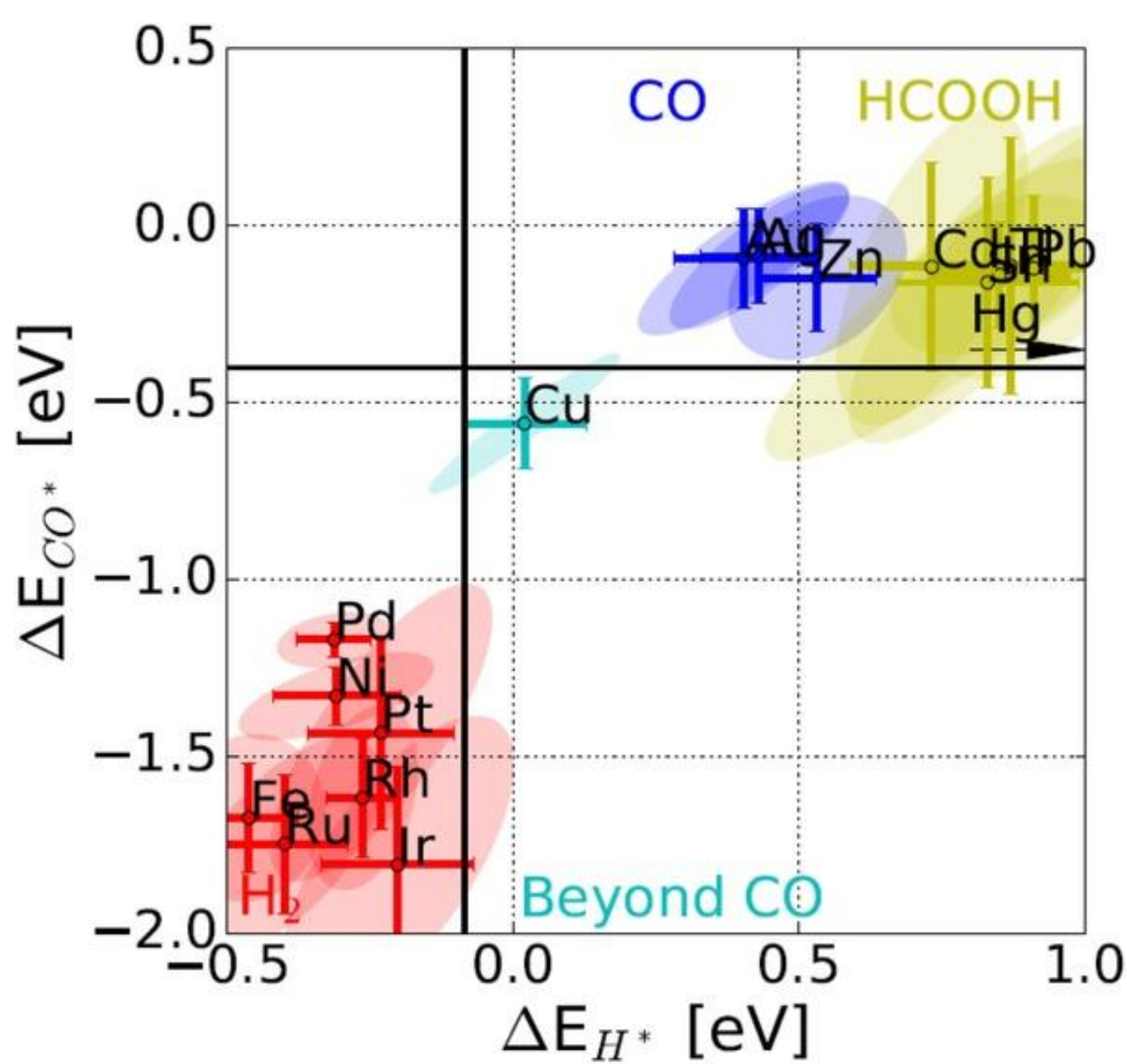

Figure 1

Calculated adsorption energies for H* and *CO on various metal (111) surfaces. These two descriptors clearly classify the Cu surface as a catalyst for highly reduced carbon products for the CO$_2$ reduction reaction. *From* [1].

Calculating adsorption energies

Generating data on surface sites and their adsorption energies requires many calculations and is only time efficient with density functional theory (DFT). We use the GPAW code and the ASE environment to calculate adsorption energies on small 2x2x4 slabs for intermediates CO and H, *see figure 2*, which have been shown to govern the selectivity in the CO$_2$ reduction reaction[1], see *figure 1*. The resulting adsorption energies are illustrated in the top row histograms of *figure 4*.
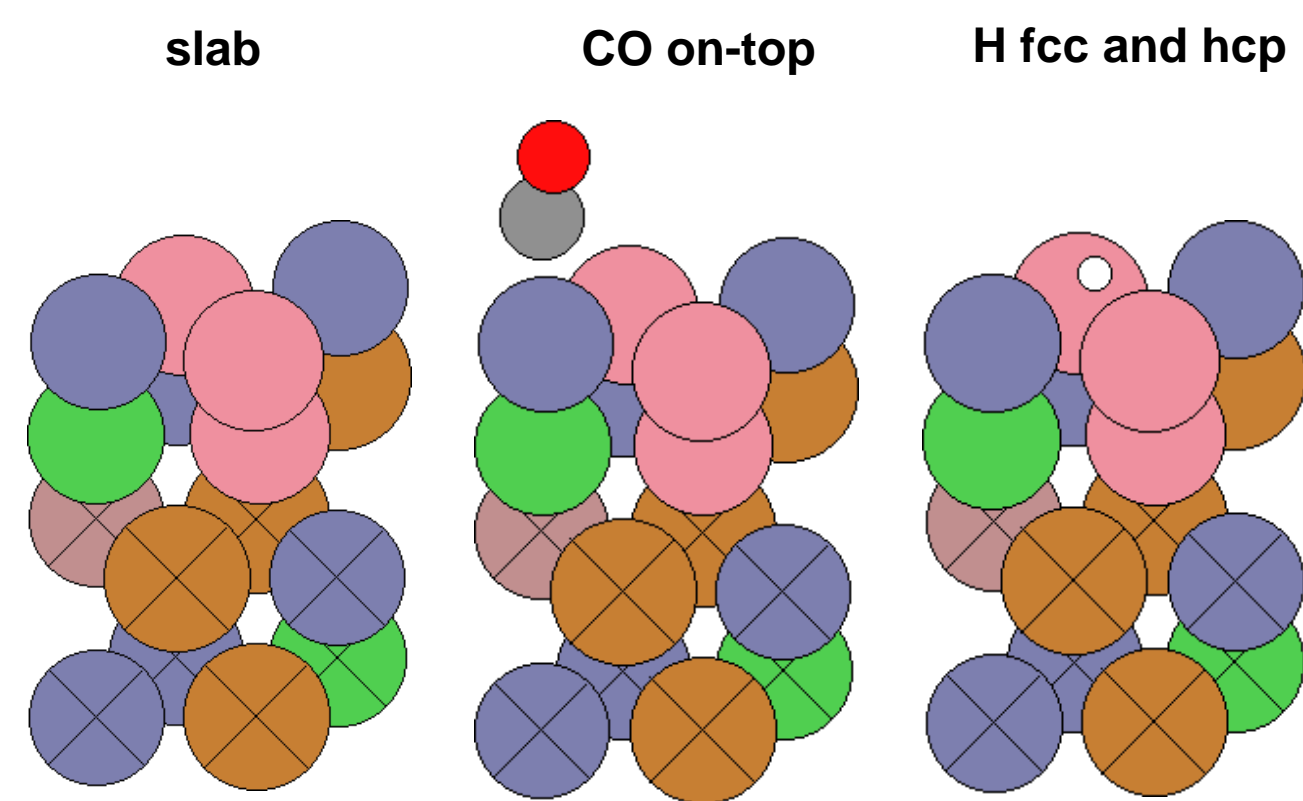

Figure 2

Example of a 2x2x4 slab of the CoCuGaNiZn HEA used for calculating adsorption energies. The two bottom layers are fixed during relaxation of the structure.

Predicting adsorption energies

Given DFT data on adsorption energies for various surfaces, a computer is coded to use the information from this data to predict adsorption energies of every surface site environment. Knowing the adsorption energy of every site allows us to favor sites which have the desired adsorption energies when optimizing the composition of the HEA.
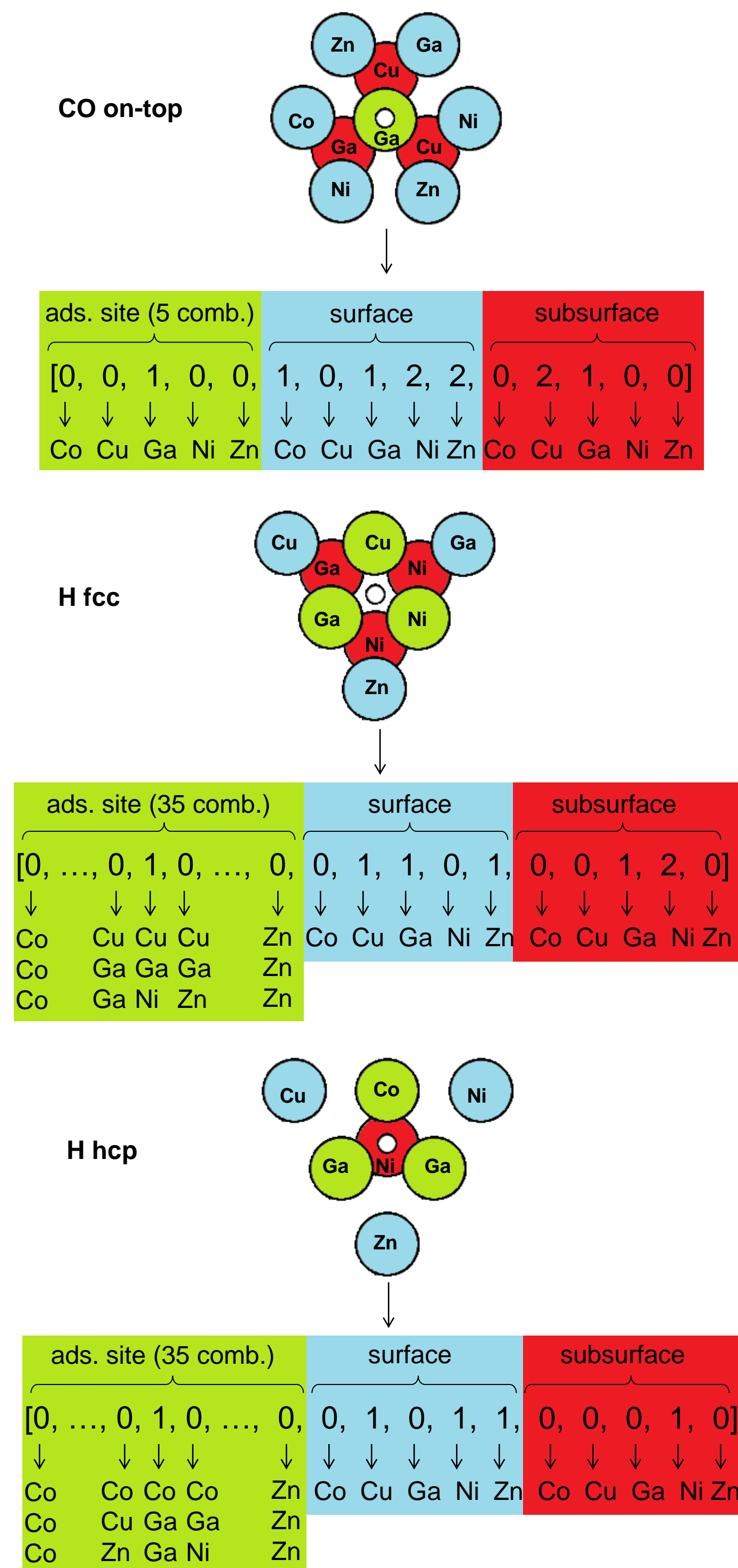

Figure 3

Description of the surface site environments for CO on-top and H hollow (fcc and hcp) adsorption as inputted to the computer. Each adsorption site is assigned a location in the fingerprint (5 or 35 combinations), and the surface and subsurface neighbor elements are counted. These representations produce input vectors of lengths 15, 45, and 45 for on-top, fcc, and hcp adsorptions respectively.
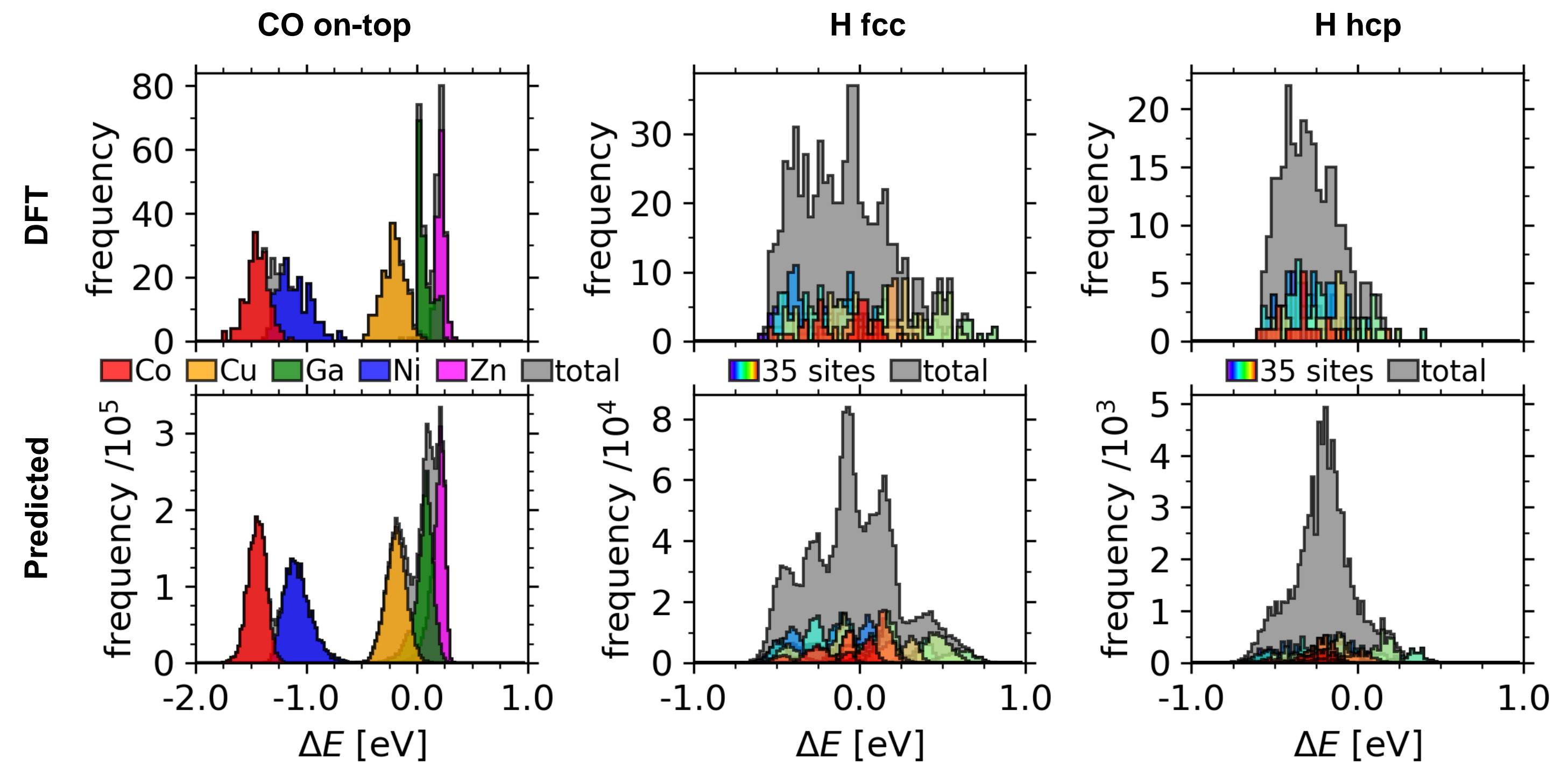
Surface site environment
Teaching a computer to predict the adsorption energy of a surface site requires a description of the surface environment. We choose a representation where the elements of the nearest neighbors of the adsorption site are inputted, see *figure 3*.

Machine learning algorithm
Having established a way for the computer to read the adsorption site environment, a machine learning algorithm must be selected. After testing a few methods we choose Gaussian process regression, the results of which are found in the second row of *figure 4*, with prediction accuracies shown in *figure 5*. This machine learning algorithm allows for a fast prediction of the entire span of possible surface sites environments, completing a multiyear DFT task in a matter of seconds, albeit with a loss in accuracy. However, this may be justified since DFT results are also inaccurate.

Optimizing the composition of the HEA

For the CO$_2$ reduction reaction, we need an optimization that limits the likelihood of adsorption energies that are known to produce H$_2$, CO, and HCOOH in order to hopefully produce highly reduced carbon products (*figure 1*). In order to achieve this, we have devised a penalty function B$_X$, X being CO, H fcc, or H hcp, such that CO adsorption energies that are too weak and H adsorption energies that are too strong will be disfavored (*figure 1*):

$$B_X(\mathbf{f}) = \sum_{i=1}^{N_X} w_i(\mathbf{f}) \times g\left(E_X^i\right)$$

where **f** is the atomic fractions of the HEA; N$_X$ the number of unique fingerprints of species X, and

$$w_i(\mathbf{f}) = \prod_k^{metals} f_k^{n_{ik}} \times m_i$$

$$g(x) = \begin{cases} \exp(x) & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}$$

$$E_{CO}^i = \Delta E_{CO}^i - \Delta E_{CO}^{upper}$$

$$E_H^i = \Delta E_H^{lower} - \Delta E_H^i$$

where n$_{ik}$ is the number of element k in fingerprint i; m$_i$ is the multiplicity of fingerprint i; ΔE_{CO}^i and ΔE_H^i are the adsorption energies of CO and H for fingerprint i. ΔE_{CO}^{upper} and ΔE_H^{lower}, as can be determined from *figure 1*, are the upper and lower boundaries for which adsorption sites are penalized.

The composition is optimized by minimizing an equally weighted sum of penalty functions with respect to the atomic fractions **f** of the HEA:

$$\mathbf{f}_{min} = \arg\min_{\mathbf{f}}(B_{CO}(\mathbf{f}) + B_{Hfcc}(\mathbf{f}) + B_{Hhcp}(\mathbf{f}))$$

Running the optimization for the CoCuGaNiZn HEA yields pure Zn with a summed penalty of 1.18 (evidently not a suitable catalyst according to *figure 1*) and three binary alloys shown in *figure 6*. One source of error when the optimized composition are no longer HEAs is the lattice constant which is fixed during the optimization.

References
[1] A. Bagger, W. Ju, A. S. Varela, P. Strasser, J. Rossmeisl, ChemPhysChem 2017, 18, 3266, DOI: 10.1002/cphc.201700736

Figure 4

DFT and predicted distributions of adsorption energies on the CoCuGaNiZn HEA for CO on-top sites, and H fcc and hcp sites. Predictions have been performed with Gaussian process regression on 891 CO, 668 H fcc, and 282 H hcp adsorption energies obtained from DFT calculations. The predictions are done on all possible surface sites using the description in *figure 3*, i.e. the predicted histograms in the second row contain 5^{10}, 5^9, and 5^7 surface sites for CO on-top, H fcc, and H hcp respectively.
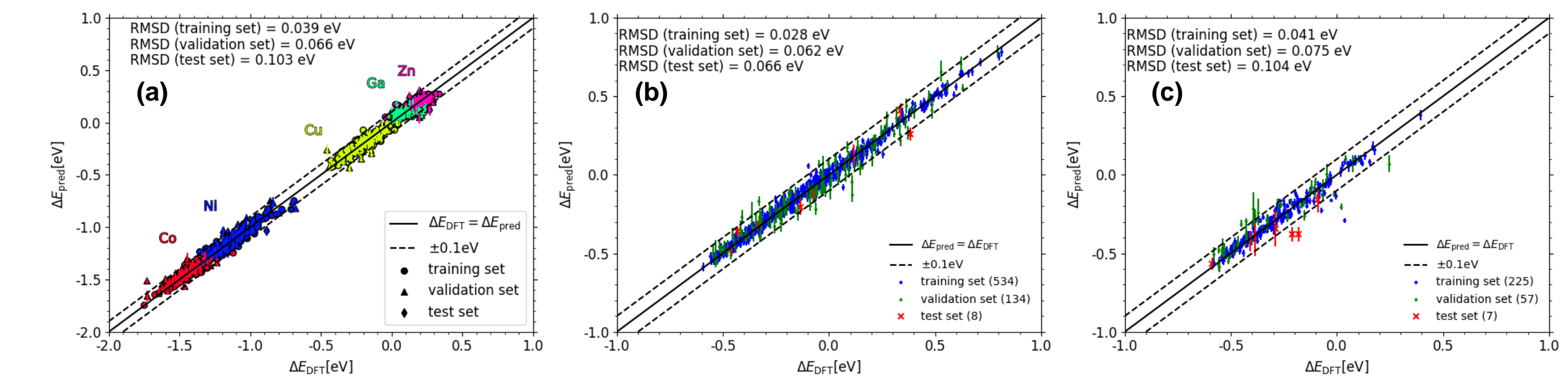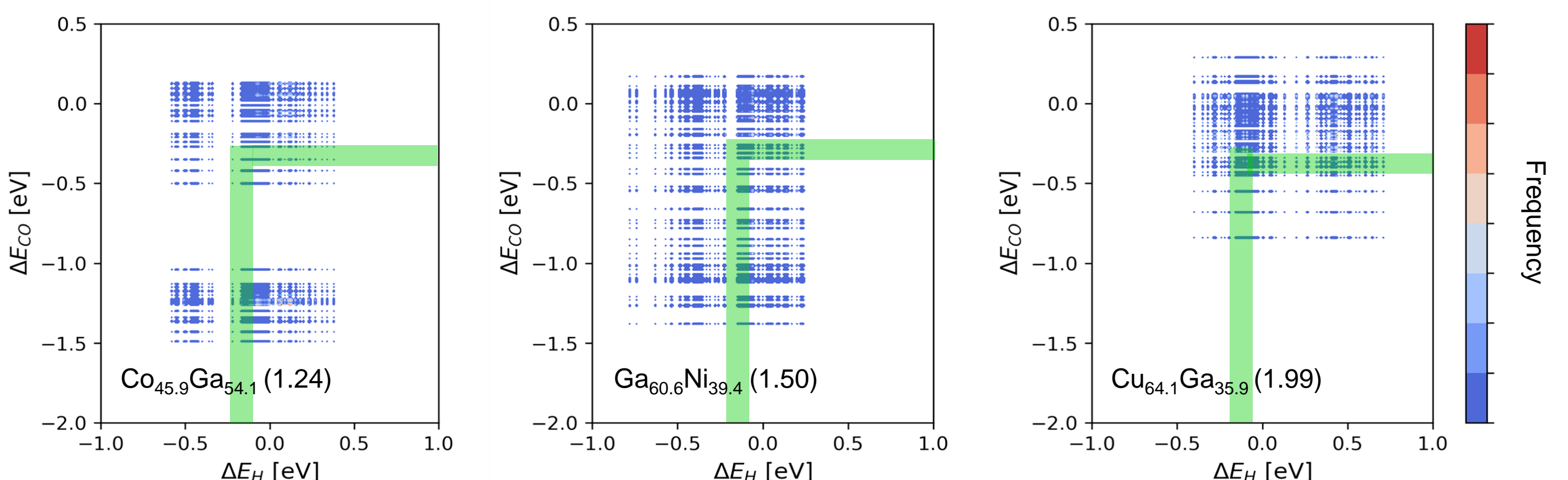

Figure 5

Predicted vs. DFT adsorption energies for **(a)** CO on-top, **(b)** H fcc, and **(c)** H hcp adsorption. 80 % of the data has been used for training and 20 % for validation. Predictions on a small test set of 3x4x4 slabs have also been performed. Root mean square deviations (RMSD) are shown. The dashed lines represent errors of ±0.1 eV. The elements displayed in color in **(a)** refer to the adsorbing atom.

Figure 6

Joint distributions of predicted CO and H adsorption energies for the three outputted compositions with penalty given in parenthesis.