

# Master Thesis

Marcus Nørgaard Weng

## Machine learning on simulated high entropy alloy catalysis for fuel cells

Supervisor: Jan Rossmeisl

Submitted on: 31 October 2023

Name of department: Department of Chemistry

Author(s): Marcus Nørgaard Weng

Title and subtitle: Machine learning on simulated high entropy alloy catalysis for fuel cells

Topic description: The project is about understanding high entropy alloy catalysis in fuel cell reactions. High entropy alloys are an increasingly researched topic partly because of their promising catalytic abilities. Fuel cells is an important use case of new catalysts, since they rely on good catalytic properties to function efficiently. Fuel cells can provide energy without greenhouse gas emissions and has applications where other green energy sources will not suffice and can therefore be a part of the technological solution to the climate crisis. The subject of the project is data treatment of DFT simulations of high entropy alloy catalysts and producing accurate models for fuel cell high entropy alloys during activity. The methods used for producing results will include machine learning and statistics.

Supervisor: Jan Rossmeisl

Submitted on: 31 October 2023

Number of characters: 108.661

# Table of contents

<b>Abstract.....</b>	<b>5</b>
<b>Acknowledgements .....</b>	<b>6</b>
<b>Introduction.....</b>	<b>7</b>
<b>Climate change .....</b>	<b>7</b>
History .....	7
Causes .....	8
Effects.....	9
<b>Fuel cells .....</b>	<b>9</b>
Structure.....	10
Reactions.....	11
Formic Acid.....	13
<b>Catalysis.....</b>	<b>20</b>
Sabatier's Principle.....	20
Computational Hydrogen Electrode.....	20
Correcting from DFT electronic energies to Gibbs free energies .....	21
<b>High-Entropy Alloys .....</b>	<b>23</b>
<b>Density Functional Theory.....</b>	<b>24</b>
Theoretical background .....	25
Practical use in catalysis .....	27
<b>Single-site structures.....</b>	<b>27</b>
<b>Machine Learning .....</b>	<b>28</b>
<b>Methods .....</b>	<b>30</b>
<b>Data.....</b>	<b>30</b>
Technical details .....	30
High-Entropy Alloy slabs with single adsorbates.....	32
High-Entropy Alloy slabs with neighbour-adsorbates .....	32
Single-site slabs with single and neighbouring adsorbates .....	32
Single molecules.....	32
<b>Setting reference energies .....</b>	<b>32</b>
<b>Predicting binding energies .....</b>	<b>36</b>

<b>Simulating surfaces .....</b>	<b>37</b>
<b>Searching for optimal compositions .....</b>	<b>38</b>
Optimization criteria .....	40
<b>Results.....</b>	<b>42</b>
Training binding energy models .....	42
Applying the models on simulated surfaces .....	43
Binding energies given neighbouring adsorbates.....	47
Applying the models to single-site structures.....	50
Searching for optimal HEA compositions at different potentials.....	52
Searching for the best single-site structured surfaces at different potentials.....	56
Comparing optimal HEA surfaces with single-site surfaces .....	58
<b>Discussion.....</b>	<b>59</b>
<b>Conclusion .....</b>	<b>60</b>
<b>References.....</b>	<b>62</b>

## Abstract

Climate change is caused by increasing CO<sub>2</sub> contents in the atmosphere largely due to the burning of fossil fuels for energy. Fuel cells running on green fuels propose a zero-emissions alternative, such as the direct formic acid fuel cell (DFAFC) using CO<sub>2</sub> as carbon feedstock for a net-zero loop. In order to get a highly efficient formic acid oxidation reaction, this work will focus on fine-tuning the anode catalyst, which is known to be limited by CO-poisoning at low and efficient potentials. High-Entropy Alloys (HEA) with PtPdCuAgAu are used as a catalyst discovery platform, where the binding energies of \*COOH and \*H will be tuned to optimize the activity at low anode potentials. Machine learning models will be trained on data from Density Functional Theory (DFT) calculations of the COOH on-top and the H hollow site binding energies on fcc(111) surfaces with random mixing and special single-site (SS) ordering. A brute-force search through the composition space will be employed with an activity estimation function that accounts for CO-poisoning. The nature of the DFAFC system dictates a trade-off between efficiency and power per site, so catalysts must be tailored to different scenarios. At anode potentials of 0.305 V and over, a pure Pt catalyst is suggested to be optimal, while Cu<sub>41</sub>Pt<sub>59</sub>, Ag<sub>69</sub>Pt<sub>31</sub> and Ag<sub>27</sub>Au<sub>12</sub>Cu<sub>33</sub>Pt<sub>27</sub> are suggested to be optimal at anode potentials of 0.27, 0.21, and 0.16 V vs RHE, respectively. The theoretical SS surfaces possess impressive abilities to avoid CO-poisoning, beating the activity of fine-tuned HEAs at low potentials, but even then, is limited by Pt \*COOH binding energies that are not as low as the optimum. Regardless, the results will show, that HEA anode catalysts can be tuned to be somewhat active, even in the CO-poisoning region, and that the theoretical SS surfaces could provide much higher power per site than their HEA counterparts at highly efficient anode potentials.

## Acknowledgements

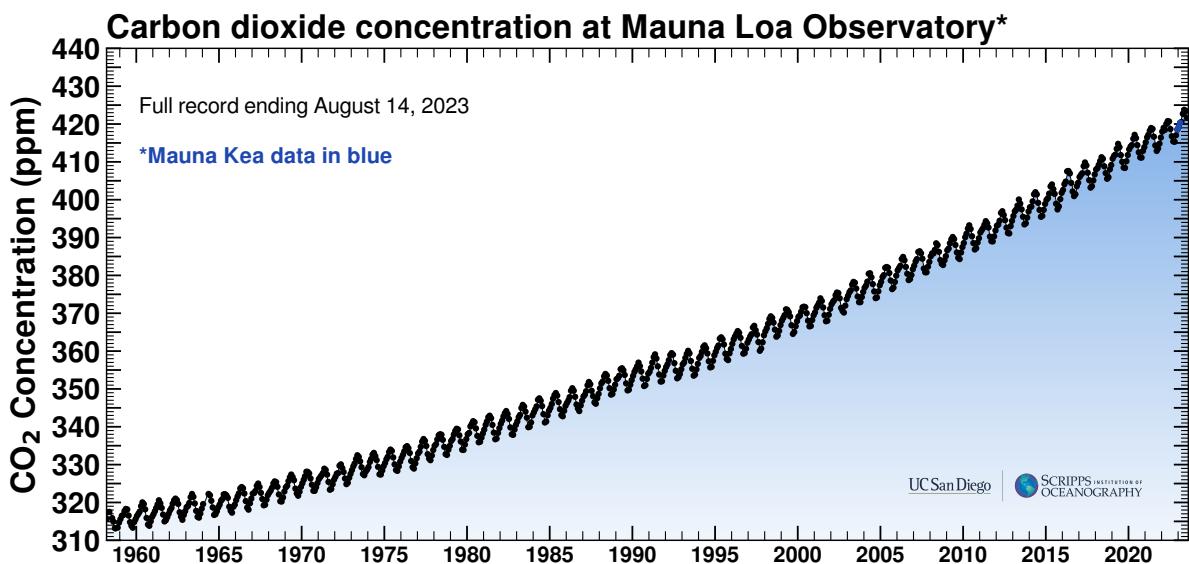
I thank Professor Jan Rossmeisl for providing a very interesting project and for inspiring supervisor meetings. I thank Postdoc Jack Kirk Pedersen for generous and patient day to day guidance. I thank the people at CHEAC for helping me with understanding the field of catalysis, particularly Jack, Henrik, Martin, Christian and Adam. I thank Alexander Bagger for his input on the thermal corrections.

# Introduction

## Climate change

### History

In 1896 Svante Arrhenius published a paper, where the link between CO<sub>2</sub> and the temperature on the ground is made for the first time<sup>1</sup>. In 1939 Guy Callendar argued, that anthropogenic emissions of CO<sub>2</sub> had raised the CO<sub>2</sub> content in the atmosphere by 10 percent since 1900<sup>2</sup>, while others believed, that the biosphere would have absorbed the anthropogenic emissions. In the 1950s a group of oceanographers and geochemists were concerned with the CO<sub>2</sub> concentration of the atmosphere because the ocean absorbs CO<sub>2</sub>, leading to more acidic surface water, which would affect sea life<sup>3</sup>. In 1957 Charles David Keeling was hired at a postdoctoral scientist position by Roger Revelle and ordered to measure the CO<sub>2</sub> concentration in the atmosphere at Mauna Loa, Hawaii. The results of Keeling's measurements were clear, the CO<sub>2</sub> concentrations rose even during the short 18-month period of measurements. The measurements suggested that at most half of the anthropogenic CO<sub>2</sub> emissions had been adsorbed by the biosphere. The plot showing CO<sub>2</sub> concentration in parts per million as a function time was coined the Keeling curve and has been measured since, showing the increase in CO<sub>2</sub> ppm rising from 313 ppm on March 29, 1958 to 421 ppm on august 2, 2023<sup>4</sup>. Historical data on CO<sub>2</sub> concentration in the atmosphere can be estimated before direct measurements were made via ice cores drilled up in Greenland and Antarctica. The past 800.000 years' atmospheric content is trapped in the ice that froze each year and can be analyzed now, giving a long-term historical perspective and data used to model climate change<sup>5,6</sup>. In figure 1, the overall increase in carbon dioxide concentration and the seasonal fluctuations can be seen, from the first measurements by Keeling at Mauna Loa Observatory in 1958 to 2023<sup>7</sup>.



**Figure 1 - Spectroscopic measurements of carbon dioxide content at the Mauna Loa Observatory on Hawaii from the first measurements in 1958 to 2023<sup>7</sup>**

## Causes

The core mechanism behind climate change is based on the interaction between infrared and visible light and so called “Greenhouse gasses” (GHGs). Electromagnetic radiation from the sun enters the atmosphere and reaches earth’s surface, where some of it is adsorbed, heating the surface, and some is reflected, possibly leaving the atmosphere again. The heated surface in turn reradiates some of the energy as infrared radiation, as described by Planck’s radiation law. GHGs interact with the infrared radiation to adsorb and re-emit infrared radiation in all directions, about half back onto the earth’s surface. This mechanism allows energy from sunlight to pass through the earth’s atmosphere unobstructed by the GHGs, while returning half of the energy from reemitted infrared radiation from the surface, creating the “greenhouse effect”<sup>8</sup>. The carbon cycle consists of large amounts (gigatons) of CO<sub>2</sub> being recycled through emitters like microbial respiration and decomposition and plant respiration and adsorbers, mainly photosynthesis. The anthropogenic carbon emissions offset the balance between these, resulting in an atmospheric carbon net annual increase. The anthropogenic carbon emissions stem mainly, 91 % in 2022, from the combustion of fossil fuels, such as oil, coal and gas<sup>9</sup>. An example of a combustion reaction of a fossil fuel is the aliphatic hydrocarbon C<sub>10</sub>H<sub>22</sub>, present in diesel fuel:



An alternative to fossil fuels is renewable energy, which harnesses the energy from renewable resources such as solar, wind, hydropower and geothermal. These renewable energy sources do

not emit CO<sub>2</sub> when producing energy, contrary to fossil fuels which, in the case of oil, returns carbon in the form of hydrocarbons from underground reservoirs back to the atmosphere.

## **Effects**

The direct effects of the anthropogenic GHG emissions are plentiful and a “threat to human wellbeing and health of the planet”, according to the IPCC report from 2022<sup>10</sup>. The average temperature of the atmosphere increases and has increased around 1.1 degrees Celsius since 1880<sup>3</sup>, projected to make 20 % of the planet a “barely livable” zone purely because of the heat<sup>10</sup>. The oceans are getting warmer and more acidic, due to the ocean absorbing 20 to 30 percent of all anthropogenic CO<sub>2</sub>, emissions posing a threat to ecosystems that inhabit the oceans<sup>3</sup>. Ice sheets on Greenland and Antarctica have declined in mass, leading to increasing sea levels, that threaten to flood the homes of 300 million people<sup>11</sup>. Furthermore, the frequency of extreme weather events is increasing<sup>12</sup>. From all this, it is apparent, that climate change is an issue on an apocalyptic scale, that must be addressed in order to secure human wellbeing and long-term existence on earth. The production of primary energy on earth increases steadily, as developing countries advance infrastructurally and technologically, highlighting the need for renewable and sustainable energy sources to replace fossil fuels. This calls for development of technology that can aid the transition to renewable energy sources. To shift all carbon emitting processes to renewable, non-carbon emitting sources of energy would require many different solutions, both in terms of energy storage and fuel cells. Container ships, cars, and airplanes, three large fossil fuel reliant transport vessels and carbon emitters, have very different needs in terms of capacity, weight limitations and power consumption. The electrical grid would also need to have a large-scale battery in order to supply electricity produced from renewable sources at all times, even when wind and solar do not produce energy and fossil fuels take over. One solution to storing energy from renewable sources is green fuels and fuel cells. Bearing in mind that 91 % of all anthropogenic emissions stem from energy production with fossil fuels, a complete transition of the energy sector to provide plenty of renewable energy, both to developed and developing countries with growing energy needs, is essential to tackling climate change<sup>9</sup>.

## **Fuel cells**

A fuel cell (FC) is an electrochemical cell that converts the stored chemical energy in the fuel into electricity through a series of chemical reactions, rather than combustion. The “green” fuels for fuel cells are made with electrochemical reactions using surplus renewable energy. The fuels

then serve as an alternative way to store excess electrical energy compared to conventional storage media, such as electrical batteries. Storing excess electrical energy as green fuels has a variety of advantages. The storage space for a liquid green fuel is cheaper than a Li-Ion battery with the same energy capacity and can be stored for long periods of time without self-discharging and cycle indefinitely without decreasing the max capacity. Furthermore, Li-ion batteries suffer from low gravimetric and volumetric energy densities. A drawback of green fuel and fuel cells is the round-trip efficiencies, where Li-ion batteries generally have round-trip efficiencies (RTEs) of 90%+, while hydrogen has an RTE from electrical energy to hydrogen, through electrolysis, back to electrical energy, via an electrochemical fuel cell, of 18-46 %<sup>13</sup>. Worthwhile green fuel production hence requires both an efficient oxidation reaction of the green fuel and an efficient reduction reaction of the precursor to the fuel. Good round-trip efficiencies are therefore highly reliant on the development of tailor-made catalysts for these reactions. Compared to combustion engines, fuel cells have the advantage, that they convert chemical energy directly into electricity, whereas combustion engines convert the chemical energy into heat, then mechanical energy, which by Carnot's theory has a limited efficiency. Carnot's theory does not apply to fuel cells since a conversion from heat to mechanical energy doesn't take place. The efficiencies of electrochemical reactions are limited by other factors, such as overpotentials due to suboptimal catalysts for various reactions.

## Structure

A proton-exchange membrane (PEM) fuel cell consists of two separate inlets for fuel and O<sub>2</sub> at fuel field plates, a gas diffusion cathode and a gas diffusion anode, a catalyst layer at each electrode and a PEM to separate the electrodes. The PEM is an electrically insulating layer, that allows the passage of protons. The catalysts at each electrode should ideally bind the intermediates of each half-cell reaction at the optimal binding energy, in order to minimize the overpotential. The green fuel enters the inlet at the gas diffusion anode, where it partakes in the oxidation half-cell reaction. When the fuel has been oxidized, the excess proton-electron pairs (H<sup>+</sup> + e<sup>-</sup>) split up, as the proton moves through the PEM to the cathode, where it will partake in the reduction half-cell reaction, where an electron will then be needed. This creates a voltage difference between the negatively charged anode and the positively charged cathode, from which an electrical current can be pulled. An assembled stack of PEMs constitutes a Membrane Electrode Assembly (MEA). In the case of direct formic acid fuel cells (DFAFCs), there are three main ways of supplying the fuel. Active DFAFCs, where fuel and oxygen in compressed air are delivered to the inlets via pumps, active air-breathing DFAFCs, where fuel is pumped in,

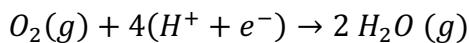
but the oxygen is supplied from exposing the cathode to the ambient air, and passive air-breathing DFAFCs, where both fuel and oxygen is supplied by diffusion from a fuel tank and ambient air<sup>14</sup>. A DFAFC PEM formic acid fuel cell has a theoretical open-circuit voltage of 1.48 V versus 1.18 V for methanol<sup>14</sup>, which is promising in terms of fuel cells.

## Reactions

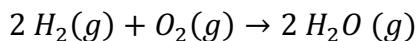
Fuel cells harness the chemical energy stored in the green fuel molecules with a set of redox-reactions, namely two half-cell reactions. In the oxidation reactions, the fuel is oxidized a number of times, each producing a proton-electron pair. In the reduction reaction a molecule, typically O<sub>2</sub>, is reduced to H<sub>2</sub>O. The simplest example is the half-cell reactions in a hydrogen fuel cell, that has the oxidation reaction:



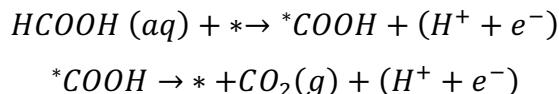
And the 4e<sup>-</sup> pathway of the oxygen reduction reaction (ORR):



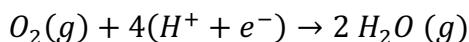
Resulting in a total reaction of hydrogen reduction:



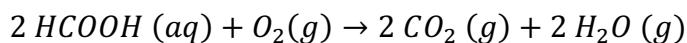
Formic acid (FA) can be reduced to CO<sub>2</sub>, providing two proton-electron pairs, as seen in the half-cell reactions. The oxidation reaction examined in this work is a two-step reaction, where each step releases a proton-electron pair:



The reduction reactions are the same as for the hydrogen fuel cell, with gaseous oxygen being reduced to water:



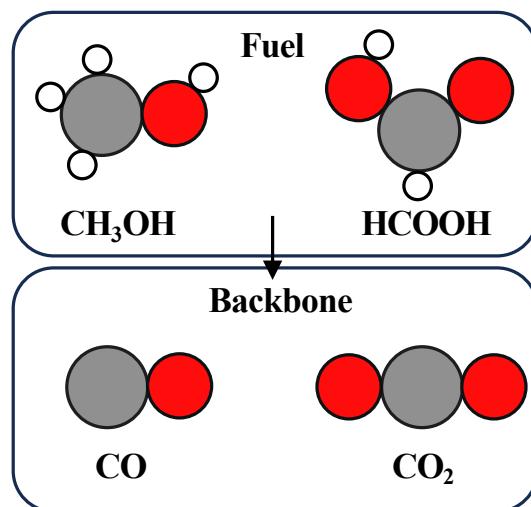
Resulting in a total reaction of formic acid reduction:



Methanol, a direct competitor to formic acid, provides 6 proton-electron pairs per reacted molecule, compared to 2 for formic acid, resulting in a higher volumetric energy density. That also leads to a more complicated oxidation reaction, that involves six reaction steps instead of the two for FA<sup>15</sup>.

Fuel	Volumetric energy density (kWh/L)	Theoretical open-circuit voltage (V)	Price (US\$/kg, retail)	Operating temperature (°C)	Storage pressure (bar)
FA	2.13	1.48	0.7	20-60 (DFAFC)	-
H <sub>2</sub>	0.53	0.9-1.0 <sup>16</sup>	2.6 - 5.1	150-200 (AFC)	700
CH <sub>3</sub> OH	4.4 - 4.9	1.18	0.2 - 0.4	30-90 (DMFC)	-

In the table above, formic acid, methanol and hydrogen is compared on a few important parameters<sup>14</sup>. Formic acid is first and foremost practical because it is a liquid and doesn't have to be stored under high pressures, which also gives it a volumetric energy density four times higher than hydrogen. The theoretical open-circuit voltage of formic acid is higher than both methanol and hydrogen, and it has a high theoretical max efficiency. The ideal methanol catalysts performance in terms of energy densities is much better than formic acid, but in terms of the current performance, formic acid is closer to its ideal catalyst. The methanol catalysts are running at a higher overpotential than formic acid.



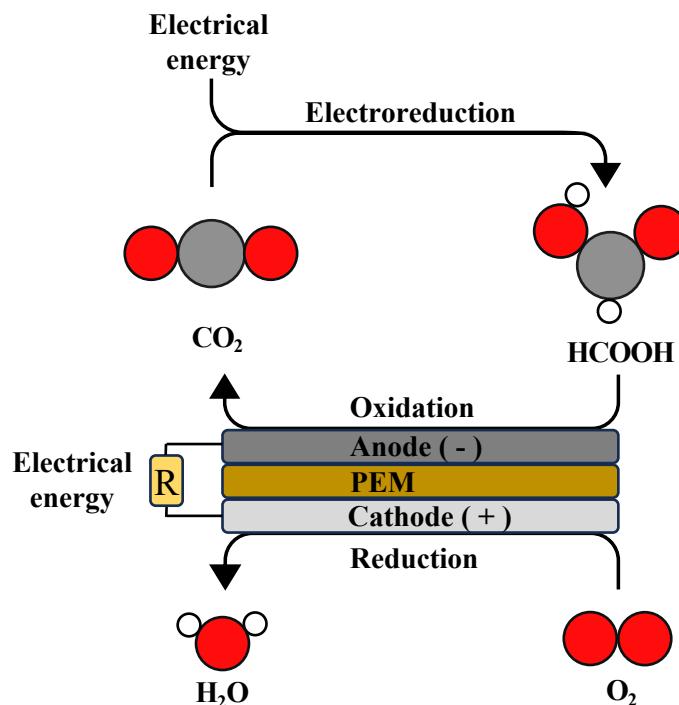
**Figure 2 – Methanol and formic acid, and their "backbones", the structure without any hydrogens. This is the structure the fuels would have after undergoing oxidation.**

The reason behind this is apparent, when observing formic acid and methanol side by side. In figure 2, the structures of methanol and formic acid are shown before (upper) and after (lower) undergoing oxidation. Methanol has a carbon monoxide backbone, which is problematic, since it

ensures carbon monoxide poisoning of the surface, when oxidized. When a carbon monoxide must be oxidized by OH or H at a larger potential, it severely limits the efficiency of a methanol fuel cell. The backbone of formic acid, however, is carbon dioxide, which is gaseous at operating conditions and will desorb by itself. This is a big advantage and allows for a more efficient fuel cell.

### Formic Acid

Formic acid is a promising green fuel, primarily because of its high volumetric energy density (2.13 kWh/L), open-circuit voltage (1.48 V) and theoretical energy efficiency (58 %)<sup>17</sup>. Formic acid is a liquid at standard conditions and an acid with a pK<sub>a</sub> of 3.7 and it is non-toxic and present in small quantities in foods and vegetables. Formic acid is therefore safe and easy to handle, as opposed to methanol, which is toxic.



**Figure 3 - The proposed formic acid cycle with  $\text{CO}_2$  formed through electroreduction. The formic acid is oxidized to  $\text{CO}_2$  on the anode, while in the other half-cell reaction, on the cathode,  $\text{O}_2$  is reduced to  $\text{H}_2\text{O}$ .**

Formic acid can be used as an energy carrier for fuel cells with  $\text{CO}_2$  as “carbon feedstock”<sup>14</sup>. FA is an obvious choice for a cycle with  $\text{CO}_2$ , since they are reasonably similar in structure, the only difference being a hydrogen bonds on C and O, respectively. This makes electroreduction and oxidation somewhat simple, as a middle ground between  $\text{H}_2$  and  $\text{CH}_3\text{OH}$ . The production of FA with electroreduction of atmospheric  $\text{CO}_2$  as the starting compound and the oxidation of FA in a fuel cell represents a closed carbon loop capable of storing energy from preferably renewable electricity sources. When FA is produced by electroreduction of  $\text{CO}_2$ , cell conditions are neutral

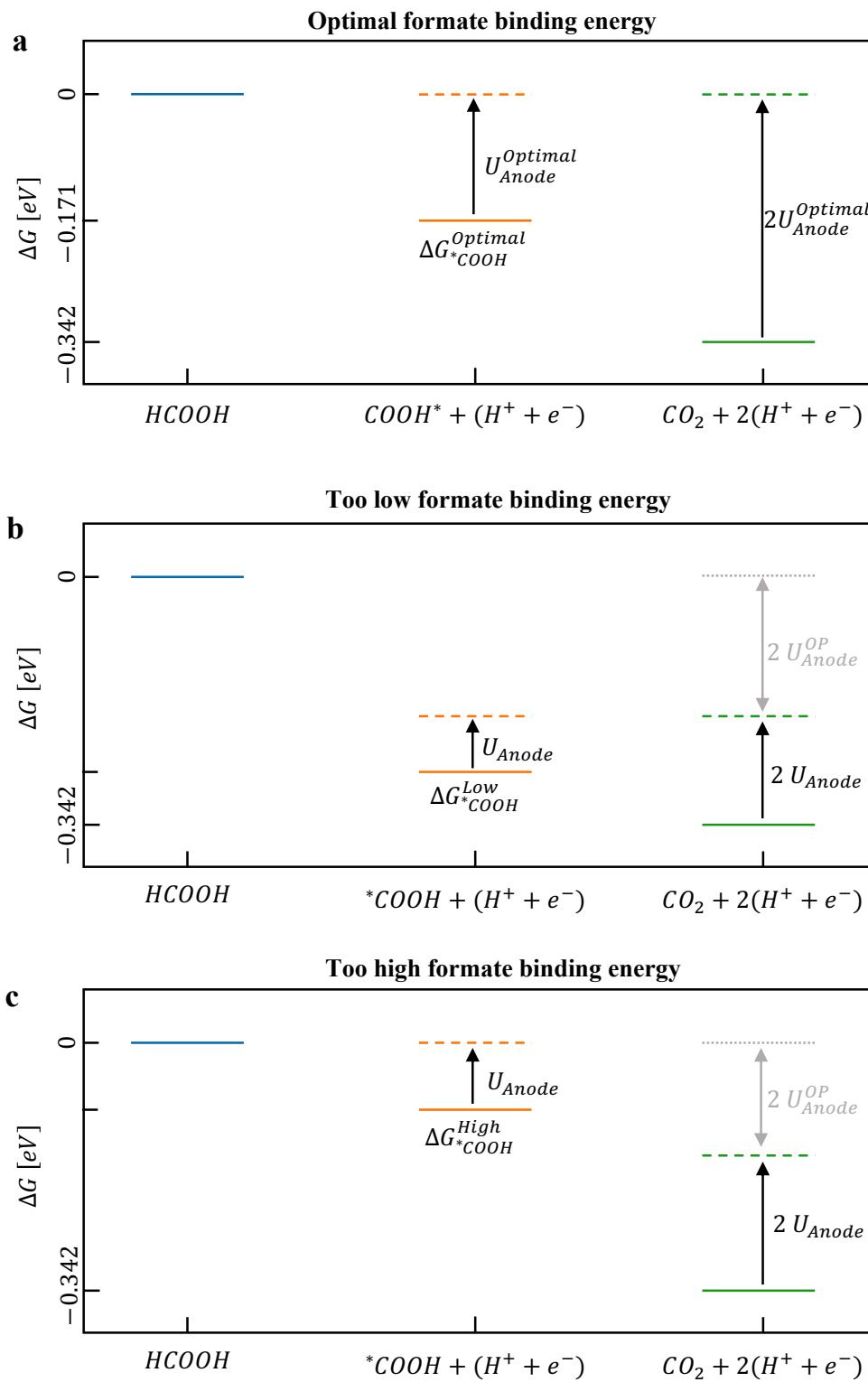
to alkaline and produce formate ( $\text{COOH}^-$ ), which becomes FA when the pH is lowered<sup>14</sup>. The closed carbon-loop is illustrated in figure 3 along with the cathode half-cell reaction and the production of electrical energy in the fuel cell. The oxidation reaction produces two proton-electron pairs, of which only the proton can move through the Proton-Exchange Membrane (PEM), and the electron remains on the anode, making it negatively charged compared to RHE. Conversely, on the cathode catalyst, the reduction reaction of  $\text{O}_2$  consumes four proton-electron pairs, leaving the cathode positively charged compared to RHE. The half-cells each contribute to a voltage-difference between the negatively charged anode and the positively charged cathode. The open-circuit voltage between the anode catalyst and the cathode catalyst is a result of surplus electrons produced by FAOR on the anode and electrons used for OER. Connecting a resistor (R) to the anode and cathode will draw a current as dictated by Ohm's Law:

$$I = \frac{U}{R}$$

The actual power drawn is a product of the voltage and the current:

$$P = U \cdot I$$

The open-circuit voltage ( $V_{\text{oc}}$ ) is the voltage that will be measured when the FC is not connected to a resistor drawing current. The voltage under load will always be lower than  $V_{\text{oc}}$ . The energy harvested from FAOR is directly proportional to the number of proton-electron pairs produced times the potential, as each electron can do the electrical “work” of one eV per V.



**Figure 4 – The Gibbs free energies of the species in the two-step pathway of formic acid oxidation at three different formate binding energies and hence three different anode potentials vs RHE. The energies are related to formic acid, which is set to 0 eV. The solid lines represent the Gibbs free energies before the energy contribution from the potential is added, calculated with the Computational Hydrogen Electrode (CHE). The dashed lines represent Gibbs free energies after the potential contribution has been added. The anode potentials are marked with arrows, while the anode overpotential,  $U_{Anode}^{OP}$ , is marked with a grey double-headed arrow.**

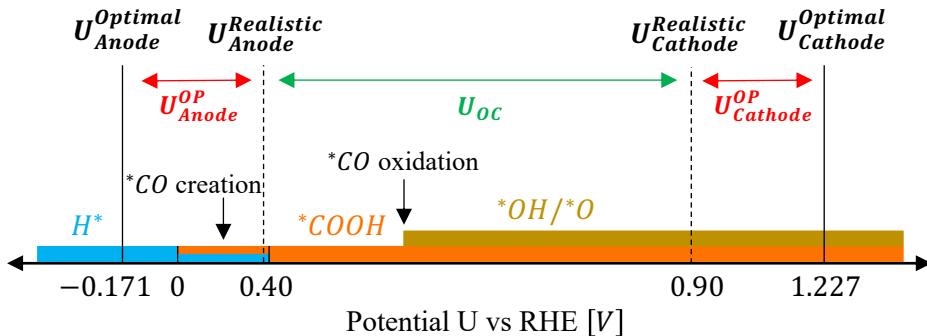
In figure 4, the Gibbs free energy of all three species involved in the two-step formic acid oxidation pathway including proton-electron pairs are plotted, as compared to the Gibbs free energy of formic acid. The starting point is formic acid at 0 eV compared to formic acid itself. The Gibbs free energies of  $^*COOH$  and  $CO_2$  including the proton-electron pairs are calculated as:

$$\Delta G_{^*COOH} = G_{^*COOH} - G_{HCOOH} - G_* + \frac{1}{2}G_{H_2} - eU$$

$$\Delta G_{CO_2} = G_{CO_2} - G_{HCOOH} - G_* + G_{H_2} - 2eU$$

Where  $\Delta G_{^*COOH}$  represents the change in Gibbs free energy from formic acid to  $^*COOH$ , the proton-electron pair contributes  $\frac{1}{2}G_{H_2} - eU$  to the energy as described by the computational hydrogen electrode (CHE), where U represents the anode potential vs RHE and e represents the elemental charge. The solid lines represent the energies at an anode potential of 0 V vs RHE, while the dashed lines represent the Gibbs free energies at the largest negative potential, where the Gibbs free energy doesn't increase between two steps. When  $CO_2$  has been produced, two proton-electron pairs have been made, and the negative potential affects that step with  $-2eU$ . The COOH step only has its Gibbs free energy increased with  $-eU$ . This mechanic determines the optimal binding energy, which makes the largest negative anode potential vs RHE possible. The optimal binding energy of formate is  $\Delta G_{^*COOH} = \frac{1}{2}\Delta G_{CO_2}$ , which enables the largest possible anode potential  $U_{Anode}^{Optimal} = -0.171\text{ V}$  vs RHE. This scenario is shown in figure 4a, where the binding energy of formate is placed optimally to ensure the maximum negative potential. In the figure, the upwards direction represents a negative potential on the anode vs RHE. In figure 4b, the binding energy of formate is too low, leading to a situation, where the Gibbs free energy of the  $CO_2$  step exceeds the formate step before the largest potential is reached, which would discourage the reaction from taking place. This results in an overpotential,  $U_{Anode}^{OP}$ . In figure 4c, the binding energy of formate is too high, leading to a situation, where the Gibbs free energy  $^*COOH$  step exceeds that of the formic acid step before the optimal potential is reached, resulting in a similar overpotential. The harvestable energy stems from the highest potential, at which the reaction will still run, meaning the energy difference between two steps cannot increase. The adsorption reaction is favoured in an absence of  $(H^+ + e^-)$ , and the binding energy increases, as the potential U falls. When the binding energy of formate increases to  $\Delta G_{^*COOH} = 0$ , the adsorption reaction is no longer favoured, and the lowest anode half-cell voltage is reached. The highest possible open-circuit voltage between anode and cathode is reached, when  $^*COOH$  has the optimal binding energy to the catalyst surface, otherwise an overpotential is needed. In

practicality, this is impossible to avoid completely since no metal exists with exactly the optimal binding energy of COOH.

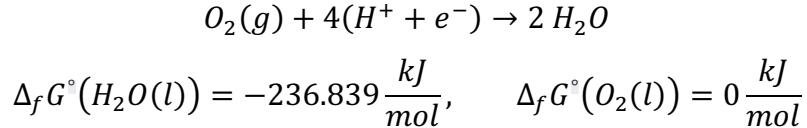


**Figure 5 – The potentials at which relevant adsorbates, H, COOH, OH and O are expected to be present on an equimolar HEA fcc(111) surface is marked. H adsorbs at small anode potentials and up to 0.40 V vs RHE for some sites, COOH starts to adsorb from 0.0 V vs RHE and up. OH, and O require larger potentials to adsorb. Four potentials are marked, indicating an optimal anode/cathode potential and dashed lines, indicating attainable anode/cathode potentials. The double-headed arrows represent overpotentials (red) and the open-circuit potential (green).**

In figure 5, the CO-poisoning problem for FAOR is highlighted, showing the adsorbates expected to be present at the anode catalyst, an equimolar simulated High-Entropy Alloy (HEA) fcc(111) surface at different anode potentials vs RHE, based on literature and DFT calculations<sup>18</sup>. H adsorbs at low potentials due to the adsorption reaction of H being favoured at negative potentials, where surplus electrons are present. H desorbs at positive voltages, predicted to be completely removed at 0.40 V vs RHE. H overlaps in presence with \*COOH, which is favoured at higher potentials, since its adsorption reaction produces a proton-electron pair. The overlap of the presence of H and COOH at voltages 0.0 V to 0.4 V is believed to be the source of carbon monoxide poisoning, due to a disproportionation reaction between a neighbouring H in a hollow site and a COOH on an on-top site, resulting in bound carbon monoxide and water. Doing formic acid oxidation at this overlap potential range is believed to result in a CO-poisoned surface, and this has been indicated by experimental observations<sup>18</sup>. At higher potentials, OH and O adsorb to the surface, which can oxidate \*CO to CO<sub>2</sub>. The potential, at which carbon monoxide is created by the disproportionation reaction is much lower than the potential at which OH and O are adsorbed and able to remove bound CO by oxidation, which requires cycling to high potentials to clean. Two optimal potentials, anode, and cathode, marked with solid lines in figure 5, represent the best possible based on thermodynamic calculations, that would give the maximum theoretical open-circuit potential and optimal efficiency. A realistic cathode potential is 0.90 V vs RHE, leading to a 0.327 V overpotential of lost energy (in eV) per electron harvested. An anode potential of 0.40 V vs RHE, where H does not bind on any sites on the HEA

surface, would lead to an anode overpotential of  $0.571\text{ V}$ , bringing the total overpotential to  $0.898\text{ V}$ , and an open-circuit voltage of  $0.50\text{ V}$ .

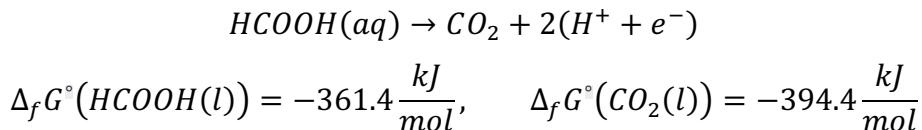
The cathode half-cell reaction potential vs RHE is calculated:<sup>19</sup>



The energy released per electron is:

$$\frac{2 \cdot (236.839 \frac{\text{kJ}}{\text{mol}}) - 0 \frac{\text{kJ}}{\text{mol}}}{4} = -1.227\text{ eV}$$

Corresponding to a cathode potential vs RHE of  $1.23\text{ V}$ . The optimal anode half-cell reaction potential vs RHE is calculated:

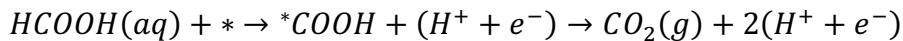


The energy released per electron is:

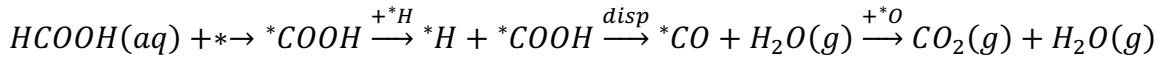
$$\frac{-394.4 \frac{\text{kJ}}{\text{mol}} - (-361.4 \frac{\text{kJ}}{\text{mol}})}{2} = -0.171\text{ eV}$$

Corresponding to an optimal anode potential vs RHE of  $-0.171\text{ V}$ , resulting in a theoretical maximum open-circuit voltage  $U_{OC}$  of  $1.227\text{ V} - (-0.171\text{ V}) = 1.398\text{ V}$ . Assuming that the cathode potential vs RHE cannot be improved beyond  $0.90\text{ V}$ , the theoretical maximum open-circuit voltage would be limited to  $1.071\text{ V}$ . The optimal anode potential is the potential at which the formate at an optimal catalytic site adsorbs to the surface. The binding energy at the optimal site can be seen in figure 4a, where the potential that increases the binding energy to  $0\text{ eV}$  is the optimal anode potential. The difference in potential between the anode potential vs RHE and the cathode potential vs RHE represents the open-circuit voltage, while the difference between the optimal anode potential and the actual anode potential represents the overpotential on the anode. The current output of the fuel cell is dependent on the activity of FAOR on the catalyst. The task is therefore to find an anode catalyst, that has a good activity on as low of a potential vs RHE as possible, while still not suffering from CO-poisoning. The efficiency of the process scales with the open-circuit voltage, which increases with lower anode potentials, but lower anode potentials also diminish the activity, as more sites become FAOR active at larger potentials, which encourage the adsorption reaction of COOH. Realistically, there will be a trade-off between efficiency and activity, where the application of the fuel cell will dictate the specific needs. In

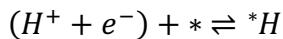
in the literature, a handful FAOR pathways are suggested, but the primary two pathways assumed to be important is the desired pathway<sup>18</sup>:



And the pathway where the disproportionation reaction produces  ${}^*CO$ :



Where H adsorbs with the adsorption reaction:



In this example oxidation of bound carbon monoxide is shown with O but can also happen with OH. The potential at which H adsorbs to the surface is lower than the potential at which O adsorbs. Poisoning the surface with carbon monoxide and “cleaning” the surface requires cycling from small potentials to higher potentials. Running the fuel cell at anode potentials, where oxidation of carbon monoxide happens is not ideal, as it requires a large anode overpotential. It is well known, that the efficiency of FAOR FCs is hindered by CO-poisoning<sup>18,20</sup>. This hypothesis is supported by single-site (SS) experiments, where catalysts are prepared with a structure, that has an on-top site with a reasonable binding energy for formate, but no feasible hollow sites next to it with good binding energies for H. An example is Al-Akraa et. al. modifying a glassy carbon (GC) electrode with multi-walled carbon nanotubes (MWCNT) and platinum nanoparticles, where the carbon nanotubes helped to separate the active Pt sites, interrupting their contiguity<sup>20</sup>. This was observed to increase the current compared to a control catalyst without MWNCTs. A similar study was conducted by Ortiz-Ortega et. al., where Pd nanobars (PdNB/C) were compared to commercial Pd/C and Pd nanoparticles (PdNP/C), where PdNB/C showed a lower poisoning-degree and a negative shift in the onset potential for FAOR at high (1 M) concentrations compared to commercial Pd/C<sup>21</sup>. Even though these studies do not directly detect CO on the surface, or prove certain reaction pathways, they indicate a link between single-sites and FAOR performance. While these are experimental results attesting to the existence of CO-poisoning, there has also been research into the specific reaction mechanisms with DFT calculations<sup>22</sup>. These are discussed further in the supplementary information. As seen in the experimental studies mentioned, the CO-poisoning has suggested solutions entailing complex nanostructures like nanotubes and nanobars for the catalysts in order to produce single-sites for adsorption. This work attempts a simpler solution, creating a HEA fcc(111) surfaces, that deters CO poisoning at as low anode potentials as possible in order to achieve a good open-circuit current and hence, efficiency. This could be possible with

single-sites on a HEA fcc(111) surface by tailoring the binding energies of the sites and neighbouring sites of sites of interest.

## Catalysis

A catalyst is a participant in chemical reactions, that can lower the energy barrier, that molecules must overcome in the transition from reactant to product, without undergoing any alterations itself. This can be done through stabilizing a reaction intermediate, possibly opening up new, previously unfeasible, pathways, or greatly increasing the reactivity. A good catalyst will exhibit a specificity towards the desired product, by virtue of the interplay between the molecular structure and electronic configurations of the catalyst and the reactants. An essential part of tuning a catalyst to a specific reaction pathway is the binding energies of relevant species.

### Sabatier's Principle

Sabatier's principle, named after Paul Sabatier, describes the optimal binding energy of an adsorbate guided by two factors limiting the activity<sup>23</sup>. The principle suggests that the catalyst should stabilize the intermediate not too weakly, so that it can partake in the reaction, but neither too strongly, limiting the desorption of products. The binding strength should then balance between limiting the activity by desorption of production and limiting the activity by activation of the reactant. Details on how to model the activity of a catalytic site is described in further details in the supplementary information.

### Computational Hydrogen Electrode

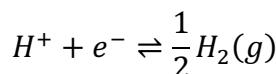
Electrochemistry revolves around chemical reactions that involve moving electrons, creating a potential or being driven by a potential. Knowing which potentials various reactions take place at is important to understand the systems and reactions examined. Meanwhile, measuring potentials precisely is a difficult task, that has to also ensure replicability across different electrodes in different laboratories. Potentials aren't used practically in isolation on a single electrode; a potential is said to exist as a potential difference between two electrodes, for example between a working electrode and a reference electrode. The "absolute" electrode potential is defined by IUPAC definition, relying on the difference between the fermi level inside the metal and an electron at rest in vacuum<sup>24</sup>. The Standard Hydrogen Electrode (SHE) is based on the reversible half-reaction of the dissociation of H<sub>2</sub> on a platinum electrode, while H<sub>2</sub> is bubbled through a strongly acidic solution, typically HCl:



The SHE has an estimated absolute potential of  $4.44 \pm 0.02$  V, but is defined by IUPAC to have a potential of 0 V.

$$E_{abs}^M = E_{SHE}^M + (4.44 \pm 0.02) V$$

Where E is the electrode potential, M is the metal of the electrode, abs is the absolute electrode potential, SHE is the electrode potential relative to the SHE. With the SHE, other half-cell reactions potentials can then be measured, and the potential can be reported relative to the SHE half-cell reaction potential<sup>24</sup>. As the SHE can be prepared similarly in different laboratories, a common reference electrode is achieved. However, a practical improvement is made in the Reversible Hydrogen Electrode, which is able to reversibly establish equilibrium under experimental conditions. In computational electrochemistry, a theoretical reference electrode is needed. The Computational Hydrogen Electrode (CHE) is the theoretical reference electrode used for electrochemical calculations. The CHE is based on the potential at which the reaction:



is at equilibrium, meaning the chemical potentials of both sides are equal:

$$\mu(H^+) + \mu(e^-) = \frac{1}{2} \mu(H_2)$$

At a potential U not equal to 0 V vs. RHE, the potential contributes with a term relating the chemical potential to the electric potential:

$$\mu(H^+) + \mu(e^-) = \frac{1}{2} \mu(H_2) - e \cdot U_{CHE}$$

When an electrochemical reaction produces a proton-electron pair, it contributes to the reaction energy with:

$$+ \frac{1}{2} E_{H_2(g)} - e \cdot U$$

Where the E represents an electronic energy from a DFT calculation or a Gibbs free energy, e is the elemental charge and U is the potential. The signs are opposite if a proton-electron pair is used as a reactant.

### **Correcting from DFT electronic energies to Gibbs free energies**

The energies obtained through DFT calculations are electronic energies. The electronic energy of a species is converted to Gibbs free energies with this formula, with approximations for the correction factors:

$$G = E + ZPE + \int C_p dT - TS + \Delta G_{solvent}$$

Where E represents the electronic energy as yielded by the DFT calculations, ZPE represents the Zero-Point Energy, Cp represents the specific heat capacity, TS represents entropy,  $\Delta G_{solvent}$  represents solvation energy. Practically, a correction constant is calculated a single time for each species:

$$C_{correction} = ZPE + \int C_p dT - TS$$

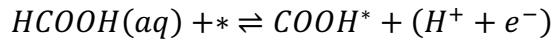
$$G = E + C_{correction}$$

The correction factors for zero-point energy, specific heat capacity and entropy are estimated in an article by Chan et. al.<sup>25</sup>. The corrections are known as thermal corrections, and collectively serve to ensure that the electronic energies, corresponding to ground state structures calculated with DFT at 0 K, are comparable to materials at realistic conditions<sup>26</sup>. Each correction factor handles a temperature-related energy contribution that is different in DFT conditions and real conditions. The Zero-Point Energy (ZPE) refers to the lowest possible energy state of a quantum-mechanical system. Even at the lowest possible energy state, the particles are still in motion, and this residual energy due to the motion is the ZPE. The ZPE correction accounts for the vibrational energy of the particles at the systems relaxed structure. The specific heat capacity ( $C_p$ ) describes the relation between the energy added to a material and the temperature increase. The specific heat capacity changes at changes from 0 K to realistic temperatures as more degrees of freedom, vibrational and rotational, are unlocked. The specific heat capacity correction accounts for the additional energy the system exhibits at realistic temperatures<sup>26</sup>. The entropy of a system describes the “disorder”, relating to the number of possible ways to arrange the particles leading to the same energy. Entropy is important to a reaction’s thermodynamics, as the entropy can drive reactions, if the entropy is higher in the products than the reactants. The Gibbs free energy change<sup>27</sup>:

$$\Delta G = \Delta H - T\Delta S$$

Incorporates the entropy to provide a measure of the thermodynamic favorability of the reaction. The entropy of a system changes from 0 K to realistic temperatures. The  $-TS$  correction transforms the electronic energy into a Gibbs free energy. For the non-adsorbed species, the chemical potentials were estimated with ideal-gas methods<sup>25,26</sup>. For the adsorbed species, the correction factors for the thermal corrections were calculated with the harmonic oscillator approximation. A simplification was made by treating all degrees of freedom as harmonic vibrations. The approximations as calculated at T=298.15 K. With these correction factors, binding energies can be calculated, by first converting all electronic energies, both from non-

adsorbed and adsorbed species. An example of an adsorption reaction and the calculations leading to the binding energy is shown underneath:



The thermal corrections for each species in the reaction is:

$$\begin{aligned} G_{HCOOH} &= E_{HCOOH} + ZPE_{HCOOH} + \int Cp_{HCOOH}dT - TS_{HCOOH} \\ G_{COOH^*} &= E_{COOH^*} + ZPE_{COOH^*} + \int Cp_{COOH^*}dT - TS_{COOH^*} \\ G_{H_2} &= E_{H_2} + ZPE_{H_2} + \int Cp_{H_2}dT - TS_{H_2} \\ G_{H_2} &= E_* \end{aligned}$$

The binding energy in Gibbs free energy change is:

$$\Delta G = G_{COOH^*} - G_{HCOOH} - G_* + \frac{1}{2}G_{H_2} - e \cdot U$$

With this routine, a binding energy can be estimated based on DFT calculations for arbitrary reactions, including all the reactions of interest to FAOR. To obtain the necessary energies, DFT calculations is used for geometry relaxation and energy calculations of a metal catalyst surface slab, with and without adsorbate, and the isolated molecules.

## High-Entropy Alloys

A high-entropy alloy (HEA) is an alloy composed of five or more metals in equimolar or near-equimolar fractions, first described in the context of catalyst-discovery in a 2004 paper, whereas previously, the design of alloy systems was based on a primary element as the matrix<sup>28</sup>. HEAs are stabilized by the configurational entropy attained when mixing the elements randomly in a single-phase solid structure. The configurational entropy change per mole can be described as:

$$\Delta S_{conf} = -k \cdot \ln (W)$$

Where k is Boltzmann's constant and W is the number of ways the elements could be mixed. The configurational entropy change of a HEA with 5 equimolar elements is 1.61R where R is the gas constant, whereas Richard's rule<sup>29</sup> predicts the entropy change in the fusion of most metals to only approximatively 1 gas constant. The resulting thermodynamic stabilization of a solid solution in solidification is described by the Gibbs energy<sup>30,31</sup>:

$$G^{mix} = H^{mix} - TS^{mix}$$

Consequently, an entropy term larger than the enthalpy of mixing indicates a stable single uniform phase HEA. A HEA with more elements has a higher configurational entropy change leading to a more stable structure. As explained in the catalysis section, the performance of a catalyst is highly dependent on the adsorption energy of the involved species, both in optimizing the activity and efficiency of the catalyst. In pure metal catalysts, the range of adsorption

energies of a given species is relatively narrow, and more importantly not tunable. HEAs are of great interest as a discovery platform for catalysts, due to the range of possible binding energies attainable with many equimolar elements leading to extremely many possible local surface environments. It is hypothesized, that a near-continuum of binding energies can be achieved with enough elements in the alloy<sup>28,32,33</sup>. Adding a new element to the composition causes a widening of the binding energy “bands” of the elements already present, covering a larger portion of binding energies. HEAs have been used experimentally to cater to specific reactions<sup>28,33</sup>. If it is possible to predict the adsorption energy of a given site based on the local surface environment, the composition of the surface can be optimized for a chosen parameter of interest, which could be a descriptor of performance of a specific reaction. With the number of different local environments around a site, if considering e.g. 13 local atoms, the brute-force approach of calculating the binding energy of a species on all combinatorically thinkable sites using DFT calculations is completely impossible. Hence, DFT is merely used to make a decent training set for a simpler binding energy prediction model to be trained on, opening up the possibility of decently accurate predictions at very high speeds. This work seeks to apply a descriptor to FAOR and tune the molar fractions of HEAs to cater to that descriptor. The metals chosen for the HEAs for this work are Platinum (Pt), Palladium (Pd), Copper (Cu), Gold (Au) and Silver (Ag). Experiments have shown both Pt and Pd to have some FAOR activity<sup>18,34,35</sup>, and Pt is known to bind COOH, the best of the chosen metals. Cu, Au, and Ag are known to bind both formate and H worse than Pt and Pd and can be used to widen the binding energy bands of Pt and Pd. Additionally, they could ideally create structures in the surface, that increase H binding energies while maintaining feasible formate binding energies. The radii of the chosen metals are reasonably similar, 145 pm to 177 pm, and are all known to form Face-centered cubic (fcc) structures to comply with the Hume-Rothery rules for solid solutions<sup>31</sup>.

## Density Functional Theory

Density Functional Theory (DFT) is an important tool to estimate observables for many-body quantum systems, such as atoms, molecules, and condensed matter. This is done with functionals, functions of functions, which operate on the estimated ground-state density distribution functions. In this work, DFT is used to estimate the energies of fcc(111) slabs with various adsorbates.

## Theoretical background

The Schrödinger Equation is the quantum-mechanical equivalent to the Newton's second law in classical dynamics, which describes the wave function of a quantum-mechanical system. The Schrödinger equation can be solved to obtain the wavefunction and hence the probability distribution of the particle in an environment described by a sum of potential energies. The Schrödinger equation is solvable for a single quantum-mechanical particle yielding an exact solution. The time-independent Schrödinger Equation is shown:

$$\left( -\frac{\hbar^2}{2m} \cdot \frac{d^2}{dx^2} + V(x) \right) \Psi(x) = E\Psi(x)$$

Where  $\Psi(x)$  represents the wavefunction in the single dimension  $x$ ,  $E$  represents the total energy of the system,  $V$  describes the environment via a sum of potentials. Solving the Schrödinger Equation for many-body systems yielding exact solutions is unfortunately not possible for many-body systems. In the time-independent Schrödinger Equation, a single spatial dimension  $x$  is shown, but the universe has three spatial dimensions, leading to a wavefunction  $\Psi(x, y, z)$ . Each quantum-mechanical particle added contributes with three extra dimensions, meaning the number of wavefunctions needed to describe an  $N$ -body problem is  $3^N$ . If the Schrödinger Equations were separable, then it would yield  $N$  coupled equations in 3D instead.

$$\begin{bmatrix} \Psi(x_1, y_1, z_1) \\ \Psi(x_2, y_2, z_2) \\ \vdots \\ \Psi(x_N, y_N, z_N) \end{bmatrix}$$

DFT seeks to combat the dimensionality problem, when adding particles to the system, by assuming the unrealistic case, where the electrons don't interact with each other.

Walter Kohn and Pierre Hohenberg's first theorem states<sup>36</sup>:

*For a system of electrons in their ground state, the properties are uniquely determined by the position density of the electrons. The external potential and hence the total energy is a unique functional of the electron density.*

Hence, a unique one-to-one mapping exists between the ground state energy and the electron density of a system. The ground state charge density represents a small portion of the wavefunction describing the system, so it is very fortunate, that it can be mapped to observables of interest e.g., position, momentum and in our case, energy. Obtaining the ground state charge density purely through the Schrödinger equation is not possible for many-body systems, so a trick has to be used, involving discarding the interaction between electrons. The unrealistic case

of non-interacting electrons is described by the Kohn-Sham Equation, a parallel to the Schrödinger Equation<sup>37</sup>:

$$\left( -\frac{\hbar^2}{2m} \nabla^2 + v_{eff}(r) \right) \Psi_i(r) = \epsilon_i \Psi(r)$$

Where  $\Psi_i$  is the Kohn-Sham orbital,  $\epsilon_i$  is the corresponding orbital energy,  $\nabla^2$  is the Laplace operator, and  $v_{eff}$  is the Kohn-Sham potential, the potential dictating non-interacting particles. The density for a system with N particles is:

$$\rho(r) = \sum_i^N |\Psi_i(r)|^2$$

In Kohn-Sham DFT, the following routine is used to overcome the impossibility of solving a many-body system, just for the ground state of the electrons:

- 1: Provide an initial guess on the ground state charge density
- 2: Calculate the potential  $v_{eff}$  based on the ground state charge density
- 3: Solve the Kohn-Sham equation based on the potential, resulting in a new guess on the ground state charge density and the expectation value of the energy.

The process iterates over step 2 and 3, while minimizing the expectation value of the energy, in order to find the ground state. The iteration is continued until the parameters ground state energy, potential and charge density converge. The resulting charge density should also obey the Schrödinger Equation. The ground state charge density can be estimated infinitely precisely with the Kohn-Sham equation, according to the second theorem of Walter Kohn and Pierre Hohenberg, which proves that the ground state charge density minimizes the energy functional for the system. The exact form of the exchange-correlation functional is not known, so approximations of the electronic interactions including exchange and correlation effects are used. In this work, the functional on the charge density returning an energy of the system is the RPBE functional<sup>38</sup>. The RPBE functional is based on the PBE (Perdew-Burke-Ernzerhof) functional<sup>39</sup>, which is a formulation of a generalized gradient approximation, relating to the exchange-correlation effects of electrons. The RPBE has been shown to improve chemisorption energies on metal fcc(111) surfaces, which is exactly what this work will be using it for. A detailed description of the GPAW implementation used in this work and DFT calculations on periodic systems is in the supplementary information<sup>40</sup>.

### **Practical use in catalysis**

In practicality, the input of a DFT calculation is an initial structure with initial guess atomic positions and the output is a relaxed structure and associated parameters such as total energy of the system. The initial structure goes through a “geometry optimization” process, where the structure is “relaxed”, meaning the positions of the atoms are moved to achieve a local minimum energy. During the iterative process of relaxing the structure, forces on each atom is approximated as the change in total energy as a function of position, practically carried out with the finite difference method with respect to the total energy change due to small atomic displacements. The forces are used to adjust the atomic positions with the goal of minimizing the forces, as this minimizes the total energy of the system, “relaxing” the structure. In reality, slabs and adsorbates go through the real process of being relaxed by forces due to electronic interactions, hence the simulated replicas should ideally achieve the same relaxed geometries in order to provide accurate total energies, that are accurate enough to approximate real life. The structure relaxation is done iteratively, until a convergence criterium is met, to ensure a low-energy relaxed structure. The convergence criterium is a maximum allowed force in eV/Å. The force is described as eV/Å instead of N, as it reflects the change in total energy when moving the atom in real space. A convergence criterium could be 0.1 eV/Å. The plane-wave cutoff is positioned at 400 eV. The binding energy calculations are sensitive to energies, and errors here compound through the data treatment, since the methods used rely solely on the binding energies based on DFT data.

### **Single-site structures**

Single site (SS) anode catalysts for FAOR have been tested in catalytic experiments with promising results<sup>21,35</sup>. These are typically based on nanostructures such as nanobars or nanoparticles. The structure of the fcc(111) single-site structures addressed in this work is a metal A surrounded by six metal B in the top layer. The lower layers are not considered and in the simulated surfaces are randomly distributed in the same molar fraction as the top layer. The central element A is an element that binds H and COOH relatively strongly (Pt and Pd) and element B is an element that binds less strongly (Au, Ag and Cu)<sup>18</sup>. A high efficiency requires a low anode potential, which can be achieved when COOH binds strongly to the catalyst, adsorbing at low potentials. This low potential unfortunately coincides with adsorbed H, making CO-poisoning a problem. From this it seems that high FAOR efficiency is guarded by scaling relations. Scaling relations hypothesize, that the binding energy of COOH and H of scale linearly across different metals. The optimal catalyst for FAOR binds COOH at -0.17 V vs RHE and only

binds H at lower potentials. This is not found in any pure element catalysts, which is in line with the prediction of the scaling relations. While the single-site anode catalysts examined experimentally relied on nanostructures, this work seeks to achieve a similar effect with fcc(111) single-site structures. This would rely on on-top sites primarily influenced by the element in the on-top site, that ensured a good binding energy of COOH, while less strongly binding surrounding elements could potentially weaken the binding of H in the neighbouring sites, seeking to exploit the fact that the characteristics of hollow sites are highly dependent on three atoms. The ideal behaviour of single-sites would be defying the scaling relations and provide active sites for FAOR that avoids CO poisoning, allowing the anode catalyst to run at low and efficient potentials. Achieving the highest possible occurrence of the fcc(111) single-sites is of interest. Arranging the atoms on the surface in a specific pattern is not possible in a HEA, the only parameters are the molar fractions of the alloy. A proof of which molar fraction statistically gives the most single-sites is shown in the supplementary information.

## Machine Learning

The key to HEA catalyst discovery is reasonably accurate predictions on the binding energy of a species on a site based on the local environment, as binding energies can be used as descriptors for activity for various reactions, which can then be the target of an optimization of the molar fractions of a HEA. The specific interactions between the electronic configurations of the atoms in the local environment around a binding site on a surface and the geometry that determines the binding energy are far too complicated to model in an exact closed-form mathematical expression. In machine learning terms we have a function  $f$ , that takes an input, for example all positions of all atoms in the surface and the adsorbate and returns the binding energy as a real number.

$$\Delta E_{binding,i} = f(X_i) \text{ where } \Delta E_{binding,i} \in \mathbb{R}$$

Where  $X_i$  represents all information about the surface and the adsorbate at site  $i$ ,  $\Delta E_{binding,i}$  is the binding energy of the adsorbate to site  $i$  belonging to the real numbers and the function  $f$  represents the complex real-world quantum-interactions between the metals in the surface and the adsorbate, which result in what we refer to as a binding energy. Modelling that function  $f$  is a regression problem, which is characterized by predicting a continuous real-numbered output, in contrast to classification problems, where the task is discerning patterns in data to categorize instances into predefined classes. The most precise method of estimating binding energies is DFT calculations, which model some of the quantum-mechanical behaviour between the

elements in the catalyst and the adsorbing species. In our case DFT data is computationally expensive, so we only make a limited number of “true” labels, energies assumed to be accurate, corresponding to simulated surface slabs with DFT calculations and use a simpler regression model to extrapolate the patterns in the data to arbitrary sites. Because of the complexity of the sites, with the number of metals and atom positions in the surface that has a contribution to the binding energy, making DFT calculations on all possible slabs with five metals is not possible, the composition space is simply too large. Extrapolation with a simpler regression model that predicts the DFT binding energy is needed to achieve rapid binding energy predictions for HEA catalyst tuning. The regression models utilized in this work are XGBoost Regressors, Gradient-Boosted Decision Trees<sup>41</sup> from the XGBoost Python package<sup>42,43</sup>. A decision tree can be imagined as a tree structure, where each internal node represents a decision based on one of the input features and each leaf node represents a predicted binding energy. In the training of a gradient-boosted decision tree, the model improves by iteratively constructing new trees to the ensemble based on the residuals from the previous trees, capturing patterns not addressed yet. The weight of each decision tree in the ensemble is based on the trees ability to reduce the residuals in the training set. Regularization of the model is implemented by limiting the complexity of the trees for example through a maximum depth or the minimum child weight, in order to prevent overfitting and ensure the generalization capabilities of the model. XGBoost is a great model for tabular data, as well as extremely quick at training and predicting compared to other regression models such as neural networks, achieving 2.1 million predicted binding energies per second on a desktop M1 CPU, making it a great choice for this work, where high throughput can be useful when testing huge amounts of molar fractions. In order to train a model on local surface environments, a suitable data format for the input is needed. This work will use the feature vectors based on surface slabs by Pedersen et. al. as shown in figure S3 in the supplementary information<sup>44</sup> with a small modification. On-top sites are described by a feature vector of length 20, that counts the number of each of the five metals in three selections of sites, including the six surrounding atoms in the top layer, three in the second layer and three in the third layer. The small modification is adding an initial number in the feature vector indicating the adsorbing atom in the on-top site. The feature vector for hollow sites follows a similar logic with five zones in the local environment in the first and second layer, and a one-hot encoding of all possible three-metal combinations at the three metals at the hollow site, resulting in a feature vector of length 55. This does not include all information about the surface and the specific site, but it does provide some important information, and in this work contributes to adequately accurate models.

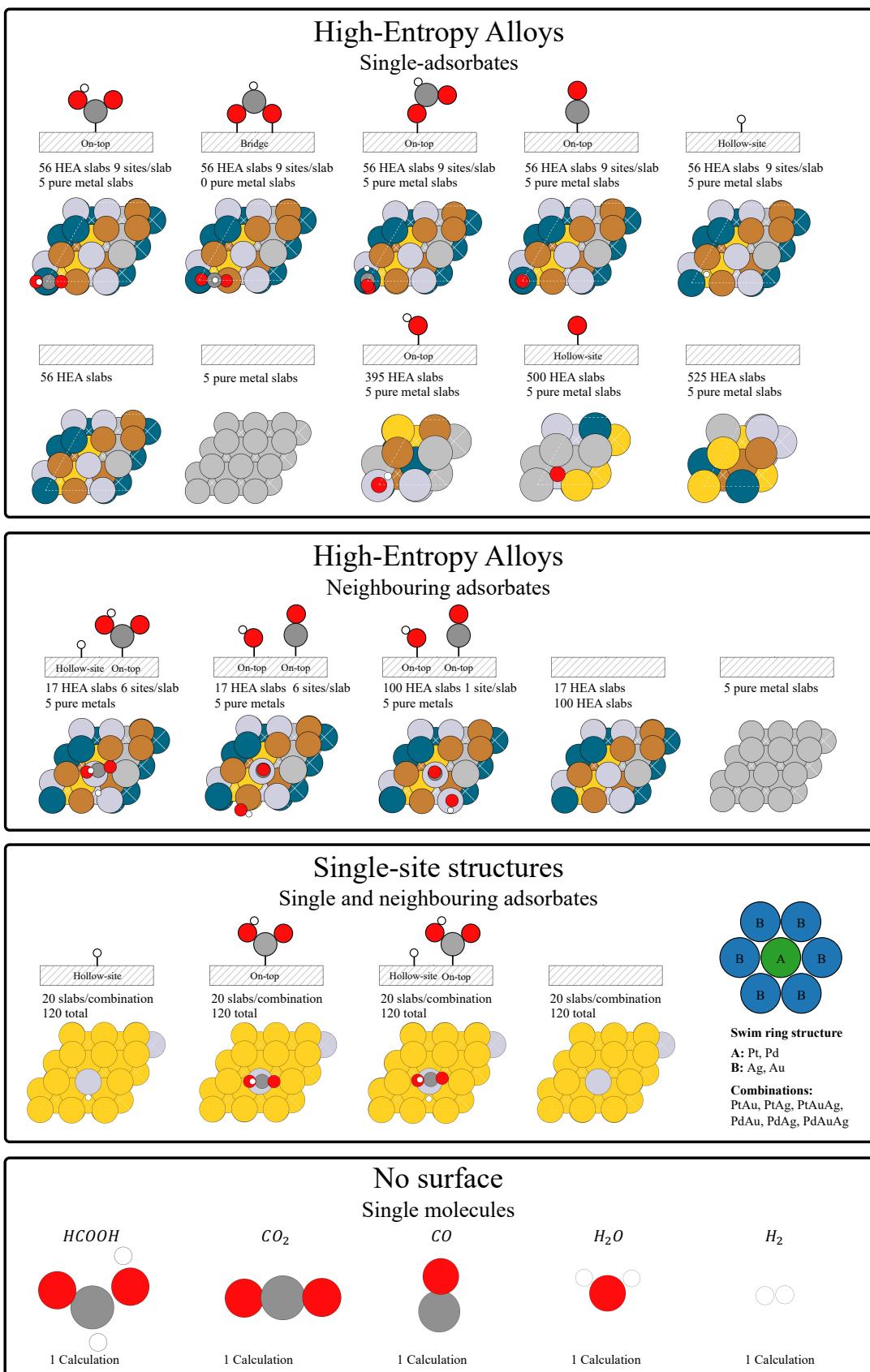
The quality and accuracy of the model can be verified and estimated by splitting up the total dataset into a training set, which is the only datapoints the model is “fitted” to, and a validation set used under training to ensure the model is not overfitting, and a test set, which the model tries to predict after training. The result on the test set is the best estimate of the accuracy of the model. More precise estimates on the accuracy of the model can be achieved by including a larger portion of the total dataset in the test set, but this also reduces the entropy in the training set, which means the model has seen fewer datapoints and will have a diminished predictive power. Ideally the model will have seen datapoints from every point in the catalyst composition space, but this is unfortunately unattainable due to computing times of making data points from DFT calculations on surface slabs.

## Methods

### Data

#### Technical details

The Density-Functional Theory (DFT) calculations are carried out with the GPAW python implementation<sup>40</sup>. A range of slabs are made with different compositions and the slabs are saved with and without the specific adsorbate. In order to isolate the energy from the adsorption, having the energy from the slab and adsorbate, the pure slab, and the molecule isolated is necessary. The energy calculations must be performed with the same functional to eliminate differences between different DFT implementations. The DFT routine calculates forces on each atom and moves them iteratively until the forces are below 0.1 eV/Å, and the electronic energy of the relaxed structure is reported. In the case of HEAs, the composition space is too large to use DFT to calculate the binding energy of an adsorbate on all possible slabs, so a compromise is made between computation time, accuracy of the calculations and entropy of the data sets. All DFT data including surface slabs with and without adsorbates, as well as single molecules are showed in figure 6. The full dataset includes HEA slabs, with all relevant adsorbates, HEA slabs with neighbouring adsorbates H+COOH and CO+OH, and single-site structured slabs with H and COOH separately and together on neighbouring on-top and hollow sites, as well as single molecules.



**Figure 6 - All DFT data including HEA slabs with single adsorbates, HEA slabs with neighbouring adsorbates, single-site structured slabs as well as single molecules.**

### **High-Entropy Alloy slabs with single adsorbates**

The HEA slabs and sample all five metals randomly at each position. 56 slabs with the adsorbate in 9 positions per slab was simulated, resulting in 504 data point. The purpose of the HEA slabs is to provide a dataset with high entropy, that encapsulates as much information as possible, that the model can extrapolate from.

### **High-Entropy Alloy slabs with neighbour-adsorbates**

HEA slabs with two adsorbates in adjacent sites are made to provide information about how the presence of a neighbouring adsorbate affects the binding energy of an adsorbate. This is relevant in the disproportionation reaction between H + COOH. A repelling effect between H and COOH on neighbouring sites could potentially discourage H from adsorbing next to COOH at slightly lower potentials than predicted by single-adsorbate models.

### **Single-site slabs with single and neighbouring adsorbates**

The single-site structured slabs contribute with training data specifically on single-site structures, which are unlikely to be present in the random HEA slabs, and there is no guarantee, that this information can be extrapolated from HEA slabs. These slabs are essential in order to be able to trust the results based on binding energy prediction models tested on simulated surfaces with single-site structures. The molar fraction of the central atom in the single-site, metal A is 1/7 in the bottom layers, and 6/7 of metal B, and a single A metal in the top layer, making the top layer 1/9 A metal. The strongly binding elements chosen as central A metals are Pt and Pd

B metals are Au and Ag. This gives rise to the following molar fractions in the bottom layers:

$Pt_1Au_6\frac{1}{7}\frac{6}{7}$	$Pt_1Ag_6\frac{1}{7}\frac{6}{7}$	$Pt_1Ag_3Au_3\frac{1}{7}\frac{6}{7}\frac{1}{7}$	$Pd_1Au_6\frac{1}{7}\frac{6}{7}$	$Pd_1Ag_6\frac{1}{7}\frac{6}{7}$	$Pd_1Ag_3Au_3\frac{1}{7}\frac{6}{7}\frac{1}{7}$
----------------------------------	----------------------------------	---	----------------------------------	----------------------------------	---

### **Single molecules**

In order to compare energies between slabs, slabs with adsorbates and molecules the energies must be calculated from the same DFT calculations, as the energies can vary between functionals. The energies of single molecules are calculated with DFT, as they are required for the calculation of some binding energies.

### **Setting reference energies**

The adsorption reaction of formic acid is based on formic acid and hydrogen as reference energies. From DFT calculations, made in collaboration with Jack Kirk Pedersen of formic acid and hydrogen separately, as single molecules, their electronic energies have been calculated to:

Species	$HCOOH$	$H_2$	$\frac{1}{2}H_2 - HCOOH$
$E_{DFT}$ (eV)	-25.7548	-6.6787	22.4154

The adsorption reaction of formic acid to bound formate is:



Hence the equation for the electronic binding energy is:

$$\Delta E_{^*COOH} = E_{^*COOH} + \frac{1}{2}E_{H_2(g)} - E_{HCOOH} - E_* - eU$$

The potential-independent, or at 0 V, electronic binding energy has the equation:

$$\Delta E_{^*COOH} = E_{^*COOH} + \frac{1}{2}E_{H_2(g)} - E_{HCOOH} - E_*$$

With the electronic energies from formic acid and hydrogen:

$$\Delta E_{^*COOH} = E_{^*COOH} - E_* + 22.4154 \text{ eV}$$

The correction from electronic energies to Gibbs free energies is based on thermal corrections calculated by Chan et. al., as shown in the table<sup>25</sup>:

Species/Thermal	ZPE	CpdT	TS
${}^*COOH$	0.62	0.10	0.19
$H_2$	0.28	0.09	0.40
$HCOOH$	0.90	0.11	0.99
${}^*H$	0.23	0.01	0.01

Furthermore, Chan et. al. notes, that they have approximated a solvation correction of the effect of solvation for species with C-OH, such as COOH and HCOOH, to -0.25 eV<sup>25,45</sup>.

$$\Delta E_{sol}(COOH) = -0.25 \text{ eV}$$

Furthermore, a functional correction of +0.09 eV for  $H_2$  is said to improve the energetics of gas-phase species<sup>25,46</sup>.

$$\Delta E_{gas-phase}(H_2) = +0.09 \text{ eV}$$

Both additional corrections are used in the correction from electronic energies to Gibbs free energies and highlighted in bold text.

The total correction factors are calculated with the formula for thermal corrections based on all species in its adsorption reaction:

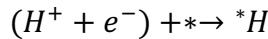
$$C_{correction} = ZPE + \int C_p dT - TS$$

$$\begin{aligned}
COOH_{corr} [eV] &= (0.62 + 0.10 - 0.19 - \mathbf{0.25}) + \frac{1}{2}(0.28 + 0.09 - 0.40 + \mathbf{0.09}) \\
&\quad - (0.90 + 0.11 - 0.99) = 0.29
\end{aligned}$$

This yields a correction:

$$\Delta G^*_{COOH} = \Delta E^*_{COOH} + 0.29 \text{ eV}$$

The adsorption reaction for H is based on  $H_2$  as a reference energy:



The equation for the electronic binding energy is:

$$\Delta E^*_H = E^*_H - E_* - \frac{1}{2}E_{H_2(g)} + eU$$

The potential-independent electronic binding energy has the equation:

$$\Delta E^*_H = E^*_H - E_* - \frac{1}{2}E_{H_2(g)}$$

The correction from electronic binding energies to Gibbs free energy is calculated with the thermal corrections from Chan et. al. and the additional gas-phase correction:

$$H_{corr} [eV] = (0.23 + 0.01 - 0.01) - \frac{1}{2}(0.28 + 0.09 - 0.40 + \mathbf{0.09}) = 0.20$$

This yields a correction:

$$\Delta G^*_H = \Delta E^*_H + 0.20 \text{ eV}$$

To evaluate the legitimacy of these thermal corrections, we look at the behaviour the Gibbs free energy dictates. We have a strong suspicion that the mechanism behind the CO-poisoning is a disproportionation reaction, that should happen at small potentials close to 0 V, where both COOH and H is adsorbed. The thermal corrections need to allow this by dictating an overlap region, where both H and COOH has negative Gibbs free adsorption energies at the same potential, in order to predict the behaviour observed in real life electrochemical experiments. An example is Bagger et. al., where CO-poisoning is seen only after visiting near-zero potentials<sup>18</sup>. In the same article in their figure 4a, the FAOR activity on a pure platinum (Pt/C) catalyst is seen increasing around 0.25 V vs RHE. With the mid-point of the steep increase at around 0.35 V. With the thermal corrections at hand, calculated from the correction terms calculated by Chan et. al., the current model dictates the following behaviour on a simulated equimolar HEA surface with equimolar fractions. Up to 0.11 V only H has negative binding energies. From 0.11 V to 0.30 V, both COOH and H have some sites with negative binding energies, opening the possibility for CO-poisoning. From 0.30 V and up, only COOH binds. The minimum visitable voltage, which will not lead to CO-poisoning of the surface, as seen in an activity decrease, is

not measured exactly in the article by Bagger et. al. The potential 0.4 V is visited without seeing a decreased activity when returning to higher potentials, and the potential 0.025 V is visited with highly decreased activity. In itself this seems reasonable, but the behaviour predicted from the binding energy models and thermal corrections, as well as the solvation correction and functional correction, doesn't align exactly with the observed behaviour from experiments on Pt/C catalyst in the article by Bagger et. al. The pure Pt surface DFT binding energies of COOH and H with the thermal corrections predict that H will desorb at a lower potential than COOH will adsorb at. The electronic binding energies of H and COOH on pure Pt, at no potential, as calculated by DFT are:

$$\Delta E_{*H}^{Pt} = -0.32 \text{ eV}$$

$$\Delta E_{*COOH}^{Pt} = -0.025 \text{ eV}$$

This means no overlap between H and COOH at any potential, and hence no CO poisoning. The middle point where both H and COOH are the same energy from  $\Delta G_{ads} = 0$  is 0.1969 V. This work is based on the assumption, that the behaviour observed by the experiments in Bagger et. al. is due to CO-poisoning, which is also the broad consensus in the literature<sup>18</sup>. It is then concluded that the corrections are a little bit off and should be corrected to align with experimental results. Both correction terms for H and COOH are adjusted with the same numeric value and should have the effect that H desorbs at slightly higher voltages and COOH adsorbs at slightly lower voltages in order to achieve the assumed overlap. If H and COOH should ad- and desorb at exactly the same voltage and have an infinitesimally small overlap, and the correction to each correction is of the same numeric value, the only solution is adding an additional correction of  $-0.0676 \text{ eV}$  to both adsorbates:

$$eU_{overlap} = 0.1969 \text{ eV}$$

The binding energies with the new correction is calculated with the formula:

$$\Delta G_{ads}^{Pt} = \Delta E_{ads} + ads_{corr} - Pt_{corr} + eU$$

$$\Delta G_{*H}^{Pt} = -0.32 \text{ eV} + 0.20 \text{ eV} - \mathbf{0.0676 \text{ eV}} + 0.1969 \text{ eV} = 0 \text{ eV}$$

$$\Delta G_{*COOH}^{Pt} = -0.025 \text{ eV} + 0.29 \text{ eV} - \mathbf{0.0676 \text{ eV}} - 0.1969 \text{ eV} = 0 \text{ eV}$$

At the overlap potential, both adsorbates now barely adsorb at platinum, which the results from experiments indicate they do. In order to get a little more overlap, both adsorbates are set to bind an arbitrary factor of 0.04 eV better than just barely binding at pure platinum at the same time. This modified correction factor including the corrections from the table values and the extra corrections to align the calculations with experimental observations are then:

$$COOH_{corr}^{mod} = 0.29 \text{ eV} - 0.0676 \text{ eV} - 0.04 \text{ eV} = \mathbf{0.1824 \text{ eV}}$$

$$H_{\text{corr}}^{\text{mod}} = 0.20 \text{ eV} - 0.0676 \text{ eV} - 0.04 \text{ eV} = \mathbf{0.0924 \text{ eV}}$$

These modified (mod) corrections are used for the further work. Furthermore, the Pd/C catalyst tested in the same article shows no signs of CO-poisoning after visiting 0.025 V vs RHE. Since the predicted COOH binding energies on Pd are positive all the way up to 0.56 V, there is no voltage at which H and COOH binds to Pd at the same time, neither pure Pd calculated directly from DFT or any Pd on-top sites in an evenly mixed HEA surface, so this behaviour is reproduced by the corrections as well.

## Predicting binding energies

The binding energy prediction models are trained on the following structured data.

**Input:** Feature vectors describing the local environment, based on the structure of the surface slabs used for DFT calculations

**Target:** DFT calculated binding energies on the surface slabs

The model is judged on its ability to predict the same electronic energies that the DFT calculations predict. The same feature vectors are used for DFT slabs for the training data and for the simulated surfaces used in the search for a good HEA catalyst. The binding energies are based on the adsorption reaction for the specific adsorbate. When training the binding energy prediction models, a single model is trained for each adsorbate on both HEA DFT data and single-site structured slabs in order to maximize the predictive power and the generalization proficiency of the models on various types of sites. DFT data with adsorbates on on-top sites with different central adsorbing atoms are included in the same mode, which is not a problem with decision trees since the boosted decision-trees can model non-linear patterns in the training data. For the models trained on both HEA and single-site structured datasets, the individual datasets are split into train, validation, and test before concatenating the train, validation, and test sets from different datasets. The model should both be trained, validated, and tested on the single-site structures, and this ensures that the datapoints from the single-site dataset doesn't end up unevenly distributed between train, validation, and test. A standard distribution of 80% training data, 10% validation data and 10 % test data are used throughout this work.

In the case of the models trained to predict the combined binding energy of two adsorbates on neighbouring sites, for example H and COOH, the so-called “mixed-site” model is trained on a combined feature vector comprised of the on-top site feature vector concatenated with the hollow site feature vector. This is the simplest solution, that allows the model to learn patterns in the local environment of both adsorbates. This results in a feature vector of length 70.

$$F_{\text{Mixed-Site}} = [F_{\text{OnTop}}, F_{\text{Hollow}}]$$

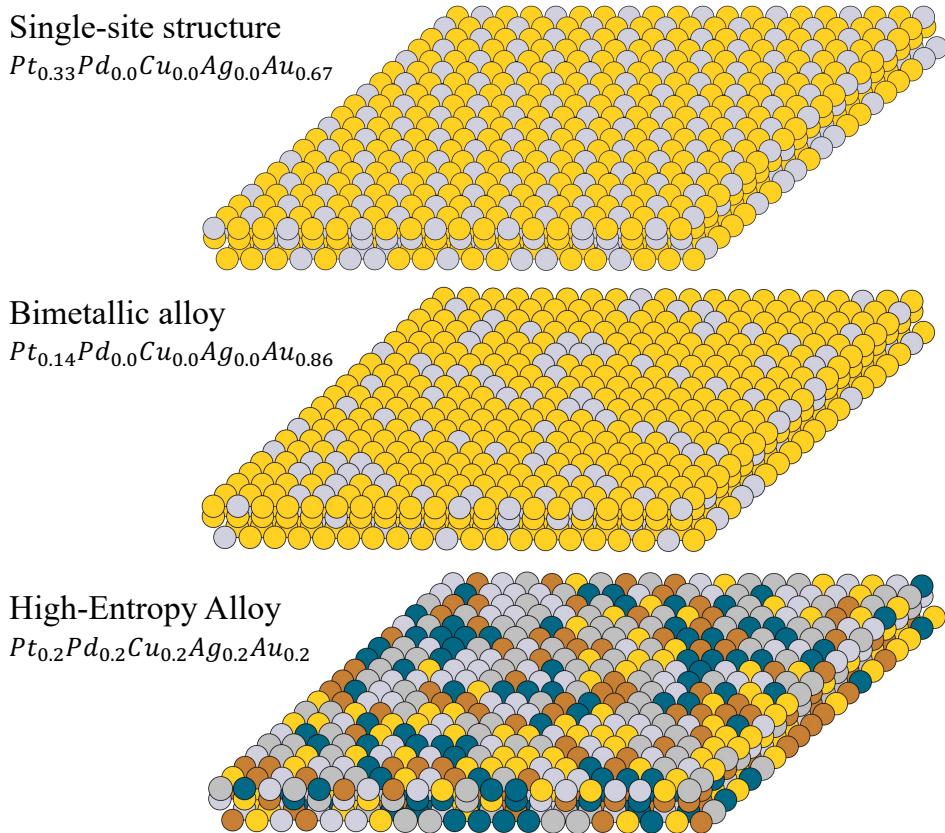
Where  $F$  represents a feature vector and mixed-site is the combined vector from an on-top site and a hollow site. All models are trained until convergence, meaning the loss on the validation data after each training round, should decrease. When the validation loss has increased for 5 rounds in a row, the model is assumed to be overtraining, and the model from the round with the lowest validation loss is saved. The model is then set to predict the electronic binding energies as estimated with DFT calculations, and the mean absolute error is used as a metric for the predictive power of the model.

## Simulating surfaces

The search for FAOR catalysts is based on simulated surfaces. The surfaces can be initialized with arbitrary dimensions and molar fractions of HEAs and bimetals, as well as specially ordered structures, such as closely packed single-site structures as shown in figure 7. The single-site structure shown is created by ordering the top layer, such that each row is comprised of two weakly binding elements and one strongly binding, repeating. The other rows are similar but offset by one towards the right side. This creates the most single-site fcc(111) structures possible, at one third of the on-top sites. The two bottom layers are of the same molar fractions as the top layer, but randomly distributed. This structure is not likely to exist in real life, as it would be completely unlikely with random mixing. The bimetallic alloy shown underneath has the molar fractions statistically most likely to lead to the highest number of single-site structures with random mixing.

## Simulating surfaces

Based on stoichiometry, structure and dimensions



**Figure 7 - Pt(111) surfaces are simulated based on stoichiometry, structure and dimensions**

The binding energy prediction models can predict the binding energy of arbitrary sites by extrapolating from the sites it has trained on from the DFT dataset. This ability is used in order to evaluate the characteristics of a suggested catalyst surface. A pipeline is prepared, in which a stoichiometry and a structure (random mixture or single-site structure) is used to simulate a surface, feature vectors are prepared based on the positions of the atoms in the surface and the models predict binding energies. The XGBoost models are able to predict the binding energy of 2.1 million sites per second yielding the ability to evaluate the binding energies of large surfaces rapidly. In order to determine the characteristics of a catalyst, various descriptors of performance can be evaluated based on the predicted binding energies of the various adsorbates.

## Searching for optimal compositions

The composition space of fine-grained molar fractions with five elements is enormous, so a structured way of exploring it is needed. Though the composition space is big, some rational choices can be employed to limit the amount of search space where good catalysts can be found.

For example, the composition-space where neither Pt or Pd is present will not be FAOR active until extremely big overpotentials, because Cu, Ag and Au bind COOH weakly.

The search for a good anode catalyst for FAOR should be based on our knowledge about the specific reaction and the limiting factor, the CO-poisoning, as it is regarded as the main hindrance for highly efficient FAOR. The overlap between adsorbed H and adsorbed COOH at low potentials is thought to be the source of the CO-poisoning, so an efficient catalyst should allow COOH to adsorb and be active at low potentials, while avoiding neighbouring H adsorbates. This is the range, where an efficient catalyst is hoped to be found. There are several different parameters to optimize in this scenario. The efficiency describes the energy harvestable from each proton-electron pair produced, which scales with the open-circuit current and increases towards smaller anode potentials. Activity, which is the number of oxidized formic acids per site per time unit, and activity is unfortunately opposed to efficiency because an overpotential helps more Pt sites bind COOH and lowers the energy of the end products  $CO_2(g) + 2(H^+ + e^-)$ , increasing the activity. Power is another important parameter, which is the result of the per-site activity, describing how many electrons are produced per site per time unit, and the open-circuit current, which describes the work each electron produced can do. The product of those two represent a per-site power measure. Depending on the application for the fuel cell, the main priority may be efficiency or power, which will have two different optimal anode catalysts. In order to find both efficient and catalysts with high power output, molar fractions are tested at different anode potentials from 0.10 V to 0.35 V vs RHE. The reasoning for the lower bound is, that 0.10 V is the lowest potential with any kind of activity on a HEA, meaning lower potentials will barely be able to facilitate FAOR. The upper bound is set to 0.35 V because a pure Pt catalyst can facilitate FAOR without experiencing CO-poisoning at and above that potential. Hence, the interesting area, where possibly new promising HEA catalysts can be found. Several “optimal” catalysts can then be imagined for different potentials. A very high-powered anode catalyst at an overpotential. For example, a very efficient anode catalyst with a great open-circuit potential, but a very low activity and hence power.

### ***Brute-force search***

A brute-force approach is applied, where all molar fractions in 5% increments are used in a pipeline, that simulates a HEA surface with dimensions 100 by 100 by 3, the binding energies of H and COOH on all sites are predicted, and a descriptor of activity is calculated. This is done for all potentials between 0.10 V and 0.35 V with a spacing of 0.01 V. The number of stoichiometries that exist for a number of elements and a step size are shown in the table below:

Molar fractions/elements	5	4	3
1 %	4598126	176851	5151
5 %	10626	1771	231
10 %	1001	286	66

This work will pick out a few potentials, where some reasonable tradeoffs between efficiency and power are made. Then, a more finely grained brute-force search can be made at the specific potential with the elements expected to be present in the optimal molar fraction, which could be less than 5. A Bayesian search with Gaussian processes could also be used and have been used for catalyst discovery<sup>44</sup>, but is not chosen in this work, as brute-force is actually feasible with 5 elements and 5% step size in the molar fractions, and finer-grained brute-force searches can be carried out at a few select potentials. Furthermore, brute-force search is useful because the results can be used to fill ternary plots, which can reveal trends in the efficiency/activity as the composition changes slightly. This could uncover or confirm hypothesized trends in the composition space.

### Optimization criteria

The goal of the descriptor used as an optimization criterion is, that it models the activity of the two-step FAOR pathway accurately while also accounting for CO-poisoning from the disproportionation reaction. The two descriptors suggested are based on only the COOH and H binding energies. A simple model for the CO-poisoning is a system, where the activity of an on-top site is assumed to be zero due to CO-poisoning if the adsorption energy of an H in one of the three neighbouring FCC hollow sites is negative. Then, the per-site current density of the two-step FAOR reaction of the remaining, un-poisoned sites could then be estimated with the following equations:

$$j_i = e^{-\frac{\Delta G_{RLS}^\dagger}{k_b \cdot T}}$$

Where the energy of the rate-limiting step is:

$$\Delta G_{RLS}^\dagger = |\Delta G_{COOH}^* + 0.17 \text{ eV}| + 0.17 \text{ eV} - e \cdot U$$

Where  $\Delta G_{COOH}^*$  is the binding energy of the COOH intermediate at the potential and  $U$  represents the anode potential vs RHE. At the optimal binding energy and the optimal potential, the energy of the rate-limiting step will be zero, which leads to the largest activity of 1, in arbitrary units.

The Koutecký-Levich equation can be used to correct for mass transport limitations:

$$\frac{1}{j_{site}} = \frac{1}{j_i} + \frac{1}{j_D} \Rightarrow j_{site} = \frac{1}{\frac{1}{j_i} + \frac{1}{j_D}} \Rightarrow j = \frac{1}{N} \sum_i^N \frac{1}{\frac{1}{j_i} + \frac{1}{j_D}}$$

Where  $j_i$  is the current density estimated from the binding energy and potential and  $j_D$  is the diffusion-limited current density. Another descriptor is the simpler measure of counting how many sites are active and un-poisoned. An on-top site should be active, when COOH binds at the given potential. This is guaranteed by the potential-dependent part of the expressions for the Gibbs free energies of  $^{*}COOH + (H^+ + e^-)$  and  $CO_2(g) + 2(H^+ + e^-)$ . If COOH adsorbs with a binding energy between 0 eV and the optimal value of  $-0.17$  eV, then the final reaction step to  $CO_2(g)$  will also be favoured, because it is lower in Gibbs free energy. Only if COOH doesn't bind will the energy increase from formic acid to COOH, and if COOH binds too strongly will the energy increase from COOH to  $CO_2(g)$ . Based on the thermal corrections and additional corrections, no HEA of Pt, Pd, Cu, Ag, and Au can bind COOH stronger than the optimal value, and hence, if COOH is predicted to bind, it should be active, given no CO-poisoning takes place. Based on this a simple measure of activity is made:

$$j_{site,(x,y)} = \begin{cases} 1 & : \Delta G_{^{*}COOH,(x,y)} < 0 \wedge (\Delta G_{^{*}H,(x,y)} \wedge \Delta G_{^{*}H,(x-1,y)} \wedge \Delta G_{^{*}H,(x,y-1)} > 0) \\ 0 & : \Delta G_{^{*}COOH,(x,y)} < 0 \wedge (\Delta G_{^{*}H,(x,y)} \wedge \Delta G_{^{*}H,(x-1,y)} \wedge \Delta G_{^{*}H,(x,y-1)} < 0) \\ 0 & : \Delta G_{^{*}COOH,(x,y)} > 0 \end{cases}$$

Where x and y represent the x and y-coordinates along the top layer of the surface. Because of the offset between the on-top sites and the FCC hollow sites, the on-top site with the coordinates (x, y) will be neighbour to (x, y), (x-1, y), and (x, y-1). Whereafter, the Koutecký-Levich equation and an averaging over all sites is employed to achieve a per-site current density for the whole simulated surface. This descriptor does not assume an exponential relation between binding energies and current density, but this can be justified, since the exponential relation might not hold true for this reaction, as it does for the Oxygen Reduction Reaction (ORR). Furthermore, sites where COOH bind will result in an active site, as seen in figure 4, at least on the HEAs studied in this work. With the exponential relation, there is very little impact from CO-poisoning to be seen on the current density because it happens at low potentials. At low potentials, the exponential relation always predicts very low current densities, because the potential in the exponent scales faster than the number of active, un-poisoned sites, which is the most important metric to measure, when designing CO-poisoning resistant catalysts.

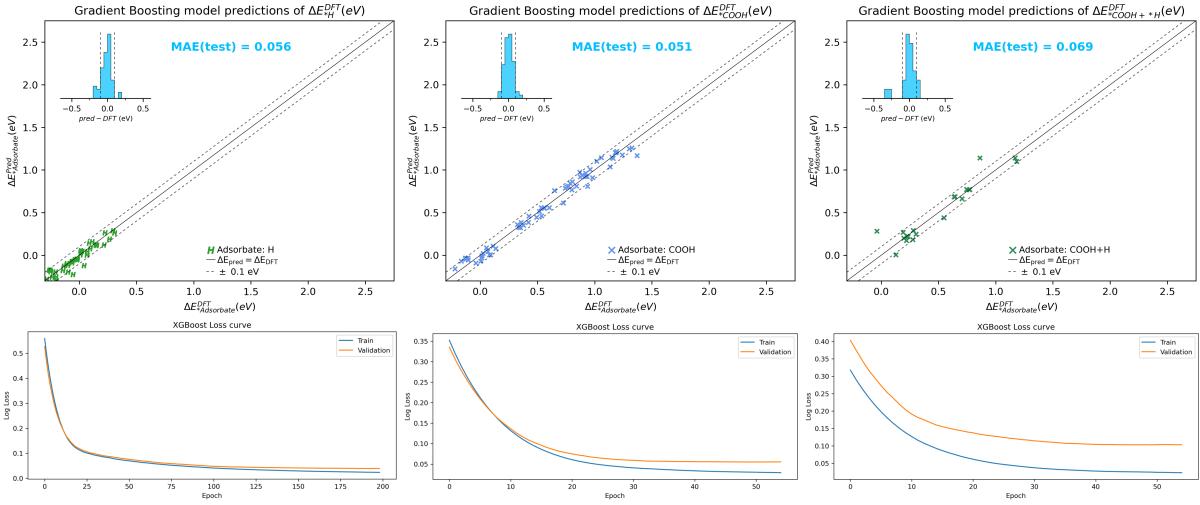
# Results

## Training binding energy models

The H, COOH and COOH+H binding energy models were trained on a combination of high-entropy alloy slabs with and without adsorbates and single-site structured slabs with and without adsorbates, as described in methods. In the sets of slabs with the H adsorbate, 490 of the 504 HEA slab DFT calculations converged, while staying in the expected site after structure relaxation, while all 120 single-site structured slabs, and all slabs with the COOH adsorbate, both single-site structured and HEA also converged. In the case of the *mixed* model that predicts the binding energy of both adsorbates, H and COOH, the features based on the hollow-site H adsorbate, were individually concatenated with the features based on the neighbouring on-top COOH adsorbate from the same slab. As shown in the table underneath, 119 of the 120 single-site structured slabs converged during DFT structure relaxation, while keeping H and COOH in the expected site, while 96 of the 102 high-entropy alloy slabs converged. In most cases, the reason for this is the H adsorbate not being in the expected hollow site position.

Mixed-model data	High-Entropy Alloy	Single-site	Total
Train	76	95	171
Validation	10	12	22
Test	10	12	22
Total	96 (Out of 102)	119 (Out of 120)	215

The models trained on two separate sets of data had both data sets split into training, validation, and test data separately, as is shown in the case of the mixed-site models in the table above, in order to guarantee the presence of both data sets in the training, validation and test set. In the case of the H, COOH and H+COOH models, the single-site data consists of 120 data points and the HEA data consists of 504 data points. Hence, just 12 data points from the single-site data set was in the test set.

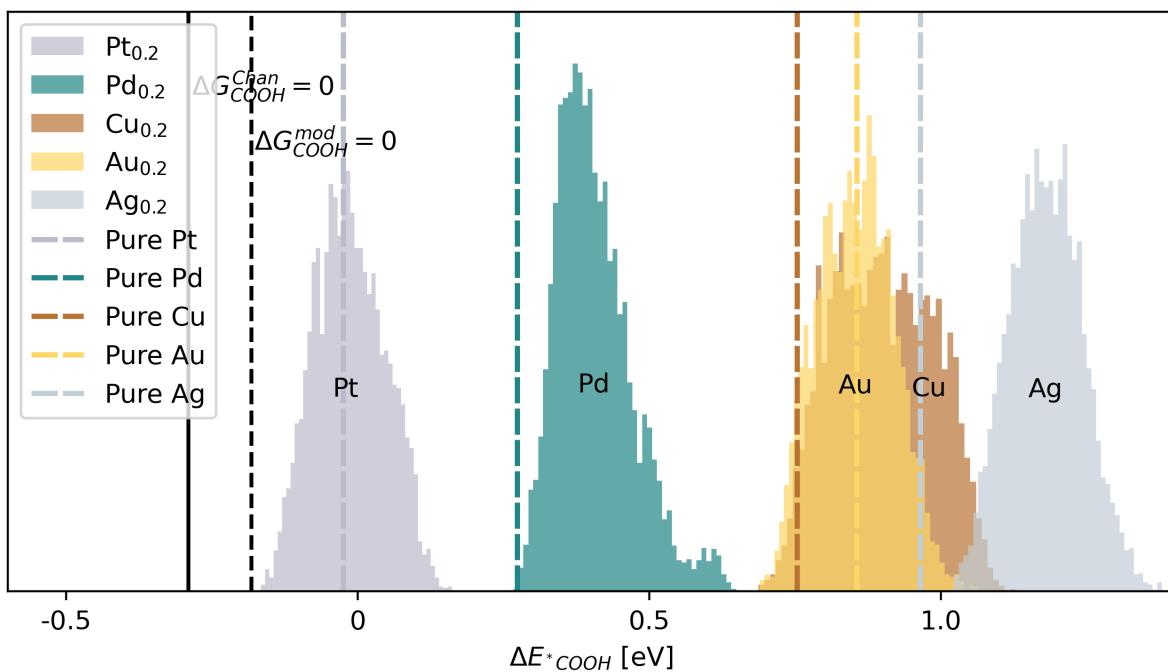


**Figure 8 – Parity plots from the binding energy prediction models for H, COOH and COOH+H on neighbouring sites on the test sets. The parity plots show the models’ predictions of DFT calculated binding energies as a function of the actual DFT calculated electronic energies. The loss curves are shown underneath.**

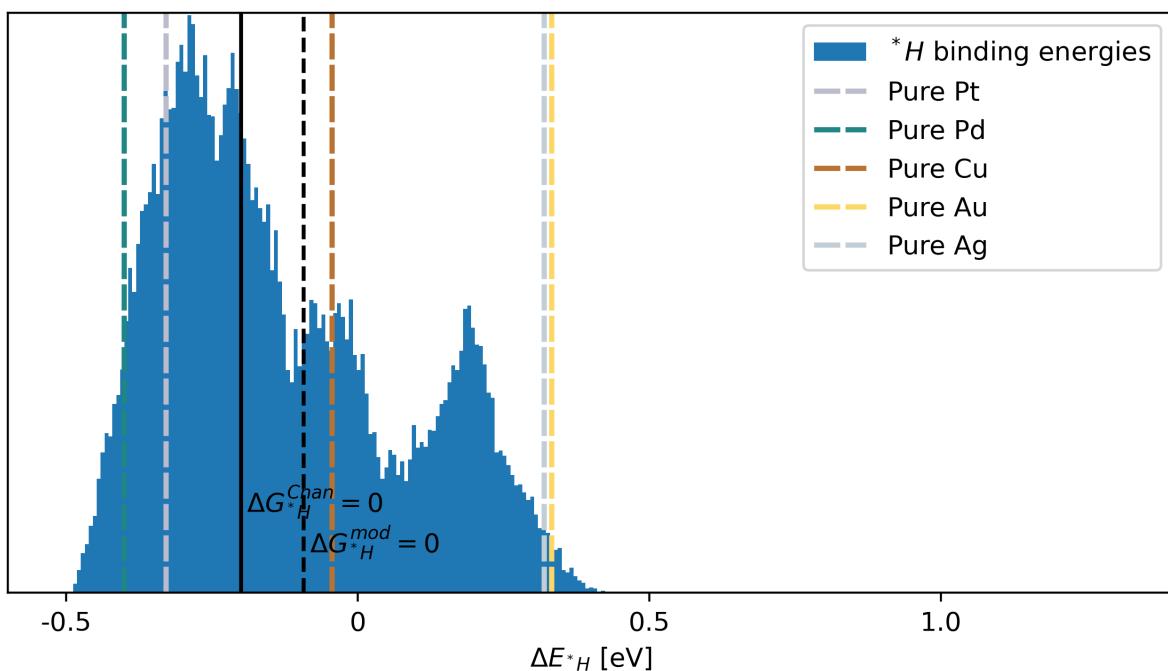
As seen in figure 8, the mean absolute errors (MAE) of the separate H and COOH models are 0.056 eV and 0.051 eV, respectively, which is useable for the further applications. The loss curves are smooth, and the training converges. For the H+COOH model, the validation loss is significantly higher than the training loss, which combined with the higher MAE of 0.069 on the test set indicates, that it is a harder regression task, than the single-adsorbate scenarios, or that it needs more data to achieve similar performance.

### Applying the models on simulated surfaces

In figure 9, the predicted binding energies of COOH adsorbed to on-top sites on an equimolar HEA simulated surface are shown. The pure metal energies are marked with dashed lines. The separation of the predicted energies on different on-top site metals is as expected, following the order of the pure metal DFT calculated binding energies, except for Au and Cu, where Au binds better on the HEA on average than Cu. The mean binding energy of COOH on Pt-sites and Au sites on the HEA surface align perfectly with the pure metal DFT calculated energy, while Pd, Cu and Ag on average bind worse on the HEA than in the pure metal DFT slabs. This might be due to perturbations of the other elements in the HEA. As Pt-sites are the only viable sites for FAOR at low, efficient, anode potentials, the accuracy of these is the most important.

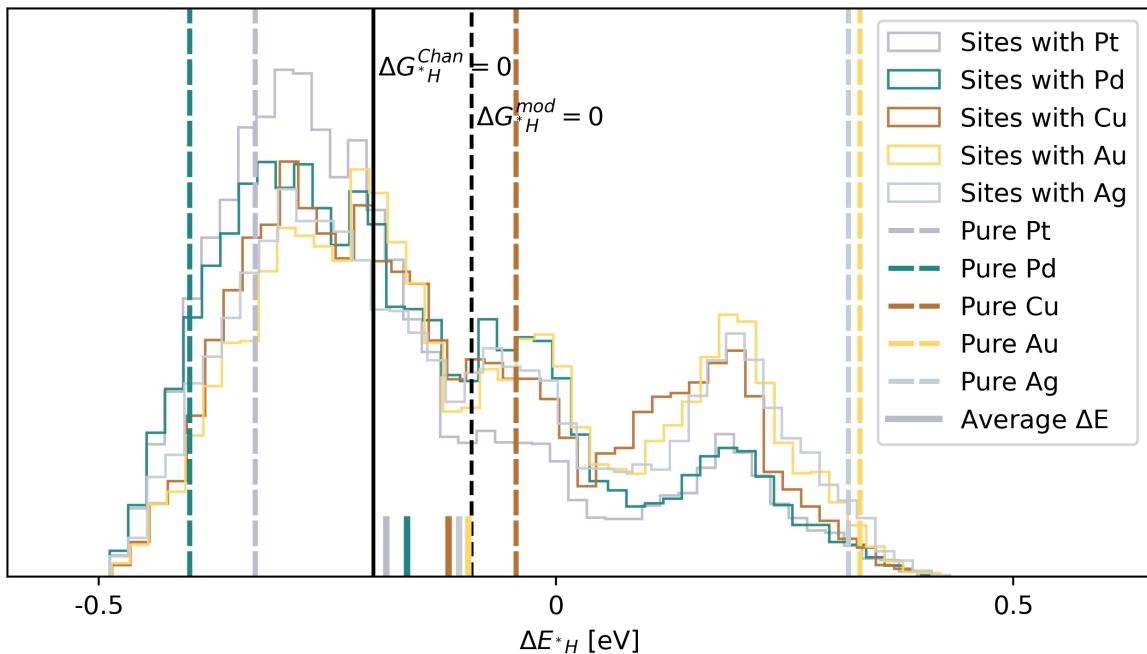


**Figure 9** - The predicted formate (COOH) binding energies on a simulated 200 by 200 by 3 equimolar HEA surface. The energies are separated by the metal on which the on-top site is. The single element binding energies directly from DFT calculations are marked with dashed lines. Black lines show the thermal correction for COOH based on calculations by Chan. et. al. (solid) and the thermal correction equal to the previous but adjusted to fit an electrocatalytic experiment (dashed)<sup>18,25</sup>, as described in methods.



**Figure 10** - The predicted H binding energies on a simulated 200 by 200 by 3 equimolar HEA surface. The single element binding energies directly from DFT calculations are marked with dashed lines. Black lines show the thermal correction for COOH based on calculations by Chan. et. al. (solid) and the thermal correction equal to the previous but adjusted to fit an electrocatalytic experiment (dashed)<sup>18,25</sup>.

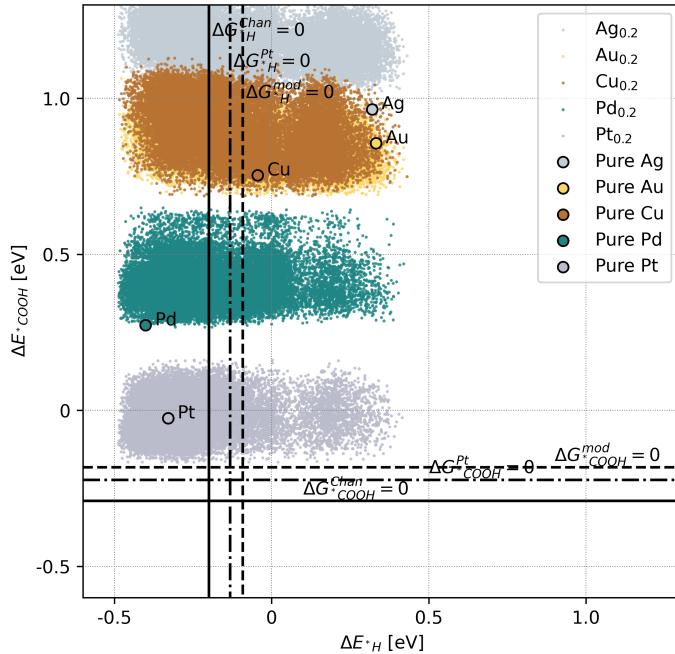
In figure 10, the predicted binding energies of H in all FCC hollow sites of an equimolar HEA simulated surface are shown. The predicted binding energies of H have a larger overlap between the peaks compared to COOH, due to the fact that hollow sites have three atoms in the top layer of the adsorbing surface, which have the largest influence on the milieu at the adsorbing site, as compared to one most influential atom at an on-top site, which, as seen in figure 9, clearly separates the binding energies. In order to investigate the influence of each of the five metals, a filter is made, where the binding energy of each H adsorbing to a hollow site is sorted into a list corresponding to each of the three metals present at the specific hollow site. This results in each binding energy being present thrice in total, distributed in one to three lists. In figure 11, the distributions can be seen, where each histogram entry represents all energies where a specific metal is present at the adsorbing site.



**Figure 11 – The predicted H binding energies on a simulated 200 by 200 by 3 equimolar HEA surface. In this plot, every binding energy of H is represented thrice, once for each of the three atoms in the FCC hollow site. The mean binding energy of all binding energies in sites where each metal is present in the three nearest atoms is shown with a short solid line. The single element binding energies directly from DFT calculations are marked with dashed lines, and the thermal corrections based on calculated values from Chan. et. al. (solid) and a thermal correction modified to fit experimental data (dashed) is shown<sup>18,25</sup>.**

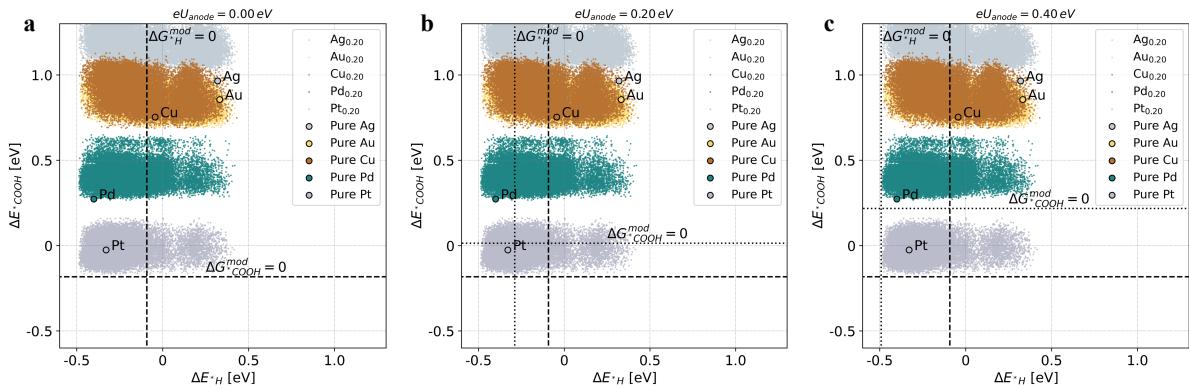
The distributions of H binding energies given the presence of a specific metal in the adsorbing hollow sites vary a bit for each metal. Pt for example has the highest prevalence of sites with low binding energies between -0.4 and -0.2 eV. The mean binding energy of hollow sites with each metal present is marked with a short solid bar, showing Pt to be the strongest binding metal, followed by Pd, Cu, Ag and Au. This aligns with the expectation that Pt and Pd are the strongest

binding metals for COOH as well as H and Cu, Ag and Au bind less strongly for both COOH and H.



**Figure 12 – Predicted binding energies of COOH on all on-top sites of the simulated 200 by 200 by 3 equimolar HEA surface and H in all hollow sites. Every binding energy is plotted thrice, as each COOH binding energy is shown with each of its three neighbouring FCC hollow site H binding energies. The binding energies directly from DFT calculations on single-element slabs for COOH and H are shown with larger circle markers with black borders. Three thermal corrections are shown in solid, dash dotted and dashed lines.**

In figure 12, the binding energy models for COOH and H have been used to predict the COOH and H binding energies on an equimolar 200 by 200 by 3 simulated surface. Every dot represents a COOH binding energy on the y-axis and the binding energy of a neighbouring H, of which there are three. Hence, all binding energies are represented thrice, in order to show all neighbour pair energies. Three corrections from electronic energies to Gibbs free energies are shown for both COOH and H. The solid line represents the correction based on calculations from Chan. et. al., the dashdotted line represents the same correction altered slightly (**-0.0676 eV**) to make both COOH and H bind to pure platinum at the anode potential  $U = 0.1969 V$  based on experiments showing that both adsorbates bind at at least one potential<sup>25</sup>. The dashed line represents the modified correction, where COOH and H bind an arbitrary 0.04 eV better, because a small overlap of H and COOH is hypothesized to cause CO-poisoning.



**Figure 13 - Predicted binding energies of COOH on all on-top sites of the simulated 200 by 200 by 3 equimolar HEA surface and H in all hollow sites. The dashed lines represent thermal corrections modified to fit experimental results<sup>18</sup>. The dotted lines represent the same correction including the potential-term in the binding energy equation for both adsorbates, -eU for COOH and eU for H. This is shown for three different potentials (a-c).**

In figure 13, the same predicted binding energies are shown, with a dotted line representing the  $\Delta E_{ads}$  value where  $\Delta G_{ads} = 0$  for the potential-dependent adsorption reaction. For COOH, an increased potential contributes with a  $-eU$  term, moving the  $\Delta G_{COOH}^{mod} = 0$  line upwards as the potential increases. Above the line,  $HCOOH(aq) + *$  is more stable and under the line  $*COOH + (H^+ + e^-)$  is more stable. For H, an increased potential contributes with an  $eU$  term, moving the  $\Delta G_H^{mod} = 0$  to the left, as the potential increases. To the left of the line  $*H$  is more stable and to the right of the line  $(H + e^-)$  is more stable. Hence, the binding energy pairs in the left square under the cross of the dotted lines are at risk of CO-poisoning, and the right square under the cross are active non-poisoned sites.

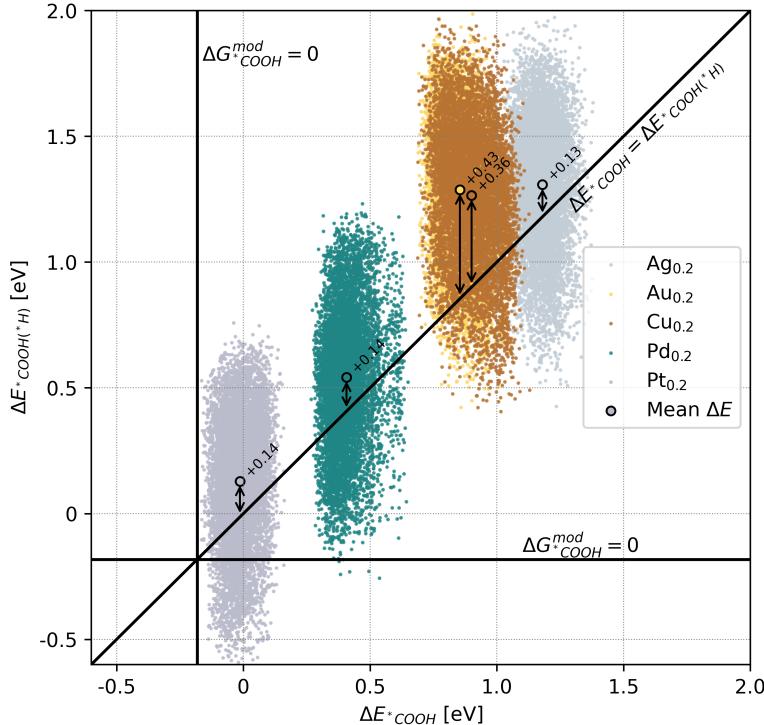
### Binding energies given neighbouring adsorbates

Two additional models are constructed by utilizing the model predicting the binding energy of both COOH and H adsorbed to neighbouring sites and the single adsorbate models. The models predict the binding energy of H on a hollow site given the presence of COOH on one of the three neighbouring on-top sites and the binding energy of COOH given the presence of H in one of the three neighbouring hollow sites. The energy contribution from the presence of a neighbouring adsorbate is assumed to be included in the mixed-site model, which is trained on DFT calculations on slabs with both adsorbates, and the binding energy of the other adsorbate, as predicted by a single-adsorbate model, is subtracted. This leaves the binding energy of the adsorbate of interest and the contribution from the presence of the neighbour-adsorbate. The equations relying on two models is shown:

$$\Delta E_{H(*COOH)}^{Predicted} = \Delta E_{H+*COOH}^{Predicted} - \Delta E_{*COOH}^{Predicted}$$

$$\Delta E_{COOH(H)}^{Predicted} = \Delta E_{H+COOH}^{Predicted} - \Delta E_H^{Predicted}$$

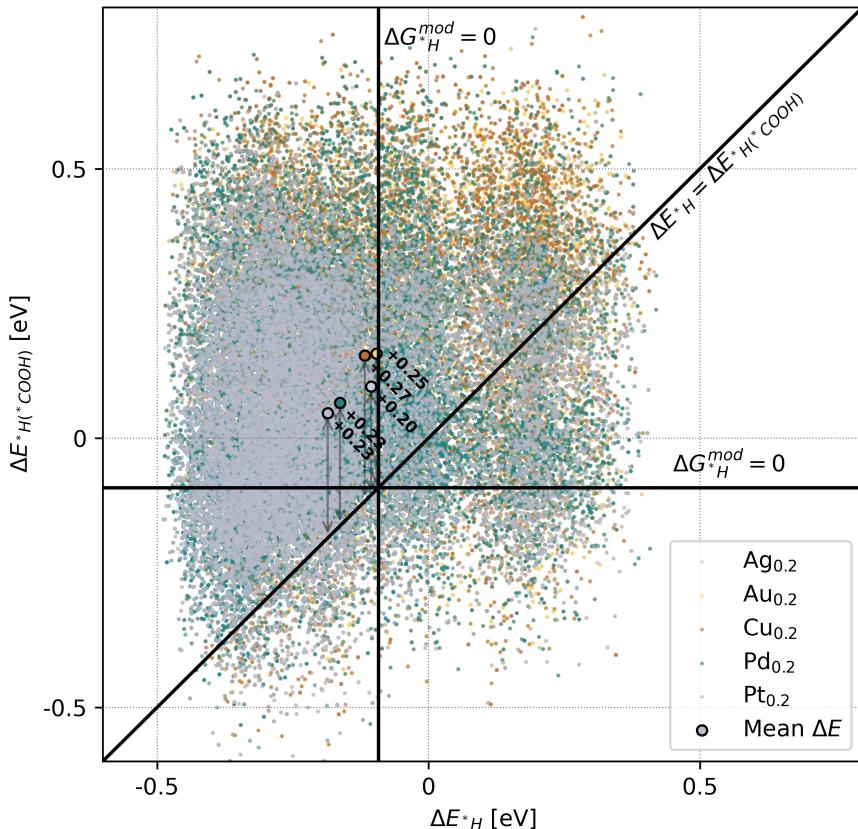
The predicted binding energies of “H given COOH” and vice versa are compared to the predictions of the binding energies from the single adsorbate models to evaluate the impact on the binding energy of an adsorbate when adsorbing to a site with and without a neighbour.



**Figure 14 -** The predicted binding energies on a simulated equimolar 200 by 200 by 3 HEA surface of COOH given the presence of H as a function of the predicted binding energies of COOH without neighbouring adsorbates. The diagonal line marks where the two binding energies are equal. The mean binding energies are marked with large circles with black borders, with the difference in mean between  $\Delta E_{COOH(H)}$  and  $\Delta E_{COOH}$  shown in text. Positive values indicate that the binding energy of COOH is higher when H sits on a neighbouring site.

In figure 14, the impact of a neighbouring H on the binding energy of COOH is investigated by using two models on the same equimolar simulated HEA surface. On the x axis, the predicted binding energies of COOH,  $\Delta E_{COOH}$ , on all on-top sites are shown, while the y axis shows the predicted binding energies of COOH given the presence of H,  $\Delta E_{COOH(H)}$ . Each binding energy is shown with the color according to the primary atom in the surface on-top site. The mean binding energies for each metal in the primary position in the on-top sites are marked with larger circles with black borders. For all metals, the mean binding energy is larger for COOH in the case where a H sits on a neighbouring site, indicating that H in a neighbouring hollow site makes the binding of COOH less favorable. For Pt, Pd and Ag, the difference is +0.14 eV, while Cu has an average difference of +0.37 eV and Au has an average difference of +0.44 eV. This indicates that a neighbouring H has a larger impact on the binding energy of COOH, when

COOH adsorbs onto an on-top site with Au and Cu than Pt, Pd, and Ag. The mean difference across all metals is +0.24 eV. The standard deviation of the binding energies in the DFT data for the mixed-site model with DFT slabs with both adsorbates is 0.56 eV for HEA slabs and 0.26 eV for single-site slabs, with a combined standard deviation of 0.42 eV. The standard deviation of the predicted binding energies of COOH given H,  $\Delta E_{COOH(H^*)}^{Predicted}$ , is a little higher at 0.55 eV. The standard deviation of the binding energies of COOH given H, between 0.25 eV and 0.27 eV, is significantly larger than the standard deviation of the binding energies of COOH, between 0.06 and 0.08, for each adsorbing atom in the on-top sites individually. While some standard deviation is expected, because the HEA surface is comprised of many different sites, the standard deviation on the predicted binding energies of COOH given the presence of H are unexpectedly high. It might indicate that the more complex local environment with H adsorbed on a neighbouring site naturally gives rise to more varying and on average a bit higher binding energy. Furthermore, the predictions are based on two separate models, which both have a variance in their predictions, increasing the total variance.

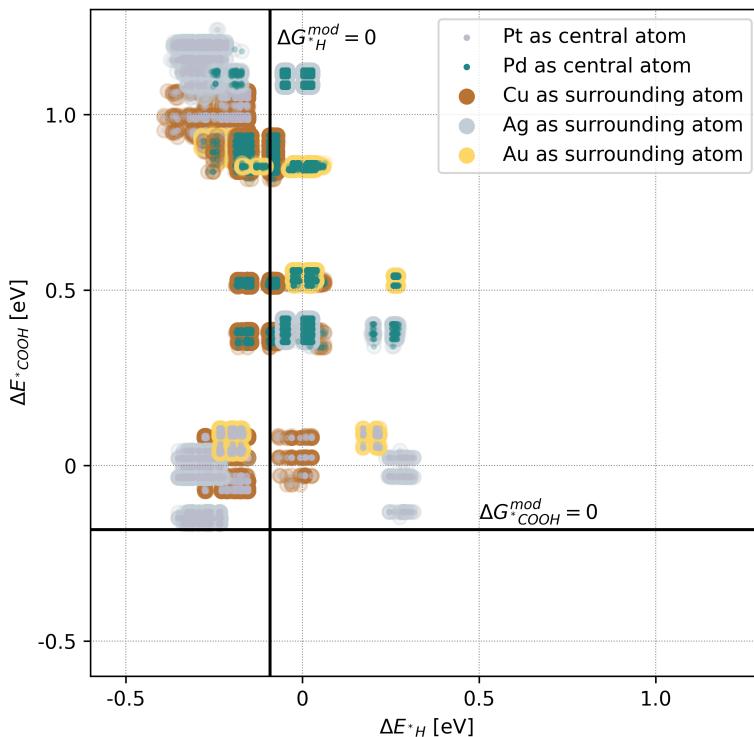


**Figure 15 – The predicting binding energies of H given the presence of COOH on a neighbouring site as a function of the binding of H on an equimolar 200 by 200 by 3 simulated HEA surface. The mean binding energies are marked with large circles with black borders, with the difference in mean between  $\Delta E_{H(COOH)}$  and  $\Delta E_H$  shown in text. The diagonal line marks where the two binding energies are equal. The thermal corrections are marked with solid lines.**

In figure 15, the binding energies of H given the presence of a neighbouring COOH is shown on the y-axis with the binding energy of H as predicted by the single-site model on the x-axis. All energies are predicted based on the same equimolar 200 by 200 by 3 HEA simulated surface. The binding site in a hollow site consists of three atoms, hence binding energies of H in a hollow site cannot be colored in the same way as an on-top site, where the single atom in the on-top site determines the color in the plot. Hence, in figure 15, each binding energy of H is represented in the colors of all metals present in the adsorbing hollow site. This can be either one, two or three colors depending on the number of unique atoms in the adsorbing site. Like in figure 14, differences in binding energies between sites with the presence of different metals can be seen. The mean H binding energies for hollow sites with the presence of a specific metal are for all metals higher, when COOH is present on a neighbouring on-top site. The surplus is calculated for each hollow site as  $E_{*H(*COOH)}^{Predicted} - E_{*H}^{Predicted}$ , and are between +0.21 eV and +0.27 eV for sites with specific metals present. The average surplus binding energy when COOH is present is estimated to be +0.24 eV. That is the same difference as when adding an H next to a COOH and the opposite, adding a COOH next to an H. Maybe logic would suggest COOH as neighbour would cause a bigger difference, but this doesn't suggest it. Although for Pt, Pd and Ag, the COOH is much less "disturbed" by H than H is by COOH through all metals. If the results are reflective of a real-world phenomenon, an average binding energy increase of H when COOH is present of -0.24 eV could be good news for efficient FAOR. This would mean that even lower anode potentials, and hence, higher efficiencies, would be reachable without H adsorbing next to COOH and resulting in CO-poisoning. Although on-top sites may not always be occupied, at which point H could adsorb with its single-adsorbate energy again. The overlap between COOH and H at a range of potentials between 0.0 V and 0.4 V might be smaller than predicted by the single-adsorbate models.

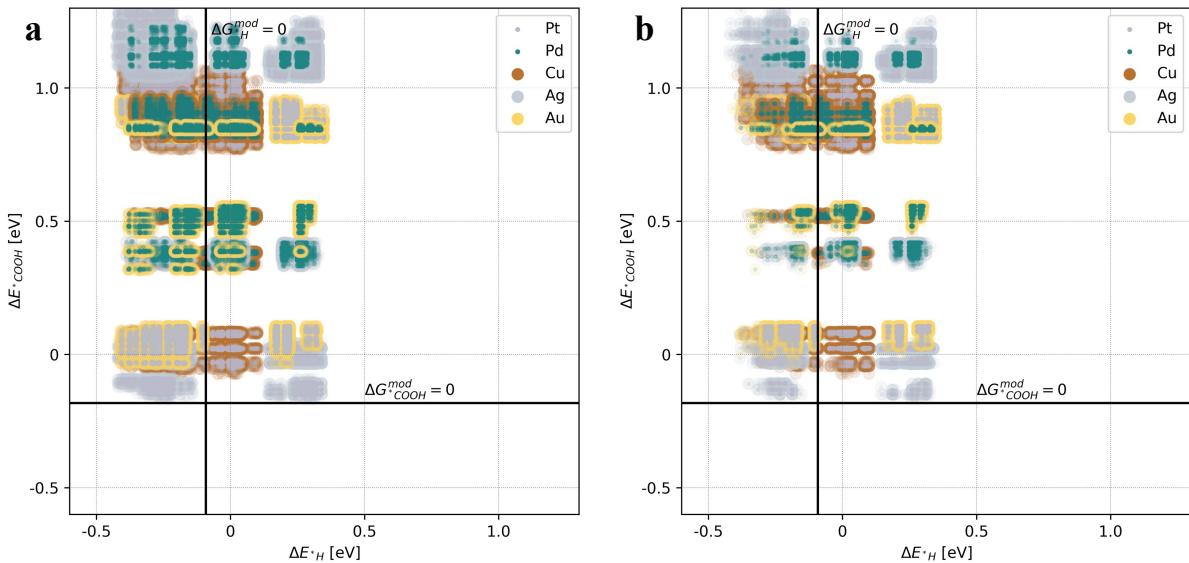
### **Applying the models to single-site structures**

Fcc(111) Single-site structures are of interest, hence whole simulated single-site structured surfaces are examined. Even though optimal single-site structured surfaces are not known to exist in reality, the simulated surfaces serve as a tool to investigate single-site structures and show a theoretical performance. The structure of the optimal single-site structured surface is shown in figure 7.



**Figure 16 - Predicted binding energies of COOH and H on a 200 by 200 by 3 simulated “single-site structured” surface with a molar fraction of 1/3 of the “central” atom and 2/3 of the “surrounding” atom. The markers have the colour of the “central” atom in a small circle, and the “surrounding” atom in a larger surrounding circle. The top layer is arranged into the optimal single-site pattern as shown in figure 7. The two bottom layers are mixed randomly. A surface with every combination with Pt and Pd as the “central” atom and Cu, Ag and Au as the “surrounding” atom was simulated, and the predicted binding energies are shown. The thermal corrections are marked with solid lines.**

In figure 16, the binding energies of COOH and H is shown for 200 by 200 by 3 simulated single-site structured surfaces are shown. All combinations of Pt or Pd as the central atom and Cu, Ag or Au as the surrounding are simulated and the energies shown in the figure, with the energies from a surface marked with a small inner circle for the central atom and a larger outer circle for the surrounding atom. As the top layer is ordered, the number of different sites is limited greatly, giving rise to fewer, distinct groups of binding energies, with a small variance from the lower two layers of the surface, which are still randomly ordered. As expected, the single-site structured surfaces with Pt as the central atom have distinctly lower binding energies than the Pd counterparts. This means Pd-based surfaces will not be FAOR active until anode potentials of 0.35 V vs RHE, which is very inefficient compared to the Pt-based surfaces. The binding energies with possible CO-resistance are the pairs with low COOH binding energies, but high H binding energies. This can be seen for both PtAg<sub>2</sub>, PtCu<sub>2</sub> and partly PtAu<sub>2</sub>.

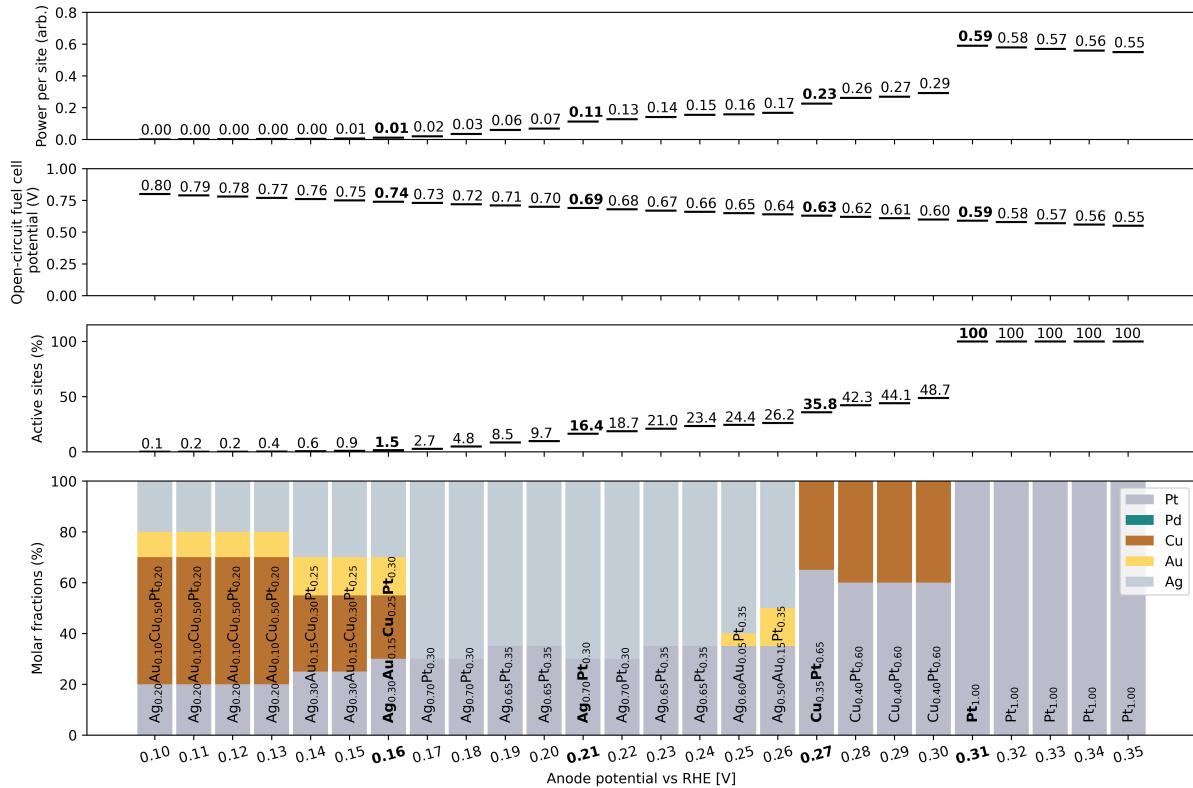


**Figure 17 – Predicted binding energies from 200 by 200 by 3 randomly mixed bimetallic simulated surfaces with a molar fraction of 1/3 for Pt/Pd and 2/3 for Ag/Au/Cu (left) and 1/7 for Pt/Pd and 6/7 for Ag/Au/Cu (right). Every combination of strongly binding and less strongly binding metal is simulated and plotted.**

In order to show the impact of the ordering of the top layer as opposed to the molar fraction in itself, binding energies from simulated HEA surfaces with the same molar fractions as the single-site structured surfaces are shown in figure 17a. In figure 17b, the molar fraction is 1/7 central atom and 6/7 surrounding atom, as this fraction has been proven to be the one that gives the largest statistical occurrence, 5.6 %, of single-sites in a randomly mixed surface. In both plots, there are more energies available due to a larger entropy of the top layer, and the same binding energies are present in 21a and 21b with the only difference being the prevalence of each binding energy-pair. If any of these bimetals with either [1/3, 2/3] or [1/7, 6/7] molar fractions are the optimal surface at any potential, it'll be found in a brute-force search through all molar fractions with a fine enough step-size.

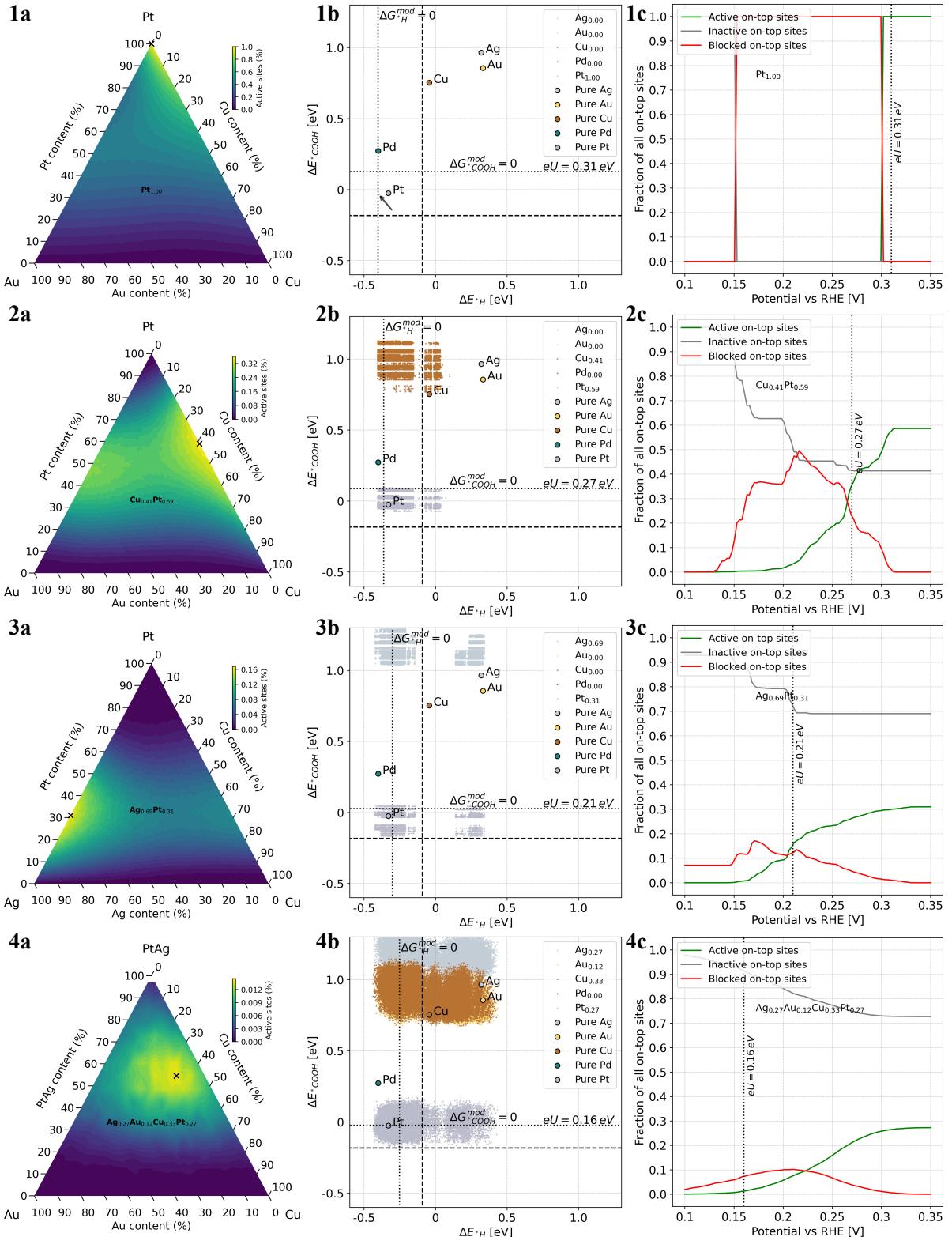
### Searching for optimal HEA compositions at different potentials

A brute-force search was employed to find the most optimal composition at each anode potential from 0.10 V vs RHE to 0.35 V vs RHE with a step size of 0.01 V. The brute-force was carried out with a step-size of 5% in the molar fractions. A 100 by 100 by 3 HEA was simulated for each molar fraction at each potential and evaluated by counting the number of “active sites” using the binding energies of COOH and H, as described in methods.



**Figure 18** – The results of a 10% step size brute-force search through all molar fractions at potentials from 0.1 to 0.35 V. The molar fraction of the most active simulated HEA surface at each potential is shown. The percentage of active sites for the best molar fraction is shown for each potential. The open-circuit fuel cell potential based on a cathode potential for 0.9 V vs RHE for ORR is shown for each potential. The power per site is a result of the current density and open-circuit fuel cell potential. Four select potentials marked with bold text are chosen for further examination.

In figure 18, the results of the brute-force search are shown, with each potential showing the most active molar fraction for a HEA. As the anode potential decreases, the percentage of active sites decreases consistently, while the open-circuit fuel cell potential increases, increasing the efficiency of the fuel cell. The open-circuit fuel cell potential is calculated as the difference between the anode potential and the cathode potential for ORR, which is assumed to be 0.9 V vs RHE in the optimistic scenario. From anode potentials of 0.30 V vs RHE and down, the percentage of active sites decreases faster than the open-circuit fuel cell potential decreases, leading to decreasing power per site. The power per site is the product of the open-circuit fuel cell potential and the current density, with the current density estimation used in this work equal to the fraction of active on-top sites. The molar fractions show that the optimal Pt content decreases as the potential decreases, in an effort to avoid CO-poisoning. The analysis poses a trade-off between power and efficiency, as no molar fraction possesses superior power and efficiency. Four potentials, marked with bold text, are chosen for further examination and represent four different trade-offs between power and efficiency.



In figure 19, an analysis is made, based on the four optimal molar fractions at the four chosen potentials. Initially, in column a, a brute-force search with finer step sizes, 1% (1-3a) and 3% (4a) was made in order to fine-tune the molar fractions. The resulting percentages of active sites are displayed in ternary plots, highlighting the changes in activity, as the molar fraction is altered. A cross marks the optimal activity in the finer brute-force search, and the most active molar fraction is written in the plot. Common for all the newly found best molar fractions is, that they only improve their power per site on the third decimal compared to those found with a 5% step size brute-force search. In 1a, where the single energy-point is marked with an arrow, the brute-force search was made for Pt, Cu and Au, but at the potential of 0.31 V, there is no reason to include any other sites than Pt, because no H binds at 0.31 V, and hence CO-poisoning is not an issue. The most active molar fraction remains  $Pt_{1.0}$ , all the way down to 0.305 V.

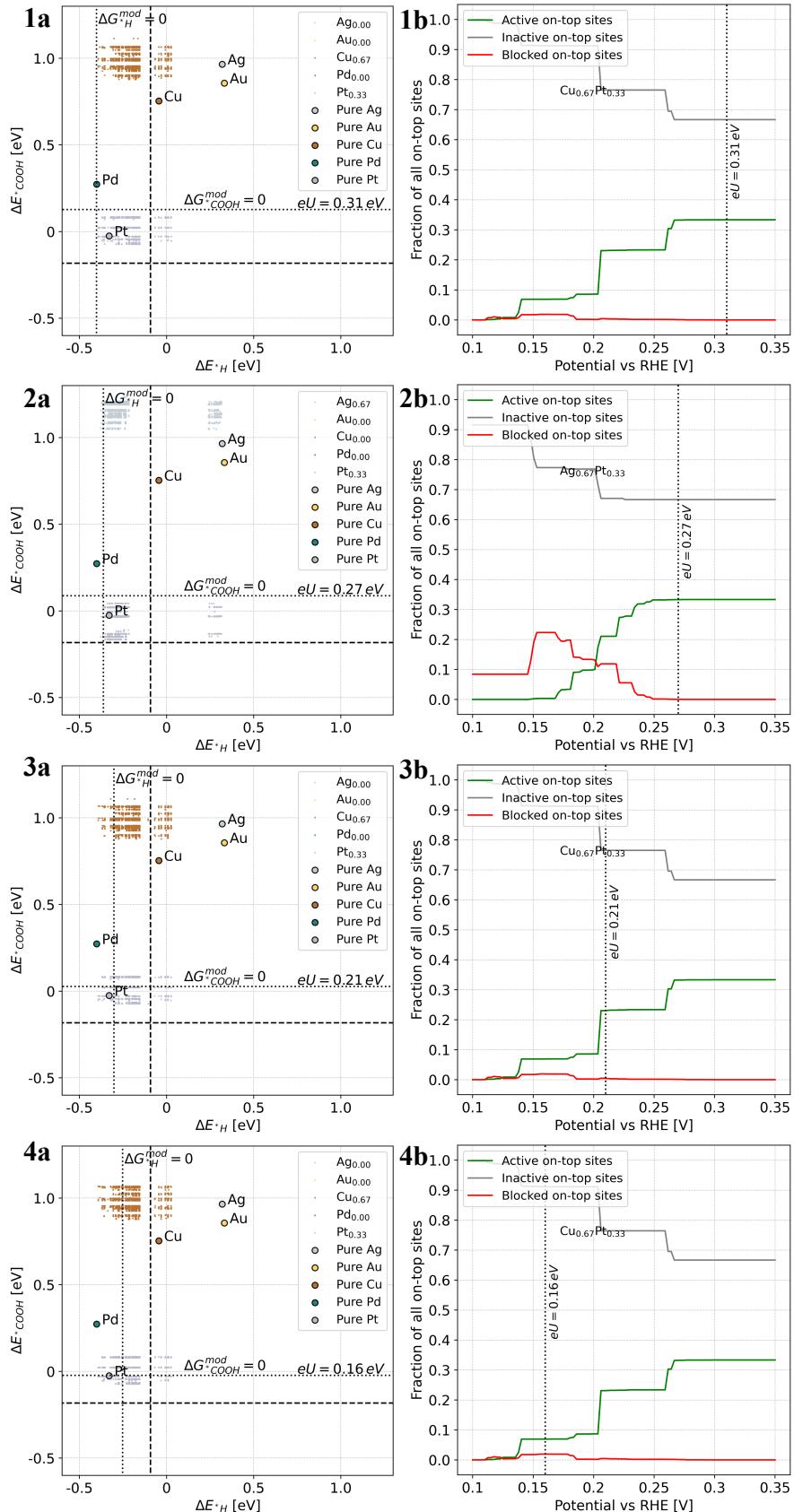
In 2a, the brute-force search was made for Pt, Cu and Au, but the optimal molar fraction at 0.27 V remains a PtCu bimetal with the molar fraction  $Cu_{0.41}Pt_{0.59}$ .

In 3a, the brute-force search was made for Pt, Cu and Ag, as Pt and Ag is in the original optimal molar fraction from the 5% search, and Cu added because it appears often in the optimal compositions, but the optimal molar fraction at 0.21 V remains a PtAg bimetal with the molar fraction  $Ag_{0.69}Pt_{0.31}$ . In 4a, the most active molar fraction of the 5% brute-force search was a quaternary alloy, and in the ternary plot of the 3% brute-force search with Pt, Cu, Ag and Au, one of the axes was set to Pt and Ag combined. The optimal molar fraction at 0.16 V was found to be the quaternary alloy  $Ag_{0.27}Au_{0.12}Cu_{0.33}Pt_{0.27}$ . In column b, the binding energies of the newly found optimal compositions are shown with two lines representing  $\Delta G^*_{ads} = 0$  eV for COOH and H. In column c, the number of expected active, inactive, and blocked sites are shown at each potential, only based on the binding energies at the specific potential, not regarding history of where the potential has been earlier, for example carrying CO from previous lower potentials. In these plots, the timing and scale of the CO-poisoning can be seen in the area at the red curve. At 0.31 V in 1c it is clear why a pure Pt surface is the optimal composition, as Pt does not become poisoned at that potential. At 0.26 V in 2c the first compromise between Pt, which is required for active sites at that potential, and Cu that perturbs the surface to provide weaker binding hollow sites. At 0.21 V in 3c, the Pt content is even lower to balance the number of possible FAOR active sites and blocked sites. At 0.16 V avoiding CO-poisoning is extremely hard, and the resulting optimal molar fraction uses even less Pt and four metals to widen the binding energies as much as possible to push a small fraction of the binding energies down into the right corner under the  $\Delta G^*_{ads} = 0$  cross in 4b.

From the pure Pt catalyst at an anode potential of 0.31 V, the efficiency can be increased by  $\frac{0.74-0.59}{0.59} * 100\% = 25.4\%$ , while reducing the power per site to 1/42<sup>th</sup> (0.59 to 0.014), by going to 0.16 V with a  $Ag_{0.27}Au_{0.12}Cu_{0.33}Pt_{0.27}$  catalyst. Alternatively, a  $\frac{0.69-0.59}{0.59} * 100\% = 17\%$  increase of the open-circuit fuel cell potential can be achieved at an anode potential of 0.21 V while only reducing the power to  $\approx 1/5^{th}$  with the  $Ag_{0.69}Pt_{0.31}$  catalyst.

### **Searching for the best single-site structured surfaces at different potentials**

Single-sites have been hypothesized to have a resistance to CO-poisoning while at the same time providing good sites for FAOR. In order to evaluate this for fcc(111) single-site structures as described in this work, all single-site structured surfaces from figure 16 were tested with the same activity estimation routine at the four selected potentials. Although these surfaces are only theoretical, and not synthesizable in a laboratory, due to the low entropy of the top layer, they serve to display the possibilities of single-site structures.



**Figure 20 – The predicted binding energies of COOH and H on the optimal single-site structured surfaces at each of the four selected potentials (1a-4a). The number of sites expected to be active, inactive, and blocked for the optimal surfaces at a smooth range of potentials without regard for prior potentials are shown, with the potential that the single-site structured surface was optimized for marked with a dotted line (1b-4b).**

In figure 20, the most active single-site structured surfaces at each of the four selected potentials are highlighted. In column a, the binding energies of COOH and H are shown. Due to the ordering of the top layer, fewer unique sites exist and therefore fewer unique binding energies. This in turn creates sharp changes in the number of active, inactive, and blocked sites, as shown in column b. For the PtCu single-site structured surface that is optimal at 0.31 V, 0.21 V, and 0.16 V, the energies are positioned in such a way that gives rise to impressively little CO-poisoning. This property would make it possible to cycle down to an anode potential of 0.145 V, at an open-circuit fuel cell potential of 0.755 V and 0.05 arbitrary units of power, while only poisoning under 4% of the Pt sites. Above 0.305 V there would be no reason to use the theoretical single-site structured surfaces, since CO-poisoning is not an issue at those potentials, while having as many Pt sites as possible will result in a more active catalyst.

### Comparing optimal HEA surfaces with single-site surfaces

In the table above, the most active catalysts found with HEA and SS surfaces at each potential is shown with their molar fractions, percentage of active sites and power.

Anode potential vs RHE [V]:	0.31	0.27	0.21	0.16
Open-circuit fuel cell potential [V]	0.59	0.63	0.69	0.74
High-Entropy Alloys	Molar fraction	<b>Pt<sub>1.0</sub></b>	<b>Cu<sub>0.41</sub>Pt<sub>0.59</sub></b>	<i>Ag<sub>0.69</sub>Pt<sub>0.31</sub></i>
	Active sites (%)	<b>100</b>	<b>35.8</b>	16.5
	Power (arb. units)	<b>0.59</b>	<b>0.23</b>	0.11
Single-site structures	Molar fraction	<i>Cu<sub>0.67</sub>Pt<sub>0.33</sub></i>	<i>Ag<sub>0.67</sub>Pt<sub>0.33</sub></i>	<b>Cu<sub>0.67</sub>Pt<sub>0.33</sub></b>
	Active sites (%)	33.3	33.3	<b>22.7</b>
	Power (arb. units)	0.20	0.21	<b>0.16</b>
				<b>0.05</b>

The most active composition at each potential is highlighted with bold text. At anode potentials of 0.31 V and 0.27 V, the HEAs found by the brute-force search are more active, than the best SS surfaces, seemingly due to the low Pt content compared with the HEAs. In the SS surfaces, the Pt sites are almost all active, but since there is only a third Pt sites, they are beaten in activity by the HEAs with higher Pt contents, even though the HEA surface already battles with CO-poisoning at 0.27 V. However, at the two lower anode potentials at 0.21 V and 0.16 V, avoiding CO-poisoning and strongly binding on-top sites becomes the most important factor in retaining FAOR activity. Here, the theoretical simulated SS surfaces beat the HEA surfaces. The SS surface beats the HEA surface with 45% higher activity at 0.21 V and 400 % higher activity at 0.16 V.

## Discussion

Knowing the precise impact of neighbouring adsorbates on the binding energies of the relevant species could perhaps alter the system modelled in this work if there is a repelling effect between H and COOH, since the overlaps are the essential mechanism in CO-poisoning. Unfortunately, the variances in the models trained to predict that are conspicuously large, though the mean binding energy increases with a neighbouring adsorbate seem sensible. As further work, more DFT data and work could be put into the mixed-site models to better understand how the adsorbates act in the overlap region. For the theoretical SS surfaces, even with an ordered top layer, there is still a spread in the binding energies due to the two random bottom layers. If one was looking for the utmost optimal theoretical surfaces, the next step would be a total search through all structures that include special ordering of the two layers under the top layer. This could include more DFT data on specific slabs of interest to guarantee the accuracy of the model used to extrapolate binding energy predictions from the slabs. This way, even more tailored theoretical surfaces could be made, possibly giving insights that could be utilized in real catalysts. As further work, a scan through crystal databases for fcc(111) surfaces with platinum and gold, silver or copper alloys with single-site resembling structures, such as the cubic  $Pd_1Au_3$  structure, to find possible real-life single-site structures is suggested. A mathematical analysis of fcc(111) single-sites showed, that the molar fraction leading to the highest statistical occurrence of single-sites is found when mixing two metals randomly is  $A_{1/7}B_{6/7}$ . This molar fraction bimetal could have been found almost exactly in the brute-force searches with a step-size of 1% in the molar fraction, but it was not. Even at the low potentials of 0.21 V and 0.16 V, the optimal compositions had more platinum and a mix of more than two metals, respectively. This indicates

that the widening of the binding energy bands from the two bottom layers also play a key role in the fine-tuning of the surfaces, and the statistical occurrence of single-sites, or single platinum-sites in general, is not the only important factor in determining good FAOR performance at low potentials. An in-depth analysis of the local environment in sites predicted to have the perfect binding energies that bind COOH at small overpotentials, while having positive H binding energies at the same potential is suggested as further work. Understanding those conditions could provide new descriptors to optimize in order to find better CO-resistant catalysts. Unfortunately, no on-top sites were predicted to have the optimal COOH binding energy, as there is a lack of metals that bind COOH even stronger at  $\Delta G_{COOH}^{ads} = -0.17$  eV, while currently the strongest binding site in an equimolar HEA is  $\Delta G_{COOH}^{ads} \approx 0$  eV. If CO-poisoning was completely solved, this would likely be the hard limit of FAOR efficiency with PtPdCuAgAu HEAs. Even with theoretical special ordering of the surface, elements binding stronger than Pt would be needed. The thermal corrections were based on experimental cyclic voltammograms, but still included a small arbitrary factor. As further work, these corrections could be adjusted more precisely by making an experiment similar to the one in Bagger et. al. figure 4a, but the potential cycles down gradually to measure the exact potential where Pt is poisoned. The method for estimating activity, while based on the free energy diagram, might not account for all factors in the activity, and the activity of each on-top site could actually depend on binding energies or other factors. In this way, the results and suggested catalysts are highly dependent on the assumptions made about the system. The only way to confirm the validity of the assumptions, corrections, and methods is experimental measurements of the activities of the suggested surfaces on different potentials.

## Conclusion

Formic acid is a promising green fuel for fuel cells with key advantages over H<sub>2</sub> and methanol, such as its non-toxicity, its high volumetric energy density due to it being liquid, and the fact that it has a CO<sub>2</sub>-backbone gives it a high theoretical efficiency, and opens the possibility of forming a closed, net zero-emissions, loop using CO<sub>2</sub> as a carbon feedstock. A limiting factor for an anode catalyst with little to no overpotential is hypothesized to be a disproportionation reaction between H bound to a hollow site and COOH bound to an on-top site, that produced a bound CO, blocking on-top sites. While the adsorption energy of H is decreased at negative potentials, and the adsorption energy of COOH increases, there is an overlap in presence of the two adsorbates at low potentials, partly blocking the most effective region at the low potentials. The

literature shows some potential solutions with synthesized nanostructures, but this work attempts to combat the problem by brute-forcing through HEA composition space, due to their tunable binding energies, and simulating theoretically optimal single-site structured surfaces on fcc(111) surface. Information about the binding energies was estimated with DFT calculations on simulated fcc(111) HEA surface slabs with and without adsorbates. The electronic DFT energies are corrected with table values for thermal corrections and modified slightly to predict the same behaviour as seen in experiments. Binding energy prediction models were trained for all adsorbates of interest based on HEA and SS DFT slabs and used on arbitrary sites on simulated HEA and SS surfaces. Binding energy models were also trained to predict the binding energy of COOH given the presence of H in a neighbouring site, which estimated the effect to an increase of +0.14 eV on the binding energy of COOH on Pt on-top sites, while models predicting the binding energy of H given the presence of COOH on a neighbouring site, estimated the average binding energy increase to +0.24 eV, both on equimolar HEA surfaces. An optimization criterion was employed to evaluate the activity of the surfaces simulated based on a brute-force search through all molar fractions with a 5 % step size. The optimization criterion estimates activity with the number of active sites by counting the number of on-top sites with negative binding energies for COOH, but positive binding energies for H in all neighbouring hollow sites, ensuring no CO-poisoning will block the site. This brute-force search is used at a range of anode potentials from 0.10 V to 0.35 V vs RHE, to find optimal HEA catalysts for potentials that represent different trade-offs between open-circuit fuel cell potential and power per site. The results predict that lower potentials require less and less Pt to avoid excessive CO-poisoning, while anode potentials above 0.305 V can use pure Pt without any CO-poisoning as H does not bind to pure Pt at that potential. The catalyst for the lowest potential at 0.16 V utilizes four elements,  $Ag_{0.27}Au_{0.12}Cu_{0.33}Pt_{0.27}$ , to achieve a small fraction of un-poisoned sites. The optimal single-site structured surfaces are able to beat the HEA surfaces at the two lowest potentials, 0.21 and 0.16 V with the same  $Cu_2Pt_1$  surface, which can sustain 6.9 % active sites all the way down to 0.145 V. This surface almost achieves complete CO-resistance at all potentials, as under 4% of all on-top sites are estimated to be CO-poisoned when scanning from high potentials to 0.10 V and back. At higher potentials with more power, the SS surfaces are limited to 33% active sites, due to the molar fraction of the top layer, hence finetuned HEA surfaces can outperform SS surfaces at potentials over 0.27 V. This work suggests that, if high power is important, a pure Pt anode catalyst could be optimal at 0.31 V, and if higher efficiency is important, HEA surfaces can be tuned to achieve better efficiencies at huge power decreases, and if very low anode potentials, under 0.16 V, are needed, special structures must be employed

if more than 1.4 % of the sites should be active. Unfortunately, even on platinum-sites in a HEA, no binding energies are as low as the optimal  $\Delta G_{COOH}^{opt} = -0.17 \text{ eV}$ , which is potentially an even bigger problem than CO-poisoning, since it might require an element with catalytic properties that does not exist in the periodic table.

## References

- (1) Arrhenius, S. “On the Influence of Carbonic Acid in the Air upon the Temperature of the Ground” (1896). In *The Future of Nature*; Robin, L., Sörlin, S., Warde, P., Eds.; Yale University Press, 2017; pp 303–315. <https://doi.org/10.12987/9780300188479-028>.
- (2) *A Brief History of Carbon Dioxide Measurements – AIRS*. <https://airs.jpl.nasa.gov/news/111/a-brief-history-of-carbon-dioxide-measurements/> (accessed 2023-08-03).
- (3) *Evidence | Facts – Climate Change: Vital Signs of the Planet*. <https://climate.nasa.gov/evidence/> (accessed 2023-08-03).
- (4) *Keeling Curve - American Chemical Society*. <https://www.acs.org/education/whatischemistry/landmarks/keeling-curve.html> (accessed 2023-08-03).
- (5) Brook, E. J.; Buzert, C. Antarctic and Global Climate History Viewed from Ice Cores. *Nature* **2018**, 558 (7709), 200–208. <https://doi.org/10.1038/s41586-018-0172-5>.
- (6) Rubino, M.; Etheridge, D.; Thornton, D.; Allison, C.; Francey, R.; Langenfelds, R.; Steele, P.; Trudinger, C.; Spencer, D.; Curran, M.; Van Ommen, T.; Smith, A. Law Dome Ice Core 2000-Year CO<sub>2</sub>, CH<sub>4</sub>, N<sub>2</sub>O and d<sub>13</sub>C-CO<sub>2</sub>, 2019, 255.62 KB. <https://doi.org/10.25919/5BFE29FF807FB>.
- (7) Keeling, R. F.; Keeling, C. D. Atmospheric Monthly In Situ CO<sub>2</sub> Data - Mauna Loa Observatory, Hawaii. In *Scripps CO<sub>2</sub> Program Data*, 2017. <https://doi.org/10.6075/J08W3BHW>.
- (8) Mitchell, J. F. B. The “Greenhouse” Effect and Climate Change. *Rev. Geophys.* **1989**, 27 (1), 115. <https://doi.org/10.1029/RG027i001p00115>.
- (9) Friedlingstein, P.; O’Sullivan, M.; Jones, M. W.; Andrew, R. M.; Gregor, L.; Hauck, J.; Le Quéré, C.; Luijkx, I. T.; Olsen, A.; Peters, G. P.; Peters, W.; Pongratz, J.; Schwingshakel, C.; Sitch, S.; Canadell, J. G.; Ciais, P.; Jackson, R. B.; Alin, S. R.; Alkama, R.; Arneth, A.; Arora, V. K.; Bates, N. R.; Becker, M.; Bellouin, N.; Bittig, H. C.; Bopp, L.; Chevallier, F.; Chini, L. P.; Cronin, M.; Evans, W.; Falk, S.; Feely, R. A.; Gasser, T.; Gehlen, M.; Grätzl, T.; Gloege, L.; Grassi, G.; Gruber, N.; Gürses, Ö.; Harris, I.; Hefner, M.; Houghton, R. A.; Hurtt, G. C.; Iida, Y.; Ilyina, T.; Jain, A. K.; Jersild, A.; Kadono, K.; Kato, E.; Kennedy, D.; Klein Goldewijk, K.; Knauer, J.; Korsbakken, J. I.; Landschützer, P.; Lefèvre, N.; Lindsay, K.; Liu, J.; Liu, Z.; Marland, G.; Mayot, N.; McGrath, M. J.; Metzl, N.; Monacci, N. M.; Munro, D. R.; Nakaoka, S.-I.; Niwa, Y.; O’Brien, K.; Ono, T.; Palmer, P. I.; Pan, N.; Pierrot, D.; Pocock, K.; Poulter, B.; Resplandy, L.; Robertson, E.; Rödenbeck, C.; Rodriguez, C.; Rosan, T. M.; Schwinger, J.; Séférian, R.; Shutler, J. D.; Skjelvan, I.; Steinhoff, T.; Sun, Q.; Sutton, A. J.; Sweeney, C.; Takao, S.; Tanhua, T.; Tans, P. P.; Tian, X.; Tian, H.; Tilbrook, B.; Tsujino, H.; Tubiello, F.; Van Der Werf, G. R.; Walker, A. P.; Wanninkhof, R.; Whitehead, C.; Willstrand Wranne, A.; Wright, R.; Yuan, W.; Yue, C.; Yue, X.; Zaehle, S.; Zeng, J.; Zheng, B. Global Carbon Budget 2022. *Earth Syst. Sci. Data* **2022**, 14 (11), 4811–4900. <https://doi.org/10.5194/essd-14-4811-2022>.

- (10) Intergovernmental Panel On Climate Change (Ipcc). *Climate Change 2022 – Impacts, Adaptation and Vulnerability: Working Group II Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, 1st ed.; Cambridge University Press, 2023. <https://doi.org/10.1017/9781009325844>.
- (11) Kulp, S. A.; Strauss, B. H. New Elevation Data Triple Estimates of Global Vulnerability to Sea-Level Rise and Coastal Flooding. *Nat. Commun.* **2019**, *10* (1), 4844. <https://doi.org/10.1038/s41467-019-12808-z>.
- (12) CSSR 2017. *USGCRP 2017 Clim. Sci. Spec. Rep. Fourth Natl. Clim. Assess. Vol. Wuebbles DJ DW Fahey KA Hibbard DJ Dokken BC Stewart TK Maycock Eds US Glob. Change Res. Program Wash. DC USA 470 Pp.*
- (13) Glenk, G.; Reichelstein, S. Reversible Power-to-Gas Systems for Energy Conversion and Storage. *Nat. Commun.* **2022**, *13* (1), 2010. <https://doi.org/10.1038/s41467-022-29520-0>.
- (14) Ma, Z.; Legrand, U.; Pahija, E.; Tavares, J. R.; Boffito, D. C. From CO<sub>2</sub> to Formic Acid Fuel Cells. *Ind. Eng. Chem. Res.* **2021**, *60* (2), 803–815. <https://doi.org/10.1021/acs.iecr.0c04711>.
- (15) Carrette, L.; Friedrich, K. A.; Stimming, U. Fuel Cells ± Fundamentals and Applications. *FUEL CELLS* **2001**, No. 1.
- (16) Laraminie, J. and A. Dicks, Fuel Cell Systems Explained. 2003, West Sussex, England. John Wiley & Sons, Inc.
- (17) Eppinger, J.; Huang, K.-W. Formic Acid as a Hydrogen Energy Carrier. *ACS Energy Lett.* **2017**, *2* (1), 188–195. <https://doi.org/10.1021/acsenergylett.6b00574>.
- (18) Bagger, A.; Jensen, K. D.; Rashedi, M.; Luo, R.; Du, J.; Zhang, D.; Pereira, I. J.; Escudero-Escribano, M.; Arenz, M.; Rossmeisl, J. Correlations between Experiments and Simulations for Formic Acid Oxidation. *Chem. Sci.* **2022**, *13* (45), 13409–13417. <https://doi.org/10.1039/D2SC05160E>.
- (19) Linstrom, P. NIST Chemistry WebBook, NIST Standard Reference Database 69, 1997. <https://doi.org/10.18434/T4D303>.
- (20) Al-Akraa, I. M.; Asal, Y. M.; Darwish, S. A. A Simple and Effective Way to Overcome Carbon Monoxide Poisoning of Platinum Surfaces in Direct Formic Acid Fuel Cells. *Int. J. Electrochem. Sci.* **2019**, *14* (8), 8267–8275. <https://doi.org/10.20964/2019.08.100>.
- (21) Ortiz-Ortega, E.; Carrera-Cerritos, R.; Arjona, N.; Guerra-Balcázar, M.; Cuevas-Muñiz, F. M.; Arriaga, L. G.; Ledesma-García, J. Pd Nanostructures with High Tolerance to CO Poisoning in the Formic Acid Electrooxidation Reaction. *Procedia Chem.* **2014**, *12*, 9–18. <https://doi.org/10.1016/j.proche.2014.12.035>.
- (22) Zhong, W.; Zhang, D. New Insight into the CO Formation Mechanism during Formic Acid Oxidation on Pt(111). *Catal. Commun.* **2012**, *29*, 82–86. <https://doi.org/10.1016/j.catcom.2012.09.002>.
- (23) Ooka, H.; Huang, J.; Exner, K. S. The Sabatier Principle in Electrocatalysis: Basics, Limitations, and Extensions. *Front. Energy Res.* **2021**, *9*, 654460. <https://doi.org/10.3389/fenrg.2021.654460>.
- (24) *The IUPAC Compendium of Chemical Terminology: The Gold Book*, 4th ed.; Gold, V., Ed.; International Union of Pure and Applied Chemistry (IUPAC): Research Triangle Park, NC, 2019. <https://doi.org/10.1351/goldbook>.
- (25) Chan, K.; Tsai, C.; Hansen, H. A.; Nørskov, J. K. Molybdenum Sulfides and Selenides as Possible Electrocatalysts for CO<sub>2</sub> Reduction. *ChemCatChem* **2014**, *6* (7), 1899–1905. <https://doi.org/10.1002/cctc.201402128>.
- (26) Atkins, P. W.; De Paula, J.; Keeler, J. *Atkins' Physical Chemistry*, Twelfth edition.; Oxford University Press: New York, NY, 2023.

- (27) Gibbs, J. W. Gibbs, Josiah Willard. “On the Equilibrium of Heterogeneous Substances.” (1879): 300-320.
- (28) Yeh, J.-W.; Chen, S.-K.; Lin, S.-J.; Gan, J.-Y.; Chin, T.-S.; Shun, T.-T.; Tsau, C.-H.; Chang, S.-Y. Nanostructured High-Entropy Alloys with Multiple Principal Elements: Novel Alloy Design Concepts and Outcomes. *Adv. Eng. Mater.* **2004**, *6* (5), 299–303. <https://doi.org/10.1002/adem.200300567>.
- (29) Swalin, R. A. *Thermodynamics of Solids*, 2d ed.; Wiley series on the science and technology of materials; J. Wiley: New York, 1972.
- (30) Miracle, D. B.; Senkov, O. N. A Critical Review of High Entropy Alloys and Related Concepts. *Acta Mater.* **2017**, *122*, 448–511. <https://doi.org/10.1016/j.actamat.2016.08.081>.
- (31) Troparevsky, M. C.; Morris, J. R.; Daene, M.; Wang, Y.; Lupini, A. R.; Stocks, G. M. Beyond Atomic Sizes and Hume-Rothery Rules: Understanding and Predicting High-Entropy Alloys. *JOM* **2015**, *67* (10), 2350–2363. <https://doi.org/10.1007/s11837-015-1594-2>.
- (32) Batchelor, T. A. A.; Pedersen, J. K.; Winther, S. H.; Castelli, I. E.; Jacobsen, K. W.; Rossmeisl, J. High-Entropy Alloys as a Discovery Platform for Electrocatalysis. *Joule* **2019**, *3* (3), 834–845. <https://doi.org/10.1016/j.joule.2018.12.015>.
- (33) Lv, Z. Y.; Liu, X. J.; Jia, B.; Wang, H.; Wu, Y.; Lu, Z. P. Development of a Novel High-Entropy Alloy with Eminent Efficiency of Degrading Azo Dye Solutions. *Sci. Rep.* **2016**, *6* (1), 34213. <https://doi.org/10.1038/srep34213>.
- (34) Pedersen, J. K.; Batchelor, T. A. A.; Bagger, A.; Rossmeisl, J. High-Entropy Alloys as Catalysts for the CO<sub>2</sub> and CO Reduction Reactions. *ACS Catal.* **2020**, *10* (3), 2169–2176. <https://doi.org/10.1021/acscatal.9b04343>.
- (35) Winjobi, O.; Zhang, Z.; Liang, C.; Li, W. Carbon Nanotube Supported Platinum–Palladium Nanoparticles for Formic Acid Oxidation. *Electrochimica Acta* **2010**, *55* (13), 4217–4221. <https://doi.org/10.1016/j.electacta.2010.02.062>.
- (36) Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *136* (3B), B864–B871. <https://doi.org/10.1103/PhysRev.136.B864>.
- (37) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140* (4A), A1133–A1138. <https://doi.org/10.1103/PhysRev.140.A1133>.
- (38) Hammer, B.; Hansen, L. B.; Nørskov, J. K. Improved Adsorption Energetics within Density-Functional Theory Using Revised Perdew-Burke-Ernzerhof Functionals. *Phys. Rev. B* **1999**, *59* (11), 7413–7421. <https://doi.org/10.1103/PhysRevB.59.7413>.
- (39) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77* (18), 3865–3868. <https://doi.org/10.1103/PhysRevLett.77.3865>.
- (40) Mortensen, J. J.; Hansen, L. B.; Jacobsen, K. W. Real-Space Grid Implementation of the Projector Augmented Wave Method. *Phys. Rev. B* **2005**, *71* (3), 035109. <https://doi.org/10.1103/PhysRevB.71.035109>.
- (41) Friedman, Jerome H. “Greedy Function Approximation: A Gradient Boosting Machine.” *Annals of Statistics* (2001): 1189–1232.
- (42) Chen, Tianqi, et al. “Xgboost: Extreme Gradient Boosting.” R Package Version 0.4-2 1.4 (2015): 1-4.
- (43) Van Rossum, G.; Drake Jr, F. L. *Python Reference Manual*; Centrum voor Wiskunde en Informatica Amsterdam, 1995.
- (44) Pedersen, J. K.; Clausen, C. M.; Krysiak, O. A.; Xiao, B.; Batchelor, T. A. A.; Löffler, T.; Mints, V. A.; Banko, L.; Arenz, M.; Savan, A.; Schuhmann, W.; Ludwig, A.; Rossmeisl, J. Bayesian Optimization of High-Entropy Alloy Compositions for

- Electrocatalytic Oxygen Reduction\*\*. *Angew. Chem.* **2021**, *133* (45), 24346–24354.  
<https://doi.org/10.1002/ange.202108116>.
- (45) Peterson, A. A.; Nørskov, J. K. Activity Descriptors for CO<sub>2</sub> Electroreduction to Methane on Transition-Metal Catalysts. *J. Phys. Chem. Lett.* **2012**, *3* (2), 251–258.  
<https://doi.org/10.1021/jz201461p>.
- (46) Studt, F.; Abild-Pedersen, F.; Varley, J. B.; Nørskov, J. K. CO and CO<sub>2</sub> Hydrogenation to Methanol Calculated Using the BEEF-vdW Functional. *Catal. Lett.* **2013**, *143* (1), 71–73. <https://doi.org/10.1007/s10562-012-0947-5>.