Simulating Cyclic Voltammograms

Project in Practice at Center for High Entropy Alloys

Written by Marcus Nørgaard Weng

1 Introduction

Good catalysts are essential for many chemical processes in green technologies that are prospective technological solutions to the climate crisis. This includes green energy solutions like fuel cell reactions and power-to-x, that are reliant on good catalytic activity. It is therefore of great importance to find new ways to improve and discover new catalysts. Cyclic voltammograms is an experimental method to examine a reaction, which shows the current from the reaction as a function of the applied voltage. In fuel cell reactions a large current at a small voltage is preferable, but measuring this experimentally requires that the catalyst material is synthesized and tested in a laboratory. Hence, large-scale testing of thousands of catalyst materials would be infeasible. The composition space of eg. high entropy alloys (HEAs) is extremely large, and thus can not be explored in a timely manner purely by experimental measurements. If cyclic voltammograms could be simulated quickly and accurately on a computer given a surface composition and a reaction, thousands of compositions' cyclic voltammograms could be simulated and compared each day. This project is an attempt at simulating cyclic voltammograms of water on an arbitrary HEA catalyst, only using the five metals Ag, Au, Cu, Pd and Pt, which are present in the data set, with variable hysteresis and steric hindrance.
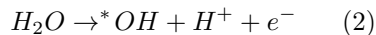
1.1 Reactions

In an aqueous solution with a reversible hydrogen electrode (RHE) three adsorbates on the catalyst is expected; hydrogen, hydroxide and oxygen. The adsorption reaction and corresponding binding energy of hydrogen is:

$$H^+ + e^- \to {}^*H \qquad (1)$$
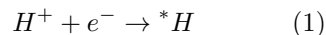
$$\Delta E_{*H} = E_{*H} - E_* - \tfrac{1}{2}E_{H_2} + eU$$

The adsorption reaction and corresponding binding energy of hydroxide is:

$$H_2O \to {}^*OH + H^+ + e^- \qquad (2)$$

$$\Delta E_{*OH} = E_{*OH} - E_* + \tfrac{1}{2}E_{H_2} - eU$$

The adsorption reaction and corresponding binding energy of oxygen is:

$$H_2O \to {}^*O + 2*H^+ + 2*e^- \qquad (3)$$

$$\Delta E_{*O} = E_{*O} - E_* + E_{H_2} - 2eU$$

The star $*$ indicates the surface and a star in front of an atom or molecule indicates an adsorbate. Because the hydrogen adsorption reaction takes an electron and the other two produce

electrons, the hydrogen adsorption reaction is encouraged under reductive conditions and the others are encouraged under oxidative conditions.

1.2 DFT training data

The DFT data consists of slabs with different compositions with an adsorbate on the surface and an associated energy. The H data set has 56 slabs with 9 hollow FCC sites used per slab at a 3x3x5 slab totaling 504 different scenarios with an associated energy. The OH data set has 395 slabs with 1 on-top site used per slab. The O data set has 500 slabs with 1 hollow FCC site used per slab. Furthermore, pristine slabs with single metals were used to find the energy of the surface without adsorbates. All DFT data was provided by Jack Kirk Pedersen.

2 Methods

2.1 Predicting binding energies of adsorbates

In order to simulate cyclic voltammograms, an accurate binding energy prediction model for arbitrary sites for the adsorbates of interest. A model for on-top sites and a model for hollow sites was trained on the DFT data. In order to train a regression model on the DFT data with the adsorbate and surface metals as input and the binding energy as output, a feature vector that converts the relevant metals from a slab to a suitable input is needed. The feature vector used takes select, important atoms from the first three layers into account as per Clausen, Christian M., et. al.[Cla+21]. Linear regression models were tested as an extremely

simple solution, which did very well, when only predicting the energy of a single adsorbate. The fact that such a simple model works well shows that the problem of predicting binding energy based on metals in the surroundings of the site is very simple. A more complex XGBoost[Che+15] (Boosted Decision Tree) model was used because it has the ability to predict binding energies of different adsorbates that binds to the same type of site when including the adsorbate in the feature vector and combining multiple data sets. This theoretically allows for transfer learning between different adsorbates and reduces the number of models to one per site type.

2.2 Predicting energies of simulated surfaces

Surfaces were simulated by randomly drawing one of the chosen metals for each atom in the surface. The metals can be chosen freely from the metals that the energy prediction models were trained on (Ag, Au, Cu, Pd, Pt). The binding energy prediction model can then be used on all possible adsorbates on all sites to get estimates of all possible binding energies on the surface. This will be necessary for both methods of simulating cyclic voltammograms, that are used in this project.

2.3 Surface binding energies to voltammogram

A phase diagram shows the binding energy (eV) as a function of the voltage (V), but the binding energy can not rise above 0, at which point a constant function y = 0 takes over. The slope of the binding energy can be seen in the equations 1-3 and is the coefficient

to the eU terms, relating to the number of electrons used or produced in the adsorption reaction. Differentiating the phase diagram yields a charge isoterm. A charge isoterm shows the charge per area ($\mu C cm^{-2}$) as a function of voltage. Differentiating the isoterm yields the cyclic voltammogram, which shows the charge transfer, amperage ($\mu A cm^{-2}$), as a function of voltage[Ros+20]. The first method of converting all surface binding energies to a cyclic voltammogram starts with creating a phase diagram for each individual adsorbate binding energy. For a 100x100 simulated surface, 10000 on-top sites, where OH can adsorb, and 10000 hollow FCC sites, where H and O exist, resulting in 30000 energies and corresponding phase diagrams. The derivative of each phase diagram is taken and summed together resulting in a singular charge isoterm, the derivative of which is the cyclic voltammogram.

2.4 Simulating surface adsorbates while varying voltage

The previous cyclic voltammogram method does not account for effects seen in experimental data, such as hysteresis and neighbour interactions. A second method where these effects are approximated is used. In this second method a linear sweep or circular sweep in voltage made, while concurrently keeping track of the adsorbates on the surface and the summed biding energies (ΔG) of the adsorbates at each voltage. The steps in the algorithm for this method is the following for each voltage:

- Use the binding-energy predic-
tion model to predict energies of all adsorbates on their appropriate sites (Both adsorbates already adsorbed and prospective adsorbates on empty sites)

- If steric hindrance is implemented: Use approximation for steric hindrance (Neighbour-interactions) to account for steric hindrance between neighbour adsorbates

- Remove all adsorbates with a binding energy above 0 or above the hysteresis energy barrier, if hysteresis is implemented

- Find the prospective adsorbate on an empty site with the smallest negative energy and adsorb the adsorbate

- Continue adsorbing prospective adsorbates until the smallest energy found is no longer negative

- Save the summed ΔG of the adsorbates divided by the number of surface atoms at this voltage for use in the phase diagram.

The phase diagram is constructed from the voltages on the x-axis and the summed ΔG of the adsorbates that are currently adsorbed on the surface divided by the number of surface atoms on the y-axis. The double derivative of this phase diagram is the cyclic voltammogram, as per [Ros+20].

2.4.1 Modeling neighbour-interactions

There are many ways to estimate neighbour-interactions, that either increase the binding energy of an adsorbate with a close neighbour or reduce the probability to 0. Since the

3

exact interactions and effects are not yet known, an estimate will have to do. The chosen estimate is a model that adds a small energy s to the binding energy for each close neighbour to a site. A close neighbour is defined as follows: On an on-top site, the close neighbours are the three surrounding hollow sites and the six closest on-top sites. The same is true for a hollow site, just with the sites switched. The small energy s can theoretically be tuned to fit the simulated cyclic voltammogram to an experimentally measured cyclic voltammogram.

$$\Delta G_{steric} = \Delta G + s * \sum neighbours \tag{4}$$

2.4.2 Modeling hysteresis

A simple model for imitating hysteresis was implemented by adsorbing adsorbates when the binding energy is negative, but only removing them when the binding energy exceeds a positive energy barrier. This causes adsorbates to adsorb at one voltage, but detach at another voltage. The hysteresis energy barrier h (eV) can also be tuned.

3 Results and discussion

3.1 Predicting binding energies of adsorbates

The results from all energy prediction models on the hollow site adsorbates are shown in figure 1. When training each model, the data set was divided into a standard 80/10/10 split between training data, validation data and test data. When training on a single adsorbate, the linear regression models surprisingly had better MAE (eV) than the Gradient Boosting model, with $MAE(H) = 0.0439$ eV and $MAE(O) = 0.0547$ eV. An XGBoost model will always be able to achieve at least similar results to a linear regression model, and it could have with more hyperparameter tuning. If the underlying connections are linear, then a linear regression model will do as well as a Gradient Boosting model, and this might be the case, when only training on one adsorbate at a time. For the models trained on both adsorbates on the hollow site, the Gradient Boosting model performed the best with an $MAE(H\&O) = 0.0583$ eV. In this case, a feature indicating which adsorbate the datapoint corresponds to is included in the feature vector. This approach of combining data from adsorbates on the same sites can be extended to more complex systems with additional adsorbates and allows for possible transfer learning between different adsorbates.

Model type	Training data	MAE(H & O)	MAE(H)	MAE(O)
	H and O	**0.0583**	0.0595	0.0569
XGBoost	H	-	0.0558	-
	O	-	-	0.0580
	H and O	0.1044	0.1161	0.0928
Linear Regression	H	-	**0.0439**	-
	O	-	-	**0.0547**

Figure 1: Results on the test data for both energy prediction models trained individually on each hollow-site adsorbate and both at once

In figure 2 the Gradient Boosting model ΔG predictions on the test set of the hollow-site adsorbates is shown as a function of the target ΔG from DFT. Almost all predictions lie within 0.1 eV of the target.
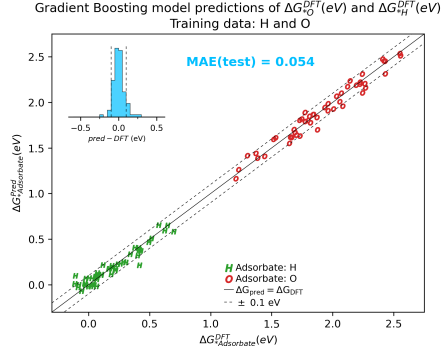
4

Figure 2: Gradient Boosting model predictions in the test data on adsorbates H and O on hollow sites

In figure 3 the Linear Regression model ΔG predictions on the test set of the hollow-site adsorbates is shown as a function of the target ΔG from DFT. It is apparent that the linear regression can not handle multiple adsorbates data sets combined because the model structure is too simple.
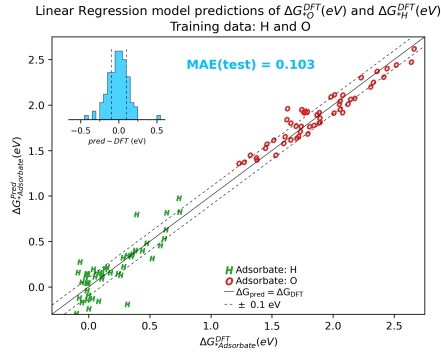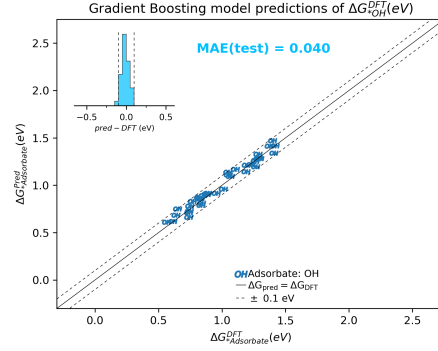

Figure 3: Linear Regression model predictions in the test data on adsorbates H and O on hollow sites

In figure 4 the Gradient Boosting model ΔG predictions on the test set of the hollow site adsorbate OH is shown as a function of the target ΔG from DFT.

Figure 4: Gradient Boosting model predictions in the test data on adsorbate OH on on-top sites

For both the hollow-site model and the on-top model, the performance is great, indicating that the task of predicting the ΔG_{DFT} from nearby surface atoms is fairly easy in terms of machine learning tasks, although the performance could likely be improved with a larger data set given that all three data sets combined are less than 1500 datapoints. Both XGBoost models are used for the rest of the data treatment.

3.2 Predicting energies of simulated surfaces

With the simulated surfaces and the energy prediction models, a surface with a chosen split between chosen metals can be simulated and the energies of all adsorbates predicted. This is shown for a 1000 by 1000 surface with an even split between the metals Ag, Au, Cu, Pd, Pt in figure 5. In the top right plot, the energies for each adsorbate is shown, with H generally having the lowest ΔG, then OH with two peaks and O generally at the highest ΔGs. This follows the expected pattern of the energies.
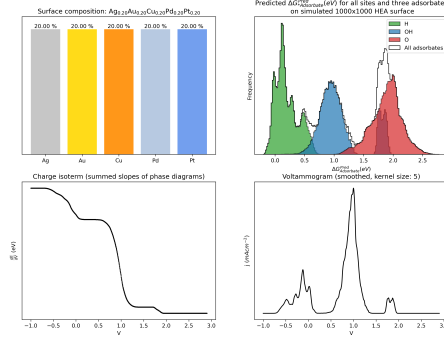
Figure 5: Results from the first method on a surface with a perfectly even distribution of all five metals

In figure 6 the atoms in a surface is drawn randomly for each atom, which might result in splits between the atoms, which are not perfect, as seen in the composition in the top left plot. In figure 5, the split between the metals is ensured to be perfect according to the specified split. Comparing the two resulting energy distributions shows that they are almost identical, hence small changes in the composition does not change the results significantly.
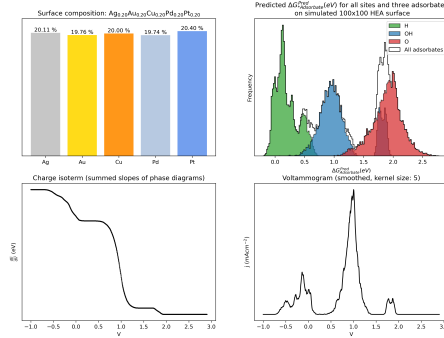
Figure 6: Results from the first method on a surface with an almost even distribution of all five metals

In figure 7, Ag is excluded from the surface, which results in different distributions of the ΔGs. Notably, the

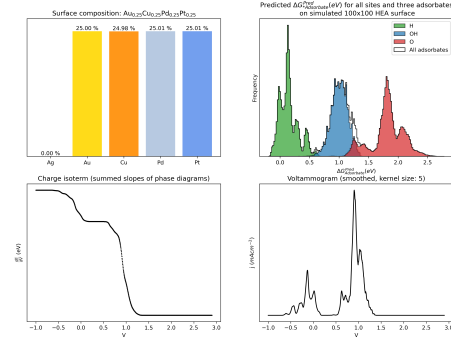OH peak at $\approx$ 1.8 eV is not present when Ag is not present in the surface.

Figure 7: Results from the first method from a surface without Ag

It is possible to create a surface with only a single metal. This results in all sites being identical, all feature vectors on each site being identical and hence all predicted ΔGs are identical for each adsorbate. This yields three delta functions, where in reality, the energies on identical sites might have a small spread.

3.3 Surface binding energies to voltammogram

In figure 5 and 7 the energies have been converted to voltammograms using the first method described. The bottom left plot shows the charge isoterms obtained from the energies and the bottom right plot shows the resulting voltammograms. Some features in the voltammogram resemble the energies, because the binding energies partially dictate which voltage is needed to make the binding energies negative.

3.4 Simulating surface adsorbates while varying voltage

In figure 8 to 12 cyclic voltammograms were simulated with the second method, where adsorbates are kept track of during a voltage sweep. The middle plots show the occupation of each adsorbate as a percentage of the maximum number of adsorbates if all sites were covered. This is useful to track which adsorbates adsorb and detach at which voltages, because the cyclic voltammogram only shows the current, not the source of the current. The voltammograms made with this method is cyclic because it is possible to use a circular voltage swipe. The simulated cyclic voltammogram of a surface with all five metals and no hysteresis or neighbour interactions is shown in figure 8. The cyclic voltammogram is in agreement with the voltammogram created with the first method on a similar surface with all metals present. This cyclic voltammogram is symmetric through a mirror plane along the linear function j = 0.
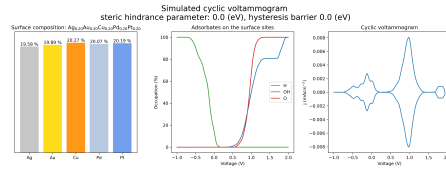

Figure 8: Results from the second method on a surface with an almost even distribution of all five metals

With this method, the hysteresis can be changed, as shown in figure 9. This results in adsorbates that adsorb at a different voltage than they detach at, leaving them at the surface for a longer time. This has an effect on the voltammogram, which is no longer symmetric. The effect is not just a simple shift in the cyclic voltammogram as expected. This might be due to the fact that the phase diagram has the summed binding energies on the y-axis and with the hysteresis energy barrier these can go above 0 eV. The observed behaviour is not suspected to be completely correct.
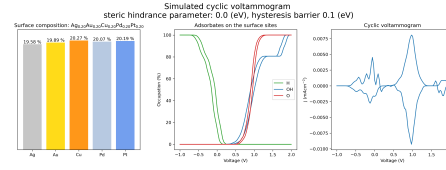

Figure 9: Results from the second method on a surface with an almost even distribution of all five metals and a hysteresis barrier of 0.1 eV

In figure 10, the steric hindrance parameter has been increased to 0.05, meaning each of the 9 "close neighbours" add 0.05 eV to the binding energy. This was supposed to make the last adsorbates require even higher or even lower voltages to adsorb, slowing the adsorption when the occupancy of the surface is high. However the adsorbates should adsorb and detach at the same voltages, which they do not. A bug is suspected to be causing this.
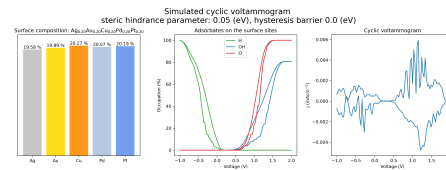

Figure 10: Results from the second method on a surface with an almost even distribution of all five metals and a steric hindrance parameter of 0.05 eV

In figure 11, the cyclic voltammogram has been simulated without the metal Ag. The results are similar to

the voltammogram made with the first method where Ag was excluded.
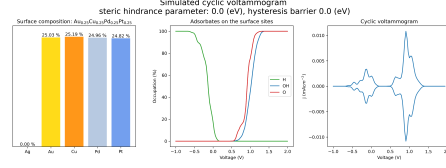

Figure 11: Results from the second method on a surface without Ag

In figure 12, a cyclic voltammogram based on a bi-metal surface consisting of Ag and Pd was simulated. It is apparent, that the energies are not as smoothly distributed as HEA energies. This might be attributed to the fact that the energy prediction model has been trained on mostly slabs where all or most metals are present in the atom positions represented in the feature vector. This probably makes it hard to extrapolate for the model, when only two metals are present, since this would be unlikely to find in the training data. It might also be due to the fact that fewer energies are present, when fewer metals are present. The less smooth energy distribution results in a more choppy cyclic voltammogram. It is hard to estimate how precise the results are without an experimentally measured cyclic voltammogram for comparison.
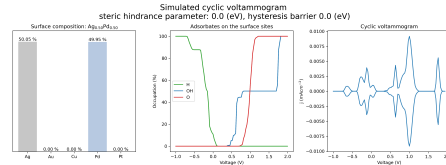

Figure 12: Results from the second method on a surface with only Ag and Pd

4   Conclusion

The energy prediction models perform well on the test data and yields precise energy predictions on the sites present in the test data, although the data set is small, especially lacking bi-metal/tri-metal slabs. Inputting identical sites in the model will give the exact same energy each time, whereas the energies in real life might follow a distribution with a small spread, which could be implemented. Voltammograms and cyclic voltammograms were simulated for varying surfaces with both presented methods, yielding agreeing results for similar surfaces and parameters. The second method has tunable hysteresis and steric hindrance. The hysteresis yields a realistic occupation but a questionable cyclic voltammogram. The steric hindrance similarly yields questionable results, since the occupation should slow down when approaching 100 % occupation and the cyclic voltammogram should have been symmetric. As further work several aspects of the project could be improved. Both the hysteresis and the steric hindrance parameter needs to be checked. More steric hindrance models could be implemented, preferably empirically backed models. When the hysteresis and steric hindrance implementations are ensured to be working properly, it would be interesting to fit the parameters to an experimental cyclic voltammogram. More complicated reactions could also be added by adding similar data sets for each new adsorbate to the training set. With the Gradient Boosting models an unlimited number of adsorbates can be added. Furthermore, adding more DFT data from slabs with 2-3 metals would be beneficial to the mod-

els ability to predict energies of adsorbates on bi- or trimetallic surfaces, since most training data is slabs with all metals.

5 Acknowledgements

6 Code

All code for this project is public at the GitHub repository: `https://github.com/MarcusNoergaardWeng/Git_PUK.git`. Note that most of the code for loading data from .db files and writing feature vectors from slabs was written by Jack Kirk Pedersen and repurposed for this project (several of the .py files in /scripts and 2_Write_features.ipynb)

References

[Che+15]   Tianqi Chen et al. "Xgboost: extreme gradient boosting". In: *R package version 0.4-2* 1.4 (2015), pp. 1–4.

[Ros+20]   Jan Rossmeisl et al. "Realistic cyclic voltammograms from ab initio simulations in alkaline and acidic electrolytes". In: *The Journal of Physical Chemistry C* 124.37 (2020), pp. 20055–20065.

[Cla+21]   Christian M Clausen et al. "What Atomic Positions Determines Reactivity of a Surface? Long-Range, Directional Ligand Effects in Metallic Alloys". In: *Advanced Science* 8.9 (2021), p. 2003357.