Department of mechanical engineering

MECH0020 Individual Project

AY 2022/23

Student:	**Fahim Ahmed**
Project Title:	**Using data science to optimise the stability of a catalyst**
Supervisor:	**Dr Enrique Galindo-Nava**

Word count: 7,484

University College London
Torrington Place
LONDON WC1E 7JE

Declaration

I, Fahim Ahmed, confirm that the work presented in this report is my own. Where information has been derived from other sources, I confirm that this has been indicated in the report.'

Abstract

Hydrogen fuel cell technology has not yet achieved 100% carbon free emissions due to current process of generating hydrogen. For the time being, hydrogen is being generating via reforming natural gas, hence releasing carbon. The cleanest method for hydrogen production involves water electrolysis which only releases H_2 and O_2 as products. Water electrolysis is currently being engineered to operate in vehicles to provide hydrogen fuel whilst in transit, hence leading to a full 'hydrogen economy'.

The main obstacle hindering the implementation of this process in a wide-scale is due to the stability of the catalyst involved in the process. Catalyst performance can drop by 20% within a 2 hours of operation, primarily due to dissolution of the species. To mitigate this problem, data science tools will be utilised to find stability trends and highly active metals. This included using 3rd party material science python packages such as pymatgen and matminer to extract further data. Also, scikit-learn was utilised to generate machine learning models, such as k-means clustering, where necessary.

The research began with a dataset containing 826 perovskites all undergoing oxygen evolution reaction (OER). Optimising OER has as much significance as hydrogen evolution reaction (HER), because the end products in OER involves a H^+ ions which is donated to the solution in the cathode, thus reacting with the OH^- product to form water. After modelling reaction intermediates and using API links to generate numerous amounts of pourbaix diagrams, it was concluded that the lanthanide series is recommended for further research into stability mitigations. Furthermore, the non-noble metals which were shown to have promising catalytic activity included iron, cobalt, or manganese:

Possible methods to incorporate lanthanides include creating a bimetallic oxide system or applying it as a coating. While the latter has already been experimented with cerium, there is still more research to be done with the other lanthanides as the pourbaix diagrams were more promising for the rest (e.g. La & Lu). Also, using bimetallic oxide systems gives advantages by allowing the best of stability from a

lanthanide in addition to the high activity from another metal. Although, lanthanides were proven to act as stable catalysts, further experimental results will be required as the pourbaix diagram involves limitations in temperatures and purity etc. Additionally, only perovskites were examined in this research, which had limited innovation into other potential design methods to improve stability.

Acknowledgements

"I would want to take this opportunity to thank everyone who helped finish this research project, especially the following people: For their crucial direction, counsel, and support during the study process, I thank my supervisor Dr Enrique Galindo-Nava and Ugne Kiudulaite who helped start me off with this project by gathering research beforehand. Their advice and skills have been crucial in forming this project. I also want to express my gratitude to the mechanical engineering department for letting me embark on this project."

Contents

List of figures

List of tables

1.Introduction

1.1. Background of project

Generating hydrogen efficiently can have a major significance in many applications, for example we can use hydrogen to power vehicles, generate electricity via nuclear energy, and create ammonia in the Haber process. Using hydrogen for such applications can reduce our reliance on fossil fuels. However, currently 95% of hydrogen is generated by reforming fossil fuels, hence creating CO_2 [Dubouis, N.(2019)]. A more environmental process is water electrolysis. The equation for the reaction is given below:

$H_2O \rightleftharpoons H_2 + \frac{1}{2}O_2$ (With a platinum catalyst)

With no greenhouse gas emissions, this will be the greatest alternative to replace the current situation of using natural gas reforming. Furthermore, with the abundance of water, electrolysis is considered a highly sustainable method for as a long-term method.

Platinum catalysts are used for a variety of reasons. One of which is its HBE (hydrogen binding energy). Hydrogen binding energy represents the strength of a material to adsorb hydrogen molecules if the value is too low it will not act as a catalyst but if it is too high the hydrogen will ‘stick’ with the metal for too long. Platinum has the optimum HBE value where it can act effectively with the hydrogen ions. Furthermore, platinum is ideal as it can remain stable for a longer duration in extreme environments (e.g. acidic environments) compared to other materials such as transition metals. However, it will be important to look into other suitable replacements as there is a scarce amount of platinum available as well as being extremely expensive. Also, platinum is centralised in certain parts of the world which will increase its difficulty to obtain in the UK.

1.2. Aims and objectives

This project aims to aid research in finding alternatives for platinum by analysing activity and stability features. Stability is a major issue as experiments have shown performance loss by approximately 20% after two hours of operation when an NPMC

(non-precious metal catalyst) is used [Banham, D.(2015)]. Activity will also be focussed as it is the main variable to determine if the catalyst operates.

Data science techniques will be utilised to proceed with this research. A dataset of over one-thousand catalysts with performance metrics were provided and will begin the process of finding factors which affect catalytic behaviour. Data manipulation and modelling will be achieved via python in Jupyter Notebook, where the basic packages which will be used include Pandas, NumPy, and Matplotlib. Other third-party packages include Scikit-learn (machine learning) & pymatgen (material property access). By taking advantage of all these technologies, the solution aims to provide suitable alternatives for platinum and validate optimum conditions for the reactions. The Jupyter notebook code file is given in the supporting materials submission. The code given in the submission cannot run without the initial dataset file but can be available upon request. Alternatively the whole folder can be downloaded via GitHub, where the link to the repository is given in appendix A [7.1].

2.Literature review

2.1. Principles of water electrolysis

The process of splitting water into hydrogen and oxygen through the use of electrical energy is known as water electrolysis. A typical setup for a water electrolysis unit includes: a cathode, anode, and an electric power supply. The cathode and anode are usually separated with an electrolyte which tends to be a solution of aqueous ions, a proton exchange membrane, or an oxygen ion exchange ceramic membrane. In order to create hydrogen, a direct current (DC) is supplied from the negative terminal of the DC source to the cathode, which serves as the site of the reduction process. The electrochemical reaction at the anode results in the production of electrons, which then travel back to the positive terminal of the DC source.

In an acid aqueous electrolyte, the reactions that take place at the electrodes, with potential differences, are given in Eq.1.1 and Eq.1.2 [Naimi Y and Antar A (2018)]:

$$Anode: 2H_2O \rightarrow O_2 + 2H^+ + 2e^- \qquad \mathrm{E}^0_{\mathrm{H_2O/H_2}} = 1.23\mathrm{V} \qquad (\mathrm{E.q.}\,1.1)$$

$$Cathode: 4H^+ + 2e^- \rightarrow H_2 \qquad \mathrm{E}^0_{\mathrm{H^+/H_2}} = 0\mathrm{V} \qquad (\mathrm{E.q.}\,1.2)$$

The half equations at each electrode can be written as:

$$Anode: 4OH^- \rightarrow O_2 + 2H_2O + 4e^- \qquad \mathrm{E}^0_{\mathrm{OH^-/O_2}} = 0.4\mathrm{V} \qquad (\mathrm{E.q.}\,1.3)$$

$$Cathode: 2H_2O + 2e^- \rightarrow H_2 + 2OH^- \qquad \mathrm{E}^0_{\mathrm{H_2O/H_2}} = -0.83\mathrm{V} \qquad (\mathrm{E.q.}\,1.4)$$

Therefore the overall reaction in water splitting is given in E.q.5:

$$H_2O \rightarrow H_2 + \frac{1}{2}O_2 \qquad \Delta\mathrm{E}^0 = \mathrm{E}^0_{\mathrm{H_2O/H_2}} - \mathrm{E}^0_{\mathrm{OH^-/O_2}} = -1.23\mathrm{V} \qquad (\mathrm{E.q.}\,1.5)$$

Because the global reaction is a negative value, this proves that the water electrolysis is not a spontaneous reaction, thus an external energy supply will be required to activate this reaction. Therefore, the reactions can also be written as:

$$water + electricity \rightarrow hydrogen + oxygen \qquad (\mathrm{E.q.}\,1.6)$$

Based on the products formed in each electrode, the reactions at the cathode are known as hydrogen evolution reactions (HER) and at the anode the reactions are denoted as oxygen evolution reactions (OER).

The schematic of a simple water electrolysis set up is shown in Figure 1.

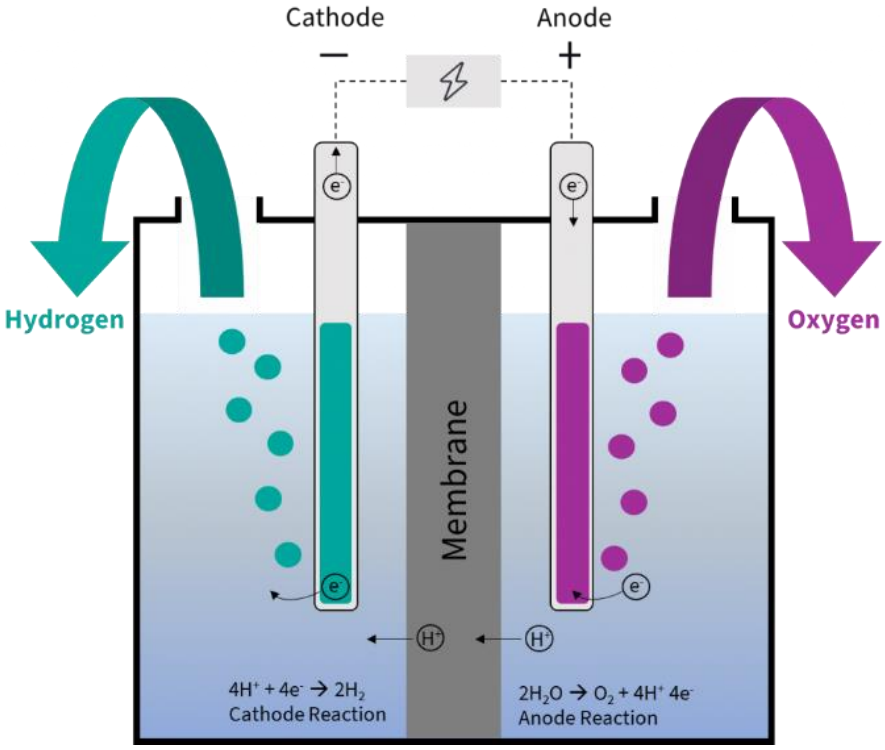


Figure 1 Schematic for a water electrolysis setup [Office of Energy Efficiency (2021)]

2.2. Instability in catalysts

Catalytic stability can be measured in terms of activity changes over a period of time, or by measuring the physicochemical properties of the catalyst. To measure activity changes, the degradation of voltage is recorded which represents how functional the

catalyst is. While physiochemical properties could include measuring how the chemical composition alters under operation. However, with a change in composition, it may not always negatively affect performance as products from the decomposition could cause an insignificant or beneficial effect to the activity [Spöri, C.(2017)].

2.2.1. Causes of instability

There are three primary causes for instability in catalysts [Banham, D.(2015)]:

1. Leaching of the metal sites – Catalysts can leach into the aqueous phase during the reaction, hence will not be effective in the long term. Occurs in both OER and HER.
2. Oxidative attack by H_2O_2. Hydrogen peroxide may be formed as a by-product in the oxygen reduction reaction at the cathode. This molecule is able to oxidise the active site of the catalyst and interrupt stability. Only occurs in HER.
3. Protonation which may be followed by anion adsorption of the active site – Due to the acidic environment, the solution is rich with H^+ ions which are capable of attacking the active site (protonation). Adding a hydrogen ion to some molecules may cause them to lose their active site. Only occurs in HER.

2.2.2. Parameters affecting stability

Many factors including properties of the catalyst and the conditions of the reaction can influence the stability of the material used. This comprises of [*Zeng, F. et al. (2022)*]:

<u>Effect of additives</u>

Integrating an additive (for example another element/component), can alter the dissolution and structural change of the catalyst, hence affecting the stability. For example, a study once incorporated titania (TiO_2) with iridium oxide (IrO_2) [Cherevko, S. *et al.* (2014)]. This experiment showed a decrease in the change of current density compared with the unmodified iridium oxide. This increase in stability may stem from a particle-support interaction between IrO_2 and TiO_2 which stabilises high Iridium oxidation states and prevents Ir^{2+} formations (dissolution).

Effect of temperature

Temperature has a major influence on the stability of catalysts, but optimum conditions are yet to still be discovered. For example, a stability rise was observed for nano-NiO_x in alkaline OER when the temperature was increased from 25℃ to 60℃ [Sayed, D.M.(2018)]. However, for the catalyst, f Ti/IrO_2 -Ta_2O_5, in an acidic electrolyte the stability decreased when increasing the temperature from 40℃ to 60℃ [Song, S.(2008)].

Effect of applied potential and current density

The applied potential and current density impact stability as the dissolution products formed is dependent on the voltage applied. One study researched the final products formed at different potentials when using iridium (Ir) as an electrolyte [Kasian, O.(2018)]. When Ir applying low potentials, iridium dissolved as Ir^{3+} in the electrolyte, with $HIrO_2$ as an intermediate product. Whereas in higher potentials, iridium dissolved into the electrolyte as IrO_4^{2-}, with IrO_3 as an intermediate product.

2.3. Reaction mechanisms

With the majority of the catalysts provided in the given dataset are OER catalysts, the reaction at the cathode will be analysed throughout. During the electrolysis process, the water splitting reaction forms intermediate products between the process of reactants and products. Analysis of intermediate products will be necessary to discover any products which may influence the activity of the catalyst. Furthermore, calculating the Gibb's free energy changes (ΔG) between two adsorbed intermediates can determine the feasibility of the reactions occurring.

There are many mechanisms which can be modelled for water splitting, one model presents the steps as shown below [E. Fabbri (2014)]:

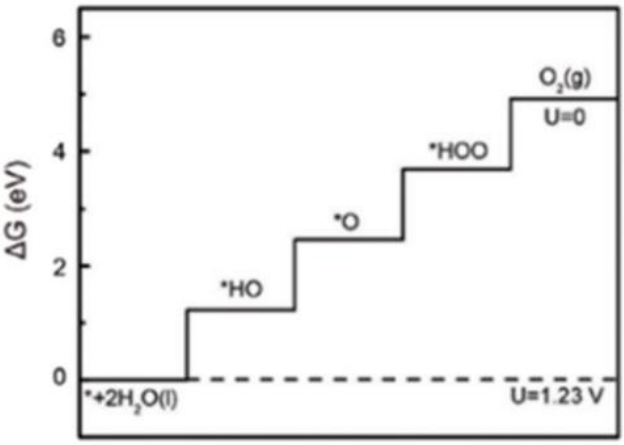
$$\mathrm{H_2O_{(l)} + M \rightleftharpoons HO^* + H^+ + e^-} \qquad (\mathrm{E.q.}\ 2.1)$$

$$\mathrm{OH^* \rightleftharpoons O^* + H^+ + e^-} \qquad (\mathrm{E.q.}\ 2.2)$$

$$\mathrm{O^* + H_2O_{(l)} \rightleftharpoons HOO^* + H^+ + e^-} \qquad (\mathrm{E.q.}\ 2.3)$$

$$\mathrm{HOO^* \rightleftharpoons M + O_{2_{(g)}} + H^+ + e^-} \qquad (\mathrm{E.q.}\ 2.4)$$

Where M characterises the catalytically active site, and the * denotes the chemical being bonded to the active site.



The reaction pathway showing the intermediate products can be visualised using a free energy diagram in Figure 2:

Figure 2 Free energy diagram for OER [E. Fabbri (2014)]

By separating the pathways as shown in Figure 2, the binding energies of each intermediate product ($\Delta G_{HO^*}, \Delta G_{O^*}, \Delta G_{OOH^*}$) can be determined. Binding energies must reach an optimum value as high values will attach the reactants too strong and will prevent further reactions, and low values will prevent reactions to occur. An ideal reaction pathway will typically have free energy changes of 1.23eV at every step (0eV overpotential) [Man, I.C (2011)]. Overpotential is the difference between the equilibrium potential for a given reaction and the potential at which the catalyst operates at a specific current. However, overpotential will always occur.

A universal linear scaling relationship between ΔG_{OOH^*} and ΔG_{HO^*} has been identified which can be applied to finding the catalyst with the minimum overpotential [Man, I.C (2011)]. The intercept of the linear trend will represent the difference in free energy change with the ideal value being 2.46eV in this case. This linear scaling indicates that the overall reaction can be represented by the adsorption energy of only one species involved, such as O^*. The O^* species can represent the activity of the overall reaction as function of oxygen binding energy, using either E.q.2.2 or E.q.2.3. The reaction step with the largest change in Gibbs free energy will indicate the potential determining step (PDS). The PDS highlights thermochemical aspects relating the OER overpotential.

The largest change in free energy change (hence PDS), can be considered to be E.q.2.2 in the reaction pathway [Wang, S (2021)]. Therefore, the difference between ΔG_{O*} and ΔG_{OH*}can be used as a universal descriptor for the activity of an OER. If this relationship was plotted onto a graph against overpotential, the catalyst with the

closest value of $\Delta G_{OH*} - \Delta G_{O*} = 1.23eV$ and an overpotential of 0V will give the optimum catalyst in terms of activity. Typically a volcano shaped scatter plot will be presented with the optimum catalyst at the tip.

2.4. Pourbaix diagrams

Analysing reaction pathways can help determine the optimum catalysts in terms of activity. However, as most catalysts cannot operate for long durations, studying activity can become meaningless as it could lead to excessive material consumption. Therefore stability is a main aspect to study along with activity. To optimise electrocatalysts reaction stability, it will be necessary to first optimise the conditions applied. Pourbaix diagrams is one of the most useful tools to measure the stability of a catalyst. These primarily focus on the long-term dissolution of the material.

A pourbaix diagram maps the regions of electrochemical stability for different redox states as a function of pH and applied potentials. An example for a pourbaix diagram can be shown in Figure 3:

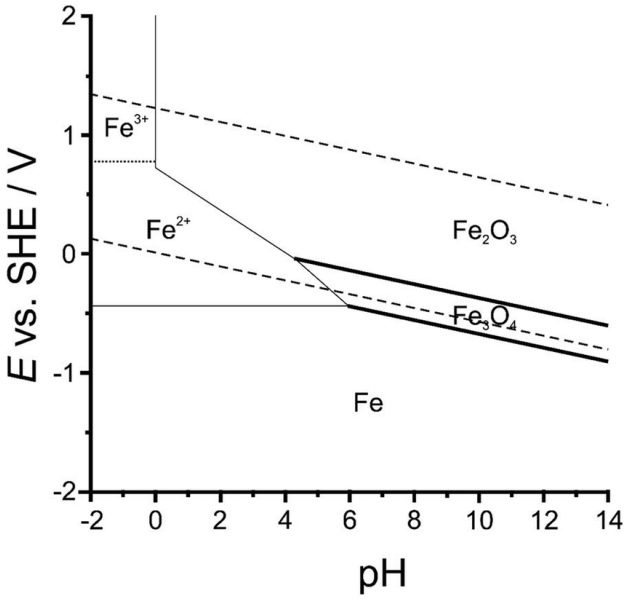


Figure 3 Pourbaix diagram for iron (Fe) (*Metallos, 2007)*

These diagrams give the ability to see how metal species transform in a solution depending on the pH & voltage applied to the electrode. The area between the dashed lines in the diagram represents the region of water stability. The bottom dashed line represents the hydrogen reduction line, whilst the top dashed line denotes water oxidation to O_2. Water is unstable below the bottom hydrogen line, hence, H_2 gas

evolves at the cathode in these conditions (HER). Similar to this, the anode produces O_2 gas above the oxygen line (OER).

In the case with Figure 3, it is visible that iron is slightly unstable in acidic media in the water stability region and under the hydrogen reduction line. However, it mostly remains stable under the hydrogen reduction line. Although oxidation of iron occurs in the water stability and oxygen evolution regions, it may favour the reactions at the electrodes as these products sometimes boost activity.

However, there are many limitations for using pourbaix diagrams given below [Ahmad, Z. (2006)]:

1) These are solely based on thermodynamic data, with no insights into the reaction itself. The information regarding the kinetics of the reaction could prevent achieving thermodynamic stability.
2) Only pure metals are considered which engineers are not very interested in.
3) It is assumed that all insoluble products are protected. As porosity, thickness, and substrate adhesion are significant elements that affect the protective efficacy of insoluble corrosion products, it is false to assume that all insoluble products are protective.
4) These diagrams are modelled in equilibrium conditions in a specified environment. However, variables such as temperature can act as a serious factor to corrosion rates.

Although the list of limitations is substantial, the benefits of applying pourbaix diagrams to research studies outweighs these disadvantages.

2.5. Material performances

Overall, to pick optimum catalysts in either OER or HER, parameters such as stability and activity are the main concern. Below are suitable options to act as a catalyst in water electrolysis.

2.5.1. Noble metals

Noble metals are the most popular catalysts in both HER and OER. In HER, the preferred metal is pure platinum due to its optimum hydrogen binding energy (HBE). HBE relates the metal-hydrogen bond strength, where a precise value should be met otherwise the hydrogen will bond too strong to the metal or not enough to proceed

with the reaction. When finding HBE values for a wide range of metals, it is found that noble metals (especially pure platinum) have the ideal value for application [Dubouis, N. and Grimaud, A. (2019)]. Furthermore, the stability for platinum in HER shows exceptional results. Usually experiments have shown performance loss by approximately 20% after two hours of operation when an NPMC (non-precious metal catalyst) is used [Banham, D.(2015)]. However, platinum can operate with no performance loss for >1200 hours [Greeley, J. et al. (2006)]. With this significant difference in stability results, it is evident that platinum is the perfect material currently as an electrode.

Contrary to the case of HER, where the favourable order for catalysts is platinum> ruthenium>Iridium, it occurs to be the opposite for the case of OER. For OER at the anode, the noble metals: iridium (Ir) and ruthenium (Ru) are the popular choices. Although ruthenium tends to be superior in activity, iridium and its oxides are typically preferred due to its slightly higher stability, both have similar stabilities to platinum (>1200 hours) [Greeley, J (2006)]. However, the limitation of applying Iridium is its scarcity, where its abundance is 1/40 times that of gold 1/10th of platinum [Obodo, K.O., Ouma, C.N. and Bessarabov, D. (2021)].

2.5.2. Bimetallic oxides

The concept of creating binary metal oxide catalysts rests on the belief that it is possible to merge the "beneficial properties of both components" into a single substance. One bimetallic oxide experimented involved combining the high stability of iridium oxide with the high activity of ruthenium oxide. This formulated $\mathrm{Ru}_x\mathrm{Ir}_{1-x}\mathrm{O}_2$, where x can vary the composition of each metal. It was found that the addition of each oxide to one another reduced the galvanostatic corrosion rate significantly. The optimal compromise between combining the two elements for activity and stability is when $0.5 < x < 0.8$ [E. Fabbri *et al.* (2014)]. There have been many combination attempts between the high stability of IrO_2 and other metals to improve its activity. One success trial involved combining the oxide with tin (Sn) to form $\mathrm{Ir}_x\mathrm{Sn}_{1-x}\mathrm{O}_2$, where stability was shown to have similar results to pure IrO_2.

2.5.3. Perovskites

A perovskite is a compound with the formula structure ABO_3. Where A is the larger cation to B. The B-cation is surrounded by 6 oxygen atoms, forming a BO_6 octahedral.

The perovskite structure allows cation substitution between A and B; hence the general formula could be shown as $(\mathrm{A}_x\mathrm{A}'_{1-x})(\mathrm{B}_y\mathrm{B}'_{1-y})\mathrm{O}_3$. This modification alters the perovskite electrical, magnetic, and optical properties thus will affect its activity performance as a catalyst. The outer orbitals of the A ions often have little bearing on the perovskite's electronic characteristics, which are thought to originate primarily from the BO_6 octahedral structure. However, this does not imply that the A ion is not significant. The conduction band's energy is influenced by the A ion's electrostatic potential, and its size dictates whether the crystal structure deviates from the ideal cubic shape. The outer s and p orbitals make up a full valence band and an empty conduction band when the B cation is a transition metal, respectively. As a result, they have little impact on the material's physical characteristics. Therefore the electrons of interest are in the d-band [E. Fabbri *et al.* (2014)]. This is evident from a research where the correlation of the number of d-electrons corresponded with the OER activity trend shown from a volcano plot in figure 4.

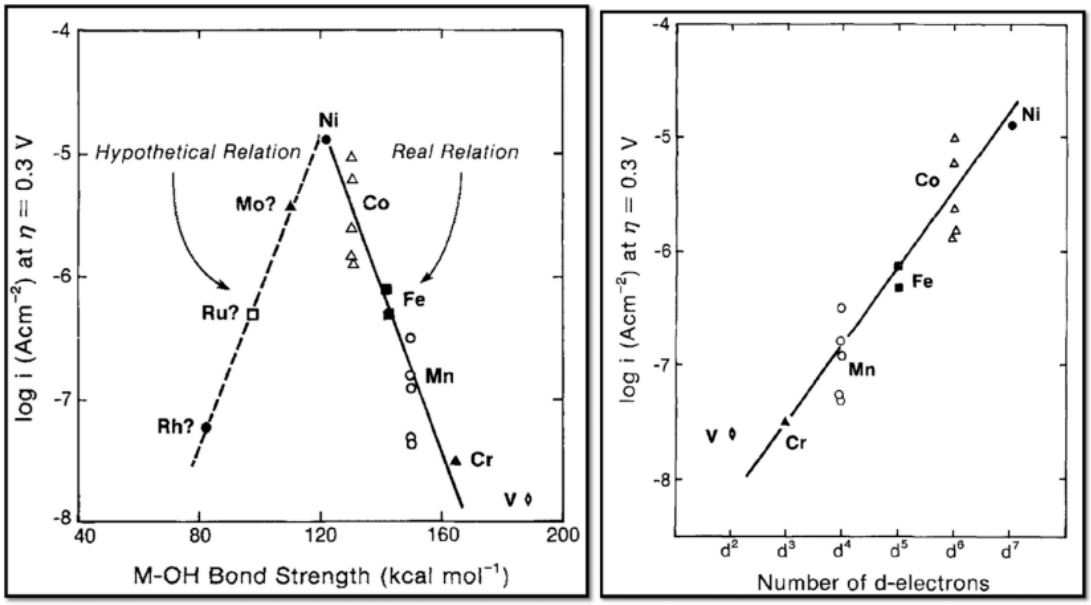


Figure 4 Volcano plot representing OER activity in comparison with the number of d-electrons [E. Fabbri et al. (2014)]

3.Methodology

3.1. Data inputting and cleaning

An initial dataset was given and imported into a Jupyter notebook. Using Python, the necessary modules required to be imported were: NumPy, Pandas, matplotlib, matminer, pymatgen and scikit-learn. The Pandas module was the key application to import, clean and modify the dataset. An overview of how the dataset is displayed is shown in Figure 5, showing all the column names and the number of catalysts present (1011):

	Catalyst Materials	Family	Substrate	Operating Conditions (T, pH, potential)	Valence state, A	Valence state, B	Ionic radius A	Ionic radius B	ΔGOH*	ΔGO*	ΔGOOH*	Adsorption strength	Stable hours/cycles	Tafel Slope, mV per decade	Overpotential at xx (nearby) current density vs RHE, V	Reaction (HER, HRR, OER, ORR)
0	(Mn-Co-Ta-Sb)Ox	NaN	(Mn-Co-Ta-Sb)Ox loaded on FTO	1M H2SO4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2 h @ 0.26 mA/cm2	-	0.55	OER
1	[Mo2S12]2-	Transition metal dichalcogenide	NaN	Acidic media	NaN	NaN	NaN	NaN	-	-	-	-0.05 eV (H)	NaN	NaN	NaN	HER
2	1T-MoS2	NaN	Exfoliated 2D 1T phase molybdenum disulfide lo...	0.5 M H2SO4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Potential increased by ~100 mV after 2 h @10 m...	322	0.42	OER
3	AgAlO3	Perovskite	-	Alkaline media	3.0	3.0	0.710	0.468	2.18	4.63	5.1	NaN	NaN	NaN	1.82	OER
4	AgBiO3	Perovskite	-	Alkaline media	1.0	5.0	1.061	0.760	1.09	3.37	4.23	NaN	NaN	NaN	1.55	OER
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1007	YTiO3	Perovskite	-	Alkaline media	3.0	3.0	0.988	0.670	1.49	3.97	4.61	NaN	NaN	NaN	1.54	OER
1008	YUO3	Perovskite	-	Alkaline media	3.0	3.0	0.988	1.025	0.81	2.59	4.09	NaN	NaN	NaN	1.72	OER
1009	YVO3	Perovskite	-	Alkaline media	3.0	3.0	0.988	0.640	0.61	1.7	3.9	NaN	NaN	NaN	1.34	OER
1010	YZnO3	Perovskite	-	Alkaline media	3.0	2.0	0.988	0.730	1.74	5.03	4.54	NaN	NaN	NaN	2.06	OER
1011	γ-MnO2	NaN	Gamma phase manganese\r\noxide loaded on FTO	1 M H2SO4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	8000 h at pH = 2	80	0.49	OER

Figure 5 Initial dataset

From this dataset, it is clear that some cleaning was required. Using the .info() function from python to investigate how many null values in each column (Figure 6):

```
RangeIndex: 1012 entries, 0 to 1011
Data columns (total 16 columns):
 #   Column                                                Non-Null Count  Dtype
---  ------                                                --------------  -----
 0   Catalyst Materials                                    1012 non-null   object
 1   Family                                                964 non-null    object
 2   Substrate                                             927 non-null    object
 3   Operating Conditions (T, pH, potential)               984 non-null    object
 4   Valence state, A                                      806 non-null    float64
 5   Valence state, B                                      806 non-null    float64
 6   Ionic radius A                                        806 non-null    float64
 7   Ionic radius B                                        806 non-null    float64
 8   ΔGOH*                                                 879 non-null    object
 9   ΔGO*                                                  879 non-null    object
 10  ΔGOOH*                                                879 non-null    object
 11  Adsorption strength                                   6 non-null      object
 12  Stable hours/cycles                                   76 non-null     object
 13  Tafel Slope, mV per decade                            112 non-null    object
 14  Overpotential at xx (nearby) current density vs RHE, V  959 non-null    float64
 15  Reaction (HER, HRR, OER, ORR)                         1000 non-null   object
dtypes: float64(5), object(11)
memory usage: 126.6+ KB
None
```

Figure 6 Null values in the dataset

The null values in each column were changed to zero values using functions that were built. Even though there are not many NaN values in Figure 6, some cells had '-' placeholders for null values which were afterwards converted to zeroes.

Another function was created to count the number of zeroes in each column as these now represented missing values, the result shows which columns have an inadequate amount of data to proceed with analysis. The following variables had fewer than 200 missing values, hence were saved for further exploration:

Family: The group/class of material the catalyst belongs to.
$(\Delta GO^*, \Delta GOH^*, \Delta GOOH^*)$: Represents the binding activity between the metal species and O^*, OH^*, OOH^* respectively.
Overpotential at xx (nearby) current density vs RHE, V: difference between the equilibrium potential for a given reaction and the potential at which the catalyst operates xx current.
Reaction (HER, HRR, OER, ORR): Type of reaction occurring at the catalyst.

An outline of how the dataset currently appears is shown in Figure 7:

	Catalyst Materials	Family	ΔGOH*	ΔGO*	ΔGOOH*	Overpotential at xx (nearby) current density vs RHE, V	Reaction (HER, HRR, OER, ORR)
0	(Mn-Co-Ta-Sb)Ox	0	0.00	0.00	0.00	0.55	OER
1	[Mo2S12]2-	Transition metal dichalcogenide	0.00	0.00	0.00	0.00	HER
2	1T-MoS2	0	0.00	0.00	0.00	0.42	OER
3	AgAlO3	Perovskite	2.18	4.63	5.10	1.82	OER
4	AgBiO3	Perovskite	1.09	3.37	4.23	1.55	OER
...	...	...	...	...	...	...	...
1007	YTiO3	Perovskite	1.49	3.97	4.61	1.54	OER
1008	YUO3	Perovskite	0.81	2.59	4.09	1.72	OER
1009	YVO3	Perovskite	0.61	1.70	3.90	1.34	OER
1010	YZnO3	Perovskite	1.74	5.03	4.54	2.06	OER
1011	γ-MnO2	0	0.00	0.00	0.00	0.49	OER

1012 rows × 7 columns

Figure 7 Dataset after cleaning

To explore the type of catalysts left in the dataset, the 'Family' and 'Reaction' column was grouped and counted. The following results were given in Table 1 & Table 2:

Catalyst family	Number of catalysts
Perovskite	826
Unknown	48

Bimetallic oxide	26
Noble metal	23
Transition metal dichalcogenide	18
Noble metal oxide	16
Other	55

Table 1 Catalysts families from dataset

Reaction type	Number of catalysts under corresponding reaction
Oxygen evolution reaction (OER)	953
Hydrogen evolution reaction (HER)	47
Unknown	12

Table 2 Reaction types from dataset

From Table 1 & Table 2, it is evident that the research conducting will be analysing stability behaviours based on perovskite structures. Also, the main focus will be optimising stability in OER (anode).

3.2. Using APIs for further data acquisition

The python libraries 'matminer' and 'pymatgen' was employed at this stage. With 'matminer', all the strings in the catalyst material column were converted to a composition type, thus creating a new column showing all the elements involved in the composition. For example, the compound '$AgAlO_3$' was converted to the form 'Ag-Al-O_3'. This function will benefit later when it will be necessary to analyse elemental composition trends.

Using 'Pymatgen', data from The Materials Project [Anubhav Jain (2013)] was necessary, so an API link was generated in the python code to retrieve some data. One issue that arose when importing the catalysts names/data was the merging errors into the current dataset. For example, on the current dataset, a compound could be labelled as '$AgAlO_3$' but from the materials project database it may be labelled as '$AlAgO_3$'. This caused a huge amount of data to not be merged into the current dataset as python cannot interpret these chemical strings correctly. Therefore, functions were written to convert the strings from The Materials Project to match the current dataset, i.e. converting '$AlAgO_3$' into '$AgAlO_3$'.

The use of pymatgen allowed us to match the catalyst materials to the corresponding material ID from the materials project database. The Materials Project API was advantageous as it has access to numerous amounts of compounds. With the material ID, it will now be possible to look up and access a pourbaix diagrams for any material in the dataset. The code for using matminer, pymatgen and cleaning the formula strings are present in appendix B [7.2].

3.3. Modelling

3.3.1. Modelling reaction steps

It is necessary to calculate how feasible and active the catalysts are to ensure top selections. By studying only the perovskites and bimetallic oxides in the dataset, the first step was to explore all the reaction pathways in the OER. From part 2.3 in the literature review section, this can be found by examining the Gibb's free energy change of all the intermediates. From the research conducted earlier, the binding energy of the oxygen element can act as a universal indicator for activity. This gives the choices to interpret equation 2.2 or 2.3. This corresponds to either using the difference of ($\Delta GOOH^* - \Delta GO^*$) or ($\Delta GO^* - \Delta GOH^*$). To determine which reaction to act as the universal indicator for activity in this case, the reaction with a greater average across the whole dataset shows it acts as the rate determining step in the overall process. Table 3 shows the average values of each reaction throughout the data.

Reaction pathway	Average Gibbs free energy change (eV)
$\Delta GOOH^* - \Delta GO^*$	1.419759
$\Delta GO^* - \Delta GOH^*$	1.772154

Table 3 Averaging values for reaction pathways

From Table 3, it is evident that $\Delta GO^* - \Delta GOH^*$ should be used as it has a higher average, hence will act as the rate determining step. The overpotential will be plotted against the reaction pathway to determine the optimum catalyst in terms of activity. This was a method used in previous studies as the ideal overpotential should be 0 and the reaction pathway should be 1.23eV for OER [Wang, S (2021)]. Figure 8 represents this plot below generated using matplotlib on python.

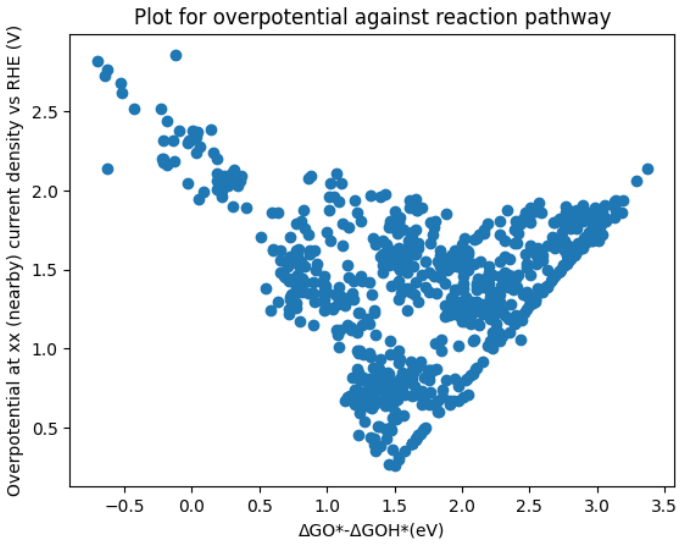


Figure 8 Plot for overpotential against reaction pathway

As expected from the literature review, a volcano plot is present with the optimum catalyst at the bottom of the curve.

3.3.2. K-means clustering

After plotting the overpotential against the reaction pathways in Figure 8, the next approach was to separate the catalysts into regions. This will segregate the optimum catalyst from the rest of the data to identify any trends that occur in different groups. Scikit-learn machine learning package on python was utilised for clustering. K-means clustering was the algorithm choice in categorising as it is efficient and appropriate for the simple model in Figure 8, due to having no outliers.

Sensitivity analysis

To identify the number of regions to input for the k-means clustering algorithm, a sensitivity analysis must be conducted. A sensitivity analysis involves measuring the errors of the output of the clustering algorithm based on the number of clusters being inputted. One approach is to use the elbow method.

For k-means clustering, the number of clusters have to be specified which generates the same number of cluster points (centroids) in the plot. One method to show the errors in this model is by calculating the distance between the data points and the closest cluster point. After the error is calculated for each data point, the results are squared and summed. This is known as calculating the sum of squared errors (SSE). Typically, an elbow plot shows the sum of squared errors against the number of

clusters used in the model. A general formula to measure the SSE is given in equation 3.0.

$$SSE = \sum_{i=1}^{K} \sum_{x \epsilon C_i} distance^2(m_i, x) \qquad \text{E. Q. 3.0}$$

Where x is the data point within cluster C_i and m_i is the centroid position for cluster C_i. K represents the number of clusters. The '$distance^2(m_i, x)$' represents a function to calculate the distance between the data point and centroid, proceeded by squaring it. This followed by summing all these values within that cluster. Followed by another summation function to sum all the errors in each cluster.

The SSE was calculated through python simply by using the 'inertia' attribute within the kmeans module in scikit-learn. Figure 9 shows the results for SSE when the clusters are ranged from 1 to 10.

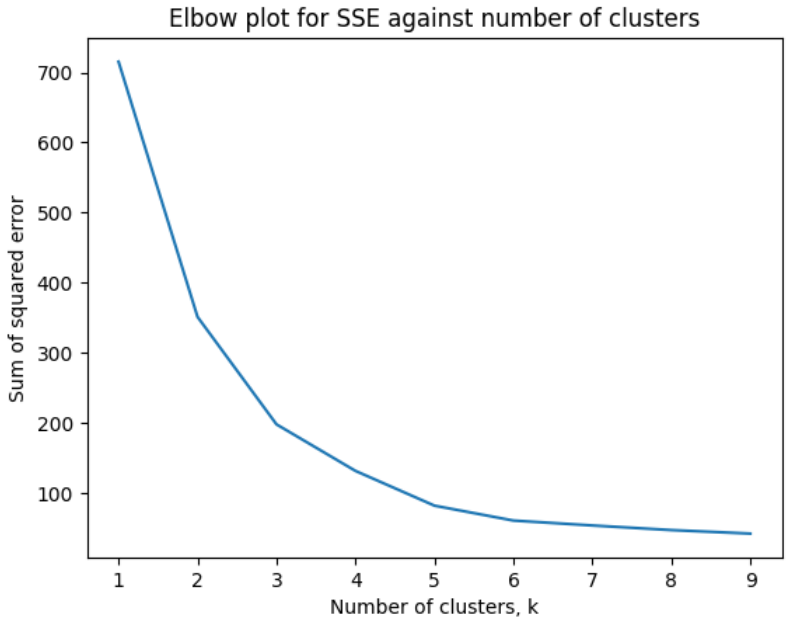


Figure 9 Elbow plot for SSE against number of clusters (k)

Generating model

From the elbow plot in Figure 9, the number of clusters shall be determined by locating the point of inflection in the graph. Therefore, the number of clusters used in k-means clustering model will be 5. Figure 10 shows the result of using 5 clusters for the unsupervised machine learning model.

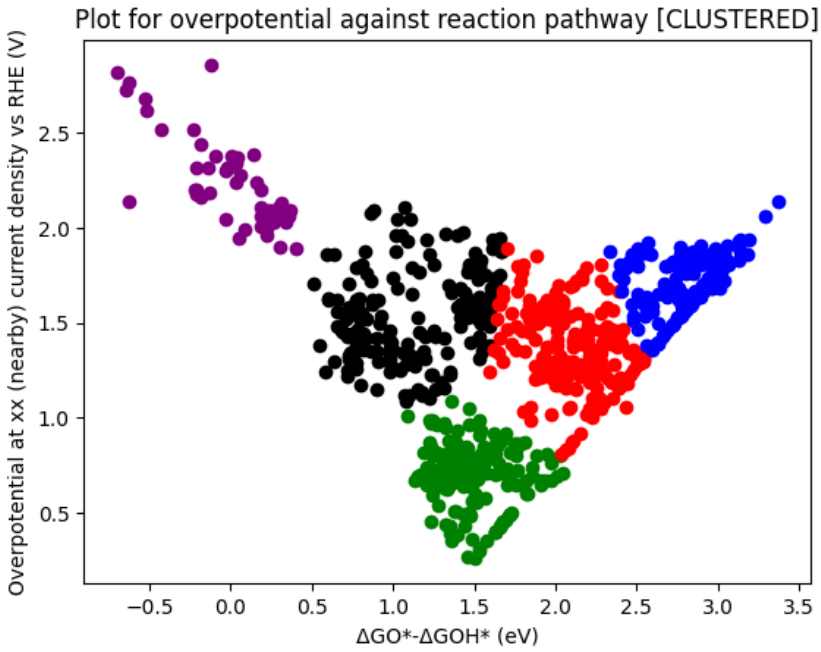


Figure 10 Plot for overpotential against reaction pathway after clustering algorithm applied

The optimum region in this case is the green indicator in the model, whereas the lowest performing catalyst are located in the blue and purple regions. The code for generating the elbow plot and k-means clustering model is given in appendix B [7.2].

3.4. Composition analysis in different regions

After separating the catalysts into different regions, the catalyst families within each region were monitored. It was found that all 26 bimetallic oxides were found in the optimum region with the rest of the catalysts being perovskites. Furthermore, a composition analysis was done amongst each group. Where each type of atom was added up and compared from every region to find atoms which represent the highest activities. This was achieved by listing the compounds in the whole dataset, followed by separating the elements and summing each type.

3.5. Generating pourbaix diagrams

To study stability results, it was decided to explore how/if these catalysts dissolute into the solution. Pourbaix diagrams were chosen to assess the catalysts stability. To retrieve these diagrams, The Materials Project API was utilised again [Anubhav Jain (2013)]. The function to generate the pourbaix diagram in python is shown in appendix B [7.2].

3.5.1. Comparison of stability in different regions

Pourbaix diagrams will need to be analysed qualitatively when presented. However, with a large dataset it is inefficient to inspect every material. Therefore to simplify this, a sample of 10 catalysts were randomly generated on python from each region. When analysing the pourbaix diagrams, the favoured catalysts were ones which remained in the same composition at any potential difference and pH value above the top dashed line mentioned in part 2.4 (as oxygen evolution occurs beyond this line). Catalysts which broken down into a different composition but remained full solid were denoted as partially stable as further research on the new composition will be required to assess how feasible it is as an electrode.

3.5.2. Comparison of single-ion stability to multi-element

To visualise how stability is affected when elements are combined together, pourbaix diagrams were generated for a number of cases. The following points gives an outline of which compounds/elements were compared against each other.

- Comparing how pourbaix diagrams alter when combining different metals together. This will be accomplished by studying single elements separately, then when combined together. For example, with '$FeNdO_3$'. 'Fe' will be studied, followed by 'Nd', then finally 'Fe-Nd'.
- Observing the effects of adding oxides to an element. For example, comparing the pourbaix diagrams of 'Nd' and 'Nd-O' to see what changes occurs.
- Comparing the pourbaix diagrams of all the lanthanide series to evaluate which one has the highest stability.

4.Results and discussion

4.1. Composition results for different regions

From the clustered plot in Figure 10, the number of elements in each region is added together and the top element in each section is plotted in a bar chart as shown in Figure 11 below.

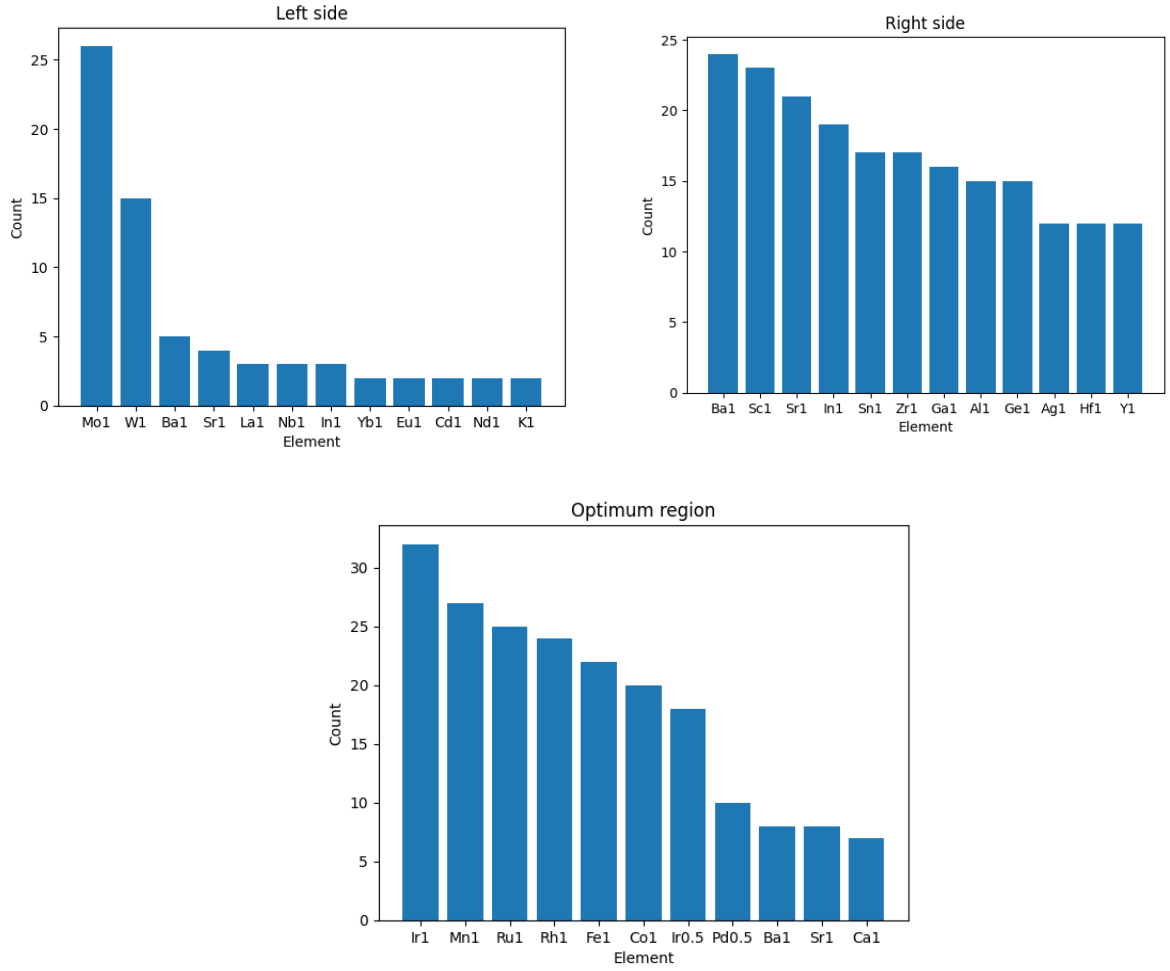


Figure 11Bar charts to show top number of elements in each region

Where the number after the element denotes if it comes from a perovskite [1] or a bimetallic oxide [0.5]. The right side and left side represent the blue and purple regions from Figure 10 respectively which act as the lowest performing catalysts. The other two regions were neglected as they contained a mixture of elements from the strongest and weakest catalyst groups, hence did not show significant results. Appendix C [7.3] reveals the rest of the bar charts for the remaining group and the full composition analysis.

Firstly, the lowest performing elements were identified. This included elements such as Tungsten, Molybdenum etc, therefore these elements are not ideal to act as a supporting material for a catalyst as they will affect its activity.

From Figure 11, it is evident that Iridium is a perfect option as a supporting catalyst as it has over 50 counts (including perovskites and bimetallic oxides) in the optimum region. As expected from the literature review in part 2.5, the noble metals have taken up the majority of the compounds in the optimum group for activity. Furthermore, platinum is not observed in the optimum group as predicted from part 2.5.1 in the literature review, where iridium and ruthenium are more active for OER. Although, from the literature review it was found that ruthenium was more active than iridium. This may be the case as more iridium compounds were sampled in the dataset.

However, these results cannot fully determine if an element is highly/weakly active, as they may appear evenly throughout different groups. Table 4 below presents a few which occur in different regions.

Element	Left side	Middle left	Middle right	Optimum region	Right side
Silver (Ag)	2	4	12	6	12
Barium (Ba)	5	5	11	8	24
Calcium (Ca)	2	4	11	7	12

Table 4 Elements which appear in multiple groups

These elements may be advantageous to study to support finding active metals. When the elements are evenly spread across all the groups, this could imply that the metal it is attached to has a strong influence on its properties acting as a catalyst. The common metals which were attached to these metals and results in them being amongst the weakest regions (right and left side in the clustered plot) include: aluminium, titanium, hafnium etc. All these metals were noted to be weak materials for catalytic activities.

The same method was applied to find the top selection of supporting elements for activity. Metals attached to calcium, barium and silver in the optimum region were listed. This list consisted of iron, cobalt, iridium, manganese, ruthenium, and rhodium. Other metals were not added to the list as they also appeared in the weaker regions, hence will be unreliable in practice. For example, vanadium combines with barium in the optimum region, however, it appears 26 times in the other groups, therefore it is not always reliable to influence the whole compound.

Between all the metals found to act as the optimum supporting catalyst, iron, cobalt, and manganese should be focussed on further research as these are not noble metals. This is due to the fact that it is desirable to study materials which are more abundant and affordable.

4.2. Comparisons of different regions

4.2.1. Random generated formulae

10 random compounds were generated in each region followed by constructing their pourbaix diagrams to see stability trends. Both acid and alkali regimes were investigated to find optimum conditions for perovskites. Below gives a brief study of each region. All pourbaix diagrams are given in appendix D [7.4]. Table 5 presents all the randomly generated formulae for each region.

Region	Formula				
Optimum region	$CdMnO_3$	$RhTbO_3$	$LuMnO_3$	$FeLuO_3$	$LaTiO_3$
	$BaMnO_3$	$CrSrO_3$	$CaORu_3$	$MnNdO_3$	$FeNdO_3$
Middle right	$BiHoO_3$	$BaBiO_3$	$HoNiO_3$	$CuLaO_3$	$AuBaO_3$
	$CsIO_3$	$LuNiO_3$	$AlYbO_3$	$DyNiO_3$	$BaCuO_3$
Middle left	$CrHoO_3$	$PbVO_3$	$NaPuO_3$	$DyVO_3$	$KOsO_3$
	$NbYbO_3$	$LaTaO_3$	$SmVO_3$	$CdVO_3$	$KTaO_3$
Left	$BaMoO_3$	$InTaO_3$	$NaWO_3$	$CaWO_3$	$AlBaO_3$
	$TlWO_3$	$BaNbO_3$	$ScSiO_3$	$MoNaO_3$	$SrVO_3$
Right	$AlGdO_3$	$BaZrO_3$	$SrTiO_3$	$ScYbO_3$	$BaScO_3$
	$InNdO_3$	$CuYO_3$	$PbSrO_3$	$EuGeO_3$	$BaInO_3$

Table 5 Randomly generated formulae for each region

4.2.2. Stability analysis for all the regions

A common trend visible between all the regions studied is their optimum conditions required for stable products to be formed. In general, dissolution was not occurring when the compound is under alkaline conditions. Table 6 shows which chemicals do not dissolve into solution. Where ‘full stability’ represents the compound not undergoing any solid transformation, and ‘fully solid’ indicates the chemical undergoing a solid transformation but does not dissolve (e.g. $AlGdO_3$ converting to $Gd_3Al_5O_{12}$). Despite a fully solid transformation alters the chemical, sometimes the altercation benefits the activity, therefore further research on these new solid compounds will need to be conducted.

Optimum		Middle right		Middle left		Right		Left	
Full stability	Fully solid	Full stability	Fully solid	Full stability	Fully solid	Full stability	Fully solid	Full stability	Fully solid
$FeNdO_3$	$FeLuO_3$	$CuLaO_3$	$BiHoO_3$	$OsKO_3$	$LaTaO_3$		$AlGdO_3$		$InTaO_3$
$LuMnO_3$	$LaTiO_3$	$CsIO_3$	$HoNiO_3$	$NaPuO_3$	$YaNbO_3$		$CuYO_3$		$AlBaO_3$
	$MnNdO_3$		$BiBaO_3$		$KTaO_3$		$PbSrO_3$		
	$RhTbO_3$		$LuNiO_3$				$TiSrO_3$		
			$DyNiO_3$				$BaInO_3$		
			$BaCuO_3$				$BaZrO_3$		
							$BaScO_3$		

Table 6 Stability results in different regions

Figure 12 shows one example of a pourbaix diagram with a stable configuration. The stable composition is visible for $E > 1V$ & pH>11 because oxygen evolution occurs above the top dashed line.

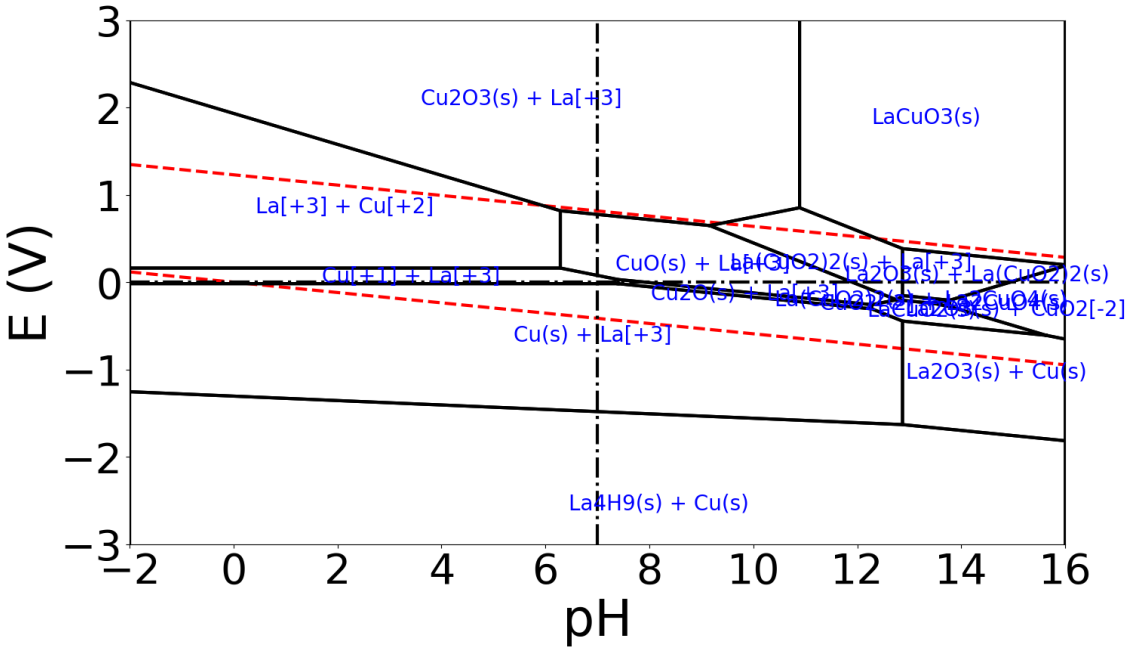


Figure 12 Pourbaix diagram for $LaCuO_3$

From the results in Table 6, it is evident that remaining in the same composition while undergoing OER is a rare phenomenon as only 10% of the compounds achieved this. Although no trend is observed from analysing the stronger regions to the weaker regions, this study was not aimed to build a correlation between activity and stability. But to find any other chemical trends. For instance, when scanning over the formulae in Table 6, it was discovered that the majority of them consisted of a lanthanide. With this information, it is valid to assume lanthanides can contribute to tackle stability issues. However, when referring back to appendix C [7.3] composition analysis, all the lanthanides are spread to different regions, therefore if these metals are applied to a system, they will need to be combined with highly active metals to improve reactivity.

4.3. Comparison of single element to compound

To compare how a single element stability is affected by combining it with other elements, the composition Fe-Nd-O will set an example.

4.3.1. Metal oxide with pure metal form

Figure 13 presents pourbaix diagrams for Nd and Nd-O.

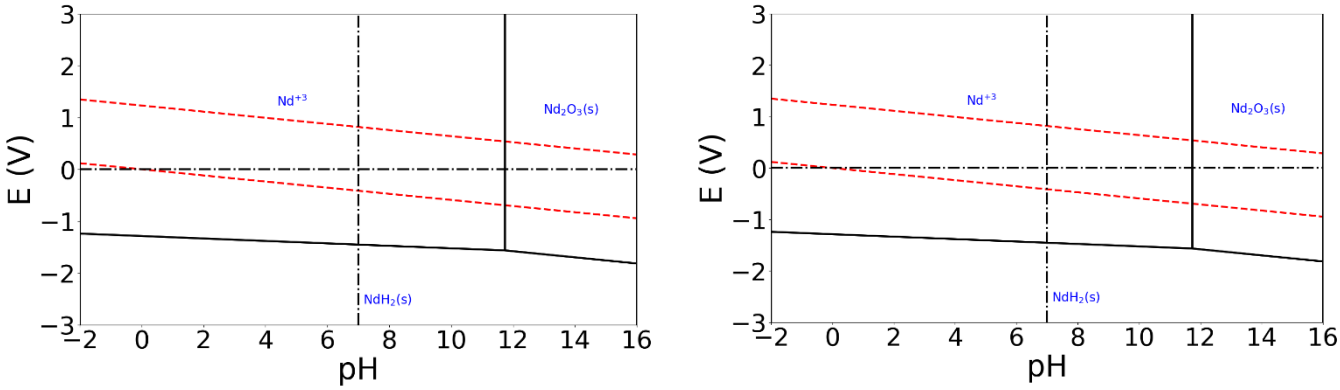


Figure 13 Pourbaix diagram comparing a metal with its metal oxide

As evident from these diagrams, oxidising a metal will not affect its stability. This is due to the fact pourbaix diagrams are solely based on thermodynamic behaviour; however, the formation of the oxide layer will not change the thermodynamic property of the metal oxide composition.

4.3.2. Different metals combined together

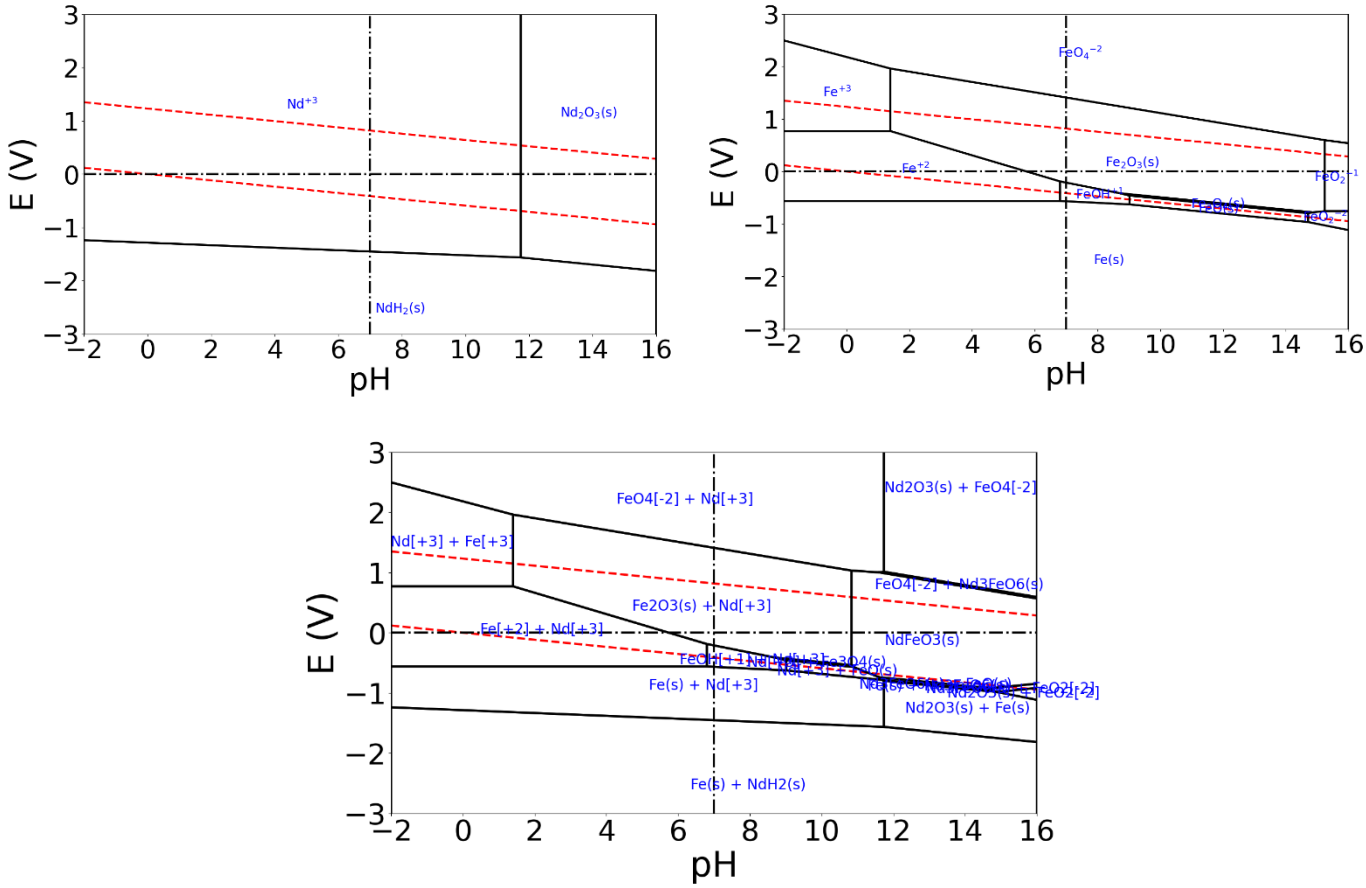


Figure 14 Pourbaix diagram for bi-metallic systems

Figure 14 shows the pourbaix diagrams for Fe (top-right), Nd (top-left), and Fe-Nd (bottom). Comparing each diagram, it is possible to determine the stability of a bi-metallic compound when each individual diagram is given. This can be done by 'overlapping' each pourbaix diagram with each other, as the end products is determined by adding together the chemicals at any position. For example, between --2<pH<1 and 1V<E<2.5V, Fe and Nd exists as Fe^{3+} and Nd^{3+} respectively. Therefore, the Fe-Nd end stability result will simply become Fe^{3+}+Nd^{3+}. With this information, it becomes clear that the most stable products can be found by finding elements with the largest area of stability in their pourbaix diagrams.

4.4. Lanthanide series stability analysis

The pourbaix diagrams for all the whole lanthanide series was generated. The top elements were determined by visually inspecting how big the area for the product M_2O_3 appeared [M = Metal]. Appendix D [7.4] displays all the pourbaix diagrams for the whole series.

In particular, the top 5 lanthanides in terms of stability are:

1. Lutetium (Lu)
2. Terbium (Tb)
3. Thulium (Tm)
4. Erbium (Er)
5. Holmium (Ho)

Although the pourbaix diagrams look similar between Tb to Ho, the diagram for Lu significantly differs to that of the rest. Lutetium covers a vast amount of area in alkali conditions compared to any other element.

Figure 15 presents the difference between Lutetium and Terbium.

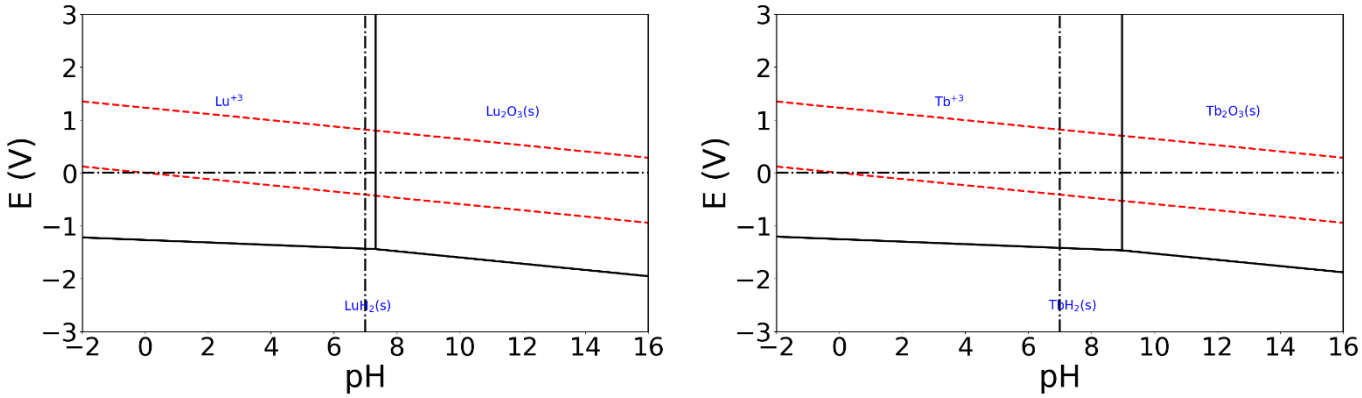


Figure 15 Pourbaix diagram for Lutetium (Left) & Terbium (right)

Despite the fact Lutetium has great stability results, if it is combined with other metals, the corresponding element will need to have satisfactory stability.

4.5. Theoretical solutions

Due to lanthanides generally showing promising results in stability, it is recommended that this series should start being implemented into current research for OER electrode materials. Design concepts can be established based on current experimentations for optimising catalysts.

One idea stimulates from utilising bimetallic oxide consisting of a lanthanide + highly active metal. With a lanthanide composition the systems stability will significantly increase, whilst the highly active metal will boost oxygen evolution. The general formula of a bimetallic system will consist of $L_xM_{1-x}O_2$. Where L is the lanthanide and M is another metal. Varying the value of *x* can allow alterations in stability/activity (increasing *x* increases stability in this case). Suitable candidates for highly active

metals can be found in part 4.1 which includes iron, cobalt, and manganese. Although in part 4.4 lutetium was decided to be the main lanthanide to focus on, other metals in the series should still be trialled as it can lead to sufficient results. In particular, Figure 14 presents an Nd-Fe-O system that has a stable region above the oxygen evolution dashed line.

Another solution involves utilising lanthanides to deploy a protection layer on the anode. This involves coating the electrode via spraying or dipping. Followed by undergoing heat treatment to promote a stable lanthanide oxide layer. Current experimentations have already considered using Cerium (Ce) to deposit CeO_2 to prevent dissolution/corrosion on electrodes [Zand, R.Z. (2015)]. These experiments have been successful in which CeO_2 has led to improved barrier properties and corrosion protection. However, this is the only lanthanide only being applied currently, even though the stability of many other lanthanides were proven to be greater in part 4.4 in the results. Therefore, other lanthanides such as lutetium, lanthanum etc. should be trialled experimentally as it is predicted to form better results.

4.6. Verification and validation

4.6.1. Limitations

The following points expands upon areas which could have limited other possible solutions.

Firstly, only perovskites were considered in this research. Even though a great amount of useful data was generated which helped acquire further information about each individual elemental property, numerous additional materials could be investigated. For example, graphene and other non-metallic materials have been researched useful to improve stability of current catalysts such as RuO_2[Bao, X. (2018)].

Furthermore, information on bimetallic oxides is very limited in research overall. As bimetallic systems allow variations in properties based on composition, it would be beneficial to analyse effects of varying the elements to see the feature differences. However, data acquisition for bimetallic systems is scarce so further experimental study would be required.

As mentioned in part 2.4, the limitations in pourbaix diagrams could have caused inaccurate results. For example, the pourbaix diagrams constructed operated at approximately 25℃ [Persson, K.A.(2012)] whereas typical water electrolysis occurs between 100-200℃ [Li, W.(2022)]. As corrosion is more likely in higher temperatures, the stability in the results have been overestimated hence may produce inaccurate solutions. In addition, pourbaix diagrams do not consider kinetics. If there were intermediate products between a catalyst and its final product, there is lack of information to show how long they are present. Therefore, if these intermediates are present for a long duration, the catalyst can be viewed as unstable.

Many catalysts composition alters as a result of water electrolysis. However, these are not concluded as unstable because it is possible the final products can benefit the activity of the reaction. Therefore, further experimental data would be required to assess the results.

4.6.2. Origin of data

Majority of data gathered for the binding energies ($\Delta GO^*, \Delta GOH^*, \Delta GOOH^*$) and overpotential were generated primarily from a machine learning model (decision tree) [Wang, X. *et al.* (2020)]. Decision trees require inputs (known as features) to predict further data for other catalysts. For example, ionic radii and valence states were inputted for all materials. As these features correlate with the binding energies, it is possible for a decision tree to compare these variable and predict unknown binding energies. The model began with 124 perovskites with experimental data. After the decision tree was modelled a further 702 perovskites were predicted.

Advantages of using decision tree models include its capability to predict variables for a wide range of catalysts which prevents experimental exhaustion. However, it can prevent observing unnatural trends if too many features are inputted, leading to over-fitting. When over-fitting occurs, it leads to the untrained data correlating with the trained data too strong, hence causing inaccuracies. For instance, this could have been the cause of the strictly straight lines in Figure 8 in the volcano plot. Typically a volcano plot with perfectly straight boundaries as shown in Figure 8is unnatural with experimental data, however, with machine learning models the correlations will be strict to fit the trends in the data.

4.6.3. Testing accuracy for predicted data

To validate the accuracy of the binding energies, the linear scaling between two of these variables will be sufficient to test. From literature, the linear relation between the binding energies of HO^* and O^* should have a slope of one half. This is due to oxygen having two bonds to the surface whereas HO^* only has one. Moreover, another linear relationship between HO^* and HOO^* exists with a unity slope due to both molecules having one bond between oxygen and the surface. Therefore, a linear regression model was generated in python for these variables,

Figure 16 and Table 7 displays the correlation and slope values for both linear relationships. The colours in Figure 16corresponds to the cluster plot in Figure 10.

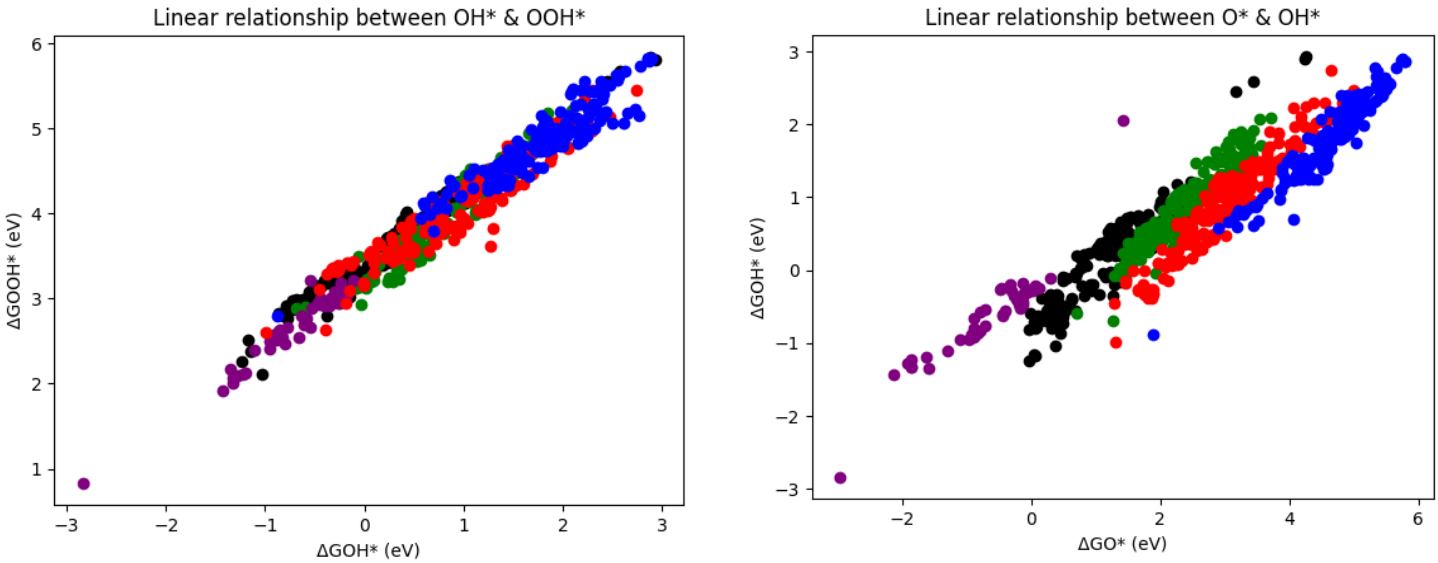


Figure 16 Linear relationship between intermediate products.

Intermediate variables to correlate	Slope value
OOH^* & OH^*	0.816452
OH^* & O^*	0.531304

Table 7 Slope values for linear relationships

With a percentage error of 6.26% between OH^* & O^*, this is a reasonable margin and valid for application. However, a percentage error of -18.45% between OOH^* & OH^* may highlight some concerns. Fortunately, the majority of the data applied in this research was using OH^* & O^* intermediates so the methodology will still be acceptable.

5.Conclusion

With the current situation in catalytic stability for water electrolysis, it is near impossible to create a clean, sustainable source for hydrogen. Currently, the main issue to solve is the dissolution of materials in solution. However, with utilising data science tools such as machine learning and APIs, it provides support to discover common stability trends.

Before studying compounds with highest stability, the materials with highest activity should be discovered to determine the most feasible chemicals in the water electrolysis. Due to having 826 perovskites all operating under OER, the reactions as the anode were analysed. A volcano plot was generated through examining the reaction step between OH^* & O^* . Using k-means clustering, the optimum catalysts were grouped away from the rest of the dataset to observe which elements had the greatest influence in activity. As a result the following metals are the most common non-noble metals: iron, cobalt, and manganese. Following this result, the stability of all the clustered regions were looked into by creating pourbaix diagrams for random samples. In addition to this, the lanthanide series were abundant in the list of stable compounds from the random samples, thus these elements were noted to be features that could enhance stability in a system.

With lanthanides noted down as the key factors to optimise the stability of a catalyst, two theoretical solutions have been designed which will need further experimental study. One solution includes incorporating lanthanides and one of the highly active metals (iron, cobalt, or manganese) in a bimetallic system. This will ensure high stability and activity. Another solution was using lanthanides as a coating against dissolution. Currently, cerium is applied in research but other lanthanides such as lutetium show promising results in this report.

The research conducted was only limited to using perovskites for study, however, there is a variety of materials that can be studied to optimise stability. For instance, carbon-based compounds such as graphene have been applied to optimise stability. Another area not studied was electronic configurations and molecular structure, which are the main aspects giving perovskites high reputation. Therefore in future study, non-metallic materials, electronic properties such as ionic radii and crystal structures

will need to be observed in hopes to find further stability trends. Nonetheless, using lanthanides should accelerate progression in replacing water electrolysis as the primary source of hydrogen production.

6.References

Dubouis, N. and Grimaud, A. (2019) The hydrogen evolution reaction: From material to interfacial descriptors, Chemical Science. The Royal Society of Chemistry. Available at: https://pubs.rsc.org/en/content/articlelanding/2019/sc/c9sc03831k#!divCitation .

Banham, D. et al. (2015) A review of the stability and durability of non-precious metal catalysts for the oxygen reduction reaction in proton exchange membrane fuel cells, Journal of Power Sources. Elsevier. Available at: https://www.sciencedirect.com/science/article/abs/pii/S0378775315004632

Naimi Y and Antar A (2018) Hydrogen Generation by Water Electrolysis. Advances In Hydrogen Generation Technologies. InTech. DOI: 10.5772/intechopen.76814

Office of Energy Efficiency & Renewable Energy (2021): 'Hydrogen Production: Electrolysis', Available: https://www.energy.gov/eere/fuelcells/hydrogen-production-electrolysis

Metallos (2007) Pourbaix diagram of iron in uncomplexed media (anions other than OH- not considered). Ion concentration 0.001 m (mol/kg water). Temperature 25°C. Available at: File:Fe-pourbaix-diagram.svg - Wikimedia Commons

Spöri, C. and Kwan, J.T.H. (2017) The stability challenges of oxygen evolving ... - wiley online library. Available at: https://onlinelibrary.wiley.com/doi/full/10.1002/anie.201608601

Wang, S., Lu, A. & Zhong, CJ. Hydrogen production from water electrolysis: role of catalysts. Nano Convergence **8**, 4 (2021). https://doi.org/10.1186/s40580-021-00254-x

Man, I.C. et al. (2011) "Universality in oxygen evolution electrocatalysis on oxide surfaces," ChemCatChem, 3(7), pp. 1159–1165. Available at: https://doi.org/10.1002/cctc.201000397 .

Ahmad, Z. (2006) "Basic concepts in corrosion," Principles of Corrosion Engineering and Corrosion Control, pp. 9–56. Available at: https://doi.org/10.1016/b978-075065924-6/50003-9 .

Zeng, F. et al. (2022) "Stability and deactivation of OER electrocatalysts: A Review," Journal of Energy Chemistry, 69, pp. 301–329. Available at: https://doi.org/10.1016/j.jechem.2022.01.025 .

Cherevko, S. et al. (2014) “Stability of nanostructured iridium oxide electrocatalysts during oxygen evolution reaction in acidic environment,” Electrochemistry Communications, 48, pp. 81–85. Available at: https://doi.org/10.1016/j.elecom.2014.08.027 .

Sayed, D.M. et al. (2018) “Activation/deactivation behavior of nano-Niox based anodes towards the OER: Influence of temperature,” Electrochimica Acta, 276, pp. 176–183. Available at: https://doi.org/10.1016/j.electacta.2018.04.175 .

Song, S. et al. (2008) “Electrochemical investigation of electrocatalysts for the oxygen evolution reaction in PEM water electrolyzers,” International Journal of Hydrogen Energy, 33(19), pp. 4955–4961. Available at: https://doi.org/10.1016/j.ijhydene.2008.06.039 .

Kasian, O. et al. (2018) “The common intermediates of oxygen evolution and dissolution reactions during water electrolysis on Iridium,” Angewandte Chemie International Edition, 57(9), pp. 2488–2491. Available at: https://doi.org/10.1002/anie.201709652 .

Greeley, J. et al. (2006) “Computational high-throughput screening of electrocatalytic materials for hydrogen evolution,” Nature Materials, 5(11), pp. 909–913. Available at: https://doi.org/10.1038/nmat1752 .

Obodo, K.O., Ouma, C.N. and Bessarabov, D. (2021) “Low-temperature water electrolysis,” Power to Fuel, pp. 17–50. Available at: https://doi.org/10.1016/b978-0-12-822813-5.00003-5

E. Fabbri et al. (2014) Developments and perspectives of oxide-based catalysts for the oxygen evolution reaction, Catalysis Science & Technology. Royal Society of Chemistry. Available at: https://pubs.rsc.org/en/content/articlehtml/2014/cy/c4cy00669k#imgfig4

Anubhav Jain et al, "Commentary: The Materials Project: A materials genome approach to accelerating materials innovation", APL Materials 1, 011002 (2013). https://doi.org/10.1063/1.4812323

Bao, X. (2018) “Preparation and characterization of ti/RGO-ruo2 electrode for hydrogen and oxygen evolution reactions,” *International Journal of Electrochemical Science*, pp. 7870–7881. Available at: https://doi.org/10.20964/2018.08.45 .

Persson, K.A. *et al.* (2012) “Prediction of solid-aqueous equilibria: Scheme to combine first-principles calculations of solids with experimental aqueous states,” *Physical Review B*, 85(23). Available at: https://doi.org/10.1103/physrevb.85.235438 .

Li, W. *et al.* (2022) “Low-temperature water electrolysis: Fundamentals, progress, and new strategies,” *Materials Advances*, 3(14), pp. 5598–5644. Available at: https://doi.org/10.1039/d2ma00185c.

Wang, X. *et al.* (2020) “First-principles based machine learning study of oxygen evolution reactions of perovskite oxides using a surface center-environment feature model,” *Applied Surface Science*, 531, p. 147323. Available at: https://doi.org/10.1016/j.apsusc.2020.147323.

Zand, R.Z., Verbeken, K. and Adriaens, A. (2015) “Synthesis and evaluation of self-healing cerium-doped silane hybrid coatings on steel surfaces,” *Intelligent Coatings for Corrosion Control*, pp. 135–194. Available at: https://doi.org/10.1016/b978-0-12-411467-8.00005-2 .

7.Appendix

7.1. Appendix A: GitHub repository link

GitHub link: https://github.com/fahim-ah/Yr3-Project.git

7.2. Appendix B: Key functions in python code

Matminer composition + The Material Project data acquisition functions

Element composition featurisation

```
#Creating a column for element composition
str_comp = StrToComposition(target_col_id='composition')
updated_data = str_comp.featurize_dataframe(updated_data, col_id='Catalyst Materials')
```

[ ] Python

Accessing material ID from material project + other properties

```
'''
This block will use an API key to gather material information from The Materials Project
'formula='**O3'' gathers data for perovskites
Data of interest that was retrieved include material_id and chemsys
print(data_list) prints all the material properties possible to gather
'''

with MPRester('CmGarHKLtPCjVpEkivPDYkhRJKJiB8A7') as mpr:
    results = mpr.summary.search(formula='**O3', fields=["material_id", "formula_pretty",'band_gap','chemsys'])
    data_list = [(result.material_id, result.formula_pretty, result.band_gap,result.chemsys) for result in results]
    #print(data_list)
```

[ ] Python

Elbow plot + K-means clustering

```
#Creating an elbow plot to determine how many cluster will be required for model

def elbow_plot(elbow_data):
    k_rng = range(1,10)
    sse = []
    for k in k_rng:
        km = KMeans(n_clusters= k)
        km.fit(elbow_data[['ΔGO*-ΔGOH*','Overpotential at xx (nearby) current density vs RHE, V']])
        sse.append(km.inertia_)     # Calculates SSE for each cluster
    print(sse)
    plt.xlabel('Number of clusters, k')
    plt.ylabel('Sum of squared error')
    plt.title('Elbow plot for SSE against number of clusters')
    plt.plot(k_rng,sse)
    plt.savefig('Scatter_plots\Elbow plot .png',bbox_inches='tight')
    plt.show()
```

[ ] Python

```
#Creating cluster plot using K-means method
def cluster_df(cluster_df,n_clusters):
    color = ['black','green','purple','red','blue']
    for k in range(0,n_clusters):
        data = cluster_df[cluster_df["cluster"]==k]
        plt.scatter(data['ΔGO*-ΔGOH*'],data["Overpotential at xx (nearby) current density vs RHE, V"],c=color[k])
        # Save datasets
        if k==0:
            data.to_csv('Cluster_csv_files\Cluster_0.csv')
        elif k==1:
            data.to_csv('Cluster_csv_files\Cluster_1.csv')
        elif k==2:
            data.to_csv('Cluster_csv_files\Cluster_2.csv')
        elif k==3:
            data.to_csv('Cluster_csv_files\Cluster_3.csv')
        else:
            data.to_csv('Cluster_csv_files\Cluster_4.csv')
    plt.xlabel('ΔGO*-ΔGOH* (eV)')
    plt.ylabel('Overpotential at xx (nearby) current density vs RHE (V)')
    plt.title('Plot for overpotential against reaction pathway [CLUSTERED]')
    plt.savefig('Scatter_plots\Overpotential against reaction step (with regions).png',bbox_inches='tight')
    plt.show()
```

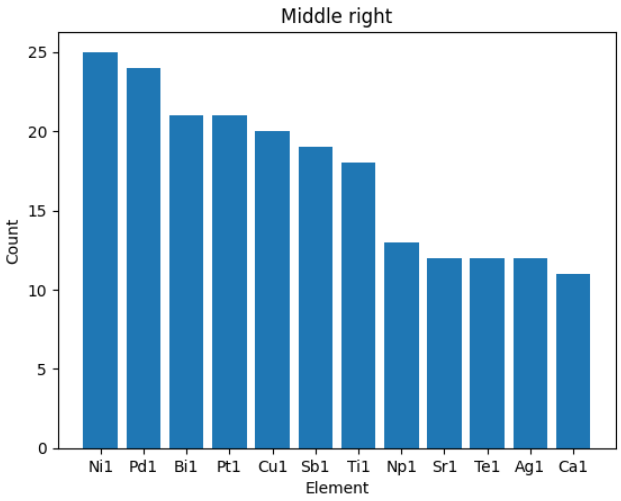
Function to generate pourbaix diagrams

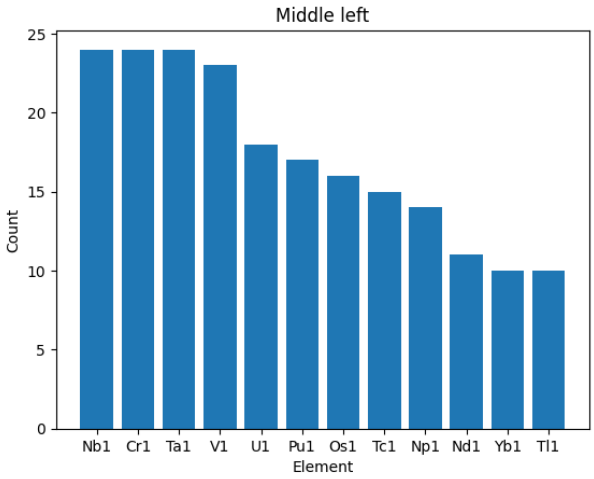
```python
# A repeatable function to generate a pourbaix diagram
# Input equals to the chemical composition and a path to save the figure

def generate_pourbaix(chemsys,path):
    with MPRester(api_key='CmGarHKLtPCjVpEkivPDYkhRJKJiB8A7') as mpr:
        pourbaix_entries = mpr.get_pourbaix_entries(chemsys)
    pd = PourbaixDiagram(pourbaix_entries)
    plotter = PourbaixPlotter(pd)
    PourbaixPlotter.get_pourbaix_plot(plotter,limits=[[-2, 16],[-3, 3]]) # Typical value for lower = -2, upper = 16
    plt.savefig(path+chemsys+'.png',bbox_inches='tight')
```

Python

7.3. Appendix C: Composition analysis

Remaining bar charts





Full elemental analysis

	Region				
Element	**Left side**	**Middle left**	**Middle right**	**Optimum region**	**Right side**
Ag1	2	4	12	6	12
Al1	1	2	8	0	15
Au1	0	0	2	0	1
B1	0	0	1	0	0
Ba1	5	5	11	8	24
Be1	0	0	0	1	0
Bi1	0	0	21	0	1
Ca1	2	4	11	7	12
Cd1	2	6	8	6	7
Ce1	1	7	8	6	6
Co0.5	0	0	0	3	0
Co1	0	2	2	20	0
Cr1	0	24	0	3	0
Cs1	0	0	1	0	0
Cu1	0	0	20	0	2
Dy1	2	9	9	6	7

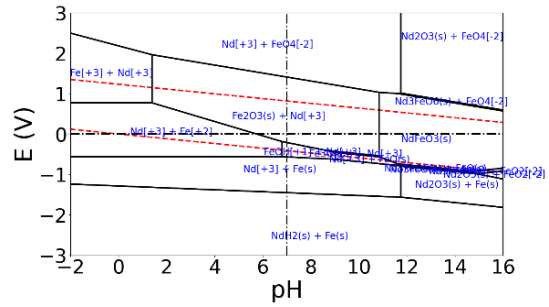
Er1	1	6	7	6	4
Eu1	2	10	9	6	5
Fe0.5	0	0	0	1	0
Fe1	0	0	2	22	0
Ga1	0	1	0	0	16
Gd1	1	7	6	6	4
Ge1	0	0	2	0	15
Hf1	1	3	5	3	12
Hg1	0	0	0	0	2
Ho1	1	7	6	6	4
I1	0	0	1	0	0
In1	3	0	0	0	19
Ir0.5	0	0	0	18	0
Ir1	0	0	0	32	0
K1	2	6	6	4	8
La1	3	9	8	6	7
Li1	0	0	1	0	4
Lu1	1	7	6	6	4
Mg1	0	0	0	0	10
Mn1	0	0	0	27	0
Mo0.5	0	0	0	4	0
Mo1	26	1	0	0	0
Na1	2	7	5	5	4
Nb1	3	24	0	1	1
Nd1	2	11	9	5	6
Ni1	0	0	25	0	0
Np1	0	14	13	0	0
Os1	0	16	0	0	0
Pa1	0	1	0	0	1
Pb1	2	9	7	6	8
Pd0.5	0	0	0	10	0
Pd1	0	0	24	0	0
Pr1	1	7	8	6	8
Pt0.5	0	0	0	1	0
Pt1	0	0	21	1	0
Pu1	0	17	7	0	0
Rb1	2	7	7	4	0
Re1	0	7	9	0	0
Rh0.5	0	0	0	7	0
Rh1	0	1	1	24	0
Ru1	0	0	0	25	0
Sb1	0	0	19	1	2
Sc1	1	0	0	1	23
Si1	1	0	1	0	2
Sm1	2	10	11	5	4
Sn1	0	1	0	0	17
Sr1	4	5	12	8	21
Ta1	1	24	2	0	1

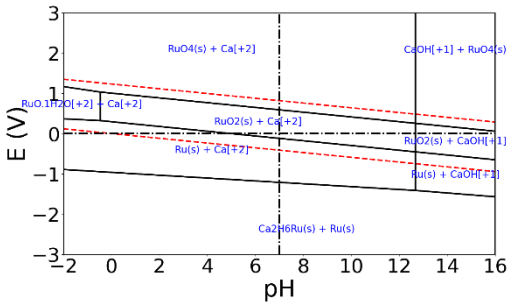
Tb1	1	7	7	6	8
Tc1	0	15	1	0	0
Te1	0	0	12	0	0
Th1	0	4	5	3	3
Ti1	1	3	18	1	9
Tl1	2	10	8	6	5
Tm1	1	9	9	6	6
U1	0	18	7	0	0
V1	2	23	1	3	0
W1	15	0	0	0	0
Y1	1	4	7	6	12
Yb1	2	10	9	6	5
Zn1	0	0	0	0	10
Zr1	0	0	0	0	17

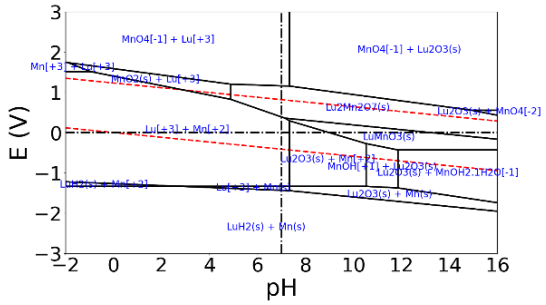
7.4. Appendix D: Pourbaix diagrams studied

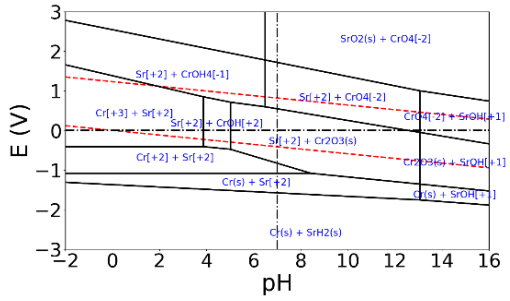
Pourbaix diagrams studied for every region

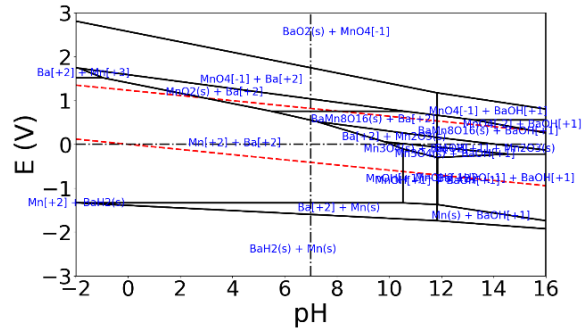
Optimum region

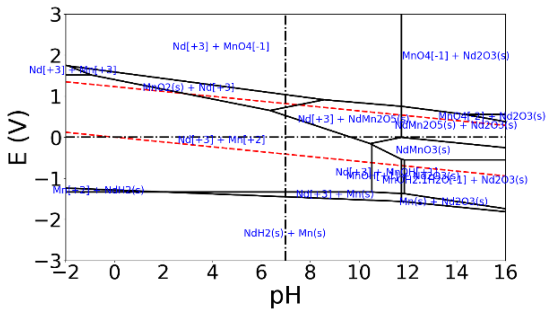


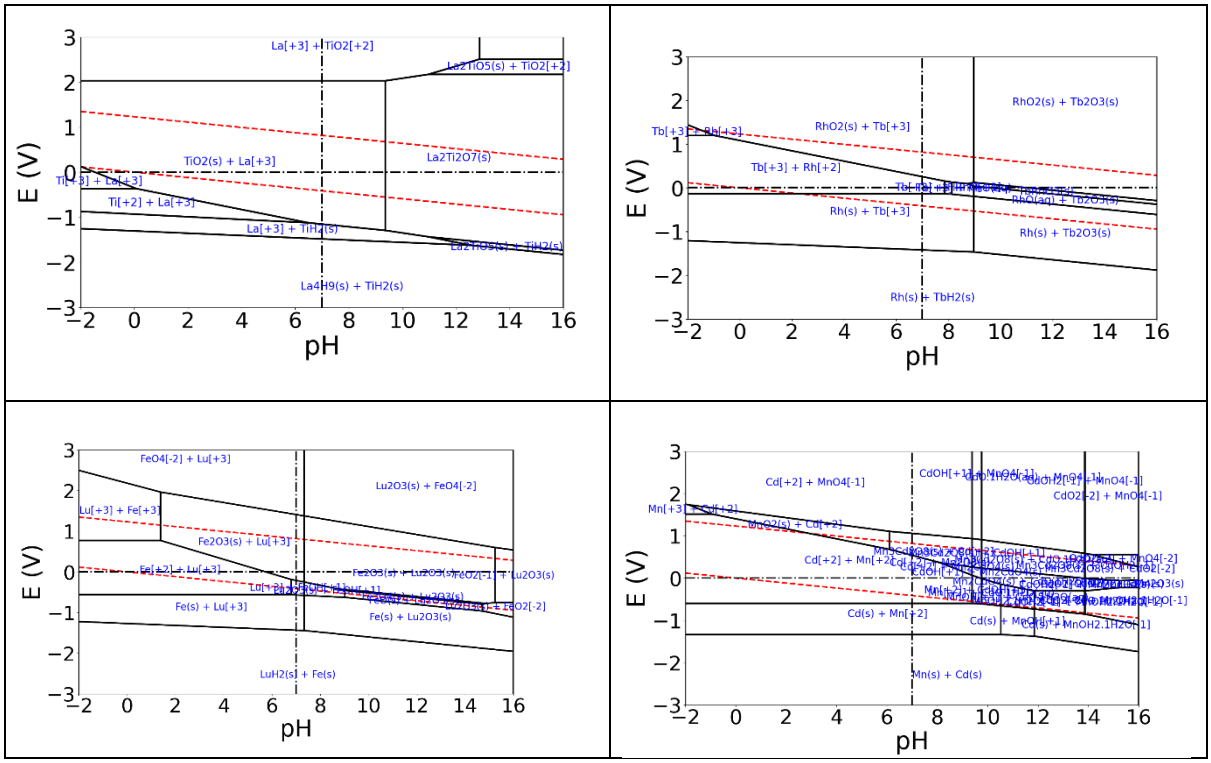




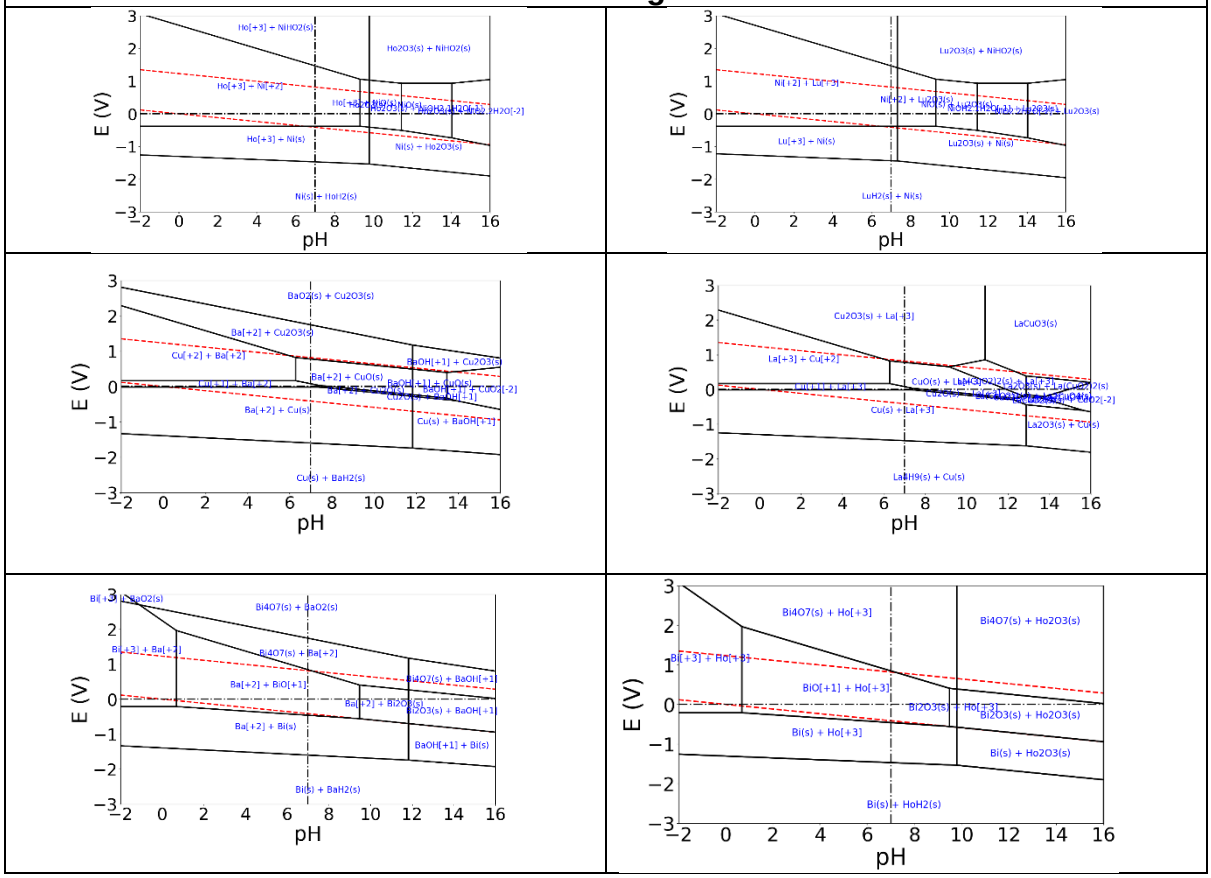


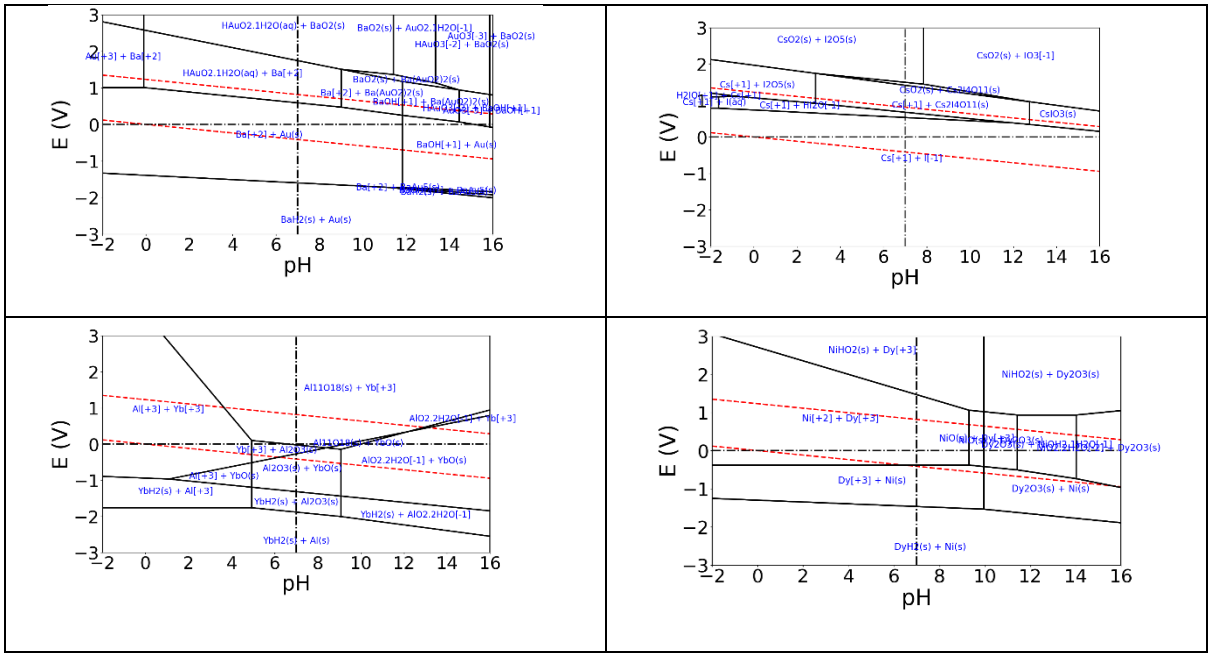




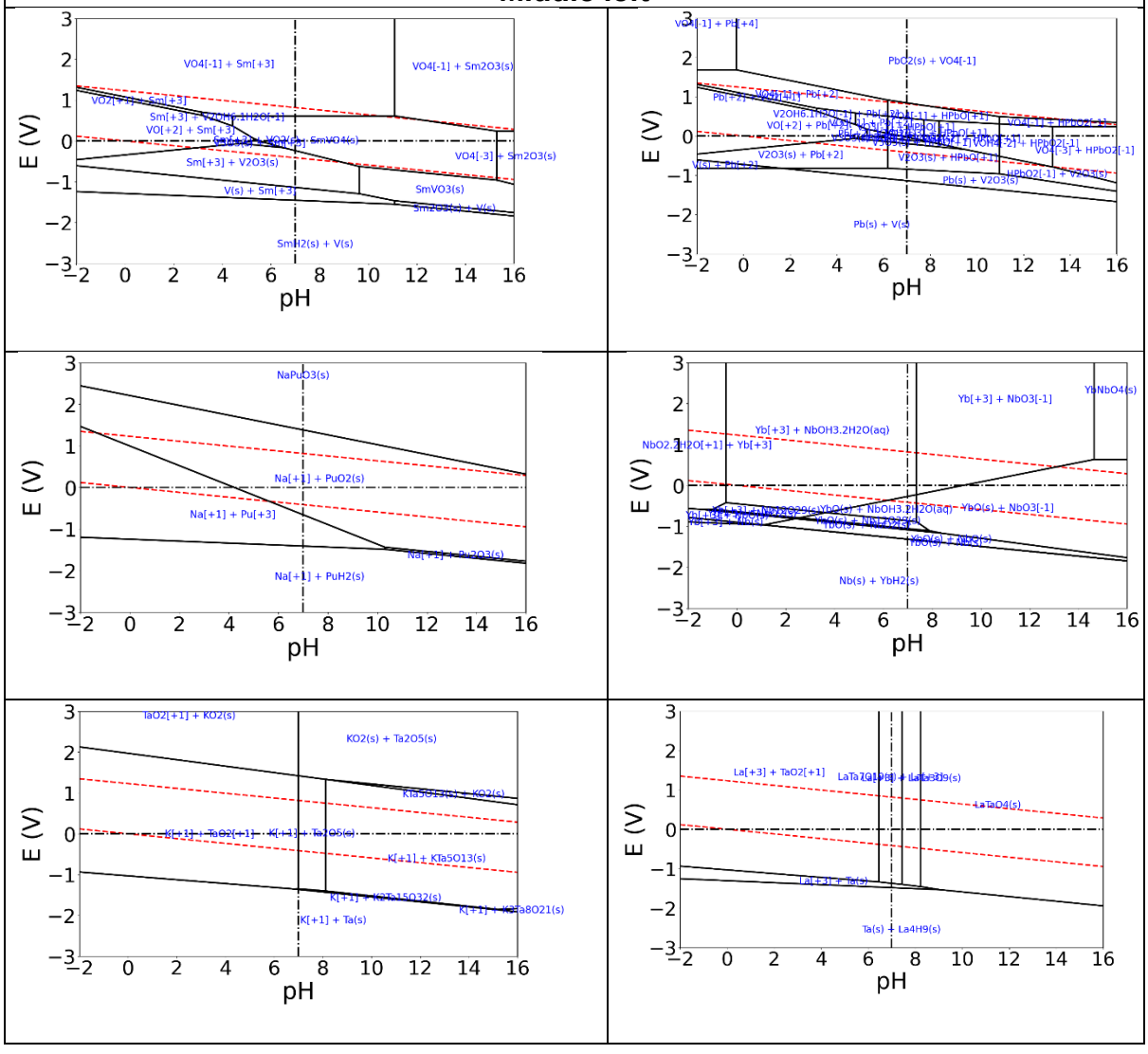


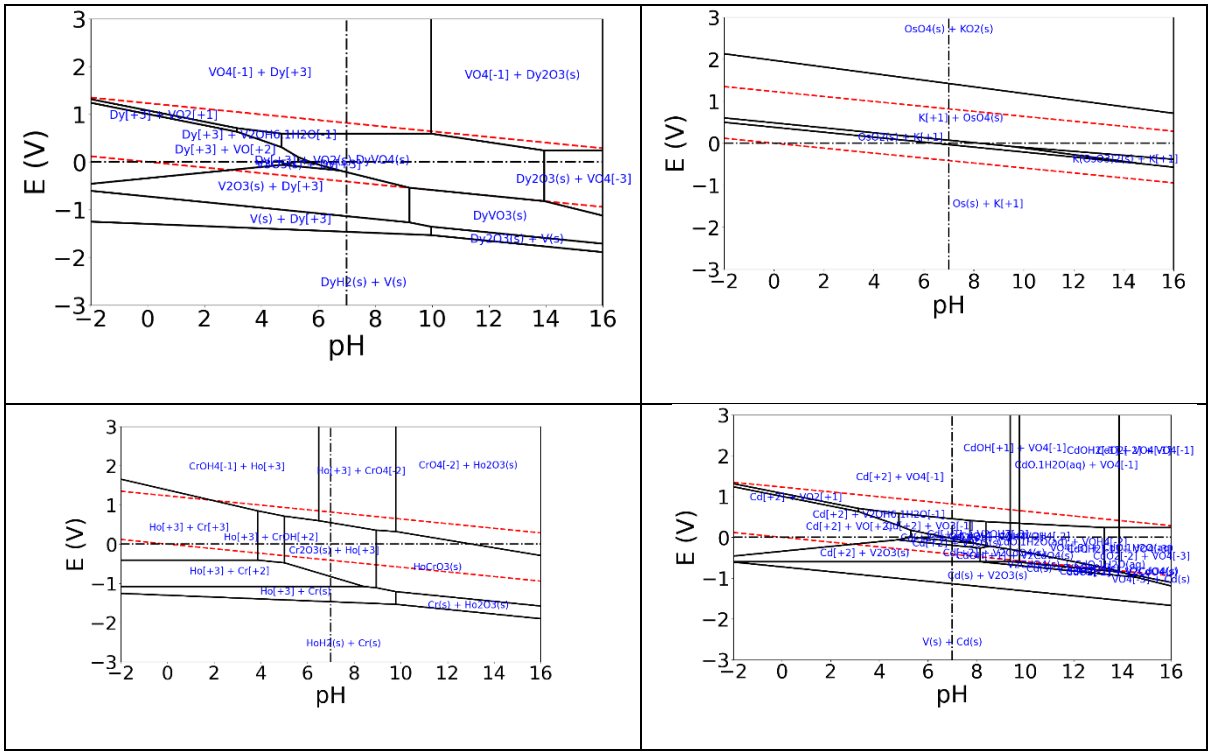
Middle right



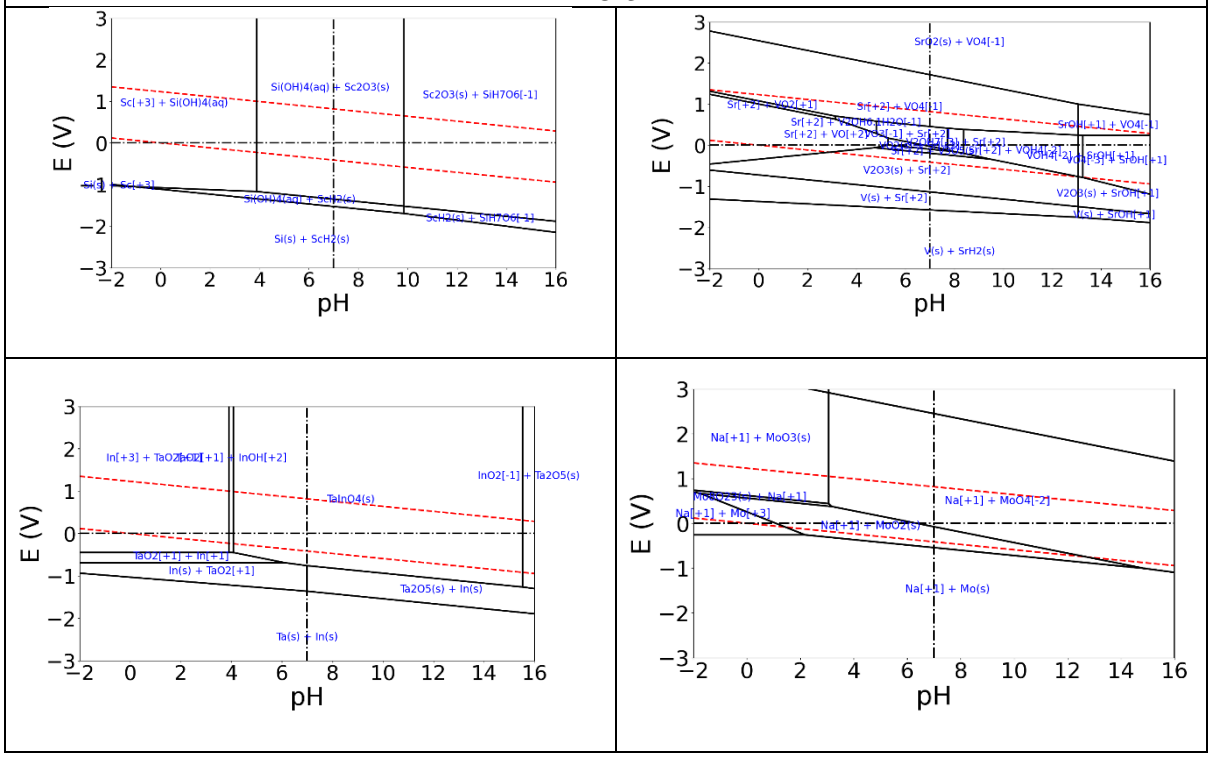

E (V)
pH
HAuO2.1H2O(aq) + BaO2(s)
BaO2(s) + AuO2.1H2O[-1]
HAuO3[-2] + BaO2(s)
Al[+3] + Ba[+2]
HAuO2.1H2O(aq) + Ba[+2]
Ba[+2] + Au(s)
BaOH[+1] + Au(s)
Ba[+2] + Au(s)
CsO2(s) + I2O5(s)
CsO2(s) + IO3[-1]
Cs[+1] + I2O5(s)
Cs[+1] + I[-1]
CsIO3(s)
Al11O18(s) + Yb[+3]
Al[+3] + Yb[+3]
Al2O3(s) + YbO(s)
YbH2(s) + Al[+3]
YbH2(s) + Al2O3(s)
YbH2(s) + Al(s)
NiHO2(s) + Dy[+3]
NiHO2(s) + Dy2O3(s)
Ni[+2] + Dy[+3]
Dy[+3] + Ni(s)
Dy2O3(s) + Ni(s)
DyH2(s) + Ni(s)

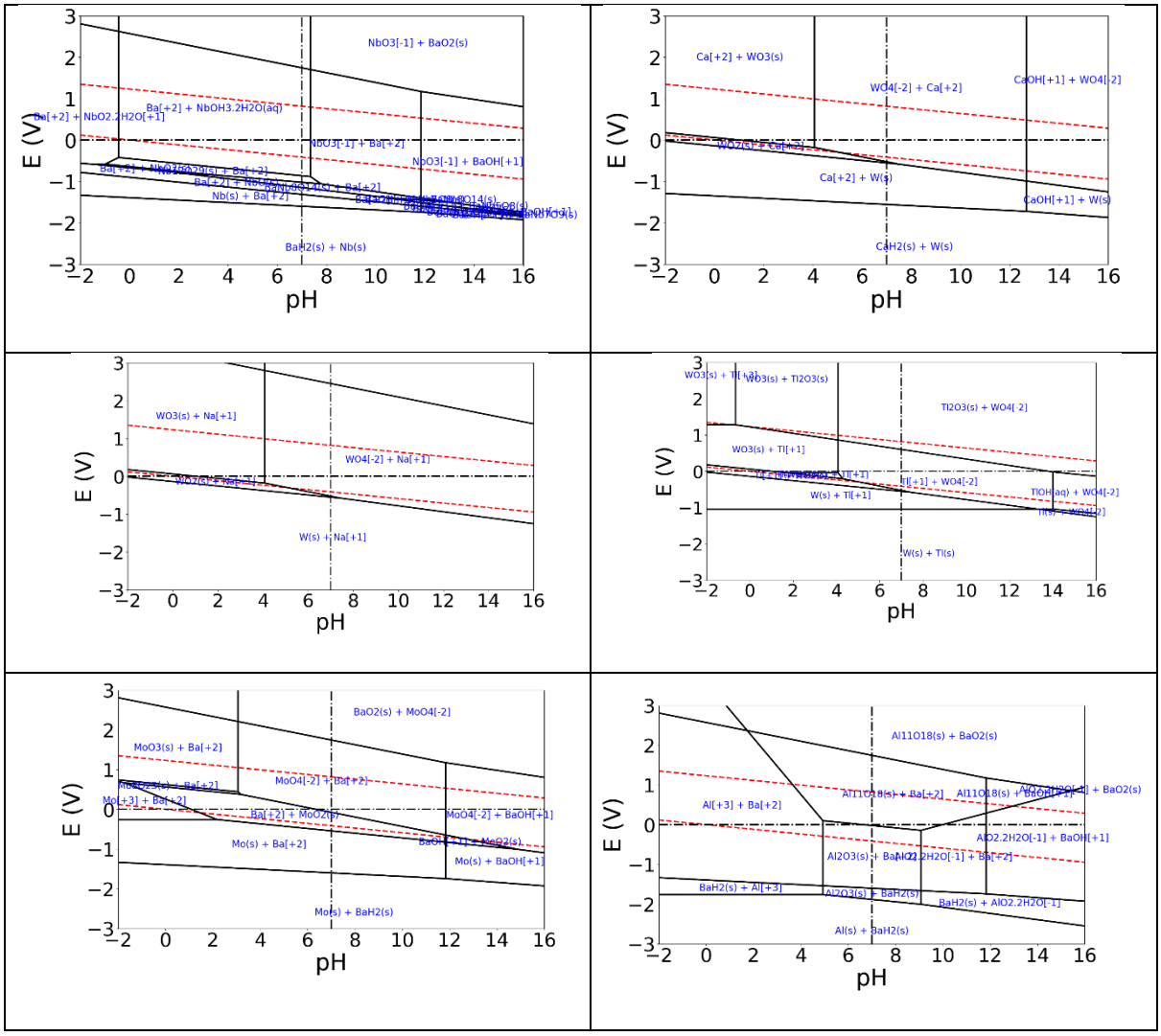
Middle left


E (V)
pH
VO4[-1] + Sm[+3]
VO4[-1] + Sm2O3(s)
VO[+2] + Sm[+3]
Sm[+3] + V2O3(s)
V(s) + Sm[+3]
SmVO3(s)
SmH2(s) + V(s)
VO4[-1] + Pb[+4]
PbO2(s) + VO4[-1]
Pb[+2] + VO4[-1]
V2O3(s) + Pb[+2]
Pb(s) + V2O3(s)
Pb(s) + V(s)
NaPuO3(s)
Na[+1] + PuO2(s)
Na[+1] + Pu[+3]
Na[+1] + PuH2(s)
Yb[+3] + NbO3[-1]
YbNbO4(s)
Yb[+3] + NbOH3.2H2O(aq)
Nb(s) + YbH2(s)
TaO2[+1] + KO2(s)
KO2(s) + Ta2O5(s)
K[+1] + KTa5O13(s)
K[+1] + Ta(s)
La[+3] + TaO2[+1]
LaTaO4(s)
La[+3] + Ta(s)
Ta(s) + La4H9(s)

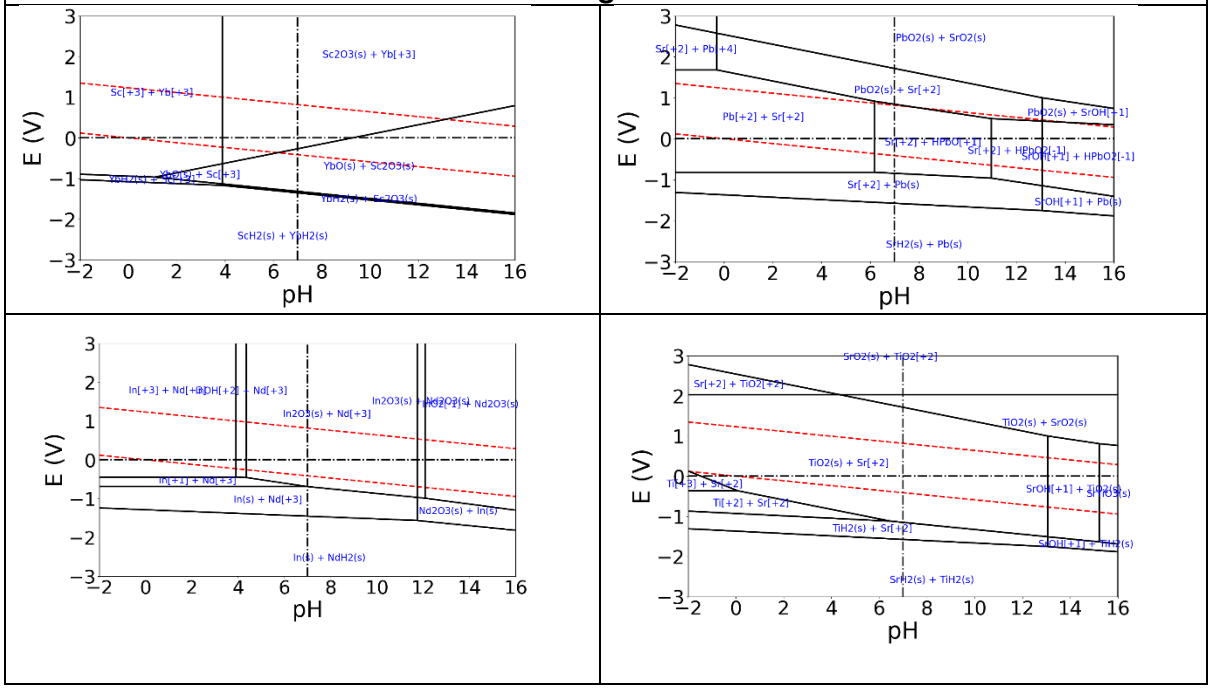

E (V)
pH
VO4[-1] + Dy[+3]
VO4[-1] + Dy2O3(s)
Dy[+3] + VO2[+1]
Dy[+3] + VO[+2]
V2O3(s) + Dy[+3]
Dy2O3(s) + VO4[-3]
V(s) + Dy[+3]
DyVO3(s)
Dy2O3(s) + V(s)
DyH2(s) + V(s)
OsO4(s) + KO2(s)
K[+1] + OsO4(s)
Os(s) + K[+1]
CrOH4[-1] + Ho[+3]
Ho[+3] + CrO4[-2]
CrO4[-2] + Ho2O3(s)
Ho[+3] + Cr[+3]
Ho[+3] + CrOH[+2]
Cr2O3(s) + Ho[+3]
Ho[+3] + Cr[+2]
HoCrO3(s)
Ho[+3] + Cr(s)
Cr(s) + Ho2O3(s)
HoH2(s) + Cr(s)
CdOH[+1] + VO4[-1]
CdO.1H2O(aq) + VO4[-1]
Cd[+2] + VO4[-1]
Cd[+2] + VO2[+1]
Cd[+2] + VO[+2]
Cd[+2] + V2O3(s)
Cd(s) + V2O3(s)
V(s) + Cd(s)

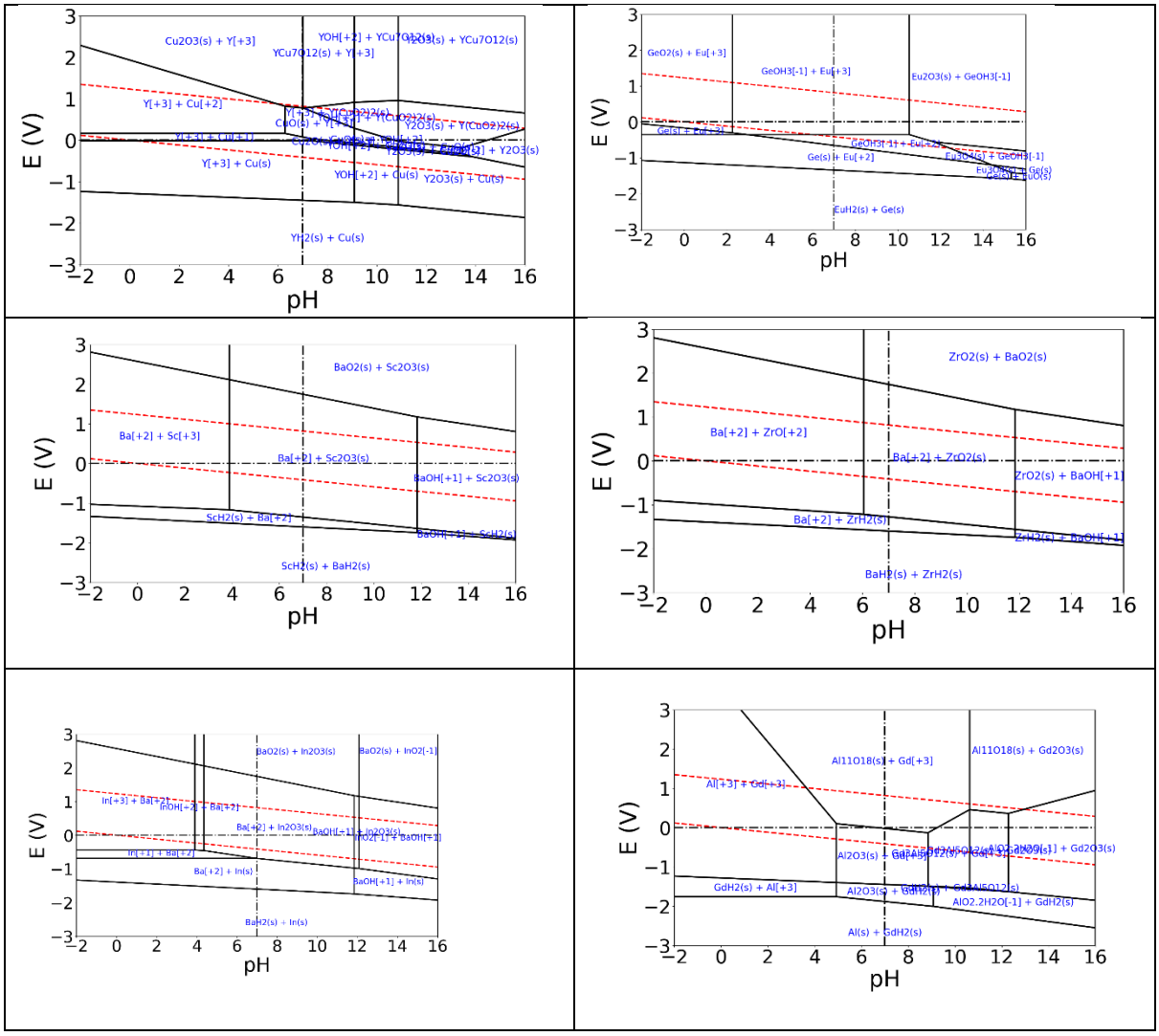
Left


E (V)
pH
Si(OH)4(aq) + Sc2O3(s)
Sc2O3(s) + SiH7O6[-1]
Sc[+3] + Si(OH)4(aq)
Si(s) + ScH2(s)
SrO2(s) + VO4[-1]
Sr[+2] + VO2[+1]
Sr[+2] + VO4[-1]
SrOH[+1] + VO4[-1]
V2O3(s) + Sr[+2]
V2O3(s) + SrOH[+1]
V(s) + Sr[+2]
V(s) + SrH2(s)
In[+3] + InOH[+2]
InO2[-1] + Ta2O5(s)
TaInO4(s)
In(s) + TaO2[+1]
Ta2O5(s) + In(s)
Ta(s) + In(s)
Na[+1] + MoO3(s)
Na[+1] + MoO4[-2]
Na[+1] + Mo[+3]
Na[+1] + MoO2(s)
Na[+1] + Mo(s)


E (V)
pH
NbO3[-1] + BaO2(s)
Ba[+2] + NbOH3.2H2O(aq)
Ba[+2] + NbO2.2H2O[+1]
NbO3[-1] + Ba[+2]
NbO3[-1] + BaOH[+1]
Nb(s) + Ba[+2]
BaH2(s) + Nb(s)
Ca[+2] + WO3(s)
WO4[-2] + Ca[+2]
CaOH[+1] + WO4[-2]
Ca[+2] + W(s)
CaOH[+1] + W(s)
CaH2(s) + W(s)
WO3(s) + Na[+1]
WO4[-2] + Na[+1]
W(s) + Na[+1]
WO3(s) + Tl2O3(s)
Tl2O3(s) + WO4[-2]
WO3(s) + Tl[+1]
W(s) + Tl[+1]
W(s) + Tl(s)
BaO2(s) + MoO4[-2]
MoO3(s) + Ba[+2]
MoO4[-2] + Ba[+2]
Mo[+3] + Ba[+2]
Ba[+2] + MoO2(s)
MoO4[-2] + BaOH[+1]
Mo(s) + Ba[+2]
Mo(s) + BaOH[+1]
Mo(s) + BaH2(s)
Al11O18(s) + BaO2(s)
Al[+3] + Ba[+2]
Al2O3(s) + Ba[+2]
BaH2(s) + Al[+3]
Al2O3(s) + BaH2(s)
BaH2(s) + AlO2.2H2O[-1]
Al(s) + BaH2(s)

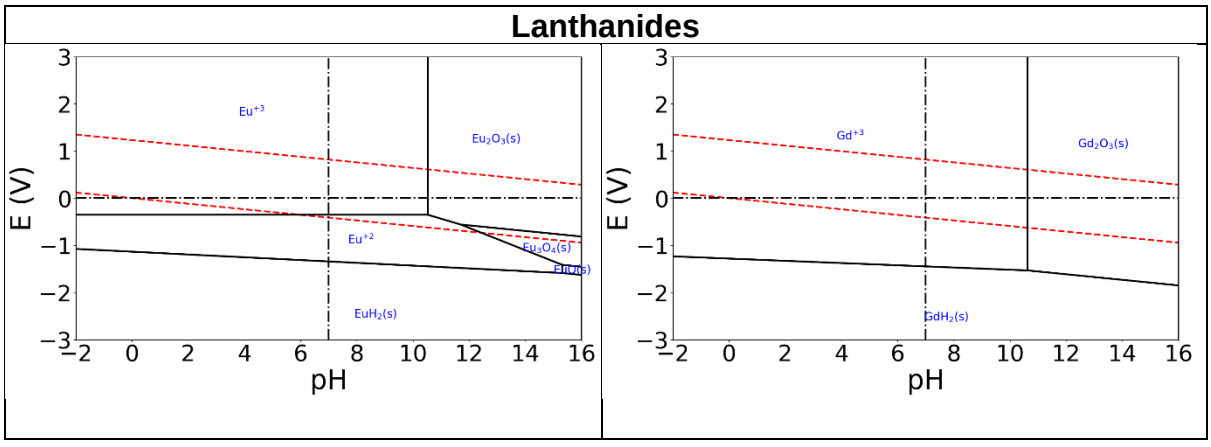
Right

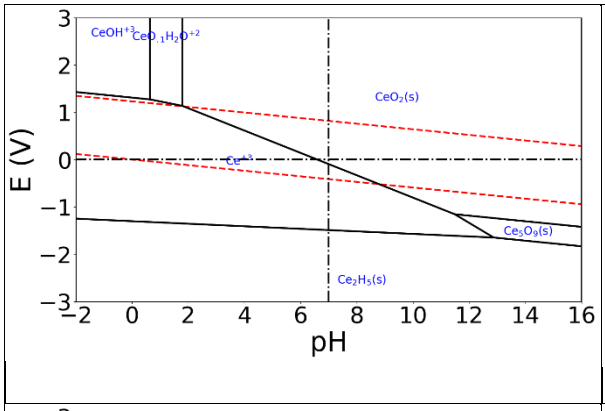

E (V)
pH
Sc2O3(s) + Yb[+3]
Sc[+3] + Yb[+3]
YbO(s) + Sc2O3(s)
ScH2(s) + YbH2(s)
PbO2(s) + SrO2(s)
Sr[+2] + Pb[+4]
PbO2(s) + Sr[+2]
Pb[+2] + Sr[+2]
Sr[+2] + HPbO2[-1]
SrOH[+1] + HPbO2[-1]
Sr[+2] + Pb(s)
SrOH[+1] + Pb(s)
SrH2(s) + Pb(s)
In[+3] + Nd[+3]
In2O3(s) + Nd[+3]
In[+1] + Nd[+3]
In(s) + Nd[+3]
Nd2O3(s) + In(s)
In(s) + NdH2(s)
SrO2(s) + TiO2[+2]
Sr[+2] + TiO2[+2]
TiO2(s) + SrO2(s)
TiO2(s) + Sr[+2]
Ti[+2] + Sr[+2]
TiH2(s) + Sr[+2]
SrOH[+1] + TiH2(s)
SrH2(s) + TiH2(s)

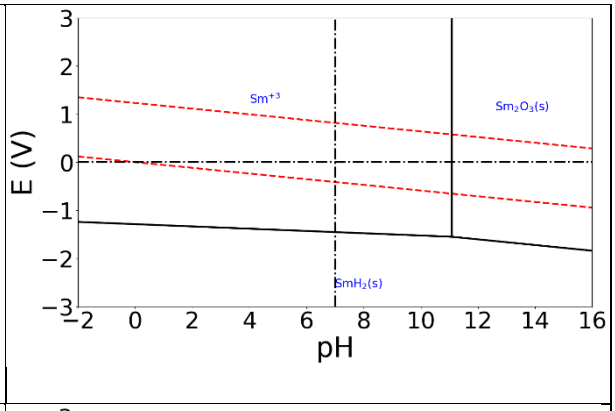


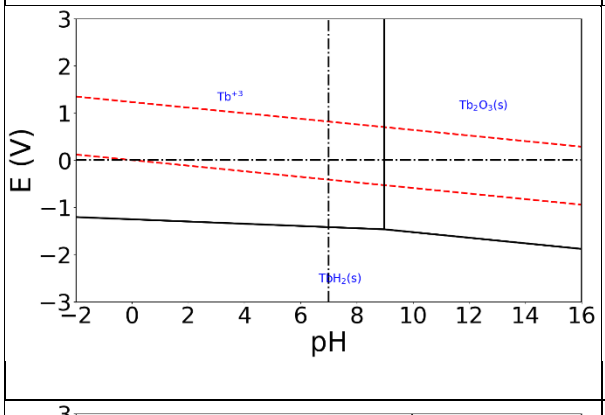
Lanthanide pourbaix diagrams

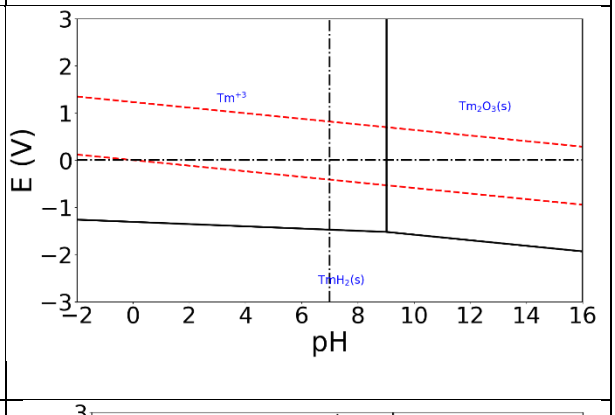
Lanthanides which were studies include: Ce, Dy, Er, Eu, Gd, Ho, La, Lu, Nd, Pr, Sm, Tb, Tm, Yb

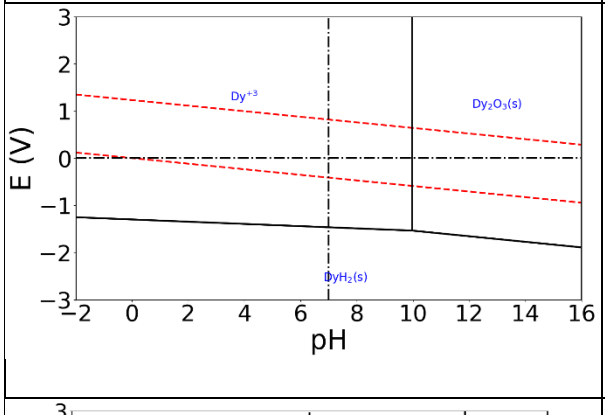


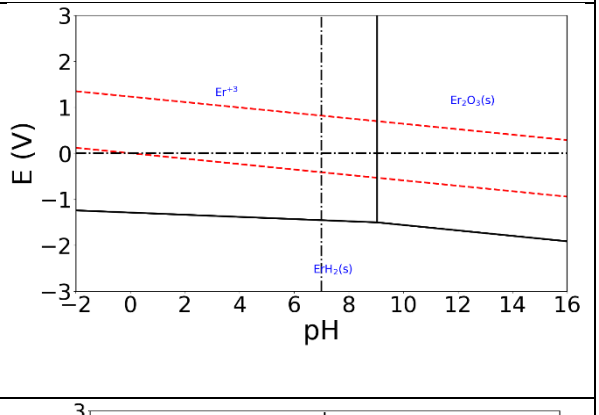

CeOH+3
CeO2H2O+2
CeO2(s)
Ce+3
Ce2H5(s)
Ce5O9(s)
E (V)
pH

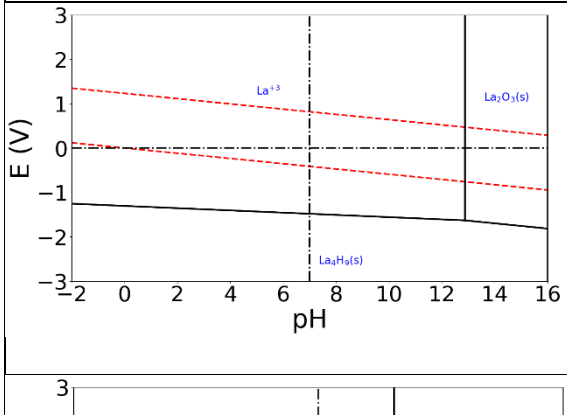

Sm+3
Sm2O3(s)
SmH2(s)
E (V)
pH

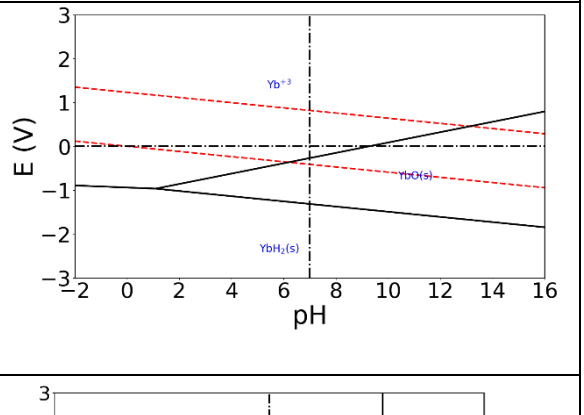

Tb+3
Tb2O3(s)
TbH2(s)
E (V)
pH

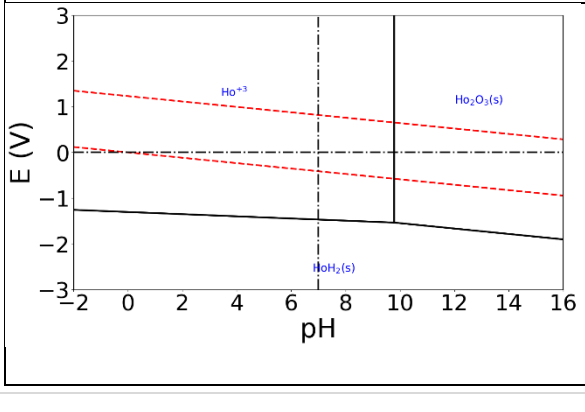

Tm+3
Tm2O3(s)
TmH2(s)
E (V)
pH

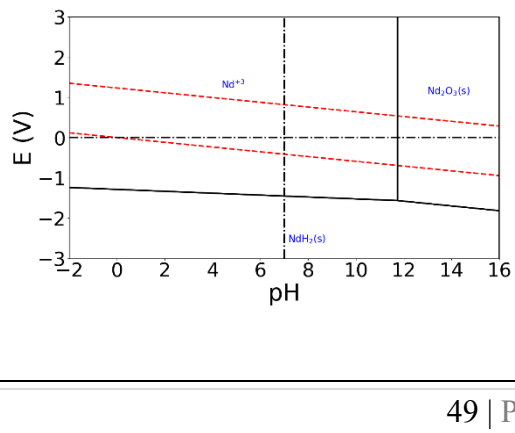

Dy+3
Dy2O3(s)
DyH2(s)
E (V)
pH

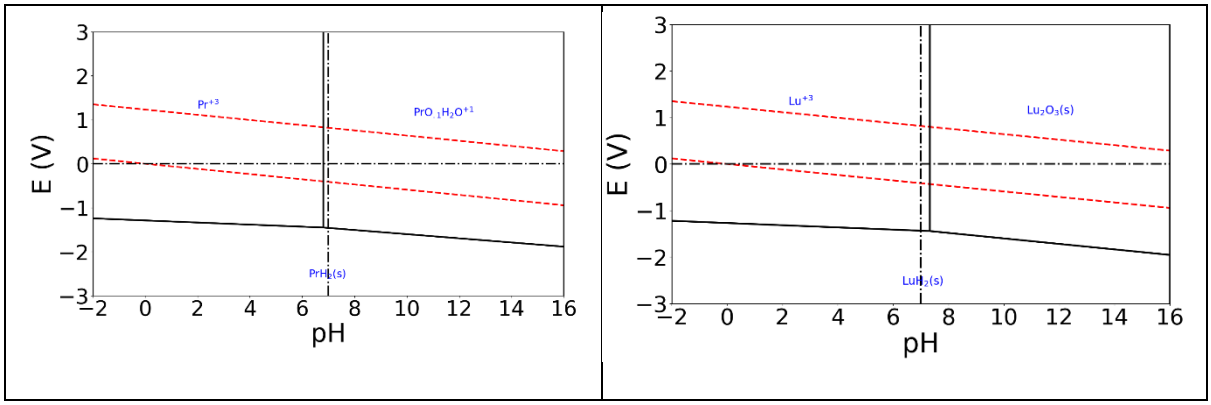

Er+3
Er2O3(s)
ErH2(s)
E (V)
pH


La+3
La2O3(s)
La4H9(s)
E (V)
pH


Yb+3
YbO(s)
YbH2(s)
E (V)
pH


Ho+3
Ho2O3(s)
HoH2(s)
E (V)
pH


Nd+3
Nd2O3(s)
NdH2(s)
E (V)
pH


E (V)
pH
Pr+3
PrO1H2O+1
PrH2(s)
Lu+3
Lu2O3(s)
LuH2(s)