

# Machine-Learning Discovery of Highly Oxidized $\text{IrO}_x$ Phases

Raul A. Flores,<sup>\*,†</sup> Christopher Paolucci,<sup>\*,‡</sup> Ankit Jain,<sup>\*,¶</sup> Muratahan Aykol,<sup>\*,§</sup>  
Jens K. Nørskov,<sup>\*,¶</sup> Michal Bajdich,<sup>\*,||</sup> and Thomas Bligaard<sup>\*,||</sup>

<sup>†</sup> *SUNCAT Center for Interface Science and Catalysis, Department of Chemical Engineering, Stanford University, Stanford 94305, California, USA*

<sup>‡</sup> *Department of Chemical Engineering, University of Virginia, Charlottesville, Virginia 22903, United States*

<sup>¶</sup> *Department of Physics, Technical University of Denmark, Lyngby, Denmark*

<sup>§</sup> *Toyota Research Institute, Los Altos, CA 94022, USA*

<sup>||</sup> *SUNCAT Center for Interface Science and Catalysis, SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA*

E-mail: flores12@stanford.edu; cp9wx@virginia.edu;

temp\_temp\_ankits\_email\_address\_temp\_temp@dtu.dk; muratahan.aykol@tri.global; jkno@dtu.dk;  
bajdich@slac.stanford.edu; bligaard@stanford.edu

## Abstract

Recent advancements in statistical methods, colloquially termed as machine learning, have revolutionized a tremendous number of fields due to the ease by which we can train models that are flexible enough to regress to data of interest while maintaining predictive power. Nowhere has this impact been felt as much as in the field of materials science, which had previously been bottle-necked by relatively computationally expensive methods. Herein, we report on a ML methodology to enumerate bulk crystal structures in the  $\text{IrO}_2$  and  $\text{IrO}_3$  space.

## Introduction

The general problem of solving for the most thermodynamically stable crystal structure for an arbitrary inorganic system remains a prohibitive challenge in material science. Experimentally synthesized inorganic materials often conform to the globally minimum energy struc-

ture (or similarly stable meta-stable phases), therefore it is desirable from a computational perspective to know the corresponding crystal structure to conduct simulations on. This problem is theoretically challenging, as the number of possible crystal structures for a given stoichiometry is prohibitive. Iterative Active Machine Learning and unique prototype identification to discover stable new materials and catalysts. Motivation for  $\text{IrO}_x$ , low representation, longstanding controversy over oxidation states and topology, and demonstrates promise for OER and Li ion batteries.

Reported +6 oxidation state phases are achievable leading to high degree of structural variability, which is the highest for transition metals. High oxidation states (low pH high anodic voltage, harsh oxidizing conditions) unexplored, need very specific structures with precise oxygen connectivity (aka high pressure  $\text{SrIrO}_3$ ) that can exist. Machine learning is the efficient way to explore this exploring Antarctica for life sparse space. What we show here...

The two key features of our algorithm that make the exploration of an expansive space pos-

sible are its use of surrogate models and active learning framework. Because DFT calculations are prohibitively computationally expensive to carry out for large data sets, herein we train a Gaussian Process (GP) machine learning model to serve as a surrogate model. In general active learning applications, the requisite training data is not available. Active learning frameworks are a means by which to generate the most valuable training data set, on the fly.

## Results and discussion

### I. Active Learning Machine Learning Methodology

The structures that comprise the candidate data set were constructed by parsing for structurally unique systems in the OQMD and Materials Project DFT databases. The structural uniqueness was performed using a space-group symmetry classification scheme developed by TEMP-Ankit-REF which can classify any arbitrary structure based on stoichiometry (composition), space-group, and Wyckoff positioning of the atoms (see SI for more details on the structural classification scheme). To focus the scope of the study, only AB<sub>2</sub> and AB<sub>3</sub> stoichiometries were parsed from the databases. The AB<sub>2</sub> formula was chosen because it includes rutile-IrO<sub>2</sub>, the known most stable polymorph of IrO<sub>2</sub>. Importantly, AB<sub>3</sub> was chosen to include high valency IrO<sub>3</sub> structures in our search. This classification scheme can directly serve as a fingerprinting scheme and has successfully been used for the prediction of formation energies. The results of the classification scheme resulted in a XYZ AB<sub>2</sub> and XYZ AB<sub>3</sub> structural prototypes for which iridium and oxygen were replaced for A and B sites, respectively. Finally, a coarse isotropic volume relaxation based on atomic radii was performed on the structures to accommodate the atomic radii of iridium and oxygen into the lattice. Finally, a Voronoi tessellation based fingerprinting scheme developed by Wolverton-paper-REF was used to encode the relevant chemical information for each structure. The Voronoi based method was used

because it is insensitive to volume relaxation. PCA was used to reduce the dimensionality of the feature space from 271 to 20 features such that 99percent of the variance is captured.

The active learning algorithm proceeds through iterative ML training, prediction, and acquisition steps and is visualized in figure TEMP. The A Gaussian process (GP) regression employing a rational quadratic kernel was chosen as the ML model because it offers a high degree of flexibility and importantly, error quantification for the predicted formation energies. Further details on the GP model, including hyper parameter information, is included in the SI.

1. Candidate space of structures is generated
2. Initial seed data is used to train a ML model
3. ML model predicts energy of entire candidate space
4. Most valuable calculation(s) is selected using acquisition function
5. DFT calculation(s) is(are) performed to obtain additional training data points
6. ML model is retrained with additional data
7. Repeat steps (TEMP) through (TEMP) until convergence criteria is reached

[Figure 1 about here.]

### II. Application of AL to stable polymorph discovery in the Ir-O space

- Describe relevant features - Physical intuition? - Describe convex hull plot (energy vs. Ir-O distance), computed amorphous phase to define synthesizability - While only 2 IrO<sub>2</sub> in MP/OQMD, we can compare our structures to other computed IrO<sub>2</sub> not in open databases.

This is a citation example.<sup>1</sup>

[Figure 2 about here.]

- XYZ unique AB<sub>3</sub> Structures, 259 unique prototypes. Substitute Ir and O, expand to minimum Ir-O distance  $\zeta$  XYZ - followed same procedure as in 3.1, Training Set of 35 structures, 8 of which are IrO<sub>3</sub> - Describe initial training and training after first 10 DFT structures

- Describe convex hull, classes of structures ( $\alpha$ -AlF<sub>3</sub> like, rutile like, and layered, should be segregated in hull plot) - briefly describe structures within each class, cite in literature where appropriate

[Figure 3 about here.]

### III. Electrochemical OER Application

In the following section we will demonstrate the merit of our stable polymorph discovery algorithm by elucidating the electrochemical properties of the four promising structures discussed in the previous section. In particular, surfaces constructed from the four polymorphs will be evaluated for their activity towards the oxygen evolution reaction (OER), an important chemistry with direct application to fuel cell devices. Additionally, the surfaces will be evaluated for their stability and equilibrium surface coverage of surface oxygen and hydroxides.

#### Bulk Pourbaix

The electrochemical stability phase diagram (E vs. pH) was constructed by considering the equilibrium conditions of the following species: Ir, rutile-IrO<sub>2</sub>,  $\alpha$ -IrO<sub>3</sub>, rutile-IrO<sub>3</sub>,  $\beta$ -IrO<sub>3</sub>, and an aqueous dissolved IrO[4-] species (See TEMP—SI for additional details). The resulting diagram is shown in Fig. S4. Importantly, under acidic conditions (pH  $\leq$  7) and in the bias region of interest for the OER (1.23 V vs. RHE)  $\alpha$ -IrO<sub>3</sub> shows a large window of stability. This indicates that the  $\alpha$ -IrO<sub>3</sub> phase may be stabilized under the highly oxidizing conditions of the OER. The stability regions for rutile-IrO<sub>3</sub> and  $\beta$ -IrO<sub>3</sub> in the absence of any other IrO<sub>3</sub>

polymorphs are also indicated by the unfilled solid lines and demonstrate a sizable stability window as well. This result implies that these metastable phases may also play a role for the OER.

[Figure 4 about here.]

#### b. OER Activities and Surfaces

Fig. S5 summarizes the major results of the electrochemical activity and surface stability analysis. First, Fig. S5 a.) shows the surface energy Pourbaix plots for the four IrO<sub>x</sub> crystals of interest. For each bulk system, the surface energy as a function of applied potential (pH=0), for various facets, and at various coverages (bare, \*OH, and \*O covered), are shown, see SI for more details. The specific facets were chosen from the highest intensity x-ray diffraction peaks from powder-diffraction spectra simulated in VESTA, as well as using physical intuition as to which facets would be most physical. Additionally, the bulk phase limits of stability from figure TEMP are included at the bottom of each subplot. In most cases, the oxygen covered surfaces dominate at the OER equilibrium potential (1.23 V vs RHE) with bare surfaces being competitive to within TEMP eV/A<sup>2</sup>, this competitiveness goes away at even modest overpotentials (eta 0.3,  $\sim$  1.5 V vs RHE), at which point the oxygen covered terminations are further overstabilized relative to the bare surfaces, making them the sole dominant termination. Therefore in our activity analysis we consider mainly oxygen terminated surfaces for the OER. The OER activity (expressed in terms of the limiting potential) for select oxygen terminated surfaces are shown in Fig. S5 as a function of the DGO-DGOH TEMP OER thermodynamic descriptor. The two rutile-IrO<sub>2</sub> surfaces (100, and 110) are located towards the strong binding side of the volcano, indicating that that they bind OER intermediate too strongly. Encouragingly, with predicted overpotentials of TEMP and TEMP, our rutile-IrO<sub>2</sub> are within the range of experimental overpotentials found in literature. The three IrO<sub>3</sub> polymorph surfaces all have DGO-DGOH descriptor towards

the top and right of the volcano, indicative of weaker binding energetics. This is evident from figure SI TEMP (scaling) which shows a clear distinction between the IrO<sub>2</sub> and IrO<sub>3</sub> polymorphs, with IrO<sub>3</sub> binding overall weaker than IrO<sub>2</sub>. The best performing systems, including the (100), (110), and (211) facets of a-IrO<sub>3</sub>, b-IrO<sub>3</sub> (101), and R-IrO<sub>3</sub> (110), have overpotentials of 0.4 V vs RHE, a 0.2 V vs RHE improvement over the rutile-IrO<sub>2</sub> system.

[Figure 5 about here.]

### c. OER Intermediate Scaling

Figure TEMP shows the scaling relations between the adsorption free energies of the OER intermediate species for the IrO<sub>x</sub> systems studied herein. It can be seen clearly that the data points corresponding to the three IrO<sub>3</sub> polymorphs are roughly 1 eV weaker binding than the rutile-IrO<sub>2</sub> points. This generally weaker binding of the IrO<sub>3</sub> stoichiometry is responsible for the observed improvement in theoretical activity. The  $\Delta G_{\text{OOH}}$  vs.  $\Delta G_{\text{OH}}$  relationship is very close to the traditional “universal scaling relations”, demonstrating that our materials do not break the infamous  $\Delta G_{\text{OOH}}$  vs.  $\Delta G_{\text{OH}}$  scaling.

[Figure 6 about here.]

## Conclusion

And in conclusion we presented work here...

**Acknowledgement** Organizations to acknowledge TRI SUNCAT Stanford NERSC etc. JAGT and MB acknowledge the support by the U.S. Department of Energy, Office of Science, Office of Basic Energy Science, via Grant DE-SC0008685 to the SUNCAT Center of Interface Science and Catalysis. The authors would like to acknowledge the use of the computer time allocation for the Transition metal-oxide and metal surfaces: applications and reactivity trends in catalysis at the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S.

# Supporting Information Available- References

(1) Smith, A.; Smith, J. TEMP title. **9999**, *1*.

## Machine Learning Algorithm Methods

Relevant details about the ML Gaussian process here

## Electrochemical OER Computational Methods

### Density Functional Theory Methods

All OER calculations were performed using density functional theory (DFT) implemented via the Vienna ab-initio simulation package (VASP) and utilizing the PBE exchange-correlation functional. Dipole corrections were imposed on all non-symmetric slabs. A 4x4x3 k-point mesh with gamma-point centered Monkshort-packing was used for all slabs. The plane-wave energy cutoff was 500 eV.

All slab calculations maintained a vacuum spacing of 15 Å. All structures were relaxed utilizing a TEMP algorithm with a stop criteria being that all atoms satisfy a maximum force threshold of 0.02 eV/Å.

### OER Thermodynamic Methodology

#### Surface Energy Pourbaix Methodology

Procedure: - For the top/most stable bulk structures the following procedure was carried out

- \* Stable stoichiometric terminations were cut from the bulk. Stable termination planes were guesstimated via intuition, and the x-ray diffraction pattern tool from Vesta

- \* Electrochemical surface coverage was elucidated via a surface Pourbaix analysis. Need to know the coverage of surface under operating conditions (1.23 V RHE)

- \* Thermodynamic/limiting potential analysis of the OER mechanistic pathway. Volcano plot, limiting potentials, etc.

# List of Figures

S1	(a) Active learning algorithm diagram. First the candidate space of structures is generated, next, the machine learning model is trained on any available DFT formation energy data. The trained ML model is then used to predict the DE of the entire candidate space. Finally, an acquisition step is performed to pick the next most valuable calculation to perform an-initio DFT on (b) Toy model demonstrating a GP model converging with each subsequent iteration. . . . .	7
S2	Gaussian process machine learning models trained initially on (a) publicly available DFT data for IrO <sub>2</sub> and (b) all of the acquired DFT calculations from the active learning algorithm. See SI for additional panels at intermediate iterations of the active learning algorithm. The Gibbs formation energy (either DFT derived or predicted from the GP model) and associated GP estimated error (2 sigmas or something TEMP) is plotted for each polymorph in the IrO <sub>2</sub> candidate space. The data points in each subset are ordered from most to least stable (lowest to largest DE formation). The individual markers are colored based on their ordering in the final converged GP model. Acquired structures are identified by their red borders and slightly larger size. The insets show the most stable TEMP structures, where several well known crystal structures are labeled. . . . .	8
S3	TEMP. . . . .	9
S4	Electrochemical bulk phase stability diagram (Pourbaix) of the Ir-O-H chemical space considering rutile-IrO <sub>2</sub> , . . . . .	10
S5	Summary of OER results for the four bulk structures of IrO <sub>x</sub> considered: rutile-IrO <sub>2</sub> (green), $\alpha$ -IrO <sub>3</sub> (purple), rutile-IrO <sub>3</sub> (orange), and $\beta$ -IrO <sub>3</sub> (pink). (a) Surface energy Pourbaix diagrams for each structure, with the surface energy of various facets and coverages shown as a function of applied potential. The bulk Pourbaix diagram's bounds of stability at pH 0 are superimposed at the bottom of each subplot. (b) OER activity volcano for IrO <sub>x</sub> systems considered utilizing the $\Delta G_O - \Delta G_{OH}$ thermodynamic descriptor. The purple dotted line corresponds to the experimental limiting potential at 10 mA cm <sup>2</sup> for IrO <sub>3</sub> , while the green band corresponds to the range of experimentally observed overpotentials for pristine IrO <sub>2</sub> catalysts. (c) Select surface facets for the four IrO <sub>x</sub> crystal systems considered. . . . .	11
S6	Relationship between the adsorption free energies of the three key OER intermediates (*OH, *O, *OOH), with $\Delta G_{OH}$ chosen as the dependent variable. Best fit lines are provided for $\Delta G_{OOH}$ vs. $\Delta G_{OH}$ and $\Delta G_O$ vs. $\Delta G_{OH}$ . Additionally, "universal scaling relations" for $\Delta G_{OOH}$ vs. $\Delta G_{OH}$ and $\Delta G_O$ vs. $\Delta G_{OH}$ are shown (black dotted lines) to emphasize our deviation from the traditionally reported scaling fits. The trivial $\Delta G_{OH}$ vs. $\Delta G_{OH}$ relationship is included for completeness. . . . .	12

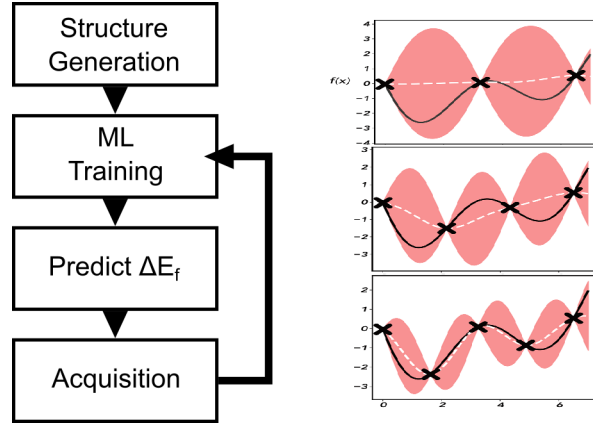


Figure S1: (a) Active learning algorithm diagram. First the candidate space of structures is generated, next, the machine learning model is trained on any available DFT formation energy data. The trained ML model is then used to predict the DE of the entire candidate space. Finally, an acquisition step is performed to pick the next most valuable calculation to perform an-initio DFT on (b) Toy model demonstrating a GP model converging with each subsequent iteration.

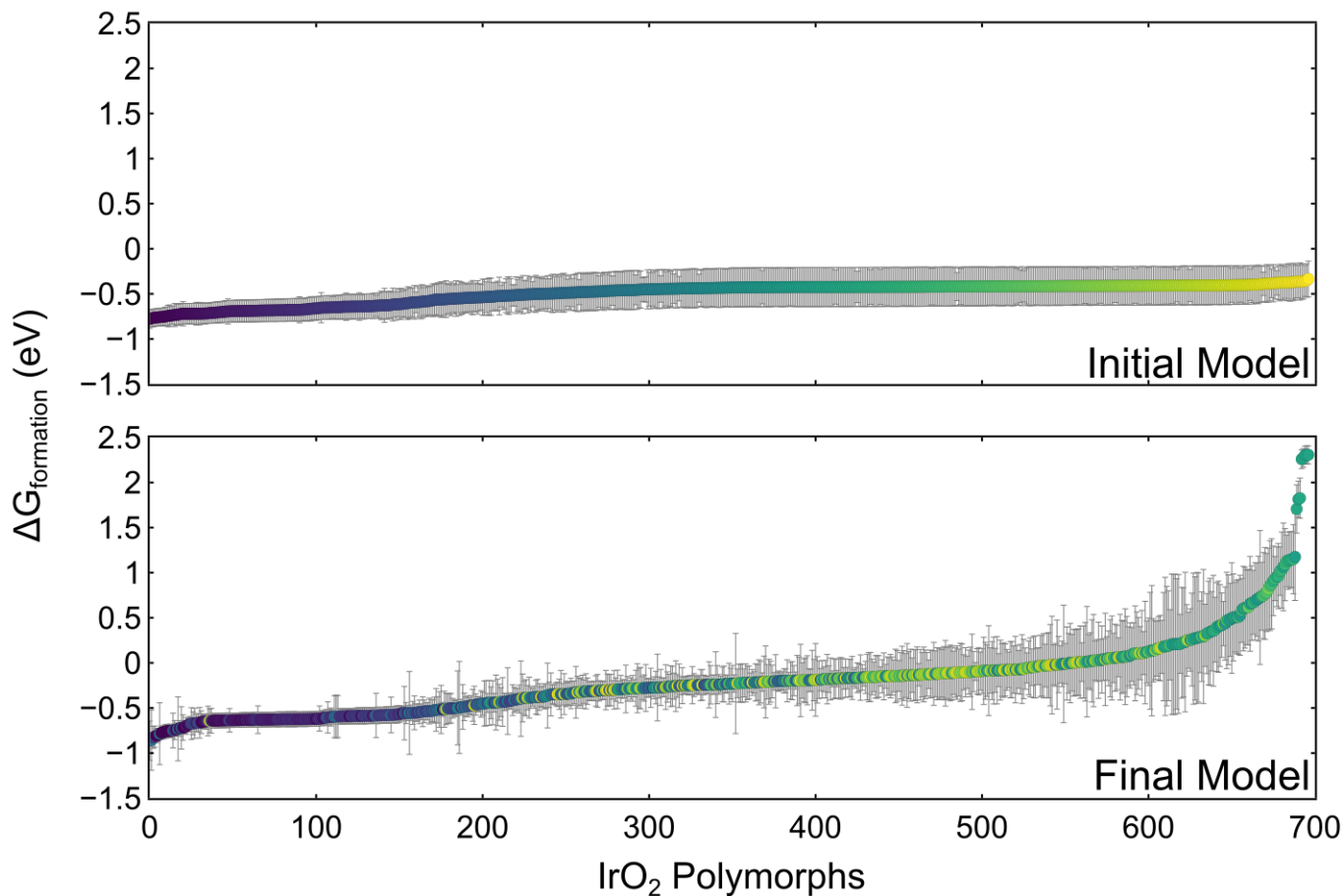


Figure S2: Gaussian process machine learning models trained initially on (a) publicly available DFT data for IrO<sub>2</sub> and (b) all of the acquired DFT calculations from the active learning algorithm. See SI for additional panels at intermediate iterations of the active learning algorithm. The Gibbs formation energy (either DFT derived or predicted from the GP model) and associated GP estimated error (2 sigmas or something TEMP) is plotted for each polymorph in the IrO<sub>2</sub> candidate space. The data points in each subset are ordered from most to least stable (lowest to largest DE formation). The individual markers are colored based on their ordering in the final converged GP model. Acquired structures are identified by their red borders and slightly larger size. The insets show the most stable TEMP structures, where several well known crystal structures are labeled.



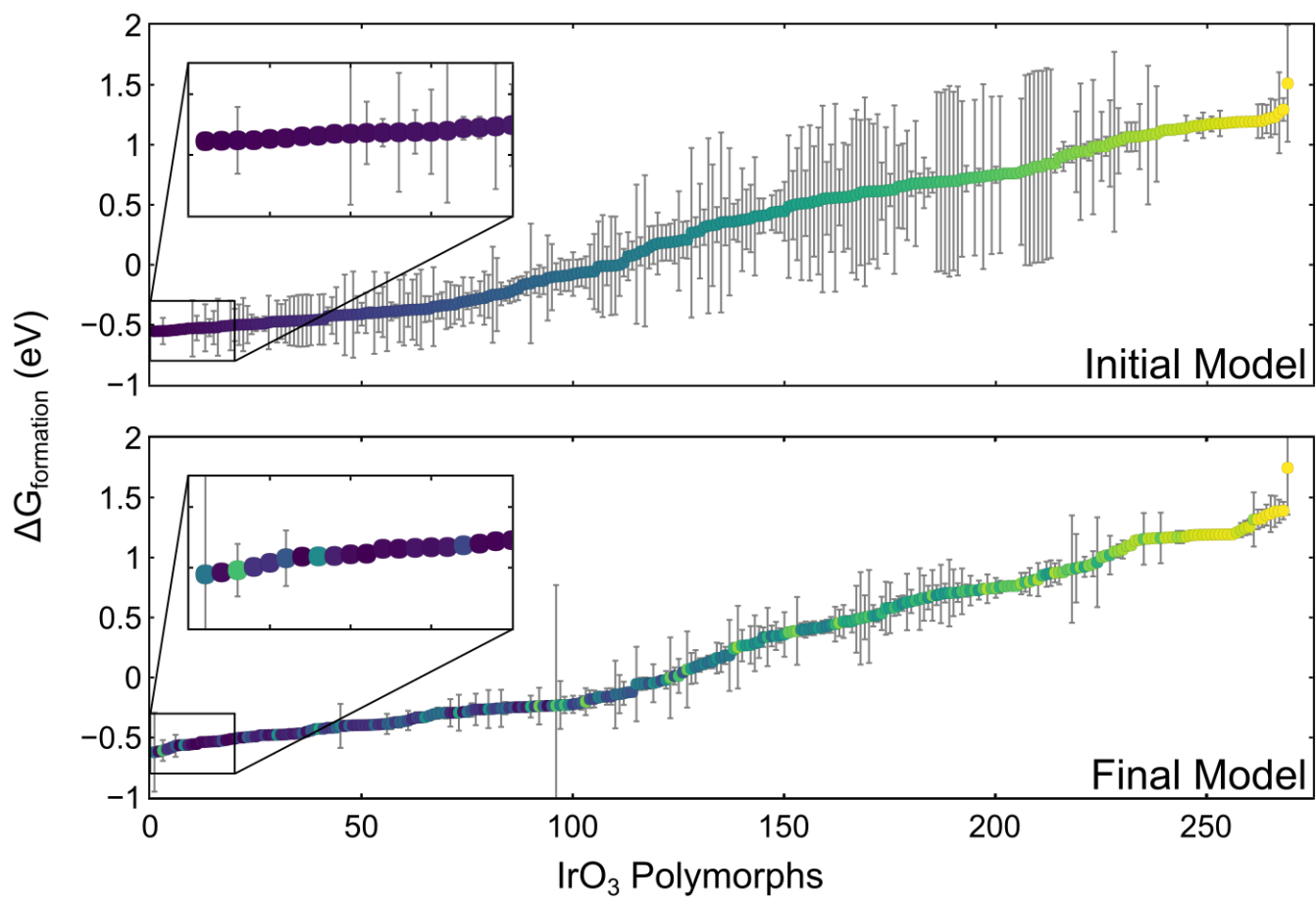


Figure S3: TEMP.

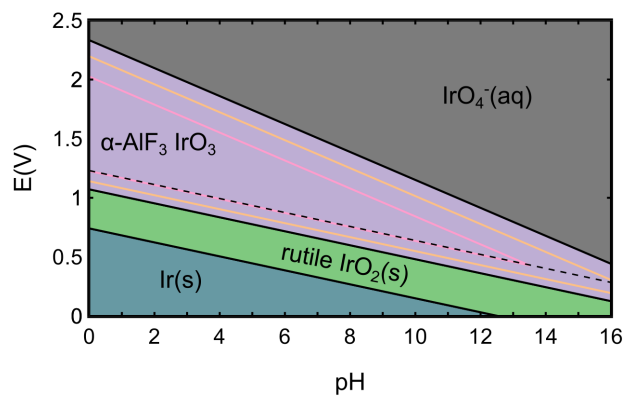


Figure S4: Electrochemical bulk phase stability diagram (Pourbaix) of the Ir-O-H chemical space considering rutile- $\text{IrO}_2$ ,

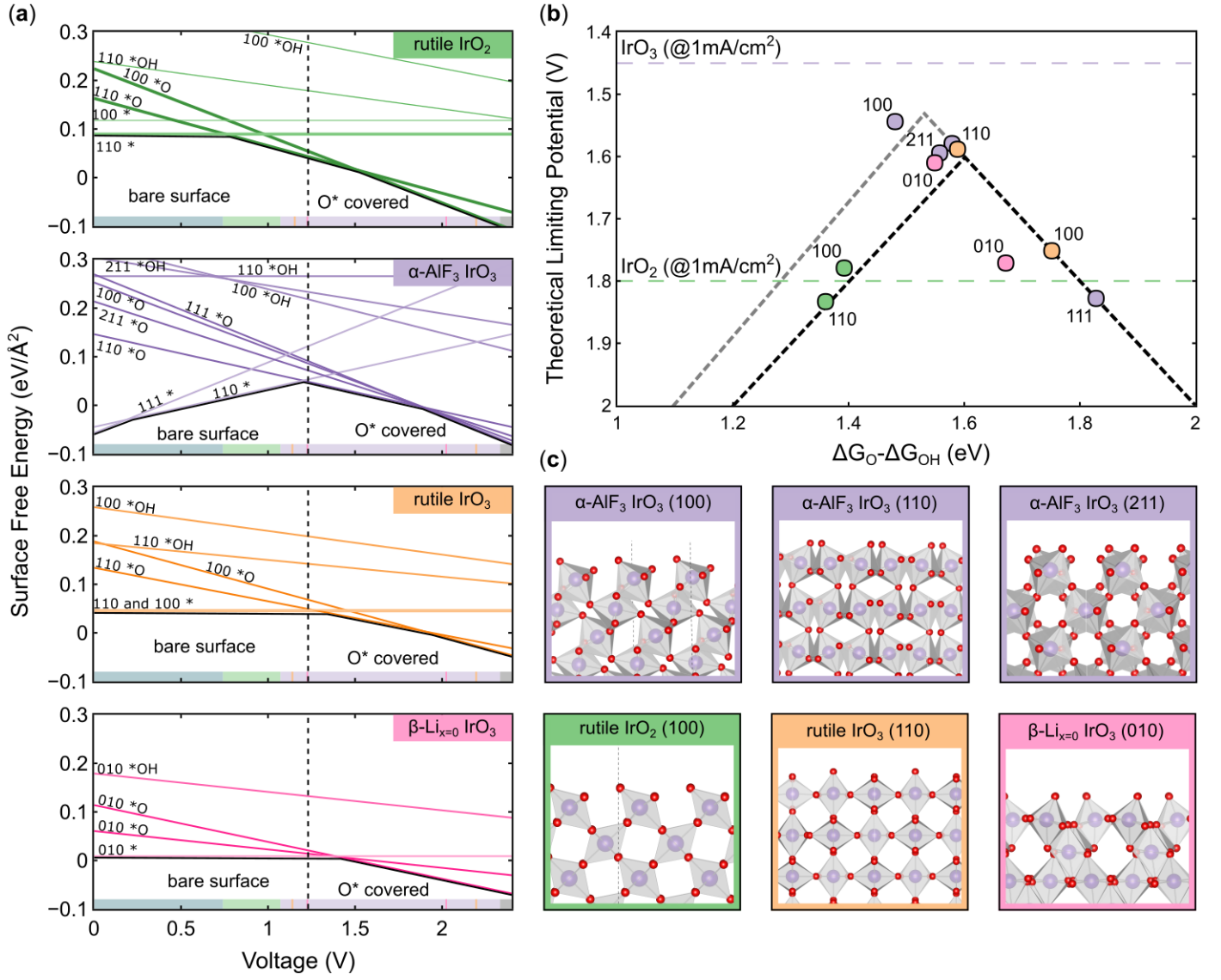


Figure S5: Summary of OER results for the four bulk structures of  $\text{IrO}_x$  considered: rutile- $\text{IrO}_2$  (green),  $\alpha\text{-IrO}_3$  (purple), rutile- $\text{IrO}_3$  (orange), and  $\beta\text{-IrO}_3$  (pink). (a) Surface energy Pourbaix diagrams for each structure, with the surface energy of various facets and coverages shown as a function of applied potential. The bulk Pourbaix diagram's bounds of stability at pH 0 are superimposed at the bottom of each subplot. (b) OER activity volcano for  $\text{IrO}_x$  systems considered utilizing the  $\Delta G_{\text{O}} - \Delta G_{\text{OH}}$  thermodynamic descriptor. The purple dotted line corresponds to the experimental limiting potential at  $10\text{ mA cm}^{-2}$  for  $\text{IrO}_3$ , while the green band corresponds to the range of experimentally observed overpotentials for pristine  $\text{IrO}_2$  catalysts. (c) Select surface facets for the four  $\text{IrO}_x$  crystal systems considered.

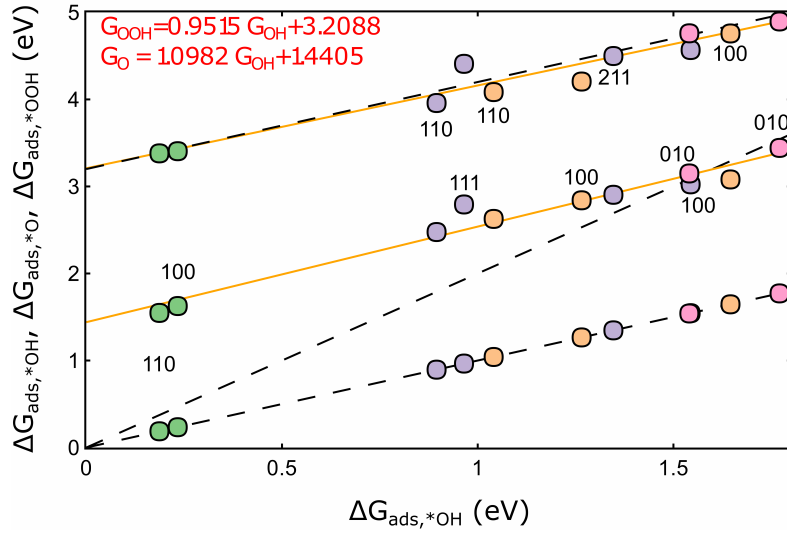


Figure S6: Relationship between the adsorption free energies of the three key OER intermediates ( $*OH$ ,  $*O$ ,  $*OOH$ ), with  $\Delta G_{\text{OH}}$  chosen as the dependent variable. Best fit lines are provided for  $\Delta G_{\text{OOH}}$  vs.  $\Delta G_{\text{OH}}$  and  $\Delta G_{\text{O}}$  vs.  $\Delta G_{\text{OH}}$ . Additionally, “universal scaling relations” for  $\Delta G_{\text{OOH}}$  vs.  $\Delta G_{\text{OH}}$  and  $\Delta G_{\text{O}}$  vs.  $\Delta G_{\text{OH}}$  are shown (black dotted lines) to emphasize our deviation from the traditionally reported scaling fits. The trivial  $\Delta G_{\text{OH}}$  vs.  $\Delta G_{\text{OH}}$  relationship is included for completeness.

## Graphical TOC Entry

