Machine-Learning Discovery of Highly Oxidized IrO$_x$ Phases

Raul A. Flores,[*,†] Christopher Paolucci,[*,‡] Ankit Jain,[*,¶] Muratahan Aykol,[*,§] Jens K. Nørskov,[*,¶] Michal Bajdich,[*,‖] and Thomas Bligaard[*,‖]

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

‡ *Department of Chemical Engineering, University of Virginia, Charlottesville, Virginia 22903, United States*

¶ *Department of Physics, Technical University of Denmark, Lyngby, Denmark*

§ *Toyota Research Institute, Los Altos, CA 94022, USA*

‖ *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 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 IrO_2 and 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 structure (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 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 $SrIrO_3$) 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 possible 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

Our machine learning accelerated materials discovery method proceeds through the following steps First, the dataset of candidate materials is constructed. This data set will define the totality of materials that will be considered by the search algorithm, this is done because the search space of materials is not a continous space but a discrite array of individual structures. Next, the dataset of materials is transformed into a vectorial representation by using a fingerprinting method that encodes the relevent chemical and structural information.

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 Jain et al.,[1] which can classify a structure based on its composition and structure. (see SI for more details on the structural classification scheme). This structural classification scheme can directly serve as a structural fingerprint and has successfully been applied towards the prediction of formation energies of inorganic compounds[1] To focus the scope of the study, only AB2 and AB3 stoichiometries were parsed from the databases. The AB2 formula was chosen because it includes rutile-IrO_2 , the known most stable polymorph of IrO2. Importantly, AB3 was chosen to include high valency IrO3 structures in our search. The results of the classification scheme resulted in a XYZ AB2 and XYZ AB3 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 Ward et al.[2] was used to encode the relevant chemical information for each structure into a vector quantity of length 271. The Voronoi based method was used because it is insensitive to volume relaxation. Separate, indpedent models were used for IrO2 and IrO3 to reduce the complexity of the space, this

3

has the addtional effect of making a large number of the fingerprint descriptors reduendant. As a result, the 271 length feature space is reduced to a TEMP length vector. PCA was used to reduce the dimensionality of the feature space from 271 to 20 features such that 99 % of the variance is captured.[3]

The active learning algorithm proceeds through iterative ML training, prediction, and acquisition steps and is visualized in figure TEMP. Although we could use practically any regression technique, we employed the Gaussian process (GP) regression model because it offers a high degree of flexibility and, most importantly, built-in error quantification. Error quantification on the predicted formation energies is important because its used in the aquisition criteria. Further details on the GP model, including hyper parameter information, is including in the SI. The aquisition is made by selecting the N systems with the lowest value of the aquisition function. The aquistion function is defined as the predicted formation energy minus the uncertainty, $a = b + c$ The value of N determines how many structures are selected for DFT calculations, and as such, determines the degree of parallization of the routine. Small values of N result in an algorithm that is slow, as every DFT calculation is performed serially. Larger values of N speed up the algorith, but potentially decreases the efficiency of the algorithm by TEMP TEMP. Here we chose a value $N = 10$. The aquired strcutres are initially volume relaxed, followed by a full relaxation of all of the atomic coordinates, see SI for additional details on the DFT methodology. The AL loop proceeds until convergence is achieved, which here is defined as the point at which the model has determined the lowest N structures and has acheived a degree of accuracy for the candidate space such that no structures have formation energies with a lower uncertainty bound lower than the most stable N structures.

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

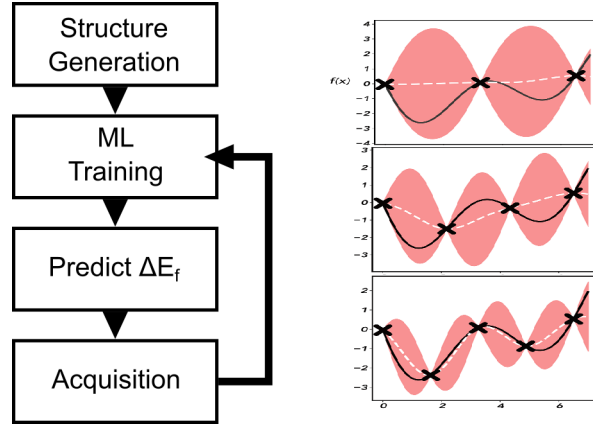7. Repeat steps (TEMP) through (TEMP) until convergence criteria is reached

Figure 1: (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.

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_2 in MP/OQMD, we can compare our structures to other computed IrO_2 not in open databases.

This is a citation example.[4]

- XYZ unique AB3 Structures, 259 unique prototypes. Substitute Ir and O, expand to minimum Ir-O distance ¿ XYZ - followed same procedure as in 3.1, Training Set of 35
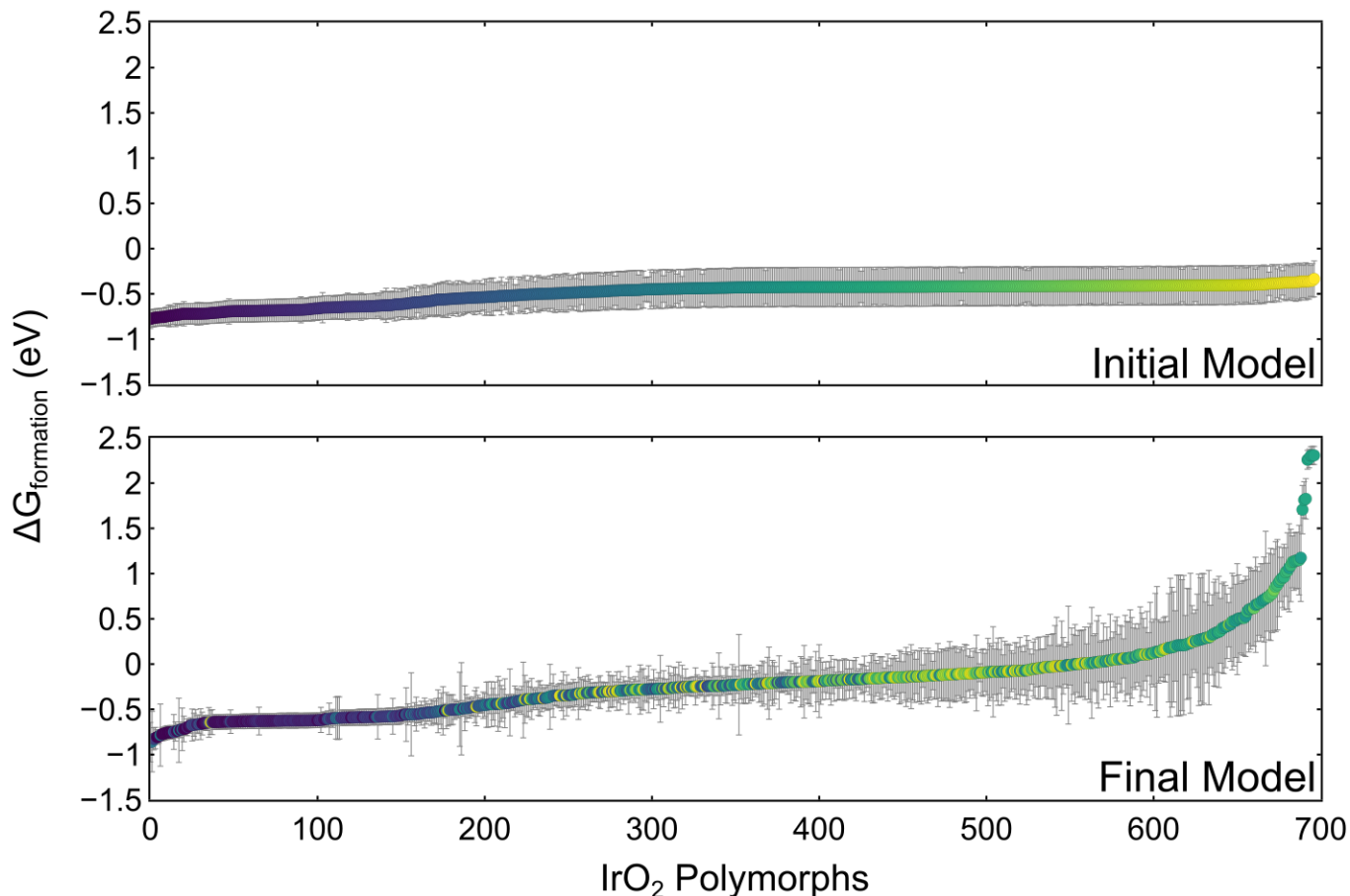
Figure 2: Gaussian process machine learning models trained initially on (a) publicly available DFT data for IrO2 and (b) all of the acquired DFT calculations from the active learning algorithm. See SI for additional panels at intermediate iterations of the active learning algorithm. The Gibbs formation 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 IrO2 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.

structures, 8 of which are IrO$_3$ - Describe initial training and training after first 10 DFT structures

- Describe convex hull, classes of structures ($\alpha-$AlF$_3$ like, rutile like, and layered, should be segregated in hull plot) - briefly describe structures within each class, cite in literature where appropriate
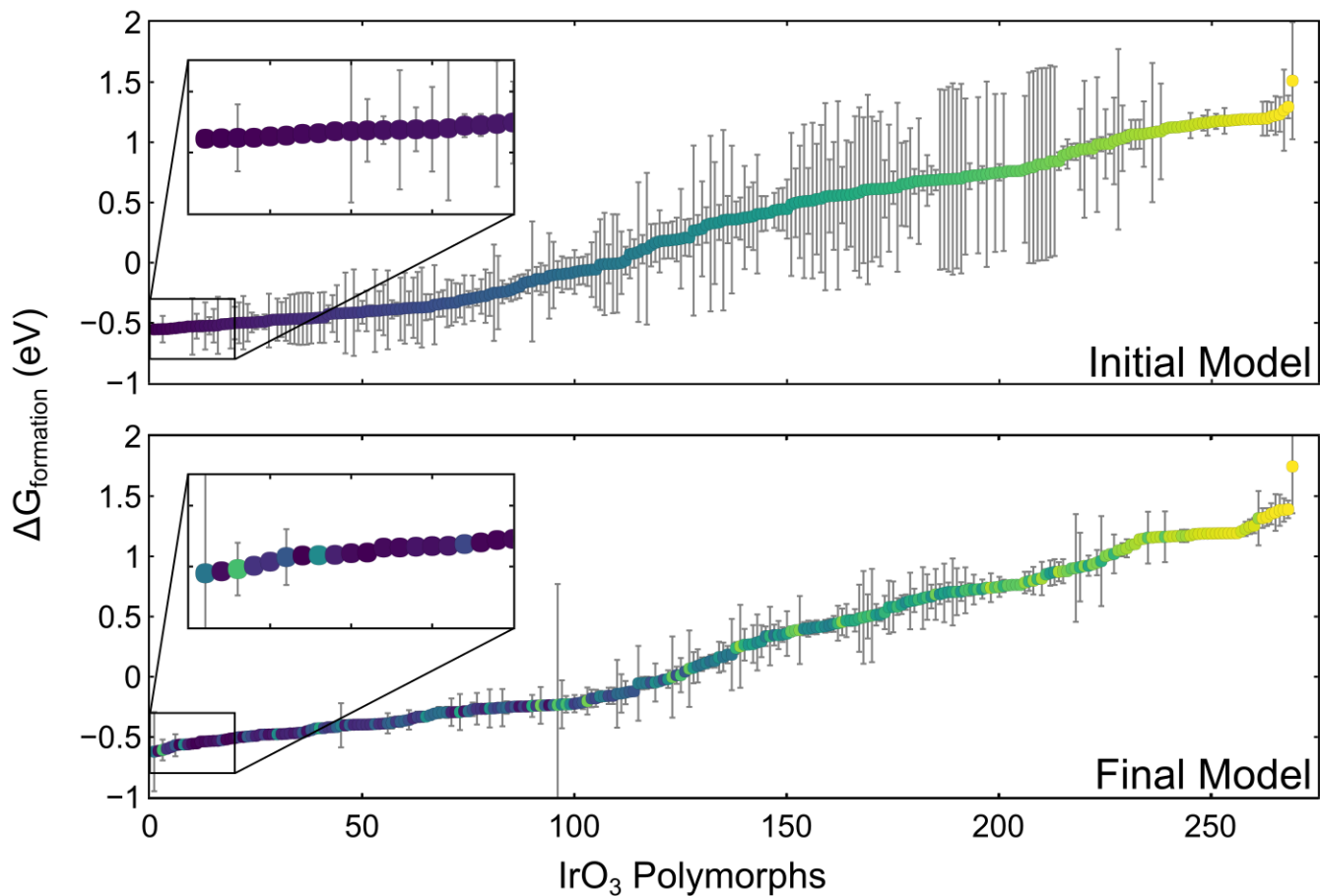

Figure 3: TEMP.

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 im-

portant 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_2 , α-IrO_3 , rutile-IrO_3 , β-IrO_3 , and an aqueous dissolved $IrO[4-]$ species (See TEMP—SI for additional details). The resulting diagram is shown in Fig. 4. Importantly, under acidic conditions (pH <7) and in the bias region of interest for the OER (1.23 V vs. RHE) α-IrO_3 shows a large window of stability. This indicates that the α-IrO_3 phase may be stabilized under the highly oxidizing conditions of the OER. For comparison, the stability regions of the metastable rutile-IrO_3 and β-IrO_3 phases, in the absence of any other competing IrO_3 polymorphs, are indicated by unfilled solid lines. As shown, these metastable phases appear to also have a wide region of stability in the OER region, due to to their formation energies being within TEMP eV of the globally stable α-IrO_3 phase, see SI table TEMP for the bulk energies of all considered phases. Because of their similar energies it is possible that some or all of these IrO_3 phases may be present and relevant for the OER. In the next section, we explore this possibility by computing the theoretical OER activity of these polymorph systems.

b. OER Activities and Surfaces

Fig. 5 summarizes the major results of the electrochemical activity and surface stability analysis. First, Fig. 5 a.) shows the surface energy Pourbaix plots for the four IrOx 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
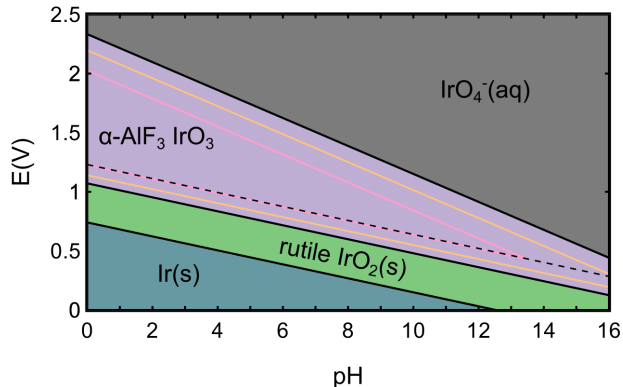
8

Figure 4: , Electrochemical bulk phase stability diagram (Pourbaix) of the Ir-O-H chemical space with respects to changes in potential and pH. We considered a bulk unoxidized Ir(s) (blue), a [+4] rutile-IrO$_2$ (green), and an aqueously dissolved IrO$_4$[4−] (gray) phase. Additionally, we considered the three IrO$_3$ polymorphs, α-IrO$_3$ (purple), rutile-IrO$_3$ (orange), and β-IrO$_3$ (pink). The water equilibrium line at 1.23 V vs RHE, which corresponds to a 0 overpotential catalyst, is shown by a dotted line.

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/A2, this competitiveness goes away at even modest overpotentials (eta 0.3, −¿ 1.5 V vs RHE), at which point the oxygen covered terminations are further over-stabilized 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. 5 as a function of the DGO-DGOH TEMP OER thermodynamic descriptor. The two rutile-IrO2 surfaces (100, and 110) are located towards the strong binding side of the volcano, indicating that that they bind OER intermediates too strongly. Encouragingly, with predicted overpotentials of TEMP and TEMP, our rutile-IrO2 are within the range of experimental overpotentials found in literature. The three IrO3 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 IrO2 and IrO3 polymorphs,

9

with IrO3 binding on average TEMP eV weaker than IrO2. The best performing systems, including the (100), (110), and (211) facets of a-IrO3, b-IrO3 (101), and R-IrO3 (110), have overpotentials of 0.4 V vs RHE, a 0.2 V vs RHE improvement over the rutile-IrO2 system. We note that the computed overpotentials for our rutile-IrO_2 system differs from that reported in $REFERENCE_{Colin_Science} by 0.2V. This discrepancy is due to our us of spin-polarization, which was neglected in Seitz et al., which strengthens the binding of IrO2.$

c. OER Intermediate Scaling

Figure TEMP shows the scaling relations between the adsorption free energies of the OER intermediate species for the IrO_x systems studied herein. It can be seen clearly that the data points corresponding to the three IrO_3 polymorphs are roughly 1 eV weaker binding than the rutile-IrO_2 points. This generally weaker binding of the IrO_3 stoichiometry is responsible for the observed improvement in theoretical activity. The ΔG_{OOH} vs. ΔG_{OH} relationship is very close to the traditional "universal scaling relations", demonstrating that our materials do not break the infamous ΔG_{OOH} vs. ΔG_{OH} scaling.
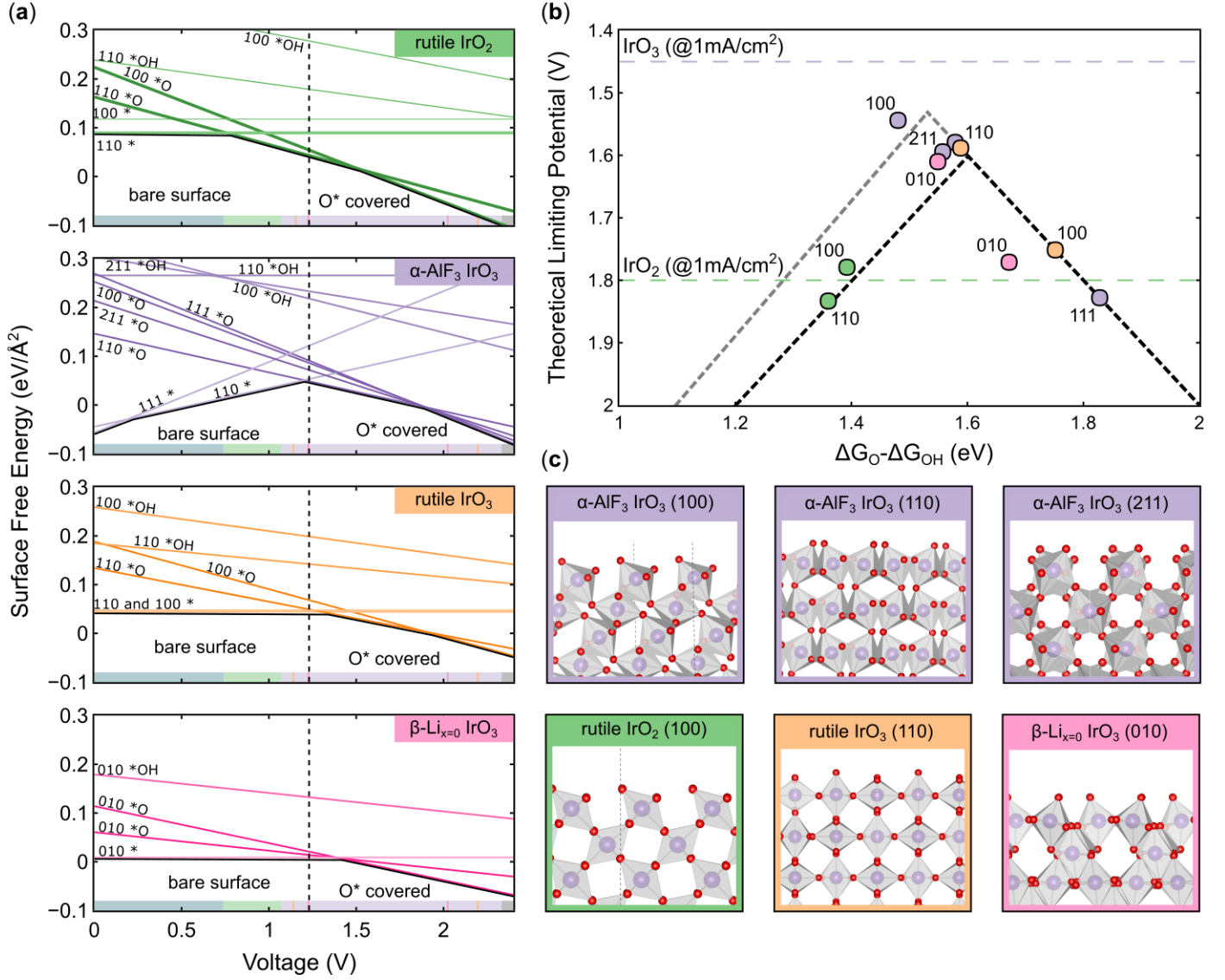
Conclusion

And in conclusion we presented work here...

Acknowledgement

Figure 5: Summary of OER results for the four bulk structures of IrO_x considered: rutile-IrO_2 (green), α-IrO_3 (purple), rutile-IrO_3 (orange), and β-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 IrO_x systems considered utilizing the ΔG_O -ΔG_{OH} thermodynamic descriptor. The purple dotted line corresponds to the experimental limiting potential at 10 mA cm^2 for IrO_3, while the green band corresponds to the range of experimentally observed overpotentials for pristine IrO_2 catalysts. (c) Select surface facets for the four IrO_x crystal systems considered.
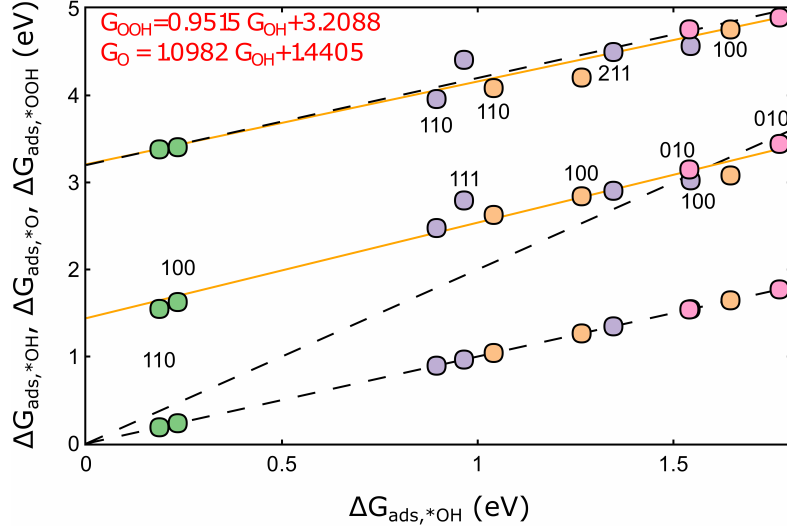
Figure 6: Relationship between the adsorption free energies of the three key OER intermediates (*OH, *O, *OOH), with ΔG_{OH} chosen as the dependent variable. Best fit lines are provided for ΔG_{OOH} vs. ΔG_{OH} and ΔG_O vs. ΔG_{OH}. Additionally, "universal scaling relations" for ΔG_{OOH} vs. ΔG_{OH} and ΔG_O vs. ΔG_{OH} are shown (black dotted lines) to emphasize our deviation from the traditionally reported scaling fits. The trivial ΔG_{OH} vs. ΔG_{OH} relationship is included for completeness.

Supporting Information Available

Machine Learning Algorithm Methods

Classifying unique prototypes in DFT databases

TEMP

Gaussian process regression model

TEMP

Gaussian process regression model

Relevant details about the ML Gaussian process here

The Gaussian Process model utilized a rational quadratic kernel with variable length scales for each dimensino of the feature space.

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 A. 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/A.

OER Thermodynamic Methodology

Surface Energy Pourbaix Methodology

Surface energy Pourbaix plots were constructed by calculating the surface energy of each slab by under standard conditions (V=0 and pH=0) and then utilizing the computational hydrogen electrode to compute the potential dependence of the surfaces.

Surface energy calculations were performed for various facets for slabs of increasing thickness. The bulk energy was then extracted by fitting the total energy of the slabs against the number of layers as explained in REF2. This was done to avoid common issues of surface energy divergence associated with using a separate bulk energy calculation.

The sensitivity of a given slab to an applied bias is dependent on the composition of the surface, in particular, the effect of coverage of electrolyte species which can deposit oxygen, hydrogen, and hydroxide species on the surface layers. These additional O and H atoms are not referenced to the atoms in the slab, but are instead referenced to the computational hydrogen electrode and water-splitting reaction. The equation for is as follows:

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.

References

(1) Jain, A. PHYSICAL REVIEW B 98 , 214112 (2018) Atomic-position independent descriptor for machine learning of material properties. **2018**, *214112*, 1–7.

(2) Ward, L.; Liu, R.; Krishna, A.; Hegde, V. I.; Agrawal, A.; Choudhary, A.; Wolverton, C. Including crystal structure attributes in machine learning models of formation energies via Voronoi tessellations. *Physical Review B* **2017**,

(3) Tipping, M. E.; Bishop, C. M. Probabilistic principal component analysis. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **1999**,

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

Graphical TOC Entry