

Active Learning Discovery of Oxidized and Active IrO_x Phases

Raul A. Flores,[†] Christopher Paolucci,[‡] Ankit Jain,[¶] Kirsten T. Winther,[†] Jose Antonio Garrido Torres,[†] 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: bajdich@slac.stanford.edu; bligaard@stanford.edu

Abstract

Machine learning (ML) based surrogate models have become an increasingly common tool to overcome the relative expense of density functional theory in the field of computational materials discovery. The application of ML surrogate models trained on expansive data sets of ab-initio DFT data has been used to discover new materials in the vast composition-space. However, surrogate model applications to the structural space of bulk-crystals, i.e. polymorphs, are very scarce due to the inherent bias of ab-initio datasets towards systems with a large variety in composition but with limited

structural diversity. Herein, we report on an active learning ML methodology that searches for the most stable crystal structures of IrO_2 and IrO_3 by utilizing surrogate models which optimize within a candidate data set of crystal motifs sourced from publicly available materials databases. We demonstrate the efficacy of this AL-accelerated methodology by discovering 70 percent of the 10 most stable crystal structures for IrO_2 and IrO_3 with less than 50 DFT calculations. For IrO_2 , we reaffirm the rutile phase as the globally stable polymorph, while for the previously unexplored IrO_3 more than **XX** unique structures are discovered within the metastability limit of 0.2 eV per atom. For IrO_3 , we discover 10 previously unknown polymorphs, e.g., $\alpha\text{-AlF}_3$ type and a rutile-like IrO_3 structure with stabilities lower than 0.2 per atom than anything known to date. With these results as inputs, we construct a new bulk Pourbaix diagram of the $\text{Ir-H}_2\text{O}$ system. We computationally test the proficiency of these phases towards oxygen evolution reaction and find that the above stable IrO_3 polymorphs have much higher activity than any IrO_2 . This work opens up an opportunity to materials/catalysts structural discovery on unprecedented scale.

Introduction

Predicting the thermodynamically favorable crystal structures for an arbitrary inorganic system remains a challenging problem in computational material science. When simulations are used to guide the search for new materials, the stable and meta-stable crystal structures, i.e. polymorphs, above the convex hull of stability must be known in order to predict the material properties. Although there have been numerous examples in recent years of machine learning algorithms applied towards the prediction of formation energies of large ab-initio data sets, these data sets are biased towards common structures and varying composition space. For example, roughly half the entries ($\sim 200,000$) in The Open Quantum Materials Database (OQMD) correspond to ternary-alloy combinations in the same close-packed cubic structure. As such, these efforts have been primarily concerned with the enumeration

of composition (elemental identity and stoichiometries) and less so with the exploration of structural diversity. The task of finding globally/locally stable crystal structures is equivalent to performing a global optimization (GO) routine within the highly dimensional potential energy surface (PES). Traditional approaches, such as simulated annealing, are only tractable for the most simple systems, such as metallic crystals which tend to adopt highly symmetric close-packed configurations, but is less suited for more complex materials. For instance, the class of structurally diverse metal-oxides, an important class of materials which tend to organize themselves into well-defined local coordination environments (octahedral, tetrahedral, etc.) which can assemble in a large variety of configurations with long-range order. In the past, various groups have put forth methodologies to address the structural-diversity problem using computation for the MnO₂ and VO_x polymorph spaces, but most of these methodologies are limited by the fact that they are operating within the highly intractable/dimensional PES. Here we report on a crystal structure discovery algorithm that leverages machine learning surrogate models and an active learning framework to accelerate the discovery of novel crystal structures at fixed composition. The algorithm avoids operating in the highly dimensional 3*N*-space by leveraging nature's propensity for symmetry by preparing data sets with a large degree of structural diversity at fixed composition.

Herein, we focus on the chemical space of iridium oxide polymorphs, an important class of materials with applications in electrochemistry. In particular, rutile-IrO₂ (Ir[4+] oxidation state), is the most stable form of iridium-oxide at standard conditions, and is a well studied electrocatalyst for the oxygen evolution reaction (OER).^{1–8} Previous studies on SrIrO₃ electrocatalyst for the OER demonstrated that Sr leaching might leave behind a highly oxidized Ir (Ir[6+] for hypothetical IrO₃) and it was argued as one possibly for observed high OER activity.¹ Other groups also observed such dissociation of IrO_x catalyst and subsequent formation of amorphous-like layer of unknown structure.⁹ Highly oxidized IrO₃ phases as also formed as the terminal structure of Li_xIrO₃ anodes.⁹ For these reasons, we focused our study to search for stable polymorphs in the standard IrO₂ stoichiometry and the more oxidized

IrO_3 stoicheometry, thereby neglecting the possibility of metal-hydroxide phases which have previously been shown to important. Purely octahedral IrO_3 leads naturally to 100 percent corner sharing octahedra, where all terminal surface Ir-oxygens are potentially OER active sites. Furthermore, such pure corner sharing octahedral crystals are known from in other systems such fluorites and chlorites.

In the first section, we define our prototype space and introduce the active-learning surrogate model. Next, we highlight the application of AL to the IrO_2 and IrO_3 prototype space. Here we discuss the acceleration/performance and practical limitations of this approach as well as the nature of the most stable polymorphs. Here, we also extract and analyze the rich structural information of our set. In the section 3, we construct a revised bulk Pourbaix diagram of the $\text{Ir}-\text{H}_2\text{O}$ system highlighting the importance of the IrO_3 phases under OER. Finally, we construct thermodynamic OER volcano of most stable phases and discuss the trends in activities.

Results and discussion

I. Candidate Space Generation and Active Learning Methodology

Active learning frameworks utilizing surrogate models have been demonstrated to successfully speed up materials discovery for alloy nanoparticles,¹⁰ structural optimizations,¹¹ and transition-state searches,¹² as well as adaptive approaches for global optimization.¹³ Historically, the most common approach to crystal strcuture prediction relied on classical global optimization schemes, in which the local/global minimum are found via numerical optimization routines that operate within the continious potential energy surface (PES). These methods are fundamentally limited by the curse of dimensionality associated with exploring the highly dimensional potential energy surface (PES), whose degrees of freedoms (and potential number of polymorphs) rises exponentialy with system size.¹⁴ As such, the global optimization approach is better suited toward smaller systems and doesn't scale efficiently

with system size. An alternative approach to GO is to first, define the set of candidate crystal polymorphs, and secondly, search through the static list of candidates with a selection-type algorithm. This approach relies on being able to prepare candidate structures that are likely to be physical and low in energy. To date, there have been various techniques to prepare candidate structures, including generating structures randomly, enumerating space groups, etc. Emperically, we know that nature tends to favor symmetric structures, and thus herein, we use construct a dataset candidate structures that leverage this intuition.

Here, we present an active learning accelerated methodology for discovering novel and stable crystal polymorphs. Our approach utilizes an active learning framework and makes use of a surrogate-model, whereby a simple regression model is trained on a set of candidate materials by iteratively and intelligently sampling structures from a candidate set. A high-level overview of algorithm is presented in figure 1. There are two primary parts of the algorithm, the first is the generation of the candidate space, which defines the list of candidate crystal structures to be searched through (1.a). The structures within the candidate space are then transformed into a vectorial representation by a fingerprinting method which encodes the relevant chemical and structural information into a numerical vector, which can then be directly used by a machine learning regression model. The second part of the algorithm is the iterative search through candidate space via a continuously retrained surrogate model (1.b).

The generation of the candidate space is a critical step in the discovery of novel crystal structures for a given system, since the initial candidate pool determines which structures that can ultimately be discovered, it is imperative to construct candidates that are sufficiently diverse, such that they encompass as much of the structural diversity in the PES as possible. Herein, the structures that comprise the candidate data sets for IrO_2 and IrO_3 were constructed by mining for AB_2 and AB_3 structures in the Materials Project¹⁵ and OQMD¹⁶ databases (in total 4528 AB_2 and 23764 AB_3 entries). Structurally redundant systems were then removed via a space-group based structural classification scheme devel-

oped by Jain et al.,¹⁷ whereby structural uniqueness is defined by the element-nonspecific stoicheometry (AB_2 , AB_3 , etc.), spacegroup symmetry and wyckoff positions of the atoms within the unit cell. Imposing structural uniqueness reduces the number of entries considerably to 697 and 259 structures for the AB_2 and AB_3 stoicheometries, respectively; this is a prime indication of the lack of structural diversity in the OQMD and MP databases. Furthermore, only bulk structures containing ≤ 75 atoms were included in the final candidate pool to reduce the computational expense of the DFT computations (further reducing the candidates spaces to 567 and 256 AB_2 and AB_3 structures respectively). Next, iridium and oxygen were substituted for the A and B sites, respectively, this was then followed by a crude isotropic expansion/contraction of the unit cell to accommodate the atomic radii of iridium and oxygen. Because of the relatively small size of the candidate space, and because we wish to validate our methods fully, we performed bulk relaxations for all 823 structures in the candidate space. See section (SI Bulk polymorph DFT optimization) for further details on the computational methodology used to relax the bulk structures. Of the 823 structures, 736 were able to be fully optimized, while 87, primarily AB_2 structures were not able to be converged for a variety of convergence issues stemming from highly voluminous and unphysical initial structures. The resulting candidate space was composed of 487 IrO_2 and 249 IrO_3 structures, all of which have unique space group/Wyckoff combinations, ensuring their structural uniqueness. While not particularly large in size, this data set will serve as a proof of concept. In addition, the size of candidate space is sufficiently small that we can tractably optimize the equilibrium structure for the entire dataset with ab-initio DFT, thus allowing us to validate the approach in an ‘after the fact’ manner.

The candidate data set was featurized using the Voronoi tessellation fingerprinting scheme developed by Ward et al.¹⁸ which produces a 271 length feature vector for each material that is invariant to isotropic expansions/contractions of the lattice and to a degree, insensitive to the exact atomic coordinates of the atoms. Herein, we apply our active learning model to the IrO_2 and IrO_3 spaces separately, because we are interested in the most stable polymorphs

at each stoichiometry. Constraining the stoicheometry per application of the AL algorithm further reduces the 271 length feature vector to 101 non-zero variance features, significantly reducing the dimensionality of the features. Further dimensionality reduction is achieved via a principle component analysis (PCA),¹⁹ which was used to reduce the remaining 101 features to 11.

The active learning algorithm proceeds through iterative generations of ML training, prediction, and acquisition steps that are visualized in figure 2. To start, we utilized Gaussian process (GP) regression with gaussian (or RBF) kernels as implemented in the CatLearn^{11,20} package for atomistic machine learning applications, and used it to train a regression model on a small seed set of DFT formation energies from randomly selected structures in the candidate space. Herein, we utilized Gaussian process regression because they are highly flexible regression models which have built-in error quantification for each prediction. The model is then used to predict the formation energies of the entire candidate space. This predicted energy landscape is then used to select the next systems to calculate via aquisition function, which defines a fitness score for each system, and systems that minimize this quantity are then selected. Herein, we use the so-called GP-UCB acquisition function

$$U = \mu - \kappa\sigma, \quad (1)$$

where μ and σ is the predicted mean and uncertainty of the formation energy, and κ is a free parameter to tune the relative weighting between exploiting low formation energy systems (small κ) and exploring high uncertainty regions of the candidate space (large κ). Here, κ is set to 1 which equally weights the energy and uncertainty. In this work we attempt to trade off exploitation and exploration by weighting the predicted formation energy and the associated uncertainty to bias systems that are both low energy and high uncertainty. Once ranked, the N systems that minimize the acquisition function are selected for full DFT calculations, which are included in the training data of subsequent AL generations. Here we chose

a bin size (N) of 5, the value of the bin size (N) determines how many structures are selected for DFT calculations, and as such, determines the degree of parallelization of the routine. The optimal value of N depends on the computational resources available, as 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 active-learning algorithm, but leads to a higher number of DFT calculations performed before convergence. Although initially unique, the structures in the candidate often relax into one another over the course of the DFT optimization, introducing duplicate structures in the post-DFT structures. These duplicate structures are removed during each generation of the AL algorithm, leaving only a single instance of lowest energy in the candidate pool. The coordination characterization function (CCF) based methodology to quantify the similarity between structures was used to identify and remove duplicate structures.²¹ The AL loop proceeds until convergence is achieved, which here is chosen to be the generation at which the structures within the range of metastability, here taken as 0.1 eV/atom, are unchanging over three consecutive generations.

II. Application of Active Learning to the discovery of stable Ir-O polymorphs

We now turn our attention of the application of this active learning scheme to the discovery of most stable forms of IrO_2 and IrO_3 . The algorithm is applied separately to both stoichiometries. Here, we report in detail the results for IrO_3 , the analogous results for IrO_2 are shown in more detail in the SI.

Figure 2a shows a sequence of plots at various generations of the active learning loop, starting with the initial generation with five randomly drawn candidates and ending with the 40th generation of the ALL (205 DFT computed structures out of the 248 total candidates). Each plot tracks the predicted (hollow grey) and DFT-derived (solid red) formation enthalpies, sorted from most to least stable. As the active learning loop acquires DFT data, the GP model becomes more accurate, as evidenced by the decreasing uncertainties

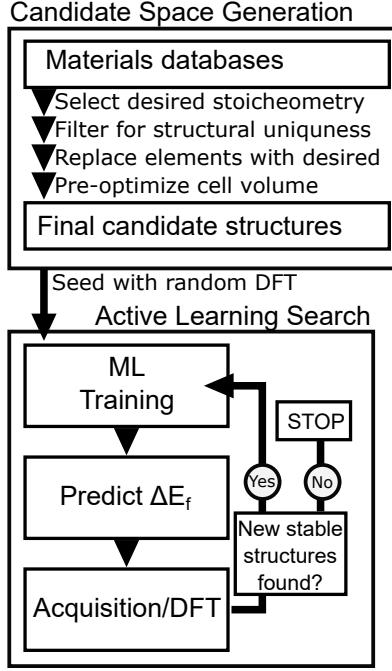


Figure 1: Process flow diagram for the active learning accelerated algorithm. The procedure is composed of (a) generation of the candidate set of considered crystal structures constructed from DFT materials databases and (b) iterative active learning surrogate search of the candidate space.

when going from left to right. At the top of each subplot of 2 the identity of top ten most stable polymorphs is tracked, with the short grey line turning into a longer red line when the structure is acquired by the ALL. At the first generation the top ten structures are randomly distributed across the entire candidate space because the GP model hasn't had enough training data to identify the most stable polymorphs as low energy systems. After only three generations (2a.ii) the GP model is sufficiently trained to have correctly identified all of the top ten systems as being low energy. Additionally, by the third generation (20 DFT calculations) 2/10 top systems have been acquired. After another 3 generations (2a.iii) (15 additional DFT calculations) the AL routine has successfully identified 7/10 of the top systems. Figure 2e plots the number of top ten structures acquired as a function of DFT calculations for the ALL with the GP-UCB acquisition function and a baseline random acquisition scheme. The results of figure 2e are averaged over independent runs of AL algorithm with the 1 sigma standard deviation between these runs shown. Overall, the GP-UCB

runs outperform the random acquisition runs, with only 50 DFT calculations on average needed to discover 7 of the top 10 systems. This is compared to the 157 DFT calculations needed to compare 7/10 top systems for the random acquisition.

The low energy region of 2a.iii is shown in 2c. And the 6 most stable structural polymorphs are shown in 2d.

Having demonstrated the efficacy of the ALL materials discovery scheme we will now turn our attention to evaluating the performance of the GP regressive model of the final generation (trained on all available IrO_3 DFT data). Figure 2b plots the GP model predicted formation enthalpy against the DFT-computed values for two special cases, 1.) predicting onto the features of the pre-optimized structures (grey), as is done in the regular operation of the ALL when acquiring new structures and 2.) predicting onto the post-DFT fingerprints (blue). It is evident from the parity plot in figure 2b that the Gaussian process model is doing an poor job of predicting the DFT formation energy of the candidate space using the pre-optimized fingerprints, with an exceptionally poor MAE of 1.5 eV/atom. The model's poor predictions are skewed towards the higher formation enthalpies, with the errors associated with low energy structures being much more robust. The same GP model does comparatively much better at predicting the formation energies of post-DFT optimized structures with an MAE of 0.2 eV/atom, which is expected since the post-DFT fingerprints directly corresponds to the DFT energies. This is to show that the GP model's poor predictive capabilities is due to the large degree of structural reorganization that occurs after DFT relaxation of the pre-optimized structure. Structures that are initialized in high energy configurations will therefore have high predicted formation enthalpies, and will then reconfigure into a lower nearby configuration, resulting in lower final energy and a large discrepancy between the predicted and final energies. It's interesting to consider why the ALL appears to perform so well as outlined previously (discovering 7/10 of the most stable candidates after only 35 DFT calculations). The reason for this is that the pre-optimized structures that are similar enough to the most stable final equilibrium structures will not restructure considerably,

meaning that their predicted formation energies will be close enough (and low enough) to be quickly picked up by the acquisition criteria.

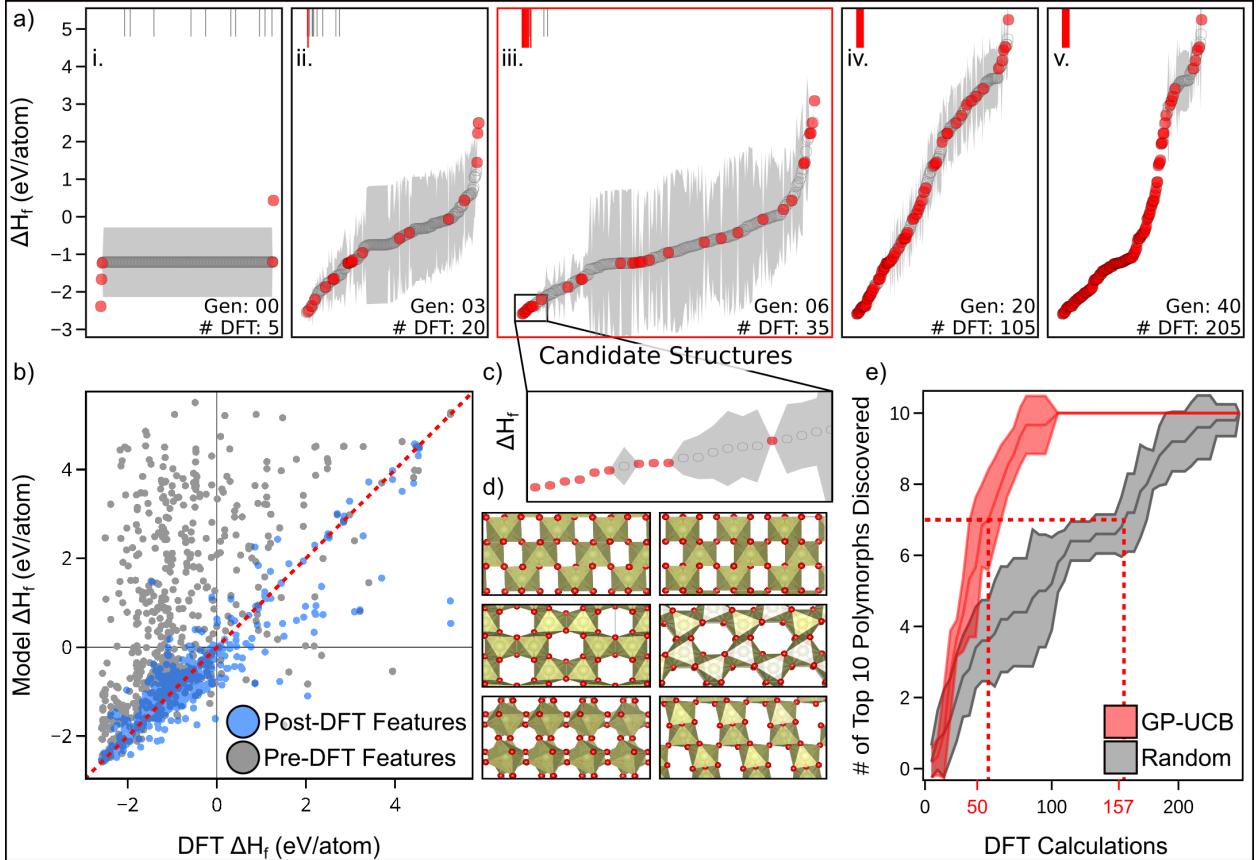


Figure 2: (a) Progress of the active learning algorithm at five subsequent generations. The AL generation and number of DFT training data at each generation is shown for each subplot. The enthalpy of formation is plotted, ordered from most to least stable, against all IrO_3 candidates. Grey markers indicate predicted formation enthalpies from the ML model while red markers correspond to DFT-computed quantities. Error bars from the GP model corresponding to 1 sigma are shown for all predictions. The vertical lines at the top of each subplot are tracking the positions of the 10 most stable polymorphs at each generation and whether they have been acquired (red) or not (grey) by the AL routine. (b) Parity plot of the final ML models for IrO_2 and IrO_3 predicting on either the pre-optimized (grey) or the post-optimized structures of IrO_2 and IrO_3 . (c) Zoomed inset of the 6th generation of the AL loop. (d) Crystal structures of 6 most stable IrO_3 polymorphs. (e) The number of the top 10 most stable polymorphs of IrO_3 that are discovered as a function of the number of DFT bulk relaxations, averaged over 5 independent runs of the AL algorithm using the GP-UCB acquisition criteria (red) and a random acquisition method (grey). Error bars indicate the standard deviation over 5 runs. Red guide lines are displayed to show how many DFT calculations are needed to discover 7/10 of most stable polymorphs for the GP-UCB and random acquisition.

III. Crystal coordination analysis of discovered phases

Next, we describe the structural variety that is present in our data set of 1000 IrO_2 and IrO_3 polymorphs. The coordination environment package ChemEnv, developed by Waroquier et al.²² and implemented in pymatgen,²³ was used to assign M-O crystal field coordination types (e.g. octahedral, square pyramidal, cubic, etc.) to each of the 1K structures in our combined data set. **Describe in two paragraphs what is shown in Figure 3 and in what the main consequences of these results** Additional esoteric coordination environments were identified manually, see SI. The resulting distribution is included in figure 3, which plots the electronic energy and volume, both normalized on a per atom basis.

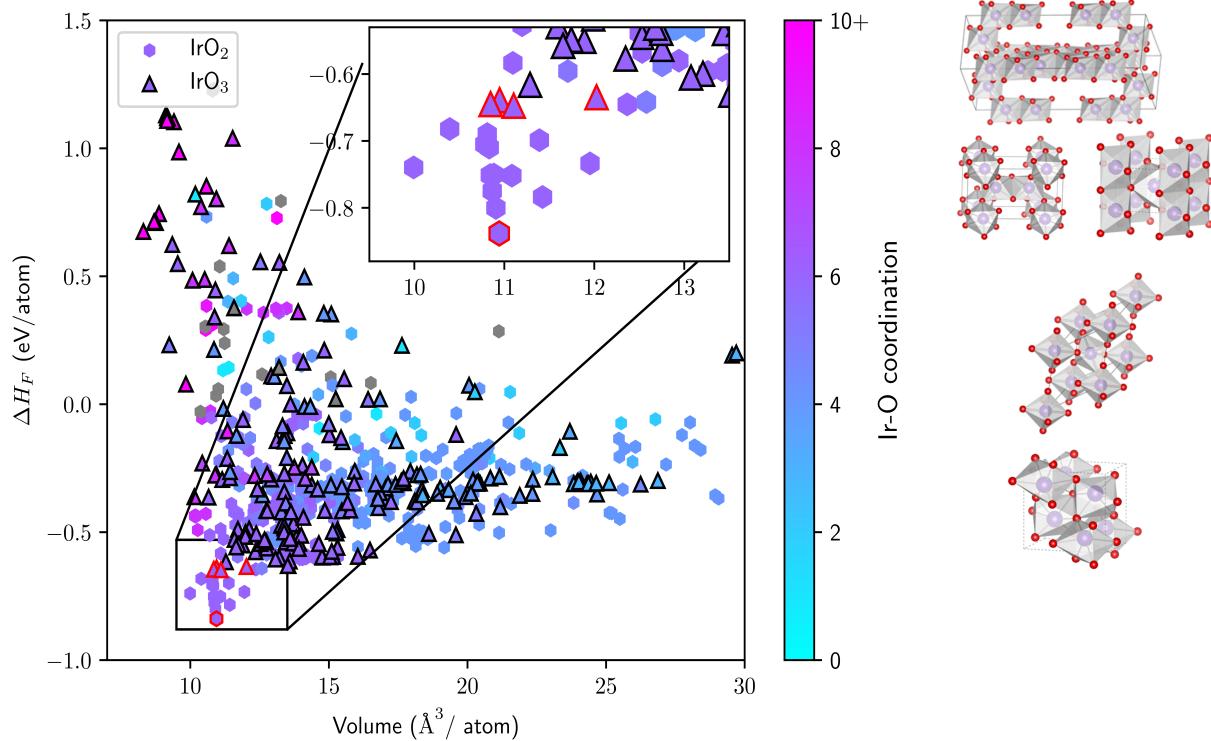


Figure 3: Enthalpy of formation for the 448 IrO_2 (circles) and 258 IrO_3 (crosses) structures in the candidate data set plotted against the volume per atom. Color overlays indicate the dominant coordination motifs as indicated by the legend. Select polymorphs systems are displayed around the plot area.

III. Electrochemical OER Application

We next performed ab-initio thermodynamic simulations to elucidate the electrochemical operational stability of IrO_x and the OER activity of the four stable polymorphs (rutile- IrO_2 , $\alpha\text{-IrO}_3$, rutile- IrO_3 , and $\beta\text{-IrO}_3$) computed previously.

Bulk Pourbaix

Figure 4 reports the IrO_x Pourbaix diagram (E vs. pH) constructed with the following species: Ir, rutile- IrO_2 , $\alpha\text{-IrO}_3$, rutile- IrO_3 , $\beta\text{-IrO}_3$, and an aqueous dissolved $\text{IrO}_4^{[4-]}$ species (See TEMP—SI for additional details). While Ir and rutile- IrO_2 are most stable at low bias, $\alpha\text{-IrO}_3$ is the most stable species under acidic conditions ($\text{pH} < 7$) and in the bias region of interest for the OER (1.23 V vs. RHE). The stability regions of the metastable rutile- IrO_3 and $\beta\text{-IrO}_3$ phases are indicated by unfilled solid lines and appear (meta?)stable in the OER relevant region of the diagram. The similar formation energies (SI TEMP) for all three IrO_3 species suggest some or all of these IrO_3 phases may be present and are stable under OER conditions.

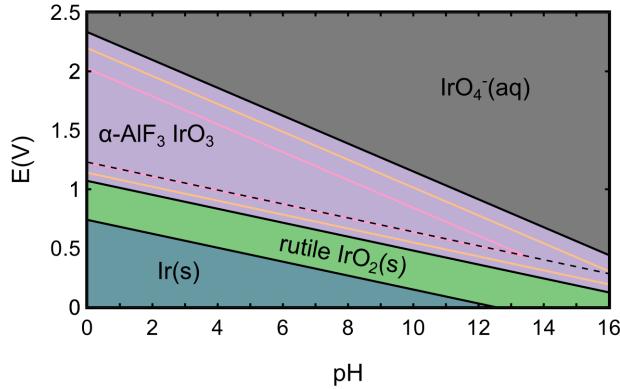


Figure 4: The revised Pourbaix diagram (electrochemical bulk phase stability) of the $\text{Ir-H}_2\text{O}$ system as a function of applied potential (vs. SHE) and pH. The most stable system studied (see Table XX in SI for a full list) are Ir-metal $\text{Ir}(s)$ (blue), a rutile- IrO_2 (green), and a dissolved $\text{IrO}_4^{[4-]}$ (gray). These are compared to the IrO_3 polymorphs, $\alpha\text{-IrO}_3$ (purple), rutile- IrO_3 (orange), and $\beta\text{-IrO}_3$ (pink). The water equilibrium line at $U=1.23$ V vs. RHE shows the ideal onset of OER.

OER Surfaces and Activities

Fig. 5 summarizes the results of the electrochemical activity and surface stability analysis, structure files and method details are reported in SI TEMP. Fig. 5 a.) reports the surface energy Pourbaix plots as a function of applied potential (at pH=0) for the four IrO_x crystals of interest. The bulk phase limits of stability from figure 4 are included at the bottom of each subplot. For each polymorph, the specific facets were chosen from the highest intensity x-ray diffraction peaks from simulated powder-diffraction spectra²⁴ (see figure TEMP SI) as well as manually cleaving the bulks along intuitive facet planes. For each facet we computed the surface free energy for three coverages, bare, $^*\text{OH}$, and $^*\text{O}$. At modest overpotentials ($\eta \sim 0.3$ or equivalently potentials of ~ 1.5 V vs RHE) the convex hull is populated solely by oxygen terminated surfaces. Consequently, we consider mainly oxygen terminated surfaces for the OER analysis.

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 $\Delta G_{\text{O}} - \Delta G_{\text{OH}}$ thermodynamic descriptor. The two rutile- IrO_2 surfaces (100 and 110) bind the OER intermediates too strongly, locating them at a theoretical limiting potential of 1.8 V vs. RHE. The predicted overpotentials of our rutile- IrO_2 systems are within the range of experimentally observed overpotentials found in literature. The three IrO_3 polymorph surfaces all have a $\Delta G_{\text{O}} - \Delta G_{\text{OH}}$ descriptor towards the top and right of the volcano, indicative of weaker binding energetics. This is evident from figure S1 in the SI, which shows a clear distinction between the IrO_2 and IrO_3 polymorphs, with IrO_3 binding on average TEMP eV weaker than IrO_2 . The best performing systems, including the (100), (110), and (211) facets of $\alpha\text{-IrO}_3$, $\beta\text{-IrO}_3$ (101), and rutile- IrO_3 (110), have overpotentials of 0.4 V vs RHE, a 0.2 V vs RHE improvement over the rutile- IrO_2 system. We note that the computed overpotentials for our rutile- IrO_2 system differs from that reported in¹ by 0.2 V. This discrepancy is due to our inclusion of spin-polarization in our ab-initio calculations, which was neglected in Seitz et al.. This discrepancy was discussed in a previous publication.²⁵

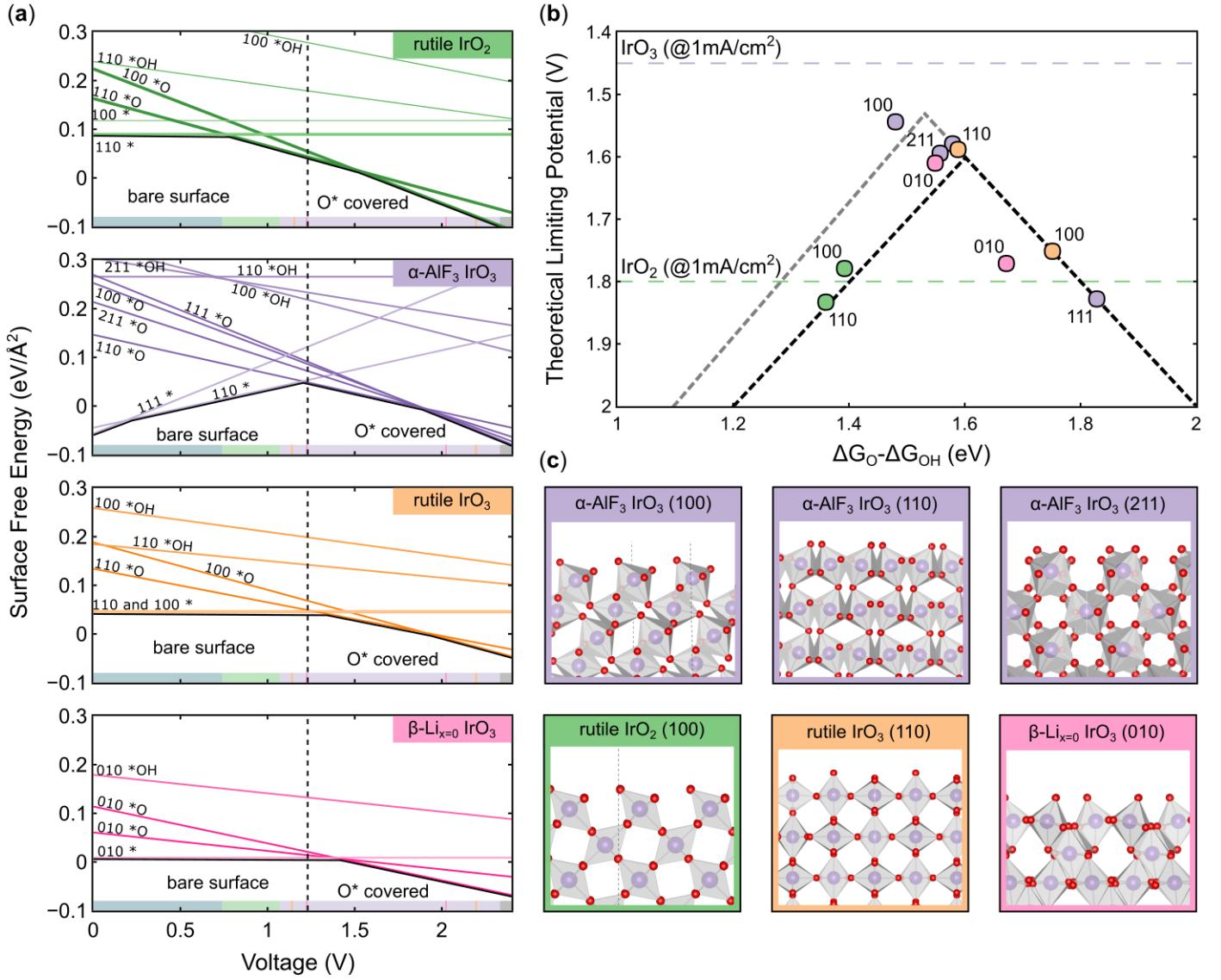


Figure 5: Summary of OER results for the four bulk structures of IrO_x considered: rutile-IrO₂ (green), α -IrO₃ (purple), rutile-IrO₃ (orange), and β -Li_{x=0}-IrO₃ (pink). (a) Surface energy Pourbaix diagrams for each structure, with the surface energy of various facets and coverages shown as a function 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 $\Delta G_O - \Delta G_{OH}$ thermodynamic descriptor. The purple dotted line corresponds to the experimental limiting potential at 10 mA cm² for IrO₃,¹ while the green band corresponds to the range of experimentally observed overpotentials for pristine IrO₂ catalysts as reported in literature. (c) Select surface facets for the four IrO_x crystal systems considered. Color legend: oxygen (red), purple (iridium), coordination motif (white)

Conclusion

In conclusion, we have demonstrated an active-learning accelerated algorithm for the discovery of stable crystal polymorphs by searching through a candidate space of structurally distinct iridium-oxide phases. The algorithm can identify 7 of the 10 most stable polymorphs of IrO_3 with only 35 DFT bulk relaxations and TEMP of the most 10 stable IrO_2 polymorphs after TEMP DFT calculations. For IrO_2 , we find.... For IrO_3 we find.... In the IrO_2 space our search failed to uncover anything more stable than the rutile- IrO_2 phase, while for IrO_3 (a much less explored stoichiometry) we found several polymorphs phases that are predicted to be stable under OER conditions. We have analyzed the local and global structural coordination and revealed a large degree of structural diversity in our dataset (octahedral, tetrahedral, square-pyramidal, cubic, and square-planar) Although octahedral coordinations are energetically preferred TEMP TEMP. The most stable systems were used to construct a revised Pourbaix diagram of Ir-H₂O system. Very importantly, we predict that IrO_3 is the thermodynamically preferred phase under OER conditions. Finally, using thermodynamical approach to OER, we show, that surfaces of selected IrO_3 have much higher relative activity than IrO_2 due to presence of high valency Ir[6+] states. The 100 percent corner-sharing octahedral structures feature maximum coverage of oxygens with optimal 2p-energy. The OER results has broader implications for related OER systems. Overall, the AL ML has tremendous potential for discovery of structurally diverse systems particularly where the known diversity is currently very small (Highly oxidized oxides). Because our method provides readily available structural information which a necessary input for any characterization/simulation analysis TEMP We envision that This opens up and new avenues of the materials/catalysis research with tailored structural properties.

Going forward we will improve

Acknowledgement

This work was supported by the Toyota Research Institute.

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. Department of Energy under Contract No. DE-AC02-05CH11231.

Supporting Information Available

Active Learning ML Section

Candidate space generation

Here we describe in more detail the methodology to construct the set of structural candidates that were considered. At a high-level, the procedure is to mine materials databases for entries with the desired non-element specific stoicheometry (i.e. AB_2 and AB_3). Since many of the entries in these databases are structurally redundant with one another, we used a structure classification scheme to reduce the data set to a structurally unique set. Once this is accomplished, the elements of interest were then substituted into the database structures, which at this point had their original elemental composition from their database. And finally, the structures are isotropically expanded or contracted to accomodate the difference between the atomic radii of the original elements in the structure and the user defined elements. Below we go into more detail for each of these steps.

Herein, we take advantage of the structural diversity already present in materials databases to construct our candidates. We utilized two materials databases, the Open Quantum Materials (OQMD) and the Materials Project (MP) databases because of their large and diverse datasets of crystalline inorganic materials. At the time that we originally mined these databases (TEMP 2018? TEMP), there were 61,471 inorganic compounds in the MP database and 435,583 entries in the OQMD database. The databases have expanded the number of entries considerably since we originally parsed them, so we'd like to caution the reader that these exact numbers are out of date. Herein, for simplicity we'll refer to the quantiative figures of these databases in the present, instead of past, tense. The structural classification scheme of Jain et al.¹⁷ was then used to classify the structure and element-nonspecific stoicheometry of each structure (497,054 in total). See section TEMP for more details on the symmetry based structural classification scheme. Once classified, we selected

all entries with the desired stoichiometry of AB_2 and AB_3 , for which MP has 2,424 and 2341 AB_2 and AB_3 entries, respectively, and OQMD has 4,736 and 28,883 AB_2 and AB_3 entries, respectively. The reason that there are considerably more AB_3 compounds in OQMD is due to the extensive compositional permutation of the cubic perovskite structure (ABO_3 ABO3) with the A cation removed. Within each stoichiometry, the structural classification was used to eliminate structurally redundant systems, i.e. systems that share their stoichiometry, space-group, and wyckoff positions. This process reduces the data set to 620 and 219 unique structures of AB_2 and AB_3 for MP and 397 and 194 structurally unique AB_2 and AB_3 OQMD entries. Combining the MP and OQMD data sets ultimately results in a dataset of 688 AB_2 and 254 AB_3 structurally unique candidates.

Table S1: (a) Number of entries in the OQMD and MP materials databases for the AB_2 and AB_3 stoichiometries. (b) Final number of unique structural candidates for AB_2 and AB_3 .

(a)

Database	Stoich.	Entries	
		Total	Unique
OQMD		61,471	
	AB_2	4,736	397
	AB_3	28,883	194
MP		435,583	
	AB_2	2,424	620
	AB_3	2,341	219

(b)

Final Candidate Set	
Stoich.	Unique Structures
AB_2	688
AB_3	254

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.¹⁷

Structure featurization and data processing

Structures are featurized using the Voronoi tessellation method developed by TEMP. 0-variance features that don't encode for any information within a set of structures with constant composition and stoicheometry are then removed (271 to 101 feature columns).

Gaussian process regression model

The Gaussian Process model utilized two isotropic Gaussian (RBF) kernels, with one kernel having a larager length scale parameter and the other one having a lengh scale parameter that operates on shorter length scales. The hyperparameters of the GP kernel are optimized every time the GP is trained such that the log marginal likelihood of the model and training data is maximized. The initial values of the hyper paramters are as follows:

GP Hyper Paramter	Initial Value
noise	0.02542
sigmaL	1.0049
sigmaF	5.19
alpha	0.018

Bulk polymorph DFT optimization

A variable k-point mesh is used such that a k-point density of at least 20 k-points per reciprocal space dimension. All bulk systems were run through the following computational recipe to converge the equilibrium structure. The recipe has 3 distinct phases, and structures are only advanced to the next phase when the previous phase completes without error.

1. A ISIF 7 calculation to optimize only the volume (initial volume of cell may be really off)
2. 3 consecutive ISIF 3 relaxations to fully converge the lattice and atomic positions
3. A final ISIF 2, calculation to relax the atomic coordinates only to avoid errors associated with changing the cell volume with a fixed plane-wave cutoff basis

The final ISIF 2 step is run with an electronic energy SCF convergence criteria of 1E-6 eV and the ionic relaxation has

a tight force convergence criteria of 1E-3 eV/Angstrom

Structural coordination motif identification

Several esoteric structural features were found in the DFT optimized structures and can be categorized as one of two types, legitimate structural motifs not characterized by the scheme of Waroquier et. al.,²² and arguably non-physical structural artifacts, including:

- unassociated oxygen atoms - molecular oxygens in the unit cell

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)²⁶⁻²⁸ and utilizing the PBE exchange-correlation functional.²⁹ Dipole corrections were imposed on all non-symmetric slabs to counteract spurious dipole interactions between the periodic cells.³⁰ A 4x4x3 k-point mesh with gamma-point centered Monkshort-packing³¹ 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 with utilizing the conjugate gradient algorithm as implemented in VASP (IBRION=2). The simulation stop criteria used was that all atoms must satisfy a maximum force threshold of 0.02 eV/Å.

OER Thermodynamic 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.

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:

OER Scaling Relations

Figure S1 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.

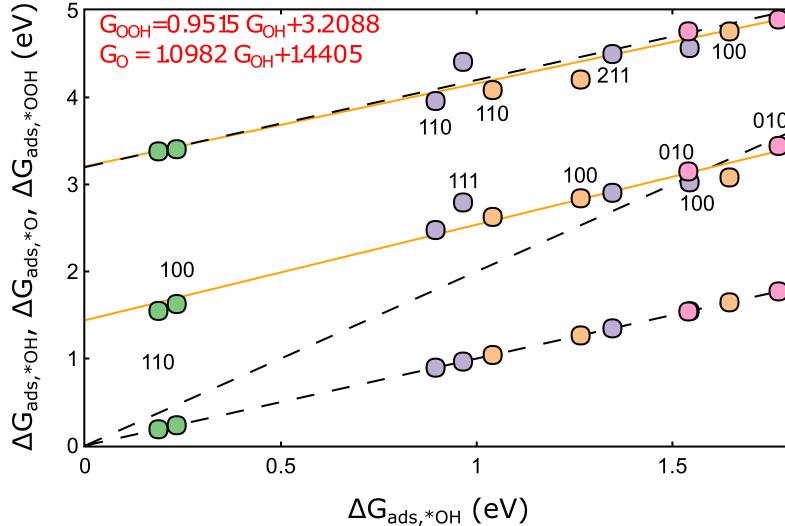


Figure S1: Relationship between the adsorption free energies of the three key OER intermediates ($*\text{OH}$, $*\text{O}$, $*\text{OOH}$), with ΔG_{OH} chosen as the dependent variable. Best fit lines are provided for ΔG_{OOH} vs. ΔG_{OH} and ΔG_{O} vs. ΔG_{OH} . Additionally, “universal scaling relations” 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 ΔG_{OH} line is shown as guide to eye.

Bulk Systems

Table of OER energetics

References

- (1) Seitz, L. C.; Dickens, C. F.; Nishio, K.; Hikita, Y.; Montoya, J.; Doyle, A.; Kirk, C.; Vojvodic, A.; Hwang, H. Y.; Norskov, J. K.; Jaramillo, T. F. A highly active and stable IrO_x/SrIrO₃ catalyst for the oxygen evolution reaction. *Science* **2016**, *353*, 1011–1014.
- (2) Lee, Y.; Suntivich, J.; May, K. J.; Perry, E. E.; Shao-Horn, Y. Synthesis and activities of rutile IrO₂ and RuO₂ nanoparticles for oxygen evolution in acid and alkaline solutions. *Journal of Physical Chemistry Letters* **2012**, *3*, 399–404.
- (3) McCrory, C. C. L.; Jung, S.; Ferrer, I. M.; Chatman, S. M.; Peters, J. C.; Jaramillo, T. F. Benchmarking Hydrogen Evolving Reaction and Oxygen Evolving

Reaction Electrocatalysts for Solar Water Splitting Devices. *Journal of the American Chemical Society* **2015**, *137*, 4347–4357.

- (4) Trotochaud, L.; Ranney, J. K.; Williams, K. N.; Boettcher, S. W. Solution-cast metal oxide thin film electrocatalysts for oxygen evolution. *Journal of the American Chemical Society* **2012**, *134*, 17253–17261.
- (5) Danilovic, N.; Subbaraman, R.; Chang, K. C.; Chang, S. H.; Kang, Y. J.; Snyder, J.; Paulikas, A. P.; Strmcnik, D.; Kim, Y. T.; Myers, D.; Stamenkovic, V. R.; Markovic, N. M. Activity-stability trends for the oxygen evolution reaction on monometallic oxides in acidic environments. *Journal of Physical Chemistry Letters* **2014**, *5*, 2474–2478.
- (6) Carmo, M.; Fritz, D. L.; Mergel, J.; Stolten, D. A comprehensive review on PEM water electrolysis. *International Journal of Hydrogen Energy* **2013**, *38*, 4901–4934.
- (7) Miles, M.; Klaus, E.; Gunn, B.; Locker, J.; Serafin, W.; Srinivasan, S. The oxygen evolution reaction on platinum, iridium, ruthenium and their alloys at 80C in acid solutions. *Electrochimica Acta* **1978**, *23*, 521–526.
- (8) Beni, G.; Schiavone, L. M.; Shay, J. L.; Dautremont-Smith, W. C.; Schneider, B. S. Electrocatalytic oxygen evolution on reactively sputtered electrochromic iridium oxide films [7]. **1979**, *282*, 281–283.
- (9) Pearce, P. E.; Perez, A. J.; Rousse, G.; Saubanère, M.; Batuk, D.; Foix, D.; McCalla, E.; Abakumov, A. M.; Van Tendeloo, G.; Doublet, M. L.; Tarascon, J. M. Evidence for anionic redox activity in a tridimensional-ordered Li-rich positive electrode β -Li₂IrO₃. *Nature Materials* **2017**, *16*, 580–586.
- (10) Jennings, P. C.; Lysgaard, S.; Hummelshøj, J. S.; Vegge, T.; Bligaard, T. Genetic algorithms for computational materials discovery accelerated by machine learning. *npj Computational Materials* **2019**, *5*, 46.

- (11) Hansen, M. H.; Torres, J. A. G.; Jennings, P. C.; Wang, Z.; Boes, J. R.; Mamun, O. G.; Bligaard, T. An Atomistic Machine Learning Package for Surface Science and Catalysis. *arXiv preprint arXiv:1904.00904* **2019**,
- (12) Torres, J. A. G.; Jennings, P. C.; Hansen, M. H.; Boes, J. R.; Bligaard, T. Low-Scaling Algorithm for Nudged Elastic Band Calculations Using a Surrogate Machine Learning Model. *Physical review letters* **2019**, *122*, 156001.
- (13) Van Den Bossche, M.; Grönbeck, H.; Hammer, B. Tight-Binding Approximation-Enhanced Global Optimization. *Journal of Chemical Theory and Computation* **2018**, *14*, 2797–2807.
- (14) Stillinger, F. H. Exponential multiplicity of inherent structures. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics* **1999**, *59*, 48–51.
- (15) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K. A. Commentary: The materials project: A materials genome approach to accelerating materials innovation. *APL Materials* **2013**, *1*.
- (16) Kirklin, S.; Saal, J. E.; Meredig, B.; Thompson, A.; Doak, J. W.; Aykol, M.; Rühl, S.; Wolverton, C. The Open Quantum Materials Database (OQMD): Assessing the accuracy of DFT formation energies. *npj Computational Materials* **2015**, *1*.
- (17) Jain, A. PHYSICAL REVIEW B 98 , 214112 (2018) Atomic-position independent descriptor for machine learning of material properties. **2018**, *214112*, 1–7.
- (18) 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**,

- (19) Tipping, M. E.; Bishop, C. M. Probabilistic principal component analysis. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **1999**,
- (20) CatLearn: An environment for atomistic machine learning in Python for applications in catalysis. 2019; <https://github.com/SUNCAT-Center/CatLearn>.
- (21) Su, C.; Lv, J.; Li, Q.; Wang, H.; Zhang, L.; Wang, Y.; Ma, Y. Construction of crystal structure prototype database: Methods and applications. *Journal of Physics Condensed Matter* **2017**, *29*.
- (22) Waroquiers, D.; Gonze, X.; Rignanese, G. M.; Welker-Nieuwoudt, C.; Rosowski, F.; Göbel, M.; Schenk, S.; Degelmann, P.; André, R.; Glaum, R.; Hautier, G. Statistical analysis of coordination environments in Oxides. *Chemistry of Materials* **2017**, *29*, 8346–8360.
- (23) Ong, S. P.; Richards, W. D.; Jain, A.; Hautier, G.; Kocher, M.; Cholia, S.; Gunter, D.; Chevrier, V. L.; Persson, K. A.; Ceder, G. Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Computational Materials Science* **2013**, *68*, 314–319.
- (24) Momma, K.; Izumi, F. VESTA 3 for three-dimensional visualization of crystal, volumetric and morphology data. *Journal of Applied Crystallography* **2011**, *44*, 1272–1276.
- (25) Strickler, A. L.; Flores, R. A.; King, L. A.; Nørskov, J. K.; Bajdich, M.; Jaramillo, T. F. Systematic Investigation of Iridium-Based Bimetallic Thin Film Catalysts for the Oxygen Evolution Reaction in Acidic Media. *ACS Applied Materials and Interfaces* **2019**, *11*, 34059–34066.
- (26) Kresse, G. Ab initio molecular dynamics for liquid metals. *Journal of Non-Crystalline Solids* **1995**, *192-193*, 222–229.

- (27) Kresse, G.; Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Computational Materials Science* **1996**, *6*, 15–50.
- (28) Kresse, G.; Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Physical Review B - Condensed Matter and Materials Physics* **1996**, *54*, 11169–11186.
- (29) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Physical Review Letters* **1996**, *77*, 3865–3868.
- (30) Neugebauer, J.; Scheffler, M. Adsorbate-substrate and adsorbate-adsorbate interactions of Na and K adlayers on Al(111). *Physical Review B* **1992**, *46*, 16067–16080.
- (31) Monkhorst, H. J.; Pack, J. D. Special points for Brillouin-zone integrations. *Physical Review B* **1976**, *13*, 5188–5192.

Graphical TOC Entry

