Data-Driven High-Throughput Rational Design of Double-Atom Catalysts for Oxygen Evolution and Reduction

Lianping Wu, Tian Guo, and Teng Li*

Surging interests exist in double-atom catalysts (DACs), which not only inherit the advantages of single-atom catalysts (SACs) (e.g., ultimate atomic utilization, high activity, and selectivity) but also overcome the drawbacks of SACs (e.g., low loading and isolated active site). The design of DACs, however, remains cost-prohibitive for both experimental and computational studies, due to their huge design space. Herein, by means of density functional theory (DFT) and topological information-based machine-learning (ML) algorithms, we present a data-driven high-throughput design principle to evaluate the stability and activity of 16 767 DACs for oxygen evolution (OER) and oxygen reduction (ORR) reactions. The rational design reveals 511 types of DACs with OER activity superior to IrO2 (110), 855 types of DACs with ORR activity superior to Pt (111), and 248 bifunctional DACs with high catalytic performance for both OER and ORR. An intrinsic descriptor is revealed to correlate the catalytic activity of a DAC with the electronic structures of the DAC and its bonding carbon surface structure. This data-driven high-throughput approach not only yields remarkable prediction precision (>0.926 R-squared) but also enables a notable 144 000-fold reduction of screening time compared with pure DFT calculations, holding promise to drastically accelerate the design of high-performance DACs.

1. Introduction

Single-atom catalysts (SACs) have attracted great interest as a new type of catalysts due to a range of advantages over their bulk counterpart, including 100% atomic utilization, high selectivity, and great catalytic activity toward diverse reactions, such as oxygen reduction reaction (ORR) and oxygen evolution reaction (OER).[1–5] However, SACs suffer from two major drawbacks. First, most catalytic reactions usually involve the co-adsorption of multiple reactants and the formation of multiple chemical bonds on the catalyst surface by intermediate molecules, but SACs only have one active site.[6–8] Second, to maintain the structural stability of SACs, the distance between adjacent SAC sites needs to be relatively large (>>1 nm). Therefore, the loading of SACs is usually low, which results in unsatisfactory catalytic activity.[9,10]

Double-atom catalysts (DACs) are emerging as promising catalysts, which inherit the advantages of SACs (e.g., ultimate atomic utilization, high activity, and selectivity) and overcome the drawbacks of SACs.[7,11] For example, DACs can anchor more metal atoms at an adsorption site in comparison with SACs, which leads to a higher metal loading than SACs. Furthermore, DACs provide adjacent unsaturated coordination sites, which can absorb multiple reactants simultaneously.[12,13] As a result, there exist multiple adsorption modes of key intermediates, which enable optimized reaction pathways and thus result in significantly improved catalytic activity and selectivity.[14] However, a significant challenge exists in designing highly efficient DACs, essentially due to the vast combinational design space. Take the design of DACs on a carbon surface as an example. There are 841 combinations of transition metal DAC compositions and more than 100 possible types of surface defects on nitrogen-doped carbon as the DAC anchoring sites, leading to tens of thousands of possible compositional and configurational designs of DACs. Such a huge design space makes both experimental explorations via trial-and-error and computational studies based on pure density functional theory (DFT) cost- and time-prohibitive. Hence, it is highly desirable to develop an efficient and effective method to predict the structural stability and electrocatalytic performance of all transition metal DACs.

Machine learning (ML) is an efficient regression method that has begun to be utilized to quickly screen catalysts with high electrocatalytic performance.[3,16] For example, ML can predict the OER/ORR and hydrogen evolution reaction performance directly from the atomic scale properties of graphene-supported SACs.[7,18] Identify active sites for CO2 reduction on alloyed gold and describe the CO2 reduction performance of SACs with the help of the atomic-scale properties of the metal.[19,20] Although the ML approach has been adopted in studying SACs and DACs on N-doped carbon materials previously,[21–26] these existing studies often use the atomic properties of the metals (e.g., electronegativity, d-orbital electron count, metal–carbon...
bond length and angles, etc.) as relevant descriptors of the electrocatalytic performance of the SACs and DACs. Such descriptors can capture the key factors that govern the SAC and DAC performance but suffer from a major limitation: they cannot distinguish between the difference in locations and species of the substrate atoms (e.g., carbon vs nitrogen doping atoms) surrounding the metal SACs and DACs. As a result, existing studies often ignore the variations of the substrate defect environment and only consider one type of substrate defect. Such missing information has a significant influence on the electrocatalytic performance of the metal SACs. For example, the DFT calculation shows that the difference in limiting potentials between the SAC surrounded by four pyridine nitrogen atoms and that surrounded by four pyrrole nitrogen atoms is as large as 38%.[25] Topological information-based ML algorithms not only eliminate the need for data preprocessing (e.g., sorting and formatting) but also allow for a precise description of complex data features (e.g., symmetry and topological structures).[26] These unique advantages have led to surging interests in applying this ML algorithm in various domains, such as social networks,[27] drug discovery,[28] materials science,[29,30] and structural biochemistry[31] (e.g., AlphaFold[32]). It is expected that such ML algorithms can be applied to the rational design of a wide range of atomic catalysts.

Here, we report a DFT-based, topological information-based ML accelerated rational design of DACs. We first perform DFT calculations to evaluate the stability, OER, and ORR catalytic performance of 164 DAC systems. Then, we use the DFT results from the 164 DAC systems as the training and testing sets of data for an ML algorithm, which in turn can predict the stability, OER, and ORR catalytic performance of DACs of 16767 possible combinations of transition metal compositions and substrate configurations. This high-throughput design approach reveals 511 types of DACs with OER activity superior to IrO$_2$ (110), the most efficient catalyst for OER, and 855 types of DACs with ORR activity superior to Pt (111), the most efficient catalyst for ORR. Furthermore, we identify 248 bifunctional DACs with catalytic performance exceeding both IrO$_2$ (110) for OER and Pt (111) for ORR. An intrinsic descriptor that correlates the catalytic activity of a DAC with the topological and electronic structures of the DAC and its bonding carbon surface structure is revealed. This approach not only yields remarkable prediction precision (> 0.926 R-squared) but also enables a drastic 144 000-fold reduction in screening time compared with pure DFT calculations. The findings of the present study hold great promise for accelerating the discovery of high-performance and low-cost DACs at a drastically reduced cost. The ML-enabled strategy demonstrated in this paper can be readily adapted to enable and accelerate the discovery of a broad range of atomic-scale catalysts.

2. Results and Discussion

2.1. Atomic Structure of DACs on a Carbon Surface

Transition metals have emerged as promising atomic catalysts because they have an incomplete d-orbital that allows the metal to exchange electrons, thus easily giving and accepting electrons. However, a challenge exists to stabilize them because these metal atoms tend to merge and form nanoparticles. Many previous experimental and theoretical studies have shown that the vacancy defect on a carbon surface can stabilize the metal atom effectively.[33–35] Besides, carbon surfaces also have other desirable advantages in electrochemistry, including lightweight, low cost, adjustable porosity, and high chemical and thermal stability.[36,37] Hence, we study the stability of transition metal DACs on a defective carbon surface and the resulting OER/ORR performance of the DACs. Moreover, nitrogen-doped (N-doped) carbon surfaces have better electrocatalytic properties than pure carbon surfaces.[38,39] Therefore, we also consider the effect of N-doping on the stability and catalytic performance of carbon-supported DACs.

Figure 1 shows the atomic structures of transition metal DACs on a defective site of various types on a carbon surface, which is formed by replacing four carbon atoms with two transition metal atoms and doping the surrounding carbon atoms with nitrogen atoms. We delete four carbon atoms and then add two metal atoms at the defective site. The initial position of

Figure 1. Schematic of the atomic structures of transition metal DACs on a carbon surface. a). Twenty-three types of carbon surface defect structures are categorized into seven groups by the N-doping level. For example, 4C-2N denotes two out of the six surrounding carbon atoms are replaced by nitrogen atoms and b). 729 (27 by 27) possible compositional combinations of transition metal DACs lead to in total of 16767 (729 by 23) atomic structures of DACs on a carbon surface.
metals is in the same plane as carbon atoms, and metal–metal atoms are neighboring binding. We use a conjugate algorithm to optimize the geometry until the atomic force’s convergence threshold is 0.01 eV Å$^{-1}$. This optimized atomic structure is confirmed to be the dominant DAC structure in experiments$^{[40]}$ and further confirmed by our modeling (see Figure S1, Supporting Information for details). Different combinations of the numbers and positions of doped nitrogen atoms lead to 23 possible defect structures on the carbon surface (Figure 1a). The 23 atomic structures of the defect site on a carbon surface can be categorized into seven groups by the N-doping level, as labeled in Figure 1a. Eight defect structures are symmetric among the 23 atomic structures (i.e., 6C, 6N, the first columns in 4C-2N and 2C-4N), and the remaining 15 defect structures are asymmetric.

In principle, all 3d, 4d, and 5d transition metal atoms can serve as atomic catalyst materials. However, due to their toxic/radioactive nature, Tc and Hg are excluded from this study (Figure 1b). Hence, there are 729 (27 by 27) possible compositional combinations of transition metal DACs. Thus, there are 16 767 (729 by 23) possible atomic structures of transition metal DACs on a defect site of a carbon surface. Next, we develop a highly efficient strategy by combining ML and DFT calculations to evaluate the stability and electrocatalytic activities of all 16 767 possible cases of DACs on a carbon surface, a highly desirable task prohibitive by other approaches.

### 2.2. Topological Information-Based ML Algorithm

Here, we propose a topological-based ML algorithm to evaluate the stability and catalytic activity of DACs. The ML algorithm is based on the topological information in twofold: the atomic properties of the metal atoms and the substrate atoms (i.e., nodal information), as well as the atomic structure of the corresponding DAC and its surrounding substrate defect (i.e., topological structure). We show that such a unified ML model can be trained with a reasonable number of DFT calculation results and then predict the structural stability and electrocatalytic performance of all 16 767 possible DACs with high efficiency and precision. Figure 2 illustrates the workflow of the topological information-based ML model to screen DACs with optimal electrocatalytic performance from a vast design space, which contains three steps:

**Step 1. Training data generation.** The training data include target data and input data. The target data are the properties of DACs that we aim to investigate and will be used for ML model training. To obtain the target data for predicting the structural stability of a DAC, we calculate the binding energy between the DACs and the carbon surface for a certain number of DAC systems. To obtain the target data for predicting electrocatalytic activity (OER and ORR) of a DAC, we calculate the overpotentials of OER and ORR for a certain number of DAC systems (see computational details in Methods). The input data is the atomic properties of the transition metal DAC and the carbon surface. We first identify eleven atomic properties based on the consideration that the electronic characteristics of catalytic centers influence the electron transfer and reaction energy in the electrocatalytic activity.$^{[41–43]}$ The eleven atomic properties can be classified into two categories (with values listed in Table S1, Supporting Information): atomic dimension and structure (atomic mass $M$, atomic radius $r_M$, $d$-electron count $d$).
Step 2. Training/testing via a topological information-based algorithm. The ML model is composed of multiple processing layers to extract the features of input data with multiple levels of abstraction. In consideration of a transition metal DAC supported on a carbon surface with different coordination environments, we use the topological information to represent the geometric structure of the transition metal atoms and surrounding substrate atoms. Specifically, we set the atomic properties identified in Step 1 as the node information and then build the links according to the bonding states of the metal atoms in the DAC with their surrounding substrate atoms. The established topological information is used to establish three sub-nets as part of the ML model (Figure S3, Supporting Information). At the end of the ML training and testing step (see details in Methods), the ML model with the highest R-squared value on the testing dataset is selected as the best performance model to predict the binding energy, overpotentials of OER and ORR (Figure 2, Step 2). Furthermore, the importance of the input features is investigated via the model agnostic technique. We use the SHapley Additive exPlanations (SHAP) value to evaluate the importance of the ten atomic properties, as shown in Figure S5 (Supporting Information). We find that the describing adsorption properties are the most important feature in predicting the binding energy. At the same time, the Pauling electronegativity is the most important feature in predicting the overpotentials of OER and ORR.

Step 3. Screening DACs based on ML prediction. The best performing ML models obtained in Step 2 are then used to predict the corresponding properties (i.e., binding energy, overpotentials of OER and ORR) of DACs of the remaining possible atomic structures. Based on the ML prediction, we sort and screen the best DACs with the desired catalytic performance. The detailed results will be introduced in the next section.

2.3. Stability of DACs on a Carbon Surface

The high stability of DACs on a carbon surface is an essential characteristic of atomic catalysts. To evaluate the stability of carbon surface supported DACs, we define the binding energy of a DAC, \( E_b \), as the difference between the energy of the bonded atomic structure and the sum of the energies of the carbon surface with the same defect structure and the freestanding metal DACs to be trapped onto the defect site (Figure 3a, see computational details in Methods). A negative value of \( E_b \) denotes a stable DAC on the carbon surface. The more negative the value of \( E_b \), the more stable the DAC on the carbon surface. To generate the database for ML training, we choose 164 DAC systems for the DFT calculation by the following two steps: 1) we first choose seven carbon surface defect structures with different numbers of doping nitrogen (0N, 1N, 2N, 3N, 4N, 5N, and 6N as labeled in green in Figure 2a); 2) we then randomly choose 27, 23, 23, 23, 23, 23 DAC metal pairs for 0N, 5C-1N, 4C-2N, 3C-3N, 2C-4N, 1C-5N, and 6N carbon surface defect structures, respectively. In order to test the effect of the N element on the catalytic performance of the DACs, we would like that the selected metal pairs for 0N carbon surface defect structure include one of the 23 metal pairs selected for 5C-1N, 4C-2N, 3C-3N, 2C-4N, 1C-5N, and 6N carbon surface defect structures, respectively. Since the selected 23 metal pairs for the 0N carbon surface already have two metal pairs which are respectively the same as the metal pairs for 4C-2N and 1C-5N, we choose four more metal pairs for 0N carbon surfaces, and these four metal pairs are randomly selected from the selected 23 metal pairs of 5C-1N, 3C-3N, 2C-4N, and 6N, respectively.

The DFT calculated database is divided into two subsets with 150 data points for training ML models and the remaining 14 data points for testing the prediction precision of the trained ML model. The ML model with the best performance shows impressive precision in predicting the binding energy between...
the DAC and the supporting carbon surface in all remaining 16603 atomic structures, yielding an R-squared of 0.983 for the training data, 0.926 for the testing data (Figure 3a). Figure 3b plot the binding energy of all DACs on a symmetric defect structure (6C). The binding energy of DACs on all 23 defect structures is also plotted in Figure S6 (Supporting Information). Three observations emerge from the results in Figure 3 and Figure S6 (Supporting Information): 1) All DACs considered in our study can be stably anchored by a defect site on a carbon surface (i.e., a negative $E_b$); 2) DACs containing Au, Zn, or Cd, in general, have lower stability than other DACs on the same carbon defect site; 3) For a DAC with two metal atoms of different species on an asymmetric defect site, the position swap of the two metal atoms results in a change of the binding energy of the DAC, ranging from 2.16% to 42.44%.

The above results reveal the pivotal role of the substrate surface defect in stabilizing a DAC, similar to that in stabilizing a SAC,[38] suggesting an effective strategy to synthesize stable atomic catalysts by engineering the substrate surface defects.

2.4. Electrocatalytic Activity of Carbon Surface Supported DACs

We next evaluate OER and ORR performance in an acidic electrolyte (see computational details in Methods).[48] We show the calculated OER and ORR free-energy diagrams at 1.23 eV for the Cu-Ni DACs at a 6C carbon surface (randomly selected as an example) in Figure S7 (Supporting Information). We choose to show the OER/ORR energy diagrams at 1.23 eV because the reaction-free energies of all four proton-transfer steps should be the same at zero potential (4.92 eV/4 = 1.23 eV) for an ideal catalyst. Via this diagram, we can directly evaluate the electrocatalytic activity using the limiting reaction barrier determined from the free energy of the rate-determining step (RDS). For example, for the OER reaction on Cu-Ni DAC, the RDS is the oxidation of O$^\ast$ to OOH$^\ast$ with a limiting barrier as large as 0.97 eV. For the ORR reaction on Cu–Ni DAC, the RDS is the desorption of OOH$^\ast$ with a limiting barrier of 0.39 eV. In other words, the overpotential of OER and ORR of Cu–Ni DAC on ON carbon surface is 0.97 and 0.39 eV, respectively. According to the definition of overpotential in this study, the lower the overpotentials of OER and ORR, the higher the OER and ORR electrocatalytic performance.[49]

To generate the database for ML training, the OER and ORR overpotentials of the selected 164 DAC systems in the previous section are calculated. The DFT calculated database is divided into two subsets with 150 data points for training ML models and the remaining 14 data points for testing the prediction precision of the trained ML model. In comparison with DFT results, the trained ML model with the best performance yields an R-squared value of the predicted overpotentials of OER of 0.957 on the training subset and 0.904 on the testing subset, respectively (Figure 4a). The corresponding ML performance indices for the prediction of ORR overpotentials are 0.982, and 0.96, respectively (Figure 4b). These high R-squared values clearly demonstrate the impressive precision of the ML predictions based on topological information-based algorithms.

We then use the ML model with the best performance to predict the OER and ORR electrocatalytic performance of all the remaining 16603 DAC atomic structures on a carbon surface. Figure 4b and 4d plot the overpotentials of OER and ORR of all DACs on a symmetric defect structure (6C), respectively. The overpotentials of OER and ORR of DACs on all 23 defect structures are also plotted in Figures S8 and S9 (Supporting Information), respectively. Three observations emerge from the results in Figure 4 and Figures S8 and S9 (Supporting Information):

1) The DACs with good OER performance also have good ORR performance. This is because both OER and ORR performance depend on metal’s adsorption and desorption ability for O, OH, and OOH. DACs with good OER performance show good metal adsorption and desorption ability for O, OH, and OOH. This good adsorption and desorption ability of DACs in turn results in their good performance in ORR. The good adsorption and desorption ability means that the adsorption energy of intermediates is sufficiently high to make sure the intermediates stay on the DAC. Meanwhile, the adsorption energy should not too high to make sure the intermediate can desorb from the DAC for the next step reaction. 2) The position swap of the two metal atoms in a DAC on an asymmetric defect site results in a change in its overpotentials of OER (ranging from 3.12% to 32.35%) and ORR (ranging from 1.06% to 30.26%), indicating the importance of surrounding defect structures on the catalytic performance of DACs. 3) The DACs with Ni, Cu, Zn, Pd, Ag, Cd, Pt, and Au are predicted to exhibit enhanced OER and ORR performance.

2.5. Screening for Best Performing DACs

Our topological information-based-ML model not only predicts the stability and OER/ORR electrocatalytic performance of DACs with high accuracy but can also drastically reduce the computation cost. As shown in Figure S10 (Supporting Information), ML-based predictions of the catalytic activity of transition metal DAC are more than 144000 times faster than the DFT-based calculation. Such a remarkable reduction in computation cost by the ML accelerated material discovery approach effectively overcomes the challenge of the low efficiency of DFT calculations in identifying high-performance transition metal DACs.

Although the binding energy and overpotentials of OER and ORR of all DACs can be predicted with high accuracy, the intrinsic characteristics of the transition metal that dominate the OER/ORR performance of DACs remain elusive. Here, we propose a descriptor $\phi$, which aims to capture the effect of the key parameters that govern the ORR and OER activity of DAC. In constructing the descriptor, we took the following into consideration: the valence electrons in the occupied $d$ orbitals of the metal element ($\ell$), the electronegativity ($\chi$), and the number of N atom and C atom around the metal pair.

$$\phi = \frac{\overline{\ell} + \overline{\chi}}{N_\text{C}\chi_{\text{C}} + N_\text{N}\chi_{\text{N}}} \quad (1)$$

where $\overline{\ell}$ represents the average of the valence electrons in the occupied $d$ orbitals of the two transition metals; $\overline{\chi}$, $\chi_{\text{C}}$, and $\chi_{\text{N}}$ represent the average of the electronegativity of the two
transition metals, the electronegativity of carbon and nitrogen atoms, respectively; \( N_C \) and \( N_N \) represent the number of the closest neighboring carbon atoms and nitrogen atoms for the DAC, respectively. The numerator and denominator of the descriptor denote the effect of the metal pair and the carbon surface, respectively. More specifically, we introduce the average valence electron and electronegativity of the metal pair into the numerator of the descriptor because the contribution of \( d \)-states to the adsorption strength is proportional to the coupling Hamiltonian matrix element\(^{[50]} \) which is correlated to the spatial extent of the metal \( d \)-orbital (associated with the \( d \)-band center\(^{[51]} \)) and the adsorption distance (can be estimated in terms of electronegativity\(^{[52]} \)). We introduce the number of C and N atoms and their electronegativity into the denominator of the descriptor because some strong ligands may become poisoning factors that deactivate catalysts\(^{[53]} \).

Figure 5a,b plots the ML predicted overpotentials of OER and ORR of all 16767 possible DACs and DFT-calculated overpotential of the randomly selected 164 DACs as a function of the descriptor \( \phi \), respectively. Such a holistic comparison reveals that the larger the value of \( \phi \) for a DAC, the better OER and ORR performance of the DAC, as verified by both the DFT calculated and ML predicted results. IrO\(_2\) (110) and Pt (111) have been unanimously considered the most efficient catalysts for OER and ORR, respectively, and thus widely used as the reference electrocatalysts in experiments. The overpotentials of OER of the IrO\(_2\) (110) surface and ORR of the Pt (111) surface are 0.43 and 0.45 eV, respectively\(^{[25]} \) As shown in Figure 5a,b, there exist 511 DACs (with \( \phi \) value in the range 13–18) that show a better OER performance than IrO\(_2\) (110) and 855 DACs (with \( \phi \) value in the range 10–18) that show a better ORR performance than Pt (111). A complete list of these high-performance DACs is provided in the Supporting Information. The constituent transition metal elements of these high-performance DACs include Cu, Zn, Pd, Ag, Cd, Ir, Pt, and Au.

Figure 5c plots the overpotentials of OER and ORR for all 16767 DACs, suggesting a positive correlation between OER and ORR activities of DACs. Certain applications, such as rechargeable zinc-air batteries, require bifunctional catalysts that are active for both OER and ORR. To this end, we also identify 248 DACs with exceptional catalytic performance for both OER and ORR in comparison with IrO\(_2\) (110) and Pt (111), respectively, as highlighted in Figure 5c and further plotted in Figure 5d for visual clarity. Among these 248 bifunctional high-performance DACs, 64 are made of non-noble transition metals; 128 are made of both noble and non-noble transition metals, and 56 are made of noble transition metals (see Supporting Information for a complete list). The constituent transition
metal elements of these bifunctional high-performance DACs are Cu, Zn, Pd, Ag, Cd, Pt, and Au. The best performing DAC for OER is Pt-Cd and the best performing DAC for ORR is Ag-Zn.

To validate the precision of our topological information-based ML algorithm, we also perform DFT calculations to compute the OER and ORR overpotentials of the nine best performing DACs on a 6C carbon defect site. Figure 5e,f compares the ML predicted and DFT calculated overpotentials of OER and ORR, respectively, which agree well with each other (e.g., with a mean absolute error of 0.07 and 0.1 eV for a mean of 0.43 and 0.45 eV for OER and ORR overpotentials, respectively), which demonstrate the high precision of our ML predictions.

To further benchmark the accuracy of the predicted electrocatalytic performance from our topological information-based ML algorithm, we also compared the predicted ORR overpotentials

Figure 5. Screening for best performing DACs. a,b) The overpotential of a) OER and b) ORR from both DFT calculations and ML predictions as a function of an intrinsic catalytic activity descriptor of $\phi$; IrO$_2$ (110) and Pt (111) have been unanimously considered the most efficient catalysts for OER and ORR, respectively, whose overpotential is labeled by a dashed line as the benchmark. c) the overpotential of OER and ORR of all transition DACs on 23 different carbon surface structures; 248 best performing bifunctional DACs with exceptional OER and ORR performance (i.e., lower overpotentials) are highlighted in red and further plotted in (d) for visual clarity; e–f) comparison of e) the OER overpotential and f) ORR overpotential of nine best performing DACs on a 6C carbon defect; and g) experimental measurement of the ORR half-wave potentials of six different DACs and the corresponding DFT calculated ORR overpotentials, which reveal the same order of the ORR catalytic activities of these six DACs: Fe-Co$^{[54]}$ > Fe-Cu$^{[55]}$ > Fe-Ni$^{[56]}$ > Fe-Mn$^{[57]}$ > Zn-Co $^{[58]}$ > Fe-Fe.$^{[59]}$
of six compositions of DACs with existing experimental measurements in acidic electrolytes. So far, there is only very limited experimental data of DACs available. Figure 5g plots the experimental measurement of the ORR half-waare potentials of six different DACs and the corresponding DFT calculated ORR overpotentials, which reveal the same decreasing order of the ORR catalytic activities of these six DACs: Fe-Co[56] > Fe-Cu[57] > Fe-Ni[58] > Fe-Mn[59] > Zn-Co[60] > Fe-Fe[61] (note that the higher the half-wave potential, the lower the overpotential, and the higher the catalytic activity). Considering that these measurements are from six independent experimental studies (from the synthesis of DACs to ORR half-wave potential measurement), the agreement shown in Figure 5g serves as strong evidence of the accuracy of our topological information-based ML algorithm.

Note that the solvent effect could result in gaps between DFT calculations and the practical electrochemical environment.[25,59] And the charge between the polar reaction intermediates and the H2O could also affect the binding strength of reaction intermediates on catalysts, especially in 2D materials.[60] However, the most important criteria for evaluating the activity of the OER and ORR electrocatalysts, namely, the overpotential, are obtained by comparing the energies of two or multiple intermediates, and the energy change caused by charges and solvation on the reaction species commonly are in the same degree,[61] leading to “error cancellation”. Furthermore, Xu et al. examined whether solvation could change the promising SACs by calculating the corresponding adsorption free energies of adsorbates, the reaction energy of each elementary step, as well as electrocatalytic activity on SACs with solvation corrections.[15] Their calculation results show that solvation effects have little influence on the application of the universal design principle for identifying promising electrocatalyst candidates. Since our study aims to identify promising electrocatalyst candidates, the solvent effect is not considered in our calculations.

The findings in the present study offer critical insight into the rational design of DACs and hold great promise for accelerating the discovery of high-performance and low-cost DACs at a drastically reduced cost. The ML-enabled material design strategy demonstrated in this paper can be readily adapted to enable and accelerate the discovery of a broad range of atomic-scale catalysts. It is worth noting that the electrochemical stability of DACs on a carbon surface should also be considered when the desired DACs are selected for the experiment. For example, a few homonuclear DACs, such as Sc-Sc, Y-Y, Zr-Zr, and HF-Hf, are electrochemically unstable at acid conditions (Figure S11, Supporting Information). Further investigations are needed to understand the electrochemical stability of heteronuclear DACs.

DACs have evolved to become a new frontier of electrochemical research. In the present study, we mainly investigate the activity of OER and ORR. The ORR proceeds via two possible pathways. The four-electron pathway involves the complete reduction of oxygen to water, whereas, in the two-electron pathway, oxygen is partially reduced to H2O2, which is directly utilized or further reduced to water. The four-electron transfer is important for the efficient operation of polymer electrolyte fuel cells, whereas the ORR with a partial two-electron transfer can serve as a versatile method for producing industrially important hydrogen peroxide (H2O2). Hence, investigating the selectivity of DACs is also important for guiding the experimental design of DACs, and deserves further studies.

3. Experimental Section

DFT Calculation: All spin-polarized density functional theory (DFT) calculations were performed using the Vienna Ab initio Simulation Package (VASP).[62] The projector-augmented wave method[63,64] and Perdew-Burke-Ernzerhof[65] functional were used. The Kohn–Sham wave functions were expanded in a plane-wave basis set with a cutoff energy of 500 eV. The Brillouin zone was sampled by the 2 × 2 × 1 Monkhorst–Pack k-point mesh. Period boundary conditions were applied in both in-plane and interlayer directions. A vacuum region of 15 Å was created to ensure negligible interaction between mirror images. 12.48 Å × 12.87 Å cubic supercells were built. The convergence threshold for the iteration in the self-consistent field (SCF) was set to be 10−8 eV and that in atomic force was set to be 0.01 eV Å−1. A spin-polarized DFT calculation was used in a ferromagnetic state. The initial magnetic moment for Fe, Co, and Ni is 3 μB, 2 μB, and 1.5 μB, respectively. The initial magnetic moment for other metals is 1 μB. The binding energy E_b is defined as

\[ E_b = E_{\text{carbonDAC}} - E_{\text{carbon}} - E_{\text{DAC}} \] (2)

where \( E_{\text{carbonDAC}} \), \( E_{\text{carbon}} \), and \( E_{\text{DAC}} \) are the total energies of the carbon surface supported DAC system, the carbon surface with no DAC, and the DAC in a vacuum, respectively. As defined, a more negative value of \( E_b \) signifies the stronger binding of a DAC to the carbon surface. The OERs occur via the following steps:

\[ \text{M} + 2\text{H}_2\text{O}(l) \rightarrow \text{MO} + \text{H}_2\text{O}_2(l) + \text{H}^+ + e^- \] (3)

\[ \text{MOH} + \text{H}_2\text{O}(l) + \text{H}^+ + e^- \rightarrow \text{MO} + \text{H}_2\text{O}_2(l) + 2\text{H}^+ + 2e^- \] (4)

\[ \text{MO} + \text{H}_2\text{O}(l) + 2\text{H}^+ + 2e^- \rightarrow \text{MOOH} + 3\text{H}^+ + 3e^- \] (5)

\[ \text{MOOH} + 3\text{H}^+ + 3e^- \rightarrow \text{M} + \text{O}_2(g) + 4\text{H}^+ + 4e^- \] (6)

The ORRs occur via the following steps:

\[ \text{M} + 2\text{O}_2(g) + 4\text{H}^+ + 4e^- \rightarrow \text{MOOH} + 3\text{H}^+ + 3e^- \] (7)

\[ \text{MOOH} + 3\text{H}^+ + 3e^- \rightarrow \text{MO} + \text{H}_2\text{O}(l) + 2\text{H}^+ + 2e^- \] (8)

\[ \text{MO} + \text{H}_2\text{O}(l) + 2\text{H}^+ + 2e^- \rightarrow \text{MOOH} + \text{H}_2\text{O}_2(l) + \text{H}^+ + e^- \] (9)

\[ \text{MOH} + \text{H}_2\text{O}(l) + \text{H}^+ + e^- \rightarrow \text{M} + 2\text{H}_2\text{O}(l) \] (10)

where \( M \) represents the preferable adsorption site for intermediates. For each step, the reaction free energy \( \Delta G \) is defined as the difference between the free energies of the initial and final states as calculated by the expression[66]

\[ \Delta G = \Delta E + \Delta ZPE - T\Delta S + \Delta G_U + \Delta G_{\text{pH}} \] (11)

where \( \Delta E \) is the total energy difference between reactants and products of reactions, \( \Delta ZPE \) is the zero-point energy correction, \( \Delta S \) is the vibrational entropy change at finite temperature \( T \), \( \Delta G_U = -eU \), where \( e \) is the elementary charge, \( U \) is the electrode potential, and \( \Delta G_{\text{pH}} \) is the correction of the \( \text{H}^+ \) free energy. \( \Delta G_{\text{pH}} = -k_B T \ln[H^+] = k_B T \ln[H^+] \).
= 0.0592 pH, where $k_b$ is the Boltzmann constant, $T$ is the temperature and $pH = 0$.

The overpotential $\eta_{\text{OER}}$ can be evaluated from the Gibbs free energy differences of each step as:

$$\eta_{\text{OER}} = \max \{ \Delta G_1, \Delta G_2, \Delta G_3, \Delta G_4 \} - 1.23,$$

where $\Delta G_1$, $\Delta G_2$, $\Delta G_3$, and $\Delta G_4$ are the free energies of the reaction steps shown in (2) to (5).

The overpotential $\eta_{\text{ORR}}$ can be evaluated from the Gibbs free energy differences of each step as:

$$\eta_{\text{ORR}} = \max \{ \Delta G_5, \Delta G_6, \Delta G_7, \Delta G_8 \} + 1.23,$$

where $\Delta G_5$, $\Delta G_6$, $\Delta G_7$, and $\Delta G_8$ are the free energies of reaction steps shown in (6) to (9).

**ML Algorithm:** The ML approach was designed to establish a relationship of nonlinear regression between the atomic parameters and binding energy, OER, and ORR catalytic activities, based on the results of DFT calculations. The trained ML models were used to predict the binding energy, OER, and ORR catalytic activities of a DAC on a defect site in a carbon surface. The ML model consists of three types of layers: graph convolutional layers, convolutional layers, and global max-pooling layers. The three graph convolutional layers of the ML model were used to extract multiple-level features between the transition metal atoms and substrate atoms. The output of the last graph convolutional layer serves as the input for three convolutional layers and a global max-pooling layer, which make a regression to predict the target properties (i.e., binding energy, onset potentials of OER and ORR) of the DACs. The detailed architecture was plotted in Figure S3 (Supporting Information). The parameters in the ML model were initialized randomly with a normal distribution. Furthermore, the rectified linear unit function (ReLU), $\text{ReLU}(x) = \max(0, x)$, where $x$ denotes the outputs of each hidden layer, was used as the activation function for each hidden layer. The initial ML models were then trained using the mean square error (MSE) loss function.

Each training dataset was divided into two subsets, with 90% for training the model and the remaining 10% for testing the performance of the trained model. The partitioning method of the training dataset into two subsets was completely random. To avoid overfitting, the validation dataset (10% of the training dataset) was used internally for the early stopping of the back-propagation trainer.

The ML model was implemented in PyTorch, and then fed the dataset (10% of the training dataset) was used internally for the early stopping of the back-propagation trainer.

The training dataset was divided into two subsets, with 90% for training the model and the remaining 10% for testing the performance of the trained model. The partitioning method of the training dataset into two subsets was completely random. To avoid overfitting, the validation dataset (10% of the training dataset) was used internally for the early stopping of the back-propagation trainer.

The ML model was implemented in PyTorch, and then fed the training and testing datasets into the ML model to optimize the model’s parameters. The optimization process was performed using the Adam algorithm with a fixed learning rate of 0.001 for 180 epochs. In each training process, mean squared errors were used as the loss function. The R-squared, $R^2$, can be defined as follow:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - f_i)^2}$$

where $y_i$ and $\bar{y}$ denote the $i$th DFT calculation result and the mean value of $n$ DFT results, respectively; $n$ is the size of the testing dataset; $f_i$ is the $i$th ML prediction result. To demonstrate the selected ML model was not overfit, the MSE loss during the training process was recorded, as shown in Figure S4 (Supporting Information).

**Supporting Information**

Supporting Information is available from the Wiley Online Library or from the author.

**Acknowledgements**

The authors acknowledge the University of Maryland supercomputing resources (http://hpcc.umd.edu) and Maryland Advanced Research Computing Center (MARCC) made available for conducting the research reported in this paper. The authors appreciate the support of Oracle for Research and the UMD New Direction Award for this research.

**Conflict of Interest**

The authors declare no conflict of interest.

**Author Contributions**

L.W. and T.G. contributed equally to this work. T.L. supervised the research. T.L., L.W., and T.G. conceived the research idea. L.W. carried out the DFT calculations. T.G. performed ML predictions. All authors contributed to the result through analysis and paper writing.

**Data Availability Statement**

The data that support the findings of this study are available in the supplementary material of this article.

**Keywords**

double-atom catalysts, machine learning, oxygen evolution reaction, oxygen reduction reaction, single-atom catalysts

Received: April 4, 2022
Published online: May 18, 2022

[1] Y. Chen, S. ji, C. Chen, P. Peng, D. Wang, Y. Li, Joule 2018, 2, 1242.
[2] W. Ma, H. Wan, L. Zhang, J. Y. Zheng, Z. Zhou, J. Energy Chem. 170, 2021.
[3] S. Ding, M. J. Hülsel, J. Pérez-Ramírez, Y. Yan, Joule 2019, 3, 2897.
[4] Z.-H. Xue, D. Luan, H. Zhang, X. W. (David) Lou, Joule 2022, 92.
[5] X. Xie, Lu Shang, X. Xiong, R. Shi, T. Zhang, Adv. Energy Mater. 2022, 12, 2102688.
[6] Y. Pan, C. Zhang, Z. Liu, C. Chen, Y. Li, Matter 2020, 2, 78.
[7] J. Liu, D. Cao, H. Xu, D. Cheng, Nano Select 2021, 2, 251.
[8] Z. W. Chen, L. X. Chen, C. C. Yang, Q. Jiang, J. Mater. Chem. A 2019, 7, 3492.
[9] P. Song, M. Luo, X. Liu, W. Xing, W. Xu, Z. Jiang, L. Gu, Adv. Funct. Mater. 2017, 27, 1700802.
[10] X. Zhou, Q. Shen, K. Yuan, W. Yang, Q. Chen, Z. Geng, J. Zhang, X. Shao, W. Chen, G. Xu, X. Yang, K. Wu, J. Am. Chem. Soc. 2018, 140, 554.
[11] Y. Ying, X. Luo, J. Qiao, H. Huang, Adv. Funct. Mater. 2021, 31, 2007423.
[12] Y.-S. Wei, L. Sun, M. Wang, J. Hong, L. Zou, H. Liu, Yu Wang, M. Zhang, Z. Liu, Y. Li, S. Horike, K. Suengna, Q. Xu, Angew. Chem. 2020, 132, 16147.
[13] L. Zhang, J. M. T. A. Fischer, Yi Jia, X. Yan, W. Xu, X. Wang, J. Chen, D. Yang, H. Liu, L. Zhuang, M. Hankel, D. J. Searles, K. Huang, S. Feng, C. L. Brown, X. Yao, J. Am. Chem. Soc. 2018, 140, 10757.
[14] X. Guo, J. Gu, S. Lin, S. Zhang, Z. Chen, S. Huang, J. Am. Chem. Soc. 2020, 142, 5709.
[15] M. Erdem Güney, R. Yıldırım, Catal. Rev. 2021, 63, 120.
[16] J. Xu, X.-M. Cao, P. Hu, Phys. Chem. Chem. Phys. 2021, 23, 11155.
[17] S. Lin, H. Xu, Y. Wang, X. C. Zeng, Z. Chen, J. Mater. Chem. A 2020, 8, 5663.
