# Journal of Materials Chemistry A

## ARTICLE

# Unravelling the Origin of Bifunctional OER/ORR Activity for Single-Atom Catalysts Supported on $C_2N$ by DFT and Machine Learning

Yiran Ying [a], Ke Fan [a], Xin Luo [b]*, Jinli Qiao [c], and Haitao Huang [a]*

Designing high-performance bifunctional oxygen evolution/reduction reaction (OER/ORR) catalysts is a newly emerged topic with wide applications in metal-air batteries and fuel cells. Herein, we report a group of (27) single-atom catalysts (SACs) supported on $C_2N$ monolayer as promising bifunctional OER/ORR catalysts by theoretical calculations. In particular, Rh@$C_2N$ exhibits lower OER overpotential (0.37 V) than $IrO_2$(110) benchmark with good ORR activity, while Au and Pd@$C_2N$ are superior ORR catalysts (with the overpotential of 0.38 and 0.40 V) than Pt(111) and their OER performance is also outstanding. More importantly, we discover the origin of the bifunctional catalytic activity by density functional theory (DFT) calculations and machine learning (ML). By DFT, we find a volcano-shaped relationship between the catalytic activity and $\Delta G_O$, and finally link them to the normalized Fermi abundance, a parameter based on electronic structure analysis. We further unravel the origin of element-specific activity by ML modelling based on random forest algorithm that takes outer electron number and oxide formation enthalpy as the two most important factors, and our model can give an accurate prediction of $\Delta G_O$ with much reduced time cost. This work not only paves the way for understanding the origin of bifunctional OER/ORR activity of SACs, but also benefits the rational design of novel SACs for other catalytic reactions by combining DFT and ML.

## Introduction

The main text of the article should appear here with headings as appropriate. Energy and environment are two crucial issues accompanying the accelerated human industrialization process, and developing clean and sustainable energy sources are urgently needed to solve the energy crisis and environmental pollution for human beings.[1, 2] Electrocatalysis plays a significant role in the energy conversion and storage process, and oxygen evolution/reduction reaction (OER/ORR) are among the most important electrocatalytic reactions related to energy and environment.[3-6] One of the bottlenecks which hinder the practical applications for electrocatalytic OER/ORR is the lack of highly efficient and stable catalysts to surpass the high reaction energy barrier, while the applications of conventional noble-metal-based catalysts ($IrO_2$ and $RuO_2$ for OER, Pt for ORR) are limited due to their high cost. Single-atom catalysts (SACs), defined as isolated transition metal (TM) atomic sites dispersed

and anchored on substrates, provide an alternative with high atom utilization efficiency, high selectivity, and stability, and the rational design of SACs for OER/ORR has become a focal point for researchers.[5, 7-9]

While monofunctional OER or ORR catalysts have been intensively studied in recent years, the development of bifunctional OER/ORR catalysts is relatively more challenging, despite the growing demand for them in rechargeable metal-air batteries and fuel cells.[10-13] Recently, several bifunctional OER/ORR SACs are reported, such as Mn-$N_2C_2$ catalysts with 0.915 V half-wave potential for ORR and 350 mV (at 10 mA cm$^{-2}$) overpotential for OER;[14] and $RuN_4C_x$ SACs with 0.372 V overpotential for OER and 0.826 V half-wave potential for ORR.[15] Theoretical calculations based on density functional theory (DFT) also predict several SACs that are active for both OER and ORR, including Rh/Co@$C_3N$,[16] Co@$N_4$,[17] Pt@$Nb_2CF_2$,[18] and Co/Rh/Ir-$N_4$-black phosphorus carbide[19]. However, these works need extensive trial-and-error experiments or high-throughput DFT calculations to screen out potential candidates, which is not cost-effective for the catalyst design. Furthermore, the intrinsic origin of the bifunctional OER/ORR activity for these SACs has not been fully elucidated yet, so that the design of such SACs is not systematic enough, and more fundamental physical and chemical studies are needed to gain a deeper understanding.

Besides experiments and DFT calculations, machine learning (ML) has recently emerged as a powerful, cross-disciplinary tool

a. Department of Applied Physics and Research Institute for Smart Energy, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, P.R. China. E-mail: aphhuang@polyu.edu.hk
b. State Key Laboratory of Optoelectronic Materials and Technologies, Centre for Physical Mechanics and Biophysics, School of Physics, Sun Yat-sen University, Guangzhou, Guangdong Province, P.R. China.E-Mail: luox77@mail.sysu.edu.cn
c. State Key Laboratory for Modification of Chemical Fibers and Polymer Materials, College of Environmental Science and Engineering, Donghua University, 2999 Renmin North Road, Shanghai 201620, P. R. China.
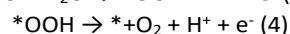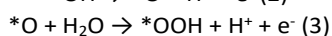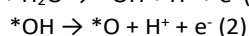
for materials design. Particularly for catalysis, by combining experiments and DFT calculations with ML, researchers can accelerate the catalyst design by analyzing the intrinsic relationships between activity descriptors and performance, as well as the importance of each descriptor, and even predict the catalytic performance of other systems based on the trained model without prior knowledge.[20-22] For instance, Tran and Ulissi identified 131 candidate alloy surfaces for $CO_2$ reduction and 258 surfaces for $H_2$ evolution by ML-guided DFT calculations.[22] They also predicted that Co-Al alloys can effectively reduce $CO_2$ to ethylene with high activity and selectivity, which is further proved by experiments.[23]

In this work, we chose 3$d$, 4$d$, and 5$d$ TM atoms supported on $C_2N$ monolayer (TM@$C_2N$) as a representative group of SACs to study their potential OER/ORR bifunctionality. $C_2N$ monolayers, which have been successfully synthesized in experiments in 2015,[24] have large and well-ordered vacancies terminated by $sp^2$-bonded nitrogen atoms, and thus is promising for anchoring TM single atoms. Besides, the unique electronic properties, outstanding chemical and thermal stability, and facile preparation routes with multiple precursors of $C_2N$ monolayers make them promising candidates in for electrocatalysis and energy storage in recent years.[25] In this regard, a comprehensive searching of potential SACs on $C_2N$ monolayers is in great demand. By combining DFT calculations, $ab$ $initio$ molecular dynamics (AIMD) simulations, and ML modelling, we systematically studied the origin of bifunctional OER/ORR activity of TM@$C_2N$.
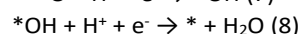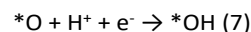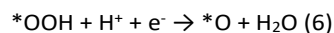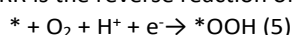
## Computational Methods

All spin-polarized DFT and AIMD calculations were performed by the projector-augmented wave method implemented in the Vienna $ab$ $initio$ Simulation Package (version 5.4.4).[26, 27] The Perdew-Burke-Ernzerhof (PBE) flavor of the generalized gradient approximation was used to describe the exchange-correlation functional,[28] and the DFT-D3 method was used to account for the van der Waals interactions.[29] The kinetic energy cutoff was set as 400 eV and gamma-centered 3×3×1 k-points mesh was used for the first Brillouin zone sampling. A vacuum layer with a thickness of more than 15 Å was used to avoid spurious interactions between adjacent image cells. All structures with and without adsorbates were fully optimized by the conjugate-gradient method, and the energy and force convergence criteria during geometric optimization were set as $10^{-5}$ eV and 0.02 eV/Å, respectively. AIMD simulations were performed by employing the Nosé-Hoover thermostat for NVT ensemble[30, 31] with the time step of 2 fs on 2×2×1 supercell of TM@$C_2N$. All geometric structures were visualized by the VESTA package.[32]

In the acidic medium, the OER was considered as a four-step process:

$$* + H_2O \rightarrow *OH + H^+ + e^- \quad (1)$$
$$*OH \rightarrow *O + H^+ + e^- \quad (2)$$
$$*O + H_2O \rightarrow *OOH + H^+ + e^- \quad (3)$$
$$*OOH \rightarrow * + O_2 + H^+ + e^- \quad (4)$$

And the 4$e^-$ ORR is the reverse reaction of OER:

$$* + O_2 + H^+ + e^- \rightarrow *OOH \quad (5)$$

$$*OOH + H^+ + e^- \rightarrow *O + H_2O \quad (6)$$
$$*O + H^+ + e^- \rightarrow *OH \quad (7)$$
$$*OH + H^+ + e^- \rightarrow * + H_2O \quad (8)$$

The OER and ORR overpotential $\eta_{OER}$ and $\eta_{ORR}$ were calculated by $\eta_{OER}$=max($\Delta G_1$, $\Delta G_2$, $\Delta G_3$, $\Delta G_4$)/e-1.23 V and $\eta_{ORR}$=max($\Delta G_5$, $\Delta G_6$, $\Delta G_7$, $\Delta G_8$)/e+1.23 V, respectively, where $\Delta G_i$ (i=1-8) represents the Gibbs free energy change for step (i). The detailed formula and data for the OER/ORR Gibbs free energy calculations in acidic medium were provided in Supplementary Note 1 and Table S1 in the Supporting Information.

## Results and Discussions

### Screening Potential Bifunctional OER/ORR SACs

A typical DFT-optimized structure of TM@$C_2N$ is shown in Fig. 1, where the TM atom forms an ionic bond with two $sp^2$-hybridized N atoms in the corner of the vacancy in $C_2N$, as suggested by the electron localization function (ELF) result (Fig. 1a), where blue regions with small ELF (<0.5) appear between TM and N. On the other hand, C-N and C-C bonds within the $C_2N$ monolayer exhibit covalent character with ELF>0.5.[33] The optimized lattice parameters for 3$d$, 4$d$, and 5$d$ TM@$C_2N$ are summarized in Table S2. The introduction of single TM atoms leads to a small deformation of the $C_2N$ cell, indicated by the slightly larger lattice parameter $b$ than $a$ (except for Co@$C_2N$) by 0.01% to 3.45% (Ir@$C_2N$). The deformation is relatively more evident for 5$d$ TM elements due to the larger atomic radius. We also list the magnetic moment for TM@$C_2N$ in Table S3 and 13 TM@$C_2N$ SACs possess the non-zero magnetic moment, and spin polarization is considered in all the DFT calculations.
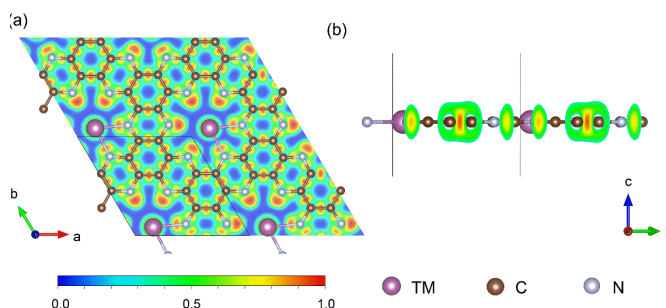


**Fig. 1** (a) Top and (b) side view of the DFT-optimized structure with the ELF contour plot (with scale bar) of TM@$C_2N$. TM, C, and N atoms are represented in purple, brown, and cyan, respectively.

Next, we move on to screen the potential bifunctional OER/ORR SACs among the TM@$C_2N$ structures. Except for Tc (without non-radioactive isotope and scarce in nature), Re (DFT calculations fail to converge and scarce in nature), and Hg (highly toxic), all the TM@$C_2N$ SACs are modelled for OER and ORR, with their overpotential values ($\eta_{OER}$ and $\eta_{ORR}$) calculated from the Gibbs free energies ($\Delta G$ values listed in Table S4) and presented in the heatmap (Fig. 2a and 2b). From the heatmap, we can conclude that for late TM elements, especially $d^7$-$d^9$ elements, TM@$C_2N$ SACs exhibit outstanding OER/ORR activity with small overpotential, while for early ($d^1$-$d^4$) TM@$C_2N$, the

activity is relatively poor. One reason for this is the dissociation of *OOH happens on some of the $d^2$-$d^4$ TM@C$_2$N structures (denoted in italic in Table S4, see Fig. S1 for examples), which will be discussed in detail in the following part. Among all the $d^7$-$d^9$ TM@C$_2$N SACs, Rh@C$_2$N and Pt@C$_2$N exhibit excellent OER overpotentials of 0.37 and 0.40 V, respectively, better than the IrO$_2$(110) benchmark (0.56 V),[34] and Rh@C$_2$N also possesses relatively good ORR activity ($\eta_{ORR}$ = 0.67 V). On the other hand, Au@C$_2$N and Pd@C$_2$N are the best ORR catalysts with overpotential values of 0.38 and 0.40 V, respectively, which are even better than Pt(111) (0.45 V),[35] and their $\eta_{OER}$ values are only slightly higher than that for IrO$_2$(110)

(0.79 and 0.71 V, respectively). We also use $\eta_{sum}=\eta_{OER}+\eta_{ORR}$ as a descriptor for OER/ORR bifunctional activity (Fig. S2), and Rh, Au, and Pd@C$_2$N with low $\eta_{sum}$ values of 1.05, 1.18, and 1.11 V are selected as three potential OER/ORR bifunctional SACs with outstanding activities.
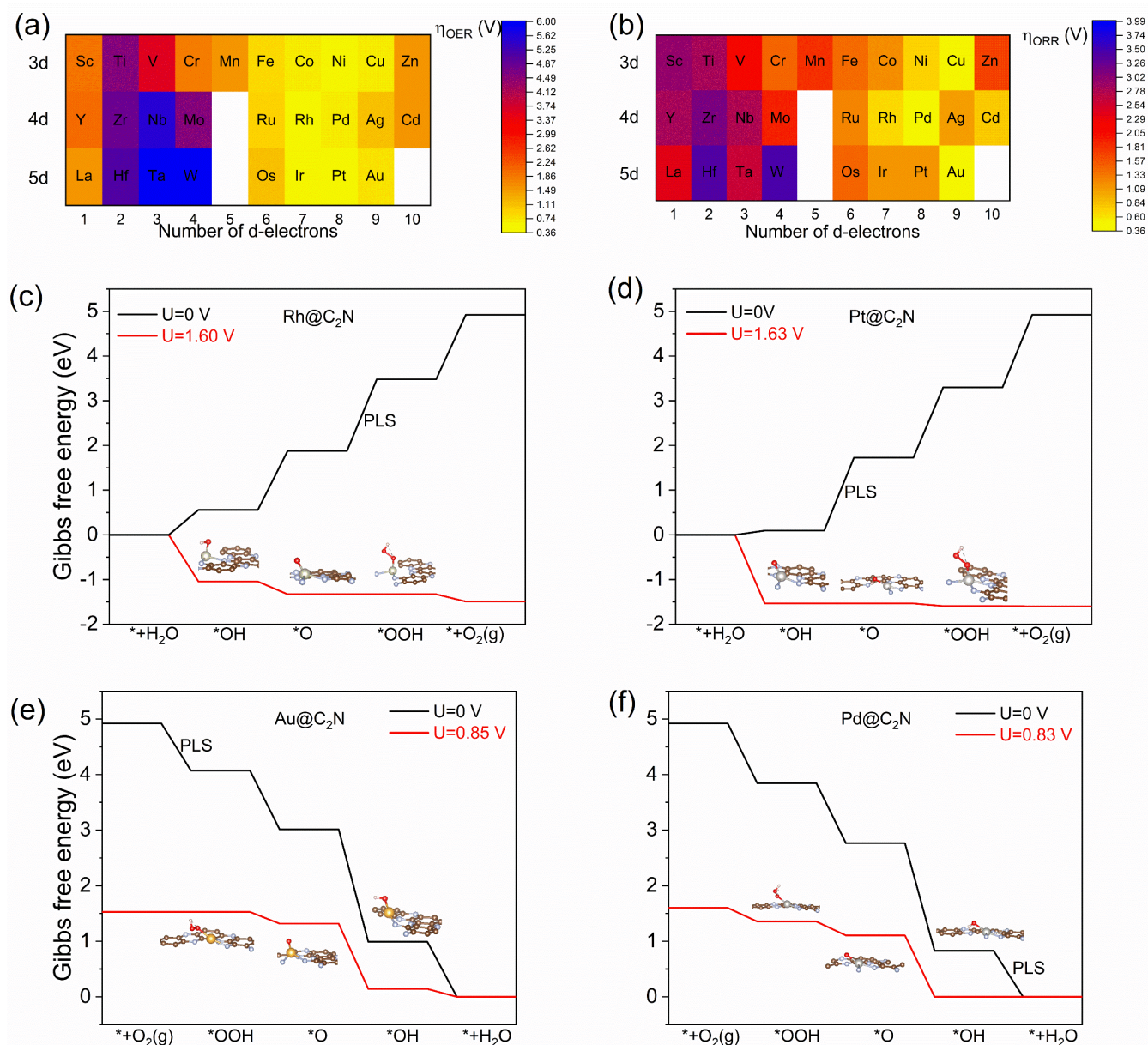


**Fig.2** Screening of potential bifunctional OER/ORR SACs. Heatmap of the overpotential η for (a) OER and (b) ORR on 3$d$-5$d$ TM@C$_2$N. Gibbs free energy diagram for OER on (c) Rh and (d) Pt@C$_2$N and ORR on (e) Au and (f) Pd@C$_2$N at zero and applied electrode potential U in acidic medium. The optimized structures of the intermediates are shown as insets of (c)-(f). The potential-limiting steps (PLS) are marked in the figures.

The 2e$^-$ ORR pathway toward $H_2O_2$ (* + $O_2$ + H$^+$ + e$^-$→ *OOH; *OOH + H$^+$ + e$^-$ → * + $H_2O_2$) is a side reaction for 4e$^-$ ORR and thus the selectivity issue needs to be taken into account.[36, 37] Considering the different products in the second step of ORR (* + $H_2O_2$ versus *O + $H_2O$), if $\Delta G_O$ < $\Delta G(H_2O_2)$-$\Delta G(H_2O)$ = 3.52 eV, the ORR is 4e$^-$ selective toward $H_2O$.[37] Herein, by comparing $\Delta G$(*O) with 3.52 eV (Table S4), we can conclude that all the SACs studied in this work are 4e$^-$ selective except for Ag@$C_2N$, which has the selectivity toward 2e$^-$ ORR with a limiting potential value of 0.15 V (Fig. S3). As a result, we will mainly discuss the 4e$^-$ ORR in this work.

We further plot the OER and ORR Gibbs free energy diagram for Rh, Pt, Au, and Pd@$C_2N$ in Fig. 2c-2f and Fig. S4. An ideal OER catalyst has 1.23 eV $\Delta G$ for all four elementary steps, while the Gibbs free energy for all steps in ORR decreases by 1.23 eV for an ideal ORR catalyst. For OER on Pt, Au, and Pd@$C_2N$, the potential-limiting step (PLS) with the largest Gibbs free energy increase lies on the second step (*OH→*O) with $\Delta G$ values of 1.63, 2.02, and 1.94 eV, respectively, while the third step (*O→*OOH) is the PLS with 1.60 eV $\Delta G$ on Rh@$C_2N$. With applied potential U larger than 1.60 V and 1.63 V for Rh and Pt@$C_2N$ respectively, all the steps in OER are exothermic (Fig. 2c,d), so the $\eta_{OER}$ can be calculated by subtracting 1.23 V. The PLS for Pd, Rh, and Pt@$C_2N$ is the fourth step (*OH→$H_2O$) for ORR with the smallest $\Delta G$ decrease, and for Au@$C_2N$ it is the first step (*+$O_2$→*OOH). At U smaller than 0.85 V for Au@$C_2N$ and 0.83 V for Pd@$C_2N$, Gibbs free energy decreases for each step (Fig. 2e,f) and these are their ORR limiting potential values.

AIMD simulations are performed to check the thermal stability of the SACs. As shown in Fig. S5, the total energies oscillate around the equilibrium values during the 10 ps AIMD simulations for Rh, Au, and Pd@$C_2N$ at 500 K, and after the simulations, all the structures do not exhibit obvious structural reconstruction, and they can restore the initial structures after structural relaxation, indicating that they are thermally stable. By calculating the formation energy ($E_f$) and dissolution potential ($U_{diss}$, vs. standard hydrogen electrode) of TM@$C_2N$ SACs, we also confirm the thermodynamic and electrochemical stabilities of Rh, Au, and Pd@$C_2N$ among 8 SACs (see Supplementary Note 2 and Table S5). To be more specific, the negative $E_f$ and positive $U_{diss}$ of Rh, Au, and Pd@$C_2N$ suggest that binding of single TM atoms on $C_2N$ is more favorable than their aggregating into clusters, and the dissolution of single TM atoms can be avoided under the electrochemical conditions.[37, 38]

### Scaling Relations and Electronic Structure Analysis

To this point, we have identified three bifunctional OER/ORR catalysts from a group of (27) 3$d$, 4$d$, and 5$d$ SACs supported on $C_2N$. However, the origin of catalytic activity for these SACs remains unclear yet, which may hinder the extension of such methods to other OER/ORR catalysts. Exploring reactivity descriptors is a possible approach to solve this issue and gain a deeper understanding of the fundamental physics and

chemistry in electrocatalysis.[6, 39] For ORR, descriptors based on $d$-band theory,[40] surface distortion,[41] and coordination number[42] have been proposed, but for OER/ORR bifunctional catalysts, such descriptors are seldom reported yet.

To unravel the origin of the outstanding bifunctional OER/ORR activity for the SACs obtained in Part 3.1, we first consider the relationship between $\Delta G_O$, $\Delta G_{OH}$, and $\Delta G_{OOH}$. As shown in Fig. 3a, $\Delta G_{OH}$ exhibit a linear scaling relation between $\Delta G_O$ ($\Delta G_{OH}$ = 0.467$\Delta G_O$–0.627) with an $R^2$ value of 0.858 for all the TM@$C_2N$ SACs. The case of the relationship between $\Delta G_{OOH}$ and $\Delta G_O$ is different: because *OOH dissociation occurs on several early TM@$C_2N$, they exhibit different linear scaling relation from those without *OOH dissociation, which is similar to the previous report of TM@g-$C_3N_4$ SACs.[43] After fitting the values from SACs with undissociated *OOH, we get $\Delta G_{OOH}$ = 0.467$\Delta G_O$–0.627 with an $R^2$ value of 0.913 (Fig. 3b). The linear scaling relations make it reasonable to deduce that the OER/ORR overpotential can be described by only one $\Delta G$, such as $\Delta G_O$. Our results in Fig. S6 indicate that for late TM@$C_2N$, a volcano-shaped relationship can be found between OER/ORR overpotential and $\Delta G_O$. For OER, the peak position is approximately $\Delta G_O$ = 1.9 eV, where Rh@$C_2N$ locates, and if $\Delta G_O$ increases or decreases, the OER performance will deteriorate. For ORR, on the other hand, the peak position is higher (around 3.0 eV). This volcano-shaped relationship is in accordance with the Sabatier principle in catalysis, that is, the optimized catalysts should bind the reaction intermediates neither too strong nor too weak to benefit both adsorption and desorption of molecules.[44]

We further use the as-obtained linear scaling relations to construct the contour plot of OER/ORR overpotentials as a function of $\Delta G_{OOH}$ - $\Delta G_O$ and $\Delta G_O$ - $\Delta G_{OH}$, which are closely related to the Gibbs free energy changes for elementary steps in OER and ORR. For OER, SACs with the optimized theoretical performance is located in the triangular red-colored region with 1.3 < $\Delta G_O$-$\Delta G_{OH}$ < 1.8 eV and 0.8 < $\Delta G_{OOH}$-$\Delta G_O$ < 1.8 eV (Fig. 3c), where Rh, Pt, and Ni@$C_2N$ locate. For ORR, on the other hand, the optimized region is 1.7 < $\Delta G_O$-$\Delta G_{OH}$ < 2.2 eV and 0.7 < $\Delta G_{OOH}$-$\Delta G_O$ < 1.6 eV (Fig. 3d), and Au and Pd@$C_2N$ are in the center of it. It is noteworthy that the fitted line (in black) by the linear scaling relations of TM@$C_2N$ pass through the center of the optimized regions for both OER and ORR with several data points in the regions, suggesting that TM@$C_2N$ SACs are a group of promising catalysts with potential bifunctional OER/ORR activity.

Next, we will investigate the electronic structure of TM@$C_2N$ and how they are related to the OER/ORR theoretical catalytic activity. The electronic structure of catalysts can form a bridge between atomic structure and catalytic activity for catalysts, and from the $d$-band theory by Nørskov et al.,[40] the $d$-states are connected to the binding strength of adsorbates and finally catalytic activity. Herein, we first calculate the partial density of states (PDOSs) of TM $d$ orbitals for TM@$C_2N$, with part of the results shown in Fig. 4a. At Fermi level $E_F$, $d$ orbital contributes to the PDOSs and finally to the metallicity of

TM@C$_2$N (except for Ag and Au, whose metallicity mainly comes from C and N, as shown in Fig. S7), which is beneficial for electrocatalysis.
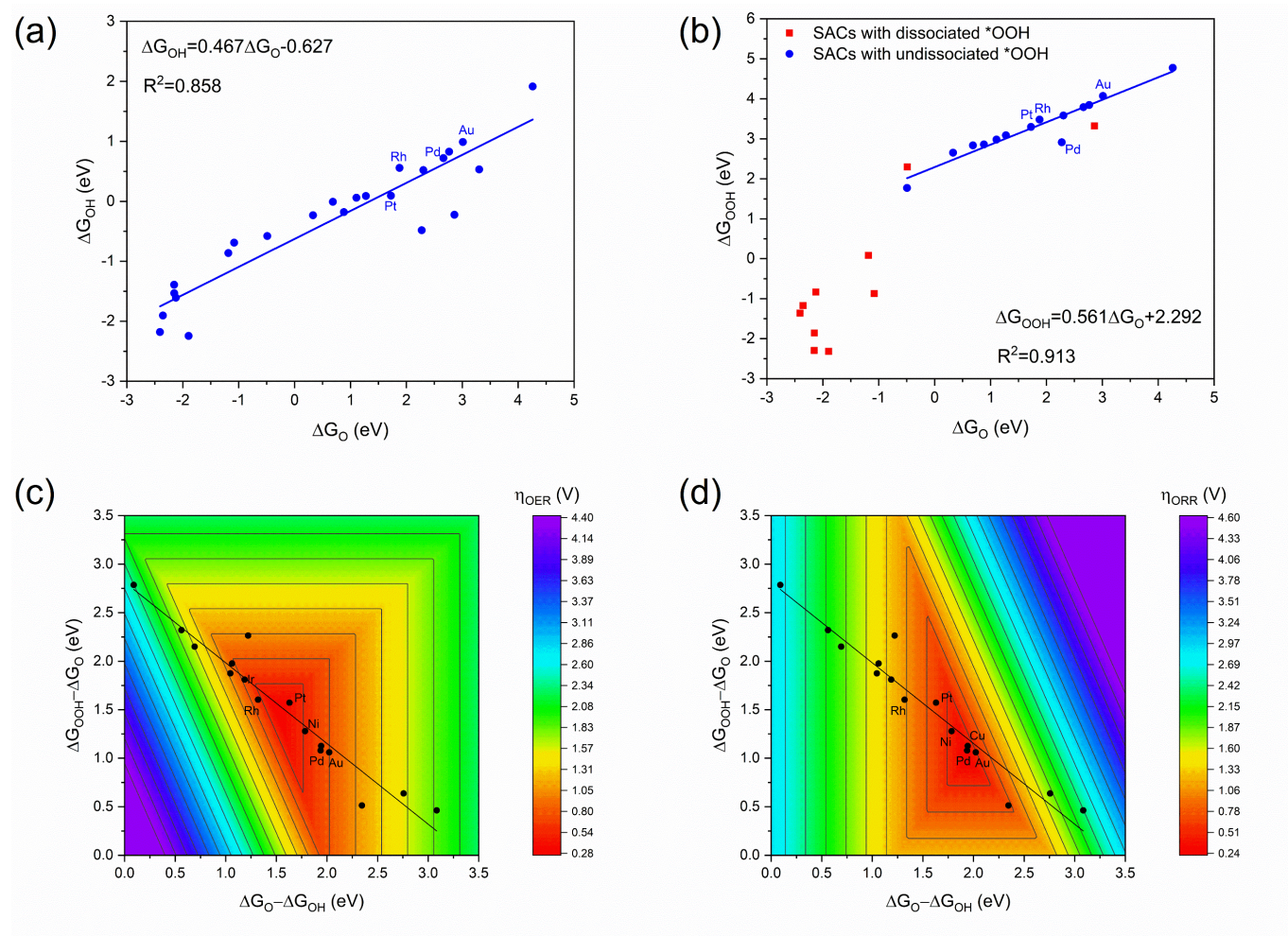


**Fig. 3** (a) $\Delta G_{OH}$ and (b) $\Delta G_{OOH}$ as a function of $\Delta G_O$ with the linear fitting results shown, and contour plot of (c) $\eta_{OER}$ and (d) $\eta_{ORR}$ as a function of $\Delta G_{OOH}$ - $\Delta G_O$ and $\Delta G_O$ - $\Delta G_{OH}$ constructed by the linear scaling relations on TM@C$_2$N SACs.
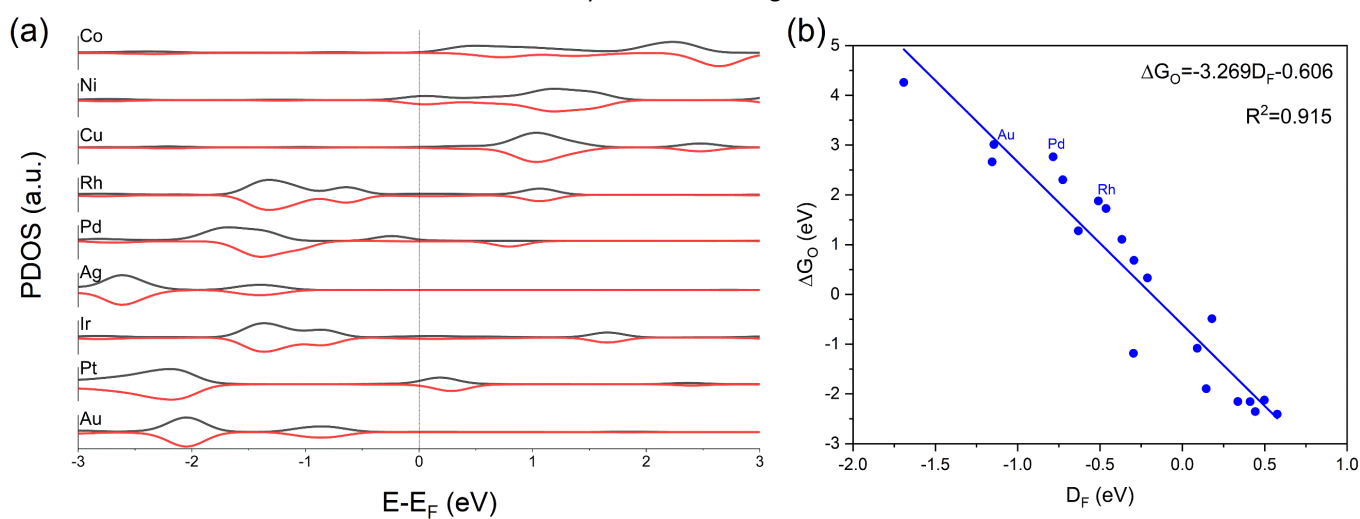
**Fig. 4** (a) Partial density of states (PDOSs) of $d$ orbitals for TM@$C_2N$. Contributions from spin-up and spin-down channels are denoted in black and red, respectively. Fermi level $E_F$ is set to zero (denoted in dotted line). (b) $\Delta G_O$ as a function of normalized Fermi abundance $D_F$ with $kT_{el}$=0.6 eV. The linear fitting results are also shown.

Since surface reactivity closely relates to TM $d$ orbitals, we first use $d$-band center $\varepsilon_d$ as an activity descriptor for OER/ORR. After plotting $\Delta G_O$ as a function of $\varepsilon_d$, we find a trend that $\Delta G_O$ decreases with increasing $\varepsilon_d$, but the correlation is weak with an $R^2$ value of only 0.647 (Fig. S8), possibly due to the fact that not all the states make an equal contribution to the surface bonding, and the states near $E_F$ make larger contributions. Thus, other descriptors considering the abundance of electronic states in the vicinity of $E_F$ should be explored. In a pioneer report, Huang *et al.* used a weight function to quantify the contribution of every state to the bonding characteristics, and the derivative of the Fermi-Dirac distribution function at non-zero temperature $f_T(E - E_F) = 1/\{\exp\left[\frac{E-E_F}{kT_{el}}\right] + 1\}$ is a well-behaved weight function for catalysis ($kT_{el}$ here is a parameter controlling the distribution of the weight function, and $T_{el}$ is defined as electronic temperature).[45] By multiplying the partial density of states of $d$-orbital $D(E)$ with the weight function, the contribution from the states near $E_F$ can be emphasized (Fig. S9), so that the correlation between surface reactivity and electronic structures of catalysts can be established.

Based on this, we define a parameter, normalized Fermi abundance $D_F$ as:

$$D_F = \frac{1}{S_F}\int_{-\infty}^{+\infty} ED(E)\frac{-\partial f_T(E-E_F)}{\partial E}dE = \frac{1}{S_F}\int_{-\infty}^{+\infty}\frac{ED(E)\exp\left[\frac{E-E_F}{kT_{el}}\right]}{kT_{el}\{\exp\left[\frac{E-E_F}{kT_{el}}\right]+1\}^2}dE$$

where

$$S_F = -\int_{-\infty}^{+\infty} D(E)\frac{\partial f_T(E-E_F)}{\partial E}dE = \int_{-\infty}^{+\infty}\frac{D(E)\exp\left[\frac{E-E_F}{kT_{el}}\right]}{kT_{el}\{\exp\left[\frac{E-E_F}{kT_{el}}\right]+1\}^2}dE$$

is the Fermi softness defined by Huang *et al*.[45] Compared with the definition of $\varepsilon_d$ (Fig. S8), $D_F$ introduces the tunable contribution term of electronic states near $E_F$, and by adjusting the value of $kT_{el}$, we can obtain weight function with different spreading. After testing different $kT_{el}$ values, we find out that there is a linear correlation between $\Delta G_O$ and $D_F$ when $kT_{el}$=0.6 eV, with an $R^2$ value of 0.915 (Fig. 4b). $\Delta G_O$ decreases with increasing $D_F$, and the optimum $\Delta G_O$ value for OER (1.9 eV) and ORR (3.0 eV) is achieved when $D_F$ reaches -0.77 and -1.10 eV, respectively. Consequently, $D_F$ can act as an activity descriptor for both OER and ORR.

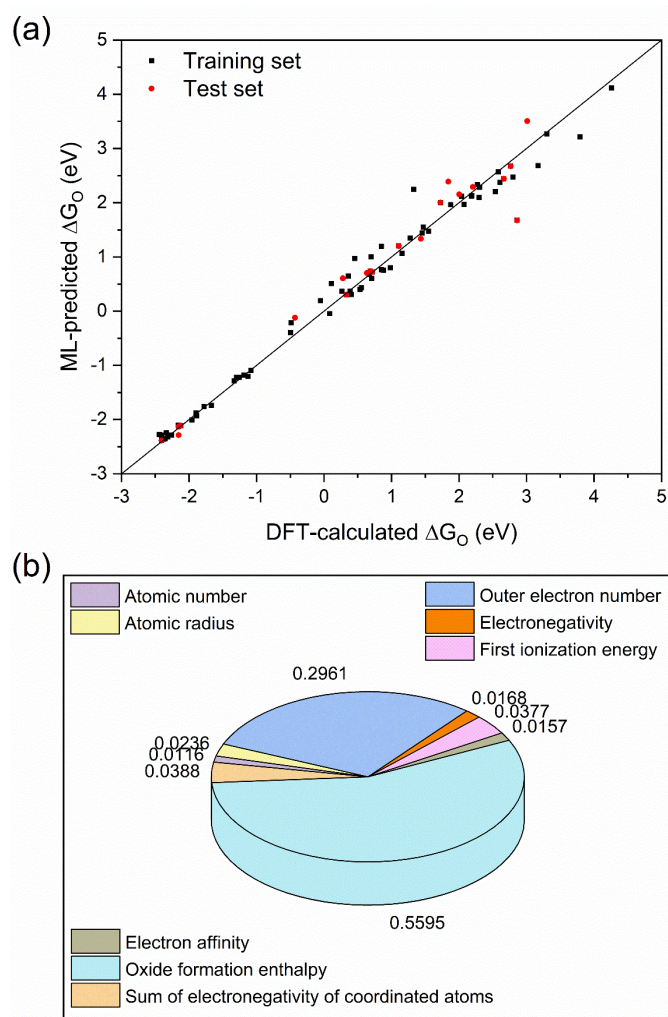**ML Analysis**

After investigating the relationship between electronic structure, adsorption energy, and catalytic activity of TM@$C_2N$, there are still two questions left: (1) how to correlate the bifunctional OER/ORR activity with more fundamental physical and chemical properties of the SACs, such as element-specific features; and (2) how to extend the approaches we build in this work to other catalyst systems for other reactions. ML can provide a solution to these through interdisciplinary investigations.

Feature engineering is one of the most important steps in ML. In our modeling, a feature set consisting of eight descriptors is selected to describe the structural and electronic properties of TM@$C_2N$, especially the intrinsic information of TM elements (see Supplementary Note 3 and Table S6). These features include TM atomic number ($Z$), atomic radius ($r$), outer electron number ($n_e$), electronegativity ($N$), first ionization energy (IE), electron affinity (EA), the sum of electronegativity of coordinated atoms of TM ($N_{sum}$), and single-atom oxide formation enthalpy ($H_{ox,f}$, defined as the difference between bulk metal cohesive energy and bulk metal oxide formation enthalpy[20, 37, 46]). To enlarge the chemical space, SACs with $N_1C_1$ and $N_1S_1$ coordinations (Fig. S10) instead of $N_2$ coordination are also considered, and the TM@$C_2N$ dataset is amplified by a factor of 2. Pearson correlation map (Fig. S11) shows that most of the chosen feature pairs for our dataset do not show a strong correlation with the Pearson coefficient smaller than 0.6, so it is reasonable to believe that a non-redundant set of features has been chosen.

We first randomly separate our dataset into the training set and test set with the ratio of 4:1. Based on grid search optimization results of parameters for the random forest algorithm, we set the maximum depth of the tree as 22 and the number of trees in the forest as 31 in our ML model. The trained random forest model performs well with the training score and test score ($R^2$ value) of 0.985 and 0.981, respectively, indicating that the modelling result is satisfactory (Table S7). The mean absolute error (MAE) and rooted mean square error (RMSE) values are also listed in Table S7. Results comparing the DFT-calculated and ML-predicted $\Delta G_O$ are shown in Fig. 5a, and the linearity between them is outstanding. This suggests that our model can predict $\Delta G_O$ values even without prior DFT calculations, providing a time-saving solution for catalyst design, considering the much reduced computational time from several days (DFT calculations) to several seconds (ML prediction).

**Fig. 5** (a) Comparison between DFT-calculated and ML-predicted $\Delta G_O$. Black and red points correspond to data in the training set and test set, respectively. (b) Pie chart for feature importance analysis in the ML model.

Furthermore, we analyze the feature importance based on our ML model. From Fig. 5b, the most significant feature is the oxide formation enthalpy, with the feature importance of 0.5595, followed by the outer electron number (0.2961). The other seven features have much smaller contributions to $\Delta G_O$ with feature importance less than 0.05. This conclusion is in accordance with our results in 'Screening potential bifunctional OER/ORR SACs' part that the number of $d$ electrons is an important factor in determining the activity origin, where late TM@C$_2$N SACs with $d^7$-$d^9$ configuration have higher bifunctional OER/ORR activity, and such periodic trend of reactivity was also observed in catalysts for methane activation.[47] Besides, the oxide formation energy has been reported as an outstanding descriptor for the oxygen adsorption behavior of materials.[20, 37, 46]

In addition to $\Delta G_O$, we further use $\eta_{sum}$, an indicator that is directly related to the overall OER/ORR activity, as the training and testing data (Table S8). The new ML model also exhibits high training and test scores of 0.946 and 0.914, respectively, and feature importance analysis results (Fig. S12) also highlight outer electron number (0.2686) and oxide formation enthalpy (0.2314) as the two most significant features. Our ML model and the methodology in this work can be extended to the design of SACs for other catalytic reactions and double-atom catalysts (DACs), a newly emerged topic in heterogeneous catalysis (Supplementary Note 4).[48-50]

## Conclusions

The conclusions section should come in this section at the end of the article, before the acknowledgements. In the present work, by means of DFT calculations, AIMD simulations, and ML, we systematically study the activity origin of a group of (27) bifunctional OER/ORR SACs—TM@C$_2$N. After identifying three SACs, namely Rh, Au, and Pd@C$_2$N with outstanding bifunctional activity, we further unravel the physical and chemical origin hidden behind them. DFT calculations exhibit a volcano-shaped relationship between the OER/ORR theoretical overpotential and $\Delta G_O$, and based on electronic structure analysis, we design a parameter, normalized Fermi abundance $D_F$, as a descriptor for OER/ORR activity because of its linear relationship between $\Delta G_O$. Finally, by ML with random forest algorithm, $\Delta G_O$ values can be predicted with much reduced time cost compared with DFT, and we attribute the OER/ORR activity of TM@C$_2$N to two element-specific factors: outer electron number and oxide formation enthalpy, which have the largest influence on $\Delta G_O$. This work opens up a new avenue for investigating the origin of OER/ORR bifunctional activity and performance for other catalytic reactions of SACs, and therefore triggers the rational design of high-performance catalysts.

## Author contributions

**Yiran Ying:** Conceptualization, Data curation, Investigation, Software, Methodology, Writing – original draft; **Ke Fan:** Data curation; **Xin Luo:** Supervision, Writing – review & editing; **Jinli Qiao:** Writing – review & editing; **Haitao Huang:** Funding acquisition, Supervision, Writing – review & editing.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

## References

1. A. Aricò, P. Bruce, B. Scrosati, J. Tarascon and W. Van Schalkwijk, *Nat. Mater.* , 2005, **4**, 366-377.
2. S. Chu and A. Majumdar, *Nature*, 2012, **488**, 294-303.
3. Y. Jiao, Y. Zheng, M. Jaroniec and S. Z. Qiao, *Chem. Soc. Rev.* , 2015, **44**, 2060-2086.
4. L. Dai, Y. Xue, L. Qu, H. J. Choi and J. B. Baek, *Chem. Rev.* , 2015, **115**, 4823-4892.
5. D. Zhao, Z. Zhuang, X. Cao, C. Zhang, Q. Peng, C. Chen and Y. Li, *Chem. Soc. Rev.* , 2020, **49**, 2215-2264.
6. J. Liu, H. Liu, H. Chen, X. Du, B. Zhang, Z. Hong, S. Sun and W. Wang, *Adv Sci (Weinh)*, 2020, **7**, 1901614.
7. X.-F. Yang, A. Wang, B. Qiao, J. Li, J. Liu and T. Zhang, *Acc. Chem. Res.* , 2013, **46**, 1740-1748.
8. J. Liu, *ACS Catal.*, 2016, **7**, 34-59.
9. Y. Pan, C. Zhang, Z. Liu, C. Chen and Y. Li, *Matter*, 2020, **2**, 78-110.
10. J. Zhang, Z. Zhao, Z. Xia and L. Dai, *Nat. Nanotechnol.* , 2015, **10**, 444-452.
11. G.-L. Chai, K. Qiu, M. Qiao, M.-M. Titirici, C. Shang and Z. Guo, *Energy Environ. Sci.* , 2017, **10**, 1186-1195.
12. Q. Wang, Y. Ji, Y. Lei, Y. Wang, Y. Wang, Y. Li and S. Wang, *ACS Energy Lett.*, 2018, **3**, 1183-1191.
13. W. Cheng, X. Zhao, H. Su, F. Tang, W. Che, H. Zhang and Q. Liu, *Nat. Energy*, 2019, **4**, 115-122.
14. H. Shang, W. Sun, R. Sui, J. Pei, L. Zheng, J. Dong, Z. Jiang, D. Zhou, Z. Zhuang, W. Chen, J. Zhang, D. Wang and Y. Li, *Nano Lett.* , 2020, **20**, 5443-5450.
15. L. Bai, Z. Duan, X. Wen, R. Si, Q. Zhang and J. Guan, *ACS Catal.*, 2019, **9**, 9897-9904.
16. Y. Zhou, G. Gao, J. Kang, W. Chu and L.-W. Wang, *J. Mater. Chem. A*, 2019, **7**, 12050-12059.
17. M. Hu, S. Li, S. Zheng, X. Liang, J. Zheng and F. Pan, *J. Phys. Chem. C*, 2020, **124**, 13168-13176.
18. D. Kan, R. Lian, D. Wang, X. Zhang, J. Xu, X. Gao, Y. Yu, G. Chen and Y. Wei, *J. Mater. Chem. A*, 2020, **8**, 17065-17077.
19. D. Chen, Z. Chen, Z. Lu, X. Zhang, J. Tang and C. V. Singh, *Nanoscale*, 2020, **12**, 18721-18732.
20. N. J. O'Connor, A. S. M. Jonayat, M. J. Janik and T. P. Senftle, *Nat. Catal.*, 2018, **1**, 531-539.
21. A. Chen, X. Zhang and Z. Zhou, *InfoMat*, 2020, **2**, 553-576.
22. K. Tran and Z. W. Ulissi, *Nat. Catal.*, 2018, **1**, 696-703.
23. M. Zhong, K. Tran, Y. Min, C. Wang, Z. Wang, C. T. Dinh, P. De Luna, Z. Yu, A. S. Rasouli, P. Brodersen, S. Sun, O. Voznyy, C. S. Tan, M. Askerka, F. Che, M. Liu, A. Seifitokaldani, Y. Pang, S. C. Lo, A. Ip, Z. Ullissi and E. H. Sargent, *Nature*, 2020, **581**, 178-183.
24. J. Mahmood, E. K. Lee, M. Jung, D. Shin, I.-Y. Jeon, S.-M. Jung, H.-J. Choi, J.-M. Seo, S.-Y. Bae and S.-D. Sohn, *Nat. Commun.*, 2015, **6**, 6486.
25. Z. Tian, N. López-Salas, C. Liu, T. Liu and M. Antonietti, *Adv. Sci.*, 2020, **7**, 2001767.
26. G. Kresse and J. Furthmüller, *Phys. Rev. B*, 1996, **54**, 11169.
27. G. Kresse and D. Joubert, *Phys. Rev. B*, 1999, **59**, 1758.
28. J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.* , 1996, **77**, 3865.
29. S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
30. S. Nosé, *J. Chem. Phys.*, 1984, **81**, 511-519.
31. G. J. Martyna, M. L. Klein and M. Tuckerman, *J. Chem. Phys.*, 1992, **97**, 2635-2643.
32. K. Momma and F. Izumi, *J. Appl. Crystallogr.* , 2011, **44**, 1272-1276.
33. T. Yu, Z. Zhao, L. Liu, S. Zhang, H. Xu and G. Yang, *J. Am. Chem. Soc.*, 2018, **140**, 5962-5968.
34. J. Rossmeisl, Z. W. Qu, H. Zhu, G. J. Kroes and J. K. Nørskov, *J. Electroanal. Chem.* , 2007, **607**, 83-89.
35. J. K. Nørskov, J. Rossmeisl, A. Logadottir, L. Lindqvist, J. R. Kitchin, T. Bligaard and H. Jonsson, *J. Phys. Chem. B*, 2004, **108**, 17886-17892.
36. A. Kulkarni, S. Siahrostami, A. Patel and J. K. Norskov, *Chem. Rev.* , 2018, **118**, 2302-2312.
37. X. Guo, S. Lin, J. Gu, S. Zhang, Z. Chen and S. Huang, *ACS Catal.*, 2019, **9**, 11042-11054.
38. <span style="color:red">J. Greeley and J. K. Nørskov, *Electrochim. Acta* 2007, **52**, 5829-5836.</span>
39. Z.-J. Zhao, S. Liu, S. Zha, D. Cheng, F. Studt, G. Henkelman and J. Gong, *Nat. Rev. Mater.*, 2019, **4**, 792-804.
40. J. K. Nørskov, T. Bligaard, J. Rossmeisl and C. H. Christensen, *Nat. Chem.*, 2009, **1**, 37-46.
41. R. Chattot, O. Le Bacq, V. Beermann, S. Kühl, J. Herranz, S. Henning, L. Kühn, T. Asset, L. Guétaz and G. Renou, *Nat. Mater.* , 2018, **17**, 827-833.
42. F. Calle-Vallejo, J. Tymoczko, V. Colic, Q. H. Vu, M. D. Pohl, K. Morgenstern, D. Loffreda, P. Sautet, W. Schuhmann and A. S. Bandarenka, *Science*, 2015, **350**, 185-189.
43. H. Niu, X. Wang, C. Shao, Y. Liu, Z. Zhang and Y. Guo, *J. Mater. Chem. A*, 2020, **8**, 6555-6563.
44. A. J. Medford, A. Vojvodic, J. S. Hummelshøj, J. Voss, F. Abild-Pedersen, F. Studt, T. Bligaard, A. Nilsson and J. K. Nørskov, *J. Catal.* , 2015, **328**, 36-42.
45. B. Huang, L. Xiao, J. Lu and L. Zhuang, *Angew. Chem. Int. Ed.* , 2016, **55**, 6239-6243.
46. S. Lin, H. Xu, Y. Wang, X. C. Zeng and Z. Chen, *J. Mater. Chem. A*, 2020, **8**, 5663-5670.
47. A. S. Rosen, J. M. Notestein and R. Q. Snurr, *ACS Catal.*, 2019, **9**, 3576-3587.
48. Y. Ying, X. Luo, J. Qiao and H. Huang, *Adv. Funct. Mater.*, 2021, **31**, 2007423.
49. <span style="color:red">X. Zhu, J. Yan, M. Gu, T. Liu, Y. Dai, Y. Gu and Y. Li, *J. Phys. Chem. Lett.* , 2019, **10**, 7760-7766.</span>
50. <span style="color:red">C. Deng, Y. Su, F. Li, W. Shen, Z. Chen and Q. Tang, *J. Mater. Chem. A*, 2020, **8**, 24563-24571.</span>