

# **Design Efficient Photo-Electro-Catalyst by First-Principle Calculations and Deep Learning**

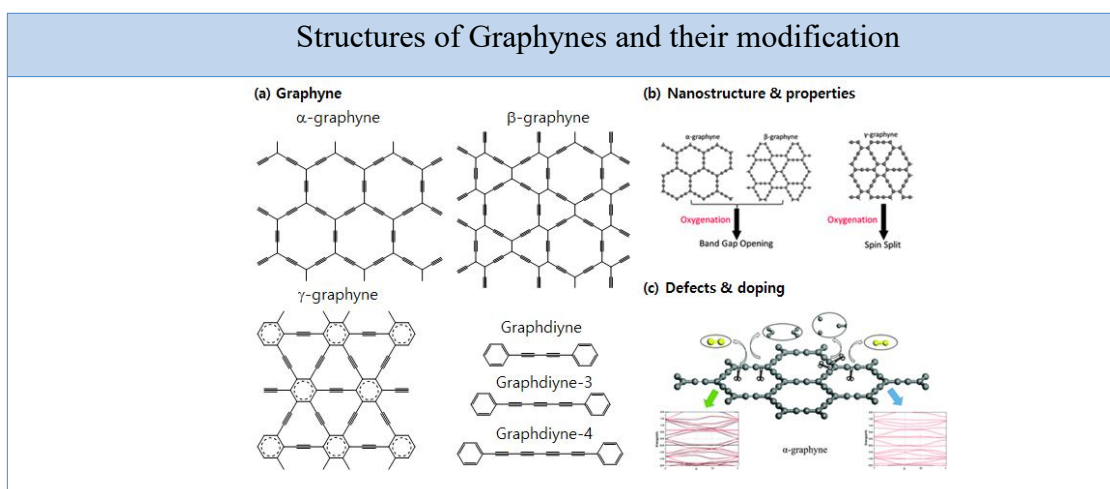
**Name:** Yingqi Tang

**Student ID:** 2020712633

Department of Chemistry

## 1. Introduction

During the past decades, the catalysts used for the hydrogen evolution reaction (HER) and oxygen evolution reaction (OER) have effectively accelerated the development of photo- and electrolytic water splitting. Due to the process of HER is relatively faster and simpler, researchers usually develop photo-electro-catalytic (PEC) materials for HER as the first step to promote water splitting performance. Low-dimensional materials, such as graphene, MoS<sub>2</sub>, graphyne, etc. are in the spotlight to replace these expensive catalysts. Graphynes electrochemical catalysts exhibit similar levels of reactivities as precious metal catalysts due to their high electrical conductivity, stability, and adsorption properties. Recent molecular orbital analysis results show that the high activity of graphynes electrochemical catalysts is due to the synergistic interaction between the  $\pi$  bonds of sp<sup>2</sup> hybrid orbitals and the d orbitals of adsorbed metals.

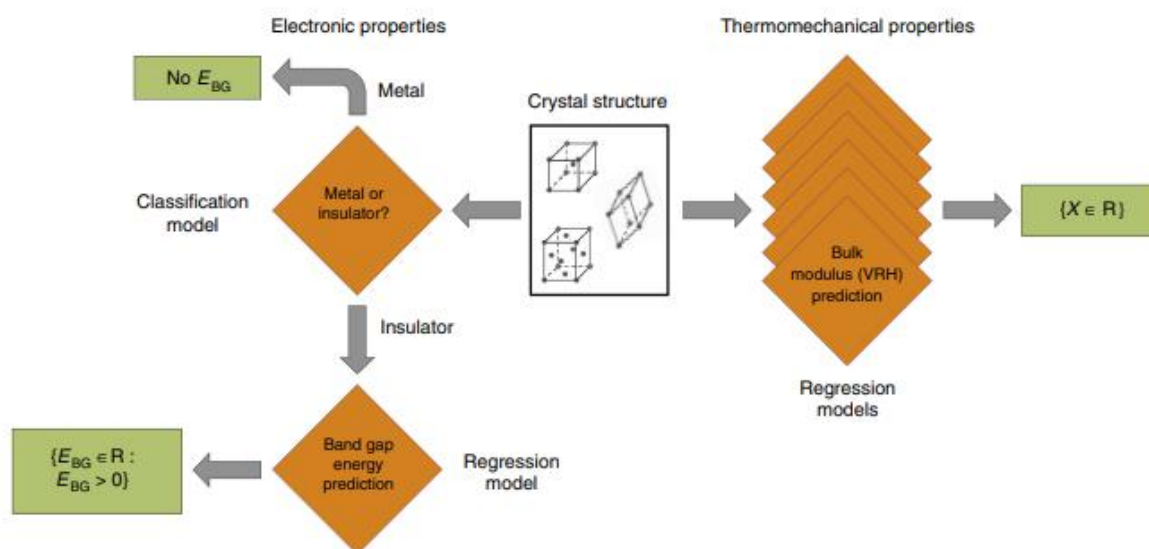


**Figure 1.** The structures of Graphynes and their modification.

With advances in technology, computational methods have tremendous advances as well. Many key properties such as cell voltage, thermal and electrolyte stability, phase diagram, and ion diffusivity can be accurately predicted by first-principle calculation. Especially, the computational approach provided insight into fundamental processes that are not easily experimentally accessible, such as bandgap, energy barrier, ionic diffusion mechanisms (ref), and capacity fading (ref). First-principles theory can not only calculate the physical and chemical properties of existing

materials but also predict the experimental results and design novel materials with excellent photo-electronic properties. However, the complexity of the structure of materials and the photo-electronic-catalytic reaction have brought severe challenges for theoretical calculation.

Recent progress in data-driven machine learning (ML) starts to stimulate great interests in the material field. ML can not only predict a series of material properties of stoichiometric inorganic crystalline materials but also start to show great power in assisting materials design and synthesis. It is anticipating that ML would assist in pushing the material revolution to a paradigm of full autonomy in the next 5-10 years, especially as emerging deep learning (DL) algorithms. However, the application of the DL in the material fields is still in its infancy. One of the main barriers is that a compatible and sophisticated descriptive system that enables correlation of the predicted properties to structures is required for materials because DL algorithms are original for imaging recognition. Combining First-principles and DL algorithms to predict the PEC materials can greatly improve the calculation efficiency and save the calculation time.



**Figure 2.** Outline of the modeling work-flow.

The objective of my research is to design Efficient PEC materials which have strong visible light absorption ability and good catalytic properties with deep learning & first-principles calculation. It involves the following issues: (1) noble metal-based

materials demonstrate the best catalytic activities but cost high, (2) most of the non-precious metal-based catalysts suffer from undesirable or unstable changes in morphology, etc. during the HER process, (3) there is no comprehensive theoretical study on the dynamics of all competitive processes in OER / HER charge transfer, (4) the catalyst activity can improve by investigating the principle of catalytic activity through the dynamic behavior of electrons and holes, (5) the dynamics of charge transfer reactions due to structural changes and chemical modifications in various catalytic systems require very large amounts of quantum calculations, so the DL algorithms can be useful.

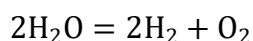
## 2. Methods

### DFT:

All density functional theory (DFT) calculations were performed in the Vienna ab initio simulation package (VASP). The generalized gradient approximation (GGA) with the Perdew–Burke–Ernzerhof (PBE) exchange–correlation functional and a 450 eV cutoff for the plane-wave basis set were employed. Due to the poor description of the weak van der Waals interactions of the popular PBE functional, an empirical dispersion-corrected density functional theory approach proposed by Grimme was used. The K-point was set to 3 \* 1 \* 1 for structural optimization and to 7\*7\*1 to calculate the electronic properties. The convergence threshold was set as 10<sup>-5</sup> eV in energy and 0.02 eV/Å in force. A vacuum space more than 20 Å was inserted in the z direction.

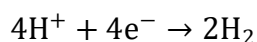
The specific reaction of electrolytic water on the two electrodes is related to the acidity and alkalinity of the electrolyte. The specific reaction is as follows:

Overall reaction:

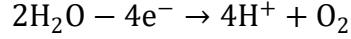


In acidic electrolyte solution:

cathode reaction:

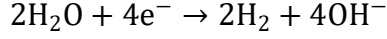


anode reaction:

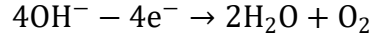


In alkaline electrolyte solution:

cathode reaction:



anode reaction:



The total reaction remains the same no matter the water electrolysis reaction takes place in acidic or alkaline solution conditions. It indicates that the thermal kinetic reaction process driven by the voltage required for water electrolysis is -1.23V, which is independent of the acid-base properties of the electrolyte. In fact, the voltage needed in the water electrolysis reaction is larger than 1.23V, so the part exceeding 1.23V was defined overpotential ( $\eta$ ), the  $\eta$  mainly compensates the electric energy consumed for resistance and contact resistance of the two electrodes and electrolyte solution. The circuit voltage can be defined as:

$$V = \varphi + \eta$$

Where  $\varphi$  is the voltage required for thermal dynamics at normal temperature and standard atmospheric pressure. Therefore, reducing overvoltage ( $\eta$ ) is the key to achieve efficiency improvement. On the one hand, a electrical catalytic for HER should includes an electrochemical process and a stripping process. So, the electrical catalytic HER can classify as Volmer - Heyrovsky reaction and Volmer - Tafel reaction. Researchers generally believe that the reaction rate is depending on the Gibbs free energy of  $\text{H}^+$  adsorbed on the catalyst surface. The exchange current density and conversion frequency of HER process from the experiment can directly reflect the catalytic properties of the catalyst. And the Stabatier Principle, which use the Gibbs free energy and exchange current density (independent of the type of catalyst) to quantitatively describe the catalytic performance of catalysts, is developed as one of the most important tools for designing and screening catalyst performance.

According to the DDMP database, the selected 86 2D-TMDs materials are used as an initial screening space to seek metallic, stable, and high-activity HER catalysts. The band gap of all these TMDs were calculated for revealing electronic conductivity.

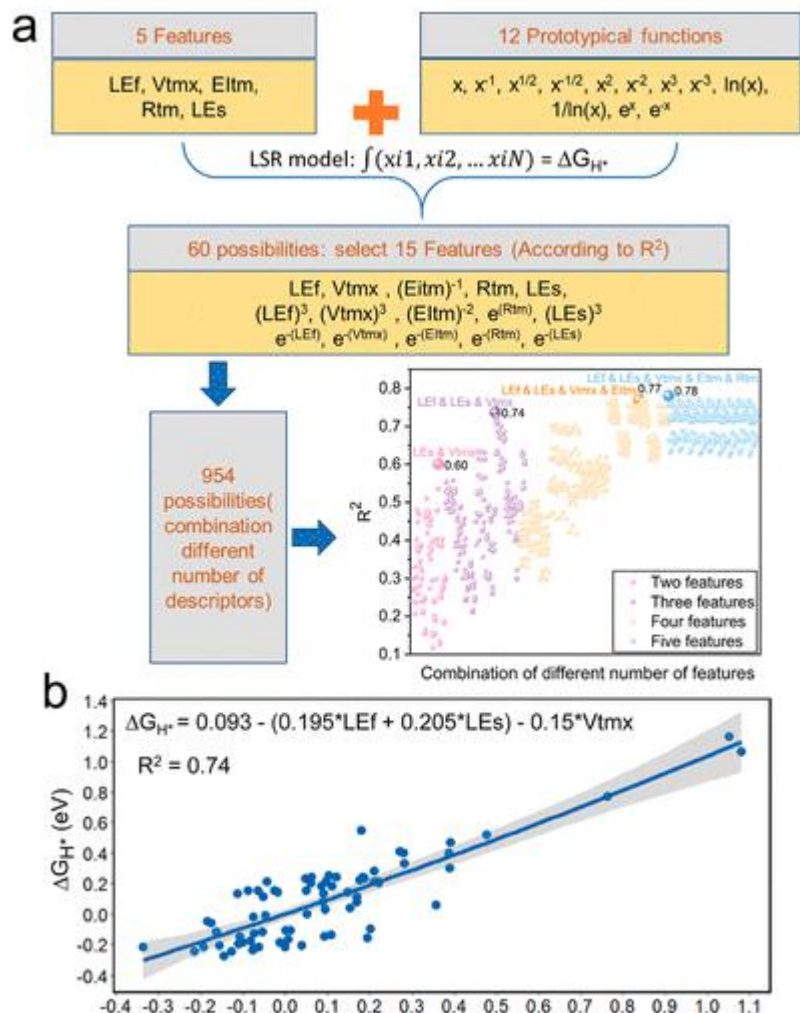
The approximately zero band gap is used to identify those TMDs with metallic phase. Subsequently, their formation energies were calculated by taking the reference of MoS<sub>2</sub> for determining whether they correspond to a lowest-energy phase among possibly existing phases (1H, 2H, 1T, 1T', 1T'', and others). Any calculated 2D-TMDs structures which meet the requirement of both lowest-energy phase and approximately zero band gap are determined to be potentially high-activity HER catalysts.

### **Deep learning:**

Determining HER Catalytic Descriptor through Machine Learning. To achieve a rational design principle for guiding the catalyst search, it is important to identify the characteristics of the active site. Although hydrogen adsorption strength ( $\Delta G_{H^*}$ ) of the reaction intermediate ( $H^*$ ) is extensively applied to correlate with the catalytic activity of HER, a more universal design principle is still lacking to describe atomic and structural characteristics for new catalyst design and activity optimization. Therefore, we performed ordinary linear regression with  $\Delta G_{H^*}$  being the HER target. A total of 27 types of features were considered; the descriptors can be categorized as (i) anion, (ii) cationic, or (iii) local area based on the previous electronic structure analysis features. It is necessary to identify which features may have significant influence on hydrogen adsorption strength. The dimensional reduction techniques such as principal component analysis and gradient boosting are used to filter out the most important features influencing  $\Delta G_{H^*}$ . The Pearson correlation coefficient ( $\gamma$ ), which accounts for the linear correlation between two features, was first employed to remove features that were highly correlated ( $\gamma > 0.90$ ). However, a high-dimensional machine learning model based on all these features not only is beyond our calculation capacity but also presents an over fitted characteristic. Because not all these features contribute equally to the target property, we employed gradient boosting, an algorithm that is designed to learn different “parts” of the training data to rank the descriptors in terms of their relative importance with respect to the HER property. Five features-nearest neighbor local electronegativity (LEf), average valence electron number of TM-X ( $V_{tmx}$ ), first ionization energy of transition metal ( $E_{itm}$ ), covalent radius of transition-metal atom

(Rtm), and next-nearest neighbor local electronegativity (LEs)—were selected to construct the surrogate models.

### 3. Discussion



**Figure 3.** Deep learning strategy.

Considering these five features, we constructed the randomForest (Breiman's random forest algorithm for classification and regression) and nnet (fit single-hidden-layer neural network, possibly with skip-layer connections) machine learning models. The fitting degree of randomForest and nnet are found to be higher than 0.90, respectively, indicating the importance of models. Although these five features are determined as the most important features to influence HER electrocatalytic activity, the random-Forest and nnet models are treated as a black box and cannot give an expression of  $\Delta G_{H^*}$  as functions of the LEf, Vtmx, LEs, Eltm, and

R<sub>tm</sub>. For experimental convenience, it is necessary to carry out further machine learning to establish a quantitative expression of  $\Delta G_{H^*}$ .

As shown in Figure 3a, 12 prototypical functions and 5 features are considered to perform the regression analysis by least-squares regression (LSR). The determination coefficient ( $R^2$ ) of the LSR fitting is used as the screening criterion of descriptors. Three rounds of LSR fitting and screening are performed. The first screening is carried out through combining each feature with one function to generate 60 possibilities, then choosing the three largest  $R^2$  combinations of each feature. Given the first round of screening of 15 features, the second screening is operated by combining two or three features to generate 306 possibilities. The L<sub>ef</sub>/L<sub>Es</sub> (local structure electronegativities) and the V<sub>tmx</sub> (average valence electron number of TM and X) jointly determine HER catalytic activity. Similarly, with the second round, the third round of screening is attempted through combining four or five features to generate 648 possibilities. If the E<sub>l</sub><sub>tm</sub> and R<sub>tm</sub> features are added, the fitting accuracy is not significantly increased, indicating that the L<sub>ef</sub>/L<sub>Es</sub> (local structure electronegativities) and the V<sub>tmx</sub> (average valence electron number of TM and X) already determine the HER catalytic activity.

The reliability of the produced analytical expression for hydrogen adsorption free energy from the machine learning is tested by comparing the predictions with experimental data available to date. The comparison among calculated, experimental data, and data calculated using the descriptor is plotted. The comparison indicates that our calculated catalytic activity is qualitatively consistent with experimental prediction. Some significant differences may be attributed to certain second-order manipulations, such as strain and support.