DFT, ML and GNN-Based Insights into Fluxional Nanoclusters for Fuel Cell Application

Ph.D. THESIS

by

RAHUL KUMAR SHARMA

DEPARTMENT OF CHEMISTRY

INDIAN INSTITUTE OF TECHNOLOGY INDORE

JANUARY 2026

DFT, ML, and GNN-Based Insights into Fluxional Nanoclusters for Fuel Cell Application

A THESIS

Submitted in partial fulfillment of the

requirements for the award of the degree

of

DOCTOR OF PHILOSOPHY

by

RAHUL KUMAR SHARMA

DEPARTMENT OF CHEMISTRY

INDIAN INSTITUTE OF TECHNOLOGY INDORE

JANUARY 2026

INDIAN INSTITUTE OF TECHNOLOGY INDORE

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled **DFT, ML, and GNN-Based Insights into Fluxional Nanoclusters for Fuel Cell Application** in the partial fulfillment of the requirements for the award of the degree of **DOCTOR OF PHILOSOPHY** and submitted in the **DEPARTMENT OF CHEMISTRY, Indian Institute of Technology Indore**, is an authentic record of my own work carried out during the time period from May 2021 to January 2026 under the supervision of **Prof. BISWARUP PATHAK**, Professor, Department of Chemistry, IIT Indore.

The matter presented in this thesis has not been submitted by me for the award of any other degree of this or any other institute.

Rahul Sharma

29-01-2026

Signature of the student with date

(RAHUL KUMAR SHARMA)

This is to certify that the above statement made by the candidate is correct to the best of my/our knowledge.

29-01-2026

Signature of Thesis Supervisor with date

(Prof. BISWARUP PATHAK)

RAHUL KUMAR SHARMA has successfully given his Ph.D. Oral Examination held on 20/05/2026

Signature of Thesis Supervisor with date

(Prof. BISWARUP PATHAK)

ACKNOWLEDGEMENTS

Completing a Ph.D. thesis is a journey shaped by individual dedication as well as the support of many people and institutions. As I reflect on these 4.5 years of transformative experience, I am filled with deep gratitude for the invaluable contributions of those who accompanied me along the way.

First and foremost, I express my deepest gratitude to my Ph.D. supervisor, ***Professor Biswarup Pathak****, whose unwavering guidance, expertise, and encouragement were instrumental in shaping every aspect of this research. Despite leading a large research group, you were always available for discussions and consistently provided valuable insights and direction. Your kindness, integrity, dedication, and compassionate approach have had a profound impact on me, both academically and personally. I will be forever grateful for the training and mentorship that contributed significantly to my academic development and personal growth. Thank you, Sir, for your constant support and trust. I extend my heartfelt thanks to the PSPC members, Professor Tridib Kumar Sarma and Professor Chelvam Venkatesh, for their insightful feedback, constructive criticism, and guidance throughout my Ph.D. journey.*

I am indebted to the Department of Chemistry at the Indian Institute of Technology Indore for providing a stimulating academic environment and access to essential research facilities. I sincerely thank the administrative staff, librarians, and technical support personnel for their assistance and cooperation. I am also thankful to my academic collaborators: Professor Rahul Banerjee (IISER Kolkata), Professor Subrata Kundu (CECRI), Professor Soumyabrata Roy (IIT Kanpur), and Professor Yuichi Negishi (Tokyo University of Science) for their valuable scientific interactions and collaborations. I gratefully acknowledge the University Grants Commission (UGC) for providing the doctoral fellowship.

I would like to extend special appreciation to Dr. Nishchal Bharadwaj for her guidance during the initial stages of my research, which helped establish a strong foundation for my work. I am also grateful to Dr. Milan

K. Jena for mentoring me during crucial phases of my Ph.D., offering constant encouragement and patiently guiding me through ML research projects. My sincere thanks to Harpriya Minhas for always being there with support, motivation, and help whenever needed, and for standing alongside me through challenging times. I am thankful to all present and former group members for their camaraderie, cooperation, and for making my time in the group both productive and enriching. Special thanks to my seniors from the MEMS department, Dr. Abhinav Sharma, Dr. Janmejai Sharma, and Dr. Vivek Rana, for their guidance and for sharing their experiences during critical stages of my journey.

Finally, I express my deepest gratitude to my beloved family. I thank my father, Mr. Roshan Lal Sharma, my mother, Mrs. Meena Kumari Sharma, and my sisters, Ms. Vijay Laxmi Sharma and Mrs. Malika Sharma, for their unconditional love, unwavering encouragement, and constant belief in me, which formed the foundation of this journey. Special thanks to my beloved Badi Maa, Mrs. Surinder Kaur, and my Mommy, Mrs. Surinder Kaur, for their invaluable support throughout this period. Their sacrifices, patience, and faith have been a continuous source of strength and motivation. I also extend my heartfelt thanks to my coach, Gajendra Singh Allsay, whose warm smile and constant efforts helped keep me mentally and physically fit throughout this long journey. In conclusion, I gratefully acknowledge all the individuals and institutions who contributed to the successful completion of this Ph.D. thesis. While the names mentioned here represent only a part of those who influenced my academic journey, their impact has been profound and enduring.

Ever Tried. Ever Failed. No Matter.
Try Again. Fail Again. Fail Better.

(Samuel Beckett)

Dedicated to my
Parents And Mentors

Synopsis

1. Introduction

Electrochemical energy conversion technology has emerged as the most promising avenue to address the growing environmental problems and global energy crisis in an eco-friendly, carbon-neutral cycle *[1]*. In this context, the development of high-performance electrocatalysts for devices such as regenerative fuel cells, metal-air batteries, and proton exchange membrane fuel cells (PEMFCs) *[2]*, remains a critical challenge despite their strong potential for clean energy conversion. In particular, the sluggish kinetics and high reaction overpotential values associated with the oxygen reduction reaction (ORR) at the cathode significantly limit the efficiency of PEMFCs and impede their large-scale commercialization. Moreover, the expensive Pt loading required and its susceptibility to poisoning and degradation under prolonged operating conditions further reduce cathodic efficiency, underscoring the need for alternative catalytic materials that minimize the use of precious metals.

Subnanometer clusters have recently emerges as an important class of heterogeneous electrocatalysts due to their exceptional catalytic performance and high atomic utilization efficiency *[3]*. At finite temperatures, these clusters exhibit a relatively flat potential energy surface (PES), which renders them highly dynamic and results in non-Arrhenius behavior. The presence of multiple undercoordinated atomic sites leads to pronounced fluxionality, giving rise to non-monotonic size-dependent activity trends that distinguish subnanoclusters from bulk and nanoparticle catalysts *[4]*. Recent studies have shown that clusters in the non-scalable regime have complex isomeric distribution and electronic structures, leading to metastability-triggered reactivity and the need for ensemble representation to accurately describe the overall activity *[5]*. Notably, the interaction of adsorbates with locally distinct coordination environments in subnano clusters can alter conventional scaling relationships, thereby enabling catalytic performance beyond the limits predicted by traditional

volcano plots. Despite these insights, a systematic understanding of the non-linear influence of transition-metal subnano clusters of varying sizes on ORR activity remains largely unexplored. This research gap persists due to the complexity of the extensive range of heterogenous catalysts, particularly within the non-scalable regime. Furthermore, exploring the vast chemical space of electrocatalysts within the size range where "each atom count" via high-throughput DFT calculations is challenging and necessitates a data-driven approach to quantify their ORR activity *[6]*.

Recently, machine learning (ML) has emerged as a powerful tool for accelerating the discovery of new catalysts by disentangling complex catalyst-intermediate interactions in various reactions, such as the CO_2 reduction reaction (CO_2RR), nitrogen reduction reaction (NRR), hydrogen evolution reaction (HER), and oxygen reduction reaction (ORR) *[7,8]*. In the context of nanoclusters, Panapitiya et al. and co-workers introduced a two-step feature- selection based on random forest models to accurately describe CO adsorption on thiolated gold nanoclusters using structural descriptors *[9]*. Similarly, Vlachos and coworkers developed surrogate energy models by integrating ML with global optimization, and cluster expansion techniques to study CO adsorption on Pd_n $(n = 1 - 55)$ clusters supported on $CeO_2(111)$ *[10].* While these studies provide important insights, the identification of fundamental descriptors governing adsorption energetics and catalytic origins in multistep ORR across different cluster sizes and transition-metal compositions in the subnanometer regime remains unresolved. In particular, the role of multifidelity electronic features in capturing the influence of d-electrons characteristics on intermediate adsorption within local coordination environments has not been systematically examined.

2. Objectives

The primary objectives of the thesis are outlined as follows:

- The first objective **(Chapter 2)** is to determine how dynamic reconstruction and fluxional behavior of Pt_{13} subnanoclusters affect

oxygen reduction reaction activity by evaluating the role of thermally accessible metastable isomers beyond the global minimum. This includes analyzing adsorption energetics, reaction mechanisms and kinetic barriers, quantifying the contribution of isomers within 0.4 eV at 300 K, and assessing stability under high oxygen coverage using ab initio thermodynamics to show how fluxionality governs overall catalytic performance.

- The second objective **(Chapter 3)** is to develop and apply a machine learning framework to predict the non-monotonic ORR activity of subnanometer clusters by capturing their fluxional behavior and size-dependent geometric and electronic features. This includes using non-ab initio descriptors to model site-specific adsorption energies, identify the electronic factors that reshape the ORR activity volcano, and screen highly active catalysts. The framework is further extended to ORR and OER to explain the shift of the volcano apex toward clusters such as Au_{11} and to identify bifunctional electrocatalysts, with Random Forest analysis defining the electronic design rules for catalyst discovery at the subnanometer scale.
- The third objective **(Chapter 4)** is to investigate how solvent-dependent fluxionality in graphene-supported Pt_7 and Pt_8 subnanoclusters influences ORR activity under electrochemical conditions. This includes comparing gas- and solvent-phase behavior, evaluating low-energy metastable ensembles, analyzing adsorption, scaling relationships and reaction mechanisms, quantifying effects of charge transfer, cluster–support interactions and the aqueous environment, and identifying stable oxidized phases via ab initio thermodynamics to understand how support- and solvent-driven metastability governs catalytic performance.
- The fourth objective **(Chapter 5)** is to develop a machine learning–based framework to quantify the size-dependent ORR activity of graphene-supported Pt_n/G $(n = 7 - 13)$ subnanoclusters by predicting site-specific adsorption energetics using non–ab initio descriptors. The study

aims to capture fluxional and non-monotonic catalytic behavior through ensemble representation, identify key geometric descriptors responsible for scaling-relation breaking, and elucidate support-induced structural evolution and activity under realistic ORR conditions to guide the design of efficient fuel cell electrocatalysts.

- The fifth objective **(Chapter 6)** is to identify the systems identify systems exhibiting non-scaling behavior and develop a machine learning–driven screening framework for heterometallic dual-atom catalysts embedded in metal surfaces to identify optimal combinations for bifunctional OER and ORR. By leveraging solid-state–derived d-band descriptors, the framework aims to capture non-scaling and non-monotonic activity trends without relying on extensive DFT calculations, evaluate potential-dependent catalytic behavior, and elucidate the role of d-states in governing electrocatalytic performance for fuel cell applications.
- The sixth objective **(Chapter 7)** is to develop a data-driven framework to screen late-transition-metal TM_{55} core–shell nanoclusters and evaluate their bifunctional electrocatalytic performance. This includes using geometric and electronic descriptors to identify TM_{55} catalysts with the lowest bifunctional overpotential, reshaping the activity volcano to explain apex shifts toward Au/Ag clusters, predicting two- or four-electron pathways for selectivity control leading to the H_2O_2 and H_2O formation, respectively. Additionally, we employed graph neural networks models with different architectures to derive structure–property relationships and clarify the role of d-states through electronic analysis.

3. Thesis Chapter Contribution

The contents of each chapter included in the thesis are discussed as follows.

3.1. Introduction (Chapter 1)

This chapter provides a comprehensive introduction to fluxional behavior, first-principles density functional theory (DFT), and a machine learning

(ML) framework, graph neural networks (GNN). The discussion focuses on the limitations of conventional catalyst representation through statistical ensemble representation, while deriving the need to account for the catalytic origins from the coexisting low-energy metastable ensemble (LEME) states. It then discusses a combined ML and DFT framework for mapping the activity of the nanoclusters, which is further explored in the subsequent chapters of the thesis. Additionally, the second part discusses **(a)** the basics of DFT employed for electronic structure calculations, **(b)** ML concepts and techniques that are used in this thesis, and **(c)** the GNN framework used in this thesis.

3.2. Role of Fluxionality and Metastability in Catalytic Interfaces (Chapter 2)

Understanding the dynamic reconstruction of active sites and the fluxional behavior of subnanoclusters is central to operando catalysis modeling. In addition to the global minimum (GM), an accurate description of catalytic phenomena must account for thermally accessible ensembles of metastable isomers. Using a Pt_{13} cluster as a model system for gas-phase oxygen reduction reaction (ORR), we investigated how metastable configurations influence catalyst dynamics and reactivity (Figure 1). Our results show that these isomers exhibit distinct adsorption energetics and ORR activity, underscoring the critical role of fluxionality in subnano cluster catalysis. The mechanistic analysis identifies several metastable structures with ORR performance comparable to the GM. Statistical ensemble treatment highlights the significant contribution of isomers within 0.4 eV of the GM, while higher-energy forms contribute negligibly at 300 K. Finally, *ab initio* thermodynamic analysis reveals the evolving stability of Pt_{13} clusters under high oxygen coverage. Overall, the study demonstrates that metastable isomers are essential for determining the cumulative catalytic activity of subnanometer clusters.

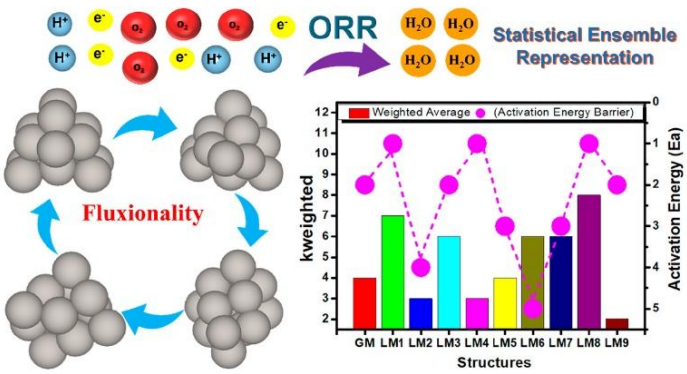


Figure 1: Schematic illustration of Pt_{13} nanoclusters exhibiting fluxional behavior during multistep ORR, represented using a statistical ensemble of metastable structures.

3.3. Reshaping the Activity Volcano in the Subnanometer Regime (Chapter 3)

In computational heterogeneous catalysis, Sabatier's principle-based activity volcano plots provide an intuitive guide to catalyst design but impose a fundamental constraint on the maximum achievable catalytic performance. Recently, subnano clusters have emerged as an exciting platform, offering high noble metal utilization and superior performance for various reactions compared to extended surfaces, reflecting a complex structure–activity relationship in the non-scalable regime. However, understanding their non-monotonic catalytic activity, attributed to the large configurational space and their fluxional identity, poses a formidable challenge. Here, we present a machine learning (ML) framework that captures the non-monotonic trends in oxygen reduction reaction (ORR) activity at the subnanometer scale, attributed to their dynamic fluxional characteristics (Figure 2). We demonstrate a size-dependent shifting and reshaping of the ORR activity volcano, with Au replacing Pt at the peak. Leveraging only the non-*ab initio* geometric and electronic properties, our trained ML model accurately captures the site-specific adsorption energies of intermediates at the subnanometer regime. To account for the inconsistent trend in activity, we analyzed the correlation between electronic and geometric properties. Our findings reveal that the d-filling

and coupling matrix of the neighboring metal atom significantly influences the intermediate adsorption on the local chemical environment compared to the d-band center. Following this analysis, we utilized ML to map the catalyst distribution in the activity volcano and identified the five best sub-nano electrocatalysts, demonstrating overpotential values lower than or comparable to the Pt (111) surface for the ORR.

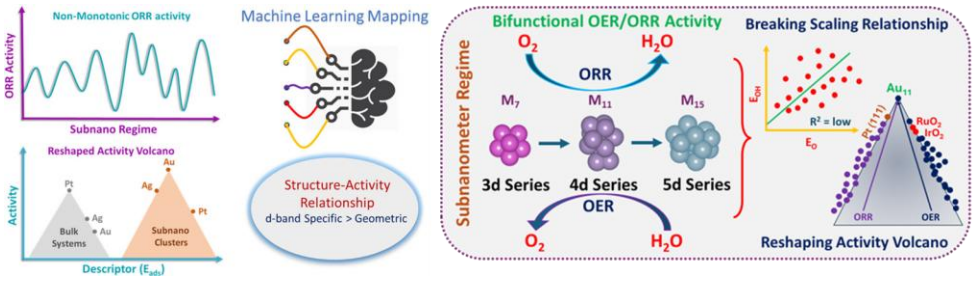


Figure 2: Schematic representation of ML-driven reshaping and mapping of the ORR activity volcano in the subnanometer regime.

We further investigated the shift in the apex of the activity volcano from Pt (111) and IrO_2 surfaces to Au_{11} clusters for both OER and ORR to identify promising bifunctional catalysts. Late transition metal subnano clusters, particularly Au_{11}, emerged as the most effective bifunctional electrocatalyst, exhibiting markedly lower overpotentials. To understand the governing factors, we grouped the catalysts into three clusters and applied Random Forest Regression to assess the influence of non-ab initio electronic features on OER and ORR activity. The analysis revealed that d-band filling is the dominant descriptor responsible for the superior bifunctional performance of these subnano clusters.

3.4. Ensemble Representation for Supported Clusters in Implicit Solvent (Chapter 4)

Subnanoclusters have gained attention in heterogeneous catalysis for their high metal utilization and strong activity. A key challenge, however, is understanding the dynamic reconstruction of active sites under realistic operating conditions. In this work, we present a multiscale investigation of solvent-dependent fluxionality in Pt_7 and Pt_8 clusters supported on graphene, analyzed in both gas and solvent phases using an implicit

solvation model (Figure 3). Under electrochemical conditions, the pronounced fluxional behavior of metastable structures requires a statistical ensemble representation of activity in addition to conventional global minima (GM) analysis. The mechanistic study of the oxygen reduction reaction (ORR) reveals distinct adsorption behavior, meaningful scaling relationships, and reduced overpotential values, highlighting superior ORR performance at the subnano scale relative to bulk and nanoclusters. We further find that strong charge transfer, cluster–support interactions, and the aqueous environment enhance the contribution of low-energy metastable ensembles (LEMEs) to overall ORR activity. *Ab initio* thermodynamic analysis identifies the stable oxidized phases of Pt clusters under ORR conditions.

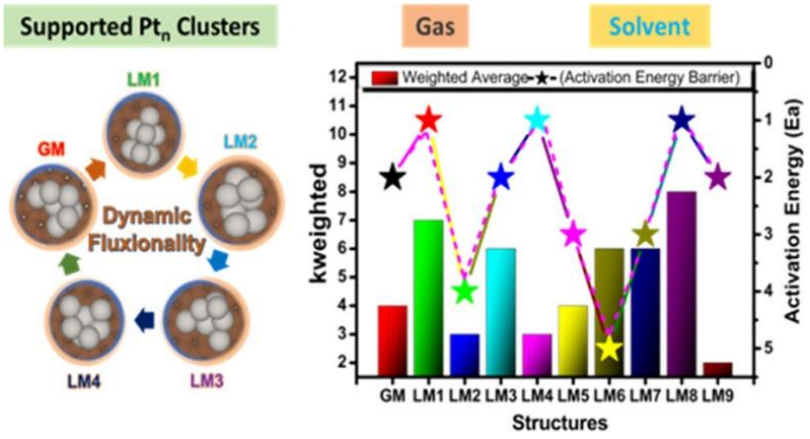


Figure 3: Illustration metastability triggered reactivity in implicit solvent medium for graphene supported Pt_7 and Pt_8 subnano clusters for ORR.

3.5. Deciphering Key Descriptors for Non-Scaling Relationships (Chapter 5)

Subnano clusters have emerged as a promising class of electrocatalysts, enabling efficient utilization of noble metals and superior activity for key electrochemical reactions. However, their fluxional nature and complex structure–activity relationships give rise to non-monotonic catalytic behavior, making activity evaluation highly challenging. Here, we develop a machine learning (ML) framework to assess the size-dependent activity of graphene-supported $Pt_n (n = 7 - 13)$ subnano clusters for the oxygen

reduction reaction (ORR) (Figure 4). Using non-*ab initio* descriptors, the trained ML model accurately predicts site-specific adsorption energies of key ORR intermediates across heterogeneous active sites.

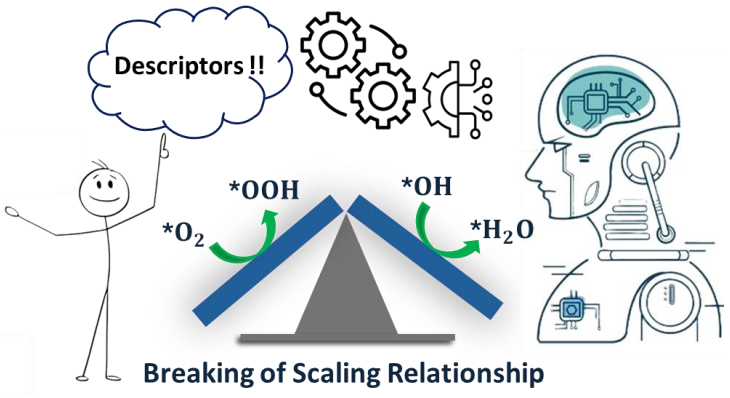


Figure 4: Schematic illustration of descriptor design and revealing non-scaling behavior assess the size-dependent activity of graphene-supported $Pt_n (n = 7 - 13)$ electrocatalysts.

Feature-driven uncertainty quantification highlights the irregular catalytic activity and the importance of ensemble representation, while identifying the key geometric descriptors responsible for breaking the scaling relationship in the subnanometer regime. Furthermore, we uncover critical insights into the dynamics of active electrocatalysts under ambient ORR conditions on support, showing their strong oxidative tendencies that drive monolayer formation of intermediates. We also derive reaction networks for the rate-determining steps, pinpointing the saturation onset point that signals changes in binding sites.

3.6. Unraveling Scaling Relationships in Dual-Atom Catalysts (Chapter 6)

Dual-atom catalysts (DACs) have emerged as a new frontier in heterogeneous catalysis, offering improved stability and superior performance in key electrocatalytic reactions. However, identifying optimal multimetallic DACs combination for a multistep reaction is challenging due to the vast chemical space. Herein, we develop a machine learning (ML) framework to expedite the screening of DACs, which consist of a

heterometallic dimer embedded in the surface layer of a metal host, for improved oxygen evolution reaction (OER) and oxygen reduction reaction (ORR) performance (Figure 5). We encode the solid-state-derived d-band descriptors to accurately train the ML model and effectively capture the non-monotonic bifunctional activity on DACs, without requiring expensive DFT calculations. Interestingly, we identify the non-scaling behaviour of these DACs, with CoPd and CoCu dimer exhibiting superior OER and ORR activity. Furthermore, we employ the surface charging method to evaluate the potential-dependent activity and reveal the non-linear relationship between catalytic activity and electrode potential.

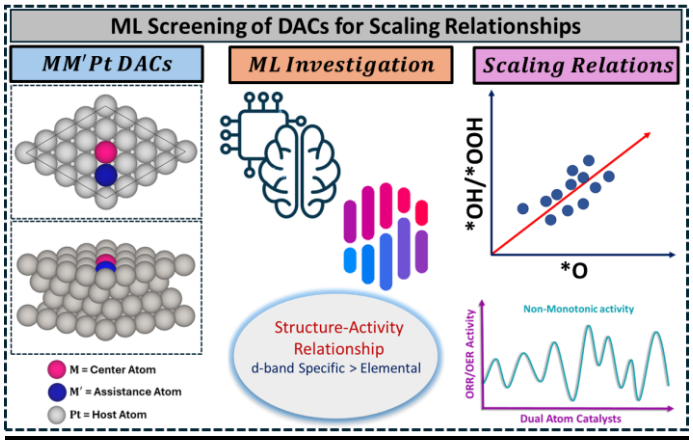


Figure 5: Illustration of investigating structure-activity relationships and scaling relationships in dual-atom catalysts for fuel cell applications.

3.7. Steering Morphology Based Activity and Selectivity Relationships (Chapter 7)

Nanoclusters engineered with atomic precision offer a promising route toward next-generation electrode materials due to their exceptional catalytic performance. In this work, we present a data-driven framework to screen late-transition-metal core–shell nanoclusters for bifunctional electrocatalysis (Figure 6). By integrating geometric and electronic descriptors, we uncover morphology-dependent trends that govern activity and selectivity, highlighting the importance of structural diversity in fuel cell applications. Our analysis reveals a unique single-cluster catalyst identity for M_{55} nanoclusters, where intermediate adsorption is dictated by

the constituent metals' electronic and elemental characteristics. Guided by the Sabatier principle, we expose limitations of conventional numerical methods and establish a reshaped activity volcano in which the peak shifts from RuO_2 and Pt toward Au/Ag-based nanoclusters. Finally, the trained ML model enables prediction of electrocatalysts favouring two- or four-electron pathways, allowing rational control over selectivity between H_2O_2 and H_2O formation.

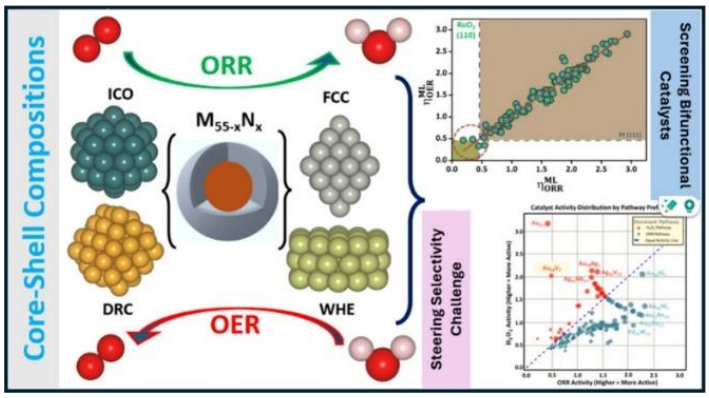


Figure 6: Illustration of ML-driven core-shell design of M_{55} nanoclusters for pathway selectivity between H_2O_2 and H_2O formation.

4. Conclusions

This thesis develops a comprehensive and unified framework to describe electrocatalysis in the subnanometer regime by moving beyond static, single-structure, and bulk-inspired models. Through the integration of global optimization, operando-relevant DFT calculations, statistical ensemble representations, and interpretable machine learning, the work demonstrates that fluxionality, metastable isomers, local coordination environments, and reaction conditions collectively dictate ORR and OER activity. Across supported clusters, bare clusters, core–shell alloys, and dual-atom catalysts, the results reveal pronounced non-monotonic, size-dependent, and morphology-dependent catalytic behavior, accompanied by systematic breaking of scaling relations and shifts in volcano plot apexes. The identification of low-energy metastable states as active contributors, together with solvent- and support-induced restructuring, provides a realistic picture of experimentally observed activity. Importantly, the use of

physically motivated, low-cost geometric and electronic descriptors enables accurate prediction of adsorption energetics and overpotentials across large chemical spaces without reliance on expensive *ab initio* electronic features. Overall, this thesis establishes metastability-triggered, data-driven catalyst design as a general and transferable strategy for discovering efficient electrocatalysts beyond the traditional volcano paradigm.

5. References

1. Lewis, N. S. Research Opportunities to Advance Solar Energy Utilization. *Science*, 2016, *351*(6271), p.aad1920 (DOI: 10.1126/science.aad1920).
2. Shi, Z.; Li, J.; Wang, Y.; Liu, S.; Zhu, J.; Yang, J.; Wang, X.; Ni, J.; Jiang, Z.; Zhang, L.; Wang, Y.; Liu, C.; Xing, W.; Ge, J. Customized Reaction Route for Ruthenium Oxide towards Stabilized Water Oxidation in High-Performance PEM Electrolyzers. Nat. Commun. **2023**, *14*, 843 (DOI: 10.1038/s41467-023-36380-9).
3. Peterson, E. J.; DeLaRiva, A. T.; Lin, S.; Johnson, R. S.; Guo, H.; Miller, J. T.; Hun Kwak, J.; Peden, C. H. F.; Kiefer, B.; Allard, L. F.; Ribeiro, F. H.; Datye, A. K. Low-Temperature Carbon Monoxide Oxidation Catalysed by Regenerable Atomically Dispersed Palladium on Alumina. Nat. Commun. **2014**, *5*, 4885 (DOI : 10.1038/ncomms5885).
4. Zhai, H.; Alexandrova, A. N. Fluxionality of Catalytic Clusters: When It Matters and How to Address It. ACS Catal. **2017**, *7*, 1905– 1911 (DOI : 10.1021/acscatal.6b03243).
5. Zandkarimi, B.; Alexandrova, A. N. Dynamics of Subnanometer Pt Clusters Can Break the Scaling Relationships in Catalysis. J. Phys. Chem.Lett. **2019**, *10*, 460– 467(DOI: 10.1021/acs.jpclett.8b03680).
6. Heiz, U.; Sanchez, A.; Abbet, S.; Schneider, W. D. Catalytic Oxidation of Carbon Monoxide on Monodispersed Platinum Clusters: Each Atom Counts. J. Am. Chem. Soc. **1999**, *121*, 3214– 3217 (DOI : 10.1021/ja983661).

7. Mai, H.; Le, T. C.; Chen, D.; Winkler, D. A.; Caruso, R. A. Machine Learning for Electrocatalyst and Photocatalyst Design and Discovery. Chem. Rev. **2022**, *122*, 13478– 13515 (DOI : 10.1021/acs.chemrev.2c00061).
8. Ma, X.; Li, Z.; Achenie, L. E. K.; Xin, H. Machine-Learning-Augmented Chemisorption Model for CO_2 Electroreduction Catalyst Screening. J. Phys. Chem. Lett. **2015**, *6*, 3528– 3533 (DOI: 10.1021/acs.jpclett.5b01660).
9. Panapitiya, G.; Avendano-Franco, G.; Ren, P.; Wen, X.; Li, Y.; Lewis, J. P. Machine-Learning Prediction of CO Adsorption in Thiolated, Ag-Alloyed Au Nanoclusters. *J. Am. Chem. Soc.* **2018**, *140*, 17508– 17514 (DOI : 10.1021/jacs.8b08800).
10. Hensen, E. J. M.; Vlachos, D. G.; Wang, Y.; Su, Y. Q. Finite-Temperature Structures of Supported Subnanometer Catalysts Inferred via Statistical Learning and Genetic Algorithm-Based Optimization. *ACS Nano* **2020**, *14* (10), 13995– 14007 (DOI : 10.1021/acsnano.0c06472).

6. List of Publications

1. **Sharma R. K.**, Nair A. S., Bharadwaj N., Roy D., Pathak B. (2023). Role of fluxionality and metastable isomers in the ORR activity: a case study. *J. Phys. Chem. C 2023, 127, 1, 217–228* (*DOI : 10.1021/acs.jpcc.2c06265*).
2. **Sharma R. K.**, Minhas H., & Pathak B. (2024). Investigating the Metastability-Triggered Reactivity of Pt7, 8 Clusters on Graphene: Unraveling Statistical Ensemble Representation for ORR in Gas and Implicit Solvent Phases. *J. Phys. Chem. C 2024, 128, 18, 7504–7517* (*DOI : 10.1021/acs.jpcc.4c00376*).
3. **Sharma R. K.**, Minhas H., Pathak B. (2024). High-throughput screening of bifunctional catalysts for oxygen evolution/reduction reaction at the subnanometer regime. *Nanoscale, 2024,16, 21340-21350* (*DOI : 10.1039/D4NR02787F*).

4. **Sharma R. K.**, Jena M. K., Minhas H., Pathak B. (2024). Machine-learning-assisted screening of nanocluster electrocatalysts: mapping and reshaping the activity volcano for the oxygen reduction reaction. *ACS Appl. Mater. Interfaces 2024, 16, 46, 63589–63601 (DOI : 10.1021/acsami.4c14076).*
5. **Sharma R. K.**, Minhas H., Pathak B. (2025). Machine Learning-Guided Discovery of Alloy Nanoclusters: Steering Morphology-Based Activity and Selectivity Relationships in Bifunctional Electrocatalysts. *ACS Appl. Mater. Interfaces 2025, 17, 28, 40488–40498 (DOI : 10.1021/acsami.5c07198).*
6. **Sharma R. K.,** Maity S., Pathak B. (2025), Deciphering Key Descriptors for Scaling Relationships in Graphene-Supported Pt_n Clusters via Machine Learning *(Small, recently accepted)*
7. **Sharma R. K.**, Minhas H., Pathak B. (2025), Unraveling Scaling Relationships in Dual-Atom Catalyst with Electronic Descriptors: A Machine Learning Investigation for OER/ORR Activity (JPCL, recently accepted).
8. **Sharma R. K.**, Minhas H., Pathak B. (2025), Graph Neural Network for High-Throughput Screening of Nanocluster Electrocatalysts: Decoding Morphology-Activity Relationship in Oxygen Reduction Reaction (manuscript communicated).
9. Minhas H., Das S., **Sharma R. K.**, Pathak B. (2024). Rattling induced bonding hierarchy in Li–Cu–Ti chalcotitanates for enhanced thermoelectric efficiency: a machine learning potential approach. *J. Mater. Chem. A, 2024,12, 25988-25999 (DOI : 10.1039/D3TA07461G).*
10. Sadhukhan A., Karmakar A., Koner K., Karak S., **Sharma R. K.**, Roy A., Sen P., Dey K. K., Mahalingam V., Pathak B., Kundu S., Banerjee R. (2024) Functionality Modulation Toward Thianthrene-based Metal-Free Electrocatalysts for Water Splitting. *Adv. Mater. 2024, 36, 2310938 (DOI : 10.1002/adma.202310938).*

11. Minhas H., Jena M. K., **Sharma R. K.**, & Pathak B. (2024). Machine learning-driven inverse design and role of dopant for tuning thermoelectric efficiency. *ACS Appl. Electron. Mater. 2024, 6, 8, 5815–5826 (DOI : 10.1021/acsaelm.4c00808).*

12. Pal H., Karmakar A., Sadhukhan A., Koner K., Karak S., **Sharma R. K.**, Ghosh M., Dey K. K., Pathak B., Kundu S., Banerjee R. (2024). Designing Salen-Based Porous Organic Polymers for Enhanced Electrolytic Water Splitting into Oxygen. *Adv. Funct. Mater. 2024, 34, 2408255 (DOI : 10.1002/adfm.202408255).*

13. Minhas H., Jena M. K., **Sharma R. K.**, Pathak, B. (2024). Insights into Thermal Conductivity of Pnictogen Chalcogenides: Machine Learning Stereochemically Active Lone Pairs and Hybridization. *Chem. Mater. 2025, 37, 1, 198–209 (DOI : 10.1021/acs.chemmater.4c02294).*

14. Minhas H., **Sharma R. K.**, Pathak B. (2025). Exploring Combined Stereochemically Active Lone-Pair and Rattling Effects in Thermoelectrics with Machine Learning Potentials. *J. Phys. Chem. Lett. 2025, 16, 20, 5051–5061 (DOI : 10.1021/acs.jpclett.5c00952).*

15. Pal H, Karmakar A., Sadhukhan A., **Sharma R. K.**, Paul S., Addicoat M. A., Nishiyama Y., Pathak B., Kundu S., Banerjee R. (2025). Cation-Assisted Water Oxidation with Crown Ether-Based Covalent Organic Frameworks. *J. Am. Chem. Soc. 2025, 147, 24, 20406–20416 (DOI : 10.1021/jacs.5c02083).*

16. Roy S. S., **Sharma R. K.**, Karmakar A., Nagappan S., Pathak B., & Kundu S. (2025). Triggering the water oxidation kinetics and reaction pathway via S-doping in layered hydroxides for enhanced electrocatalytic performance. *Applied Catalysis B: Environment and Energy 371 (2025) 125227 (DOI : 10.1016/j.apcatb.2025.125227).*

17. Minhas H., **Sharma R. K.**, Pathak B. (2025). Antibonding States Drive Anharmonicity and Low Thermal Conductivity in Edge-Sharing Metal Chalcogenides. *ACS Appl. Mater. Interfaces 2025, 17, 44, 60669–60681 (DOI : 10.1021/acsami.5c17211).*

18. Jayan P., Jana A., Li Z., **Sharma R.K.**, Yadav V., Mata A.C., Li M.H., Zeraati A. S., Base T., Hung S. F., Wu J., Pathak B., Pradeep T., Roy S. CO2 Electroreduction to CO Over Silver Nanoclusters: The Impact of Nuclearity on Synergistic Activity Modulation. *Small 2025, e05305 (DOI : 10.1002/smll.202505305).*

19. Ghosh P., Roy D., Das A., **Sharma R. K.**, Pathak B. (2025). Dynamic Surface Evolution and O Diffusion in High-Index Cu2O Surfaces for Enhanced Electrochemical CO2 Reduction. *J. Phys. Chem. C 2025, 129, 49, 21622–21633 (DOI : 10.1021/acs.jpcc.5c05945).*

20. Minhas H., **Sharma R. K.**, Pathak B. (2025). Unravelling Lone Pair Induced Bonding Effects on Thermal Conductivity in Metal Chalcogenides using Machine Learning Potentials. *J. Mater. Chem. A (recently accepted)*

Table of Contents

List of Figures and Schemes

Chapter 3

each intermediate are outlined with a rectangular dotted box, with their performance metrics marked in bold. Parity plots comparing DFT-calculated and ML-predicted (d) E_{*OOH}, (e) E_{*O}, and (f) E_{*OH} values are shown for the best-performing GBR model. (B) ML Models Explainability and Interpretability: Global interpretation of the best-performing models for extracted features in predicting (g) E_{*OOH}, (h) E_{*O}, and (i) E_{*OH}. Beeswarm plot that displays an information-dense summary of how the most important features affect the prediction of (j) E_{*OOH}, (k) E_{*O}, (l) and E_{*OH}. A solitary dot on the feature row represents each datapoint with its SHAP value on the x-axis. The color gradient along the y-axis represents the feature values, ranging from low to high.

Figure 6.4 (A) OER and ORR activity plot. Distribution of overpotentials for OER and ORR activity derived from (a) ML-predicted (η_{ML}^{OER}, η_{ML}^{ORR}), and (b) DFT-calculated (η_{DFT}^{OER}, η_{DFT}^{ORR}) screening. The intersection of the horizontal and vertical dashed lines indicates the low-overpotential regions, serving as a reference point to identify active bifunctional electrocatalysts from the ML and DFT dataset. Radar plots represent the predicted (c) η_{ML}^{OER} and (d) η_{ML}^{ORR} values from the trained GBR model for late-transition metal-based DACs (M = Ag, Au, Ir, Os, Rh, and Ru) included as the unseen dataset. (B) Electronic States Analysis. Partial density of states (PDOS) for key OER/ORR **270**

marked by η_{ORR}^{GNN} = 0.00 V. Kernel density estimation (KDE) plots illustrating the correlation between electronic descriptors and η_{ORR}^{GNN}: Coupling matrix vs η_{ORR}^{GNN} for (d) 3d: $Co_{42}M_{13}$, $Ni_{42}M_{13}$, $Cu_{42}M_{13}$, (e) 4d: $Rh_{45}M_{10}$, $Pd_{45}M_{10}$, $Ag_{42}M_{13}$, and (f) 5d: $Ir_{48}M_7$, $Pt_{46}M_9$, and $Au_{48}M_7$; d-band center vs η_{ORR}^{GNN} for the same metal groups are represented in (g-i). KDE contours represent smoothed density distributions, highlighting regions of high data concentration in descriptor-overpotential space. Marginal histograms display distribution trends of individual variables, offering insight into underlying structure-activity relationships.

Figure 7.16 Statistical analysis of ORR activity. (a) Scatter plot comparing GNN-predicted ORR overpotentials η_{ORR}^{GNN} with DFT-computed values η_{ORR}^{DFT} for various core-shell nanocluster configurations. Each marker represents a unique nanocluster, and the orange and blue dashed lines indicate the benchmark overpotentials of Pt(111) from DFT and GNN, respectively. The shaded orange region highlights catalysts outperforming Pt(111) by both methods. (b) Pearson correlation coefficients showing relationships among input features and their correlation with E_{ads} and η_{ORR}^{DFT}. (c) t-SNE projection of NC alloys coloured by Cluster ID, illustrating distinct zones corresponding to high and low-overpotential (η) catalysts based on E_{ads} and η_{ORR}^{DFT}. (d) Silhouette score analysis used to **330**

List of Tables

Chapter 5

Chapter 6

Acronyms

PEMFCs	Proton-Exchange Membrane Fuel Cells
ORR	Oxygen Reduction Reaction
AFCs	Alkaline Fuel Cells
DMFCs	Direct Methanol Fuel Cells
SMSI	Strong Metal-Support Interaction
AIMD	*ab initio* Molecular Dynamics
DFT	Density Functional Theory
KS	Kohn-Sham
XC	Exchange-Functional
LDA	Local Density Approximation
GGA	Generalized Gradient Approximation
VASP	Vienna *ab initio* simulation package
CINEB	Climbed image nudged elastic band
MEP	Minimum energy path
HTST	Harmonic transition state theory
CHE	Computational Hydrogen Electrode
ML	Machine Learning
PCC	Pearson Correlation Coefficient
SCC	Spearman Correlation Coefficient
MI	Mutual Information
GNN	Graph Neural Network
CGCNN	Crystal Graph Convolutional Neural Network
ALIGNN	Atomistic Line Graph Neural Network
GATGNN	Global Attention Graph Neural Network
OER	Oxygen Evolution Reaction
ORR	Oxygen Reduction Reaction
HER	Hydrogen Evolution Reaction
GO	Global Optimization
GM	Global Minima
LM	Local Minima

PGOPT	Parallel Global Optimization Pathway Toolkit
LEME	Low Energy Metastable Ensembles
GCN	Generalized Coordination Number
CN	Coordination Number
BEP	Bell-Evans-Polanyi
RDS	Rate-Determining Step
CDD	Charge Density Difference
DOS	Density of States
pMuTT	Python Multiscale Thermochemistry Toolbox
PBE	Perdew-Burke-Ernzerhof
DACs	Dual-Atom Catalysts
QSPR	Quantitative Structure Property Relationships
CPM	Constant Potential Method
AEM	Adsorbate Evolution Mechanism
XGBR	Extreme Gradient Boosting Regression
KRR	Kernel Ridge Regression
LR	Linear Regression
GPR	Gaussian Process Regression
RFC	Random Forest Classifier
SVM	Support Vector Machine
STO	Slater Type Orbitals
GTO	Gaussian Type Orbitals

CHAPTER 1

Introduction

CHAPTER 1

1.1 Electrochemical Energy Conversion

Electrochemical energy conversion has become one of the most viable strategies for mitigating environmental degradation and addressing the global energy demand through sustainable and carbon-neutral processes *[1,2]*. Within this context, the development of efficient electrocatalysts is central to technologies such as regenerative fuel cells *[3]*, metal–air batteries *[4]*, and proton-exchange membrane fuel cells (PEMFCs) *[5]* yet remains a major scientific and technological challenge. Critical limitation arises from the oxygen reduction reaction (ORR) occurring at the cathode, where slow reaction kinetics and large overpotentials prevent operation close to thermodynamic efficiency and impede the large-scale deployment of PEMFC systems *[6,7]*. Moreover, the high cost of platinum and its vulnerability to deactivation during prolonged operation further reduce cathodic efficiency, highlighting the need to minimize precious-metal usage through alternative catalyst design *[8]*.

In recent years, nanoclusters and subnano clusters have gained considerable attention as a new class of heterogeneous electrocatalysts due to their exceptional atomic utilization and enhanced catalytic performance, enabling a substantial reduction in noble-metal content *[9]*. At finite temperatures, these clusters are characterized by shallow and highly anharmonic potential energy surfaces (PES), which render them intrinsically dynamic and give rise to deviations from conventional Arrhenius-type behavior *[10]*. The presence of numerous undercoordinated atomic sites leads to pronounced fluxionality, resulting in size-dependent and non-monotonic activity trends that differ fundamentally from those observed in bulk catalysts *[11,12]*. Furthermore, clusters operating in the non-scalable regime exhibit intricate structural and electronic heterogeneity, where metastable configurations play an active role in governing reactivity, thereby necessitating an ensemble-based description to capture the true catalytic response *[13,14]*. Notably, interactions between adsorbates and the local coordination environment of subnano clusters can

break the traditional scaling relationships, opening pathways to surpass the performance constraints predicted by volcano-type correlations.

1.2 Overview of Fuel Cells

The high operational and maintenance costs associated with energy systems, combined with the inherent intermittency of renewable sources such as wind and solar, have shifted significant research attention toward efficient energy storage and conversion technologies. Among the available options, fuel cells stand out as highly adaptable energy devices due to their high energy density, portable nature, and zero-emission operation *[15]*. As a result, they have found widespread use in transportation, aerospace, portable electronics, and hybrid power systems. A notable recent development is the introduction of the hydrogen-powered Coradia iLint by Alstom, recognized as the world's first commercial train operated entirely using fuel cell technology, marking a major advancement in sustainable transportation.

Fuel cells function as electrochemical systems that directly convert chemical energy into electrical energy via redox reactions at the electrodes. At the anode, fuel oxidation releases electrons, which travel through an external circuit to the cathode, where they participate in the corresponding reduction reaction. This direct electron-driven energy conversion bypasses the intermediate thermal and mechanical steps required in conventional engines, thereby minimizing energy losses and enhancing overall efficiency. Based on their operating temperature, fuel cells are generally categorized into low-temperature and high-temperature systems. Low-temperature fuel cells include proton-exchange membrane fuel cells (PEMFCs), alkaline fuel cells (AFCs), and direct methanol fuel cells (DMFCs), while phosphoric acid fuel cells, molten carbonate fuel cells, and solid oxide fuel cells belong to the high-temperature class *[16,17]*. These fuel cell types differ in their operating principles, electrode and electrolyte materials, and application domains, as summarized in Table 1.1.

Table 1.1 Types of fuel cells and their features and their applications.

Type of Fuel Cell	Working Temperature	Anode Reaction	Cathode Reaction	Applications
Solid-Oxide Fuel Cell (SOFC)	600-1000 °C	$H_2 + O^{2-} \rightarrow H_2O + 2e^-$	$\frac{1}{2}O^{2-} + 2e^- \rightarrow O^{2-}$	Automobiles Range extenders
Molten Carbonate Fuel Cell (MCFC)	500-650 °C	$CO_3^{2-} + H_2 \rightarrow CO_2 + H_2O + 2e^-$	$\frac{1}{2}O^{2-} + 2e^- + CO_2 \rightarrow CO_3^{2-}$	Industrial Waste Heat
Direct Methanol Fuel Cell (DMFC)	50-120 °C	$CH_3OH + H_2O \rightarrow 6H^+ + 6e^+ + CO_2$	$\frac{3}{2}O_2 + 6H^+ + 6e^- \rightarrow 3H_2O$	Electronic devices Military applications
Alkaline Fuel Cell (AFC)	60-120 °C	$H_2 + 2(OH) \rightarrow H_2O + 2e^-$	$\frac{1}{2}O_2 + H_2O + 2e^- \rightarrow 2\,(OH)$	Space endeavours
Phosphoric acid fuel cell (PAFC)	160-200 °C	$H_2 \rightarrow 2H^+ + 2e^-$	$\frac{1}{2}O_2 + 2H^- + 2e^- \rightarrow H_2O$	Stationary power
Proton exchange membrane fuel cell (PEMFC)	60-80 °C	$H_2 \rightarrow 2H^+ + 2e^-$	$\frac{1}{2}O_2 + 2H^- + 2e^- \rightarrow H_2O$	Automobiles Stationary power

Among these technologies, PEMFCs are the most extensively deployed due to their low operating temperature, broad fuel availability, and environmentally benign operation, with water as the sole reaction by-product *[16]*. In PEMFCs, the polymer electrolyte membrane plays a critical role by selectively conducting protons from the anode to the cathode while preventing electron crossover. Commonly used membranes include Nafion® (DuPont™), Aciplex®, and Flemion®, which are valued for their high proton conductivity and strong resistance to chemical and mechanical degradation *[17]*. Owing to these advantages, PEMFCs are widely employed in automotive propulsion systems as well as stationary power generation. A schematic representation of the working mechanism and major application areas of PEMFCs is provided in Figure 1.1.

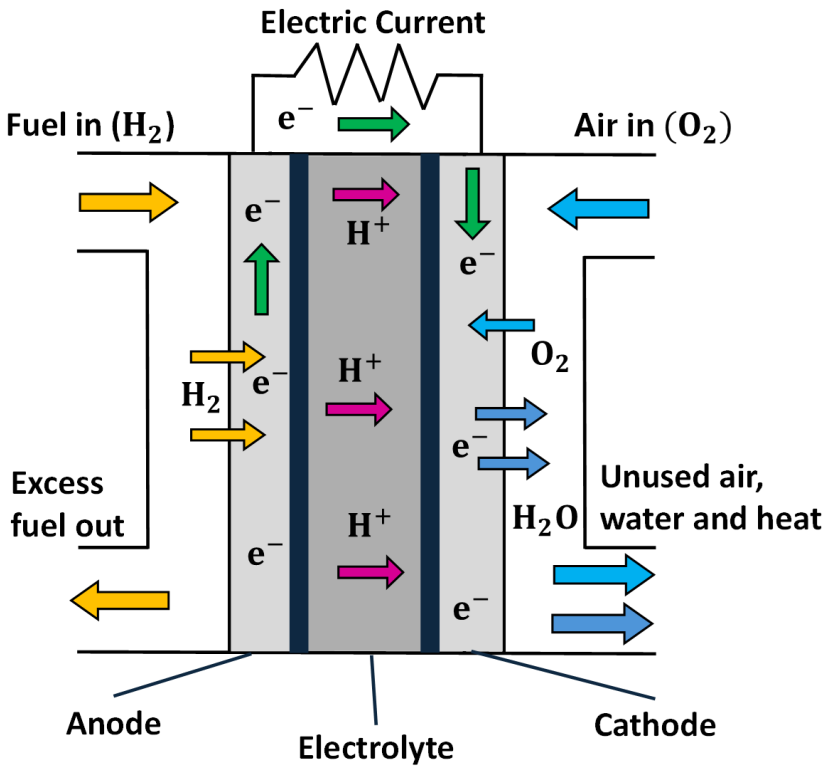


Figure 1.1: Schematic representation of the working mechanism of a proton exchange membrane fuel cells (PEMFCs).

1.3 Nanoclusters Regime: Each Atom Counts

Although the term nanotechnology was formally introduced by Norio Taniguchi in 1974 *[18]*, the foundational idea of manipulating matter at the atomic scale was presented much earlier by Richard Feynman during a talk at the American Physical Society meeting at the California Institute of Technology in 1959, several years before he was awarded the Nobel Prize in Physics. Despite this early conceptualization, experimental realization of nanoscale control became possible only decades later with the invention of the scanning tunnelling microscope (STM) by Gerd Binnig and Heinrich Rohrer at IBM Zurich in 1981. This breakthrough enabled direct access to the nanoscale regime, initiating extensive research across multiple scientific fields, including catalysis. One of the earliest and most influential demonstrations of nanoscale effects in catalysis was reported by Haruta and co-workers in 1987 *[19]*. Their work revealed that gold nanoparticles

exhibit remarkable catalytic activity, despite bulk gold being well known to be chemically inert *[20,21]*. Remarkably, these nanoparticles were shown to catalyze CO oxidation at temperatures as low as −70 °C. This finding highlighted the profound impact of size reduction on material properties, establishing nanocatalysis as a distinct field of research.

When the particle size is further reduced to the subnanometer regime, typically consisting of fewer than 30 atoms, catalytic behavior becomes even more unconventional. In this regime, often referred to as the subnano-cluster regime, every atom plays a crucial role in determining the cluster's overall properties *[22]*. Importantly, trends observed for larger nanoparticles cannot be extrapolated to these ultrasmall systems, rendering this size domain fundamentally non-scalable. As a result, clusters of different sizes exhibit distinct electronic structures, geometries, and reactivities. This size sensitivity has been clearly demonstrated for platinum clusters supported on various substrates. For example, Pt_7 and Pt_8 clusters deposited on α-alumina, despite their similar sizes, show markedly different catalytic activities toward ethylene dehydrogenation *[14]*. In this case, the addition of a single Pt atom significantly suppresses the dehydrogenation activity, illustrating how even minimal changes in cluster size can dramatically alter catalytic performance. Similarly, MgO-supported Pt_n clusters (n=8–15) exhibit size-dependent activity for ethylene hydrogenation that cannot be inferred from trends in larger particles *[23]*. This behavior contrasts sharply with bulk platinum, where different crystal facets display nearly identical turnover frequencies for hydrogenation reactions, indicating structure-insensitive catalysis *[23,24]*.

Further studies have demonstrated that subnano Pt clusters dispersed on SnO/Al_2O_3 can exhibit catalytic activities up to two orders of magnitude higher than conventional Pt or V-based catalysts used in oxidative dehydrogenation of propane *[25]*. Size-selected Pt clusters supported on rutile TiO_2 (110) have also been shown to catalyze CO oxidation *[26]*. Notably, a pronounced enhancement in catalytic activity is observed for Pt_n

clusters at n=8, coinciding with a structural transition from two-dimensional to three-dimensional geometries, as confirmed by STM measurements. This again emphasizes the critical role of individual atoms in determining catalytic behavior at these sizes. The exceptional properties observed in the subnano regime are not limited to platinum clusters. Hutchings and co-workers investigated size-selected gold clusters supported on iron oxide using aberration-corrected scanning transmission electron microscopy (STEM) and found that clusters containing approximately ten Au atoms exhibit the highest activity toward CO oxidation *[27]*. The variation in activity among clusters of different sizes was attributed to differences in cluster morphology, such as monolayer versus bilayer structures on the support, which are directly governed by the number of atoms in the cluster. In another study, Vajda and colleagues demonstrated that Au_6-Au_{10} clusters are highly active for propylene epoxidation, attributing this enhanced activity to the large fraction of undercoordinated Au atoms present in these small clusters *[28]*. Similarly, Pd clusters supported on reduced graphene oxide display unique size-dependent behavior and have been successfully employed for the selective oxidation of alcohols in the subnano regime *[29]*. These studies collectively underscore that the precise number of atoms in a subnano cluster is a determining factor for its electronic structure and catalytic properties. While this pronounced size sensitivity offers a powerful avenue for fine-tuning catalyst activity and selectivity, it also poses significant experimental challenges. Achieving atomically precise size control during catalyst synthesis and maintaining cluster stability against sintering under reaction conditions remain major obstacles. From a theoretical standpoint, subnano clusters challenge conventional computational approaches, as their fluxional nature and the presence of multiple low-energy isomers necessitate ensemble-based descriptions rather than reliance on a single global minimum structure. This dynamic behavior has also highlighted the need for advanced operando experimental

techniques to accurately characterize cluster catalysts under realistic reaction environments.

1.4 Modelling Challenges: Beyond the Global Minima

A central challenge in modelling catalytic reactions lies in identifying the true structure of the catalyst and the nature of its active site under realistic operating conditions. This task is inherently complex and becomes even more demanding for surface-supported cluster catalysts. Elevated temperatures, typically around 300 °C in thermal catalysis, high adsorbate coverage, and the presence of electrolytes and applied potentials in electrocatalysis can substantially alter both the nanocluster and the supporting material *[10,30-32]*. Nanoclusters are intrinsically dynamic, while supports may undergo defect formation or reconstruction under reaction conditions. Consequently, developing computational models that are both physically accurate and computationally feasible is essential *[33]*. Historically, cluster catalysis modelling often relied on a simplified approach in which the lowest-energy structure of an isolated gas-phase cluster was first identified, then placed onto a support and locally optimized. However, studies dating back to the 1970s demonstrated that clusters can undergo pronounced structural rearrangements upon interaction with supports, a phenomenon now widely recognized as strong metal–support interaction (SMSI) *[34]*. Adsorbates further contribute to this restructuring *[35]*. For example, Pt_{13} clusters deposited on CeO_2 experience substantial geometry changes compared to their gas-phase counterparts *[36]*. Even the crystallographic orientation of the support plays a decisive role: *ab initio* molecular dynamics (AIMD) simulations revealed that the CeO_2 (110) surface provides greater stabilization for Pt_{13} than other facets (Figure 1.2). Moreover, when solvent effects were considered, the supported Pt_{13} cluster was found to be more oxidized in aqueous environments than in a vacuum, due to enhanced electron transfer from the cluster to the support.

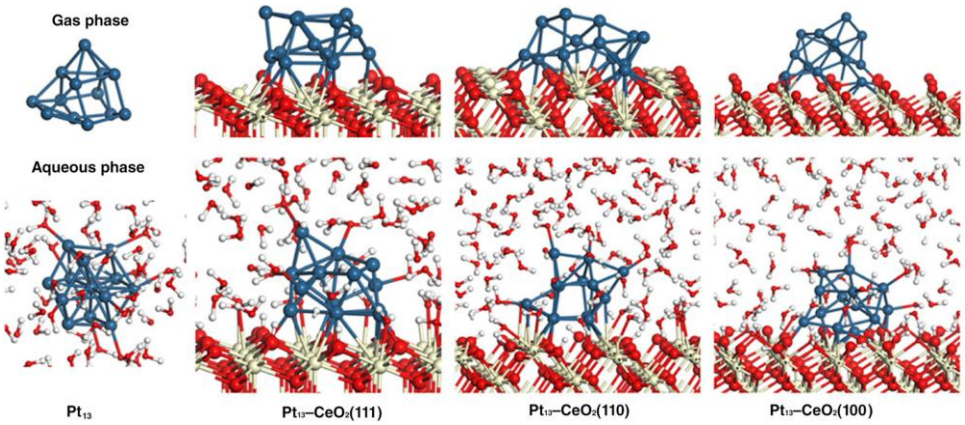


Figure 1.2: Representative optimized structures of free and CeO_2-supported Pt_{13} clusters obtained under vacuum and aqueous conditions, along with configurations on the CeO_2 surface after 10 ps of AIMD simulation. Cerium atoms are shown in light yellow, oxygen in red, platinum in cyan, and hydrogen in white. Adapted and modified with permission from reference 36.

Support-induced restructuring is not limited to noble metal clusters. Keller and co-workers examined VO_4 clusters supported on SiO_4, Nb_2O_5, and ZrO_2 and observed that the classical model, assuming three shared oxygen atoms with the support, does not hold *[37]*. Instead, VO_4 binds through a single oxygen atom on SiO_2, while on Nb_2O_5 and ZrO_2 the cluster adopts increasingly distorted geometries, particularly on ZrO_2. The study further showed that both V–O bond lengths and V–support distances depend sensitively on the surface structure. In contrast, some supports exert minimal influence on cluster geometry. For instance, Pt_{13} clusters maintain similar structures in the gas phase and when supported on graphene *[37]*. These examples highlight that the extent of structural modification depends strongly on the specific cluster–support combination.

Detailed studies of anchoring sites further illustrate this complexity. Sasahara et al. demonstrated that on rutile $TiO_2(110)$, only oxygen vacancy sites can effectively stabilize isolated Pt atoms, accompanied by electron transfer from Pt to the oxide *[38]*. Additionally, doping carbon-based supports with heteroatoms such as N, B, P, or S provides a powerful strategy

to tune cluster properties through modified electronic structure, geometry, and dispersion interactions arising from enhanced metal–support coupling *[39]*. Given this diversity of behaviors, it is generally unsafe to assume that a cluster retains its gas-phase structure upon adsorption.

As a result, computational catalysis shifted toward identifying the most stable structure, or global minimum, of clusters on specific supports and in the presence of adsorbates. These efforts were motivated by the assumption that the global minimum hosts the catalytically active site. However, predicting this structure a priori is rarely possible because chemical bonding in small clusters is highly non-intuitive. Consequently, stochastic global optimization techniques combined with first-principles calculations became essential. These include particle swarm optimization *[40,41]*, random sampling *[42,43]*, genetic algorithms *[44-48]*, basin hopping *[49-51]*, and simulated annealing *[52]*. Due to the high computational cost of these methods, especially for surface-supported clusters, empirical force fields *[53,54]* and potential energy surface fitting approaches *[55-57]* have been employed to accelerate structure searches and enable simulations of larger systems *[58]*.

Global optimization studies have conclusively shown that supported clusters can differ drastically from their gas-phase analogues in geometry, charge state, spin configuration, bonding, and chemical reactivity *[13,14,28,53]*. For instance, whether a cluster is open-shell or closed-shell can strongly influence its interaction with adsorbates and its ability to activate chemical bonds via homolytic pathways, thereby affecting reaction mechanisms and selectivity. Additionally, cluster–adsorbate interactions at higher coverage can alter adsorption modes of subsequent species, with further mechanistic implications *[23,59-61]*.

More recent computational efforts have pushed beyond fixed stoichiometry models by allowing cluster composition to vary during optimization. Under reactive atmospheres, metallic clusters may incorporate heteroatoms such as oxygen, with the extent of uptake depending on temperature, pressure,

and cluster geometry. This introduces a grand-canonical aspect to global optimization. Hensen and co-workers employed basin hopping within a Gibbs ensemble framework (GCMC-DFT) to determine stable structures of Pd_8 clusters on CeO_2 in the presence of O_2, without presupposing the oxide stoichiometry *[50]*. This approach revealed the formation of stable oxidized Pd_8 species on CeO_2 (111).

The gas-phase bicapped octahedral Pd_8 cluster undergoes significant restructuring upon deposition on ceria. Although the support structure itself was assumed to remain static, a necessary approximation for computational feasibility, the oxidation state and geometry of the cluster changed markedly. By comparing Pd_8, Pd_8O_6, and Pd_8O_{12} clusters, the study demonstrated enhanced CO oxidation activity with increasing cluster oxidation. Notably, this work also suggested that linear scaling relations may extend to cluster catalysis, revealing a linear relationship between activation barriers and CO and O binding energies for ceria-supported Pd clusters.

Related work by Janik and colleagues investigated Pd cluster stability during methane activation using grand-canonical Monte Carlo combined with ReaxFF *[62,63]*. Their results showed that Pd_7O_x species with Pd in high oxidation states can be thermodynamically stabilized on ceria, leading to reduced methane activation barriers. These findings align with earlier experimental observations on enhanced oxide formation of Pd clusters on CeO_2 *[64,65]*. Despite these methodological advances, identifying the global minimum structure alone does not provide a complete picture of the catalytically active site. Unlike extended surfaces, clusters do not remain fixed in a single configuration during catalysis. Their composition may evolve with temperature and environment, as demonstrated by Vajda and co-workers, who observed changes in oxygen content of Cu and Pd clusters under reaction conditions *[66]*. Clusters may also undergo partial poisoning *[14,67-72]*, sintering *[73-76]*, or structural rearrangements even at fixed stoichiometry. The concept of structural fluxionality, introduced by

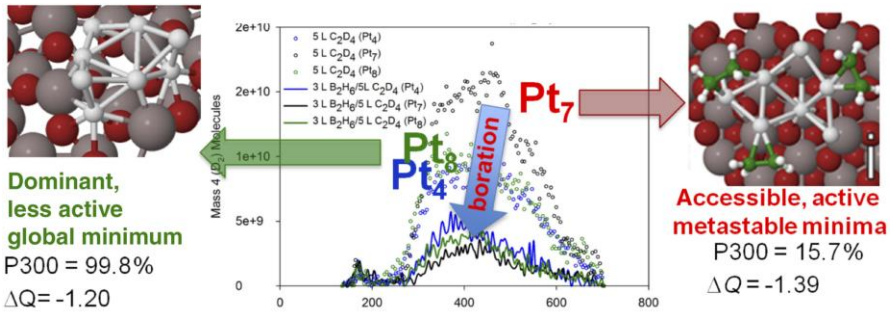
Landman and co-workers, captures this inherent dynamism *[77]*. Their studies on $MgO-$supported Au clusters showed that structural adjustments during O_2 binding and activation enhance catalytic performance in CO oxidation *[78,79]*. Similarly, Fabris and co-workers demonstrated that fluxionality and mobility of Au clusters on stoichiometric and defective CeO_2 (111), particularly near oxygen vacancies, play a critical role in facilitating CO oxidation *[30]*.

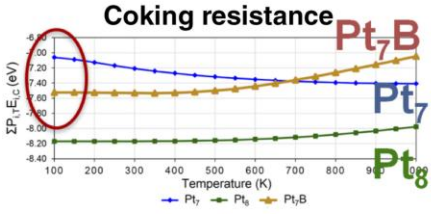
These insights led to a fundamental realization: under catalytic conditions, the active catalyst is not a single structure but an ensemble of thermally accessible states *[10,13,14,80]*. Recent reports have also shown that key properties of cluster-decorated interfaces, including ionization potentials *[55]*, heat capacities *[55]*, poisoning behavior *[13]*, sintering mechanisms *[71,75]*, and catalytic activity and selectivity *[13,14,60]*, differ substantially when described using a statistical ensemble rather than only the global minimum. Metastable, higher-energy structures are often more reactive and can dominate catalytic turnover *[60]*. This challenges the traditional assumption that the global minimum is the most catalytically relevant state, supporting the idea that rare, transient structures may be responsible for most catalytic events. This behavior arises from the characteristically flat potential energy surfaces of nanoclusters, which enable rapid interconversion among numerous isomers on nanosecond timescales at catalytic temperatures *[80]*. Consequently, the catalyst must be viewed as a dynamic population of structures rather than a static entity. This ensemble-based perspective fundamentally reshapes how cluster catalysis should be understood and modelled, motivating a paradigm shift toward statistical mechanical descriptions of catalytic activity.

1.5 Ensemble Representation of Cluster Catalysts

The necessity of describing cluster catalysts using a statistical ensemble framework became evident through combined theoretical and experimental investigations of alkane dehydrogenation on Al_2O_3-supported Pt clusters *[13,14]*. These studies focused on the selective conversion of alkanes to

alkenes, with ethylene serving as a representative intermediate that may either desorb or undergo further dehydrogenation. Despite their close similarity in size, geometry, and charge transfer in their respective global minimum structures, Pt_7 and Pt_8 clusters deposited on alumina exhibit strikingly different catalytic activities in experiments. In particular, Pt_7 shows significantly higher ethylene dehydrogenation activity than Pt_8 (Figure 1.3) *[14]*.





		1 C_2H_4	2 C_2H_4	3 C_2H_4
Pt_7	450 K	7.35%	17.51%	84.26%
	700 K	14.93%	29.42%	89.22%
Pt_7B	450 K	96.41%	2.20%	0.17%
	700 K	89.51%	8.00%	1.70%

Figure 1.3: (Top) Temperature-programmed desorption measurements for the dehydrogenation of deuterated ethylene on size-selected Pt_n clusters supported on Al_2O_3. The observed variation in reactivity is attributed to the availability of highly active metastable configurations in Pt_7, in contrast to Pt_8, which is dominated by a single global minimum structure. P_{300} denotes the population of a specific isomer at 300 K, while ΔQ represents the support-to-cluster charge transfer obtained from Bader analysis. The blue arrow indicates the experimentally observed influence of cluster boration on catalytic performance, where reduced activity corresponds to suppressed coking (copyright). (Bottom) Computational modeling of coking trends captures the experimentally observed decrease in activity and coke formation upon boration, but only at elevated temperatures, where the

cluster ensemble broadens to include metastable states. Adapted and modified with permission from references 13 and 14.

Theoretical analysis revealed that this behavior cannot be explained by considering only the global minimum structures. Instead, enhanced activity of Pt_7 arises from its ability to thermally access a set of higher-energy isomers that adopt single-layer configurations on the support. These metastable structures acquire approximately 0.2 additional electrons from the alumina surface, strengthening ethylene adsorption and activation. At 300 K, the onset temperature for dehydrogenation, roughly 15% of the Pt_7 population occupies these highly reactive states, with their population increasing further at elevated temperatures. In contrast, Pt_8 exhibits a particularly stable and less reactive global minimum, which dominates the ensemble population (≈99%) at 300 K. As a result, its more reactive isomers are not thermally accessible, leading to inferior catalytic performance. This comparison clearly demonstrates that a global-minimum-only description fails to capture the experimentally observed size-dependent activity differences between Pt_7 and Pt_8 clusters. Indeed, earlier theoretical studies based solely on single optimized structures were unable to reproduce this pronounced size sensitivity widely observed in experiments.

Building on these insights, it was theoretically proposed that the selectivity of dehydrogenation reactions could be improved by incorporating boron into Pt clusters. Subsequent experiments confirmed this prediction *[13]*. Computational results showed that the global minimum of Pt_7B supported on Al_2O_3 exhibits a higher affinity for carbon species than pure Pt_7. However, when thermal effects are considered and the ensemble of accessible metastable structures is considered, the carbon-binding behavior changes markedly. At elevated temperatures, the ensemble population of borated clusters shifts toward configurations with reduced coking tendency compared to pure Pt clusters *[58]*. Similarly, Pt_7B clusters display lower ethylene dehydrogenation activity than Pt_7, but this difference only becomes apparent when ensemble effects are considered. Notably, the

global minimum structures of alumina-supported Pt_7 and Pt_7B exhibit comparable reactivity, further emphasizing that catalytic behavior emerges from the ensemble rather than a single structure.

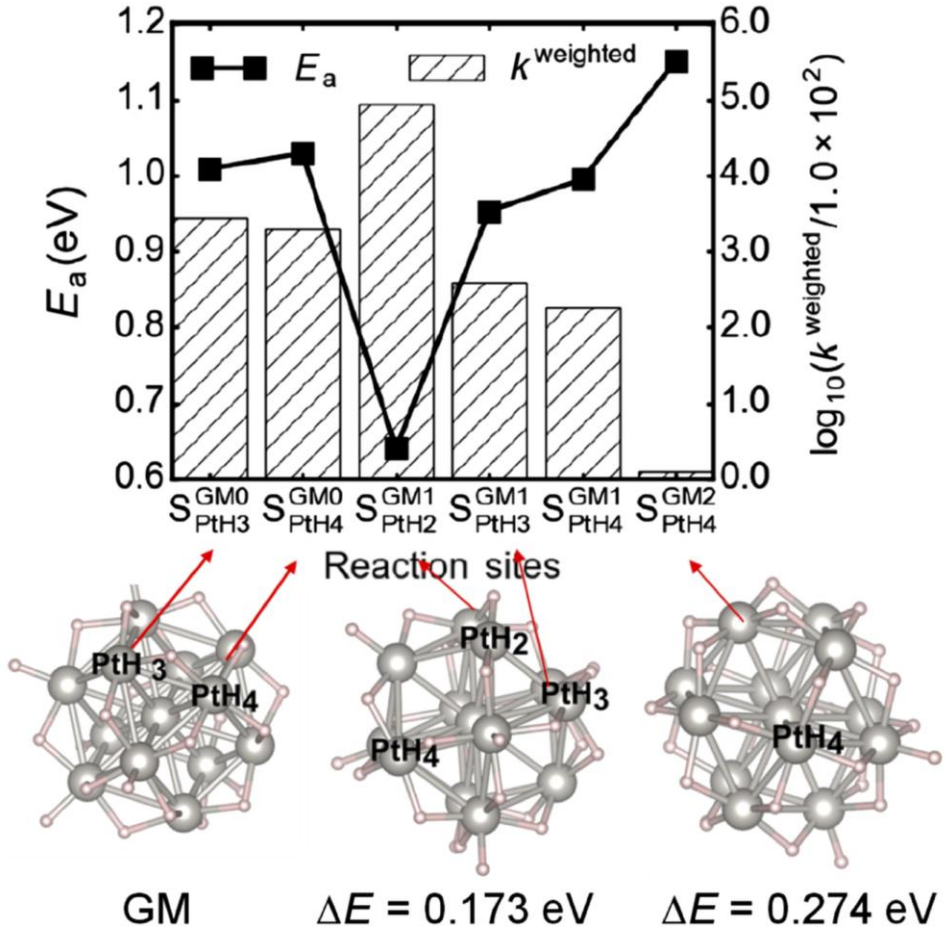


Figure 1.4: Three local minima structures of $Pt_{13}H_{26}$ that can catalyze methane activation. The activation energies (E_a [eV]) and the relative contribution to the reaction rate (rectangle bars) are shown for these three structures. The weighted reaction rate constant is given with a reference of 1×10^2 at 400°C. Adapted and modified with permission from reference 81.

Additional evidence supporting the ensemble-based view comes from a purely theoretical study by Sun and Sautet, who examined hydrogen evolution and methane activation on hydrogen-covered Pt_{13} clusters (gas-phase $Pt_{13}H_{26}$) *[81]*. Their results showed that the second-lowest-energy isomer, rather than the global minimum, possesses the highest activity for

methane activation and dominates the overall catalytic response despite representing a minority fraction of the population (Figure 1.4). Collectively, these findings underscore the importance of evaluating observable properties of cluster-based catalysts as ensemble averages, a concept consistently emphasized by Alexandrova and co-workers *[10,13,14,55,60,71,72,75]*. In this framework, all energetically accessible structures contribute to experimentally measurable properties, and the catalyst is no longer described by a single optimized geometry but by a dynamic collection of interconverting isomers. At finite temperatures, ensemble-averaged properties, such as energy, can be expressed as a weighted sum over all contributing structures:

$$\langle E \rangle = \sum_i E_i P_i \tag{1.1}$$

where the probability P_i of each isomer depends on its electronic, vibrational, and rotational partition functions:

$$P_i = \frac{Z_{elec,i} Z_{vib,i} Z_{rot,i}}{Z} \tag{1.2}$$

The translational contribution can typically be neglected, as it remains nearly identical for all isomers. Incorporating Boltzmann weighting is particularly crucial for capturing kinetic behavior, as it accounts for the influence of thermally populated metastable states on reaction pathways and rates, yielding a more realistic description of catalytic performance. In summary, the active state of a cluster catalyst under operating conditions is best described as a thermally populated ensemble of dynamically interconverting structures rather than a single static geometry. The number of such accessible minima can be substantial, often reaching several tens for small clusters *[80]*, and it is generally impossible to predict in advance which structures will dominate a given catalytic property. Therefore, comprehensive sampling aimed at identifying all thermally relevant minima, rather than only the global minimum, is essential for accurately modeling and understanding cluster catalysis.

1.6 Fluxionality and the Ensemble Nature of the Active Catalysts

As catalytic reactions proceed, cluster geometries can adjust in response to adsorbed intermediates, enabling structural rearrangements to occur at various stages along the reaction pathway. Consequently, the ensemble of accessible cluster configurations is not static but evolves continuously, with each minimum on the reaction free-energy profile corresponding to a distinct equilibrium ensemble. This situation raises two fundamental mechanistic questions. First, does cluster rearrangement occur simultaneously with the elementary reaction step, or does it take place before or after it? Second, does the system have sufficient time to reach equilibrium within each reaction intermediate well, or does it retain a degree of structural memory and evolve under non-equilibrium conditions? Each of these scenarios has major implications for how cluster catalysis should be understood and theoretically described. In any case, neglecting cluster dynamics would lead to an incomplete and potentially misleading description of the catalytic process.

Whether cluster rearrangement is dynamically coupled to the reaction step depends on the relative magnitude of the reaction barrier and the barrier associated with cluster reorganization. If these barriers differ significantly, the likelihood of strong coupling is low. However, when the barriers are comparable, it becomes difficult to determine a priori whether such coupling occurs *[80,82]*. Furthermore, some local minima may be kinetically inaccessible or trapped due to large reorganization barriers. Therefore, a comprehensive description of cluster catalysis must incorporate not only thermodynamics but also kinetics and explicit dynamical considerations to properly account for competing timescales and potential deviations from equilibrium behavior.

Quantifying cluster fluxionality is therefore an important step toward understanding catalytic behavior. Yang and co-workers introduced a measure of structural flexibility, which they referred to as fluxionality, based on the displacement of atoms within the cluster upon adsorption *[83]*.

As shown in Figure 1.5, Pt_6 clusters supported on $TiO_2(101)$ undergo substantial geometric rearrangements upon CO_2 adsorption. Their study revealed a linear correlation between atomic displacement and CO_2 adsorption energy across Pt_n clusters (n = 4, 6, 8) supported on pristine and partially reduced anatase $TiO_2(101)$. For Pt-only sites, similar levels of flexibility were observed across three-dimensional, quasi-three-dimensional, and planar two-dimensional cluster geometries. In contrast, at interfacial edge sites, three-dimensional clusters exhibited the largest atomic displacements, followed by intermediate structures and then planar clusters. This indicates that three-dimensional geometries undergo more pronounced structural changes during CO_2 adsorption compared to two-dimensional ones. A positive correlation between structural flexibility and binding strength was also observed, suggesting that clusters capable of greater geometric adaptation achieve improved orbital overlap with adsorbates, leading to stronger binding. While this stabilization through relaxation is expected, it should be noted that adsorption-induced reconstruction differs from fluxionality involving isomer interconversion, which requires overcoming energy barriers.

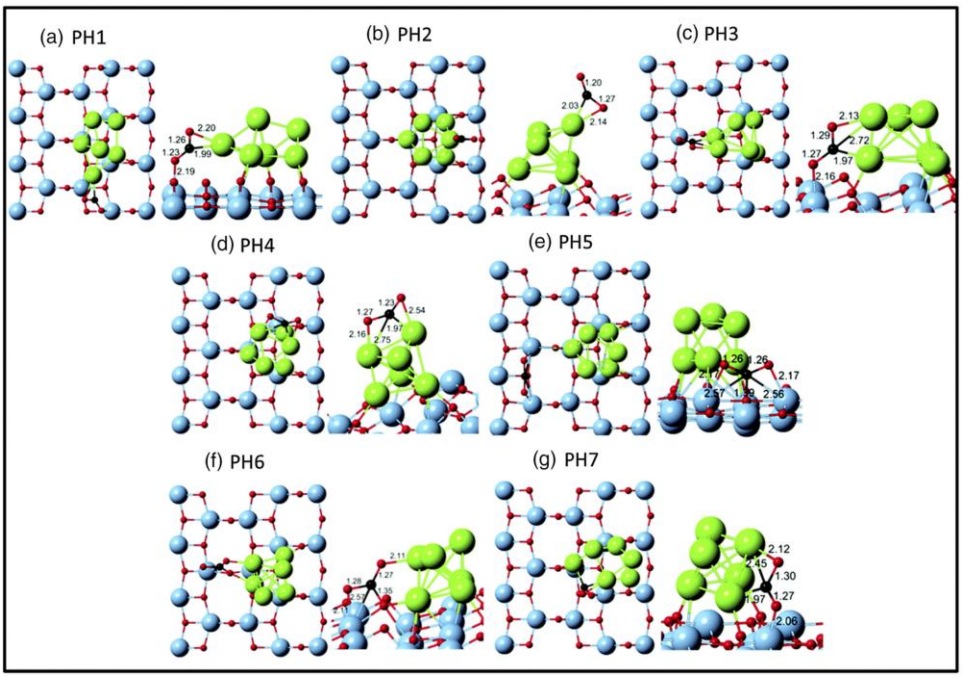


Figure 1.5: Side and top views of energetically stable CO_2 adsorption geometries on the pristine anatase TiO_2(101) surface decorated with Pt hexamers (PH). Oxygen atoms are shown in red, carbon in black, titanium in blue, and platinum in green. Labeled values correspond to interatomic distances in Å. Adapted and modified with permission from reference 83.

Similar conclusions were reached in a comparative study of gold cluster isomers in the gas phase and when supported on partially reduced anatase TiO_2 (101) *[84]*. In that work, the icosahedral isomer was found to bind CO_2 more strongly due to its greater ability to undergo structural rearrangements that optimize adsorption geometry. This observation is consistent with the higher flexibility predicted for the icosahedral structure using the fluxionality metric proposed by Yang et al. *[83]*. If cluster rearrangements occur concurrently with reaction dynamics, then the conventional definition of the reaction coordinate must be extended to explicitly include cluster motion. Such a scenario would significantly complicate catalytic simulations. To date, most studies identify transition states by allowing only the reacting species to move while the catalyst remains largely stationary or undergoes limited relaxation. The system illustrated in Figure 1.6 exemplifies this approach. However, transition states involving substantial cluster rearrangement may remain undiscovered because the appropriate collective motions of cluster atoms are not known in advance. That said, there are strong arguments suggesting that cluster isomerization and reaction dynamics are often decoupled. Because metal atoms are relatively heavy, large-scale cluster rearrangements are expected to occur on slower timescales than the motion of lighter reacting species. Indeed, molecular dynamics studies conducted so far show no clear evidence of strong coupling between reaction events and cluster isomerization. Nevertheless, conventional molecular dynamics may still fail to capture low-probability but kinetically relevant pathways involving cluster rearrangements. Developing robust methods to identify such pathways remains an active area of research. Even if cluster dynamics are not directly coupled to

reaction steps, an important unresolved question is whether cluster ensembles can fully or partially re-equilibrate during the lifetime of reaction intermediates. If equilibration does occur, an important conclusion follows: thermodynamics and kinetics may be governed by different catalyst states. While reaction intermediates may preferentially bind to the global minimum structures, the kinetics could be dominated by rare, highly reactive metastable states. This perspective departs fundamentally from the traditional view of catalysis, which has historically assumed a single dominant active structure.

Interestingly, this complexity offers new opportunities rather than limitations. Because different intermediate and transition states interact with different catalyst configurations, it becomes possible to disrupt linear scaling relationships in catalysis. Such scaling relations, which link the binding energies of chemically related intermediates, are well known in processes such as ammonia synthesis (e.g., NH versus NH_2) and the oxygen reduction reaction (e.g., O, OH, and OOH) *[85-87]*. While scaling relations simplify catalyst design, they also impose intrinsic performance limits. In a recent study, we demonstrated that small Pt clusters can deviate significantly from linear scaling behavior, thereby breaking conventional correlations and enabling enhanced catalytic performance *[12]*. For Pt clusters of varying sizes, both in the gas phase and supported on graphene, substantial structural rearrangements occur upon binding ORR intermediates, accompanied by changes in adsorption sites (atop versus bridge) and coverage-dependent effects. As a result, poor correlations were observed between adsorption energies, as illustrated in Figure 1.6. These findings suggest that cluster dynamics can serve as a powerful design parameter for overcoming scaling limitations, for example in the development of more efficient and cost-effective catalysts for ammonia synthesis.

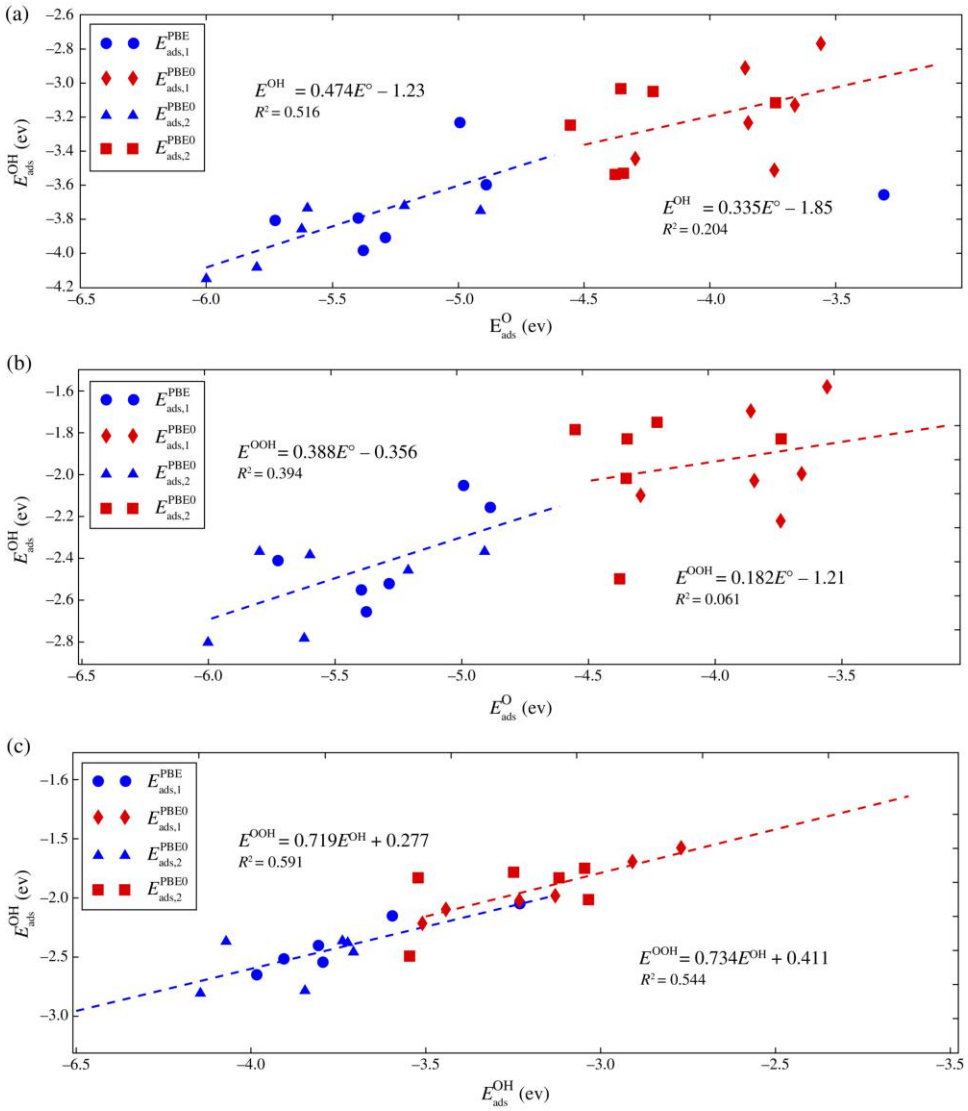


Figure 1.6: Calculated relationships between the adsorption energies of (a) O–OH, (b) O–OOH, and (c) OH–OOH on gas-phase Pt_n clusters (n = 1–6), considering cases with one or two co-adsorbed species. Results obtained using the PBE functional are shown in blue, while PBE0 data are shown in red. Weak correlations arise from structural rearrangements of the clusters and changes in preferred adsorption sites. Adapted and modified with permission from reference 12.

More broadly, cluster dynamics can be viewed as a general lever in catalyst design. Structural flexibility enables stabilization of adsorbate-bound states, leading to deeper energy minima and lower transition-state barriers. As a result, highly fluxional clusters can transform otherwise inactive materials

into effective catalysts. This flexibility also provides a means to tune catalytic selectivity, as demonstrated in our forthcoming work. Finally, these concepts extend beyond metallic clusters. While clusters with more directional bonding, such as metal oxides, may exhibit reduced flexibility, ensemble effects remain relevant even if the number of accessible isomers is substantially smaller. Ensemble behavior is also expected to be important in catalytic processes occurring at lower temperatures, such as electrocatalysis or oxidative dehydrogenation assisted by oxygen. More generally, ensemble effects are a defining feature of highly dynamic catalytic interfaces. For example, hexagonal boron nitride under oxidative dehydrogenation conditions develops a highly fluxional, under-stoichiometric $BO_x(OH)_\gamma$ overlayer *[88,89]*. Solid-state NMR measurements indicate that most of the interface consists of inactive boron oxide, implying that minority species are responsible for the observed high selectivity toward propene formation *[90]*. Similar behavior has been reported for $Rh_{0.5}Pd_{0.5}$ nanoparticles, which undergo reversible changes in composition and chemical state in response to oxidizing and reducing environments *[91]*. Even in relatively rigid systems, ensemble effects may arise from different configurations of co-adsorbates surrounding active sites, as observed by Xu and co-workers *[92]*. These examples collectively highlight the necessity of incorporating fluxionality and structural diversity into theoretical models of catalytic interfaces.

1.7 Theoretical Methodology

Density functional theory (DFT) has become a cornerstone of modern computational modelling across disciplines such as chemistry, physics, materials science, biology, and engineering. Owing to its favourable balance between predictive accuracy and computational cost, DFT is particularly well-suited for probing ground-state electronic properties and atomic-scale phenomena. This chapter outlines the fundamental principles and key theoretical foundations of DFT that are relevant to the present thesis. In addition, the specific DFT methodologies employed for atomic-

level simulations and electronic structure calculations are briefly introduced, with emphasis on their application to fuel cell applications in Vienna *ab initio* simulation package.

1.7.1 Schrödinger Equation for Many-Body Problem

The many-body problem represents a central challenge in both physics and chemistry, as it concerns systems composed of multiple interacting particles whose collective behavior determines observable material properties. In atomic, molecular, and solid-state systems, these particles are primarily electrons, and their mutual interactions play a decisive role in defining chemical bonding, stability, and functionality. As a result, a fundamental understanding of materials relies on an accurate description of their electronic structure, particularly the ground-state configuration. The intrinsic difficulty of the many-body problem stems from the fact that every particle interacts with all others simultaneously. This leads to a rapid increase in computational complexity as the system size grows, making exact solutions impractical for all but the simplest cases. From a quantum mechanical perspective, the behavior of a many-body system is governed by the Schrödinger equation. For a system containing N electrons and M nuclei, the time-independent Schrödinger equation can be written as:

$$\hat{H}\Psi(r_1, r_2, \dots, r_N, R_1, R_2, \dots, R_M) = E\Psi(r_1, r_2, \dots, r_N, R_1, R_2, \dots, R_M) \quad (1.3)$$

where $\Psi(r_1, r_2, r_3, \dots, r_N, R_1, R_2, R_3, \dots, R_M)$ denotes the many-electron wavefunction, which depends on the spatial coordinates of all electrons (r_i) and the nuclei R_I, E is the energy of the system. The operator $\hat{H}$ is the Hamiltonian operator of the system, and encompasses both kinetic and potential energy contributions arising from electrons and nuclei, as detailed in equation 1.4:

$$\hat{H} = \underbrace{-\frac{\hbar^2}{2m_e}\sum_i \nabla_i^2}_{\hat{T}_e} \underbrace{-\frac{\hbar^2}{2}\sum_I \frac{\nabla_I^2}{M_I}}_{\hat{T}_n} + \underbrace{\underbrace{\frac{1}{2}\sum_{i\neq j}\frac{e^2}{4\pi\varepsilon_0 r_{ij}}}_{\hat{V}_{ee}} \underbrace{-\sum_{i,I}\frac{e^2 Z_i}{4\pi\varepsilon_0 R_{iI}}}_{\hat{V}_{en}} + \underbrace{\frac{1}{2}\sum_{I\neq J}\frac{e^2 Z_I Z_J}{4\pi\varepsilon_0 R_{IJ}}}_{\hat{V}_{nn}}}_{\hat{V}} \quad (1.4)$$

where m_e and M_I are the mass of the i^{th} electron and I^{th} nuclei, respectively, while e represents the elementary charge and Z_I is the atomic number of the I^{th} atom. In equation (1.4), the first two terms ($\hat{T}_e$ and $\hat{T}_n$) account for the kinetic energy contributions of electrons and nuclei, respectively. The remaining terms, $\hat{V}_{ee}$, $\hat{V}_{en}$, and $\hat{V}_{nn}$, describe the Coulombic interactions arising from electron-electron repulsion, electron-nucleus attraction, and nucleus-nucleus repulsion, respectively.

1.7.2 Born-Oppenheimer Approximation

The full many-electron Hamiltonian in Equation 1.4 is mathematically intractable for all but the simplest systems, such as hydrogen atoms or small molecules. When extended to larger systems, including solids, direct solutions become computationally impractical, making the use of suitable approximations unavoidable. A key physical insight enabling such simplifications arises from the large disparity between nuclear and electronic masses, with a proton being approximately 1836 times heavier than an electron. As a result, nuclear motion occurs on a much slower timescale than electronic problems. This separation of electronic and nuclear dynamics forms the basis of the Born-Oppenheimer approximation *[93]*. To provide an intuitive picture, Prof. Rod Bartlett likens this separation to flies moving around a wedding cake: the nuclei resemble the cake, while the electrons behave like flies that rapidly adjust their motion as the cake itself moves (Figure 1.7).

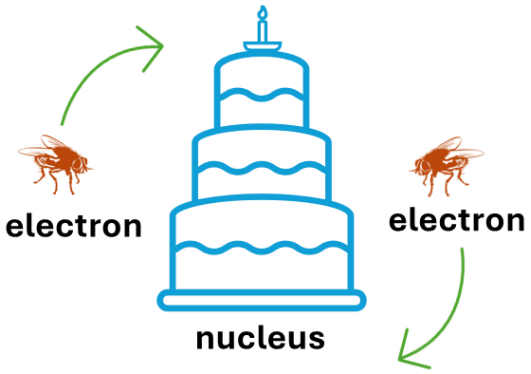


Figure 1.7: Prof. Rod Bartlett's wedding cake (nuclei) and flies (electrons) analogy to explain the Born-Oppenheimer approximation.

From a physical standpoint, this approximation implies that electrons respond almost instantaneously to changes in nuclear positions. Consequently, nuclei may be treated as classical particles that remain effectively fixed with respect to the much faster quantum motion of electrons. Under this assumption, the total wavefunction of the system can be factorized into independent electronic and nuclear components,

$$\Psi_{\text{total}} = \Psi_{\text{electronic}}\,\Psi_{\text{nuclear}} \tag{1.5}$$

Accordingly, many-body wavefunction in equation (1.1) can be expressed as a product of electronic and nuclear wavefunctions, as shown in equation 1.6.

$$\Psi(r_1, r_2, \dots, r_i, R_1, R_2, \dots, R_I) = \Psi(r_1, r_2, \dots, r_i)\Psi(R_1, R_2, \dots, R_I) \tag{1.6}$$

This separation enables the Hamiltonian operator to be decomposed into electronic and nuclear contributions. The nuclear Hamiltonian may be written as:

$$\hat{H}_{\text{nucl}} = \hat{T}_{\text{nucl}} + \hat{U}_{\text{nucl}} \tag{1.7}$$

where the nuclear kinetic energy term can be neglected due to the substantially larger mass of nuclei compared to electrons. The remaining contribution corresponds to classical electrostatic ion-ion interactions, which can be effectively evaluated using methods such as Ewald summation. The electronic Hamiltonian of the many-body system is given by:

$$\hat{H}_{\text{el}} = \hat{T}_e + \hat{V}_{\text{ext}} + \hat{U}_{ee} \tag{1.8}$$

Under the Born–Oppenheimer approximation, the total Hamiltonian can therefore be rewritten as:

$$\hat{H} = \underbrace{-\frac{\hbar^2}{2m_e}\sum_i \nabla_i^2}_{\hat{T}_e} + \underbrace{\underbrace{\frac{1}{2}\sum_{i\neq j}\frac{e^2}{4\pi\varepsilon_0 r_{ij}}}_{\hat{V}_{ee}} - \underbrace{\sum_{i,I}\frac{e^2 Z_i}{4\pi\varepsilon_0 R_{iI}}}_{\hat{V}_{en}}}_{\hat{V}} \tag{1.9}$$

where the terms represent the electronic kinetic energy, electron-electron repulsion, and electron-nucleus attraction, respectively.

The Born–Oppenheimer approximation thus leads to a substantial reduction in the dimensionality of the original problem by decoupling nuclear and

electronic degrees of freedom. Nevertheless, even with this simplification, the Schrödinger equation remains difficult to solve because the electronic wavefunction still depends on the coordinates of all electrons, resulting in a problem with 3N variables arising from electron-electron interaction. Since the number of electrons in realistic systems is typically large, a direct wavefunction-based solution remains computationally prohibitive. To overcome this challenge, it is more practical to shift the focus from individual electron coordinates to the electron density. This perspective forms the foundation of density functional theory (DFT), which reformulates the many-body problem in terms of the electron density rather than the full many-electron wavefunction. In the following section, the DFT framework is introduced as an efficient and conceptually powerful alternative for studying complex many-electron systems.

1.7.3 Density Functional Theory

Density functional theory (DFT) is a quantum mechanical framework for treating interacting electron systems, where the electron density is the fundamental variable, rather than the many-electron wavefunction. This represents a conceptual departure from traditional wavefunction-based approaches. By formulating the problem in terms of electron density, the complexity of the system is drastically reduced, as the number of variables scales from 3N for an N-electron wavefunction to just three spatial coordinates. Figure 1.8 schematically contrasts the many-body wavefunction description with the density-based perspective adopted in DFT.

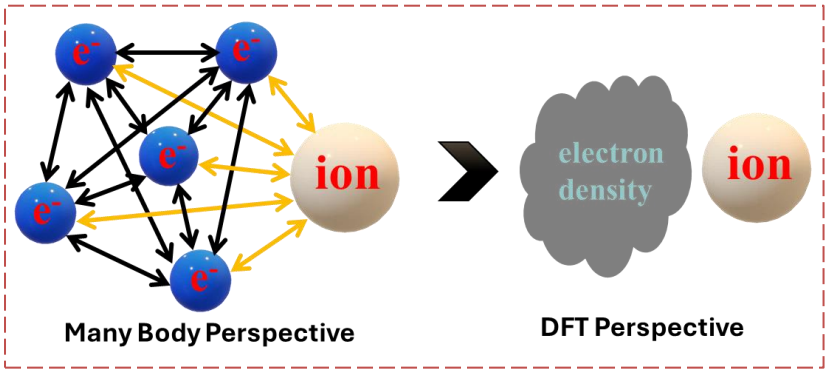


Figure 1.8: Comparative illustration of many-body and DFT perspectives.

The fixed dimensionality of electron density greatly simplifies the underlying equations, enabling DFT to be applied efficiently to systems comprising hundreds or even thousands of atoms. This balance between accuracy and computational feasibility has established DFT as the dominant theoretical tool in condensed matter physics. In practice, DFT is primarily used to evaluate the ground-state properties of many-body systems, including isolated atoms, molecules, and extended solids. The origins of DFT can be traced back to the Thomas–Fermi model, in which quantum systems were described using a uniform, non-interacting electron density *[94,95]*. While this early approach captured some qualitative features, it lacked the ability to properly account for electron–electron interactions and therefore showed limited accuracy for real materials. A rigorous foundation for modern DFT was later provided by Kohn and Hohenberg through the Hohenberg–Kohn theorems, which formally established electron density as a complete descriptor of the ground state in interacting many-body systems. Building on this framework, the practical implementation known as Kohn–Sham DFT was subsequently developed, forming the basis of most contemporary DFT calculations *[96]*.

1.7.3.1. The Hohenberg-Kohn Theorem

In 1964, Hohenberg and Kohn laid the theoretical groundwork for density functional theory by demonstrating that the properties of an interacting electronic system can be uniquely determined from the electron density. Their formulation introduced two key theorems, both of which are based on the spatial distribution of electron density $\rho(\mathbf{r})$ in the presence of an external potential.

Theorem I: "For a system of interacting particles in an external potential $\hat{V}(\mathbf{r})$ can be uniquely determined by its ground-state electron density $\rho(\mathbf{r})$ within an additive constant."

This implies that the Hamiltonian, and thus all ground-state properties, can be described entirely by the electron density $\rho(\mathrm{r})$.

Theorem II: "A universal functional for the Energy (E[n]) can be described in terms of electron density $\rho(\mathbf{r})$ which could be defined for any external potential (V_{ext}). For any particular potential V_{ext}, the exact ground state energy is the global minima of the energy functional, and the electron density $\rho(\mathbf{r})$ that minimizes the function is the exact ground-state electron density $\rho_0(\mathbf{r})$."

Here, $\rho_0(\mathbf{r})$ uniquely determines the external potential of the system, which leads to the ground state wavefunction.

By combining the two Hohenberg–Kohn theorems, the ground state total energy can be expressed as a functional of the electron density, as shown in equation 1.10:

$$E_{HK}[\rho] = F_{HK}[\rho] + \int d^3r V_{ext}(r)\rho(r) \tag{1.10}$$

where F_{HK} includes all the internal energies, including kinetic and potential energies of the interacting system. Hence, the total internal energy functional can be expressed as given below in equation 1.11:

$$E_{HK}[\rho] = T[\rho] + E_{int}[\rho] \tag{1.11}$$

as given in Equation 1.6, this functional is independent of the external potential V_{ext} and depends only on the electron density $\rho(r)$. The Hohenberg–Kohn theorems do not specify the explicit mathematical form of this functional, which makes practical calculations reliant on suitable approximations. The exact ground-state energy E_0 of the system corresponds to the global minimum of the energy functional equation (1.11). The electron density $\rho_0(r)$ that achieves this minimum represents the true ground-state density. This condition follows from the variational principle and is expressed as:

$$\frac{\delta}{\delta n} E_{HK}[\rho(r)]|_{\rho=\rho_0} = 0 \tag{1.12}$$

Overall, the Hohenberg–Kohn theorems provide the conceptual foundation of modern density functional theory by establishing a one-to-one correspondence between the ground-state electron density and the ground-state energy. However, while they guarantee the existence of such a

functional, they do not offer a practical prescription for constructing it. To overcome this limitation, Kohn and Sham later introduced a practical framework that treats exchange and correlation effects approximately, enabling the accurate calculation of ground-state densities for interacting electron systems.

1.7.3.2. Kohn-Sham Formalism

In 1965, Kohn and Sham introduced a key idea showing that the true ground-state electron density of an interacting many-electron system can be exactly reproduced by a fictitious system of non-interacting particles. This auxiliary system is constructed such that its ground-state density matches that of the original interacting system, as schematically depicted in Figure 1.9.

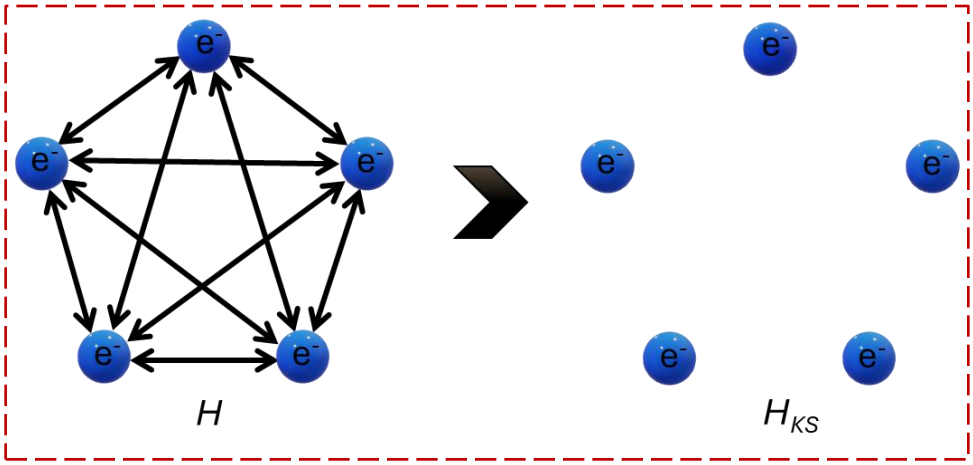


Figure 1.9: Illustration of interacting and non-interacting many-electron systems having the same ground-state electron density.

A central refinement introduced through the Hohenberg–Kohn framework is the substitution of the fully interacting electron problem with an equivalent auxiliary system of non-interacting particles that reproduces the same electron density. Within this Kohn–Sham (K–S) formulation, the total energy is expressed as a functional of the electron density, as shown in equation (1.13):

$$E_{K-S}[n] = T_S[n] + \int d^3\, r V_{ext}(r) n(r) + \frac{1}{2} \int d^3\, r d^3 r' \frac{n(r) n(r')}{|r - r'|} + E_{xc}[n] \quad (1.13)$$

Here, the first term corresponds to the kinetic energy of a fictitious system of non-interacting electrons, the second term accounts for the interaction of electrons with the external potential arising from nuclei and other applied fields, and the third term represents the classical electron–electron Coulomb interaction, commonly referred to as the Hartree energy. All remaining many-body contributions, including exchange and correlation effects beyond the classical description, are incorporated into the exchange-correlation functional $E_{xc}[n]$. The explicit analytic form of $E_{xc}[n]$ is not known, and the practical success of the Kohn-Sham approach relies on the development of accurate approximations for this term. In addition, the Coulomb repulsion between nuclei is treated separately and added as a constant contribution to total energy. According to the second Hohenberg–Kohn theorem, the ground state of the auxiliary Kohn–Sham system is obtained by minimizing the Kohn–Sham energy functional with respect to the electron density [n(r)]. This variational procedure leads to a set of Schrödinger-like equations for the Kohn–Sham orbitals, given in equation 1.14:

$$\widehat{H}_{K-S}\emptyset_i(r) = \left[-\frac{\hbar^2}{2m_e}\nabla^2 + V_{K-S}(r)\right]\emptyset_i(r) = \varepsilon_i\emptyset_i(r) \qquad (1.14)$$

where $\emptyset_i(r)$ denotes the K-S orbital, ε_i are their corresponding eigenvalues, and $V_{K-S}(r)$ is the effective single-particle potential. The Kohn–Sham potential is defined as:

$$V_{K-S}(r) = V_{ext} + \int d^3r' \frac{n(r)}{|r-r'|} + V_{xc} \qquad (1.15)$$

where the exchange-correlation potential V_{xc} is obtained as the functional derivative of the exchange-correlation energy with respect to the density:

$$V_{xc} = \frac{\delta E_{xc}[n]}{\delta n(r)} \qquad (1.16)$$

It is important to emphasize that the Kohn–Sham orbitals $\emptyset_i(r)$ are not true many-electron wavefunctions and do not have a direct physical interpretation as such. Their primary role is to reproduce the correct ground-state electron density, which is constructed from these orbitals as:

$$\rho(r) = \sum_i |\emptyset_i^{KS}(r)|^2 \qquad (1.17)$$

This formalism enables the practical calculation of ground-state properties of interacting electron systems while retaining the exact electron density in principle.

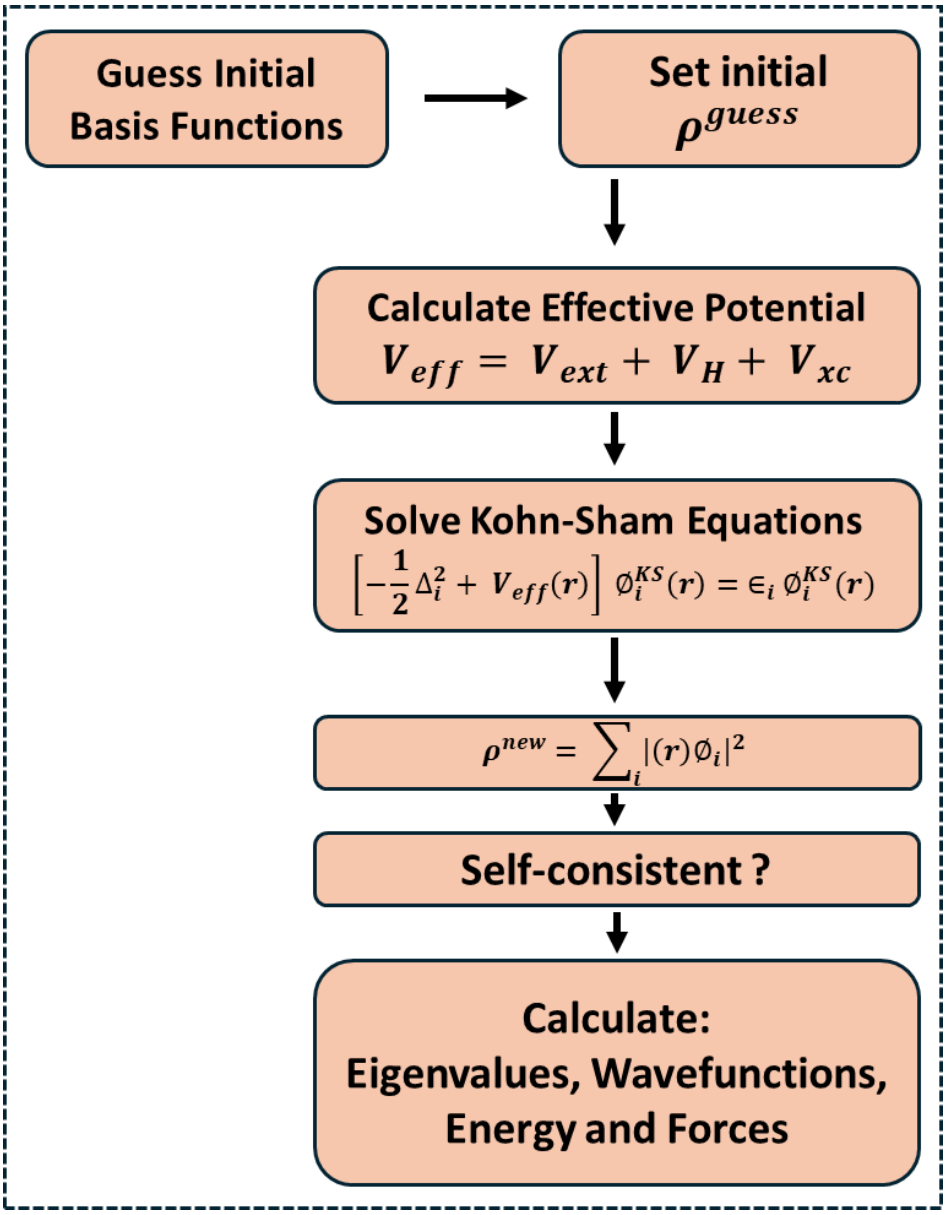


Figure 1.10: Flow chart of the self-consistent loop for the solution of Kohn-Sham equations.

To obtain the ground-state electron density, the Kohn–Sham equations are solved using a self-consistent iterative procedure, as schematically shown in Figure 1.10. The process begins with an initial trial electron density $\rho(r)$. Based on this density, the effective potential $V_{eff}(r)$ is constructed. The

Kohn–Sham equations are then solved within this potential to generate a set of Kohn–Sham orbitals, $\emptyset_i^{KS}(r)$. These orbitals are subsequently used to compute an updated electron density. The newly obtained density serves as the input for the next iteration, and this cycle is repeated until self-consistency is achieved, and this cycle is repeated until self-consistency is achieved, meaning that the input and output densities differ by less than a predefined threshold.

1.7.4. Exchange-Correlation Functionals

Since the exact analytical form of the exchange-correlation energy is not known, a variety of approximate exchange-correlation (xc) functionals have been developed. These approximations aim to balance computational cost with predictive accuracy. In practice, the reliability of any DFT calculation is strongly governed by the quality of the chosen E_{xc} functional.

1.7.4.1. Local Density Approximation (LDA)

The earliest and most straightforward exchange-correlation functional within the Kohn–Sham framework is the local density approximation (LDA) *[96]*. In LDA, the exchange-correlation energy at each point in space is assumed to depend only on the local value of the electron density, treating the system as if it were a uniform electron gas at that density. Under this assumption, the exchange-correlation energy is expressed as:

$$E_{xc}^{LDA} = \int n(\mathbf{r})\varepsilon_{xc}^{uni}[n(\mathbf{r})]d\mathbf{r} \quad (1.18)$$

Here, ε_{xc}^{uni} denotes the exchange-correlation energy per electron of a homogeneous electron gas evaluated at the local density $n(\mathbf{r})$. This quantity can be separated into exchange (ε_x) and correlation (ε_c) contributions. While the exchange term admits an analytical expression, the correlation part cannot be derived exactly and is instead obtained from extensive quantum Monte Carlo simulations combined with suitable interpolation schemes *[97]*. Despite its simplicity, LDA has been shown to perform remarkably well for a wide range of model systems. It was originally expected to be most accurate for systems with slowly varying electron densities, such as nearly free-energy metals *[98]*. However, subsequent

studies demonstrated that LDA can also provide reasonable predictions for semiconductors and insulators, yielding experimentally relevant structural and energetic properties *[99,100]*. Nevertheless, LDA has well-known limitations. It typically overestimates cohesive and binding energies, tends to overly delocalize electrons in real space, which is acceptable for systems dominated by s and p electrons but problematic for materials involving localized d and f states, and fails to capture long-range interactions such as van der Waals forces due to its inherently local character.

1.7.4.2. Generalized-Gradient Approximation (GGA)

The extension beyond the local density approximation (LDA) was motivated by the Hohenberg–Kohn framework *[96]*, in which the exchange–correlation energy per particle is described not only by the local electron density but also by its spatial variation. In this approach, the exchange–correlation energy density depends explicitly on the electron density $n(r)$ and its gradient $\nabla n(r)$. Within the generalized gradient approximation (GGA), the exchange–correlation energy is obtained by modifying the LDA energy density and can be expressed as:

$$E_{xc}^{GGA} = \int n(\mathbf{r})\varepsilon_{xc}^{GGA}[n(\mathbf{r}), \nabla n(\mathbf{r})]d\mathbf{r} \quad (1.19)$$

$$E_{xc}^{GGA} = \int \varepsilon_x^{uni}\, n(\mathbf{r})F_{xc}[n(\mathbf{r}), \nabla n(\mathbf{r})]n(\mathbf{r})d\mathbf{r}, \quad (1.20)$$

Here, ε_x^{uni} denotes the exchange energy density of a homogeneous electron gas with density $n(r)$, while F_{xc} is a dimensionless enhancement factor that depends on the local density and its gradient *[101,102]*. This enhancement factor is commonly separated into exchange and correlation contributions. Several functional forms have been proposed for the exchange component, including Perdew-Wang (PW92) *[103]*, Becke (B88) *[104]*, B95 *[105]*, Lee, Yang, and Parr (LYP) *[106,107]*, and Perdew-Burke-Ernzerhof (PBE) *[108]*, all of which are widely employed in practical calculations.

A brief comparison between the LDA and GGA formalisms highlights their respective strengths and weaknesses. While LDA successfully reproduces many fundamental electronic properties, it often performs poorly in

predicting accurate chemical properties, particularly in determining binding energies *[105]*. In contrast, GGA typically yields lower exchange–correlation energies, leading to reduced binding energies and an overall improvement over LDA predictions. This generally brings calculated values closer to experimental observations, although it may sometimes result in slight under binding. As a result, GGA overcomes several deficiencies inherent to the LDA approach and provides a more reliable description of reaction energetics and energy barriers on metallic and semiconducting surfaces *[109,110]*. Nevertheless, GGA remains limited in its ability to accurately describe long-range dispersion interactions, such as van der Waals forces.

1.7.4.3. van der Waals Density Functional (vdW-DF) Method

Accurate evaluation of electronic structure and related properties in sparse systems requires a balanced treatment of two distinct interactions: strong, short-range chemical bonding and weak, long-range dispersion forces. Systems such as van der Waals (vdW) complexes, soft condensed matter, and biomolecular assemblies are characterized by relatively large interatomic separations, where non-local dispersion interactions play a decisive role in determining structure and stability. To properly capture these effects, this thesis adopts the van der Waals density functional (vdW-DF) framework for electronic structure calculations *[111,112]*.

The vdW-DF was first introduced in 2004 as a non-local extension of density functional theory, specifically designed to incorporate long-range dispersion interactions within the exchange correlation energy functional *[113]*. A key advantage of this approach is that vdW interactions are treated on the same theoretical footing as other bonding mechanisms, including covalent, ionic, hydrogen, and metallic interactions. As a result, vdW-DF has demonstrated reliable performance across a wide range of weakly and moderately bonded systems. Within the vdW-DF framework, the total energy is expressed in terms of the Kohn–Sham scheme as:

$$E^{vdW-DF}[n] = T_S[n] + E_H[n] + \int_r V_{ext}(r)n(r) + E_{xc}^{vdW-DF} \quad (1.21)$$

where T_S denotes the kinetic energy of non-interacting electrons, E_H is the Hartree electrostatic energy, and V_{ext} represents the external potential due to nuclei and other fields. The distinguishing feature of vdW-DF, compared to LDA or GGA, lies in the treatment of the exchange-correlation term, which explicitly includes non-local density contributions.

The exchange-correlation energy in vdW-DF is written as:

$$E_{xc}^{vdW-DF} = E_x^{GGA} + E_C^{LDA} + E_c^{nl} \tag{1.22}$$

In this expression, the exchange energy is taken from a generalized gradient approximation, while the local part of the correlation energy follows the LDA form. The additional term E_c^{nl} accounts for non-local electron correlation effects responsible for vdW interactions. Thus, vdW-DF combines semi-local exchange and local correlation with an explicitly non-local correlation contribution.

Although the kinetic and electrostatic components of the Kohn–Sham formalism are themselves non-local operators, the uniqueness of vdW-DF arises from the explicit non-local dependence introduced directly into the correlation energy functional. The non-local correlation term is defined through a six-dimensional integral,

$$E_C^{nl}[n] = \frac{1}{2}\int_r \int_{r'} n(r)\,\phi(r,r')\,n(r') \tag{1.23}$$

where $\phi(r,r')$ is a kernel function describing the interaction between density fluctuations at two spatial points. In the asymptotic regime, this kernel reproduces the characteristic $1/r^6$ decay associated with vdW forces. The kernel function can be expresses as:

$$\phi(r,r') = \frac{2me^4}{\pi^2}\int_0^\infty a^2\,da \int_0^\infty b^2\,dbW(a,b) \times T(v(a),v(b),v'(a),v'(b)) \tag{1.24}$$

here $T(w,x,y,z) = \frac{1}{2}\left[\frac{1}{w+x} + \frac{1}{y+z}\right]\left[\frac{1}{(x+z)(w+y)} + \frac{1}{(y+z)(w+z)}\right]$ and

$$W(a,b) = 2[(3-a^2)b\cos b \sin a + (3-b^2)a\cos a \sin b + (a^2+b^2-3)\sin a \sin b - 3ab\cos a \cos b]/a^3b^3$$

where $v(y) = y^2/2h(\frac{y}{d})$ and $v'(y) = y^2/2h(\frac{y}{d'})$ and $d = |r - r'|q_0(r)$ and $d' = |r - r'|q_0(r')$

The parameters $q_0(r)$ can be written as: $q_0(r) = \frac{\epsilon_{xc}^0}{\epsilon_x^{LDA}} k_F(r)$ (1.25)

The quantity ϵ_{xc}^0 corresponds to an LDA-based exchange-correlation energy augmented by a gradient correction:

$$\epsilon_{xc}^0 \approx \epsilon_{xc}^{LDA} - \left[\frac{Z_{ab}}{9}\left(\frac{\nabla n}{2k_F n}\right)^2\right] \quad (1.26)$$

The LDA exchange energy entering this expression is given by

$$\epsilon_{xc}^{LDA} = -3e^2 k_F/4\pi \quad (1.27)$$

with the Fermi wavevector defined as

$$k_F^3 = 3\pi^2 n \quad (1.28)$$

The parameter Z_{ab} controls the magnitude of the gradient correction and is determined from first-principles considerations. For the vdW-DF functional, $Z_{ab} = -0.8491$, although it should be noted that this parameter is, in general, density-dependent rather than strictly constant. Through this formalism, vdW-DF provides a physically motivated and internally consistent description of dispersion interactions within density functional theory, making it particularly well-suited for the study of sparse and weakly bonded systems.

1.7.5. Electronic Structure of Periodic Solids

The electronic structure approaches discussed above are formally suited to systems containing a finite number of electrons, such as isolated atoms or molecules. In contrast, crystalline solids are inherently periodic and contain an effectively infinite number of electrons. As a result, standard DFT formulations cannot be applied directly to crystals without further methodological extensions. This limitation is overcome by introducing periodic boundary conditions, which exploit the translational symmetry of crystalline material. Owing to this periodicity, the crystal potential is itself periodic, and consequently the effective potential appearing in the Kohn-Sham (K-S) equations satisfies:

$$V_{eff}(\mathbf{r} + \mathbf{R}) = V_{eff}(\mathbf{r}) \quad (1.29)$$

where R denotes a lattice translation vector.

Under these conditions, the single particle wavefunctions can be expressed using the Bloch theorem *[114]*. According to this theorem, the electronic wavefunction takes the form:

$$\Psi_{\mathbf{k}}(\mathbf{r}) = e^{\mathbf{ikr}} u_{\mathbf{ik}}(\mathbf{r}) \tag{1.30}$$

where k is the crystal wavevector and $u_{ik}(\mathbf{r})$ is a lattice-periodic function satisfying $u_{ik}(\mathbf{r}) = u_{ik}(\mathbf{r} + \mathbf{R})$. The periodic part of the wavefunction can be expanded as a Fourier series in terms of reciprocal lattice vectors G with corresponding plane-wave coefficients (c_{iG}),

$$u_{ik}(\mathbf{r}) = \sum_{\mathbf{G}} c_{i\mathbf{G}} e^{i\mathbf{G}.\mathbf{r}} \tag{1.31}$$

The reciprocal lattice vectors obey the condition $G.R = 2\pi m$, where m is an integer reflecting the translational symmetry of the lattice. Substituting this expansion into the Bloch form leads to an alternative representation of the wavefunction as a superposition of plane waves,

$$\Psi_{ik}(\mathbf{r}) = \sum_{\mathbf{G}} c_{\mathbf{ik+G}} e^{i(\mathbf{k+G}).\mathbf{r}} \tag{1.32}$$

This formulation transforms the many-electron problem of a solid, which involves infinitely many electrons, into a problem defined over an infinite set of k-points in reciprocal space. While this may appear counterintuitive, it enables a major simplification. The electronic wavefunction vary smoothly with respect to nearby k-points, allowing the Brillouin zone to be sample using a finite and discrete k-point mesh. Consequently, accurate electronic structure calculations for crystalline solids can be achieved by considering a suitably chosen to set of k-points rather than the full continuum.

1.7.6. Pseudopotentials

In nanoscale electronic device simulations, a clear understanding of the Kohn–Sham (K–S) equations and the treatment of exchange–correlation effects is essential. As the size of a device increases, the computational cost rises rapidly because of the large number of electrons involved, particularly tightly bound core electrons. These core electrons, however, play a minimal role in chemical bonding and most relevant physical or chemical properties,

which are largely governed by valence electrons. This observation motivates the use of pseudopotentials (PPs), which reduce the effective number of electrons treated explicitly in the calculation. Pseudopotentials provide an effective description of the chemically inert core states while retaining an explicit and accurate treatment of valence electrons. In this approach, the strong Coulomb potential of the nucleus together with the core electrons is replaced by an effective ionic potential acting only on the valence states. By removing the need to represent the rapidly oscillating core-electron wavefunctions, pseudopotentials significantly lower the number of basis functions required and thereby improve computational efficiency.

The construction of a pseudopotential typically involves two main steps.

1. First, the atomic core electrons are excluded from calculation, and their effect is absorbed into an effective potential.
2. Second, the resulting ionic potential is smoothed so that the valence wavefunctions remain well behaved and free of the strong oscillations imposed by orthogonality to the core states.

A rigorous framework for pseudopotential construction was established by Hamann and co-workers through the development of norm-conserving pseudopotential *[115]*. In this formalism, several conditions are imposed to ensure accuracy and transferability:

(i) The eigenvalues of the real and pseudo valence states are identical for a chosen atomic reference configuration.

(ii) The real and pseudo wavefunctions coincide beyond a selected core radius $r \geq r_C$.

(iii) The integrated charge density of the real and pseudo wavefunctions matches outside the core region, thereby conserving the norm; and

(iv) The logarithmic derivatives of the real and pseudo wavefunctions, along with their first energy derivatives, agree beyond the core radius, ensuring reliable behavior across different chemical environments.

Various first-principles schemes have been proposed for generating pseudopotentials, differing mainly in how the effective potential is defined and how the pseudo wavefunctions are smoothed. In VASP, pseudopotentials are provided in the form of POTCAR files and are available for all elements through the VASP portal. Most of these pseudopotentials are constructed within the projector augmented wave (PAW) framework *[116]*. The PAW datasets distributed with VASP are generated following well-established construction protocols described in the literature *[117]*.

VASP also provides supporting documentation to assist users in selecting and preparing pseudopotentials. This includes practical guidelines for generating POTCAR files in the correct format, as well as detailed recommendations for choosing an appropriate pseudopotential variant depending on the target system and property of interest. The available pseudopotential libraries are broadly divided into two categories. Standard PAW potentials are designed to accurately describe predominantly occupied electronic states and are well-suited for routine DFT calculations. In contrast, PAW potentials labeled with "_GW" suffix are optimized for a more reliable treatment of unoccupied states at energies well above the Fermi level. These GW-type potentials are therefore recommended for optical property calculations and for simulations based on many-body perturbation theory. For each pseudopotential, the accompanying tables specify key information, including the potential label, the number of valence electrons, the reference valence electronic configuration, and the recommended plane-wave cutoff energy (ENMAX). This information provides a practical basis for selecting suitable pseudopotentials and ensuring consistent and reliable calculations.

1.7.7. Basis Sets

The Kohn–Sham formalism expresses the single-particle states in terms of an appropriate set of basis functions. Selecting a basis set that is computationally efficient while still delivering reliable accuracy is,

therefore, a key step in any DFT calculation. For molecular systems, Gaussian-based basis sets are among the most employed. In this framework, molecular orbitals ψ_i are expanded as linear combinations of atom-centred basis functions ϕ_k, with expansion coefficients C_{ki}, as shown in Equation. (1.33):

$$\psi_i = \sum_{k=1}^{n} C_{ki}\, \phi_k \tag{1.33}$$

Both nature and the number of basis functions strongly influence the quality of the electronic structure description. In practical molecular calculations, contracted Gaussian-type orbitals (GTOs) are widely used because they offer a favorable balance between accuracy and computational cost, while closely approximating the behavior of Slater-type orbitals (STOs). Since chemically relevant properties are primarily governed by valence electrons, higher accuracy is typically achieved by increasing the flexibility of the valence basis. This leads to split-valence basis sets, such as double-zeta and triple-zeta expansions, which provide an improved description of bonding and electronic distributions.

For crystalline and other periodic systems, a different strategy is commonly adopted. When the lattice can be defined by translation vectors a_1, a_2, and a_3, plane-wave basis functions of the form $\exp(ik \cdot r)$ are often used. These basis functions naturally satisfy Bloch's theorem, where the single-particle states are written as a product of a plane wave and a lattice-periodic function, as expressed in Equation (1.33). Here, k denotes the wave vector, and $u_k(r)$ has the periodicity of the supercell *[118]*. The use of plane-wave basis sets offers several advantages for periodic systems. The electronic structure problem can be formulated in reciprocal space, allowing the Schrödinger equation to be solved independently at each k-point. Moreover, integrations are restricted to the first Brillouin zone rather than the entire real-space lattice, which substantially reduces the computational effort while preserving accuracy.

1.7.8. Projector Augmented Wave (PAW) Methods

The rapid oscillations of electronic wavefunctions in the vicinity of atomic cores make an explicit description of core electrons particularly demanding in plane-wave based approaches. Accurately resolving these features requires a very large plane-wave cutoff, which significantly increases the computational cost for periodic systems. In most problems of chemical and physical interest, however, the key contributions to bonding and material properties arise mainly from the valence electrons. As a result, it is common practice to treat only the valence states explicitly with plane-wave basis functions, while the combined effect of the core electrons with plane-wave basis functions, while the combined effect of the core electrons and nuclei is incorporated through an effective potential, referred to as a pseudopotential *[119]*. This strategy is widely known as the frozen-core approximation.

Several classes of pseudopotentials have been developed for plane-wave DFT calculations, including norm-conserving and ultrasoft pseudopotentials, which mainly differ in their required energy cutoffs and computational efficiency. To further minimize the reliance on empirical choices in pseudopotential construction, the projector augmented wave (PAW) formalism was originally introduced by Bloch and subsequently implemented in practical electronic structure codes by Kresse and Joubert *[116,117]*. In the PAW framework, the strongly varying all-electron wavefunction in the core region is mapped onto a smoother pseudo-wavefunction through a linear transformation, as expressed in Equation 1.34:

$$|\psi_n\rangle = |\tilde{\psi}_n\rangle + (|\phi_n\rangle - |\tilde{\phi}_n\rangle) \langle \tilde{p}^i|\tilde{\psi}_n\rangle \qquad (1.34)$$

In this representation, the true all-electron state Ψ_n is reconstructed from the pseudo-wavefunction $\tilde{\Psi}_n$ using a combination of all-electron partial waves, corresponding pseudo partial waves, and localized projector functions. This approach enhances the accuracy and transferability of

pseudopotentials while retaining computational efficiency comparable to pseudopotentials methods, and it allows results that closely match all-electron calculations. In the present thesis, periodic systems are therefore treated using the PAW method as implemented in the Vienna *ab initio* Simulation Package (VASP).

1.7.9. Dispersion Correction Method

A significant drawback of conventional DFT approaches is the limited capability of standard exchange-correlation (XC) functionals to describe the non-local, long-range dispersion interactions. These weak interactions play a decisive role in accurately modelling catalyst-surface interactions, adsorption energies, and reaction mechanisms. To address this deficiency without substantially increasing computation cost, several dispersion-correction schemes have been developed, each based on a different theoretical strategy.

One widely used class of materials is the semi-empirical DFT-D approach proposed by Grimme, in which an explicit dispersion correction is added to the total DFT energy to account for van der Waals interactions *[120,121]*. In the DFT-D3 formulation, the corrected total energy is expressed as:

$$E_{DFT-D3} = E_{KS-DFT} - E_{disp}^{D3} \tag{1.35}$$

where E_{KS-DFT} is the energy obtained from solving the Kohn-Sham equations using a chosen XC functional, and E_{disp}^{D3} is the dispersion correction term. The DFT-D3 dispersion energy consists of both pairwise and three-body contributions,

$$E_{disp}^{D3} = E_{disp}^{D2} + E_3 \tag{1.36}$$

The two-body dispersion term, inherited from the DFT-D2 scheme, is given by:

$$E_{disp}^{D2} = -s_6 \sum_{i=1}^{N_{at}-1} \sum_{j=i+1}^{N_{at}} \frac{C_6^{ij}}{R_{ij}^6} f_{dmp}(R_{ij}) \tag{1.37}$$

Here, s_6 is a scaling factor that depends on the selected XC functional, N_{at} is the total number of atoms, C_6^{ij} denotes the dispersion coefficient for the atomic pair i and j, and R_{ij} is the interatomic distance. The damping function

$f_{dmp}(R_{ij})$ ensures that the dispersion correction smoothly vanishes at short distances, preventing unphysical divergences.

Beyond pairwise interactions, the DFT-D3 method also incorporates a three-body dispersion term to account for collective effects among atomic triplets. This contribution is written as:

$$E_3 = \sum_{A,B,C} f_d^{(3)}(r_{ABC})\, E_{ABC} \tag{1.38}$$

where the summation runs over all atomic triplets A, B, and C, and the damping function depends on the geometric mean distance r_{ABC} of the triplet. An alternative approach for treating dispersion interactions is provided by van der Waals density functional (vdW-DF), in which non-local correlation effects are incorporated directly into the XC functional. Within this framework, the exchange-correlation energy is expressed as:

$$E_{XC}[n] = E_{GGA} - E_{GGA}^{c} + E_{LDA}^{c} + E_{nl}^{c} \tag{1.39}$$

where E_{GGA} is the total GGA energy, E_{GGA}^{c} and E_{LDA}^{c} are the GGA and LDA correlation energies, respectively, and E_{nl}^{c} represents the nonlocal correlation term responsible for capturing dispersion interactions [113,122]. The nonlocal correlation energy is evaluated as:

$$E_{nl}^{c} = \frac{1}{2} \int dr \int dr'\; n(r)\, \phi(r, r')\, n(r') \tag{1.40}$$

where $n(r)$ is the electron density and $\phi(r, r')$ is a kernel function describing the nonlocal interaction between electron densities at different spatial points.

Within the vdW-DF framework, several functionals have been proposed by combining different exchange terms with the same nonlocal correlation component. The original vdW-DF and vdW-DF2 formulations employ revPBE and rPW86 exchange functionals, respectively *[123]*. Optimized variants such as optPBE, optB88, and optB86b use modified exchange functionals to improve predictions of structural parameters, binding energies, and adsorption characteristics *[124]*. In this thesis, dispersion interactions are treated using the DFT-D3 correction scheme unless explicitly stated otherwise.

1.8 Hubbard Model

Standard DFT approaches often fail when applied to strongly correlated materials because commonly used exchange–correlation functionals cannot fully eliminate the self-interaction error associated with localized electrons. While hybrid functionals offer a partial remedy by mixing exact exchange, their practical use is limited by the high computational cost. To overcome this limitation, the Hubbard model introduces an additional interaction term into the electronic Hamiltonian to explicitly account for electron–electron interactions within localized orbitals *[125]*.

The DFT+U methodology incorporates this idea by adding an on-site Coulomb correction, denoted by the parameter U, to the conventional DFT framework. Within the GGA+U formalism, the total energy of the system is expressed as

$$E_{GGA} + U[\rho(r)] = E_{GGA}[\rho(r)] + E_{Hub}[\{n^{I\sigma}_{mm}\} - E_{dc}[n^{I\sigma}]] \quad (1.41)$$

Here, E_{GGA} corresponds to the energy obtained from the standard GGA exchange–correlation functional. The Hubbard correction term, E_{Hub}, depends on the occupation matrix $n^{I\sigma}_{mm'}$, which describes localized orbitals characterized by the atomic site index I, orbital indices m, m', and spin σ. The final term, E_{dc}, removes the double-counting contribution arising from interactions already included in the GGA functional. By explicitly treating on-site Coulomb interactions, the DFT+U approach provides a more reliable description of the electronic structure, magnetic properties, and chemical behavior of systems containing localized d- or f-electron states.

1.9 Climbing Image Nudged Elastic Band (CINEB) Methods

Accurate determination of kinetic quantities such as activation energies, reaction rates, and transition-state geometries is essential for a mechanistic understanding of heterogeneous catalytic processes. The nudged elastic band (NEB) technique is widely used to identify the minimum energy path (MEP) connecting reactant and product states on a potential energy surface *[126]*. In the NEB approach, a series of intermediate structures, referred to as images, are generated between the initial and final configurations. These

images are linked by artificial spring forces to maintain continuity along the reaction coordinate, while the atomic forces are iteratively optimized to relax the pathway toward the true MEP.

In this thesis, an improved variant of NEB, namely the climbing-image NEB (CI-NEB) method, is adopted. In CI-NEB, the image with the highest energy along the reaction path is systematically driven toward the saddle point by maximizing its energy along the tangent direction while minimizing it in all perpendicular directions *[127]*. The force acting on the climbing image is given by:

$$F_{i,max} = -\nabla E\left(R_{i,max}\right) + 2\nabla E\left(R_{i,max}\right).\hat{T}_{i,max}\,\hat{T}_{i,max} \qquad (1.42)$$

where Erepresents the total energy of the system as a function of the atomic coordinates R, and $\hat{T}_{i,max}$ denotes the normalized local tangent vector at the climbing image. For surface reactions on crystalline catalysts, a relatively small number of images, typically between four and eight, is generally sufficient to resolve the MEP. In several chapters of this thesis, NEB calculations are complemented by the dimer method to efficiently locate transition states *[128]*. The dimer approach identifies the saddle point starting from two closely spaced high-energy configurations, collectively referred to as a dimer, obtained from a CI-NEB calculation. The algorithm drives the dimer uphill on the potential energy surface while performing rotational searches to identify the direction of lowest curvature. This strategy significantly lowers the computational expense compared to full NEB calculations, as it avoids explicit path optimization and relies on Hessian information combined with gradient-based minimization. The transition states obtained from either NEB or dimer methods are subsequently employed within the harmonic transition state theory (HTST) framework to evaluate kinetic parameters.

1.10 Crystal Orbital Hamilton Population Method

Crystal Orbital Hamilton Population (COHP) analysis is a widely used technique for examining chemical bonding in solid-state systems by evaluating the bond-resolved density of states between selected atomic pairs

[129]. Within the COHP framework, the total band energy is partitioned into bonding, nonbonding, and antibonding contributions, allowing a detailed interpretation of orbital interactions in crystalline materials. In plane-wave based electronic structure calculations, a projected density matrix P^{proj} can be constructed for each electronic band j at every k-point. This projected density matrix is expressed as:

$$P_{\mu\nu}^{j,proj}(k) = T_{j\mu}^{*}(k)T_{j\nu}^{*}(k) \quad (1.43)$$

Using this projection, the corresponding Hamiltonian matrix elements can be written as:

$$H_{\mu\nu}^{j,proj}(k) = \sum_j \varepsilon_j(k)\, T_{j\mu}^{*}(k)\, T_{j\nu}^{*}(k) \quad (1.44)$$

where $\varepsilon_j(k)$ denotes the eigenvalue of the band j at wave vector k, and $T_{j\mu}(k)$ represents the projection coefficient of the plane-wave state onto the local atomic orbital basis. These projected quantities enable extraction of orbital-resolved information, which forms the basis for quantitative bonding analysis.

Analogous to the integrated density of states, the integrated COHP (ICOHP) can be evaluated to quantify the cumulative contribution of individual bands to the total band energy. The ICOHP value serves as a useful metric for estimating the bond strength between specific atom pairs. COHP analysis is performed using fully converged electronic wavefunctions obtained from self-consistent calculations. In this work, COHP calculations were carried out using the LOBSTER package interfaced with the VASP code *[130]*.

1.11 Computational Hydrogen Electrode (CHE) Model

A central difficulty in first-principles simulations of electrocatalytic processes is the evaluation of reaction free energies for electrochemical steps, since standard DFT cannot directly describe isolated protons. Within the computational hydrogen electrode (CHE) framework *[131]*, the combined chemical potential of a proton and an electron participating in a concerted proton–electron transfer is approximated by half the chemical potential of a hydrogen molecule at the equilibrium potential, given by:

$$\mu H^+ + \mu e^- = \frac{1}{2}\mu H_2 \quad (1.45)$$

This approximation allows the energy of a proton–electron pair to be referenced to the DFT-calculated total energy of gas-phase H_2. When an external electrode potential U is applied, the proton chemical potential is shifted by $-eU$. Accordingly, the free energy change of an electrochemical reaction step can be written as:

$$\Delta G = \Delta E + \Delta ZPE - T\Delta S - neU \quad (1.46)$$

where ΔE represents the electronic energy difference obtained from DFT, ΔZPE is the zero-point energy correction, T is the temperature, ΔS is the entropy change, n is the number of transferred electrons, and e denotes the elementary charge.

1.12 Machine Learning

Machine learning (ML), also described as statistical learning or predictive modeling, is an interdisciplinary field at the interface of artificial intelligence, statistics, and computer science. Its primary goal is to design computational models that can learn from data, identify underlying patterns, and make predictions or decisions without being explicitly programmed for each task *[132,133]*. Data form the core of any ML framework and may be structured, such as tabulated datasets and databases, or unstructured, including text, images, and other complex data formats. By extracting patterns from these data, ML models can generalize their learning to previously unseen cases. As additional data become available, the models can be updated iteratively, leading to progressively improved predictive performance.

Broadly, machine learning approaches are categorized into four major paradigms: supervised, semi-supervised, unsupervised, and reinforcement learning *[134,135]*. In supervised learning, models are trained using labeled datasets, where both input features and corresponding target outputs are known, enabling the algorithm to learn explicit input–output relationships. Semi-supervised learning combines a limited number of labeled samples with a large pool of unlabeled data and is particularly useful when data

annotation is costly or time-consuming. Reinforcement learning relies on an agent interacting with an environment and learning optimal decision-making strategies through trial and error by maximizing a cumulative reward signal. This paradigm is especially effective in problems such as control systems, robotics, and game playing. In contrast, unsupervised learning focuses on discovering intrinsic structures or hidden patterns in unlabeled data, such as clustering and dimensionality reduction. Beyond these core categories, several specialized branches of machine learning exist. Deep learning is a prominent subset that employs multilayer neural network architectures, often referred to as deep neural networks, to achieve high accuracy in complex tasks including image analysis and speech recognition. In this thesis, machine learning methodologies are employed to capture and interpret the non-monotonic size-dependent trends observed in oxygen reduction reaction (ORR) activity within the nanocluster regime.

1.12.1 Supervised Learning

Supervised machine learning is a fundamental learning paradigm in which models are trained using labeled datasets. In this framework, each training sample is composed of an input feature vector paired with a known target output. The objective of the model is to learn the functional relationship between inputs and outputs so that it can reliably predict outcomes for previously unseen data. During the training process, the model adjusts its internal parameters to reduce the discrepancy between predicted and true values, typically measured through a predefined loss function. The effectiveness of supervised learning largely depends on the quality and representativeness of the training data, as well as the relevance of the selected input features. When sufficient high-quality labeled data are available, supervised models can achieve high predictive accuracy across a wide range of practical problems. Feature selection, data preprocessing, and appropriate model tuning play critical roles in determining overall model performance. Supervised learning problems are commonly divided into two major categories: classification and regression *[136]*. Classification tasks

involve assigning inputs to discrete classes, for example, identifying whether an email is spam or legitimate. Regression tasks, on the other hand, aim to predict continuous numerical values, such as property prices or financial indicators. A more detailed discussion of regression- and classification-based learning methods is presented in the subsequent sections.

1.12.2 Regression

Regression is a key component of supervised machine learning that focuses on estimating a continuous output variable using one or more input variables, commonly referred to as features or descriptors *[137]*. In contrast to classification problems, where predictions belong to discrete categories, regression problems involve the prediction of numerical values. The primary goal of regression is to establish a mathematical relationship between the response variable and the explanatory variables. Once this relationship is learned, the model can be used to forecast future values, analyze trends, and gain insights into the structure and behavior of the underlying dataset. In this thesis, we utilized the regression algorithm for the prediction of adsorption energy of principle reaction intermediates to predict the activity of the nanoclusters.

1.12.3 Unsupervised Learning

Unsupervised learning is a class of machine learning techniques in which models are trained using data that do not contain predefined labels or target outputs. Unlike supervised learning, there is no explicit notion of a "correct answer." Instead, the objective is to automatically discover hidden patterns, structures, or relationships that are inherent in the data itself. The model relies solely on the statistical properties of the input features to organize, group, or represent the data in a meaningful way. In unsupervised learning, the algorithm attempts to identify similarities, correlations, or regularities among data points, which can reveal underlying structures that may not be apparent through direct inspection. Common tasks include clustering, where data points are grouped based on similarity; dimensionality reduction,

where high-dimensional data are transformed into a lower-dimensional representation while preserving essential information; and density estimation, which focuses on modeling the distribution of data in feature space. These approaches are particularly valuable for exploratory data analysis, as they help researchers gain insights into complex datasets without prior assumptions. Unsupervised learning is especially useful in situations where labelled data are scarce, expensive, or impractical to obtain. It plays a crucial role in fields such as materials science, chemistry, and physics, where large datasets are often generated through simulations or experiments without clear target labels. By uncovering intrinsic trends and structural organization within the data, unsupervised learning can guide hypothesis generation, feature extraction, anomaly detection, and the development of more accurate supervised models in subsequent analyses *[138]*.

1.12.4 Dataset

A dataset is an organized collection of data points used for analysis or modeling, where each sample is described by a set of features and, in some cases, corresponding labels or target values. In machine learning, datasets provide the information from which models learn patterns, make predictions, and are evaluated for accuracy and reliability.

1.12.4.1 Training set

A training dataset is a subset of the overall dataset that is used to fit and optimize a machine learning model by learning the relationship between input features and output targets or labels. It serves as the foundation for adjusting model parameters so that the model can make accurate predictions on new, unseen data.

1.12.4.2 Test set

A test dataset is a separate portion of the dataset used to evaluate the performance of a trained machine learning model on unseen data. It provides an unbiased assessment of how well the model generalizes beyond the training dataset.

1.12.4.3 Validation set

A validation dataset is a subset of data used during model development to tune hyperparameters and make decisions about model architecture without influencing the final evaluation. It helps prevent overfitting by providing an intermediate check on model performance before testing.

1.12.4.4 Feature

In machine learning, features are the measurable input variables or descriptors that represent the underlying properties of a system and are used by a machine learning model to make predictions. They encode relevant information from the raw data and strongly influence the accuracy and interpretability of the model *[139]*.

1.12.4.5 Feature Engineering

Feature engineering refers to the process of creating, selecting, transforming, and refining features from raw data to improve the performance and robustness of a machine learning model. It combines domain knowledge with statistical insight to ensure that the features capture the most relevant patterns and relationships in the data *[140,141]*.

1.12.4.6 Pearson Correlation Coefficient (PCC)

The Pearson correlation coefficient (PCC) is a standard statistical measure used to evaluate the strength and direction of a linear relationship between two continuous variables *[141]*. It is denoted by r and varies between −1 and 1, where a value of 1 signifies a perfect positive linear correlation, −1 represents a perfect negative linear correlation, and 0 indicates no linear dependence. PCC is computed by first evaluating the covariance between the two variables and then normalizing it using the product of their standard deviations. Owing to this normalization, PCC is dimensionless, which makes it convenient for comparing correlations across different datasets and variables. In data analysis and machine learning studies, PCC is commonly employed to identify highly correlated or redundant features, aiding in feature selection and reducing multicollinearity. The mathematical expression for PCC is given in Equation (1.91):

$$PCC = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})}} \quad (1.47)$$

where x_i and y_i correspond value of x and y, respectively. The $\bar{x}$ and $\bar{y}$ are the mean values of x and y, respectively.

1.12.4.7 Spearman Rank Correlation Coefficient (SCC)

The Spearman rank correlation coefficient (SCC) is a non-parametric statistical metric used to quantify the strength and direction of association between two variables based on their ranks [142]. In contrast to the Pearson correlation coefficient, SCC does not rely on assumptions of linearity or normal data distribution. Instead, it measures how consistently the variables vary with respect to each other in a monotonic manner. Since the calculation is performed using ranked data rather than raw values, SCC is less sensitive to outliers and is well suited for ordinal or non-normally distributed datasets. Owing to these properties, SCC is widely applied in cases where the conditions required for PCC are not satisfied. The mathematical formulation of SCC is given in Equation 1.48:

$$\rho = 1 - \frac{6\sum {d_i}^2}{n(n^2-1)} \quad (1.48)$$

where d_i is the difference in the ranks given to the two variables and n is the total number of samples.

1.12.4.8 Mutual Information (MI)

Mutual Information (MI) is a commonly used feature selection approach that captures the degree of statistical dependence between two variables *[142]*. It measures how much knowledge of one random variable reduces the uncertainty associated with another, making it suitable for identifying both linear and non-linear relationships. In this framework, the mutual dependence between variables X and Y is expressed as:

$$MI(X,Y) = H(X) - H(X|Y) \quad (1.49)$$

Here, $H(X)$ represents the Shannon entropy, which quantifies the uncertainty of a discrete random variable X, and is defined as:

$$H(X) = -\sum_{x \in X} p(x) \log p(x) \quad (1.50)$$

where $p(x)$ denotes the probability distribution of X. The term $H(X|Y)$ is the conditional entropy of X given Y, reflecting the remaining uncertainty in X once Y is known. Consequently, mutual information can be interpreted as the decrease in uncertainty of X achieved by incorporating information from Y.

1.12.4.9 Hyperparameter Tuning

Hyperparameter tuning plays a critical role in machine learning by identifying the most suitable set of hyperparameters that lead to optimal model performance. In contrast to model parameters, which are learned directly from the training data, hyperparameters are predefined and govern how the learning process is carried out *[143]*. Several strategies exist for this purpose, among which Grid Search and Randomized Search are two widely adopted approaches for systematic hyperparameter optimization *[144]*.

1.12.4.10 Grid Search

Grid search is a systematic hyperparameter tuning approach in which predefined sets of values are assigned to each hyperparameter, and the model is trained and evaluated for every possible combination of these values. By exhaustively exploring the entire parameter space, grid search ensures that the best-performing hyperparameter combination within the specified ranges is identified, albeit at the cost of increased computational expense.

1.12.4.11 Randomized Search

Randomized search is a hyperparameter tuning strategy in which probability distributions are defined for each hyperparameter rather than fixed value sets. The algorithm randomly draws combinations from these distributions and evaluates the model using the sampled configurations. By testing a limited number of randomly selected combinations, this method can cover a wider region of the hyperparameter space while requiring significantly fewer model evaluations than an exhaustive grid search.

1.12.5 Machine Learning Algorithms

1.12.5.1 Linear Regression

Linear regression is a supervised learning approach employed to model a straight-line relationship between an input variable x and a corresponding output variable y *[145]*. The objective is to determine the optimal linear function that best represents the observed data trend. This relationship can be mathematically expressed as:

$$y = wx + b \tag{1.51}$$

where y denotes the predicted response, w represents the slope of the fitted line indicating the sensitivity of y to changes in x, and b is the intercept that defines the value of y when x = 0.

1.12.5.2 Logistic Regression

Logistic regression is a widely used classification technique that estimates the probability of a data point belonging to a particular categorical class *[146]*. It employs a logistic, or sigmoid, activation function σ, which maps input values to a range between 0 and 1. The sigmoid function is defined as

$$\sigma(t) = \frac{1}{1+ e^{-t}} \tag{1.52}$$

This probabilistic output makes logistic regression especially suitable for binary classification problems.

1.12.5.3 K-Nearest Neighbors

The K-Nearest Neighbors (KNN) algorithms is a non-parametric learning method that makes predictions based on the proximity of data points in the feature space *[147]*. For a given input, the algorithm identifies the 'k' closest training samples and assigns a class label or prediction based on the collective information from these neighbors.

1.12.5.4 Kernel Ridge Regression

Kernel ridge regression integrates ridge regression with kernel-based techniques to capture nonlinear dependencies between input variables and target outputs *[148]*. By implicitly mapping data into a higher-dimensional feature space, this method enables effective modeling of complex relationships that cannot be represented using linear functions.

1.12.5.5 Support Vector Machines

Support Vector Machines (SVM) are supervised learning algorithms that aim to separate data points of different classes by constructing an optimal decision boundary, known as the maximum-margin hyperplane *[149]*. The hyperplane is chosen such that the distance between the nearest data points of each class and the boundary is maximized. The mathematical form of the hyperplane is given by:

$$W^T x + b = 0 \tag{1.53}$$

where W is the normal vector defining the orientation of the hyperplane, x represents the feature vector, and b controls the offset of the hyperplane from the origin.

1.12.5.6 Random Forest

Random forest is an ensemble learning technique that constructs multiple decision trees using randomly sampled subsets of the training data and features *[150]*. The final prediction is obtained by aggregating the outputs from all individual trees, which improves robustness and reduces overfitting.

1.12.5.7 Adaptive Boosting

Adaptive boosting, commonly known as AdaBoost, is an ensemble method that sequentially combines several weak learners to form a strong predictive model *[151]*. During training, higher weights are assigned to misclassified samples in subsequent iterations, enabling the model to focus on difficult cases and progressively enhance performance.

1.12.5.8 Extreme Gradient Boosting

Extreme Gradient Boosting (XGBoost) is an optimized implementation of the gradient boosting framework designed for computational efficiency and scalability *[152]*. It builds an ensemble of decision trees in a sequential manner, where each new tree corrects the errors introduced by the previous models.

1.12.5.9 Gaussian Process Regression

Gaussian Process Regression (GPR) is a probabilistic regression approach rooted in Bayesian inference *[153]*. Instead of assuming a fixed functional form, GPR defines a distribution over possible functions f(x), allowing for uncertainty quantification in predictions while modeling nonlinear relationships.

1.12.5.10 Decision Tree

A decision tree is a hierarchical model composed of internal decision nodes, branches corresponding to decision rules, and terminal leaf nodes representing final outcomes *[154]*. It provides an intuitive and interpretable framework for both classification and regression tasks.

1.12.5.11 Extra Tree Regression

Extra tree regression is an ensemble-based regression technique that generates multiple decision trees using randomly selected features and split points *[155]*. This additional randomness increases diversity among the trees, often leading to improved generalization and reduced prediction variance.

1.12.6 Model Evaluation Metrics

1.12.6.1 Coefficient of determination (R^2)

The coefficient of determination R^2 quantifies how well a regression model explains the variance in the target variable using the input features *[156]*. Its value lies between 0 and 1 and is defined as

$$R^2 = \frac{\sum_{i=1}^{n}(M_i - \bar{m})^2}{\sum_{i=1}^{n}(m_i - \bar{m})^2} \quad \text{where} \quad \overline{m} = \frac{\sum_{i=1}^{n} m_i}{n} \tag{1.54}$$

where M_i are predicted values, m_i are the reference DFT values, $\bar{m}$ is their mean, and n is the dataset size.

1.12.6.2 Mean Absolute Error (MAE)

MAE represents the average magnitude of prediction errors without considering their direction:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|M_i - m_i| \tag{1.55}$$

1.12.6.3 Mean Squared Error (MSE)

MSE emphasizes larger errors by squaring the deviations between predicted and actual values:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(M_i - m_i)^2 \quad (1.56)$$

1.12.6.4 Root Mean Squared Error (RMSE)

RMSE is the square root of MSE and provides an error metric in the same units as the target variable:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(M_i - m_i)^2} \quad (1.57)$$

1.12.7 Cross-Validation

Cross-validation assesses model robustness and generalization by repeatedly training and testing on different data partitions. In this work, K-fold cross-validation is employed, where the dataset is divided into K subsets and each subset is used once as a test set (Figure 1.11) *[157]*.

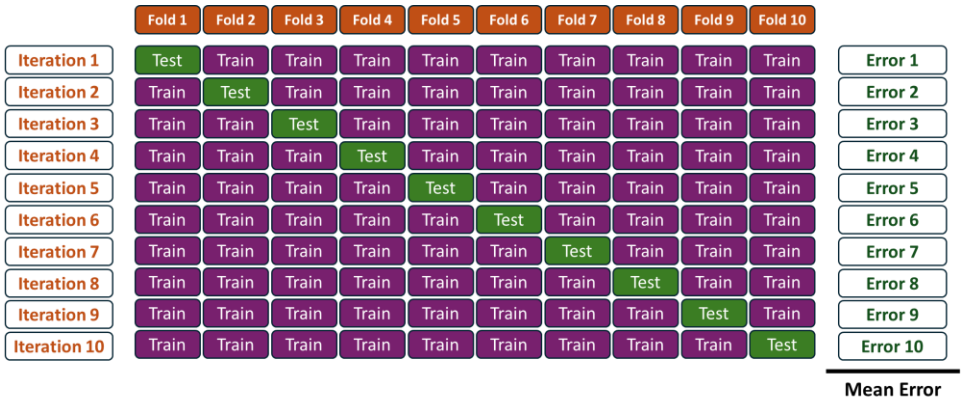


Figure 1.11: Schematic diagram showing K-fold (K=10) cross-validation.

1.12.8 Feature Importance

Feature importance in machine learning refers to the systematic evaluation of how strongly each input feature influences the prediction of the target variable. This analysis plays a key role in interpreting model behavior, selecting relevant descriptors, and improving overall predictive performance. Feature importance techniques can be broadly categorized into model-specific approaches and model-agnostic methods, both of which

offer valuable insights into the contribution of individual features within the adopted ML models.

1.12.9 Permutation Feature Importance

Permutation feature importance is a model-independent approach used to quantify the relevance of individual features by examining how random rearrangement of their values affects model performance *[150]*. In this method, the values of a selected feature are permuted while keeping all other features unchanged, and the resulting change in prediction error is evaluated. A significant increase in error upon permutation indicates that the model strongly depends on that feature, classifying it as important. In contrast, minimal or no change in error suggests a weak influence of the feature on the model's predictions.

1.12.10 Machine Learning Explainability

The need for machine learning explainability arises from the opaque nature of many conventional ML models, which often function as black boxes and provide limited insight into how predictions are generated. Explainable Artificial Intelligence (XAI) encompasses a collection of methods and frameworks designed to interpret and clarify the decision-making processes of ML models. Model explainability enables human users to understand, trust, and validate model outputs by offering transparent reasoning behind predictions (Figure 1.12). Core principles of XAI include transparency, fairness, causality, transferability, and accountability *[158]*.

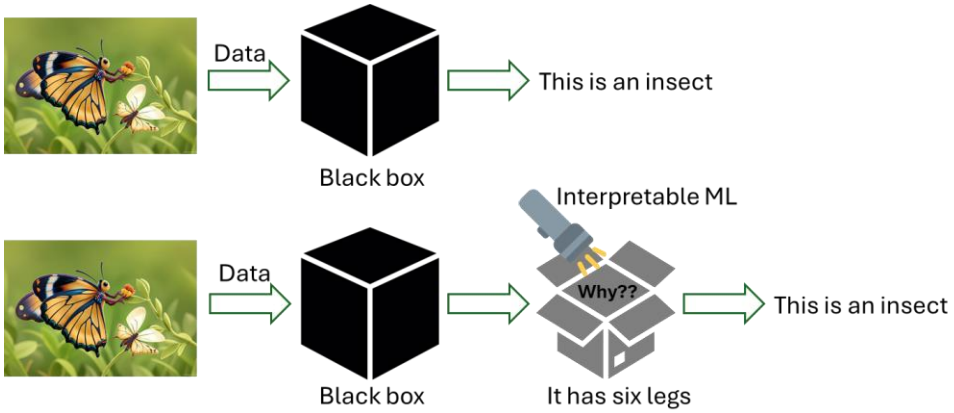


Figure 1.12: Schematic illustration of the black box model and explainable AI interpreting the decision behind ML prediction.

1.12.11 SHapley Additive exPlanations (SHAP)

SHapley Additive exPlanations (SHAP) provide a unified framework for interpreting predictions generated by machine learning models by quantifying the contribution of each input feature. The method is grounded in cooperative game theory, where each feature is treated as a player contributing to the final prediction. SHAP is model-agnostic in nature and can therefore be applied to a broad class of ML algorithms. It enables interpretation at both the local level, for individual predictions, and the global level, by summarizing feature contributions across the entire dataset *[159]*.

At the local scale, SHAP values decompose a single model prediction by distributing the deviation of the predicted output from a reference or baseline value among the input features. This provides detailed insight into how specific features influence a particular prediction. At the global scale, aggregating SHAP values over all samples yields an overall ranking of feature importance, offering a comprehensive understanding of model behavior. The Shapley value $\emptyset_x(p)$ associated with feature x, among a total of n features, for a model prediction p is computed using the following expression:

$$\emptyset_x(p) = \sum_{s \subseteq N/x} \frac{|S|!(n-|S|-1)!}{n!} (p(S \cup x) - p(S)) \qquad (1.58)$$

where S represents all possible subsets of the feature set excluding feature x, n denotes the total number of features, $p(S \cup x)$ is the model prediction obtained when feature x is included, and $p(S)$ corresponds to the prediction made without feature x.

1.12.12 Graph Neural Network

A broad spectrum of machine learning methods, including naïve Bayes *[160]*, linear regression *[161]*, symbolic regression *[162]*, decision trees *[163,164]*, support vector machines *[165,166]*, and neural networks *[167-*

169], has been widely applied in materials science and engineering. While these approaches have delivered valuable insights, they encounter significant difficulties when applied to complex crystalline systems. One major challenge arises from the reliance on handcrafted structural descriptors, which are often tailored to specific problems and do not generalize well across large and diverse crystal datasets. Additionally, crystal structures are inherently non-Euclidean, making them incompatible with conventional ML models unless complex transformations are introduced, which can degrade interpolation quality and predictive reliability. Graph neural networks (GNNs) address these limitations by directly operating on graph-structured data, as represented in Figure 1.13 *[167-177]*. In GNN-based frameworks, atomic or molecular structures are naturally represented as graphs, enabling a direct connection between structure and material properties such as formation energies, elastic constants, and electronic characteristics. This capability has led to the widespread adoption of GNNs in tasks including crystal structure prediction, high-throughput screening, and materials property estimation. As a result, GNNs are now extensively used across multiple length scales, from mesoscale microstructures to atomistic modeling.

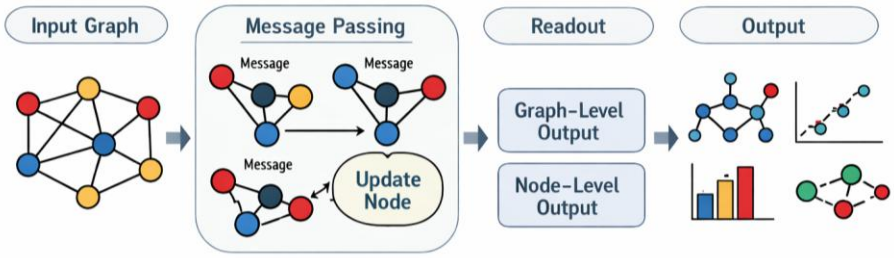


Figure 1.13: Schematic illustration of a graph neural network (GNN). The graph is represented by nodes and edges encoding atomic or structural features. Through iterative message passing, information is exchanged between neighboring nodes and node representations are updated. A readout operation aggregates node-level embeddings to produce either node-level or graph-level outputs.

Research activity related to GNN applications in materials science has grown rapidly, surpassing roughly 600 publications per year since 2022. At the mesoscale, GNNs have been successfully applied to polycrystalline systems, where they capture grain-to-grain interactions more effectively than traditional ML methods, enabling improved predictions of phenomena such as abnormal grain growth. GNNs have also been used in conjunction with image-based datasets to extract microstructural features and generate synthetic microstructures, leading to enhanced performance predictions. At the atomic scale, GNNs enable explicit modeling of atomic connectivity, bonding environments, and reaction pathways, supporting accurate predictions of electronic, thermodynamic, and structural properties *[178]*. These capabilities make GNNs particularly powerful tools for rational materials design and optimization. Accordingly, this review primarily focuses on atomic-scale GNN applications, where material structures and properties are mapped directly onto graph representations *[168]*.

In general, GNN-based materials modeling follows three main stages: graph construction, feature learning, and property prediction *[169]*. During graph construction, atoms, unit cells, surfaces, or other structural motifs are represented as nodes, while chemical bonds or geometric relationships form the edges. Feature learning is carried out through graph convolution operations, which iteratively update node and edge features by combining topological and chemical information. Finally, the learned representations are passed through fully connected layers to predict target properties. Accurate encoding of crystal structure information is critical at this stage. Common strategies include one-hot encoding, bag-of-bonds descriptors *[180]*, atom-centered symmetry functions, Coulomb matrices *[181,182]*, and crystal graph convolutional neural networks (CGCNN) *[167]*. Continuous methodological improvements aim to enhance both predictive accuracy and computational efficiency.

Overall, GNNs provide a flexible and physically meaningful framework for materials modeling. Their strength lies in their ability to handle irregular,

variable-sized structures while explicitly capturing atomic and molecular relationships. GNNs can often achieve high accuracy with limited labeled data, offer improved interpretability compared to black-box models, and exhibit strong transferability to unseen materials sharing similar structural motifs. Based on application focus, GNNs can be broadly classified into several categories. The different representative GNN models includes CGCNN *[167]* and its variants such as iCGCNN and OGCNN *[171,172]*, MEGNet *[170]*, GATGNN (Global Attention Graph Neural Network) *[183]*, ALIGNN (Atomistic Line Graph Neural Network) *[184]*, and BonDNet *[185]*. These models have robust graph representations, strong predictive performance, interpretability, and transferability across diverse materials systems.

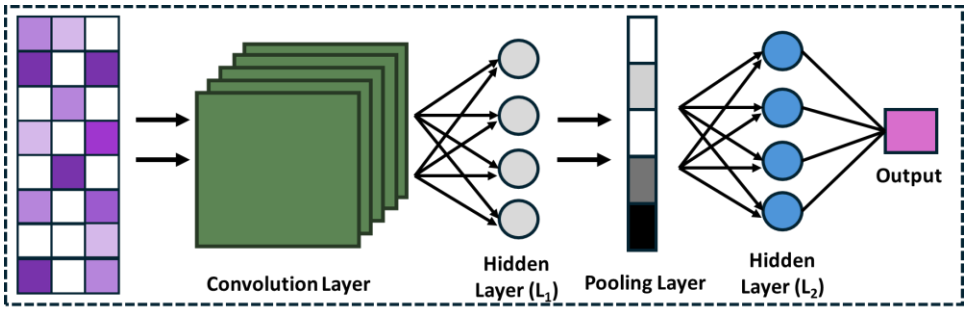


Figure 1.14: Architecture of the crystal graph convolutional neural network model.

1.12.12.1 Crystal Graph Convolutional Neural Network (CGCNN)

The crystal graph convolutional neural network (CGCNN) *[167]* is a foundational graph-based framework that represents crystalline materials as graphs, explicitly encoding atomic identities and interatomic connections. In CGCNN, convolution operations are performed directly on the crystal graph, allowing the model to automatically learn structure-aware features for property prediction using datasets typically derived from density functional theory (DFT) calculations, as represented in Figure 1.14. Developed by Xie et al. and released as an open-source model *[167]*, CGCNN treats atoms in the unit cell as nodes, while edges represent bonding or proximity relationships. Both nodes and edges are described

using feature vectors that capture elemental properties and bonding characteristics. The CGCNN architecture consists of an input layer, multiple graph convolution layers, a pooling operation, fully connected hidden layers, and an output layer. Among these components, the graph convolution layers are central, as they iteratively aggregate information from neighboring atoms to build increasingly expressive representations of local chemical environments. This hierarchical feature learning enables CGCNN to extract chemically meaningful descriptors that are subsequently used for accurate property prediction *[186]*. Zhou et al. employed CGCNN to identify high-performance cathode materials for zinc-ion batteries by combining data from the Materials Project and AFLOW databases. Through large-scale screening of approximately 700,000 inorganic compounds, they identified more than 70 candidates exhibiting both high operating voltage and high specific capacity, demonstrating the effectiveness of CGCNN for accelerating energy materials discovery. In another study, Wang et al. constructed a borate bandgap database using text-mining techniques applied to experimental literature and applied CGCNN to predict electronic band gaps. Their model achieved strong agreement with experiments, with an average absolute error of 0.40 eV, and was integrated into a web-based platform that enables direct bandgap prediction from CIF files. Additionally, Ahmad et al. used CGCNN to screen solid electrolyte materials capable of suppressing lithium dendrite growth. By employing bulk and shear moduli as screening metrics, they evaluated nearly 13,000 bulk materials and over 15,000 interfacial systems, identifying more than 20 promising interfaces and six solid electrolytes with strong dendrite suppression potential *[187]*. Overall, CGCNN consistently outperforms conventional ML models and traditional computational approaches in terms of predictive accuracy and generalization. This advantage stems from its ability to learn directly from atomic connectivity, enabling the development of transferable representations across diverse material classes. Furthermore, CGCNN offers a degree of interpretability by linking learned features to

local atomic environments, providing valuable insights into structure–property relationships and supporting rational materials design.

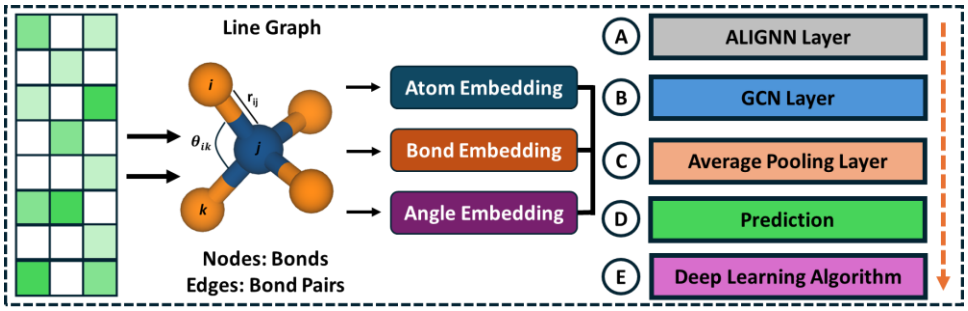


Figure 1.15: Architecture of the atomistic line graph neural network model.

1.12.12.2 Atomistic Line Graph Neural Network (ALIGNN)

ALIGNN [184] extends conventional graph neural network frameworks by introducing a line graph representation, as illustrated in Figure 1.15. In this approach, the material structure is first represented as an atomic graph, and a corresponding line graph is constructed where each node represents an interatomic bond from the atomic graph. The edges of the line graph capture relationships between bonds, typically associated with atomic triplets or bonded pairs. Information is exchanged through alternating convolution operations between the atomic and line graphs, enabling simultaneous learning of atomic-level and bond-level features. By explicitly incorporating both bond lengths and bond angles, ALIGNN captures detailed geometric information that is often missed by simpler graph representations, leading to improved prediction accuracy for material properties. In addition, ALIGNN employs a multi-head self-attention mechanism to model long-range interactions within materials. By dividing the feature space into multiple attention heads, the model learns diverse attention patterns, enhancing robustness and representational capacity.

An extended version, ALIGNN-d, was introduced by Hsu et al. *[188]*, which incorporates dihedral angles in addition to bond angles. This extension enables the model to encode full three-dimensional geometric information while maintaining memory efficiency, making it particularly suitable for complex molecular and crystalline systems. Gurunathan et al.

[189] applied ALIGNN to predict phonon density of states and derived thermodynamic properties, including heat capacity, vibrational entropy, and phonon–isotope scattering rates, achieving high accuracy. Similarly, Kaundinya et al. *[189]* demonstrated the use of ALIGNN for electronic density of states prediction using two complementary graph-based representations, enabling efficient high-throughput DOS screening. Hsu et al. [188] further showed that ALIGNN-d can accurately predict infrared absorption spectra for different Cu(II) aqua complex configurations. Beyond property prediction, Choudhary et al. *[190]* developed ALIGNN-based interatomic force fields capable of modeling solids containing arbitrary combinations of 89 elements. Their model was benchmarked against DFT for lattice constants, formation energies, and energy–volume curves across a wide range of materials. In another development, an ALIGNN-based framework incorporating deep transfer learning was proposed to enable property prediction across both three-dimensional and two-dimensional systems, covering inorganic and organic materials *[191]*. Despite its strong performance, ALIGNN has several limitations. Training can be computationally expensive, and model performance is sensitive to hyperparameter choices, often requiring extensive tuning. Interpretability remains limited, making it difficult to directly associate learned features with physical insights. In addition, predictive performance can be dataset-dependent, highlighting the need for broader validation. Ongoing research efforts aim to improve the efficiency, robustness, and interpretability of ALIGNN-based models.

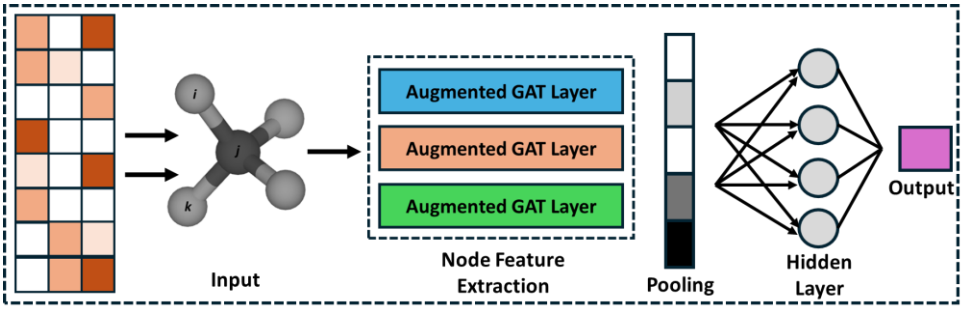


Figure 1.16: Architecture of the global attention graph neural network model.

1.12.12.3 Global Attention Graph Neural Network (GATGNN)

Traditional graph neural networks primarily emphasize local message passing, which restricts their ability to explicitly model long-range interactions. To overcome this limitation, the global attention graph neural network (GATGNN) was introduced as a hybrid framework that integrates graph convolution with a global attention mechanism. As illustrated in Figure 1.16, GATGNN consists of multiple augmented graph attention (AGAT) layers followed by a global attention layer. The AGAT layers focus on learning local atomic interactions by assigning adaptive weights to neighboring nodes during feature aggregation. The global attention layer enables the model to capture interactions between all nodes in the graph, making GATGNN particularly effective for irregular structures where the importance of nodes varies significantly. Through multi-layer convolution and multi-head attention operations, GATGNN jointly learns node and edge representations while dynamically quantifying interaction importance. This design substantially improves predictive accuracy and generalization. However, the global attention representation alone is not sufficient for direct node-level classification tasks and requires additional processing.

To address this issue, Sadam et al. proposed DeeperGATGNN *[183]*, which extends the original architecture by incorporating differentiable group normalization and skip connections to alleviate over-smoothing in deep GNNs. DeeperGATGNN achieves state-of-the-art performance across multiple benchmark datasets, often requiring minimal hyperparameter tuning. The architectural principles introduced in DeeperGATGNN also provide a general strategy for improving other deep GNN models. Nevertheless, its scalability is lower than that of MEGNet, particularly for very large systems. Louis et al. *[170]* demonstrated that GATGNN outperforms CGCNN in predicting formation energies, elastic moduli, and electronic band structures, consistently yielding lower mean absolute errors.

Kong et al. further showed that GATGNN can accurately predict phonon and electronic density of states using crystal structures from the Materials Project *[192]*. DeeperGATGNN has also been applied to screen piezoelectric materials within the ABC_3 family *[193]*. Additional studies have used GATGNN-based models to predict vibrational frequencies *[194]*, dielectric constants *[195]*, and other electronic properties directly from crystal structural data.

1.13 References

1. Lewis, N. S. (2016), Research opportunities to advance solar energy utilization, *Science*, 351, aad1920 (DOI:10.1126/science.aad1920)
2. She, Z. W., Kibsgaard, J., Dickens, C. F., Chorkendorff, I., Nørskov, J. K., Jaramillo, T. F. (2017), Combining theory and experiment in electrocatalysis: Insights into materials design, *Science*, 355, eaad4998 (DOI:10.1126/science.aad4998)
3. Huang, Z. F., Wang, J., Peng, Y., Jung, C. Y., Fisher, A., Wang, X. (2017), Design of efficient bifunctional oxygen reduction/evolution electrocatalyst: Recent advances and perspectives, *Adv. Energy Mater.*, 7, 1700544 (DOI:10.1002/aenm.201700544)
4. Shinde, S. S., Lee, C. H., Jung, J. Y., Wagh, N. K., Kim, S. H., Kim, D. H., Lin, C., Lee, S. U., Lee, J. H. (2019), Unveiling dual-linkage 3D hexaaminobenzene metal–organic frameworks towards long-lasting advanced reversible Zn–air batteries, *Energy Environ. Sci.*, 12, 2 (DOI:10.1039/c8ee02679c)
5. Shi, Z., Li, J., Wang, Y., Liu, S., Zhu, J., Yang, J., Wang, X., Ni, J., Jiang, Z., Zhang, L., Wang, Y., Liu, C., Xing, W., Ge, J. (2023), Customized reaction route for ruthenium oxide towards stabilized water oxidation in high-performance PEM electrolyzers, *Nat. Commun.*, 14, 36380 (DOI:10.1038/s41467-023-36380-9)
6. Wang, Y., Wang, D., Li, Y. (2021), A fundamental comprehension and recent progress in advanced Pt-based ORR nanocatalysts, *SmartMat*, 2, e1023 (DOI:10.1002/smm2.1023)

7. Zhang, L., Roling, L. T., Wang, X., Vara, M., Chi, M., Liu, J., Choi, S. I., Park, J., Herron, J. A., Xie, Z., Mavrikakis, M., Xia, Y. (2015), Platinum-based nanocages with subnanometer-thick walls and well-defined, controllable facets, *Science*, 349, 6246 (DOI:10.1126/science.aab0801)
8. Shu, Q., Zhang, J., Hu, B., Deng, X., Yuan, J., Ran, R., Zhou, W., Shao, Z. (2022), Rational design of a high-durability Pt-based ORR catalyst supported on Mn/N codoped carbon sheets for PEMFCs, *Energy Fuels*, 36, 3 (DOI:10.1021/acs.energyfuels.1c04306)
9. Peterson, E. J., DeLaRiva, A. T., Lin, S., Johnson, R. S., Guo, H., Miller, J. T., Kwak, J. H., Peden, C. H. F., Kiefer, B., Allard, L. F., Ribeiro, F. H., Datye, A. K. (2014), Low-temperature carbon monoxide oxidation catalysed by regenerable atomically dispersed palladium on alumina, *Nat. Commun.*, 5, 5885 (DOI:10.1038/ncomms5885)
10. Zhai, H., Alexandrova, A. N. (2017), Fluxionality of catalytic clusters: When it matters and how to address it, *ACS Catal.*, 7, 3 (DOI:10.1021/acscatal.6b03243)
11. Zandkarimi, B., Alexandrova, A. N. (2020), Can fluxionality of subnanometer cluster catalysts solely cause non-Arrhenius behavior in catalysis?, *J. Phys. Chem. C*, 124, 36 (DOI:10.1021/acs.jpcc.0c04136)
12. Zandkarimi, B., Alexandrova, A. N. (2019), Dynamics of subnanometer Pt clusters can break the scaling relationships in catalysis, *J. Phys. Chem. Lett.*, 10, 3 (DOI:10.1021/acs.jpclett.8b03680)
13. Ha, M. A., Baxter, E. T., Cass, A. C., Anderson, S. L., Alexandrova, A. N. (2017), Boron switch for selectivity of catalytic dehydrogenation on size-selected Pt clusters on Al_2O_3, *J. Am. Chem. Soc.*, 139, 33 (DOI:10.1021/jacs.7b05894)
14. Baxter, E. T., Ha, M. A., Cass, A. C., Alexandrova, A. N., Anderson, S. L. (2017), Ethylene dehydrogenation on $Pt_{4,7,8}$ clusters on Al_2O_3:

Strong cluster size dependence linked to preferred catalyst morphologies, *ACS Catal.*, 7, 5 (DOI:10.1021/acscatal.7b00409)

15. Lubitz, W., Tumas, W. (2007), Hydrogen: An overview, *Chem. Rev.*, 107, 3900–3903 (DOI:10.1021/cr050200z)
16. Albarbar, A., Alrweq, M. (2017), Proton exchange membrane fuel cells: Design, modelling and performance assessment techniques, *Springer*, (DOI:10.1007/978-3-319-70727-3)
17. Ryan, O'H., Cha, S.-W., Colella, W., Prinz, F. B. (2016), *Fuel Cell Fundamentals*, 3rd ed., Wiley
18. Zandkarimi, B., Alexandrova, A. N. (2019), Surface-supported cluster catalysis: Ensembles of metastable states run the show, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 9, e1420 (DOI:10.1002/wcms.1420)
19. Haruta, M., Kobayashi, T., Sano, H., Yamada, N. (1987), Novel gold catalysts for the oxidation of carbon monoxide at a temperature far below 0 °C, *Chem. Lett.*, 16, 405–408 (DOI:10.1246/cl.1987.405)
20. Hammer, B., Nørskov, J. K. (1995), Why gold is the noblest of all the metals, *Nature*, 376, 238–240 (DOI:10.1038/376238a0)
21. Hammer, B., Nørskov, J. K. (1995), Electronic factors determining the reactivity of metal surfaces, *Surf. Sci.*, 343, 211–220 (DOI:10.1016/0039-6028(96)80007-0)
22. Heiz, U., Sanchez, A., Abbet, S., Schneider, W. D. (1999), Catalytic oxidation of carbon monoxide on monodispersed platinum clusters: Each atom counts, *J. Am. Chem. Soc.*, 121, 3214–3217 (DOI:10.1021/ja983616l)
23. Crampton, A. S., Rötzer, M. D., Ridge, C. J., Schweinberger, F. F., Heiz, U., Yoon, B., Landman, U. (2016), Structure sensitivity in the nonscalable regime explored via catalysed ethylene hydrogenation on supported platinum nanoclusters, *Nat. Commun.*, 7, 10389 (DOI:10.1038/ncomms10389)

24. Keppeler, M., Bräuning, G., Radhakrishnan, S. G., Liu, X., Jensen, C., Roduner, E. (2016), Reactivity of diatomics and of ethylene on zeolite-supported 13-atom platinum nanoclusters, *Catal. Sci. Technol.*, 6, 18 (DOI:10.1039/c6cy00182c)
25. Vajda, S., Pellin, M. J., Greeley, J. P., Marshall, C. L., Curtiss, L. A., Ballentine, G. A., Elam, J. W., Catillon-Mucherie, S., Redfern, P. C., Mehmood, F., Zapol, P. (2009), Subnanometre platinum clusters as highly active and selective catalysts for the oxidative dehydrogenation of propane, *Nat. Mater.*, 8, 213–216 (DOI:10.1038/nmat2384)
26. Watanabe, Y., Wu, X., Hirata, H., Isomura, N. (2011), Size-dependent catalytic activity and geometries of size-selected Pt clusters on TiO_2(110) surfaces, *Catal. Sci. Technol.*, 1, 8 (DOI:10.1039/c1cy00204j)
27. Herzing, A. A., Kiely, C. J., Carley, A. F., Landon, P., Hutchings, G. J. (2008), Identification of active gold nanoclusters on iron oxide supports for CO oxidation, *Science*, 321, 5894 (DOI:10.1126/science.1159639)
28. Lee, S., Molina, L. M., López, M. J., Alonso, J. A., Hammer, B., Lee, B., Seifert, S., Winans, R. E., Elam, J. W., Pellin, M. J., Vajda, S. (2009), Selective propene epoxidation on immobilized Au_{6-10} clusters: The effect of hydrogen and water on activity and selectivity, *Angew. Chem. Int. Ed.*, 48, 8 (DOI:10.1002/anie.200804154)
29. Lei, L., Wu, Z., Wang, R., Qin, Z., Chen, C., Liu, Y., Wang, G., Fan, W., Wang, J. (2017), Controllable decoration of palladium sub-nanoclusters on reduced graphene oxide with superior catalytic performance in selective oxidation of alcohols, *Catal. Sci. Technol.*, 7, 23 (DOI:10.1039/c7cy01732d)
30. Ghosh, P., Farnesi Camellone, M., Fabris, S. (2013), Fluxionality of Au clusters at ceria surfaces during CO oxidation: Relationships among reactivity, size, cohesion, and surface defects from DFT simulations, *J. Phys. Chem. Lett.*, 4, 14 (DOI:10.1021/jz4009079)

31. Hook, A., Massa, J. D., Celik, F. E. (2016), Effect of tin coverage on selectivity for ethane dehydrogenation over platinum–tin alloys, *J. Phys. Chem. C*, 120, 48 (DOI:10.1021/acs.jpcc.6b08407)
32. Jimenez-Izal, E., Alexandrova, A. N. (2018), Computational design of clusters for catalysis, *Annu. Rev. Phys. Chem.*, 69 (DOI:10.1146/annurev-physchem-050317-014216)
33. Pidko, E. A. (2017), Toward the balance between the reductionist and systems approaches in computational catalysis: Model versus method accuracy for the description of catalytic systems, *ACS Catal.*, 7 (DOI:10.1021/acscatal.7b00290)
34. Tauster, S. J., Fung, S. C., Garten, R. L. (1978), Strong metal–support interactions. Group 8 noble metals supported on TiO_2, *J. Am. Chem. Soc.*, 100, 1 (DOI:10.1021/ja00469a029)
35. Tang, H., Su, Y., Guo, Y., Zhang, L., Li, T., Zang, K., Liu, F., Li, L., Luo, J., Qiao, B., Wang, J. (2018), Oxidative strong metal–support interactions (OMSI) of supported platinum-group metal catalysts, *Chem. Sci.*, 9, 32 (DOI:10.1039/c8sc01392f)
36. Ren, Z., Liu, N., Chen, B., Li, J., Mei, D. (2018), Theoretical investigation of the structural stabilities of ceria surfaces and supported metal nanoclusters in vapor and aqueous phases, *J. Phys. Chem. C*, 122, 9 (DOI:10.1021/acs.jpcc.7b10208)
37. Keller, D. E., Koningsberger, D. C., Weckhuysen, B. M. (2006), Molecular structure of a supported VO_4 cluster and its interfacial geometry as a function of the SiO_2, Nb_2O_5, and ZrO_2 support, *J. Phys. Chem. B*, 110, 29 (DOI:10.1021/jp060749h)
38. Sasahara, A., Pang, C. L., Onishi, H. (2006), Probe microscope observation of platinum atoms deposited on the TiO_2(110)-(1 × 1) surface, *J. Phys. Chem. B*, 110, 27 (DOI:10.1021/jp062000c)
39. Campisi, S., Chan-Thaw, C. E., Villa, A. (2018), Understanding heteroatom-mediated metal–support interactions in functionalized

carbons: A perspective review, *Appl. Sci.*, 8 (DOI:10.3390/app8071159)

40. Call, S. T., Zubarev, D. Y., Boldyrev, A. I. (2007), Global minimum structure searches via particle swarm optimization, *J. Comput. Chem.*, 28, 7 (DOI:10.1002/jcc.20621)
41. Avendaño-Franco, G., Romero, A. H. (2016), Firefly algorithm for structural search, *J. Chem. Theory Comput.*, 12, 7 (DOI:10.1021/acs.jctc.5b01157)
42. Pickard, C. J., Needs, R. J. (2011), Ab initio random structure searching, *J. Phys.: Condens. Matter*, 23 (DOI:10.1088/0953-8984/23/5/053201)
43. Pickard, C. J., Needs, R. J. (2009), Structures at high pressure from random searching, *Phys. Status Solidi B*, 246, 3 (DOI:10.1002/pssb.200880546)
44. Alexandrova, A. N. (2010), $H\cdot(H_2O)_n$ clusters: Microsolvation of the hydrogen atom via molecular ab initio gradient embedded genetic algorithm (GEGA), *J. Phys. Chem. A*, 114, 48 (DOI:10.1021/jp1092543)
45. Alexandrova, A. N., Boldyrev, A. I. (2005), Search for the $Li_n^{0/+1/-1}$ (n = 5–7) lowest-energy structures using the ab initio gradient embedded genetic algorithm (GEGA): Elucidation of the chemical bonding in the lithium clusters, *J. Chem. Theory Comput.*, 1, 4 (DOI:10.1021/ct050093g)
46. Kanters, R. P. F., Donald, K. J. (2014), CLUSTER: Searching for unique low energy minima of structures using a novel implementation of a genetic algorithm, *J. Chem. Theory Comput.*, 10, 12 (DOI:10.1021/ct500744k)
47. Davis, J. B. A., Shayeghi, A., Horswell, S. L., Johnston, R. L. (2015), The Birmingham parallel genetic algorithm and its application to the direct DFT global optimisation of Ir_n (n = 10–20) clusters, *Nanoscale*, 7, 33 (DOI:10.1039/c5nr03774c)

48. Deaven, D. M., Ho, K. M. (1995), Molecular geometry optimization with a genetic algorithm, *Phys. Rev. Lett.*, 75, 2 (DOI:10.1103/PhysRevLett.75.288)
49. Wales, D. J., Doye, J. P. K. (1997), Global optimization by basin-hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms, *J. Phys. Chem. A*, 101, 28 (DOI:10.1021/jp970984n)
50. Oganov, A. R. (2010), *Modern methods of crystal structure prediction*, Wiley-VCH, Weinheim (DOI:10.1002/9783527632831)
51. Zhai, H. J., Zhao, Y. F., Li, W. L., Chen, Q., Bai, H., Hu, H. S., Piazza, Z. A., Tian, W. J., Lu, H. G., Wu, Y. B., Mu, Y. W., Wei, G. F., Liu, Z. P., Li, J., Li, S. D., Wang, L. S. (2014), Observation of an all-boron fullerene, *Nat. Chem.*, 6, 8 (DOI:10.1038/nchem.1999)
52. Wang, J., Ma, L., Zhao, J., Jackson, K. A. (2009), Structural growth behavior and polarizability of Cd_nTe_n (n = 1–14) clusters, *J. Chem. Phys.*, 130, 21 (DOI:10.1063/1.3147519)
53. Heiles, S., Johnston, R. L. (2013), Global optimization of clusters using electronic structure methods, *Int. J. Quantum Chem.* (DOI:10.1002/qua.24462)
54. Ferrando, R., Fortunelli, A., Rossi, G. (2005), Quantum effects on the structure of pure and binary metallic nanoclusters, *Phys. Rev. B*, 72, 8 (DOI:10.1103/PhysRevB.72.085449)
55. Zhai, H., Alexandrova, A. N. (2016), Ensemble-average representation of Pt clusters under catalytic conditions accessed through GPU-accelerated deep neural network fitting global optimization, *J. Chem. Theory Comput.*, 12, 12 (DOI:10.1021/acs.jctc.6b00994)
56. Zhai, H., Ha, M. A., Alexandrova, A. N. (2015), AFFCK: Adaptive force-field-assisted ab initio coalescence kick method for global minimum search, *J. Chem. Theory Comput.*, 11, 5 (DOI:10.1021/acs.jctc.5b00065)

57. Ouyang, R., Xie, Y., Jiang, D. E. (2015), Global minimization of gold clusters by combining neural network potentials and the basin-hopping method, *Nanoscale*, 7, 36 (DOI:10.1039/c5nr03903g)
58. Pittaway, F., Paz-Borbón, L. O., Johnston, R. L., Arslan, H., Ferrando, R., Mottet, C., Barcaro, G., Fortunelli, A. (2009), Theoretical studies of palladium–gold nanoclusters: Pd–Au clusters with up to 50 atoms, *J. Phys. Chem. C*, 113, 21 (DOI:10.1021/jp9006075)
59. Sattler, J. J. H. B., Ruiz-Martinez, J., Santillan-Jimenez, E., Weckhuysen, B. M. (2014), Catalytic dehydrogenation of light alkanes on metals and metal oxides, *Chem. Rev.* (DOI:10.1021/cr5002436)
60. Jimenez-Izal, E., Zhai, H., Liu, J. Y., Alexandrova, A. N. (2018), Nanoalloying MgO-deposited Pt clusters with Si to control the selectivity of alkane dehydrogenation, *ACS Catal.*, 8, 9 (DOI:10.1021/acscatal.8b02443)
61. Hook, A., Celik, F. E. (2017), Predicting selectivity for ethane dehydrogenation and coke formation pathways over model Pt–M surface alloys with ab initio and scaling methods, *J. Phys. Chem. C*, 121, 33 (DOI:10.1021/acs.jpcc.7b03789)
62. Senftle, T. P., Meyer, R. J., Janik, M. J., Van Duin, A. C. T. (2013), Development of a ReaxFF potential for Pd/O and application to palladium oxide formation, *J. Chem. Phys.*, 139, 4 (DOI:10.1063/1.4815820)
63. Senftle, T. P., Van Duin, A. C. T., Janik, M. J. (2017), Methane activation at the Pd/CeO_2 interface, *ACS Catal.*, 7, 1 (DOI:10.1021/acscatal.6b02447)
64. Nilsson, J., Carlsson, P. A., Fouladvand, S., Martin, N. M., Gustafson, J., Newton, M. A., Lundgren, E., Grönbeck, H., Skoglundh, M. (2015), Chemistry of supported palladium nanoparticles during methane oxidation, *ACS Catal.*, 5, 4 (DOI:10.1021/cs502036d)

65. Farrauto, R. J., Lampert, J. K., Hobson, M. C., Waterman, E. M. (1995), Thermal decomposition and reformation of PdO catalysts: Support effects, *Appl. Catal. B*, 6, 3 (DOI:10.1016/0926-3373(95)00015-1)
66. Timoshenko, J., Halder, A., Yang, B., Seifert, S., Pellin, M. J., Vajda, S., Frenkel, A. I. (2018), Subnanometer substructures in nanoassemblies formed from clusters under a reactive atmosphere revealed using machine learning, *J. Phys. Chem. C*, 122, 37 (DOI:10.1021/acs.jpcc.8b07952)
67. Baldyga, L. M., Blavo, S. O., Kuo, C. H., Tsung, C. K., Kuhn, J. N. (2012), Size-dependent sulfur poisoning of silica-supported monodisperse Pt nanoparticle hydrogenation catalysts, *ACS Catal.*, 2, 12 (DOI:10.1021/cs300625m)
68. Kolpin, A., Jones, G., Jones, S., Zheng, W., Cookson, J., York, A. P. E., Collier, P. J., Tsang, S. C. E. (2017), Quantitative differences in sulfur poisoning phenomena over ruthenium and palladium: An attempt to deconvolute geometric and electronic poisoning effects using model catalysts, *ACS Catal.*, 7, 1 (DOI:10.1021/acscatal.6b02765)
69. Düll, F., Späth, F., Bauer, U., Bachmann, P., Steinhauer, J., Steinrück, H. P., Papp, C. (2018), Reactivity of CO on sulfur-passivated graphene-supported platinum nanocluster arrays, *J. Phys. Chem. C*, 122, 28 (DOI:10.1021/acs.jpcc.8b01589)
70. Taccardi, N., Grabau, M., Debuschewitz, J., Distaso, M., Brandl, M., Hock, R., Maier, F., Papp, C., Erhard, J., Neiss, C., Peukert, W., Görling, A., Steinrück, H. P., Wasserscheid, P. (2017), Gallium-rich Pd–Ga phases as supported liquid metal catalysts, *Nat. Chem.*, 9, 9 (DOI:10.1038/NCHEM.2822)
71. Dadras, J., Shen, L., Alexandrova, A. (2015), Pt–Zn clusters on stoichiometric MgO(100) and TiO_2(110): Dramatically different sintering behavior, *J. Phys. Chem. C*, 119, 11 (DOI:10.1021/jp512277x)

72. Dadras, J., Jimenez-Izal, E., Alexandrova, A. N. (2015), Alloying Pt sub-nanoclusters with boron: Sintering preventative and coke antagonist?, *ACS Catal.*, 5, 10 (DOI:10.1021/acscatal.5b01513)
73. Zhao, G., Wu, X. P., Chai, R., Zhang, Q., Gong, X. Q., Huang, J., Lu, Y. (2015), Tailoring nano-catalysts: Turning gold nanoparticles on bulk metal oxides to inverse nano-metal oxides on large gold particles, *Chem. Commun.*, 51, 27 (DOI:10.1039/c5cc00016e)
74. Su, Y. Q., Liu, J. X., Filot, I. A. W., Hensen, E. J. M. (2017), Theoretical study of ripening mechanisms of Pd clusters on ceria, *Chem. Mater.*, 29, 21 (DOI:10.1021/acs.chemmater.7b03555)
75. Ha, M. A., Dadras, J., Alexandrova, A. (2014), Rutile-deposited Pt–Pd clusters: A hypothesis regarding the stability at 50/50 ratio, *ACS Catal.*, 4, 10 (DOI:10.1021/cs5011426)
76. Wettergren, K., Schweinberger, F. F., Deiana, D., Ridge, C. J., Crampton, A. S., Rötzer, M. D., Hansen, T. W., Zhdanov, V. P., Heiz, U., Langhammer, C. (2014), High sintering resistance of size-selected platinum cluster catalysts by suppressed Ostwald ripening, *Nano Lett.*, 14, 10 (DOI:10.1021/nl502686u)
77. Häkkinen, H., Abbet, S., Sanchez, A., Heiz, U., Landman, U. (2003), Structural, electronic, and impurity-doping effects in nanoscale chemistry: Supported gold nanoclusters, *Angew. Chem., Int. Ed.*, 42, 11 (DOI:10.1002/anie.200390334)
78. Eichler, A., Hafner, J. (1999), Reaction channels for the catalytic oxidation of CO on Pt(111), *Surf. Sci.*, 433 (DOI:10.1016/S0039-6028(99)00061-8)
79. Allers, K. H., Pfnür, H., Feulner, P., Menzel, D. (1994), Fast reaction products from the oxidation of CO on Pt(111): Angular and velocity distributions of the CO_2 product molecules, *J. Chem. Phys.*, 100, 5 (DOI:10.1063/1.466332)

80. Zhai, H., Alexandrova, A. N. (2018), Local fluxionality of surface-deposited cluster catalysts: The case of Pt_7 on Al_2O_3, *J. Phys. Chem. Lett.*, 9, 7 (DOI:10.1021/acs.jpclett.8b00379)
81. Sun, G., Sautet, P. (2018), Metastable structures in cluster catalysis from first-principles: Structural ensemble in reaction conditions and metastability-triggered reactivity, *J. Am. Chem. Soc.*, 140, 8 (DOI:10.1021/jacs.7b11239)
82. Pisliakov, A. V., Cao, J., Kamerlin, S. C. L., Warshel, A. (2009), Enzyme millisecond conformational dynamics do not catalyze the chemical step, *Proc. Natl. Acad. Sci. U.S.A.*, 106, 41 (DOI:10.1073/pnas.0909150106)
83. Yang, C. T., Wood, B. C., Bhethanabotla, V. R., Joseph, B. (2015), The effect of the morphology of supported subnanometer Pt clusters on the first and key step of CO_2 photoreduction, *Phys. Chem. Chem. Phys.*, 17, 38 (DOI:10.1039/c5cp03674g)
84. Liu, L., Liu, Z., Sun, H., Zhao, X. (2017), Morphological effects of Au_{13} clusters on the adsorption of CO_2 over anatase TiO_2(101), *Appl. Surf. Sci.*, 399 (DOI:10.1016/j.apsusc.2016.12.089)
85. Fernández, E. M., Moses, P. G., Toftelund, A., Hansen, H. A., Martínez, J. I., Abild-Pedersen, F., Kleis, J., Hinnemann, B., Rossmeisl, J., Bligaard, T., Nørskov, J. K. (2008), Scaling relationships for adsorption energies on transition metal oxide, sulfide, and nitride surfaces, *Angew. Chem., Int. Ed.*, 47, 25 (DOI:10.1002/anie.200705739)
86. Montemore, M. M., Medlin, J. W. (2014), Scaling relations between adsorption energies for computational screening and design of catalysts, *Catal. Sci. Technol.* (DOI:10.1039/c4cy00335g)
87. Greeley, J. (2016), Theoretical heterogeneous catalysis: Scaling relationships and computational catalyst design, *Annu. Rev. Chem. Biomol. Eng.* (DOI:10.1146/annurev-chembioeng-080615-034413)

88. Zhang, Z., Jimenez-Izal, E., Hermans, I., Alexandrova, A. N. (2019), Dynamic phase diagram of the catalytic surface of hexagonal boron nitride under conditions of oxidative dehydrogenation of propane, *J. Phys. Chem. Lett.*, 10, 1 (DOI:10.1021/acs.jpclett.8b03373)
89. Grant, J. T., Carrero, C. A., Goeltl, F., Venegas, J., Mueller, P., Burt, S. P., Specht, S. E., McDermott, W. P., Chieregato, A., Hermans, I. (2016), Selective oxidative dehydrogenation of propane to propene using boron nitride catalysts, *Science*, 354, 6319 (DOI:10.1126/science.aaf7885)
90. Love, A. M., Thomas, B., Specht, S. E., Hanrahan, M. P., Venegas, J. M., Burt, S. P., Grant, J. T., Cendejas, M. C., McDermott, W. P., Rossini, A. J., Hermans, I. (2019), Probing the transformation of boron nitride catalysts under oxidative dehydrogenation conditions, *J. Am. Chem. Soc.*, 141, 1 (DOI:10.1021/jacs.8b08165)
91. Tao, F., Grass, M. E., Zhang, Y., Butcher, D. R., Renzas, J. R., Liu, Z., Chung, J. Y., Mun, B. S., Salmeron, M., Somorjai, G. A. (2008), Reaction-driven restructuring of Rh–Pd and Pt–Pd core–shell nanoparticles, *Science*, 322, 5903 (DOI:10.1126/science.1164170)
92. Chen, Z., Wang, H., Su, N. Q., Duan, S., Shen, T., Xu, X. (2018), Beyond mean-field microkinetics: Toward accurate and efficient theoretical modeling in heterogeneous catalysis, *ACS Catal.*, 8, 7 (DOI:10.1021/acscatal.8b00943)
93. Born, M., Oppenheimer, R. (1927), Zur quantentheorie der molekeln, *Ann. Phys.*, 389, 20 (DOI:10.1002/andp.19273892002)
94. Nesbet, R. K. (2002), Thomas–Fermi theory revisited, *Int. J. Quantum Chem.*, 90, 1 (DOI:10.1002/qua.10061)
95. Thomas, L. H. (1927), The calculation of atomic fields, *Math. Proc. Cambridge Philos. Soc.*, 23, 5 (DOI:10.1017/S0305004100011683)
96. Rajagopal, A. K., Callaway, J. (1973), Inhomogeneous electron gas, *Phys. Rev. B*, 7, 5 (DOI:10.1103/PhysRevB.7.1912)

97. Ceperley, D. M., Alder, B. J. (1980), Ground state of the electron gas by a stochastic method, *Phys. Rev. Lett.*, 45, 7 (DOI:10.1103/PhysRevLett.45.566)
98. Zupan, A., Blaha, P., Schwarz, K. (1998), Pressure-induced phase transitions in solid Si and Fe: Performance of local-spin-density and generalized-gradient-approximation density functionals, *Phys. Rev. B*, 58, 17 (DOI:10.1103/PhysRevB.58.11266)
99. Grossman, J. C., Mitas, L., Raghavachari, K. (1995), Structure and stability of molecular carbon: Importance of electron correlation, *Phys. Rev. Lett.*, 75, 21 (DOI:10.1103/PhysRevLett.75.3870)
100. Jones, R. O., Gunnarsson, O. (1989), The density functional formalism, its applications and prospects, *Rev. Mod. Phys.*, 61, 3 (DOI:10.1103/RevModPhys.61.689)
101. Perdew, J. P., Chevary, J. A., Vosko, S. H., Jackson, K. A., Pederson, M. R., Singh, D. J., Fiolhais, C. (1992), Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation, *Phys. Rev. B*, 46, 11 (DOI:10.1103/PhysRevB.46.6671)
102. Perdew, J. P., Yue, W. (1986), Accurate and simple density functional for the electronic exchange energy: Generalized gradient approximation, *Phys. Rev. B*, 33, 12 (DOI:10.1103/PhysRevB.33.8800)
103. Perdew, J. P., Wang, Y. (1992), Accurate and simple analytic representation of the electron-gas correlation energy, *Phys. Rev. B*, 45, 23 (DOI:10.1103/PhysRevB.45.13244)
104. Becke, A. D. (1988), Density-functional exchange-energy approximation with correct asymptotic behavior, *Phys. Rev. A*, 38, 6 (DOI:10.1103/PhysRevA.38.3098)
105. Becke, A. D. (1993), Density-functional thermochemistry. III. The role of exact exchange, *J. Chem. Phys.*, 98, 7 (DOI:10.1063/1.464913)

106. Lee, C., Yang, W., Parr, R. G. (1988), Development of the Colle–Salvetti correlation-energy formula into a functional of the electron density, *Phys. Rev. B*, 37, 2 (DOI:10.1103/PhysRevB.37.785)
107. Miehlich, B., Savin, A., Stoll, H., Preuss, H. (1989), Results obtained with the correlation energy density functionals of Becke and Lee, Yang and Parr, *Chem. Phys. Lett.*, 157, 3 (DOI:10.1016/0009-2614(89)87234-3)
108. Perdew, J. P., Burke, K., Ernzerhof, M. (1996), Generalized gradient approximation made simple, *Phys. Rev. Lett.*, 77, 18 (DOI:10.1103/PhysRevLett.77.3865)
109. Hammer, B., Scheffler, M., Jacobsen, K. W., Nørskov, J. K. (1994), Multidimensional potential energy surface for H_2 dissociation over Cu(111), *Phys. Rev. Lett.*, 73, 10 (DOI:10.1103/PhysRevLett.73.1400)
110. Penev, E., Kratzer, P., Scheffler, M. (1999), Effect of the cluster size in modeling the H_2 desorption and dissociative adsorption on Si(001), *J. Chem. Phys.*, 110, 8 (DOI:10.1063/1.478279)
111. Berland, K., Jiao, Y., Lee, J. H., Rangel, T., Neaton, J. B., Hyldgaard, P. (2017), Assessment of two hybrid van der Waals density functionals for covalent and noncovalent binding of molecules, *J. Chem. Phys.*, 146, 23 (DOI:10.1063/1.4986522)
112. Berland, K., Chakarova-Käck, S. D., Cooper, V. R., Langreth, D. C., Schröder, E. (2011), A van der Waals density functional study of adenine on graphene: Single-molecular adsorption and overlayer binding, *J. Phys.: Condens. Matter*, 23, 13 (DOI:10.1088/0953-8984/23/13/135001)
113. Dion, M., Rydberg, H., Schröder, E., Langreth, D. C., Lundqvist, B. I. (2004), van der Waals density functional for general geometries, *Phys. Rev. Lett.*, 92, 24 (DOI:10.1103/PhysRevLett.92.246401)
114. Bloch, F. (1929), Über die quantenmechanik der elektronen in kristallgittern, *Z. Phys.*, 52, 7–8 (DOI:10.1007/BF01339455)

115. Hamann, D. R., Schlüter, M., Chiang, C. (1979), Norm-conserving pseudopotentials, *Phys. Rev. Lett.*, 43, 20 (DOI:10.1103/PhysRevLett.43.1494)

116. Blöchl, P. E. (1994), Projector augmented-wave method, *Phys. Rev. B*, 50, 24

117. Joubert, D. (1999), From ultrasoft pseudopotentials to the projector augmented-wave method, *Phys. Rev. B*, 59, 3 (DOI:10.1103/PhysRevB.59.1758)

118. Luckey, G. W. (1957), Introduction to solid state physics, *J. Am. Chem. Soc.*, 79, 12 (DOI:10.1021/ja01569a094)

119. Sholl, D. S., Steckel, J. A. (2009), *Density functional theory: A practical introduction*, Wiley (DOI:10.1002/9780470447710)

120. Grimme, S. (2006), Semiempirical GGA-type density functional constructed with a long-range dispersion correction, *J. Comput. Chem.*, 27, 15 (DOI:10.1002/jcc.20495)

121. Grimme, S., Antony, J., Ehrlich, S., Krieg, H. (2010), A consistent and accurate *ab initio* parametrization of density functional dispersion correction (DFT-D) for the 94 elements H–Pu, *J. Chem. Phys.*, 132, 15 (DOI:10.1063/1.3382344)

122. Thonhauser, T., Cooper, V. R., Li, S., Puzder, A., Hyldgaard, P., Langreth, D. C. (2007), van der Waals density functional: Self-consistent potential and the nature of the van der Waals bond, *Phys. Rev. B*, 76, 12 (DOI:10.1103/PhysRevB.76.125112)

123. Lee, K., Murray, É. D., Kong, L., Lundqvist, B. I., Langreth, D. C. (2010), Higher-accuracy van der Waals density functional, *Phys. Rev. B*, 82, 8 (DOI:10.1103/PhysRevB.82.081101)

124. Klimeš, J., Bowler, D. R., Michaelides, A. (2010), Chemical accuracy for the van der Waals density functional, *J. Phys.: Condens. Matter*, 22, 2 (DOI:10.1088/0953-8984/22/2/022201)

125. Hubbard, J. (1963), Electron correlations in narrow energy bands, *Proc. R. Soc. Lond. A*, 276, 1365 (DOI:10.1098/rspa.1963.0204)

126. Henkelman, G., Jónsson, H. (2000), Improved tangent estimate in the nudged elastic band method, *J. Chem. Phys.*, 113, 22 (DOI:10.1063/1.1323224)
127. Henkelman, G., Uberuaga, B. P., Jónsson, H. (2000), Climbing image nudged elastic band method, *J. Chem. Phys.*, 113, 22 (DOI:10.1063/1.1329672)
128. Heyden, A., Bell, A. T., Keil, F. J. (2005), Efficient methods for finding transition states, *J. Chem. Phys.*, 123, 22 (DOI:10.1063/1.2104507)
129. Deringer, V. L., Tchougréeff, A. L., Dronskowski, R. (2011), Crystal orbital Hamilton population analysis, *J. Phys. Chem. A*, 115, 21 (DOI:10.1021/jp202489s)
130. Maintz, S., Deringer, V. L., Tchougréeff, A. L., Dronskowski, R. (2016), LOBSTER: A tool to extract chemical bonding, *J. Comput. Chem.*, 37, 11 (DOI:10.1002/jcc.24300)
131. Nørskov, J. K., Rossmeisl, J., Logadottir, A., Lindqvist, L., Kitchin, J. R., Bligaard, T., Jónsson, H. (2004), Origin of the overpotential for oxygen reduction, *J. Phys. Chem. B*, 108, 46 (DOI:10.1021/jp047349j)
132. Artrith, N., Butler, K. T., Coudert, F. X., Han, S., Isayev, O., Jain, A., Walsh, A. (2021), Best practices in machine learning for chemistry, *Nat. Chem.* (DOI:10.1038/s41557-021-00716-z)
133. Keith, J. A., Vassilev-Galindo, V., Cheng, B., Chmiela, S., Gastegger, M., Müller, K. R., Tkatchenko, A. (2021), Combining machine learning and computational chemistry, *Chem. Rev.* (DOI:10.1021/acs.chemrev.1c00107)
134. Pugliese, R., Regondi, S., Marini, R. (2021), Machine learning-based approach, *Data Sci. Manag.*, (DOI:10.1016/j.dsm.2021.12.002)
135. Sarker, I. H. (2021), Machine learning: Algorithms, applications and research directions, *SN Comput. Sci.* (DOI:10.1007/s42979-021-00592-x)

136. Sen, P. C., Hajra, M., Ghosh, M. (2020), Supervised classification algorithms, In *Adv. Intell. Syst. Comput.*, Vol. 937 (DOI:10.1007/978-981-13-7403-6_11)

137. Li, J. (2019), Regression and classification in supervised learning, In *ACM Int. Conf. Proc. Ser.* (DOI:10.1145/3366650.3366675)

138. Usama, M., Qadir, J., Raza, A., Arif, H., Yau, K. L. A., Elkhatib, Y., Hussain, A., Al-Fuqaha, A. (2019), Unsupervised machine learning for networking, *IEEE Access*, 7 (DOI:10.1109/ACCESS.2019.2916648)

139. Liu, H., Motoda, H. (1998), *Feature selection for knowledge discovery and data mining*, Springer (DOI:10.1007/978-1-4615-5689-3)

140. Bishop, C. M. (2007), *Pattern recognition and machine learning*, *J. Electron. Imaging*, 16, 4 (DOI:10.1117/1.2819119)

141. Dembczynski, K., Waegeman, W., Hüllermeier, E. (2012), An analysis of chaining in multi-label classification, In *Front. Artif. Intell. Appl.*, Vol. 242 (DOI:10.3233/978-1-61499-098-7-294)

142. Spearman, C. (1904), The proof and measurement of association between two things, *Am. J. Psychol.*, 15, 1 (DOI:10.2307/1412159)

143. Bergstra, J., Bengio, Y. (2012), Random search for hyper-parameter optimization, *J. Mach. Learn. Res.*, 13

144. Hutter, M., Zaffalon, M. (2005), Distribution of mutual information, *Comput. Stat. Data Anal.*, 48, 3 (DOI:10.1016/j.csda.2004.03.010)

145. Su, G. X. (2009), *Linear regression analysis: Theory and computing*, World Scientific (DOI:10.1142/6986)

146. LaValley, M. P. (2008), Logistic regression, *Circulation*, 117, 2395–2399 (DOI:10.1161/CIRCULATIONAHA.106.682658)

147. Nigsch, F., Bender, A., van Buuren, B., Tissen, J., Nigsch, E., Mitchell, J. B. O. (2006), Melting point prediction, *J. Chem. Inf. Model.*, 46, 6 (DOI:10.1021/ci060149f)

148. Nadaraya, E. A. (1964), On estimating regression, *Theory Probab. Appl.*, 9, 1 (DOI:10.1137/1109020)

149. Pisner, D. A., Schnyer, D. M. (2020), Support vector machine, In *Academic Press*

150. Breiman, L. (2001), Random forests, *Mach. Learn.*, 45, 1 (DOI:10.1023/A:1010933404324)

151. Hastie, T., Rosset, S., Zhu, J., Zou, H. (2009), Multi-class AdaBoost, *Stat. Interface*, 2, 3 (DOI:10.4310/sii.2009.v2.n3.a8)

152. Chen, T., Guestrin, C. (2016), XGBoost: A scalable tree boosting system, *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 13–17 August 2016 (DOI:10.1145/2939672.2939785)

153. Challis, E., Hurley, P., Serra, L., Bozzali, M., Oliver, S., Cercignani, M. (2015), Gaussian process classification of Alzheimer's disease and mild cognitive impairment from resting-state fMRI, *Neuroimage*, 112 (DOI:10.1016/j.neuroimage.2015.02.037)

154. Myles, A. J., Feudale, R. N., Liu, Y., Woody, N. A., Brown, S. D. (2004), An introduction to decision tree modeling, *J. Chemom.*, (DOI:10.1002/cem.873)

155. Geurts, P., Ernst, D., Wehenkel, L. (2006), Extremely randomized trees, *Mach. Learn.*, 63, 1 (DOI:10.1007/s10994-006-6226-1)

156. Glantz, S. A., Slinker, B. K., Neilands, T. B. (2016), *Primer of applied regression and analysis of variance*, 3rd ed.; McGraw-Hill

157. Soper, D. S. (2021), Greed is good: Rapid hyperparameter optimization and model selection using greedy k-fold cross validation, *Electronics*, 10, 16 (DOI:10.3390/electronics10161973)

158. Longo, L., Brcic, M., Cabitza, F., Choi, J., Confalonieri, R., Del Ser, J., Guidotti, R., Hayashi, Y., Herrera, F., Holzinger, A., Jiang, R., Khosravi, H., Lecue, F., Malgieri, G., Páez, A., Samek, W., Schneider, J., Speith, T., Stumpf, S. (2024), Explainable artificial intelligence (XAI) 2.0: A manifesto of open challenges and interdisciplinary research directions, *Inf. Fusion* (DOI:10.1016/j.inffus.2024.102301)

159. Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S. I. (2020), From local

explanations to global understanding with explainable AI for trees, *Nat. Mach. Intell.*, 2, 1 (DOI:10.1038/s42256-019-0138-9)

160. Addin, O., Sapuan, S. M., Mahdi, E., Othman, M. (2007), A naïve-Bayes classifier for damage detection in engineering materials, *Mater. Des.*, 28, 8 (DOI:10.1016/j.matdes.2006.07.018)
161. Kusano, M., Miyazaki, S., Watanabe, M., Kishimoto, S., Bulgarevich, D. S., Ono, Y., Yumoto, A. (2020), Tensile properties prediction by multiple linear regression analysis for selective laser melted and post heat-treated Ti-6Al-4V with microstructural quantification, *Mater. Sci. Eng. A*, 787 (DOI:10.1016/j.msea.2020.139549)
162. Wang, Y., Wagner, N., Rondinelli, J. M. (2019), Symbolic regression in materials science, *MRS Commun.* (DOI:10.1557/mrc.2019.85)
163. Rajan, A. C., Mishra, A., Satsangi, S., Vaish, R., Mizuseki, H., Lee, K. R., Singh, A. K. (2018), Machine-learning-assisted accurate band gap predictions of functionalized MXene, *Chem. Mater.*, 30, 12 (DOI:10.1021/acs.chemmater.8b00686)
164. Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O., Walsh, A. (2018), Machine learning for molecular and materials science, *Nature* (DOI:10.1038/s41586-018-0337-2)
165. Baumes, L. A., Serra, J. M., Serna, P., Corma, A. (2006), Support vector machines for predictive modeling in heterogeneous catalysis, *J. Comb. Chem.*, 8, 4 (DOI:10.1021/cc050093m)
166. Ortegon, J., Ledesma-Alonso, R., Barbosa, R., Vázquez Castillo, J., Castillo Atoche, A. (2018), Material phase classification by means of support vector machines, *Comput. Mater. Sci.*, 148 (DOI:10.1016/j.commatsci.2018.02.054)
167. Xie, T., Grossman, J. C. (2018), Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties, *Phys. Rev. Lett.*, 120, 14 (DOI:10.1103/PhysRevLett.120.145301)

168. Reiser, P., Neubert, M., Eberhard, A., Torresi, L., Zhou, C., Shao, C., Metni, H., van Hoesel, C., Schopmans, H., Sommer, T., Friederich, P. (2022), Graph neural networks for materials science and chemistry, *Commun. Mater.* (DOI:10.1038/s43246-022-00315-6)

169. Kulichenko, M., Smith, J. S., Nebgen, B., Li, Y. W., Fedik, N., Boldyrev, A. I., Lubbers, N., Barros, K., Tretiak, S. (2021), The rise of neural networks for materials and chemical dynamics, *J. Phys. Chem. Lett.* (DOI:10.1021/acs.jpclett.1c01357)

170. Louis, S. Y., Zhao, Y., Nasiri, A., Wang, X., Song, Y., Liu, F., Hu, J. (2020), Graph convolutional neural networks with global attention for improved materials property prediction, *Phys. Chem. Chem. Phys.*, 22, 32 (DOI:10.1039/d0cp01474e)

171. Chen, C., Ye, W., Zuo, Y., Zheng, C., Ong, S. P. (2019), Graph networks as a universal machine learning framework for molecules and crystals, *Chem. Mater.*, 31, 9 (DOI:10.1021/acs.chemmater.9b01294)

172. Karamad, M., Magar, R., Shi, Y., Siahrostami, S., Gates, I. D., Barati Farimani, A. (2020), Orbital graph convolutional neural network for material property prediction, *Phys. Rev. Mater.*, 4, 9 (DOI:10.1103/PhysRevMaterials.4.093801)

173. Park, C. W., Wolverton, C. (2020), Developing an improved crystal graph convolutional neural network framework for accelerated materials discovery, *Phys. Rev. Mater.*, 4, 6 (DOI:10.1103/PhysRevMaterials.4.063801)

174. Benjamin, S. L., Alán, A. G. (2018), Inverse molecular design using machine learning: Generative models for matter engineering, *Science*, 361, 6400

175. Pablo-García, S., Morandi, S., Vargas-Hernández, R. A., Jorner, K., Ivković, Ž., López, N., Aspuru-Guzik, A. (2023), Fast evaluation of the adsorption energy of organic molecules on metals via graph neural networks, *Nat. Comput. Sci.*, 3, 5 (DOI:10.1038/s43588-023-00437-y)

176. Isayev, O., Oses, C., Toher, C., Gossett, E., Curtarolo, S., Tropsha, A. (2017), Universal fragment descriptors for predicting properties of inorganic crystals, *Nat. Commun.* (DOI:10.1038/ncomms15679)

177. Wu, X., Wang, H., Gong, Y., Fan, D., Ding, P., Li, Q., Qian, Q. (2023), Graph neural networks for molecular and materials representation, *J. Mater. Inform.*, 3, 2 (DOI:10.20517/jmi.2023.10)

178. Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., Sun, M. (2020), Graph neural networks: A review of methods and applications, *AI Open* (DOI:10.1016/j.aiopen.2021.01.001)

179. Gori, M., Monfardini, G., Scarselli, F. (2005), A new model for learning in graph domains, *Proc. IEEE Int. Joint Conf. Neural Netw.*

180. Montavon, G., Rupp, M., Gobre, V., Vazquez-Mayagoitia, A., Hansen, K., Tkatchenko, A., Müller, K. R., von Lilienfeld, O. A. (2013), Machine learning of molecular electronic properties in chemical compound space, *New J. Phys.*, 15 (DOI:10.1088/1367-2630/15/9/095003)

181. Montavon, G., Hansen, K., Fazli, S., Rupp, M., Biegler, F., Ziehe, A., Tkatchenko, A., von Lilienfeld, O. A., Müller, K. R. (2012), Learning invariant representations of molecules for atomization energy prediction, In *Adv. Neural Inf. Process. Syst.*, 1

182. von Lilienfeld, O. A. (2018), Quantum machine learning in chemical compound space, *Angew. Chem. Int. Ed.*, 57, 16 (DOI:10.1002/anie.201709686)

183. Omee, S. S., Louis, S. Y., Fu, N., Wei, L., Dey, S., Dong, R., Li, Q., Hu, J. (2022), Scalable deeper graph neural networks for high-performance materials property prediction, *Patterns*, 3, 5 (DOI:10.1016/j.patter.2022.100491)

184. Choudhary, K., DeCost, B. (2021), Atomistic line graph neural network for improved materials property predictions, *npj Comput. Mater.*, 7, 1 (DOI:10.1038/s41524-021-00650-1)

185. Wen, M., Blau, S. M., Spotte-Smith, E. W. C., Dwaraknath, S., Persson, K. A. (2021), BonDNet: A graph neural network for the prediction of bond dissociation energies for charged molecules, *Chem. Sci.*, 12, 5 (DOI:10.1039/d0sc05251e)

186. Zhou, L., Yao, A. M., Wu, Y., Hu, Z., Huang, Y., Hong, Z. (2021), Machine learning assisted prediction of cathode materials for Zn-ion batteries, *Adv. Theory Simul.*, 4, 9 (DOI:10.1002/adts.202100196)

187. Ahmad, Z., Xie, T., Maheshwari, C., Grossman, J. C., Viswanathan, V. (2018), Machine learning enabled computational screening of inorganic solid electrolytes for suppression of dendrite formation in lithium metal anodes, *ACS Cent. Sci.*, 4, 8 (DOI:10.1021/acscentsci.8b00229)

188. Hsu, T., Pham, T. A., Keilbart, N., Weitzner, S., Chapman, J., Xiao, P., Qiu, S. R., Chen, X., Wood, B. C. (2022), Efficient and interpretable graph network representation for angle-dependent properties applied to optical spectroscopy, *NPJ Comput. Mater.*, 8, 1 (DOI:10.1038/s41524-022-00841-4)

189. Gurunathan, R., Choudhary, K., Tavazza, F. (2023), Rapid prediction of phonon structure and properties using the atomistic line graph neural network (ALIGNN), *Phys. Rev. Mater.*, 7, 2 (DOI:10.1103/PhysRevMaterials.7.023803)

190. Choudhary, K., De Cost, B., Major, L., Butler, K., Thiyagalingam, J., Tavazza, F. (2023), Unified graph neural network force-field for the periodic table: Solid state applications, *Digital Discovery*, 2, 2 (DOI:10.1039/d2dd00096b)

191. Gupta, V., Choudhary, K., DeCost, B., Tavazza, F., Campbell, C., Liao, W. K., Choudhary, A., Agrawal, A. (2024), Structure-aware graph neural network based deep transfer learning framework for enhanced predictive analytics on diverse materials datasets, *NPJ Comput. Mater.*, 10, 1 (DOI:10.1038/s41524-023-01185-3)

192. Kong, S., Ricci, F., Guevarra, D., Neaton, J. B., Gomes, C. P., Gregoire, J. M. (2022), Density of states prediction for materials discovery via contrastive learning from probabilistic embeddings, *Nat. Commun.*, 13, 1 (DOI:10.1038/s41467-022-28543-x)

193. Li, Q., Dong, R., Fu, N., Omee, S. S., Wei, L., Hu, J. (2023), Global mapping of structures and properties of crystal materials, *J. Chem. Inf. Model.*, 63, 12 (DOI:10.1021/acs.jcim.3c00224)

194. Nguyen, N., Louis, S. Y. V., Wei, L., Choudhary, K., Hu, M., Hu, J. (2022), Predicting lattice vibrational frequencies using deep graph neural networks, *ACS Omega*, 7, 30 (DOI:10.1021/acsomega.2c02765)

195. Williams, L., Mukherjee, A., Dongol, R., Rajan, K. (2023), Machine learning-aided property prediction of hybrid organic-inorganic perovskites using Hirshfeld surface representations of crystal structures, *J. Phys. Chem. C*, 127, 24 (DOI:10.1021/acs.jpcc.3c00802)

CHAPTER 2

Role of Fluxionality and Metastability in Catalytic Interfaces

CHAPTER 2

2.1 Introduction

The proton exchange membrane fuel cell (PEMFC) has become a promising power source to meet rising global energy needs *[1–3]*. Its high energy efficiency, zero carbon emissions, high energy density, low operating temperature, and negligible exhaust make it a focus of renewable energy research *[4–7]*. Within PEMFCs, the hydrogen evolution reaction (HER) occurs at the anode and the oxygen reduction reaction (ORR) at the cathode. Although HER has fast kinetics, ORR remains the rate-limiting step due to its sluggish kinetics and the resulting high overpotential at the cathode *[8–10]*. Platinum-based catalysts are widely used because they optimally bind ORR intermediates within the Sabatier range. However, high Pt cost and limited durability restrict large-scale commercialization of PEMFCs *[11,12]*. Considerable efforts have therefore been directed toward designing catalysts that lower the cathodic overpotential. Both Pt-based and non-Pt-based materials have been explored. Pt nanoclusters *[13–15]*, nanotubes *[16]*, nanoshells *[17,18]*, and nanofilms *[19]* have shown strong ORR activity. For non-Pt catalysts, transition metal alloys *[20–25]*, and inorganic materials such as graphene, carbon nanotubes *[26–28]*, and metal–organic frameworks (MOFs) *[29]* have also demonstrated ORR activity. However, the limited stability of platinum-free systems under harsh fuel cell conditions continues to shift research attention toward platinum-based electrocatalysts *[30]*. The ORR mechanism has been extensively studied, and platinum nanoparticles remain among the most efficient ORR catalysts *[31–33]*. These particles possess well-defined bulk-like facets. Still, reducing Pt loading without compromising performance remains a major challenge *[34–36]*.

Subnanometer Pt_n clusters have recently drawn attention after Imaoka et al. synthesized clusters containing 10–20 atoms and reported significantly enhanced ORR activity for Pt_{19} *[34–36]*. The improvement was attributed to low-coordination edge sites, previously assumed to diminish activity. Our recent work also showed that ORR activity in nanoclusters is highly

sensitive to atomicity, with exceptional activity emerging from specific cluster sizes *[37]*. These findings highlight the promise of subnanometer clusters, driven by their unique structural and electronic characteristics. Borna et al. further reported high fluxionality for Pt_n $(n = 1 - 6)$ clusters, indicating strong structural dynamics that could enhance ORR activity *[38]*.

Understanding the structural evolution of catalysts under reaction conditions is critical for rational design. At elevated temperatures, electrochemical potentials, electrolyte environments, high surface coverage, or in the presence of strong metal–support interactions (SMSI), clusters undergo major structural rearrangements *[39–42]*. Global optimization (GO) techniques such as genetic algorithms (GA) *[43]*, particle-swarm optimization (PSO) *[44]*, basin hopping *[45]*, and simulated annealing *[46]* coupled with *ab initio* calculations are widely used to predict the global minimum (GM) structure. Although these methods provide a putative global minimum (PGM), relying only on GM may not give a complete description of catalysis. At finite temperatures, thermally accessible metastable local minima (LM) structures contribute collectively to catalytic turnover *[47]*. This ensemble perspective suggests that catalysis arises not from a single static structure but from a swarm of transient, fluxional configurations. Recent work by Alexandrova and co-workers demonstrated this for Pt clusters on alumina surfaces, where including metastable isomers explained the superior catalytic activity of Pt_7 compared to Pt_8 in ethylene dehydrogenation *[48,49]*.

In this study, Pt_{13} subnanoclusters were used to investigate the role of both GM and LM structures and their fluxional behavior during ORR. Global optimization identified multiple Pt_{13} isomers, and their structural characteristics were analyzed. A comprehensive ORR investigation was performed, including adsorption interactions with intermediates, reaction mechanisms, and kinetic analysis of metastable isomers. A statistical ensemble description was developed by accounting for energetics and

kinetics through weighted-average contributions. High-coverage conditions were further examined to determine thermodynamically and kinetically active clusters. The combined results reveal how both GM and LM structures contribute to the overall ORR activity of Pt_{13} subnanoclusters.

2.2 Computational Details

2.2.1 Global Optimization

For global optimization of Pt_{13} subnanoclusters, initial cluster geometries were generated using the Parallel Global Optimization Pathway Toolkit (PGOPT) *[50]*. PGOPT employs the bond length distribution algorithm (BLDA) to efficiently sample the configuration space and generate metal cluster structures, which are then optimized to identify GM and LMs isomers. A population of 500 unique geometries was generated with an order parameter of 3 and a minimum bond length difference of 0.25 Å. The order parameter defines the binding pattern during structure generation. With order 3, the first half of the atoms follow order 1, where each added atom binds to at least one existing atom, while the remaining atoms follow order 2, binding to at least two existing atoms. This setup was chosen to improve structural accuracy for Pt_{13} clusters. Duplicate structures were removed using a similarity check based on depth-first search (DFS), and only unique geometries were retained for optimization. All structures were optimized in a cubic box of 18 × 18 × 18 $Å^3$ to eliminate interactions between periodic images, and calculations were performed across spin multiplicities from singlet to nonet. The isomer with the lowest total energy was designated as the GM, while other low-energy isomers close in energy were categorized as LM.

2.2.2 Electronic Structure Calculations.

The plane wave DFT calculations were performed using the Vienna *ab initio* simulation package (VASP) *[51]*. The exchange-correlation energy was treated using the Generalized Gradient Approximation with the Perdew-Burke-Ernzerhof (GGA-PBE) functional *[51,52]*. The Projector Augmented Wave (PAW) method was applied to describe core–valence

interactions, and the plane-wave cutoff energy was set to 470 eV *[53]*. A Γ-centered (1 × 1 × 1) A Γ-centered (1 × 1 × 1) k-point grid was used for Brillouin zone sampling *[38,54]*. Geometry optimization was carried out using the conjugate-gradient algorithm with an electronic energy convergence threshold of 10^{-4} eV and Hellmann–Feynman force convergence criteria below 0.02 eV $Å^{-1}$. DFT-D3 method was used to correctly define van der Waals interaction *[55]*. Gaussian smearing of 0.02 eV has been employed. All spin-polarized calculations were performed for all the molecular species and oxygen adsorbed intermediates. The climbing image-nudged elastic band (CINEB) was employed to locate the transition states by creating six images in the pathway of the rate-determining step. And the transition states were confirmed by the presence of one imaginary frequency *[56]*. Zero-point energy (ZPE) was calculated using the following equation:

$$ZPE = \sum_i \frac{1}{2} h\nu_i \quad (2.1)$$

where h is Planck's constant and ν_i is the frequency of the i^{th} vibrational mode. The Gibbs free energies of elementary ORR steps were computed within the computational hydrogen electrode (CHE) model proposed by Nørskov and coworkers *[57]*. The free energy change was determined using:

$$\Delta G = \Delta E + \Delta ZPE - T\Delta S - eU \quad (2.2)$$

where ΔE is the reaction energy, ΔZPE is the change in zero-point energy, T is the temperature (300 K), ΔS is the entropy change, U is the electrode potential referenced to SHE, and e is the number of transferred electrons. Zero-point energies and entropies were estimated using the harmonic oscillator approximation *[57]*. U is the electrode potential references to the standard hydrogen electrode, and e is the transferred electronic charge. We have employed gas-phase model of Pt_{13} cluster, without considering the effect of support and electrode potential, as investigated in previous studies *[58]*.

2.3 Results and Discussion

2.3.1 Structural Exploration of the Pt_{13} Clusters

To examine the low-energy metastable ensembles (LEME) obtained from global optimization, an ensemble of ten isomers was constructed and evaluated for their collective contribution to ORR activity. The relative energies of the nine LM, referenced to the GM, along with their corresponding structures, are shown in Figure 2.1. The first LM is located 0.26 eV above the GM, and the number of isomers increases toward the higher-energy end, consistent with the potential energy surface characteristics of Lennard-Jones clusters described by Wales via disconnectivity graphs *[59]*. All clusters are stable in the singlet state except LM3 and LM6, which stabilize in the triplet state. Because multiple thermodynamically accessible metastable isomers can coexist and contribute under operating conditions, these isomers were included in the LEME for subsequent ORR analysis.

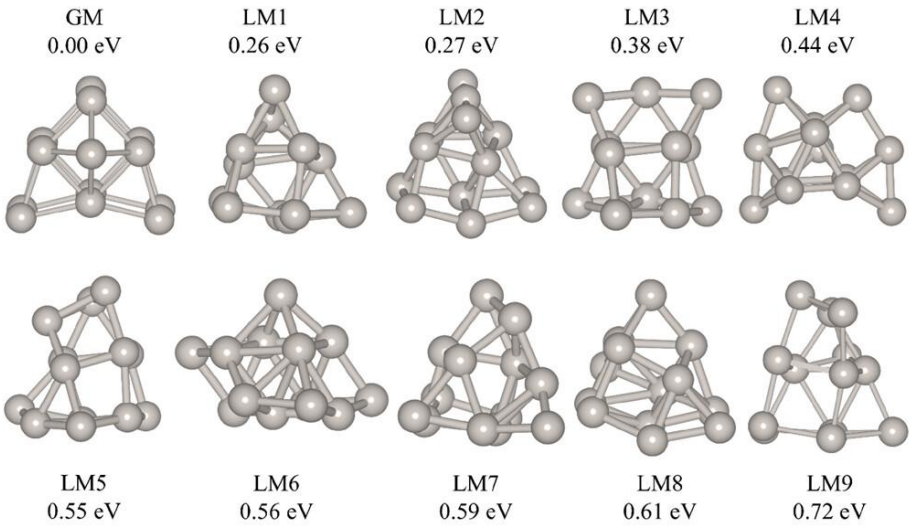


Figure 2.1 Relative energies (eV) and corresponding geometries of the isomers that form the low-energy metastable ensemble (LEME) identified from global optimization. The global minimum (GM) is taken as the reference at 0 eV.

The GM of Pt_{13} features a highly symmetric structure composed of tetrahedral units with capped atoms. A closely related geometry has been

described earlier as a disordered structure (DIS) by Bunau et al. and as a tricapped pentagonal prism by Zhai et al. *[50,60]*. Many studies have reported icosahedral or cuboctahedral motifs for Pt_{13} due to the enhanced stability of magic-number clusters satisfying closed-shell configurations. *[61,62]* However, no such high-symmetry structures emerged from our global optimization. The GM has an average coordination number (CN) of 4, lower than the compact cuboctahedral configuration with an average CN of 5.5 (Table 2.1). This reduced coordination strengthens Pt–Pt bonding by decreasing d–d interactions between 5d electrons, thereby enhancing structural stability.

The geometries of the low-energy LM isomers differ significantly from the GM, exhibiting no symmetrical elements. All identified Pt_{13} isomers were three-dimensional; planar two-dimensional structures were not obtained due to their much higher energies, making them thermally inaccessible, consistent with earlier findings *[37]*. To evaluate coordination environments, we adopted an approach like Lan et al. *[63]* and calculated CN by counting neighboring Pt atoms within 3.0 Å. The CN of individual Pt atoms spans a broad range from 2 to 7 across the ensemble, although the average CN of each isomer remains nearly constant (Table 2.1). The lowest CN value (2) appears in LM7 and LM8, while the highest (7) occurs in LM3 and LM4. Moreover, only minor differences were observed in the average CN across the clusters, consistent with the very small variation in their average Pt–Pt bond lengths (0.058 Å).

Table 2.1. Properties of LEME (Zero Represents the Energy of Global Minima) Obtained from Global Optimization of Pt_{13} Clusters. The cluster binding energies (CBE) in eV/atom, maximum and minimum coordination number (max. CN and min. CN respectively), average Pt-Pt bond length ($\overline{Pt-Pt}$) in Å, average coordination number (avg. CN) and average generalized coordination number (avg. GCN) values corresponding to each isomer in LEME are listed respectively.

Clusters	GM	LM1	LM2	LM3	LM4	LM5	LM6	LM7	LM8	LM9
RE	0.00	0.26	0.27	0.38	0.44	0.55	0.56	0.59	0.61	0.72
Multiplicities	Singlet	Singlet	Singlet	Triplet	Singlet	Singlet	Triplet	Singlet	Singlet	Singlet
$\overline{Pt-Pt}$	2.57	2.57	2.62	2.59	2.59	2.56	2.62	2.57	2.62	2.56
CBE	-3.88	-3.86	-3.86	-3.85	-3.85	-3.84	-3.84	-3.84	-3.84	-3.81
Max. CN	6	5	6	7	7	5	6	6	6	6
Min. CN	3	3	3	3	3	3	3	2	2	3
Avg. CN	4	4	4.46	4.31	4.46	3.85	4.46	4	4.46	3.85
Avg. GCN	1.41	1.39	1.81	1.69	1.82	1.28	1.75	1.43	1.79	1.35

To quantify the structural diversity within the LEME, we employed the generalized coordination number (GCN) proposed by Calle-Vallejo et al. *[64,65]* and defined as:

$$\mathrm{GCN_i} = \sum_{j=1}^{n_i} \frac{\mathrm{CN_j}}{\mathrm{CN_{max}}} \tag{2.3}$$

where $\mathrm{CN_j}$ is the coordination number of each neighboring atom j, and $\mathrm{CN_{max}}$ is the maximum coordination number in the bulk crystal structure (12 for fcc metals such as Pt). GCN is widely used as a structural descriptor of catalytic activity because it accounts for the coordination environment of neighboring atoms, allowing better differentiation between geometries than conventional CN measures. For the Pt_{13} clusters, GCN values span from 0.58 to 2.66, indicating significant structural variation among the LEME (Figure 2.2). The most frequently observed GCN values fall between 1.25 and 1.37, corresponding to the average GCN of LM5 and LM9. The GM has an average GCN of 1.41, slightly higher but close to the dominant range, reflecting the high fluxional behavior of Pt_{13} clusters. This observation aligns with previous findings by Fung et al. *[66]*, who linked fluxionality to variations in coordination environments. As the first member of a magic-number series, Pt_{13} is known for its strong fluxional nature, making structural assignments challenging *[66]*. These results highlight the

importance of including metastable isomers when evaluating the catalytic properties of subnanometer Pt clusters.

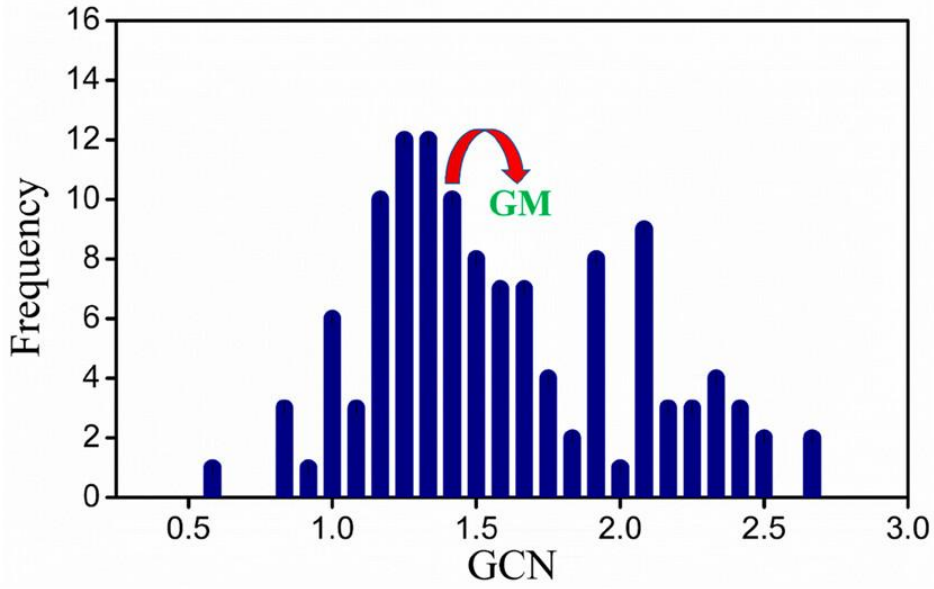


Figure 2.2 Generalized coordination numbers (GCN) of metastable isomers within 0.8 eV of the global minimum obtained from global optimization of the Pt_{13} cluster. The GCN of the GM is highlighted.

To evaluate the stability of the isomers in the LEME of Pt_{13} subnanoclusters, we calculated the cluster binding energy (CBE) for each structure using the following expression:

$$\mathrm{CBE} = [\mathrm{E}(\mathrm{Pt}_{13}) - 13 \times \mathrm{E}(\mathrm{Pt})]/13 \qquad (2.4)$$

where $\mathrm{E}(\mathrm{Pt}_{13})$ and $\mathrm{E}(\mathrm{Pt})$ are the total energies of the Pt_{13} cluster and an isolated Pt atom, respectively. The global minimum (GM) structure shows the highest binding energy, −3.88 eV per atom, indicating the greatest stability. As shown in Table 2.1, the metastable isomers exhibit binding energies close to that of the GM, implying comparable stability. Based on the structural analysis, no direct correlation is observed between the coordination number (CN) and cluster stability, suggesting that CN alone does not govern the stability of the Pt_{13} isomers.

2.3.2 Adsorption of ORR Intermediates

To understand the ORR mechanism, we examined the adsorption behavior of the key intermediates O_2^*, O^*, OH^*, OOH^*, and H_2O^*. In subnanometer

clusters, each atomic site exhibits a distinct electronic structure and coordination geometry, making it challenging to pinpoint the active site and the most stable adsorption configuration. Therefore, multiple adsorption sites were tested for every intermediate, sampling top, bridge, hollow, and disigma positions, and the most stable adsorption configuration were utilized for further mechanistic investigation. For example, the most stable adsorption configuration for different intermediates on GM is represented in Figure 2.3, representing the irregular geometries of these isomers. For the O* intermediate, several three-fold hollow sites were also tested, but they were consistently less stable than top or bridge configurations.

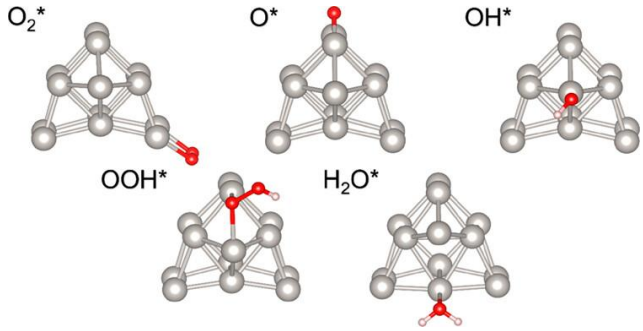


Figure 2.3. Most stable adsorption configuration of ORR intermediates on global minima of Pt_{13} clusters. The colors gray, red, and pink represent platinum, oxygen, and hydrogen atoms, respectively.

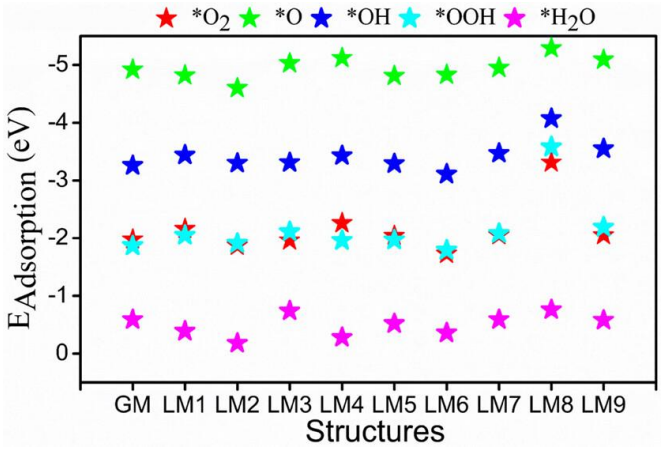


Figure 2.4. Adsorption energy of ORR intermediates on global minima (GM) and different local minima (LM) corresponding to its most stable binding sites.

During the screening, several distinct adsorption configurations showed only small energy differences due to variations in geometry, coordination environment, and electronic structure. For instance, LM1 and LM3 bind all intermediates exclusively at top sites. In LM1, the most stable configurations of O_2^*, O^*, OH^*, OOH^* were only 1.00, 0.03, 0.19, and 0.19 eV more stable than the next-best top sites. This indicates that the symmetry-based elimination of equivalent sites traditionally applied to larger clusters cannot be used here, because adsorption breaks site indistinguishability. Therefore, exhaustive site sampling is required for subnanoclusters. To evaluate the conventional notion that low-coordinated atoms poison catalyst surfaces by binding intermediates strongly *[67]*, we inspected O^* adsorption (the strongest binder among ORR intermediates). The most stable O^* structures were distributed over Pt atoms with a wide range of coordination numbers. In some cases, the most stable binding occurred on Pt atoms with CN = 2 or 6, whereas most often the stable sites showed CN = 3, 4, or 5. This suggests that in the subnanometer regime the coordination environment can either drastically enhance activity or severely poison active sites, depending on the isomer. Adsorption energies were computed using:

$$E_{ads} = E(Pt_{13} - X) - E(Pt_{13})_{ads} - E(X)_{ads} \quad (2.5)$$

where $E(Pt_{13} - X)$ is the total energy of the adsorbed system, and $E(Pt_{13})_{ads}$ and $E(X)_{ads}$ are the energies of the isolated cluster and isolated intermediate at the adsorbed geometry. The adsorption energies for all intermediates across all isomers are shown in Figure 2.4. Adsorption energies showed large variations across isomers with no systematic trend. O^* always bound most strongly, while H_2O^* exhibited the weakest binding. Compared to bulk Pt(111) and the larger Pt_{79} nanocluster *[68]*, all Pt_{13} isomers displayed significantly higher binding energies for ORR intermediates due to their undercoordinated Pt atoms. For the GM, O^* adsorption falls between Pt(111) and Pt_{79}, whereas OH^*, OOH^*, and H_2O^* adsorption is stronger than both bulk systems (Table 2.2). Among all

isomers, LM8 showed the highest binding energies for every intermediate due to its highly coordinated Pt environment; LM7 demonstrated similar behavior.

Bridge, top, disigma, and three-fold hollow sites emerged as the main adsorption motifs. Adsorption trends varied substantially across isomers and intermediates, as represented in Table 2.2. O_2^* favored disigma binding for most metastable isomers, but top sites were also competitive for LM1 and LM3. O^* showed similar preference for bridge and top positions across clusters. OH^* and OOH^* were predominantly absorbed at top sites, except in GM and LM8 where bridge sites were most stable. None of the intermediates preferred the three-fold site; for example, in GM and LM7 the most stable three-fold O^* site was 0.80 and 0.39 eV less stable than the best bridge and top sites, respectively. This contrasts with bulk surfaces or larger nanoparticles, where hollow sites are typically favored. The same qualitative behavior was observed for all remaining isomers.

Table 2.2. Adsorption Energies (eV) of ORR Intermediates at the Most Stable Site over Different Metastable Isomers. Here t, b, and f denote top, bridging, and fcc sites respectively.

Isomer	O_2^*	O^*	OH^*	OOH^*	H_2O^*
GM	-1.98 (b)	-4.92 (b)	-3.26 (t)	-1.87 (b)	-0.59 (t)
LM1	-2.16 (t)	-4.82 (t)	-3.44 (t)	-2.05 (t)	-0.39 (t)
LM2	-1.87 (b)	-4.60 (t)	-3.30 (t)	-1.91 (t)	-0.18 (b)
LM3	-1.96 (t)	-5.03 (t)	-3.31 (t)	-2.11 (t)	-0.74 (t)
LM4	-2.26 (b)	-5.12 (b)	-3.43 (t)	-1.96 (t)	-0.28 (t)
LM5	-2.04 (b)	-4.81 (t)	-3.29 (t)	-1.97 (t)	-0.52 (b)
LM6	-1.73 (b)	-4.83 (b)	-3.11 (t)	-1.79 (t)	-0.36 (t)
LM7	-2.05 (b)	-4.95 (t)	-3.47 (t)	-2.08 (t)	-0.59 (t)
LM8	-3.31 (b)	-5.29 (b)	-4.07 (b)	-3.58 (b)	-0.76 (t)
LM9	-2.05 (b)	-5.09 (b)	-3.55 (t)	-2.19 (t)	-0.58 (t)
Pt (111)	-0.79 (b)	-4.54 (t)	-2.40 (t)	-1.22 (b)	-0.34 (b)
Pt_{79}(NC)	-1.56 (b)	-5.19 (t)	-2.49 (t)	-1.40 (b)	-0.34 (b)

Table 2.3: Root-mean square displacement (RMSD) between pristine and adsorption-induced deformed Pt_{13} clusters geometries for different ORR intermediates on different isomers in low energy metastable ensembles (LEME).

Isomer	O_2^*	O^*	OH^*	OOH^*	H_2O^*
GM	0.02	0.11	0.01	0.13	0.24
LM1	0.07	0.04	0.12	0.06	0.04
LM2	0.04	0.08	0.04	0.08	0.07
LM3	0.46	0.32	0.09	0.30	0.27
LM4	0.19	0.09	0.03	0.09	0.13
LM5	0.06	0.16	0.18	0.15	0.28
LM6	0.05	0.12	0.05	0.12	0.05
LM7	0.08	0.05	0.04	0.06	0.03
LM8	0.57	0.29	0.34	0.64	0.18
LM9	0.16	0.15	0.07	0.06	0.05

We have also quantified the response of the different ORR intermediates toward metastable isomers by calculating the root-mean-square displacement (RMSD) between the pristine cluster and the adsorbed cluster through the Kabsch algorithm available in the program Charnley *[69]*. Large adsorption-induced structural distortions were observed for most isomers. LM8 showed the highest RMSD for OOH^*, while the smallest distortions across all clusters occurred during H_2O^* adsorption (Table 2.3). Intermediates such as O_2^*, O^*, and OOH^* contributed most to structural deformation. This demonstrates the fluxional nature of metastable Pt_{13} isomers during ORR. d-band center (dbc) analysis (Table 2.4) showed that all LMs exhibit a dbc upshift toward the Fermi level relative to GM, consistent with stronger adsorption. For example, dbc shifts of 0.19, 0.67, and 0.27 eV were observed for LM6 (O_2^*), LM4 (O^*), and LM6 (OOH^*), respectively.

Table 2.4: d-band center values of different adsorbates on different isomers of Pt_{13} clusters.

Isomer	O_2^*	O^*	OH^*	OOH^*
GM	-1.70	-1.82	-1.33	-1.78
LM1	-1.68	-1.75	-1.20	-1.69
LM2	-1.51	-1.76	-1.20	-1.60
LM3	-1.67	-1.22	-1.22	-1.72
LM4	-1.59	-1.15	-1.42	-1.69
LM5	-1.69	-1.79	-1.19	-1.72
LM6	-1.51	-1.65	-1.12	-1.51
LM7	-1.64	-1.75	-1.50	-1.69
LM8	-1.66	-1.68	-1.52	-1.76
LM9	-1.69	-1.71	-1.49	-1.72

Finally, we tested whether scaling relationships exist between the adsorption energies of ORR intermediates across different Pt_{13} isomers (Figure 2.5) *[38]*. A weak correlation was observed between O^* and OH^* adsorption ($R^2 = 0.52$), indicating scaling breakdown. In contrast, OH* vs OOH* showed strong correlation ($R^2 = 0.89$), which arises from their similar binding positions (mainly top sites) and comparable coordination and electronic environments. Because LM8 consistently behaved as an outlier due to its high stability and high coordination, we recalculated the correlations by excluding LM8. The R^2 values dropped to 0.24 and 0.69 for O^* vs OH^* and OH^* vs OOH^*, respectively, confirming that scaling relationships in subnanoclusters are sensitive to specific high-energy isomers.

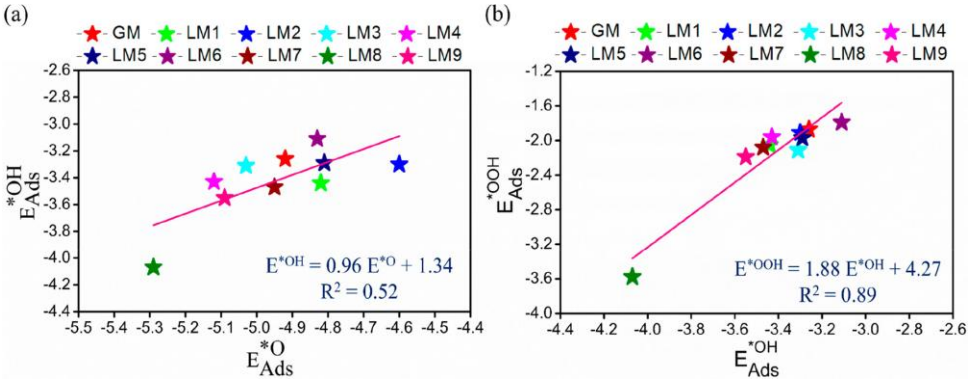


Figure 2.5. Scaling relationship between (a) O^* and OH^* adsorption energies and (b) OH^* vs OOH^* adsorption energies. The isomers are represented by their respective colored mentioned in the figure.

2.3.3 Free Energy Analysis

To investigate the ORR energetics for each Pt_{13} isomer, free energy diagrams were constructed using the most stable adsorption configurations. ORR generally proceeds through two competing pathways: (i) O_2 dissociation (dissociative pathway) and (ii) peroxyl formation (associative pathway) *[70,71]*. The elementary steps for both mechanisms are summarized below:

Associative Pathway:

$$O_2\,(g) + (H^+ + e^-) + * \rightarrow OOH^* \quad \text{(R1)}$$

$$OOH^* + (H^+ + e^-) \rightarrow O^* + H_2O \quad \text{(R2)}$$

$$O^* + (H^+ + e^-) \rightarrow OH^* \quad \text{(R3)}$$

$$OH^* + (H^+ + e^-) \rightarrow H_2O + * \quad \text{(R4)}$$

Dissociative Pathway

$$O_2\,(g) + (H^+ + e^-) + * \rightarrow O^* + O^* + (H^+ + e^-) \quad \text{(R5)}$$

$$O^* + O^* + (H^+ + e^-) \rightarrow O^* + OH^* \quad \text{(R6)}$$

$$O^* + OH^* + (H^+ + e^-) \rightarrow O^* + H_2O \quad \text{(R7)}$$

$$O^* + (H^+ + e^-) \rightarrow OH^* \quad \text{(R8)}$$

$$OH^* + (H^+ + e^-) \rightarrow H_2O + * \quad \text{(R9)}$$

We calculated the free energy change for every isomer at 0 V (no applied potential) and at 1.23 V, the equilibrium potential of ORR. For an ideal catalyst, the total reaction free energy of 4.92 eV would be evenly distributed across all elementary steps *[12,72]*. In practice, one or more steps become endothermic at 1.23 V, and therefore require an additional potential (overpotential) to drive the reaction. This endothermic step corresponds to the rate-determining step. Because each Pt_{13} isomer exhibits a distinct coordination environment, large differences in ORR activity were observed across the metastable structures (Table 2.5). LM4 and LM9

display the highest ORR activity with the lowest overpotentials, marking them as the most thermodynamically favorable among all isomers. Other isomers show higher overpotentials for their respective rate-determining steps and are therefore less active than LM4 and LM9.

Table 2.5: Rate determining step (RDS) in both associative and dissociative pathway and overpotential associated with RDS for all isomers. The activation energy (E_a) is also provided for RDS for all isomers for the dissociative pathway.

Isomers	*Associative*	*Associative*	*Dissociative*	*Dissociative*	*Activation Energy*
	RDS	η_{ORR} (V)	RDS	η_{ORR} (V)	E_a (eV)
GM	$O_2^* \rightarrow OOH^*$	1.61	$O^* \rightarrow OH^*$	1.00	0.00
LM1	$O_2^* \rightarrow OOH^*$	1.57	$OH^* - O^* \rightarrow O^*$	1.07	1.54
LM2	$O_2^* \rightarrow OOH^*$	1.48	$O^* \rightarrow OH^*$	2.13	0.53
LM3	$O_2^* \rightarrow OOH^*$	1.39	$O^* \rightarrow OH^*$	1.19	0.00
LM4	$O_2^* \rightarrow OOH^*$	1.84	$OH^* - O^* \rightarrow O^*$	0.89	0.67
LM5	$O_2^* \rightarrow OOH^*$	1.59	$OH^* \rightarrow H_2O$	1.05	0.48
LM6	$O_2^* \rightarrow OOH^*$	1.69	$OH^* - O^* \rightarrow O^*$	1.01	1.19
LM7	$O_2^* \rightarrow OOH^*$	1.52	$OH^* - O^* \rightarrow O^*$	1.17	0.94
LM8	$OH^* \rightarrow H_2O$	1.87	$OH^* \rightarrow H_2O$	1.01	0.00
LM9	$O_2^* \rightarrow OOH^*$	1.41	$OH^* - O^* \rightarrow O^*$	0.87	0.00

Table 2.6: Gibbs free energy change (in eV) of ORR elementary steps associated with the associative pathway (R1-R4) of different Pt_{13} clusters in LEME. The calculated theoretical overpotential values (η_{ORR} in V) measured at 1.23 V are also included.

Cluster	*GM*	*LM1*	*LM2*	*LM3*	*LM4*	*LM5*	*LM6*	*LM7*	*LM8*	*LM9*
ΔG_{R1}	-1.42	-1.62	-1.44	-1.62	-1.48	-1.49	-1.36	-1.58	-3.15	-1.70
ΔG_{R1}	-2.63	-2.36	-2.32	-2.54	-2.87	-2.51	-2.62	-2.53	-1.27	-2.53
ΔG_{R1}	-0.74	-0.98	-0.78	-0.65	-0.69	-0.78	-0.62	-0.88	-1.13	-0.83
ΔG_{R1}	-0.12	0.05	-0.37	-0.11	0.03	-0.14	-0.32	0.07	0.64	0.14
η_{ORR}	1.61	1.57	1.48	1.39	1.84	1.59	1.69	1.52	1.87	1.41

Table 2.7 Gibbs free energy change (in eV) of ORR elementary steps associated with the associative pathway (R5-R9) of different Pt_{13} clusters in LEME. The calculated theoretical overpotential values (η_{ORR} in V) measured at 1.23 V are also included.

Cluster	*GM*	*LM1*	*LM2*	*LM3*	*LM4*	*LM5*	*LM6*	*LM7*	*LM8*	*LM9*
ΔG_{R5}	-2.68	-1.58	-2.33	-3.01	-2.07	-2.41	-1.87	-2.66	-2.61	-2.61
ΔG_{R6}	-0.67	-0.99	-1.09	-0.17	-0.77	-0.90	-1.22	-1.04	-0.60	-0.41
ΔG_{R7}	-0.55	-0.16	-1.95	-1.07	-0.33	-0.63	-1.13	-0.05	-0.52	-0.36
ΔG_{R8}	-0.23	-1.56	0.73	-0.04	-0.78	-0.79	-1.33	-0.87	-1.12	-0.54
ΔG_{R8}	-0.82	-0.62	-0.28	-0.63	-0.97	-0.18	-0.36	-0.30	-0.07	-1.00
η_{ORR}	1.00	1.07	2.13	1.19	0.89	1.05	1.10	1.17	1.01	0.87

A comparison of ΔG_{R1} and ΔG_{R5} in Table 2.6-2.7 indicates that most isomers favor the dissociative mechanism, where the O–O bond breaking step is more exergonic. In the dissociative pathway, the O_2 dissociation step (ΔG_{R5}) is the rate-determining step for all clusters. In the associative pathway, $O_2 \rightarrow$ OOH is the rate-limiting step for most isomers, except LM8, where OH $\rightarrow H_2O$ governs the energetics. From Table 2.5, the smaller overpotentials (η) associated with the dissociative pathway highlight its preference over the associative route. Notably, LM4 and LM9 show working potentials of 0.34 and 0.36 V *[12]*, close to the benchmark Pt(111) surface, making their activity comparable to bulk Pt. This underscores the potential of subnanoclusters to deliver high catalytic performance while reducing precious-metal usage and improving atom efficiency. The corresponding free energy diagrams for GM is shown in Figure 2.6.

A Langmuir–Hinshelwood reaction network was considered, where adsorbed protons react on the cluster surface to form ORR products. Since calculating kinetic barriers for every elementary step is computationally intensive, kinetic analysis was performed only for the rate-determining step of the dissociative mechanism (Table 2.5). Along with GM, LM3, LM8,

and LM9 exhibit a barrierless rate-determining step, while LM1 and LM6 show high activation barriers of 1.54 and 1.19 eV for the water-formation step, making them kinetically less accessible. Overall, the thermodynamic and kinetic analyses show that metastable isomers can actively participate in ORR alongside the GM. Therefore, a statistical ensemble approach is required to quantify the collective contribution of all isomers to the overall ORR activity of Pt_{13}.

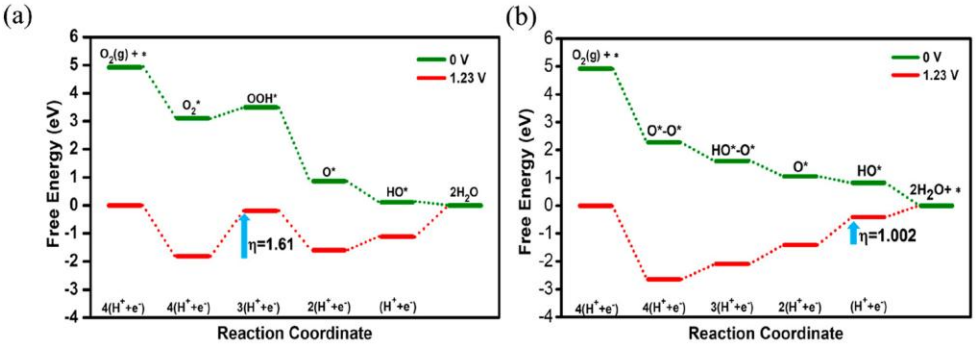


Figure 2.6. Free energy diagrams of ORR elementary steps on the GM of the Pt_{13} cluster via (a) the associative pathway and (b) the dissociative pathway at applied potentials of 0 and 1.23 V. The overpotential corresponding to the rate-determining step (RDS) at 1.23 V is indicated by the blue arrow. The asterisk (*) denotes surface-adsorbed species.

2.3.4 Statistical Ensemble Representation of Overall Catalysis

The above ORR mechanistic investigations led us to the conclusion that along with GM, different metastable isomers of Pt_{13} clusters also show significant catalytic activity. Therefore, a more accurate description of subnanoclusters includes the cumulative sum of activity from all isomers in the statistical pool of the ensemble. Hence, to quantify the contribution of metastable isomers, we have calculated a normalized rate constant ($k_{Weighted}$) of these isomers weighted by their Boltzmann distribution, according to the method recently introduced by Sun and Sautet. *[73]* as:

$$k_{Weighted} = A \exp\left(\frac{-E_a^{LM_i}}{k_B T}\right) \times \exp\left(-\frac{E^{LM_i} - E^{GM}}{k_B T}\right) \qquad (2.6)$$

where A is the Arrhenius parameter for which a uniform value of 10^{13} s^{-1} is considered for the reactions, $E_a^{LM_i}$ is the activation barrier of isomer, E^{LM_i} is the energy of the isomer, E^{GM} is the energy of the GM, and the temperature (T) is 300 K. The activation energies and relative energies of LM compared to GM were combined to acquire a weight averaged activity contribution from each isomer. The calculated values of $k_{Weighted}$ for each cluster are represented in Figure 2.7. The kinetic energy barrier associated with GM, LM3, LM8, and LM9 for the rate-determining step was 0.00 eV. Among all isomers, the order of contribution toward ensemble catalytic activity is as follows:

$$GM \gg LM3 > LM8 > LM2 > LM5 > LM9 > LM4 > LM7 > LM6 > LM1$$

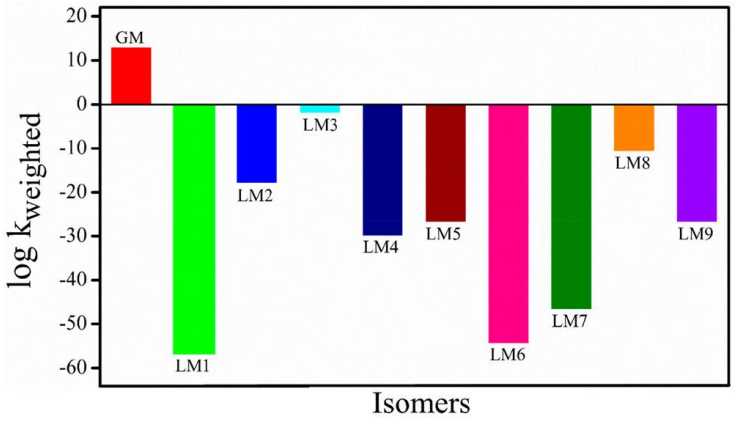


Figure 2.7. Normalized weighted contribution of different metastable isomers in overall ORR activity representing ensemble-averaged ORR activity of Pt_{13} clusters.

This trend confirms that the GM of Pt_{13} dominates the overall ORR activity, even though several metastable isomers show high thermodynamic and kinetic activity. While some LMs exhibit a barrierless rate-determining step and therefore display activity comparable to the GM, their relative stability is much lower, which limits their contribution. For example, LM1 and LM6 have high activation barriers (E_a = 1.54 and 1.19 eV), making them kinetically inaccessible and resulting in negligible catalytic contribution. LM3 and LM8 have zero kinetic barriers for the rate-

determining step and are intrinsically active, yet their contribution remains low due to excessively strong adsorption of intermediates, which suppresses their activity under fuel cell operating temperatures. In electrocatalysis, the probability of forming a metastable state is inherently lower than that of the GM, which further reinforces GM dominance. Even so, LM2, LM3, and LM8 contribute noticeably to the overall ORR activity compared to LM1, LM6, and LM7. The results also indicate that although certain metastable isomers may exhibit high intrinsic activity due to their dynamic nature, their limited thermodynamic accessibility prevents them from outperforming the GM.

LM6, LM7, and LM9 lie significantly higher in energy than GM, making them thermodynamically inaccessible. The large population of the GM relative to the LMs further explains its superior contribution. Overall, the impact of an LM depends strongly on cluster atomicity and the structural landscape available to the cluster. For Pt_{13}, LM2 and LM3 fall within approximately 0.4 eV of the GM and therefore contribute meaningfully to ORR activity alongside the GM. Isomers beyond this 0.4 eV cutoff provide only marginal enhancement due to their low thermodynamic accessibility. These findings align well with previous studies by Alexandrova and co-workers, Sun et al., and Nair et al. *[48,49,54,73]*, which also identified catalytically relevant metastable isomers within ~0.4 eV of the GM. In summary, LMs that (i) are energetically close to the GM (within ~0.4 eV) and (ii) possess low activation barriers can significantly contribute to ORR activity in subnanoclusters, whereas higher-energy isomers have negligible effect on overall performance.

2.3.5 *Ab Initio* Thermodynamics Analysis

Under catalytic reaction conditions, Pt clusters readily oxidize, leading to high coverage of reaction intermediates on surface adsorption sites. Since the reactivity and contribution of individual isomers varied under single-adsorbate conditions, we extended our analysis to high-coverage scenarios. GM, LM2, LM3, and LM8 were selected because they contributed most

significantly to ORR activity at low coverage. Given that O binds most strongly to all clusters, we constructed a series of $Pt_{13}O_x$ models (x = 0, 1, 2, 3, 4, 8, 12) for each of these isomers. Multiple configurations were generated for each oxidation level, and the most stable structure identified from DFT optimization was used for subsequent thermodynamic and kinetic stability analysis. To evaluate thermodynamic stability under oxidizing conditions, we applied the *ab initio* thermodynamics framework *[74,75]*. In this approach, the chemical potential of gas-phase oxygen $\mu_o(T, P)$ is treated as a function of temperature and pressure and expressed as:

$$\mu_o(T, P) = \frac{1}{2}\left[E_{O_2}\,(T = 0\ K, P^o)\ \mu_{O_2}(T, P^o) + k_B T \ln \frac{P}{P^o}\right] \qquad (2.7)$$

where T is the temperature, P is the oxygen partial pressure, P^o is the standard atmospheric pressure, and k_B is Boltzmann's constant. The Gibbs free energies of various coverages were calculated using the following equation:

$$\Delta G\,(T, P) = G_{Pt_{13}O_x}\,(T, P) - G_{Pt_{13}}\,(T, P) - N_o \mu_o(T, P) \qquad (2.8)$$

where $G_{Pt_{13}O_x}$ is the free energy of Pt_{13} subnanoclusters with *x* number of oxygen atoms adsorbed on it, and $G_{Pt_{13}}$ is the free energy of the bare Pt_{13} clusters having no adsorbates. The phase diagrams were constructed using pMuTT (Python Multiscale Thermochemistry Toolbox) *[76]*, a python-based software library by utilizing the vibrational degrees of freedom associated with both Pt and O atoms are included in the free energy calculations *[77]*. The second-order phase diagrams associated with the most stable configuration of multiple adsorbates for active isomers are represented in Figure 2.8.

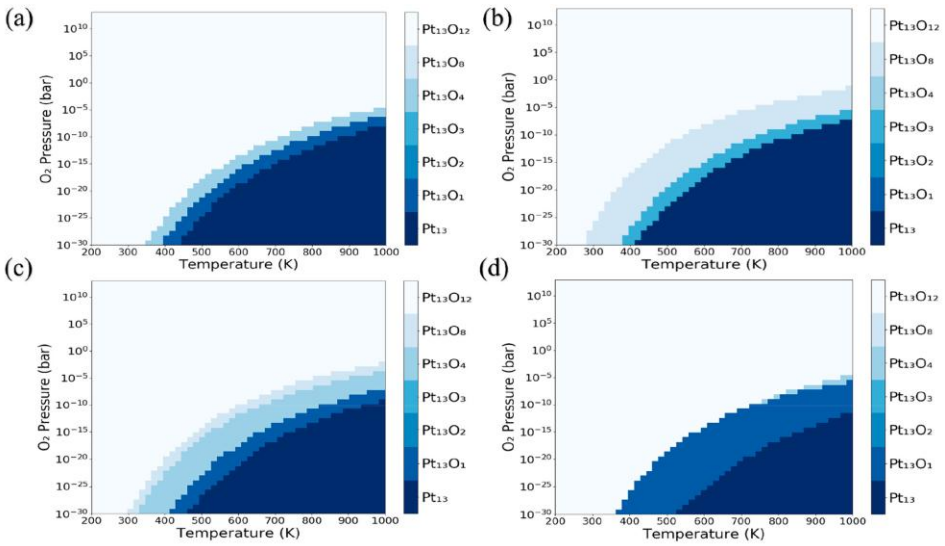


Figure 2.8. *Ab initio* thermodynamics analysis representing second-order phase diagrams of (a) GM, (b) LM2, (c) LM3, and (d) LM8 for different coverages of the Pt_{13} cluster, respectively.

Significant geometric transformations were observed across multiple oxygen coverages, indicating the high fluxionality of Pt_{13} clusters under reaction conditions. Thermodynamic stability at fuel cell operating temperatures was evaluated by computing the reaction free energy using:

$$\Delta G_r = G(Pt_{13}O_x) - \frac{x}{2} G(O_2) - G(Pt_{13})) \quad (2.9)$$

where $G(Pt_{13}O_x)$ and $G(Pt_{13})$ were obtained from the harmonic oscillator approximation, and $G(O_2)$ was calculated using DFT combined with ideal-gas statistical thermodynamics. Both first- and second-order phase diagrams show that $Pt_{13}O_x$ is highly stable across a broad temperature and pressure range.

Although only five O^* coverages were analyzed and individual phase diagrams were not generated for every coverage, a full coverage-dependent investigation would be required to map the energetics of each isomer in detail. Since 1 ML coverage ($Pt_{13}O_{12}$) already provides strong thermodynamic stabilization at fuel cell operating temperatures, it is sufficient for understanding cluster activity under high-coverage conditions. For $Pt_{13}O_{12}$ on all selected isomers, O atoms preferentially occupy bridging

sites. $Pt_{13}O_8$ was found to be stable only for LM2 and LM3 at 300 K; however, higher temperature and pressure again favor $Pt_{13}O_{12}$ formation. Coverage strongly alters the ORR energetics. While the rate-determining step was barrierless for these clusters at low coverage, the activation energy at $Pt_{13}O_{12}$ increases significantly for GM, LM2, LM3, and LM8 (Table 2.8). These results highlight that high coverage induces a fundamentally different energetic landscape, which is necessary to capture the realistic dynamics of Pt subnanoclusters under an oxidizing environment. Although the explicit ensemble contribution of high-coverage isomers was not computed, the clear structural and stability trends at large coverage indicate that coverage effects play a major role in determining the catalytic behavior of subnanometer clusters.

Table 2.8: Free energy associated with the multiple coverage on metastable isomers and the activation energy (E_a, in eV units) associated with the rate-determining step (RDS) for each cluster for $Pt_{13}O_{12}$ clusters.

Isomer	$Pt_{13}O_1$	$Pt_{13}O_2$	$Pt_{13}O_3$	$Pt_{13}O_4$	$Pt_{13}O_8$	$Pt_{13}O_{12}$	E_a
GM	-5.62	-11.37	-16.12	-20.87	-45.86	-71.13	0.66
LM2	-5.37	-11.68	-16.44	-21.02	-46.76	-71.49	1.11
LM3	-5.12	-10.98	-16.42	-21.23	-45.93	-70.38	1.61
LM8	-5.93	-11.83	-15.93	-20.56	-46.17	-71.47	0.72

2.4 Conclusion

Uncovering the morphology and active site reconstruction of Pt_{13} subnanoclusters under realistic conditions has provided valuable insights into the origin of their catalytic behavior. Global optimization using DFT identified the global minimum (GM) along with several low-energy metastable isomers that contribute collectively to catalysis. Restructuring, fluxionality, and ensemble effects play a central role in reactivity at the subnanometer scale. Distinct local coordination environments across isomers induce different adsorption-driven interactions, resulting in high structural dynamics. Scaling relationships constructed for the adsorption of key intermediates across the metastable isomers offered important insights

into subnano cluster catalysis. Detailed ORR analysis showed that several metastable isomers exhibit thermodynamic and kinetic activity comparable to GM through the dissociative pathway. The rate-determining step, associated transition states, and the extent of structural distortions during the reaction helped rationalize the catalytic performance. Overpotentials close to bulk Pt(111) highlight opportunities for efficient utilization of all atoms in a cluster.

A statistical ensemble analysis using the weight-averaged normalized rate constant emphasized the dominance of GM due to its highest occurrence probability. Metastable isomers within ~0.4 eV of GM and with low activation barriers contribute meaningfully to overall ORR activity, whereas higher-energy isomers (beyond the 0.4 eV cutoff) offer only marginal enhancements at low temperature. Ab initio thermodynamics further revealed that oxygen-rich conditions lead to energetics that differ fundamentally from the low-coverage regime. The negligible contribution of metastable isomers relative to GM for ORR is specific to the Pt_{13} system. A broader investigation is required to understand size-dependent ensemble effects as a function of cluster atomicity. Although cluster- and reaction-specific, these results illustrate how operando-level modeling that incorporates metastable-state reactivity enables a cumulative framework for interpreting experimentally observed catalytic activity.

2.5 References

1. Debe, M. K. (2012), Electrocatalyst Approaches and Challenges for Automotive Fuel Cells, Nature, 486, 43–51 (DOI: 10.1038/nature11115)
2. Lubitz, W., Tumas, W. (2007), Hydrogen: An Overview, Chem. Rev., 107, 3900–3903 (DOI: 10.1021/cr050200z)
3. Watanabe, M., Tryk, D. A., Wakisaka, M., Yano, H., Uchida, H. (2012), Overview of Recent Developments in Oxygen Reduction Electrocatalysis, Electrochim. Acta, 84, 187–201 (DOI: 10.1016/j.electacta.2012.04.035)

4. Qin, C., Wang, J., Yang, D., Li, B., Zhang, C. (2016), Proton Exchange Membrane Fuel Cell Reversal: A Review, Catalysts, 6, 197 (DOI: 10.3390/catal6120197)
5. Kraytsberg, A., Ein-Eli, Y. (2014), Review of Advanced Materials for Proton Exchange Membrane Fuel Cells, Energy Fuels, 28, 7303–7330 (DOI: 10.1021/ef501977k)
6. De Bruijn, F. A., Dam, V. A. T., Janssen, G. J. M. (2008), Review: Durability and Degradation Issues of PEM Fuel Cell Components, Fuel Cells, 8, 3–22 (DOI: 10.1002/fuce.200700053)
7. Gewirth, A. A., Thorum, M. S. (2010), Electroreduction of Dioxygen for Fuel-Cell Applications: Materials and Challenges, Inorg. Chem., 49, 3557–3566 (DOI: 10.1021/ic9022486)
8. Shao, M., Chang, Q., Dodelet, J. P., Chenitz, R. (2016), Recent Advances in Electrocatalysts for Oxygen Reduction Reaction, Chem. Rev., 116, 3594–3657 (DOI: 10.1021/acs.chemrev.5b00462)
9. Viswanathan, V., Hansen, H. A., Rossmeisl, J., Nørskov, J. K. (2012), Universality in Oxygen Reduction Electrocatalysis on Metal Surfaces, ACS Catal., 2, 1654–1660 (DOI: 10.1021/cs300227s)
10. Bing, Y., Liu, H., Zhang, L., Ghosh, D., Zhang, J. (2010), Nanostructured Pt-Alloy Electrocatalysts for PEM Fuel Cell Oxygen Reduction Reaction, Chem. Soc. Rev., 39, 2184–2202 (DOI: 10.1039/b912552c)
11. Gewirth, A. A., Varnell, J. A., DiAscro, A. M. (2018), Nonprecious Metal Catalysts for Oxygen Reduction in Heterogeneous Aqueous Systems, Chem. Rev., 118, 2313–2339 (DOI: 10.1021/acs.chemrev.7b00335)
12. Kulkarni, A., Siahrostami, S., Patel, A., Nørskov, J. K. (2018), Understanding Catalytic Activity Trends in the Oxygen Reduction Reaction, Chem. Rev., 118, 2302–2312 (DOI: 10.1021/acs.chemrev.7b00488)

13. Sementa, L., Andreussi, O., Goddard, W. A., Fortunelli, A. (2016), Catalytic Activity of Pt38 in the Oxygen Reduction Reaction from First-Principles Simulations, Catal. Sci. Technol., 6, 688–696 (DOI: 10.1039/c6cy00750c)
14. Jennings, P. C., Aleksandrov, H. A., Neyman, K. M., Johnston, R. L. (2014), A DFT Study of Oxygen Dissociation on Platinum Based Nanoparticles, Nanoscale, 6, 1153–1165 (DOI: 10.1039/c3nr04750d)
15. Toyoda, E., Jinnouchi, R., Hatanaka, T., Morimoto, Y., Mitsuhara, K., Visikovskiy, A., Kido, Y. (2011), The D-Band Structure of Pt Nanoclusters Correlated with the Catalytic Activity for an Oxygen Reduction Reaction, J. Phys. Chem. C, 115, 21236–21240 (DOI: 10.1021/jp206360e)
16. Nair, A. S., Mahata, A., Pathak, B. (2018), Multilayered Platinum Nanotube for Oxygen Reduction in a Fuel Cell Cathode: Origin of Activity and Product Selectivity, ACS Appl. Energy Mater., 1, 4150–4158 (DOI: 10.1021/acsaem.8b00641)
17. Cheng, D., Qiu, X., Yu, H. (2014), Enhancing Oxygen Reduction Reaction Activity of Pt-Shelled Catalysts via Subsurface Alloying, Phys. Chem. Chem. Phys., 16, 19845–19855 (DOI: 10.1039/c4cp02863e)
18. Oezaslan, M., Hasché, F., Strasser, P. (2013), Pt-Based Core–Shell Catalyst Architectures for Oxygen Fuel Cell Electrodes, J. Phys. Chem. Lett., 4, 3273–3291 (DOI: 10.1021/jz4014135)
19. Du, M., Cui, L., Cao, Y., Bard, A. J. (2015), Mechanoelectrochemical Catalysis of the Effect of Elastic Strain on a Platinum Nanofilm for the ORR, J. Am. Chem. Soc., 137, 7397–7403 (DOI: 10.1021/jacs.5b03034)
20. Xiao, W., Liutheviciene Cordeiro, M. A., Gong, M., Han, L., Wang, J., Bian, C., Zhu, J., Xin, H. L., Wang, D. (2017), Optimizing the ORR Activity of Pd Based Nanocatalysts by Tuning Their Strain and Particle Size, J. Mater. Chem. A, 5, 9877–9885 (DOI: 10.1039/c7ta02479g)

21. Gupta, R. K., Verma, A. D., Sinha, I., Malviya, M. (2017), Oxygen Reduction Reaction on Anisotropic Silver Nanoparticles in Alkaline Media, Chem. Phys. Lett., 680, 90–95 (DOI: 10.1016/j.cplett.2017.05.031)
22. Nie, Y., Li, L., Wei, Z. (2015), Recent Advancements in Pt and Pt-Free Catalysts for Oxygen Reduction Reaction, Chem. Soc. Rev., 44, 2168–2201 (DOI: 10.1039/c4cs00484a)
23. Holewinski, A., Idrobo, J. C., Linic, S. (2014), High-Performance Ag–Co Alloy Catalysts for Electrochemical Oxygen Reduction, Nat. Chem., 6, 828–834 (DOI: 10.1038/nchem.2032)
24. Tang, W., Zhang, L., Henkelman, G. (2011), Catalytic Activity of Pd/Cu Random Alloy Nanoparticles for Oxygen Reduction, J. Phys. Chem. Lett., 2, 1328–1331 (DOI: 10.1021/jz2004717)
25. Kuttiyiel, K. A., Sasaki, K., Su, D., Wu, L., Zhu, Y., Adzic, R. R. (2014), Gold-Promoted Structurally Ordered Intermetallic Palladium Cobalt Nanoparticles for the Oxygen Reduction Reaction, Nat. Commun., 5, 5185 (DOI: 10.1038/ncomms6185)
26. Del Cueto, M., Ocón, P., Poyato, J. M. L. (2015), Comparative Study of Oxygen Reduction Reaction Mechanism on Doped Graphene Surfaces, J. Phys. Chem. C, 119, 10912–10922 (DOI: 10.1021/jp512588r)
27. Li, Y., Zhang, S., Yu, J., Wang, Q., Sun, Q., Jena, P. (2015), A New C=C Embedded Porphyrin Sheet with Superior Oxygen Reduction Performance, Nano Res., 8, 2902–2912 (DOI: 10.1007/s12274-015-0795-x)
28. Chen, Z., Higgins, D., Tao, H., Hsu, R. S., Chen, Z. (2009), Highly Active Nitrogen-Doped Carbon Nanotubes for Oxygen Reduction Reaction, J. Phys. Chem. C, 113, 21008–21013 (DOI: 10.1021/jp908067v)
29. Strickland, K., Miner, E., Jia, Q., Tylus, U., Ramaswamy, N., Liang, W., Sougrati, M. T., Jaouen, F., Mukerjee, S. (2015), Highly Active

Oxygen Reduction Non-Platinum Group Metal Electrocatalyst, Nat. Commun., 6, 7343 (DOI: 10.1038/ncomms8343)

30. Keeley, G. P., Cherevko, S., Mayrhofer, K. J. J. (2016), The Stability Challenge on the Pathway to Low Platinum Loading, ChemElectroChem, 3, 51–54 (DOI: 10.1002/celc.201500425)
31. Wu, J., Yang, H. (2013), Platinum-Based Oxygen Reduction Electrocatalysts, Acc. Chem. Res., 46, 1848–1857 (DOI: 10.1021/ar300359w)
32. Huang, X., Zhao, Z., Cao, L., Chen, Y., Zhu, E., Lin, Z., Li, M., Yan, A., Zettl, A., Wang, Y. M., Duan, X., Mueller, T., Huang, Y. (2015), High-Performance Transition Metal-Doped Pt3Ni Octahedra, Science, 348, 1230–1234 (DOI: 10.1126/science.aaa8765)
33. Bu, L., Zhang, N., Guo, S., Zhang, X., Li, J., Yao, J., Wu, T., Lu, G., Ma, J. Y., Su, D., Huang, X. (2016), Biaxially Strained PtPb/Pt Core/Shell Nanoplate Boosts ORR, Science, 354, 1410–1414 (DOI: 10.1126/science.aah6133)
34. Imaoka, T., Kitazawa, H., Chun, W. J., Yamamoto, K. (2015), Finding the Most Catalytically Active Platinum Clusters, Angew. Chem. Int. Ed., 54, 9810–9815 (DOI: 10.1002/anie.201504473)
35. Imaoka, T., Kitazawa, H., Chun, W. J., Omura, S., Albrecht, K., Yamamoto, K. (2013), Magic Number Pt13 and Misshapen Pt12 Clusters, J. Am. Chem. Soc., 135, 13089–13095 (DOI: 10.1021/ja405922m)
36. Yamamoto, K., Imaoka, T., Chun, W. J., Enoki, O., Katoh, H., Takenaga, M., Sonoi, A. (2009), Size-Specific Catalytic Activity of Platinum Clusters, Nat. Chem., 1, 397–402 (DOI: 10.1038/nchem.288)
37. Nair, A. S., Anoop, A., Ahuja, R., Pathak, B. (2021), Role of Atomicity in ORR Activity of Platinum Subnanometer Clusters, J. Comput. Chem., 42, 1975–1985 (DOI: 10.1002/jcc.26725)

38. Zandkarimi, B., Alexandrova, A. N. (2019), Dynamics of Subnanometer Pt Clusters Break Scaling Relations, J. Phys. Chem. Lett., 10, 460–466 (DOI: 10.1021/acs.jpclett.8b03680)
39. Jimenez-Izal, E., Alexandrova, A. N. (2018), Computational Design of Clusters for Catalysis, Annu. Rev. Phys. Chem., 69, 377–400 (DOI: 10.1146/annurev-physchem-050317-014216)
40. Zhai, H., Alexandrova, A. N. (2017), Fluxionality of Catalytic Clusters, ACS Catal., 7, 1905–1911 (DOI: 10.1021/acscatal.6b03243)
41. Hook, A., Massa, J. D., Celik, F. E. (2016), Effect of Tin Coverage on Ethane Dehydrogenation, J. Phys. Chem. C, 120, 27307–27317 (DOI: 10.1021/acs.jpcc.6b08407)
42. Ghosh, P., Farnesi Camellone, M., Fabris, S. (2013), Fluxionality of Au Clusters at Ceria Surfaces, J. Phys. Chem. Lett., 4, 2256–2263 (DOI: 10.1021/jz4009079)
43. Deaven, D. M., Ho, K. M. (1995), Molecular Geometry Optimization with a Genetic Algorithm, Phys. Rev. Lett., 75, 288–291 (DOI: 10.1103/PhysRevLett.75.288)
44. Avendaño-Franco, G., Romero, A. H. (2016), Firefly Algorithm for Structural Search, J. Chem. Theory Comput., 12, 3416–3428 (DOI: 10.1021/acs.jctc.5b01157)
45. Wales, D. J., Doye, J. P. K. (1997), Global Optimization by Basin-Hopping, J. Phys. Chem. A, 101, 5111–5116 (DOI: 10.1021/jp970984n)
46. Wang, J., Ma, L., Zhao, J., Jackson, K. A. (2009), Structural Growth Behavior of Cd_nTe_n Clusters, J. Chem. Phys., 130, 214307 (DOI: 10.1063/1.3147519)
47. Jimenez-Izal, E., Zhai, H., Liu, J. Y., Alexandrova, A. N. (2018), Nanoalloying Pt Clusters with Si, ACS Catal., 8, 8346–8356 (DOI: 10.1021/acscatal.8b02443)

48. Baxter, E. T., Ha, M. A., Cass, A. C., Alexandrova, A. N., Anderson, S. L. (2017), Ethylene Dehydrogenation on Pt Clusters, ACS Catal., 7, 3322–3335 (DOI: 10.1021/acscatal.7b00409)
49. Ha, M. A., Baxter, E. T., Cass, A. C., Anderson, S. L., Alexandrova, A. N. (2017), Boron Switch for Selectivity of Catalytic Dehydrogenation, J. Am. Chem. Soc., 139, 11568–11575 (DOI: 10.1021/jacs.7b05894)
50. Zhai, H., Alexandrova, A. N. (2016), Ensemble-Average Representation of Pt Clusters, J. Chem. Theory Comput., 12, 6213–6226 (DOI: 10.1021/acs.jctc.6b00994)
51. Kresse, G., Hafner, J. (1994), Ab Initio Molecular-Dynamics Simulation of Liquid Metal–Semiconductor Transition, Phys. Rev. B, 49, 14251–14269 (DOI: 10.1103/PhysRevB.49.14251)
52. Perdew, J. P., Chevary, J. A., Vosko, S. H., Jackson, K. A., Pederson, M. R., Singh, D. J., Fiolhais, C. (1992), Generalized Gradient Approximation, Phys. Rev. B, 46, 6671–6687 (DOI: 10.1103/PhysRevB.46.6671)
53. Joubert, D. (1999), From Ultrasoft Pseudopotentials to PAW Method, Phys. Rev. B, 59, 1758–1775 (DOI: 10.1103/PhysRevB.59.1758)
54. Nair, A. S., Anoop, A., Ahuja, R., Pathak, B. (2022), Relativistic Effects in Platinum Nanocluster Catalysis, J. Phys. Chem. A, 126, 1531–1541 (DOI: 10.1021/acs.jpca.1c09981)
55. Teter, M. P., Payne, M. C., Allan, D. C. (1989), Solution of Schrödinger's Equation for Large Systems, Phys. Rev. B, 40, 12255–12263 (DOI: 10.1103/PhysRevB.40.12255)
56. Grimme, S., Antony, J., Ehrlich, S., Krieg, H. (2010), Density Functional Dispersion Correction (DFT-D), J. Chem. Phys., 132, 154104 (DOI: 10.1063/1.3382344)
57. Nørskov, J. K., Rossmeisl, J., Logadottir, A., Lindqvist, L., Kitchin, J. R., Bligaard, T., Jónsson, H. (2004), Origin of the Overpotential for Oxygen Reduction, J. Phys. Chem. B, 108, 17886–17892 (DOI: 10.1021/jp047349j)

58. Zhang, Z., Zandkarimi, B., Munarriz, J., Dickerson, C. E., Alexandrova, A. N. (2022), Fluxionality of Subnano Clusters Reshapes Activity Volcano, ChemCatChem, 14, e202200345 (DOI: 10.1002/cctc.202200345)
59. Doye, J. P. K., Wales, D. J. (1999), Dynamics of Structural Transitions in NaCl Clusters, J. Chem. Phys., 111, 4800–4811 (DOI: 10.1063/1.480465)
60. Bunău, O., Bartolomé, J., Bartolomé, F., Garcia, L. M. (2014), Large Orbital Magnetic Moment in Pt13 Clusters, J. Phys.: Condens. Matter, 26, 196006 (DOI: 10.1088/0953-8984/26/19/196006)
61. Lim, D. H., Wilcox, J. (2012), Mechanisms of ORR on Graphene-Supported Pt Nanoparticles, J. Phys. Chem. C, 116, 3653–3660 (DOI: 10.1021/jp210796e)
62. Ramos-Sanchez, G., Balbuena, P. B. (2013), Interactions of Platinum Clusters with Graphite, Phys. Chem. Chem. Phys., 15, 11950–11960 (DOI: 10.1039/c3cp51791h)
63. Lan, J., Rybkin, V. V., Iannuzzi, M. (2020), Ionization of Water at Aqueous Electrode Interfaces, J. Phys. Chem. Lett., 11, 3663–3669 (DOI: 10.1021/acs.jpclett.0c01025)
64. Calle-Vallejo, F., Martínez, J. I., García-Lastra, J. M., Sautet, P., Loffreda, D. (2014), Fast Prediction of Adsorption Properties for Pt Nanocatalysts, Angew. Chem. Int. Ed., 53, 8316–8319 (DOI: 10.1002/anie.201402958)
65. Rück, M., Bandarenka, A., Calle-Vallejo, F., Gagliardi, A. (2018), Rapid Prediction of Mass Activity of Pt Electrocatalysts, J. Phys. Chem. Lett., 9, 4463–4469 (DOI: 10.1021/acs.jpclett.8b01864)
66. Fung, V., Jiang, D. E. (2017), Structural Diversity and Fluxionality of Pt_n Clusters, J. Phys. Chem. C, 121, 10796–10802 (DOI: 10.1021/acs.jpcc.6b11968)

67. Zibordi-Besse, L., Tereshchuk, P., Chaves, A. S., Da Silva, J. L. F. (2016), Adsorption on Transition-Metal 13-Atom Clusters, J. Phys. Chem. A, 120, 6159–6168 (DOI: 10.1021/acs.jpca.6b03467)
68. Mahata, A., Rawat, K. S., Choudhuri, I., Pathak, B. (2016), Shape-Dependent Catalytic Activity of Pt Nanoclusters, Catal. Sci. Technol., 6, 8210–8220 (DOI: 10.1039/c6cy01709f)
69. Kabsch, W. (1976), A Solution for the Best Rotation to Relate Two Sets of Vectors, Acta Crystallogr. A, 32, 922–923 (DOI: 10.1107/S0567739476001873)
70. Duan, Z., Wang, G. (2011), ORR on Subsurface-Modified Pt(111), Phys. Chem. Chem. Phys., 13, 20178–20187 (DOI: 10.1039/c1cp21687b)
71. Sha, Y., Yu, T. H., Liu, Y., Merinov, B. V., Goddard, W. A. (2010), Solvent Effects on ORR, J. Phys. Chem. Lett., 1, 856–861 (DOI: 10.1021/jz9003153)
72. Man, I. C., Su, H. Y., Calle-Vallejo, F., Hansen, H. A., Martínez, J. I., Inoglu, N. G., Kitchin, J., Jaramillo, T. F., Nørskov, J. K., Rossmeisl, J. (2011), Universality in Oxygen Evolution Electrocatalysis, ChemCatChem, 3, 1159–1165 (DOI: 10.1002/cctc.201000397)
73. Sun, G., Sautet, P. (2018), Metastable Structures in Cluster Catalysis, J. Am. Chem. Soc., 140, 2816–2827 (DOI: 10.1021/jacs.7b11239)
74. Reuter, K. (2016), Ab Initio Thermodynamics and Microkinetics, Catal. Lett., 146, 541–563 (DOI: 10.1007/s10562-015-1684-3)
75. Reuter, K., Scheffler, M. (2002), Composition and Stability of RuO_2(110), Phys. Rev. B, 65, 035406
76. Lym, J., Wittreich, G. R., Vlachos, D. G. (2020), Python Multiscale Thermochemistry Toolbox (PMuTT), Comput. Phys. Commun., 247, 106864 (DOI: 10.1016/j.cpc.2019.106864)

CHAPTER 3

Reshaping the Activity Volcano in the Subnanometer Regime

CHAPTER 3

Section: A

3.1 Introduction

Electrochemical energy conversion technology has emerged as the most promising avenue to address the growing environmental problems and global energy crisis in an eco-friendly, carbon-neutral cycle *[1,2]*. In this regard, developing high-performance electrocatalysts, such as regenerative fuel cells *[3]*, metal-air batteries *[4]*, and proton exchange membrane fuel cells (PEMFCs) *[5]*, presents a formidable challenge despite their potential for sustainable solutions. Specifically, the sluggish kinetics and high reaction overpotential values associated with the oxygen reduction reaction (ORR) at the cathode make it challenging to achieve the ideal conditions and hinder the widespread commercialization of PEMFC technology *[6,7]*. Additionally, the expensive Pt loading and its susceptibility to poisoning under long-term working conditions further decrease the cathodic conversion efficiency, necessitating the development of new materials to reduce the usage of precious metals *[8]*.

To address this challenge, subnano clusters have recently emerged as an important class of electrocatalysts in heterogenous catalysts due to their superior performance and high atomic efficiency, significantly reducing the utilization of expensive metals *[9]*. At finite temperatures, the subnano clusters possess a relatively flat potential energy surface (PES), making them exceptionally dynamic and leading to their non-Arrhenius behavior *[10]*. Owing to their multiple undercoordinated sites, these dispersed molecular units exhibit fluxional behavior, resulting in non-monotonic catalytic activity with respect to cluster size, further distinguishing their catalytic origins from the bulk counterparts *[11,12]*. Recently, Alexandrova and coworkers pioneered the concept of metastability-triggered reactivity in Pt clusters within the subnanometer regime *[13,14]*. Their work demonstrated the breaking of scaling relationships and shifting of the volcano peak from Pt/Pd to Au/Ag for the ORR with limited M_n $(M = Pt,\ Pd,\ Au, \text{and } Ag;\ n = 1 - 6)$ clusters *[15]*. Although important,

a systematic exploration of the non-linear impact of various-sized transition metal subnano clusters on ORR is still lacking. This research gap persists due to the complexity of the extensive range of heterogenous catalysts, particularly within the non-scalable regime. Furthermore, exploring the vast chemical space of electrocatalysts within the size range where 'each atom counts' via high-throughput DFT calculations is challenging and necessitates a data-driven approach to quantify their ORR activity *[16]*.

Recently, machine learning (ML) has emerged as a powerful tool for accelerating the discovery of new catalysts by disentangling complex catalyst-intermediate interactions in various reactions, such as the CO_2 reduction reaction (CO_2RR) *[17,18]*, nitrogen reduction reaction (NRR) *[19]*, hydrogen evolution reaction (HER) *[20,21]*, and oxygen reduction reaction (ORR) *[22]*. For nanoclusters, Panapitiya et al. devised a two-step feature selection using random-forest model to map the CO adsorption on thiolated gold nanoclusters ($Au_{15}(SR)_{13}$) accurately with structural descriptors *[23]*. Similarly, Vlachos and coworkers developed a surrogate energy model by combining ML, global optimization, and cluster expansion (CE) techniques for CO adsorption on Pd_n ($n = 1 - 55$) clusters supported on $CeO_2(111)$ *[24-26]*. Despite offering crucial insights, identifying fundamental determinants of adsorption and catalytic origins of a multistep ORR with varying size and transition elements at a subnanometer regime has remained elusive. Furthermore, the exploration of multifidelity electronic features using ML algorithms to understand the crucial role of d-electron characteristics during intermediate adsorption within the local environment of subnano clusters has not been previously investigated.

Herein, we introduce an ML framework to screen the active electrocatalysts at the subnanometer regime for ORR within a vast chemical space of 3d, 4d, and 5d transition metal clusters, as shown in Scheme 3.1. We demonstrate that the fluxional identity of these clusters causes significant variations in the intermediate adsorption strength, leading to a complex relationship between their activity and size/metal. Subsequently, we utilize

ML algorithms to predict adsorption energy and deduce the complex relationship between the local environment's intrinsic geometric and electronic properties. The versatility of our ML approach is illustrated by accurately mapping the size-dependent shift in the apex of the volcano plot from Pt to Au for ORR activity at the subnanometer regime, compared to the bulk counterparts. Furthermore, we have also traced the electronic origins of the three best-performing catalysts for ORR activity through density functional theory (DFT) calculations.

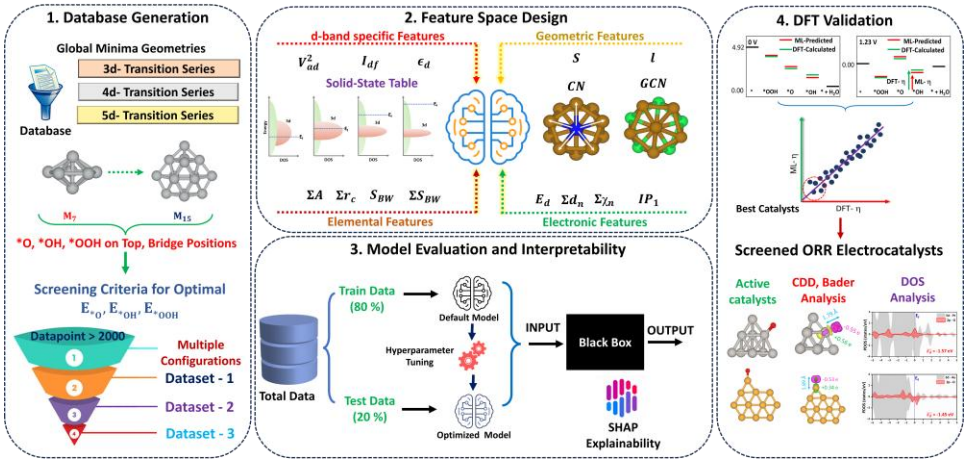


Figure 3.1: Schematic illustration of the ML framework to screen the active electrocatalysts for ORR at the subnanometer regime. The four streamlined processes include (1) the generation of adsorption energy database of *O, *OH, and *OOH intermediate across different transition metals subnano clusters, (2) feature space design to encode the structure-activity relationship of the local chemical environment of subnano clusters, (3) ML model evaluation and interpretability, and (4) screening and DFT validation of the best electrocatalyst.

3.2 Computational Methodology

3.2.1 DFT Details

All the plane-wave spin-polarized density functional calculations were conducted using the Vienna *ab initio* simulation package (VASP) *[27]* with the Perdew-Burke-Ernzerhof (PBE) functional within the generalized gradient approximation (GGA) *[28]*. Ion core and valence electron

interactions were described using the projector augmented wave (PAW) method *[29]*. For relaxation calculations, kinetic energy cut-offs of 500 eV with energy convergence criteria of 10^{-6} eV were employed, and geometric relaxation continued until forces on atoms were smaller than 0.02 eV/Å *[30]*. Gaussian smearing with a sigma value of 0.2 eV was applied and the Γ-centered (1×1×1) k-point grids were considered for the sampling of the Brillouin zone in congruence with the previous reports *[12,31]*. A higher (3 × 3 × 3) k-mesh was used to calculate the density of states (DOS). Bader atomic charges were determined using the Henkelman code with the near-grid algorithm refine edge method *[32]*. All the isosurface values for the charge density difference analysis were set to be 0.006 eV $Å^{-3}$, where pink and green colors represent the charge accumulation and depletion zones, respectively. A sufficiently large box with dimensions 20 × 20 × 20 $Å^3$ was utilized to optimize the distinct geometries to avoid the possibility of spurious interaction between the adjacent images along each axis.

3.2.2 Machine Learning Details

For ML investigation, we have employed five different ML algorithms, namely, catboost regression (CR), gradient boosting regression (GBR), adaboost regression (ABR), random forest regression (RFR), and kernel ridge regression (KRR), as available in the Scikit-learn open-source library *[33]*. The stability and accuracy of all models were evaluated coefficient of determination (R^2), the root-mean-square error (RMSE), and the mean absolute error (MAE), with formulae:

Coefficient of determination (R^2): $$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y}_i)^2}$$

Root Mean Square (RMSE): $$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Mean absolute error (MAE): $$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

where y_i and $\hat{y}_i$ are the DFT calculated and ML predicted values, respectively, $\bar{y}_i$ is the average of the DFT data, and n is the sample size.

RMSE, MAE, and R^2 describe model performance from related but distinct angles, and their interplay helps diagnose both accuracy and reliability. MAE measures the average magnitude of errors in a linear, easy-to-interpret way, while RMSE penalizes larger errors more strongly, making it sensitive to outliers. When RMSE is much higher than MAE, it usually signals occasional large prediction errors. Low values of both RMSE and MAE indicate good predictive accuracy, with RMSE close to MAE suggesting stable error behavior. R^2 captures how well the model explains variance in the data rather than absolute error; high R^2 combined with low RMSE and MAE reflects a strong, well-calibrated model. In contrast, a high R^2 with large errors can indicate that trends are captured but predictions lack precision, while low or negative R^2, even with moderate errors, suggests poor explanatory power and limited generalization.

3.3 Results and Discussion

3.3.1 Adsorption Energy Landscape on Subnano Clusters

Considering the vast chemical space arising from the diverse sizes and shapes of subnano clusters, identifying the most stable geometries at the subnanometer regime is a significant challenge. To address this, we initially extracted the previously reported optimized geometries of global minima (GM) of 3d, 4d, and 5d transition metal subnano clusters *[34]*, denoted as TM_n, where n ranges from 7 to 15, as represented in Figure 3.2a. Notably, the catalysts within this size range represent a diverse array of coordination environments and electronic properties while exhibiting significant fluxionality due to the non-planar structure and absence of bulk facets predominantly, setting them apart from their bulk counterparts. Metals such as Tc and La were excluded due to their toxicity and radioactivity *[35]*.

According to the Bell-Evans-Polanyi (BEP) principle, the stable adsorption energy (E_{ads}) of an intermediate can serve as an important descriptor to evaluate the catalyst's performance *[36,37]*. However, given the inherent fluxional identity and high sensitivity of the local atomic coordination environment, identifying the most stable configuration is essential.

Therefore, to generate our E_{ads} database, we systematically sampled multiple configurations of single intermediate-adsorbed subnano cluster involving principle ORR intermediate (*O, *OH, and *OOH) across various top and bridge positions at a distance (l) (Figure 3.2b-c). Subsequently, the most stable adsorption configurations were considered for our further investigation. The adsorption energies of the ORR intermediates, hereon represented as E_{*O}, E_{*OH} and E_{*OOH}, at site S (top/bridge) of the cluster is computed as:

$$E^{S}_{M_n-{}^{*}O,{}^{*}OH,{}^{*}OOH} = E_{M_n(*O,*OH,*OOH)-}(E_{M_n} + E_{*O,*OH,*OOH}) \quad (3.1)$$

where $E_{M_n(*O,*OH,*OOH)}$, E_{M_n}, and $E_{*O,*OH,*OOH}$ represent the electronic energies of the cluster with *O, *OH, and *OOH at site S, bare cluster and the corresponding intermediates in the gas phase, respectively.

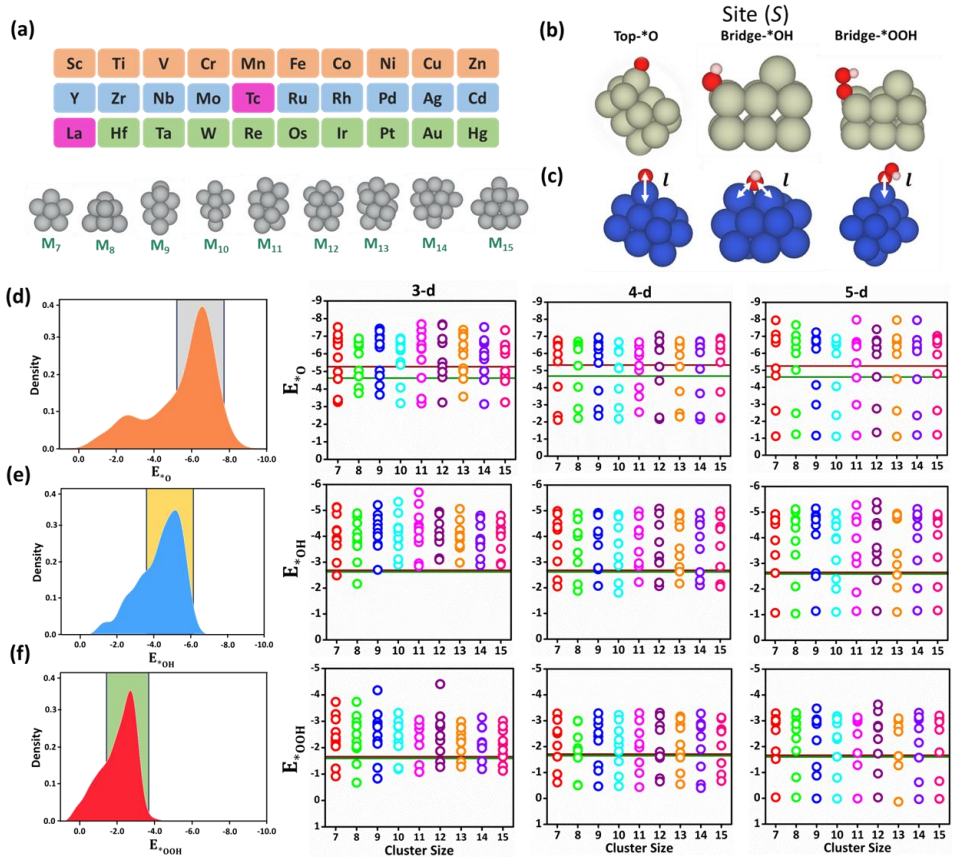


Figure 3.2: Initial Database Generation: (a) A cropped section of the periodic table representing the transition metal (TM_n) subnano clusters selected across 3d, 4d, and 5d series within the size range (n) of 7-15. Pink

blocks denote the excluded TM_n, (b) distinct top and bridge positions (S) considered for intermediate adsorption, (c) bond-length (l) between the intermediate and the adsorption sites on M_n subnano clusters. Range of adsorption energy distribution for (d) E_{*O}, (e) E_{*OH}, and (f) E_{*OOH}, demonstrating variations specific to different sizes and transition metal series in eV. The green and brown solid horizontal lines represent the E_{*O}, E_{*OH}, and E_{*OOH} for Pt (111) surface and Pt_{79} nanocluster, respectively, for comparison with bulk systems and undercoordinated subnano clusters.

The trend of E_{ads} for each intermediate is summarized in Figure 3.2d-f, where the left and right sides represent the density distribution and its variation across the 3d, 4d, and 5d transition series, respectively. Four interesting observations emerged within the diverse distribution of the E_{ads} dataset: (1) the E_{ads} values for each intermediate exhibit a lack of consistent trend, with density of distributions falling within different energy ranges. The peak density distribution of E_{*O}, E_{*OH}, E_{*OOH} ranges from -5.2 to -7.6 eV, -3.7 to -6.0 eV, -1.4 to -3.7 eV, respectively. (2) Compared to bulk systems such as Pt (111) surface and Pt_{79} nanocluster *[38]*, represented by blue and green horizontal lines, respectively, the subnano cluster exhibits stronger binding to *O, *OH, and *OOH intermediates, owing to the presence of undercoordinated atoms. (3) Interestingly, few catalysts exhibit significant negative E_{ads} values, indicating an over-binding of intermediates with active sites, potentially leading to the poisoning of the catalyst. (4) Transitioning from the 3d to 5d TM series reveals a decrease in E_{ads} values with a continuous shift towards the less negative range for each intermediate, highlighting the weaker binding affinity of the 5d catalysts. Within this chemical space, we have curated a comprehensive dataset of 252 subnano catalysts with three ORR intermediates, resulting in 756 E_{ads} datapoints (252 × 3), represented as Dataset-1 in Figure 3.3.

3.3.2 Screening of Optimal Adsorption Energy for ML Design

After acquiring our initial E_{ads} dataset, the subsequent step involved extracting the optimum adsorption energy for each intermediate. According to the Sabatier principle *[39]*, the intermediate-cluster interaction should neither be too strong (E_{*O} = -7.97 as in Ta_{11}) nor too weak (E_{*OOH} = -0.03 eV as in Hg_7). Therefore, to identify the active ORR electrocatalysts within Dataset-1, we employed three screening criteria based on the E_{ads} of each intermediate to eliminate inactive catalysts, as shown in Figure 3.3.

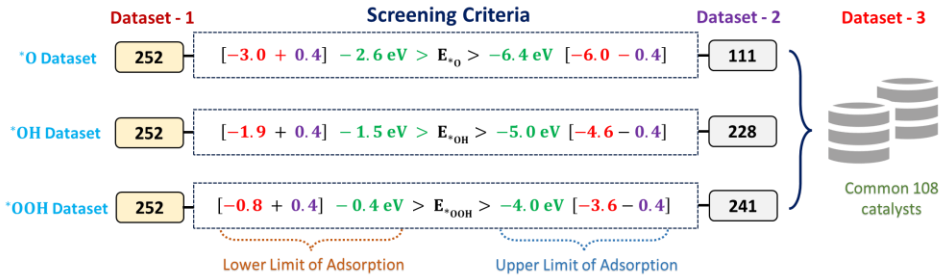


Figure 3.3: Workflow of screening criteria to extract active electrocatalysts for ORR from the adsorption energy database following the Sabatier principle.

Herein, the upper and lower limits for E_{*O}, E_{*OH}, and E_{*OOH} were selected based on the previously reported E_{ads} values of intermediates across various surfaces such as Pt (111) surface, Pt_{79} nanoclusters, and Pt_{1-10} subnanoclusters *[12,38,40-43]*. Additionally, we extended this limit by 0.4 eV on both ends (a total of 0.8 eV), weighted by Boltzmann statistics to account for metastability-triggered reactivity and ensure the robustness of the dataset *[11,12,31,44]*. Following these screening criteria, we extracted our Dataset-2 encompassing 111, 228, and 241 datapoints for E_{*O}, E_{*OH}, and E_{*OOH}, respectively (Figure 3.3). We highlight that this E_{ads} dataset was uniformly sampled across a wide range of chemical elements (28 elements), catalyst sizes (9 atomic sizes in the range of 7-15), and three intermediates, which is pivotal for developing a robust ML model.

3.3.3 Electronic and Geometric Feature Design for Local Adsorption

Prior to the construction of ML models, mining effective features that uniquely define the local environment and accurately reveal the structure-activity relationship is crucial *[19]*. Therefore, to encode the electronic and geometric properties during intermediate adsorption on the subnano clusters, we divided the extracted features into four different categories: (1) Elemental, (2) Electronic, (3) Geometric, and (4) d-band Specific features, tabulated in Table 3.1.

Table 3.1: List of Features Including Elemental, Electronic, Geometric, and d-band Specific Indicators.

Category	Features	Symbol
Elemental	Sum of atomic weight	ΣA
	Sum of covalent radii	Σr_c
	Bulk Wigner Seitz radii	S_{BW}
	Sum of bulk Wigner Seitz radii	ΣS_{BW}
Electronic	d orbital energy	E_d
	Sum of d orbital energy	ΣE_d
	Sum of d electrons	Σd_n
	Sum of electronegativity	$\Sigma \chi_n$
	First ionization potential (eV)	IP_1
Geometric	Number of metal atoms involved in site (top, bridge)	S
	Coordination number	CN
	Generalized coordination number	GCN
	M-O bond length	l
d-band Specific	Coupling matrix	V_{ad}^2
	Idealized d band filling (size dependent)	I_{df}
	Size-dependent d-band center	ϵ_d

The elemental features (ΣA, $\Sigma r_c, S_{BW}, \Sigma S_{BW}$) pertain to the physical description of different TM involved in adsorption, where $\sum A$ and $\sum r_c$ relates to the size and composition of the clusters. However, S_{BW} relates to the neutral radius of atoms in the perfect solid. The electronic features (E_d, ΣE_d, Σd_n, $\Sigma \chi_n$, IP_1) correlates the donation/acceptance capabilities of the active site. Subsequently, the geometric features (S, CN, GCN and l) provides fundamental insights into the influence of the local atomic arrangement of subnano clusters during adsorption. Here, CN represents the

number of atoms directly bonded to the site ensemble, while GCN was introduced to incorporate the impact of the second coordinating sphere *[45]*. Notably, we have included the summation of features ($\sum A, \sum r_c, \sum S_{BW}, \sum E_d, \sum d_n,$ and $\sum \chi_n$) itself as features to distinguish different-sized TM clusters (Table 3.1).

Considering the significant role of the moments of the electronic density distribution of d-states onto site atoms, Hammer and Nørskov introduced a physics-inspired d-band theory to conceptualize the adsorption on transition-metal systems *[46]*. According to this theory, the contribution of metal sp-states is approximated as constant, while the variations in E_{ads} are predominantly governed by the metal d-states, thereby establishing a link between the electronic properties of TM surfaces and E_{ads}. In addition to the d-band center (ϵ_d) and d-band filling (d_f), recent studies have also underscored the effectiveness of reactivity descriptors like d-band center plus half the d-bandwidth (ϵ_d^w), and the maximum of the d-band Hilbert-transform (ϵ_u), that explicitly accounts for the higher-order characteristics of the d-band moments *[47,48]*. Foster and coworkers investigated such descriptors (ϵ_d, ϵ_d^w, and ϵ_u) by establishing a linear correlation between E_{ads} and d_f for bimetallic nanoclusters comprising 55 atoms, for the hydrogen evolution reaction (HER) *[49]*. Although these features have the potential to accelerate the data-driven exploration of catalytic activity of bulk surfaces, their generalizability across subnano clusters has yet to be explored. Moreover, many prototypical electronic descriptors associated with d-band characteristics require self-consistent quantum calculations. To address this limitation, we attempted to encode the d-band model using elemental-specific numeric values provided in the Solid-State Table for surfaces *[50]* and categorized them as a "d-band specific" feature, as shown in Table 3.1. This includes coupling matrix (V_{ad}^2), idealized d band filling (I_{df}) and d-band center with respect to size (ϵ_d), and S_{BW} represents the Bulk Wigner-Seitz radius. Notably, these features govern the inherent

characteristics of d-band electronic states and are easily substitutable without resorting to expensive DFT calculations.

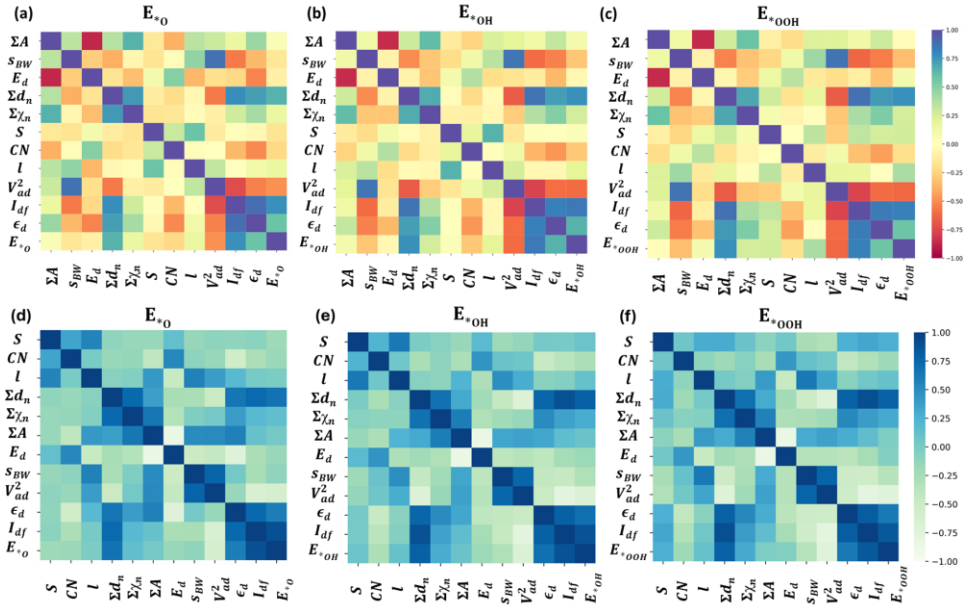


Figure 3.4: Feature-feature and feature-output correlation matrix displaying (a-c) Pearson's correlation coefficient (PCC) and (d-f) Spearman's correlation coefficient (SCC) for E_{*O}, E_{*OH}, and E_{*OOH} datasets in eV. The scale on the right represents the range of correlation coefficients.

To ensure the rationality of the feature set, we quantified the Pearson correlation coefficient (PCC) *[51]* and Spearman's rank correlation coefficient (SCC) *[52]* to assess linear and monotonic dependencies between features and target properties, as represented in Figure 3.4. The variable distribution of correlation between the electronic, geometric, and d-band specific features within both positive and negative ranges in PCC analysis implies that no single feature can exclusively capture the E_{ads} pattern across subnano clusters (Figure 3.4). The pairwise feature distributions are observed to be consistent for each intermediate. A strong positive correlation was observed between CN and GCN, $\Sigma\chi_n$ and IP_1, Σr_c and ΣS_{BW} and E_d and ΣE_d for each intermediate, as shown in Figure 3.4a-c. Except Σd_n, all summation features (Σr_c, ΣS_{BW}, ΣE_d, Σd_n, $\Sigma \chi_n$) considered in our investigation exhibit low correlations with the as E_{*O}, E_{*OH}, and

E_{*OOH}. This suggests that intermediate adsorption on subnano clusters is primarily influenced by site-specific chemical composition, further reflecting their dispersive nature akin to single-atom catalysts. Interestingly, the high correlation observed between the V_{ad}^2, $I_{df,}\epsilon_d$ features and E_{ads} of each intermediate complements the crucial role of the d-band model while investigating the adsorption on TM subnano clusters *[46-48]*. Similarly, a comparable result was observed in the SCC analysis for E_{*O}, E_{*OH}, and E_{*OOH}, as represented in Figure 3.4d-f. To streamline our database and avoid redundancy, we eliminated Σr_c, GCN, ΣS_{BW}, ΣE_d, and IP_1 features, which exhibited strong correlations and minimal contribution towards the output (E_{ads}). Consequently, a final set of features were chosen for the ML analysis where the E_{ads} for each intermediate on different TM_n clusters can be expressed as:

$$E_{TM_n}^{*O,*OH,*OOH} = f\,(S, CN, l, \Sigma d_n, \Sigma\chi_n, \Sigma A, E_d, S_{BW}, V_{ad}^2, \epsilon_d, I_{df}) \qquad (3.2)$$

Ultimately, our final dataset represents the inherent complexities while characterizing the E_{ads} across diverse sites, sizes, and TM_n clusters at the subnanometer regime, thereby demanding a data-driven framework to address this combinatorial problem.

3.3.4 Machine Learning Model Training and Evaluation

Following feature engineering, model training involves identifying suitable parameters for the mathematical formulation of the respective model to minimize the gap between the model's validation results on the specified dataset and real predictions. In our study, the entire dataset was divided into training and testing sets in an 80:20 ratio, ensuring a well-sampled representation of the overall dataset in both sets, rather than using a random splitting method. Subsequently, we employed five different ML algorithms, namely catboost regression (CR), gradient boosting regression (GBR), adaboost regression (ABR), random forest regression (RFR), and kernel ridge regression (KRR), as available in the Scikit-learn open-source library *[33]*. The models are optimized utilizing the randomized searchCV method and evaluated with standard metrics (MAE/RMSE and R^2 score).

Consequently, the best models were employed to predict the E_{*O}, E_{*OH}, and E_{*OOH} values for all the electrocatalysts present in the Dataset-2.

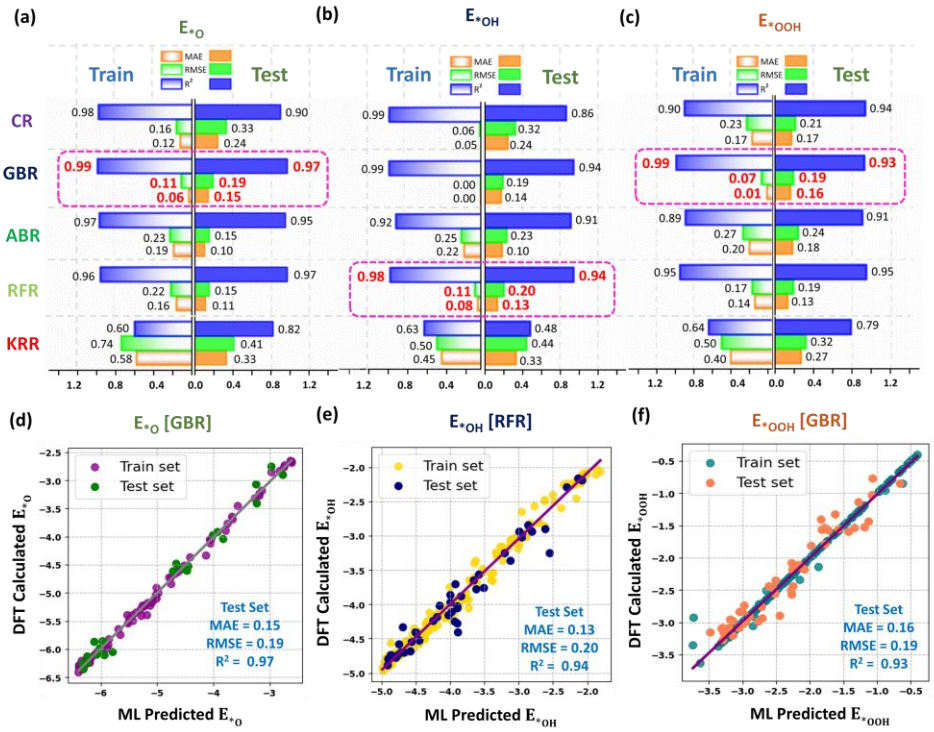


Figure 3.5: Evaluation of ML Models performance: Mean absolute error (MAE), root-mean-squared error (RMSE), and coefficient of determination (R^2 score) for CR, GBR, ABR, RFR, and KRR with (a) E_{*O}, (b) E_{*OH}, and (c) E_{*OOH} datasets in eV. The best-performing model for each intermediate is marked in the rectangular pink-colored box. Parity plots comparing the DFT-calculated and ML-predicted (d) E_{*O}, (e) E_{*OH}, and (f) E_{*OOH} values by the best-performing models.

Figure 3.5a-c represents the final test scores of various ML models for each intermediate. Given the unique distribution of the dataset for each intermediate, different best algorithms were identified for different intermediates. The GBR model exhibited superior performance while predicting E_{*O} and E_{*OOH} compared to the other five ML models, achieving notably low MAE/RMSE values of 0.15/0.19 eV and 0.16/0.19 eV, along with high R^2 scores of 0.97 and 0.93 for test datasets, respectively. Although models such as ABR and RFR exhibited decent performance in

predicting E_{*O} in the test datasets, the high values of MAE and RMSE in training datasets indicate significant instances of underfitting. Similarly, underfitting was observed for the E_{*OOH} datasets with the RFR models. Conversely, the RFR model emerged as the best predictor for E_{*OH}, displaying low MAE/RMSE values of 0.11/0.20 eV, and a high R^2 value of 0.94, compared to the other five models.

Furthermore, the linear correlation between the DFT-calculated and ML-predicted E_{ads} values in the parity plots also show the prediction accuracy of best models for different intermediates with training and testing datasets (Figure 3.5d-f). Therefore, GBR for E_{*O} and E_{*OOH}, and RFR for E_{*OH} were chosen for further investigation to identify the most effective catalyst in our investigation.

3.3.5 Machine Learning Explainability with SHAP Analysis

Further, we conducted a Shapley Additive exPlanations (SHAP) *[53]* analysis with the trained model to improve its interpretability and understand the interplay between extracted features towards the prediction of E_{*O}, E_{*OH}, and E_{*OOH}, as represented in Figure 3.6.

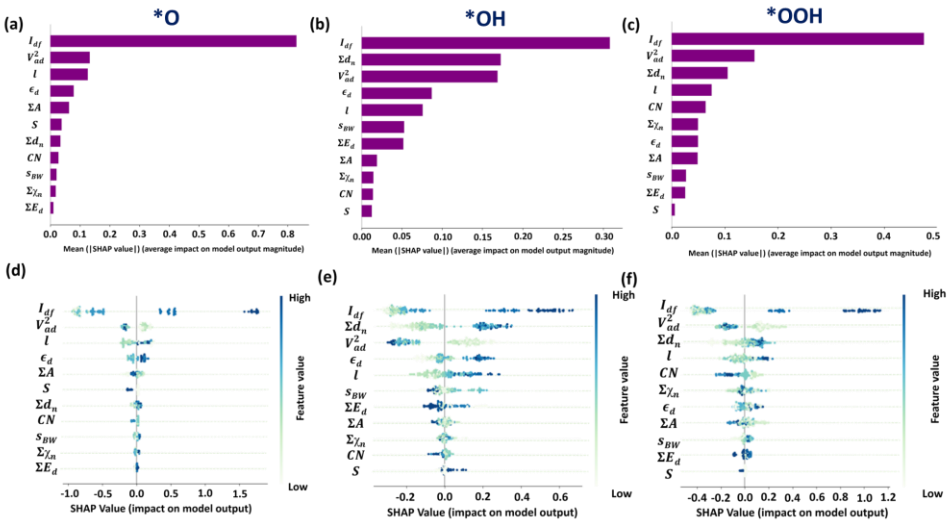


Figure 3.6: SHAP analysis: Global interpretation of the best-performing models for extracted features in predicting (a) E_{*O}, (b) E_{*OH}, and (c) E_{*OOH}. Beeswarm plot that displays an information-dense summary of how the

most important features affect the prediction of (d) E_{*_O}, (e) $E_{*_{OH}}$, (f) and $E_{*_{OOH}}$, respectively. A solitary dot on the feature row represents each datapoint with its SHAP value on the x-axis. The color gradient along the y-axis represents the feature values, ranging from low (white) to high (blue).

In SHAP analysis, the features I_{df} and V_{ad}^2 consistently appeared within the top three features of the bar plots, indicating a significant influence of d-states during intermediate adsorption on subnano clusters (Figure 3.6a-c). For each intermediate, the higher feature values of I_{df} exhibit a positive correlation with the E_{ads} (Figure 3.6d-f). In contrast, higher feature values of V_{ad}^2 demonstrate a negative impact on the model prediction. It is worth noting that I_{df} and V_{ad}^2 values offset each other for each catalyst. Interestingly, these observations for the subnano clusters align with the principles of the d-band theory, where the increase in I_{df} and decrease in V_{ad}^2 values result in repulsive interaction and weak coupling between the metal and adsorbates, ultimately reducing the E_{ads} *[54]*. On the other hand, the bond length (l) has emerged as the primary contributor among the geometric features, with its high feature values driving E_{ads} toward positive directions, suggesting an inverse relationship between l and adsorption strength. Therefore, it can be inferred that both the electronic and geometric features jointly delineate the adsorption characteristics of intermediates within the local chemical environment of subnano clusters.

3.3.6 ML Prediction and Screening of Active Electrocatalysts

3.3.6.1 Validation of ML Predicted Adsorption Energy

To validate the accuracy of our best-performing model, we predicted the E_{*_O}, $E_{*_{OH}}$, and $E_{*_{OOH}}$ values for all catalysts falling within our defined Sabatier range. Subsequently, to screen the active electrocatalysts, we extracted 108 subnano catalysts that possessed E_{*_O}, $E_{*_{OH}}$, and $E_{*_{OOH}}$ values for all intermediates from DFT-calculated and ML-predicted datasets and derived an ORR mechanism via an associative pathway, compiled as

Dataset-3. Further, utilizing the computational hydrogen electrode (CHE) model *[55]* proposed by Nørskov and coworkers, we calculated the reaction energy at 0 V and 1.23 V following the elementary steps outlined below:

$$O_{2(g)} + (H^+ + e^-) + * \rightarrow OOH^* \qquad (\Delta R_1)$$

$$OOH^* + (H^+ + e^-) \rightarrow O^* + H_2O\,(l) \qquad (\Delta R_2)$$

$$O^* + (H^+ + e^-) \rightarrow OH^* \qquad (\Delta R_3)$$

$$OH^* + (H^+ + e^-) \rightarrow H_2O\,(l) + * \qquad (\Delta R_4)$$

The overpotential values (η) for the rate-determining step (RDS) at 1.23 V were computed for the screened 108 catalysts. Note that the $2\,e^-$ pathway leading to the formation of H_2O_2 was excluded due to its unstable adsorption on subnano clusters *[31,43]*.

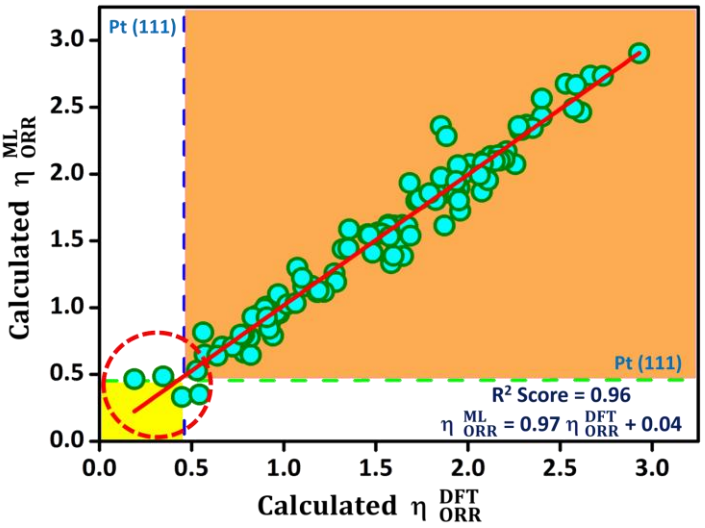


Figure 3.7: Linear relationship between the DFT calculated and ML predicted overpotential values (η) for the screened catalysts. The intersection of green and blue dashed lines corresponds to the η values of Pt (111), i.e., 0.45 V, serving as a reference point to screen the best five catalysts with higher catalytic efficiency, represented within the red circle.

The high correlation (R^2=0.96) and minimal error between the DFT-calculated η values (η_{DFT}^{ORR}) and ML-predicted η values (η_{ML}^{ORR}) for RDS

underscores the capability of our ML framework in capturing the non-monotonic ORR activity for subnano clusters, as represented in Figure 3.7. Compared to the Pt (111) surface, our analysis identified the five best electrocatalysts demonstrating higher (or similar) ORR activity with lower η values (Figure 3.7). Specifically, $Au_9, Au_{11}, Au_{12}, Au_{13}$ subnano clusters emerged as the predominant active electrocatalysts alongside Pt_{10}. Interestingly, these ML observations align with the recent findings by Alexandrova and colleagues, which demonstrated a shift in the apex of the volcano plot from Pd/Pt to Ag/Au using DFT calculations for isolated M_n subnano clusters (M = Pd, Pt, Ag, Au; n= 1-6) for ORR *[15]*. In contrast, Y_{15}, Hf_{14}, and W_8 were identified as inactive catalysts for ORR due to their high η values for RDS as follows: $\eta_{DFT}^{ORR} : \eta_{ML}^{ORR} = 2.93\ V : 2.90\ V$, $\eta_{DFT}^{ORR} : \eta_{ML}^{ORR} = 2.73\ V : 2.73\ V$, and $\eta_{DFT}^{ORR} : \eta_{ML}^{ORR} = 2.66\ V : 2.73\ V$.

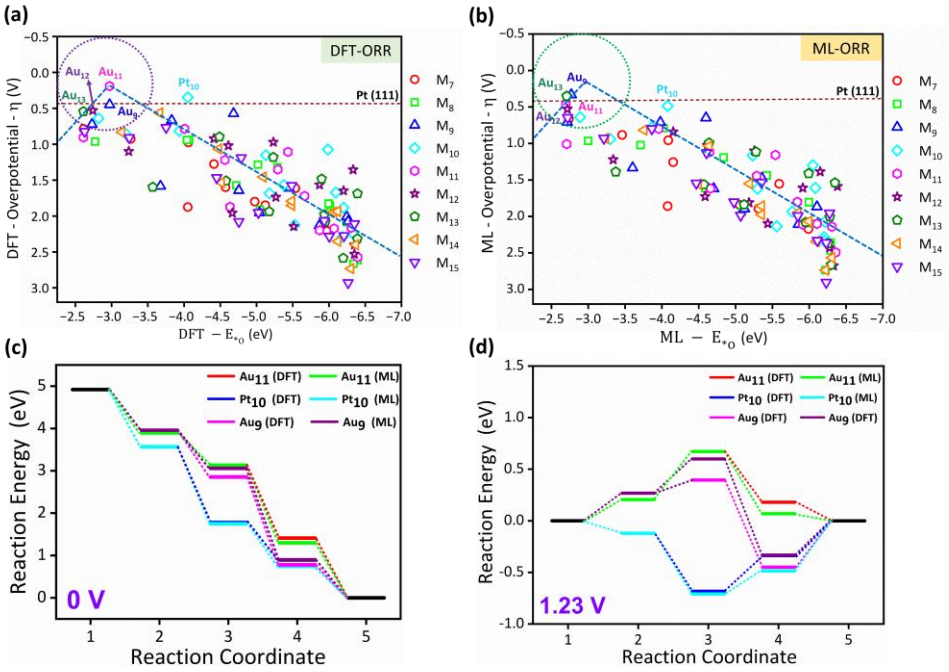


Figure 3.8: Construction of volcano plots for ORR activity using (a) DFT-calculated, and (b) ML-predicted η values in Dataset-3. Catalysts positioned at the apex outperform the Pt (111) surface (marked horizontally in brown color) at the subnanometer regime with the lower η values. The

reaction energy diagram for Au_{11}, Pt_{10}, and Au_9 ORR catalysts at (c) 0 V and (d) 1.23 V, respectively.

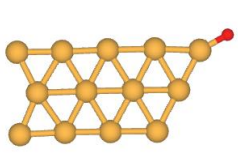
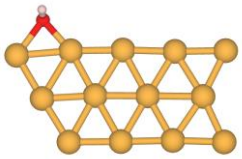
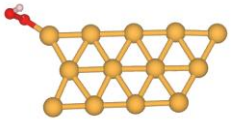
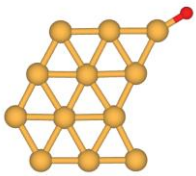
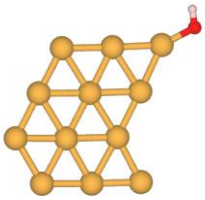
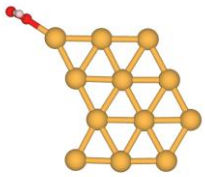
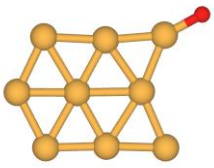
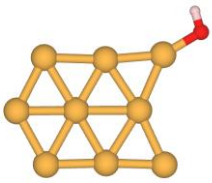
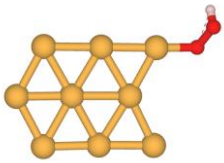
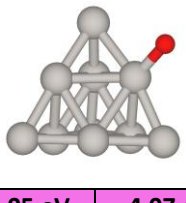
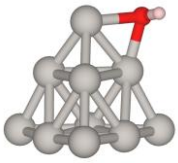
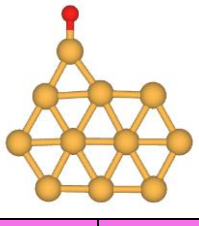
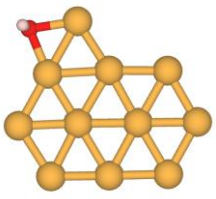
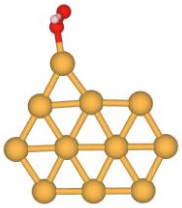
The irregular trends in the η values of the screened subnano clusters in Dataset-3 potentially reconstruct the apex of the volcano plots for ORR at the subnanometer regime, as shown in Figure 3.8a-b. Additionally, our ML framework can accurately map the positions of the best catalysts at the peak of the volcano plots and the distribution of catalysts on both sides of the volcanic distribution, with slight variations from the DFT dataset owing to the small errors in the E_{ads} prediction. Interestingly, within each size range, there exists an active electrocatalyst outperforming Pt and Pd subnano clusters, except for Pt_{10} with $\eta_{DFT}^{ORR}: \eta_{ML}^{ORR} = 0.35\,V : 0.48\,V$. This demonstrates the size-specific ORR activity at the subnanometer regime, which is effectively captured by our ML framework. The good alignment between the DFT and ML reaction energy profile diagrams for Au_{11}, Pt_{10} and Au_9 across different reaction steps, as shown in Figure 3.8c-d, underscores the effectiveness of our approach in identifying potential electrocatalysts.

For enhanced ORR performance in fuel cells, the catalyst at the apex of the activity volcano should exhibit near-optimal adsorption energetics for each intermediate. The difference between DFT-calculation and ML-predicted E_{*O}, E_{*OH}, and E_{*OOH} for the five most promising catalysts is quite acceptable (Table 3.2). These values indicate under-binding compared to the Pt (111) surface, suggesting that O_2 activation is no barrier for these catalysts. Upon closer examination, the E_{*O}, E_{*OH}, and E_{*OOH} values of Au_{11}, Pt_{10}, and Au_9, located at the peak of the volcano, exhibits an optimal energy range within the E_{ads} interval of

$$-4.92\ \text{eV} < \mathbf{E_{ads}} < -1.23\ \text{eV} \qquad (3.3)$$

Table 3.2: DFT-calculated and ML-predicted E_{*O}, E_{*OH}, and E_{*OOH} (in eV) of the five most promising catalysts. The overpotential values (η) for the thermodynamically rate-determining step (RDS), calculated by combining

the E_{*O}, E_{*OH}, and E_{*OOH} values in DFT and ML datasets in Dataset-3 for an associative pathway.

Catalysts	E_{*O} (DFT)	E_{*O} (ML)	E_{*OH} (DFT)	E_{*OH} (ML)	E_{*OOH} (DFT)	E_{*OOH} (ML)
Au_{13}						
	-2.61	-2.71	-2.05	-2.05	-1.27	-1.13
Au_{12}						
	-2.74	-2.72	-2.34	-2.24	-0.95	-0.95
Au_9						
	-2.97 eV	-2.77 eV	-2.49 eV	-2.38 eV	-1.22 eV	-1.22 eV
Pt_{10}						
	-4.05 eV	-4.07 eV	-2.39 eV	-2.53 eV	-1.61 eV	-1.61 eV
Au_{11}						
	-2.97 eV	-2.69 eV	-1.87 eV	-1.89 eV	-1.28 eV	-1.28 eV

Within this specified range, subnano clusters demonstrate an optimum binding affinity with each intermediate and emerge as the most promising candidates, surpassing the Pt (111) surface. Moreover, when compared to both the bulk Pt (111) surface and Pt_{79} nanoclusters *[38]*, the E_{ads} of intermediates is weaker for Au_{11}, Pt_{10}, and Au_9, potentially leading to a decrease in their η values. Therefore, this E_{ads} range in the subnanometer regime could be harnessed to extract active electrocatalysts, potentially disrupting the conventional scaling relationship and volcano plots for the ORR, further replacing the Pt/Pd elements at the peak.

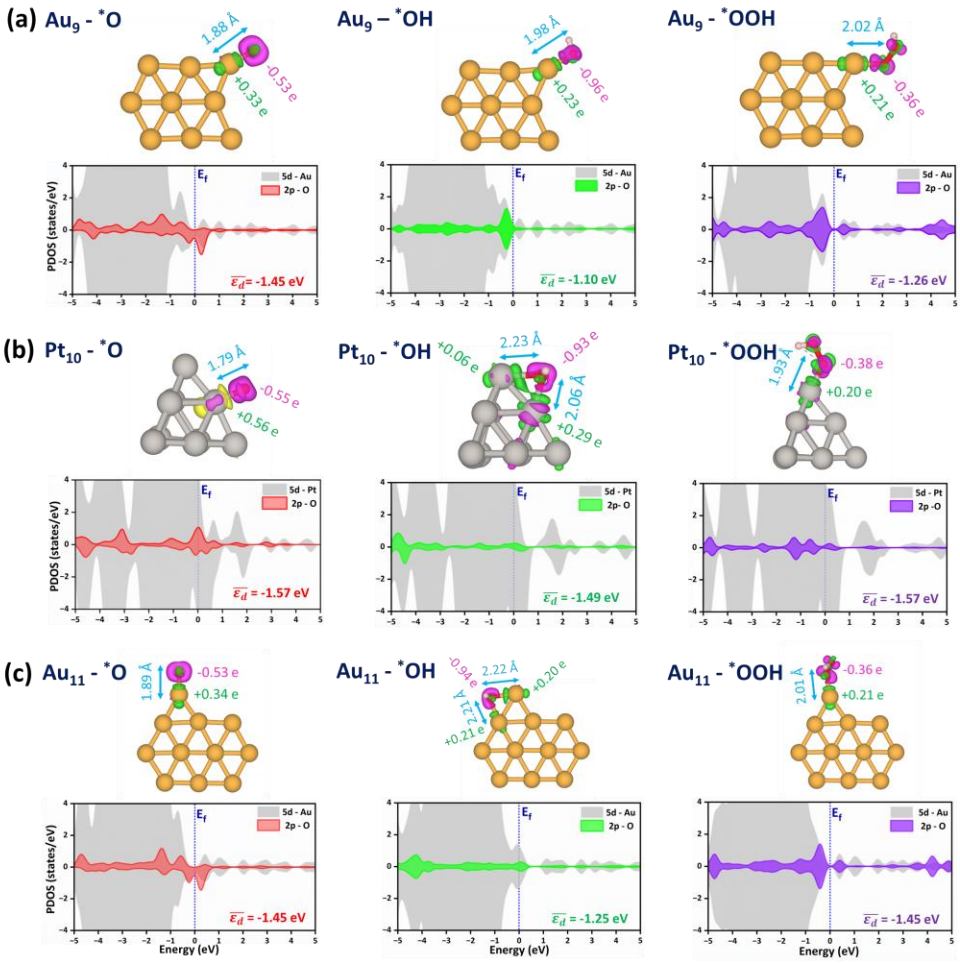


Figure 3.9: The partial density of states (PDOS) and charge density difference (CDD) plots for (a) Au_9, (b) Pt_{10}, and (c) Au_{11} during adsorption of *O, *OH, and *OOH. The Fermi level and the d-band center were represented by E_f and $\overline{\varepsilon_d}$, respectively. In CDD plots, the pink and green

regions represent charge accumulation and depletion, respectively (isosurface = 0.006 e/bohr^3), where the numeric values indicate the charge transfer quantified through the Bader charge analysis in |e| units. The bond length (l) between the intermediate and local adsorption sites is represented in blue for each optimized configuration.

3.3.6.2 Electronic Properties Analysis of Active Catalysts

Since the E_{ads} on TM surfaces are primarily influenced by electronic states, charge transfer, and orbital coupling, we selected Au_{11}, Pt_{10}, and Au_9, exhibiting optimal binding with each intermediate and analyzed their electronic structures to understand the origins of their ORR activity, as shown in Figure 3.9. From partial density of states (PDOS) analysis, the narrow distribution of 5d states near the Fermi level (E_f) in Au_9 and Au_{11} subnano clusters indicate a weaker coupling between $O\ (2p) - Au\ (5d)$ bonding states, resulting in decreased E_{ads} (Figure 3.9a,c). In contrast, the broader distribution of 5d states in Pt_{10} subnano clusters suggest a stronger $O\ (2p) - Pt\ (5d)$ interactions, leading to increased E_{ads} values (Figure 3.9b) *[54]*. Charge density difference (CDD) and Bader charge analysis demonstrate moderate charge transfer from $Au_{11}, Pt_{10}, \text{and } Au_9$ to the O atom of $^*O$ and $^*OOH$ intermediates, signifying their optimal binding strengths. However, during $^*OH$ adsorption on Pt_{10} and Au_{11}, significant charge transfer from the metal atoms to the O atom occurs, likely due to binding at bridging positions where two metal atoms donate electrons to the intermediates. Similarly, notable charge transfers during $^*OH$ adsorption on Au_9 at the top positions aligns with increased overlap between $O\ (2p) - Au\ (5d)$ orbitals near the E_f (Figure 3.9a). These observations are further supported by longer $M - O$ bond in $Au_{11}, Pt_{10}, \text{and } Au_9$ subnano clusters, indicating weaker binding strengths contributing to the optimal adsorption energetics, further reducing η values in the subnanometer regime.

3.3.7 Conclusion

In summary, we represent an ML-assisted approach to break the conventions in electrocatalysis by reshaping the ORR activity volcano plots

at the subnanometer regime. Our findings reveal a shift in the apex of the volcano plot from Pt/Pd to Au, identifying $Au_9, Au_{11}, Au_{12}, Au_{13}$ alongside Pt_{10}, which emerged as the best five catalysts for ORR at the subnanometer regime. We mapped the non-monotonic catalytic distribution in the activity volcano plots, deducing a significant correlation (R^2=0.96) between the ML-predicted and DFT-calculated overpotential values. The non-linear catalytic distribution with respect to size and elemental composition in volcano plots is attributed to the fluxional identity and high sensitivity of the local coordinating environment during the adsorption of $^*O, ^*OH,$ and $^*OOH$ across 252 catalysts in the 3d, 4d, and 5d transition metal series (TM_n), with n = 7-15. By leveraging easily accessible, non-ab initio geometric and electronic features such as bond length, coordination number, idealized d-band filling, and coupling matrix, our trained ML model demonstrated prediction accuracy for *O, *OH, and *OOH adsorption with MAEs of 0.14, 0.11 and 0.16 eV, respectively. Interestingly, the neighboring metal atom's d-band filling and coupling matrix significantly influences the adsorption on the local chemical environment, compared to the d-band center and other electronic/geometric features, bypassing the need for expensive calculations for high-fidelity d-state characteristics. Our observations align with the principles of the d-band theory, where an increase in I_{df} and a decrease in V_{ad}^2 values reduce the E_{ads} values due to repulsive interaction and weak coupling. Finally, we provide an optimal adsorption energy cutoff for each intermediate confirmed by moderate charge transfer and orbital overlap between intermediate and active clusters, enabling rapid screening of the active ORR electrocatalysts within the large chemical space. Our data-driven approach can accelerate the catalyst design and unravel the hidden design principles of the partially explored subnanometer regime.

CHAPTER 3

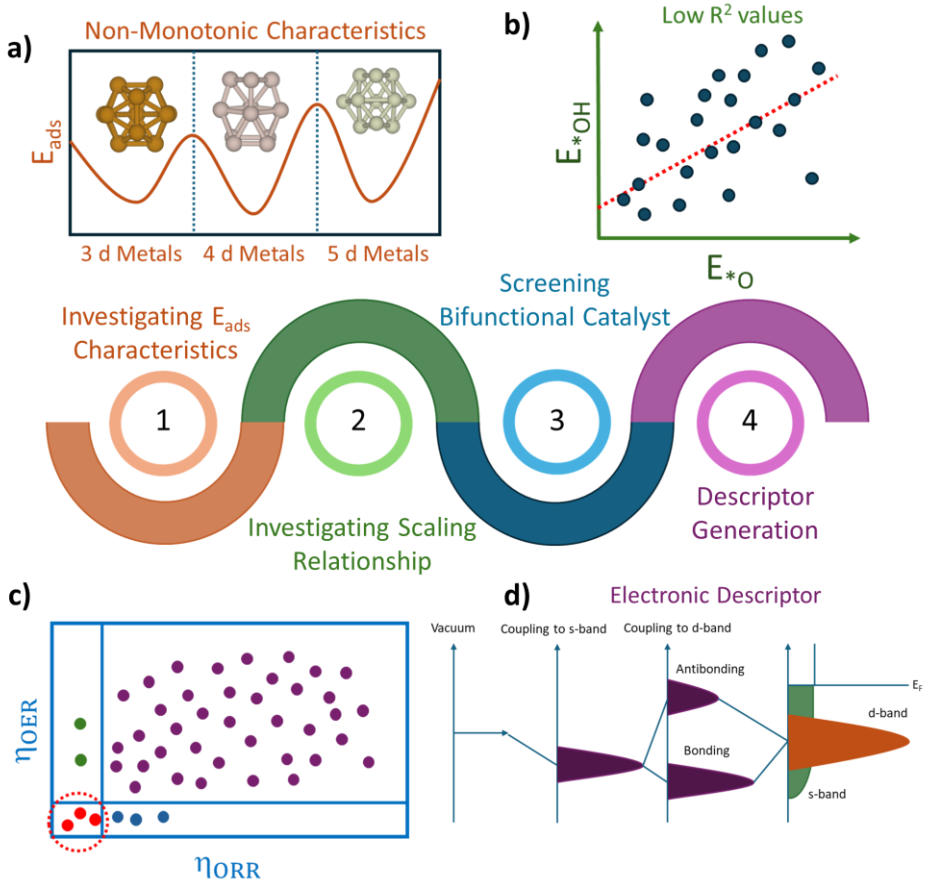
Section: B

3.4 Introduction

Developing sustainable and green technologies for efficient energy conversion is essential to meet the growing energy demands and ensure a secure, sustainable future *[1,2]*. In this context, electrochemical energy conversion technology is the most promising technology, with applications in metal-air batteries, hydrogen-producing devices, and fuel cells *[3-5]*. Notably, catalysts play a crucial role in these devices, facilitating a series of complex multistep reactions at the different electrodes. For instance, the oxygen evolution reaction (OER) and oxygen reduction reaction (ORR) are two core reactions in these energy storage devices that occur at the anode and cathode, respectively, determining the electrochemical reaction rates of these devices *[56,57]*. However, these electrode reactions typically exhibit high overpotential values and sluggish kinetics owing to the multi-electron transfer process, which impedes their widespread commercialization *[6,7]*. Therefore, developing suitable catalysts with high activity is essential for the practical application of energy-related devices. Currently, state-of-the-art electrocatalysts for OER and ORR primarily rely on precious metals such as Pt and IrO_2 *[58-60]*. However, these precious metal's scarcity, low selectivity, and high costs restrict their large-scale application, motivating researchers to develop low-cost electrocatalysts with excellent activity and stability under electrochemical conditions *[61-63]*. Designing efficient bifunctional electrocatalysts that facilitate both the OER and ORR in the same electrolyte is indispensable for the rapid development of regenerative fuel cells.

In this section, we focus on screening bifunctional electrocatalysts for the OER and ORR in the subnanometer regime using density functional theory (DFT) (Figure 3.10). We characterized the fluxional identity of the subnano catalysts by systematically investigating the adsorption energy characteristics across 3d, 4d, and 5d transition metal clusters (TM_n), where n = 7-15. The size-dependent scaling relationship between the adsorption

energies indicates that the catalytic activity of subnano clusters significantly differs from their bulk counterparts. Furthermore, our systematic exploration of the four-electron OER and ORR mechanism in an acidic medium reveals a shift in the apex of the activity volcano plots for OER/ORR activity. Further, to understand the origins of bifunctional activity, we categorized the clusters into three groups and analyzed the impact of the electronic properties of the local chemical environment on the OER/ORR activity. Our study screens potential bifunctional catalysts and establishes a correlation between electronic properties and catalytic activities of subnano clusters, providing valuable guidance for designing efficient catalysts for OER and ORR activities.



Scheme 3.10: Schematic illustration of our DFT framework for screening active bifunctional electrocatalysts for OER/ORR activity in the subnanometer regime. The four streamlined processes include (a)

investigating adsorption energies (E_{ads}) of principal OER/ORR intermediates across different transition metal subnano clusters of varying sizes (TM_n), (b) investigating the scaling relation between OER/ORR intermediates, (c) screening active bifunctional catalysts, and (d) deriving a correlation between electronic descriptors and OER/ORR activity.

3.5 Results and Discussion

3.5.1 Adsorption Characteristics of OER/ORR Intermediates

To understand the OER/ORR activity of the subnano clusters, we analyzed the adsorption characteristics of the principal intermediates ($^*O$, $^*OH$, and $^*OOH$) within the subnanometer regime. Identifying the most stable geometries and adsorption positions for these intermediates is challenging owing to the diverse distribution and vast chemical space of subnano clusters. To address this, we extracted the previously reported optimized global minima (GM) geometries for 3d, 4d, and 5d transition metal subnano clusters, denoted as TM_n (where n = 7-15) *[34]*. Tc and La were excluded from our investigation due to their radioactivity and toxicity *[35]*. TM_n clusters in this size regime predominantly exhibit non-bulk-like facets, with diverse coordination environments and electronic properties, distinguishing them from bulk counterparts like the Pt (111) surface. Unlike the highly symmetric configurations of larger TM_n clusters (n = 38, 55, 79) *[38,49]*, their stable geometries show extensive structural diversity, including non-planar, planar, pentagonal bipyramidal, Mackay icosahedral, distorted icosahedral, pentagonal antiprism, and capped square or triangular bipyramids, among others. The irregular facets provide both saturated (highly coordinated) and unsaturated (low-coordinated) sites, contributing to fluxional behavior and disrupting conventional scaling relationships due to varying binding affinities *[12]*. These structural variations lead to a non-monotonic change in catalytic activity with size and composition, offering opportunities to reshape conventional ORR volcano plots.

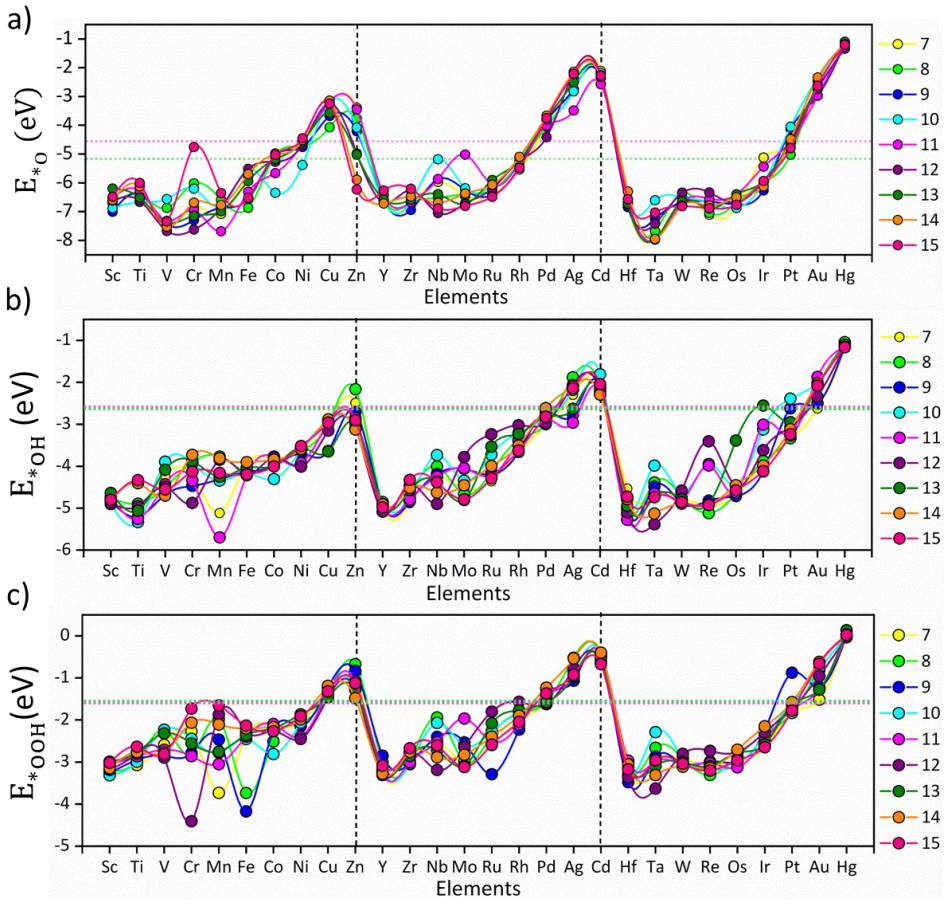


Figure 3.11: Adsorption energy trends of $^*O$, $^*OH$ and $^*OOH$ intermediates: (a) E_{*_O}, (b) $E_{*_{OH}}$ and (c) $E_{*_{OOH}}$ on the investigated TM_n subnano clusters, represented by circles. The colors represent different-sized subnano clusters (n = 7-15), as denoted in the legends. The vertical lines separate the 3d, 4d, and 5d transition metals, and the solid lines connecting the different dot markers are guides for the eye. The pink and green dotted horizontal lines represent the adsorption energy of intermediates on the Pt (111) surface and Pt_{79} nanoclusters, respectively, for comparison with the bulk surfaces.

Given the fluxionality of subnano clusters, resulting from multiple heterogeneous sites, we optimized multiple geometries of the single-intermediate adsorbed onto different active sites (top and bridge) of the clusters (> 2000 configurations). Subsequently, using the Bell-Evans-Polanyi (BEP) principle, which states that a lower activation energy accompanies the most stable adsorption energy (E_a) *[36,37]*, we have

extracted the most stable configurations for further investigation of the OER/ORR reaction mechanism. Notably, 3-fold hollow positions are unstable adsorption sites for OER/ORR intermediates at the subnanometer regime and were excluded from our investigation *[31,43]*. All calculations were performed for optimization using density functional theory (DFT) with the Vienna *ab initio* simulation (VASP) package *[28]*. The adsorption energy of each intermediate (E_{ads}) on the TM_n is computed as follows:

$$E_{ads} = E(TM_nX) - E(TM_n) - E(X) \quad (3.4)$$

where $E(TM_nX)$ is the energy of the intermediate adsorbed geometry of the different TM_n clusters, $E(TM_n)$ is the energy of the bare clusters, and $E(X)$ is the energy of the intermediate in the gas phase.

The size-specific values of E_{ads} for $^*O$, $^*OH$ and $^*OOH$ intermediates across the different TM series are summarized in Figure 3.11. The non-monotonic distribution of E_{*_O}, $E_{*_{OH}}$, and $E_{*_{OOH}}$ represents the fluxional behavior of the subnano clusters, contributing to their variable OER/ORR activity; however, a periodic pattern is observed for each intermediate across different TM series (Figure 3.11). Specifically, E_{*_O}, $E_{*_{OH}}$, and $E_{*_{OOH}}$ generally decrease upon transition from v^1 (metals with one valence electron in the d-orbitals) to v^{10}, indicating that strong electronic repulsion weakens the coupling between the cluster and the intermediates. Compared to the bulk Pt (111) surface and Pt_{79} nanoclusters *[38]*, the majority of the subnano clusters exhibited a shift towards more negative E_{ads}, indicative of stronger binding affinities due to their under-coordinated sites, except for TM_n clusters with v^{10} configurations. The highest E_{ads} are observed for Ta_{13} (-7.96 eV for E_{*_O}), Mn_{11} (-5.69 eV for $E_{*_{OH}}$), and Cr_{12} (-4.41 eV for $E_{*_{OOH}}$), while the lowest are found for Hg_{13} (-1.09 eV for E_{*_O}), Hg_8 (-1.04 eV for $E_{*_{OH}}$), and Hg_{14} (+0.03 eV for $E_{*_{OOH}}$). These extremes represent a non-Sabatier range, where catalysts with very strong or weak E_{ads} may lead to either poisoning or physisorption, thus reducing the overall catalytic

activity. Overall, the distribution of E_{ads} in the subnanometer regime demonstrates a strong dependence on the valence electrons in the d-orbitals.

3.5.2 Analyzing Scaling Relationship at Subnanometer Regime

Scaling relationships are simple linear correlations between the thermodynamic properties of intermediates across different surfaces. However, these relationships impose intrinsic limitations on the maximum activity of the catalyst. Zandkarimi et al. previously demonstrated the breaking of the conventional scaling relationship between the ORR intermediates, attributed to the isomeric diversity and fluxional behavior of subnano clusters *[12]*. Similarly, in our study, we examined the correlation between OER/ORR for our subnano catalysts, as depicted in Figure 3.12.

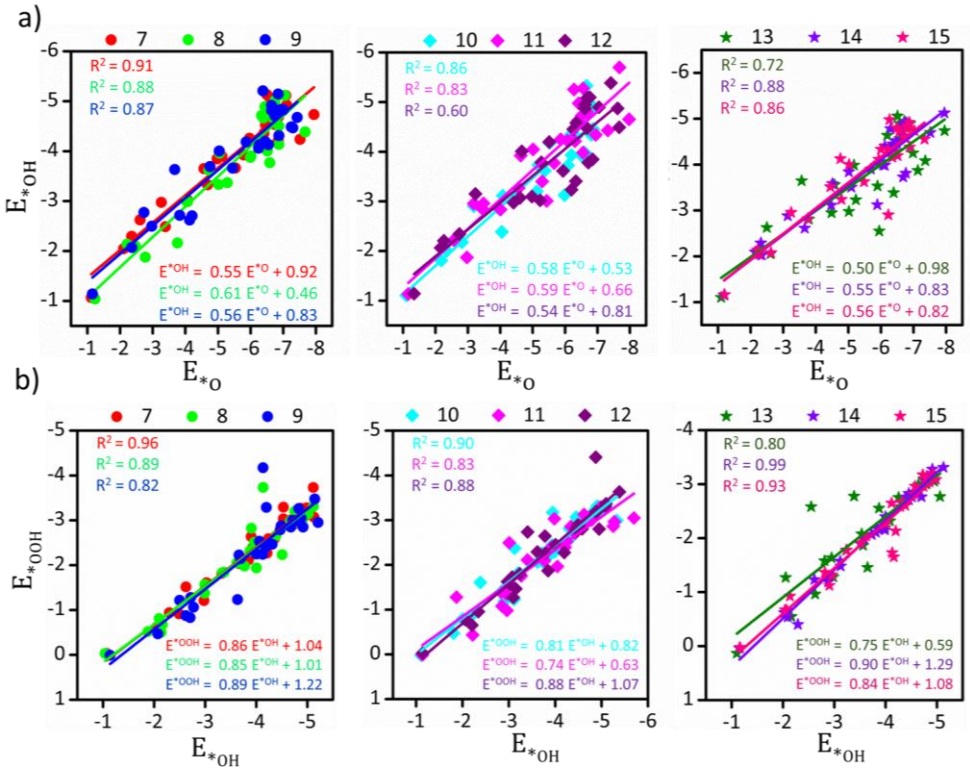


Figure 3.12: Scaling relationship between (a) E_{*O} vs. E_{*OH} and (b) E_{*OH} vs. E_{*OOH} for different-sized TM_n subnano clusters. The sizes are represented in colors on the top panel. The coefficient of determination (R^2) and linear fit equations representing the slope of the best-fitted line corresponding to each size range are shown in the plots.

From Figure 3.12, it is evident that the scaling relationship in the non-scalable regime is size-dependent across different TM series, with TM_7 and TM_{12} representing the highest and lowest values of the coefficient of determination (R^2 = 0.91 and 0.60 for E_{*O} vs E_{*OH}, respectively). Observing the overall distribution (Figure 3.13), we note a scaling relationship for subnano clusters with R^2 = 0.84 and 0.88 (compared to R^2 = 0.91 for bulk surfaces) between E_{*O} vs E_{*OH} and E_{*OH} vs E_{*OOH} intermediates, respectively *[12]*. The variable correlation across different sizes could be attributed to differences in the adsorption sites and changes in the electronic structure, resulting in varying E_{ads} across different surfaces. In the investigation of scaling relationship, the slope of the best-fit line indicates the optimal electron density contribution from the clusters to the bound intermediates, specifically oxygen in our case *[64]*. The computed slopes of E_{*O} vs E_{*OH} and E_{*OH} vs E_{*OOH} varied from the expected values of 0.5 and 1.0, reflecting a decreased electron contribution to the bound oxygen atom (Figure 3.12). This also signifies the inadequacy of the effective medium theorem for small clusters in generalizing E_{ads} across subnano clusters. These results demonstrate the potential of size and transition metal variation to modify the scaling relationship and reshape the activity volcano plots in the subnanometer regime.

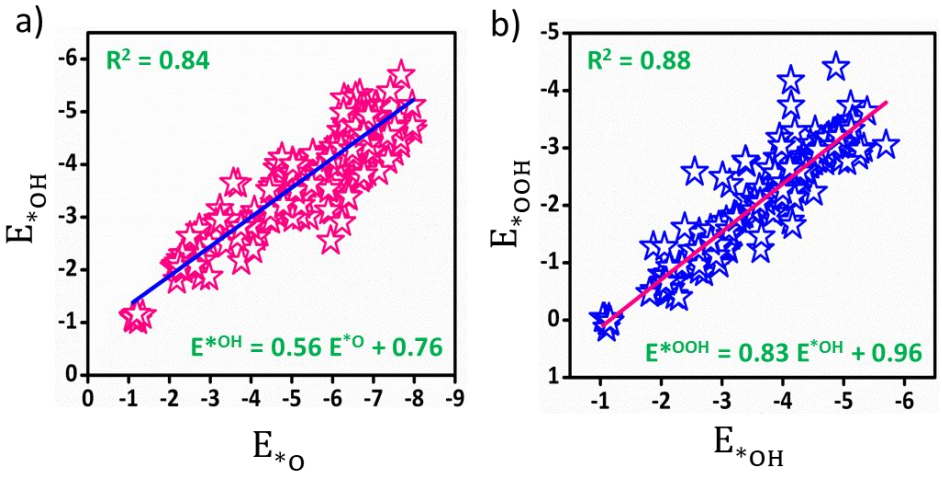


Figure 3.13: Scaling relationship between (a) E_{*_O} vs $E_{*_{OH}}$, and (b) $E_{*_{OH}}$ vs $E_{*_{OOH}}$ for different-sized TM_n subnano clusters.

3.5.3 Evaluating the OER/ORR Electrocatalytic Performance

Next, we derived the OER and ORR reaction mechanisms to screen for active electrocatalysts by simulating a four-electron reaction at 0 V (without any external potential) and 1.23 V (equilibrium potential) in an acidic medium *[55]*. The elementary steps of both reactions are as follows:

In an acidic medium, the OER is considered a four-step process as follows:

$$* + H_2O\,(l) \rightarrow OH^* + (H^+ + e^-) \qquad \mathbf{(\Delta R_1)}$$

$$OH^* \rightarrow O^* + (H^+ + e^-) \qquad \mathbf{(\Delta R_2)}$$

$$O^* + H_2O\,(l) \rightarrow OOH^* + (H^+ + e^-) \qquad \mathbf{(\Delta R_3)}$$

$$OOH^* \rightarrow O_2(g) + (H^+ + e^-) + * \qquad \mathbf{(\Delta R_4)}$$

And the $4e^-$ ORR is the reverse reaction of the OER with four-step as follows:

$$O_2(g) + (H^+ + e^-) + * \rightarrow OOH^* \qquad \mathbf{(\Delta R_5)}$$

$$OOH^* + (H^+ + e^-) \rightarrow O^* + H_2O\,(l) \qquad \mathbf{(\Delta R_6)}$$

$$O^* + (H^+ + e^-) \rightarrow OH^* \qquad \mathbf{(\Delta R_7)}$$

$$OH^* + (H^+ + e^-) \rightarrow H_2O\,(l) + * \qquad \mathbf{(\Delta R_8)}$$

To compare the catalytic activity of the TM_n subnano clusters, the overpotential values (η) of the rate-determining step (RDS) at 1.23 V were utilized to evaluate the OER/ORR performance of the catalyst *[55]*, as depicted in Figure 3.14. The theoretical overpotential at 1.23 V is calculated by the equation:

$$\eta_{OER} = \frac{\max(\Delta R_1, \Delta R_2, \Delta R_3, \Delta R_4)}{e} \qquad (3.5)$$

$$\eta_{ORR} = \frac{\max(\Delta R_5, \Delta R_6, \Delta R_7, \Delta R_8)}{e} \qquad (3.6)$$

The two-electron pathway leading to peroxide formation was not investigated because of the unstable adsorption (breaking) of the H_2O_2 intermediates on subnano clusters, as reported in our previous investigations *[31,43]*. The OER/ORR activity with respect to the size and element in the subnanometer regime is summarized in Figure 3.14.

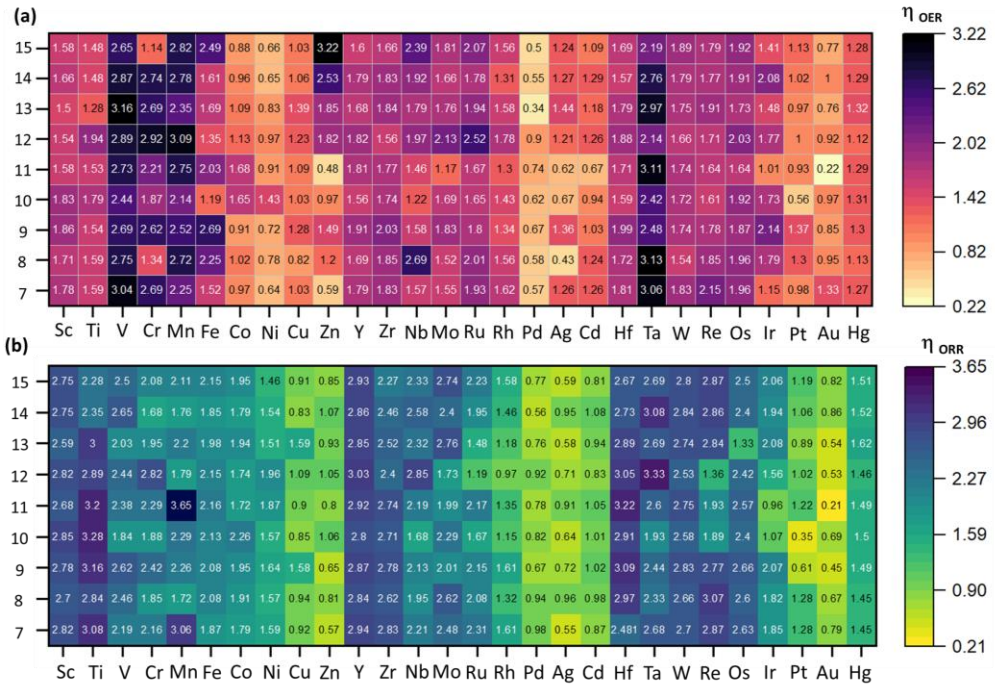


Figure 3.14: Activity heap maps for (a) OER, and (b) ORR activities across different catalysts. The numeric values correspond to the overpotential values (η) of the rate-determining step (RDS) calculated at 1.23 V to measure the catalytic activity. The scale on the right represents the range of η values, where light orange and yellow represent higher activity, and the dark color represents lower activity.

As depicted in Figure 3.14, catalytic activity varies non-monotonically with size and elemental composition in the subnanometer regime. For the OER, Au_{11}, Pd_{13}, and Ag_8 emerged as active electrocatalysts with lower η_{OER} values of 0.22, 0.34, and 0.43 V, respectively, with the RDS involving $^*OH \rightarrow {}^*O$ (ΔR_2) formation (Figure 3.14, 3.15). These η_{OER} values are lower (or comparable) to traditional active OER catalysts, such as RuO_2 (η_{OER} = 0.42

V) and IrO_2 (η_{OER} = 0.56 V) *[65]*. Interestingly, Au_{11} also exhibits superior OER activity compared to previously theoretically reported carbon-based single-atom catalysts like N/C-coordinated graphene (Co-doped), C_2N (Ni-doped), C_3N_4 (Ni-doped), graphdiyne (Co-doped), covalent organic framework (Cu-doped), and metal-organic framework (Co-doped), which show η_{OER} values of 0.46 V, 0.34 V, 0.96 V, 0.64 V, 0.69 V, and 0.29 V, respectively *[66]*. In contrast, Zn_{15}, V_{13}, and Ta_8 are identified as inactive OER electrocatalysts with significantly higher η_{OER} values of 3.22 V, 3.16 V, and 3.13 V respectively, with the RDS of $^*O \rightarrow {}^*OOH$ (ΔR_3) formation.

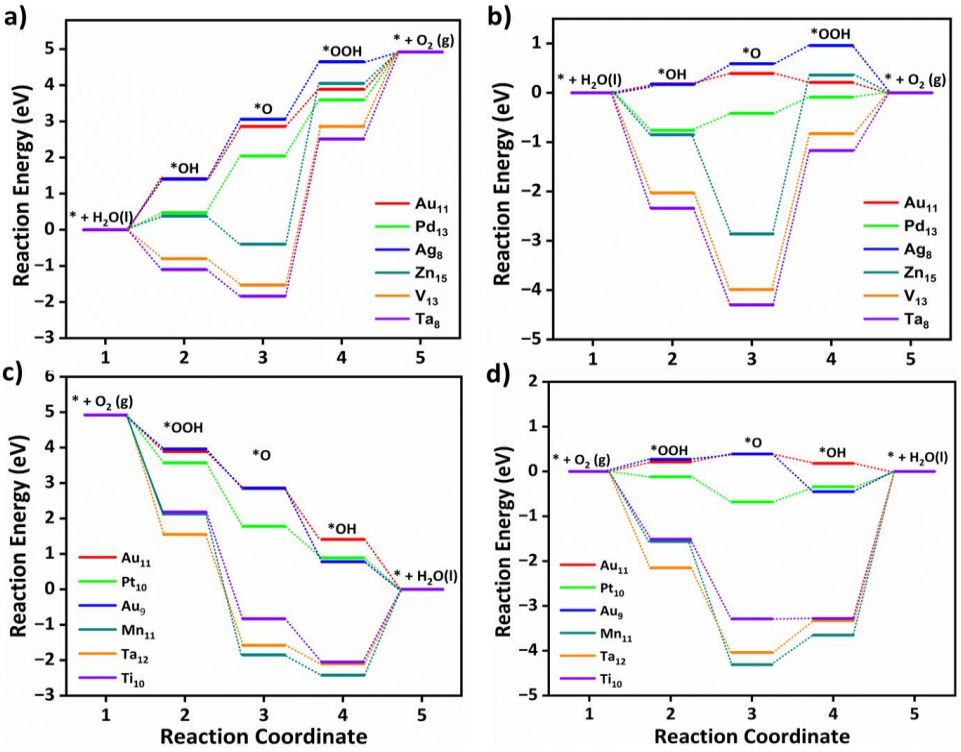


Figure 3.15: Free-energy diagrams of OER elementary steps at (a) 0 V and (b) 1.23 V, and ORR elementary steps at (c) 0 V and (d) 1.23 V.

For the ORR, Au_{11}, Pt_{10}, and Au_9 have emerged as active electrocatalysts with lower η_{ORR} values of 0.21, 0.35, and 0.45 V, respectively. In contrast, Mn_{11}, Ta_{12}, and Ti_{10} are categorized as the poor electrocatalysts with higher η_{ORR} values of 3.65 V, 3.33 V, and 3.28 V, respectively. For most electrocatalysts, the RDS for the ORR involves of $^*OH \rightarrow H_2O$ (l) (ΔR_8)

formation. However, for Ag_8, Ag_{10}, Au_{12}, and Au_{15}, the RDS is $O_2(g) \rightarrow {}^*OOH$ formation (ΔR_5). Compared to an ideal system such as Pt (111) surface with η_{ORR} = 0.45 V *[67]*, these catalysts exhibit enhanced ORR activity. Conversely, for Au_{13}, Au_7, and Au_{11}, the RDS involves the ${}^*OOH \rightarrow {}^*O + H_2O\ (l)$ (ΔR_6) formation, while for Ir_{11}, Ru_{12}, Pt_{10} and Ir_{13}, the RDS constitutes the ${}^*O \rightarrow {}^*OH$ (ΔR_7) formation.

Interestingly, we observe size-dependent reshaping of the OER/ORR activity volcano in the subnanometer regime, where the apex of the plot shifts from benchmarked systems such as RuO_2 and Pt (111) surface to the Au_9, Au_{11}, and Ag_8 subnano clusters. Apart from Au_{11} clusters, Pd_{13} and Pt_{10} clusters have also emerged as active catalysts for OER and ORR activity, respectively. For each active OER/ORR electrocatalysts in the subnanometer regime, the adsorption energy for each intermediate lies within $-4.10 < E_{ads} < -1.23$ eV range (Table 3.3). This range of E_{ads} could potentially be utilized to extract subnano clusters exhibiting optimal adsorption energetics and higher activity in the subnanometer regime.

Table 3.3: Adsorption energies (E_{ads}) of ${}^*O$, ${}^*OH$, and ${}^*OOH$ for the OER/ORR active and inactive electrocatalysts. Each catalyst's overpotential values (η) were calculated at 1.23 V for the rate-determining step (RDS).

Reaction	Activity	Catalysts	*O	*OH	*OOH	η (V)
OER	Active	Au_{11}	-2.97	-1.87	-1.28	0.22
		Pd_{13}	-3.79	-2.81	-1.58	0.34
		Ag_8	-2.77	-1.88	-0.52	0.43
	Inactive	Zn_{15}	-6.23	-2.90	-1.12	3.22
		V_{13}	-7.36	-4.08	-2.32	3.16
		Ta_8	-7.67	-4.39	-2.66	3.13
ORR	Active	Au_{11}	-2.97	-1.87	-1.28	0.21
		Pt_{10}	-4.05	-2.39	-1.61	0.35
		Au_9	-2.97	-2.49	-1.22	0.45
	Inactive	Mn_{11}	-7.68	-5.69	-3.05	3.65
		Ta_{12}	-7.41	-5.38	-3.63	3.33
		Ti_{10}	-6.66	-5.33	-2.99	3.28

From Figure 3.14, it is evident that TM_n clusters with v^1-v^5 electronic configurations exhibited higher η_{OER}/η_{ORR} values, rendering them unsuitable for both reactions in the subnanometer regime. Interestingly, compared to the early transition metal subnano clusters, the late transition metal subnano with $v^8 - v^{10}$ configurations exhibited a significant decrease in η_{OER} and η_{ORR} values (Figure 3.14), making them suitable for fuel cell electrocatalysis.

3.5.4 Screening of Potential Bifunctional OER/ORR Subnano Catalysts

In fuel cells, OER corresponds to the charging process, while the ORR is associated with the battery discharge process. Therefore, to screen bifunctional subnano catalysts that facilitate both the OER and ORR in the subnanometer regime, we computed a widely accepted descriptor $\eta_{Bi} = \eta_{OER} + \eta_{ORR}$ *[68]*, where catalysts with lower η_{Bi} exhibit high electrocatalytic performance, as shown in Figure 3.16. The bifunctional catalytic activity demonstrates nonlinear variation with size and elemental composition, with subnano clusters featuring v^9/v^{10} configuration exhibit lower η_{Bi} values. Conversely, subnano clusters with v^1-v^5 configuration exhibit significantly higher η_{Bi} values, indicating their inactivity as electrocatalysts in the subnanometer regime. Interestingly, Au_{11} emerged as the best bifunctional electrocatalysts with a low $\eta_{Bi} = 0.43$ V. This value is even lower than the 1.01 V of Ru/C *[69]*, which has been reported as the best bifunctional electrocatalyst. In contrast, Mn_{11} emerged as an inactive bifunctional electrocatalysts with a high $\eta_{Bi} = 6.40$ V.

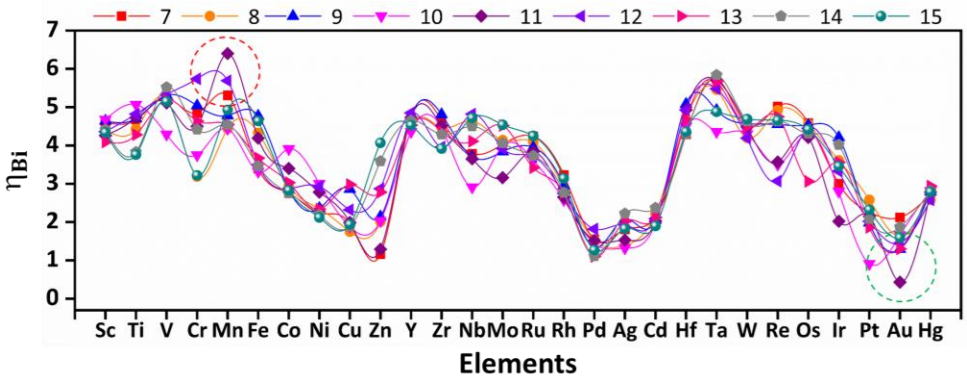


Figure 3.16: Trends of bifunctional activity across 3d, 4d, and 5d TM_n subnano clusters, computed as $\eta_{Bi} = \eta_{OER} + \eta_{ORR}$. The solid lines connecting different dot markers are guides for the eye, representing different sizes and elements. The colors represent different-sized subnano clusters (n = 7-15), as denoted in the legends. The active and inactive bifunctional catalysts are enclosed within green and red circles, respectively.

3.5.5 Descriptors for Bifunctional Catalytic Activity

Furthermore, to screen the potential bifunctional electrocatalysts and gain deeper insights into the electronic descriptors affecting their activity, we categorized the bifunctional catalysts into three clusters: active, medium active, and inactive catalysts based on the η ranges listed in Table 3.4. The higher and lower limits of the η_{OER} and η_{ORR} values were approximated based on the results of previous literature *[3,22,31,40,42,43,68-71]*.

Table 3.4: Range for OER and ORR active electrocatalysts for screening active bifunctional active catalysts.

Cluster	OER Range (V)	ORR Range (V)	Instances
Cluster 1	$0.00 < \eta < 0.40$	$0.00 < \eta < 0.45$	1
Cluster 2	$0.45 < \eta < 0.80$	$0.45 < \eta < 0.80$	10
Cluster 3	$0.80 < \eta$	$0.80 < \eta$	251

Furthermore, to elucidate the origin of bifunctional activities in our investigation, it is crucial to evaluate the descriptors that significantly influence OER and ORR activities. Therefore, to encode the electronic characteristics of TM_n in our investigation, we extracted descriptors into three different categories: (1) elemental, (2) electronic, and (3) d-band specific features, as tabulated in Table 3.5. The elemental features provide a physical description of the TM. However, the electronic descriptors pertain to the electron acceptance/donation capability of different TM. Previous investigations have evidenced the d-band center (ϵ_d) as an effective descriptor to relate the catalyst's electronic structure to the

intermediates' adsorption strength *[46,47,49,50]*. However, considering the vast chemical space in our investigations, it becomes prohibitively expensive to extract ϵ_d with self-consistent quantum calculations. To circumvent this, we attempted to encode the elemental-specific numeric values of the d-band characteristics from the Solid-State Table (relative to Cu) provided for the surfaces *[50]*. Each feature regulates the inherent d-band electronic characteristics, which can be substituted without expensive DFT calculations. Note that the individual features (S_{BW}, E_d, IP_1, V_{ad}^2, I_{df}, and ϵ_d) correlates with the single metal atoms of the subnano clusters; however, the summation features (ΣA, Σr_c, ΣS_{BW}, ΣE_d, Σd_n, and $\Sigma \chi_n$) are included to differentiate between the different-sized metal clusters.

Table 3.5: List of Descriptors Including Elemental, Electronic, and d-band Specific Features.

Category	Features	Symbol
Elemental	Sum of atomic weight	ΣA
	Sum of covalent radii	Σr_c
	Bulk Wigner Seitz radii	S_{BW}
	Sum of bulk Wigner Seitz radii	ΣS_{BW}
Electronic	d orbital energy	E_d
	Sum of d orbital energy	ΣE_d
	Sum of d electrons	Σd_n
	Sum of electronegativity	$\Sigma \chi_n$
	First ionization potential (eV)	IP_1
d-band Specific	Coupling matrix	V_{ad}^2
	Idealized d band filling (size dependent)	I_{df}
	Size-dependent d-band center	ϵ_d

We analyzed the correlation plots of OER/ORR catalysts after categorizing them into active, medium active, and inactive catalysts, as shown in Figure 3.17a. The linear correlation between the feature-feature and feature-OER/ORR activity can be assessed using the Pearson correlation coefficient (PCC) *[51,52]*. Most pairwise feature distributions, as shown in Figure 3.17b, exhibited low correlation ($|PCC| < 0.8$), indicating their independent influence on catalytic activity, and were allowed to coexist. Note that

features such as ΣA, Σr_c, E_d, and ΣE_d exhibited a significantly low correlation with the OER and ORR activity (|PCC| ~ 0), suggesting they do not predict the changes in OER and ORR activity. Consequently, we removed highly correlated features that exhibited a low impact on ORR/OER activity from our final dataset to avoid data redundancy. The final list of features and their correlations is provided in Table 3.6. Overall, the OER and ORR activities of the subnano clusters can be described as follows:

$$\eta_{\mathrm{OER/ORR}} = f\left(S_{BW}, \Sigma S_{BW}, \Sigma E_d, \Sigma d_n, \Sigma \chi_n, IP_1, I_{df}, \epsilon_d\right) \tag{3.7}$$

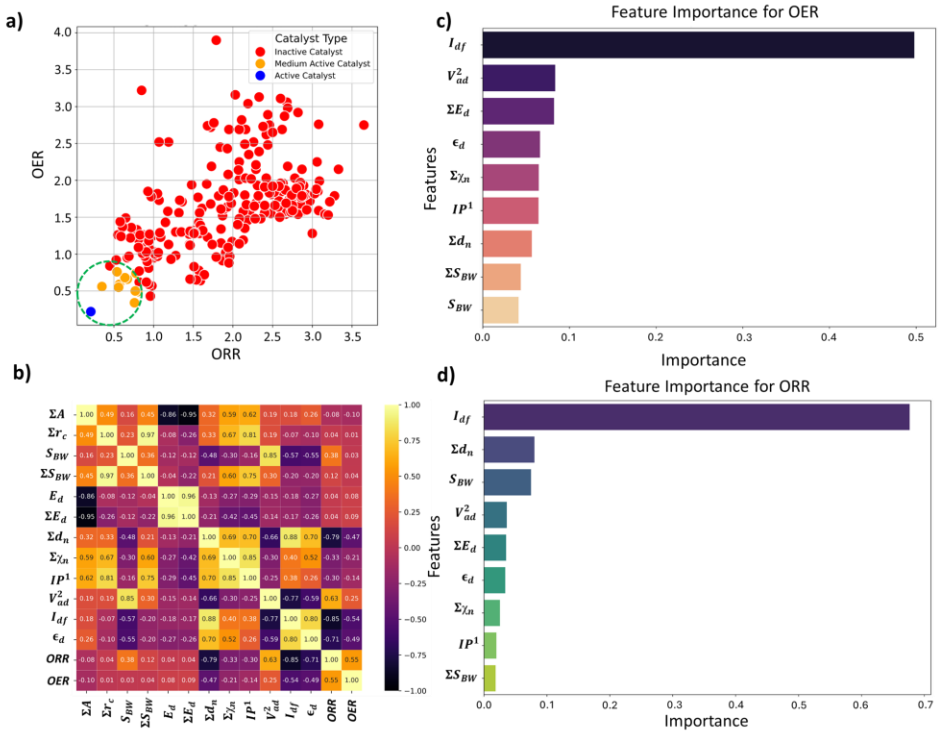
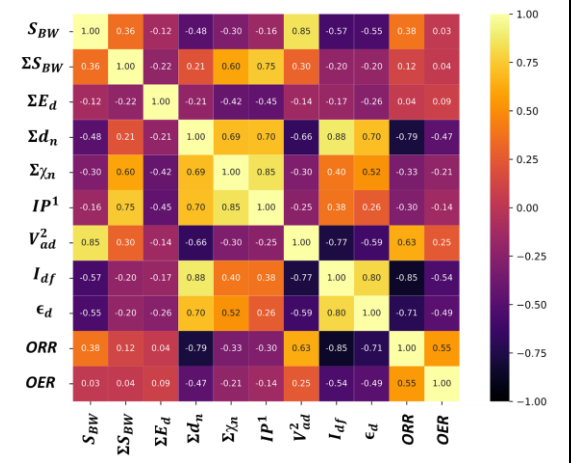


Figure 3.17: (a) Distribution of OER and ORR overpotential values (η) to screen active bifunctional catalysts. (b) Feature-feature and feature-output correlation matrix displaying Pearson's correlation coefficient (PCC) for OER and ORR datasets. The scale on the right represents the range of correlation, where the yellow and black colors represent perfect positive and negative correlations, respectively. Feature importance for (c) OER and (d) ORR activity.

Table 3.6: List of Descriptors Including Elemental, Electronic, and d-band Specific Features Exhibiting Low Correlations.

Features	Symbol	Pearson Correlation Coefficient
Bulk Wigner Seitz radii	S_{BW}	
Sum of bulk Wigner Seitz radii	ΣS_{BW}	
Sum of d orbital energy	ΣE_d	
Sum of d electrons	Σd_n	
Sum of electronegativity	$\Sigma \chi_n$	
First ionization potential (eV)	IP_1	
Coupling matrix	V_{ad}^2	
Idealized d band filling (size dependent)	I_{df}	
Size-dependent d-band center	ϵ_d	

Following the feature selection process, we employed the Random Forest regression (RFR) to assess the feature importance towards OER and ORR activity, as shown in Figure 3.17c-d. Introduced by Breiman and Cutler, the RFR method is based on an ensemble of decision trees from which the prediction of a continuous variable is provided as the average of the prediction of all trees *[72]*. Most importantly, the RFR model evaluates the significance of descriptors in the model by sequentially replacing each descriptor with random noise and observing the resulting decline in model performance, which is measured by the change in the mean-square-error (MSE) for the out-of-bag (OOB) validation data when the descriptor is replaced. Interestingly, I_{df} demonstrates the highest and equal contribution towards the OER and ORR activities, suggesting its effectiveness in identifying bifunctional active catalysts in the subnanometer regime. However, the contributions of the other features are low and offset each other, reflecting their unidirectional utilization towards either OER or ORR activity. Note that the d-band model developed by Hammer and Nørskov systematically correlates the perturbations in the adsorbate to the position of the ϵ_d relative to the Fermi level (E_f), is limited to pure transition metals

and certain alloys; however, it fails to describe interactions on more complex metal surfaces adequately *[73-76]*. In contrast, the I_{df} outperformed the ϵ_d at the subnanometer regime, effectively capturing the elemental-specific d-state contributions of the unique local coordination environment while accounting for the perturbations introduced by the complex chemical environment of subnano clusters.

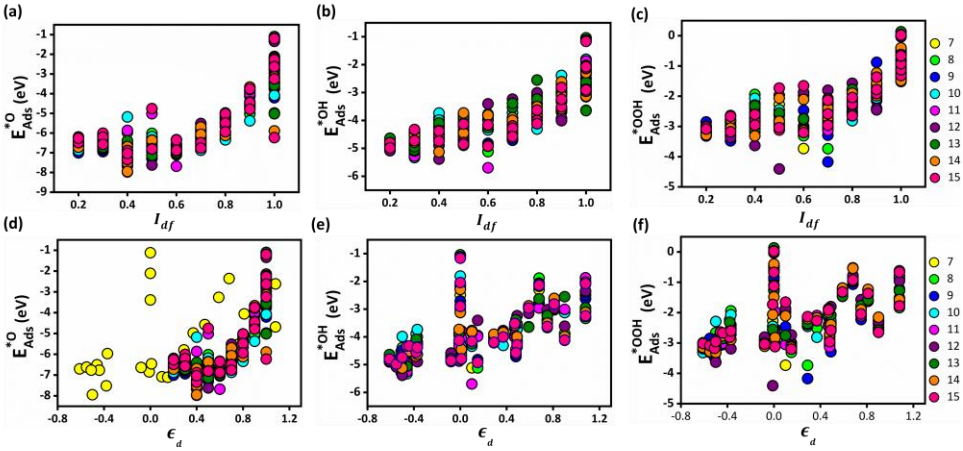


Figure 3.18: Distribution analysis between **(a-c)** E_{*_O}, $E_{*_{OH}}$, and $E_{*_{OOH}}$ and I_{df}, and **(d-f)** E_{*_O}, $E_{*_{OH}}$, and $E_{*_{OOH}}$ and ϵ_d.

Upon closer examination of the active bifunctional catalyst Au_{11}, a higher I_{df} value (1.0) resulted in weak coupling between the metal and intermediates, driving the E_{ads} values of intermediates towards lower levels, leading to optimum binding energetics (Figure 3.18). In contrast, lower I_{df} values (0.6) observed for inactive bifunctional catalysts, such as Mn_{11}, resulted in stronger binding, which ultimately increased the η values. Similar trends were observed for high values of ϵ_d, which reduced binding strength and consequently lowered η values (Figure 3.18). The feature ΣS_{BW} consistently appeared at the bottom of the plots with its minimal impact, demonstrating its lowest contribution to the OER and ORR activities. For enhanced OER/ORR bifunctional activity, catalysts at the apex of activity volcano plots should exhibit near-optimal adsorption energetics for each reaction intermediate. Similarly, Au_{11} bifunctional catalyst achieves optimal adsorption energies for $^*O$, $^*OH$, and $^*OOH$ within

the range of -4.92 eV < E_{ads} < -1.23 eV, suggesting weaker binding than the Pt (111) surface. Additionally, to understand electronic structures during the OER/ORR intermediate adsorption on Au_{11} subnano clusters, we performed a partial density of states (PDOS) analysis, represented in Figure 3.19. The narrow distribution of 5d states near the Fermi level (E_f) indicates weaker coupling between Au (5d) and O (2p) states, further driving the E_{ads} values towards the optimal range. Overall, our study emphasizes the significant contribution of idealized d-band filling while encoding the complex relations of OER/ORR activities at the subnanometer regime, presenting a new pathway for designing efficient bifunctional electrocatalysts.

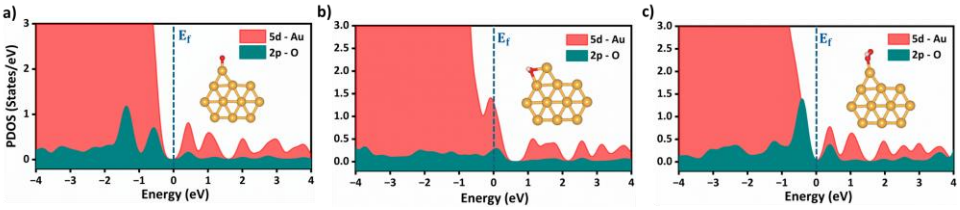


Figure 3.19: Partial density of states (PDOS) for *O, *OH and *OOH adsorbed Au_{11} bifunctional catalysts. The Fermi level at 0.00 eV is represented as (E_f). The inset represents the stable adsorption configurations of the reaction intermediates onto Au_{11} subnano clusters.

3.5.6 Conclusions

In summary, we demonstrate a breaking of conventions for OER and ORR activity across 3d, 4d, and 5d transition metal TM_n subnano clusters. Owing to the fluxionality and non-monotonic catalytic activity at the subnanometer regime, we observe a shift in the apex of the activity volcano from Pt (111)/IrO_2 to Au_{11} clusters. Compared to early transition metal clusters (with $v^1 - v^5$ configurations), the late transition metal clusters (with $v^8 - v^{10}$ configurations) exhibited significantly reduced overpotential values for OER/ORR activity, highlighting their potential in the theoretical and experimental screening of bifunctional catalysts. Interestingly, Au_{11} emerged as a top bifunctional electrocatalyst exhibiting low bifunctional

overpotential values (η_{Bi}) of 0.43 V. Further, impinging on the non-ab *initio* electronic descriptors, we demonstrate the significant contribution of the idealized d-band filling (I_{df}) feature towards the bifunctional activity at the subnanometer regime. Additionally, our observations also align with the d-band theory where higher values of I_{df} and d-band center (ϵ_d) lead to optimal reaction energetics, thereby reducing the η_{Bi} values due to weak intermediate binding strengths. Thus, our work correlates the electronic descriptors of the subnano clusters with their OER and ORR activities, opening new avenues for effectively designing and screening efficient bifunctional cluster catalysts.

3.6. References

1. Lewis, N. S. (2016), Research Opportunities to Advance Solar Energy Utilization, *Science*, 351, aad1920 (DOI: 10.1126/science.aad1920)
2. She, Z. W., Kibsgaard, J., Dickens, C. F., Chorkendorff, I., Nørskov, J. K., Jaramillo, T. F. (2017), Combining Theory and Experiment in Electrocatalysis: Insights into Materials Design, *Science*, 355, eaad4998 (DOI: 10.1126/science.aad4998)
3. Huang, Z. F., Wang, J., Peng, Y., Jung, C. Y., Fisher, A., Wang, X. (2017), Design of Efficient Bifunctional Oxygen Reduction/Evolution Electrocatalyst: Recent Advances and Perspectives, *Adv. Energy Mater.*, 7, 1700544 (DOI: 10.1002/aenm.201700544)
4. Shinde, S. S., Lee, C. H., Jung, J. Y., Wagh, N. K., Kim, S. H., Kim, D. H., Lin, C., Lee, S. U., Lee, J. H. (2019), Unveiling Dual-Linkage 3D Hexaiminobenzene Metal–Organic Frameworks toward Long-Lasting Reversible Zn–Air Batteries, *Energy Environ. Sci.*, 12, 727–738 (DOI: 10.1039/c8ee02679c)
5. Shi, Z., Li, J., Wang, Y., Liu, S., Zhu, J., Yang, J., Wang, X., Ni, J., Jiang, Z., Zhang, L., Wang, Y., Liu, C., Xing, W., Ge, J. (2023), Customized Reaction Route for Ruthenium Oxide toward Stabilized Water Oxidation in High-Performance PEM Electrolyzers, *Nat. Commun.*, 14, 1260 (DOI: 10.1038/s41467-023-36380-9)

6. Wang, Y., Wang, D., Li, Y. (2021), Fundamental Comprehension and Recent Progress in Advanced Pt-Based ORR Nanocatalysts, *SmartMat.*, 2, e1023 (DOI: 10.1002/smm2.1023)
7. Zhang, L., Roling, L. T., Wang, X., Vara, M., Chi, M., Liu, J., Choi, S. I., Park, J., Herron, J. A., Xie, Z., Mavrikakis, M., Xia, Y. (2015), Platinum-Based Nanocages with Subnanometer-Thick Walls and Well-Defined Facets, *Science*, 349, 412–416 (DOI: 10.1126/science.aab0801)
8. Shu, Q., Zhang, J., Hu, B., Deng, X., Yuan, J., Ran, R., Zhou, W., Shao, Z. (2022), High-Durability Pt-Based ORR Catalyst on Mn/N Codoped Carbon Sheets for PEMFCs, *Energy Fuels*, 36, 1881–1891 (DOI: 10.1021/acs.energyfuels.1c04306)
9. Peterson, E. J., DeLaRiva, A. T., Lin, S., Johnson, R. S., Guo, H., Miller, J. T., Kwak, J. H., Peden, C. H. F., Kiefer, B., Allard, L. F., Ribeiro, F. H., Datye, A. K. (2014), Low-Temperature Carbon Monoxide Oxidation Catalyzed by Atomically Dispersed Palladium on Alumina, *Nat. Commun.*, 5, 4885 (DOI: 10.1038/ncomms5885)
10. Zhai, H., Alexandrova, A. N. (2017), Fluxionality of Catalytic Clusters: When It Matters and How to Address It, *ACS Catal.*, 7, 1905–1911 (DOI: 10.1021/acscatal.6b03243)
11. Zandkarimi, B., Alexandrova, A. N. (2020), Can Fluxionality of Subnanometer Cluster Catalysts Solely Cause Non-Arrhenius Behavior in Catalysis?, *J. Phys. Chem. C*, 124, 25008–25016 (DOI: 10.1021/acs.jpcc.0c04136)
12. Zandkarimi, B., Alexandrova, A. N. (2019), Dynamics of Subnanometer Pt Clusters Can Break the Scaling Relationships in Catalysis, *J. Phys. Chem. Lett.*, 10, 460–467 (DOI: 10.1021/acs.jpclett.8b03680)
13. Ha, M. A., Baxter, E. T., Cass, A. C., Anderson, S. L., Alexandrova, A. N. (2017), Boron Switch for Selectivity of Catalytic Dehydrogenation

on Size-Selected Pt Clusters on Al_2O_3, *J. Am. Chem. Soc.*, 139, 11568–11575 (DOI: 10.1021/jacs.7b05894)

14. Baxter, E. T., Ha, M. A., Cass, A. C., Alexandrova, A. N., Anderson, S. L. (2017), Ethylene Dehydrogenation on $Pt_{4,7,8}$ Clusters on Al_2O_3: Strong Cluster Size Dependence Linked to Preferred Catalyst Morphologies, *ACS Catal.*, 7, 3322–3335 (DOI: 10.1021/acscatal.7b00409)
15. Zhang, Z., Zandkarimi, B., Munarriz, J., Dickerson, C. E., Alexandrova, A. N. (2022), Fluxionality of Subnano Clusters Reshapes the Activity Volcano of Electrocatalysis, *ChemCatChem*, 14, e202200345 (DOI: 10.1002/cctc.202200345)
16. Heiz, U., Sanchez, A., Abbet, S., Schneider, W. D. (1999), Catalytic Oxidation of Carbon Monoxide on Monodispersed Platinum Clusters: Each Atom Counts, *J. Am. Chem. Soc.*, 121, 3214–3217 (DOI: 10.1021/ja983616l)
17. Ma, X., Li, Z., Achenie, L. E. K., Xin, H. (2015), Machine-Learning-Augmented Chemisorption Model for CO_2 Electroreduction Catalyst Screening, *J. Phys. Chem. Lett.*, 6, 3528–3533 (DOI: 10.1021/acs.jpclett.5b01660)
18. Roy, D., Mandal, S. C., Pathak, B. (2021), Machine Learning-Driven High-Throughput Screening of Alloy-Based Catalysts for Selective CO_2 Hydrogenation to Methanol, *ACS Appl. Mater. Interfaces*, 13, 56512–56523 (DOI: 10.1021/acsami.1c16696)
19. Mai, H., Le, T. C., Chen, D., Winkler, D. A., Caruso, R. A. (2022), Machine Learning for Electrocatalyst and Photocatalyst Design and Discovery, *Chem. Rev.*, 122, 13478–13534 (DOI: 10.1021/acs.chemrev.2c00061)
20. Ge, L., Yuan, H., Min, Y., Li, L., Chen, S., Xu, L., Goddard, W. A. (2020), Predicted Optimal Bifunctional Electrocatalysts Using Chalcogenide Heterostructures, *J. Phys. Chem. Lett.*, 11, 1941–1948 (DOI: 10.1021/acs.jpclett.9b03875)

21. Pandit, N. K., Roy, D., Mandal, S. C., Pathak, B. (2022), Rational Designing of Bimetallic and Trimetallic HER Catalysts Using Supervised Machine Learning, *J. Phys. Chem. Lett.*, 13, 7484–7492 (DOI: 10.1021/acs.jpclett.2c01401)
22. Niu, H., Wan, X., Wang, X., Shao, C., Robertson, J., Zhang, Z., Guo, Y. (2021), Single-Atom Rhodium on Defective g-C_3N_4 as a Bifunctional Oxygen Electrocatalyst, *ACS Sustain. Chem. Eng.*, 9, 2862–2871 (DOI: 10.1021/acssuschemeng.0c09192)
23. Panapitiya, G., Avendano-Franco, G., Ren, P., Wen, X., Li, Y., Lewis, J. P. (2018), Machine-Learning Prediction of CO Adsorption in Thiolated Ag-Alloyed Au Nanoclusters, *J. Am. Chem. Soc.*, 140, 17508–17514 (DOI: 10.1021/jacs.8b08800)
24. Hensen, E. J. M., Vlachos, D. G., Wang, Y., Su, Y. Q. (2020), Finite-Temperature Structures of Supported Subnanometer Catalysts via Statistical Learning, *ACS Nano*, 14, 14173–14183 (DOI: 10.1021/acsnano.0c06472)
25. Wang, Y., Su, Y. Q., Hensen, E. J. M., Vlachos, D. G. (2022), Insights into Supported Subnanometer Catalysts Exposed to CO, *Chem. Mater.*, 34, 1751–1762 (DOI: 10.1021/acs.chemmater.1c03616)
26. Liao, V., Cohen, M., Wang, Y., Vlachos, D. G. (2023), Deducing Subnanometer Cluster Size and Shape Distributions, *Nat. Commun.*, 14, 2031 (DOI: 10.1038/s41467-023-37664-w)
27. Kresse, G., Hafner, J. (1994), Ab Initio Molecular-Dynamics Simulation of Liquid–Metal–Amorphous-Semiconductor Transitions in Ge, *Phys. Rev. B*, 49, 14251–14269 (DOI: 10.1103/PhysRevB.49.14251)
28. Perdew, J. P., Chevary, J. A., Vosko, S. H., Jackson, K. A., Pederson, M. R., Singh, D. J., Fiolhais, C. (1992), GGA Applications for Exchange and Correlation, *Phys. Rev. B*, 46, 6671–6687 (DOI: 10.1103/PhysRevB.46.6671)

29. Joubert, D. (1999), From Ultrasoft Pseudopotentials to the PAW Method, *Phys. Rev. B*, 59, 1758–1775 (DOI: 10.1103/PhysRevB.59.1758)
30. Teter, M. P., Payne, M. C., Allan, D. C. (1989), Solution of Schrödinger's Equation for Large Systems, *Phys. Rev. B*, 40, 12255–12263 (DOI: 10.1103/PhysRevB.40.12255)
31. Sharma, R. K., Nair, A. S., Bharadwaj, N., Roy, D., Pathak, B. (2023), Role of Fluxionality and Metastable Isomers in ORR Activity, *J. Phys. Chem. C*, 127, 134–145 (DOI: 10.1021/acs.jpcc.2c06265)
32. Tang, W., Sanville, E., Henkelman, G. (2009), Grid-Based Bader Charge Analysis without Lattice Bias, *J. Phys.: Condens. Matter*, 21, 084204 (DOI: 10.1088/0953-8984/21/8/084204)
33. Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011), Scikit-Learn: Machine Learning in Python, *J. Mach. Learn. Res.*, 12, 2825–2830
34. Chaves, A. S., Piotrowski, M. J., Da Silva, J. L. F. (2017), Evolution of Properties of Transition-Metal Clusters, *Phys. Chem. Chem. Phys.*, 19, 15484–15502 (DOI: 10.1039/c7cp02240a)
35. Zhang, L., Guo, X., Zhang, S., Huang, S. (2022), Genome of Bi-Atom Catalysts for HER/OER/ORR, *J. Mater. Chem. A*, 10, 11518–11532 (DOI: 10.1039/d2ta02050e)
36. Bligaard, T., Nørskov, J. K., Dahl, S., Matthiesen, J., Christensen, C. H., Sehested, J. (2004), Brønsted–Evans–Polanyi Relations in Catalysis, *J. Catal.*, 224, 206–217 (DOI: 10.1016/j.jcat.2004.02.034)
37. Ferrin, P., Simonetti, D., Kandoi, S., et al. (2009), Ethanol Decomposition on Transition Metals, *J. Am. Chem. Soc.*, 131, 5809–5815 (DOI: 10.1021/ja8099322)
38. Nair, A. S., Pathak, B. (2019), Screening ORR Activity of Core–Shell Clusters, *J. Phys. Chem. C*, 123, 3013–3022 (DOI: 10.1021/acs.jpcc.8b11483)

39. Ooka, H., Huang, J., Exner, K. S. (2021), Sabatier Principle in Electrocatalysis, *Front. Energy Res.*, 9, 654460 (DOI: 10.3389/fenrg.2021.654460)
40. Nair, A. S., Mahata, A., Pathak, B. (2018), Multilayered Platinum Nanotubes for ORR, *ACS Appl. Energy Mater.*, 1, 3858–3866 (DOI: 10.1021/acsaem.8b00641)
41. Nair, A. S., Anoop, A., Ahuja, R., Pathak, B. (2022), Relativistic Effects in Platinum Nanocluster Catalysis, *J. Phys. Chem. A*, 126, 1485–1494 (DOI: 10.1021/acs.jpca.1c09981)
42. Nair, A. S., Anoop, A., Ahuja, R., Pathak, B. (2021), Role of Atomicity in ORR Activity of Pt Clusters, *J. Comput. Chem.*, 42, 2001–2010 (DOI: 10.1002/jcc.26725)
43. Sharma, R. K., Minhas, H., Pathak, B. (2024), Metastability-Triggered Reactivity of $Pt_{7,8}$ Clusters on Graphene, *J. Phys. Chem. C*, 128, 8245–8256 (DOI: 10.1021/acs.jpcc.4c00376)
44. Poths, P., Li, G., Masubuchi, T., et al. (2023), Self-Limiting Poisoning in Sub-Nano Cluster Catalysts, *ACS Catal.*, 13, 1625–1635 (DOI: 10.1021/acscatal.2c05634)
45. Calle-Vallejo, F., Martínez, J. I., García-Lastra, J. M., Sautet, P., Loffreda, D. (2014), Prediction of Adsorption Properties Using Coordination Numbers, *Angew. Chem. Int. Ed.*, 53, 8316–8319 (DOI: 10.1002/anie.201402958)
46. Hammer, B., Nørskov, J. K. (2000), Theoretical Surface Science and Catalysis, *Adv. Catal.*, 45, 71–129 (DOI: 10.1016/S0360-0564(02)45013-4)
47. Xin, H., Linic, S. (2010), Exceptions to the d-Band Model of Chemisorption, *J. Chem. Phys.*, 132, 221101 (DOI: 10.1063/1.3437609)
48. Vojvodic, A., Nørskov, J. K., Abild-Pedersen, F. (2014), Electronic Structure Effects in Surface Chemistry, *Top. Catal.*, 57, 25–32 (DOI: 10.1007/s11244-013-0159-2)

49. Jäger, M. O. J., Ranawat, Y. S., Canova, F. F., Morooka, E. V., Foster, A. S. (2020), ML-Aided Screening of Hydrogen Adsorption, *ACS Comb. Sci.*, 22, 680–688 (DOI: 10.1021/acscombsci.0c00102)
50. Ruban, A., Hammer, B., Stoltze, P., Skriver, H. L., Nørskov, J. K. (1997), Surface Electronic Structure and Reactivity of Metals, *J. Mol. Catal. A*, 115, 421–429 (DOI: 10.1016/S1381-1169(96)00348-2)
51. Li, B., Zhang, P., Liang, S., Ren, G. (2008), Feature Extraction for Gear Fault Diagnosis, *Proc. ICSP*, 234–238 (DOI: 10.1109/ICOSP.2008.4697740)
52. Wang, N. N., Dong, J., Deng, Y. H., et al. (2016), ADME Prediction Using Boosting, *J. Chem. Inf. Model*, 56, 763–773 (DOI: 10.1021/acs.jcim.5b00642)
53. Lundberg, S. M., Erion, G., Chen, H., et al. (2020), Explainable AI for Tree Models, *Nat. Mach. Intell.*, 2, 56–67 (DOI: 10.1038/s42256-019-0138-9)
54. Li, B., Gao, W., Jiang, Q. (2021), Electronic and Geometric Determinants of Adsorption, *JPhys Energy*, 3, 022003 (DOI: 10.1088/2515-7655/abd295)
55. Nørskov, J. K., Rossmeisl, J., Logadottir, A., et al. (2004), Origin of ORR Overpotential, *J. Phys. Chem. B*, 108, 17886–17892 (DOI: 10.1021/jp047349j)
56. Cui, X., Ren, P., Deng, D., Deng, J., Bao, X. (2016), Graphene-Encapsulated Non-Precious Metals for OER, *Energy Environ. Sci.*, 9, 123–129 (DOI: 10.1039/c5ee03316k)
57. Bai, L., Hsu, C. S., Alexander, D. T. L., Chen, H. M., Hu, X. (2021), Double-Atom Catalysts for OER, *Nat. Energy*, 6, 1054–1066 (DOI: 10.1038/s41560-021-00925-3)
58. Gao, R., Wang, J., Huang, Z. F., et al. (2021), Pt–Fe Pair Sites for ORR, *Nat. Energy*, 6, 614–623 (DOI: 10.1038/s41560-021-00826-5)

59. Liu, Z., Ma, Y., Gu, W., Yuan, C., Teng, F. (2020), MoxC–FeNi@NC for OER, *J. Phys. Chem. Solids*, 147, 109578 (DOI: 10.1016/j.jpcs.2020.109578)
60. Wang, Y. J., Zhao, N., Fang, B., Li, H., Bi, X. T., Wang, H. (2015), Carbon-Supported Pt Alloy Electrocatalysts for ORR, *Chem. Rev.*, 115, 3433–3467 (DOI: 10.1021/cr500519c)
61. Fu, Q., Draxl, C. (2019), Hybrid Perovskites as Substrates for Single-Atom Catalysts, *Phys. Rev. Lett.*, 122, 046101 (DOI: 10.1103/PhysRevLett.122.046101)
62. Fu, Z., Yang, B., Wu, R. (2020), Frontier Orbitals in Single-Atom Catalysis, *Phys. Rev. Lett.*, 125, 156001 (DOI: 10.1103/PhysRevLett.125.156001)
63. Jiang, B., Zhang, F., Wang, Y., et al. (2023), Dynamically Confined Single-Atom Sites for CO Oxidation, *Phys. Rev. B*, 107, 205421 (DOI: 10.1103/PhysRevB.107.205421)
64. Montemore, M. M., Medlin, J. W. (2014), Scaling Relations in Catalyst Design, *Catal. Sci. Technol.*, 4, 3748–3761 (DOI: 10.1039/c4cy00335g)
65. Man, I. C., Su, H. Y., Calle-Vallejo, F., et al. (2011), Universality in OER Electrocatalysis, *ChemCatChem*, 3, 1159–1165 (DOI: 10.1002/cctc.201000397)
66. Wang, Y., Huang, X., Fu, H., Shang, J. (2022), Activity Origin of HER on Single-Atom Catalysts, *J. Mater. Chem. A*, 10, 23840–23852 (DOI: 10.1039/d2ta07167c)
67. Zhang, J., Zhao, Z., Xia, Z., Dai, L. (2015), Metal-Free Bifunctional ORR/OER Catalyst, *Nat. Nanotechnol.*, 10, 444–452 (DOI: 10.1038/nnano.2015.48)
68. Wei, B., Fu, Z., Legut, D., et al. (2021), MXene-Based Double-Atom Catalysts, *Adv. Mater.*, 33, 2102595 (DOI: 10.1002/adma.202102595)

69. Prabu, M., Ketpang, K., Shanmugam, S. (2014), Hierarchical $NiCo_2O_4$ for Zn–Air Batteries, *Nanoscale*, 6, 3173–3181 (DOI: 10.1039/c3nr05835b)
70. Kulkarni, A., Siahrostami, S., Patel, A., Nørskov, J. K. (2018), Trends in ORR Activity, *Chem. Rev.*, 118, 2302–2312 (DOI: 10.1021/acs.chemrev.7b00488)
71. Zhang, T., Zhang, B., Peng, Q., Zhou, J., Sun, Z. (2021), Mo_2B_2MBene-Supported Single-Atom Catalysts, *J. Mater. Chem. A*, 9, 230–240 (DOI: 10.1039/d0ta08630d)
72. Palmer, D. S., O'Boyle, N. M., Glen, R. C., Mitchell, J. B. O. (2007), Random Forest Models for Solubility Prediction, *J. Chem. Inf. Model*, 47, 150–158 (DOI: 10.1021/ci060164k)
73. Saini, S., Halldin Stenlid, J., Deo, S., Plessow, P. N., Abild-Pedersen, F. (2024), Modeling Surface Site Stabilities on Multimetallic Catalysts, *ACS Catal.*, 14, 1223–1235 (DOI: 10.1021/acscatal.3c04337)
74. Saini, S., Halldin Stenlid, J., Abild-Pedersen, F. (2022), Electronic Structure Factors in Chemisorption, *npj Comput. Mater.*, 8, 156 (DOI: 10.1038/s41524-022-00846-z)
75. Gao, W., Chen, Y., Li, B., Liu, S. P., Liu, X., Jiang, Q. (2020), Determining Adsorption Energies from Intrinsic Properties, *Nat. Commun.*, 11, 14969 (DOI: 10.1038/s41467-020-14969-8)
76. Li, B., Li, X., Gao, W. (2022), Predictive Model of Surface Adsorption on Transition Metals and Alloys, *J. Mater. Chem. A*, 10, 6394–6404 (DOI: 10.1039/d1ta10795j)

CHAPTER 4

Ensemble Representation for Supported Clusters in Implicit Solvent

CHAPTER 4

4.1. Introduction

Proton exchange membrane fuel cells (PEMFCs) are emerging as a promising power source to meet increasing energy demands *[1-3]*. Their high energy density, low operating temperature, and minimal exhaust emissions have positioned them at the forefront of renewable energy research. The key electrochemical reactions in PEMFCs are the hydrogen evolution reaction (HER) at the anode and the oxygen reduction reaction (ORR) at the cathode. The sluggish ORR kinetics, however, remain the primary barrier to large-scale commercialization due to high overpotentials *[4]*. This challenge has motivated extensive efforts to design Pt-based electrocatalysts with reduced cathodic overpotential *[5,6]*. Toward this goal, Pt nanostructures with various dimensions and morphologies, including nanoclusters, nanofilms, nanoshells, and nanotubes, have been actively explored *[7-10]*. Yet, minimizing Pt loading while maintaining high catalytic activity continues to be a major challenge.

Imaoka et al. synthesized Pt_n subnanoclusters containing 10−20 atoms and reported a notable enhancement in ORR activity for Pt_{19}, attributed to its low-coordinated edge sites *[11,12]*. Our group also demonstrated that ORR activity in subnanoclusters is highly sensitive to atomicity and that such clusters can outperform the periodic Pt(111) surface *[13]*. Despite their promise, subnanoclusters are difficult to study due to their structural and electronic complexity. One distinctive feature of these size-specific supported nanoclusters is the presence of a competing low-energy metastable ensemble (LEME) that becomes relevant under high electrochemical potential, intermediate coverage, and elevated temperatures, resulting in fluxional behavior *[14,15]*.

Earlier work has primarily relied on first-principles global optimization methods such as basin-hopping (BH), genetic algorithms (GA), and particle swarm optimization (PSO) to identify the global minimum (GM) structure and explain catalytic behavior solely on its basis *[16-18]*. At finite temperatures, however, thermally accessible metastable isomers can

contribute significantly to catalysis due to their increased lifetime and population *[19]*. Alexandrova and co-workers highlighted this through a statistical ensemble representation of supported Pt_n $(n = 4, 7, 8)$ clusters on alumina for ethylene dehydrogenation *[20,21]*. They showed that Pt_7 exhibits higher activity than Pt_4 and Pt_8 because of its single-layered active local minima featuring increased binding sites and enhanced charge transfer from the support. More recently, our group also reported a substantial contribution from low-lying isomers of Pt_{13} toward overall ORR activity *[22]*. These findings emphasize the necessity of accounting for metastable isomers alongside the GM to accurately describe catalytic behavior.

Numerous theoretical studies have investigated nanoclusters and Pt(111) surfaces to understand the thermodynamics and kinetics of ORR and to assess the influence of gas and solvent environments *[23-27]*. For instance, Nair et al. used a hybrid solvation model to show substantial differences between gas- and solvent-phase ORR energetics on Pt(111) due to dispersion effects *[28]*. Sha et al. also reported major changes in adsorption sites, free-energy pathways, and kinetic barriers for ORR intermediates on Pt(111) when moving from gas-phase to solvent-phase simulations using an implicit solvation model *[29]*. More recently, Munarriz and Zhang et al. demonstrated potential-dependent electrochemical behavior of Ptn clusters on graphite using the surface charging (SC) method combined with the linearized Poisson–Boltzmann implicit solvation model *[30,31]*. This approach provides a promising route to evaluate solvent-mediated ensemble-averaged ORR activity of low-energy isomers.

Motivated by these studies, the present work investigates the solvent effect on the ensemble composition of size-selected Pt_n $(n = 7, 8)$ subnanoclusters supported on graphene (denoted Pt_7/G and Pt_8/G). Clusters in the 7–8 atom range are of particular interest due to their unusually high catalytic activity *[32,33]*. We combined global optimization and density functional theory (DFT) to explore the structural diversity of Pt_7/G and Pt_8/G isomers and assess their structural properties. To probe

the ORR mechanism, we analyzed adsorption-induced interactions between intermediates and clusters, followed by free-energy and activation barrier evaluations in both gas and solvent phases. The fluxional behavior of low-energy metastable isomers was quantified using a statistical ensemble representation based on normalized rate constants. An implicit solvation model was then used to determine the contribution of individual isomers in the solvent medium. The observed solvation trends of ORR intermediates were rationalized in terms of size, polarizability, and hydrogen-bonding capability. To highlight the role of the support and solvent, we compared the computed ORR activity with that of bare subnanoclusters and bulk Pt (111), revealing dynamic changes in catalytic behavior across different environments. After establishing the significant influence of metastable states under electrochemical conditions, we emphasize the necessity of an ensemble-based perspective in catalysis simulations. Lastly, *ab initio* thermodynamic analysis was performed to probe the possible formation of Pt oxides (Pt_nO_x) under ambient ORR conditions.

4.2. Computational Details

4.2.1. Electronic Structure Calculations

Plane-wave density functional theory calculations were carried out using the Vienna *ab initio* simulation Package (VASP) *[34]*. The generalized gradient approximation with the Perdew–Burke–Ernzerhof (GGA-PBE) functional was used *[35]*, and ion core–valence electron interactions were described using the projector augmented wave (PAW) method *[36]*. A kinetic energy cutoff of 470 eV and an electronic energy convergence criterion of 10^{-4} eV were applied for all relaxation calculations. Geometry optimizations were performed until the atomic forces were below 0.02 eV $Å^{-1}$ *[37]*. Gaussian smearing with a sigma value of 0.2 eV was used and van der Waals corrections were included via the DFT-D3 approach *[38]*. All molecular species and oxygen-bound intermediates were treated with spin-polarized calculations. Transition states were located using the climbing image nudged elastic band (CI-NEB) method with six images

generated along the reaction pathway. The identification of a single imaginary frequency confirmed each transition state (TS) *[39]*. The Gibbs free energies of ORR elementary steps were computed using the computational hydrogen electrode (CHE) model proposed by Nørskov and co-workers, according to the following equation *[40]*:

$$\Delta G = \Delta E + \Delta ZPE - T\Delta S - eU_{RHE} \quad (4.1)$$

where ΔE is the total reaction energy, ΔZPE is the zero-point energy correction, T is the temperature (300 K), ΔS is the entropy change, U_{RHE} is the applied electrode potential (vs reversible hydrogen electrode, RHE), and e is the number of transferred electrons. The ZPE and entropy values were obtained using the harmonic oscillator approximation according to the following expression:

$$ZPE = \sum_i \frac{1}{2} h\nu_i \quad (4.2)$$

$$S_v = R \sum \frac{\frac{\hbar\nu_i}{k_B T}}{\left(\exp\left(\frac{\hbar\nu_i}{kT}\right)-1\right)} - \ln\left(1 - \exp\left(\frac{\hbar\nu_i}{kT}\right)\right) \quad (4.3)$$

where h is the Planck constant, S_v is the vibrational entropy, R is the gas constant, k_B is the Boltzmann constant, ν_i is the frequency of the i^{th} vibrational mode, and $\hbar = \frac{h}{2\pi}$. The adsorption energies of intermediates were calculated using the following equation:

$$E_{ads} = E(Pt_n/G - X) - E(Pt_n/G) - E(X) \quad (4.4)$$

where $E(Pt_n/G - X)$ is the energy of the intermediate-adsorbed geometry on different isomers, $E(Pt_n/G)$ is the energy of the $Pt_{7/8}$ isomers adsorbed on the graphene sheet, and $E(X)$ is the energy of the ORR intermediates at the adsorbed geometry. The Γ-centered (1 × 1 × 1) k-point grid was used for Brillouin zone sampling, consistent with previous studies *[41,42]*. A denser (5 × 5 × 1) k-mesh was applied for density of states (DOS) calculations. Bader charge analysis was performed using the Henkelman code with the near-grid algorithm refine-edge method *[43,44]*. To account for solvent effects, we employed the VASPsol package, which incorporates a self-consistent continuum solvation model within VASP. Because a fully

explicit solvation model is computationally prohibitive, the linearized Poisson–Boltzmann implicit solvation model was adopted. The relative permittivity of the aqueous electrolyte was set to 78.4, and a Debye screening length of 3.0 Å was used, corresponding to an ionic strength of 1 M at 298 K.

4.2.2. Global Optimization

Initial geometries of Pt_7/G and Pt_8/G for global optimization were generated using the parallel global optimization pathway toolkit (PGOPT) *[45]*. PGOPT employs the bond-length distribution algorithm (BLDA) for efficient sampling of configurational space. A total of 800 starting structures (400 each for Pt_7/G and Pt_8/G) were produced, and only unique geometries were optimized. Duplicate structures were removed using a depth-first search (DFS) similarity check. During optimization, an 8 × 6 graphene supercell (a = 2.46 Å) with a 20.00 Å vacuum gap was used as support. The lowest-energy structure was identified as the global minimum (GM), and all other low-energy near-degenerate structures were labeled as local minima (LMs).

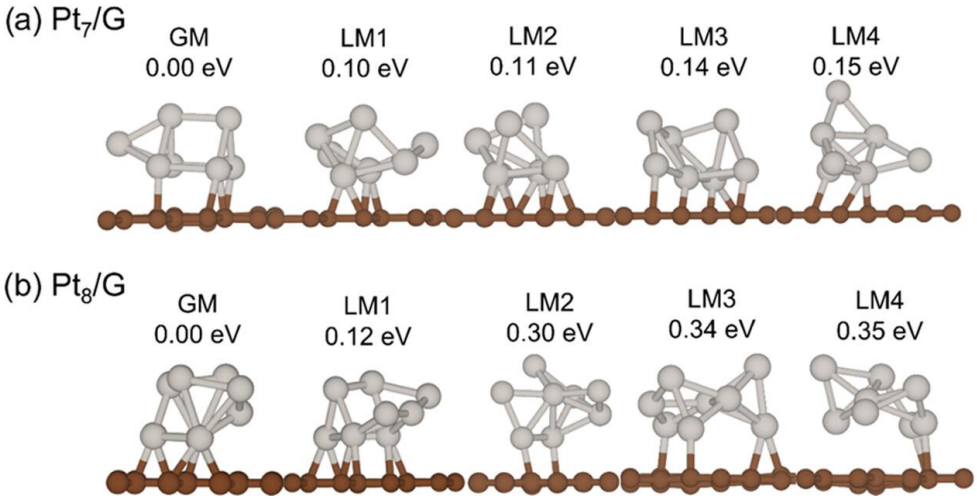


Figure 4.1. Thermally accessible structures within 0.4 eV of the global minimum (GM) for (a) Pt_7/G and (b) Pt_8/G. Each isomer is labeled with its relative energy difference from the GM (in eV), where zero denotes the GM.

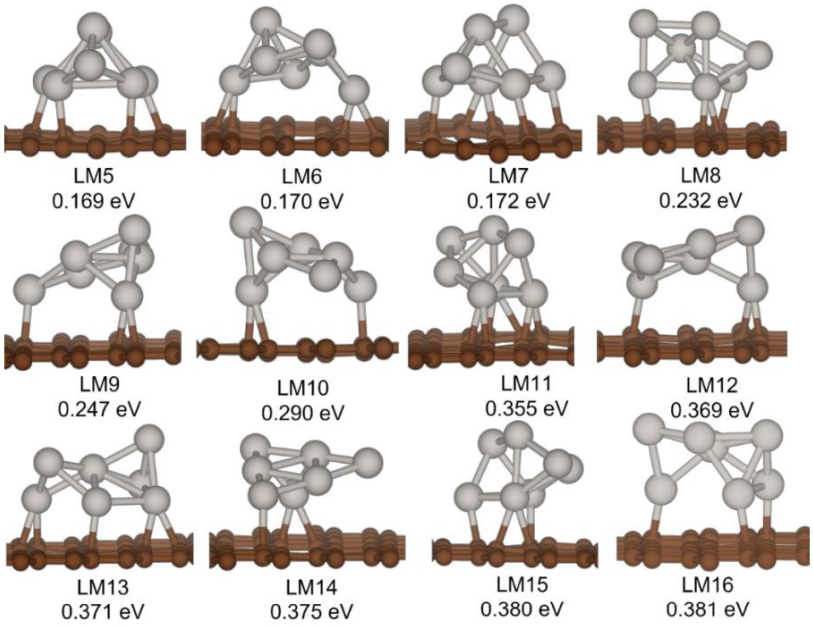


Figure 4.2. Thermally accessible isomers of Pt_7/G (LM5-LM16) with non-zero thermal population, within 0.4 eV energy difference from GM. Structural isomers are represented by their relative energy difference from GM (in eV).

4.3. Results and Discussion

4.3.1. Structural Analysis

4.3.1.1. Structural Explorations of Pt_7/G and Pt_8/G LEME.

LEME of Pt_7/G: Exploration of the low-energy metastable ensemble (LEME) for Pt_7/G identified 16 isomers within 0.4 eV of the global minimum (GM), all with non-zero thermal populations. On graphene, these isomers adopt three-dimensional nonplanar configurations, in contrast to the planar isomers of the bare Pt_7 cluster reported in gas-phase optimization *[41]*. Considering the narrow energy window relevant to ORR under ambient conditions, the ensemble was restricted to the eight most stable Pt_7/G isomers (within 0.17 eV of the GM) and the five most stable Pt_8/G isomers (within 0.4 eV), as shown in Figure 4.1a; remaining Pt_7/G isomers are provided in Figure 4.2. The metastable isomers display substantial structural diversity driven by variations in spatial arrangement and Pt coordination numbers ranging from 2 to 5 within a 3.0 Å cutoff. The first

metastable isomer lies 0.10 eV above the GM, and the number of accessible isomers increases toward the higher-energy limit. Pt_7/G structures adopt bi- or trilayered distributions and show net electron transfer from the cluster to the graphene support, producing positively charged clusters (Table 4.1), consistent with Pt_{55}/G and Pt_{147}/G *[46]*. Apart from LM1, which shows minimal charge transfer, all Pt_7/G isomers exhibit lower charge transfer than the GM of Pt_6/G (+0.19|e|). Charge transfer for Pt_7/G ranges from 0.00|e| to +0.13|e|, while LM1 shows +0.18|e| *[47]*, like a Pt adatom on graphene. The GM of Pt_7/G is in a singlet state, whereas most LMs stabilize in higher-spin configurations such as nonet and quintet *[48]*.

LEME of Pt_8/G: Adding one atom to Pt_7/G produces a more complex potential energy surface (PES) for Pt_8/G. Only five topologically distinct isomers fall within 0.4 eV of the GM (Figure 4.1b), indicating a comparatively rigid morphology due to enhanced metal-metal bonding, which restricts fluxionality. The first metastable state lies 0.12 eV above the GM in the quintet state. Like bare Pt_8 clusters *[32]*, Pt_8/G primarily adopts trilayered, three-dimensional arrangements with net electron transfer from the support to the cluster, producing negatively charged Pt clusters except for the GM and LM4 (Table 4.1). Charge transfer in LM1 (−0.30|e|), LM2 (−0.03|e|), and LM3 (−0.01|e|) shows an opposite trend to Pt_4, Pt_6, Pt_{13}, and Pt_{19} on graphite, where clusters gain positive charge (+0.11, +0.11, +0.15, and +0.19|e|) *[46-48]*. Structural inspection of LM1–LM3 shows that Pt atoms not directly bonded to graphene accumulate more negative charge than those in contact with the surface, indicating that charge-transfer direction depends on atomic arrangement. Pt coordination numbers for Pt_8/G range from 2 to 6 within a 3.0 Å cutoff, confirming the unsaturated coordination environment characteristic of subnanoclusters compared with bulk Pt(111). Like Pt_7/G, the density of accessible Pt_8/G isomers increase at higher energies, consistent with the PES behavior of Lennard-Jones clusters described via the disconnectivity graph by Wales et al *[49]*.

4.3.1.2. Stability Analysis of LEME.

To evaluate the structural and electrochemical stability of the isomers, we calculated their binding energy (BE) and dissolution potential U_{Diss}, summarized in Table 4.1 *[50,51]*. For a Pt cluster supported on graphene, the binding energy is defined as:

$$BE = \left(E_{Pt_n/G} - E_G - n \times E_{Pt,isolated}\right)/n \quad (4.5)$$

where $E_{Pt_n/G}$ is the total energy of the adsorbed system, E_G is the energy of the bare graphene support, n is the number of Pt atoms in the cluster, and $E_{Pt,isolated}$ is the energy of an isolated Pt atom. The dissolution potential U_{Diss}^{all} was estimated following the method introduced by Morimoto and co-workers, based on the cohesive and interaction energies of supported Pt particles *[51]*:

$$U_{diss}^{all} = U_{diss}^{bulk} + \frac{E_{coh} - E_{coh}^{bulk}}{2e} \quad (4.6)$$

where E_{coh} and E_{coh}^{bulk} are the cohesive energies per atom of the nanocluster and bulk Pt (111) surface, respectively. To include the effect of the support, the total cohesive energy combines the cohesive and interaction energies such as:

$$E_{coh} = \frac{nE_{tot}[Pt(g) + E_{tot}[C(g)] - E_{tot}[Pt_n/C(g)]}{n} \quad (4.7)$$

$$E_{int} = \frac{E_{tot}[Pt_n(g) + E_{tot}[C(g)] - E_{tot}[Pt_n/C(g)]}{n} \quad (4.8)$$

where n is the number of Pt atoms, $E_{tot}[Pt(g)]$ and $E_{tot}[C(g)]$ are the isolated Pt atom and graphene energies, $E_{tot}[Pt_n/C(g)]$ is the energy of the supported isomer, and $E_{tot}[Pt_n(g)]$ is the gas-phase cluster energy. Reference values of E_{coh}^{bulk} and U_{Diss}^{bulk} are 5.07 eV/atom and 1.19 V, respectively *[51]*.

Table 4.1. Properties of Low-Energy Metastable Ensembles (LEMEs) of Pt_7/G and Pt_8/G obtained from global optimization. Zero indicates the energy of the global minimum (GM), and the energies of the local minima (LM) are listed relative to the GM. $\Delta Q|e|$ denotes the net charge on the clusters, 2S + 1 is the spin multiplicity, and Min. CN and Max. CN are the

minimum and maximum coordination numbers, respectively. The average Pt–Pt bond length ($\overline{Pt-Pt}$) is reported in Å. BE refers to the binding energy (eV), and U_{Diss} corresponds to the dissolution potential (eV).

Prop.	Pt_7/G								Pt_8/G				
	GM	**LM1**	**LM2**	**LM3**	**LM4**	**LM5**	**LM6**	**LM7**	**GM**	**LM1**	**LM2**	**LM3**	**LM4**
ΔE(eV)	0.00	0.10	0.11	0.14	0.15	0.17	0.17	0.17	0.00	0.12	0.30	0.34	0.35
ΔQ|e|	+0.13	0.00	+0.18	+0.12	+0.11	+0.10	+0.05	+0.06	+0.08	-0.30	-0.03	-0.01	+0.03
2S+1	1	9	3	5	5	5	3	3	3	5	5	3	3
Min.CN	3	2	2	2	2	2	2	2	3	2	2	2	2
Max.CN	4	5	4	5	5	4	5	4	4	5	6	5	6
$\overline{Pt-Pt}$	2.59	2.63	2.63	2.64	2.62	2.62	2.61	2.62	2.68	2.60	2.58	2.55	2.56
BE	-3.62	-3.61	-3.60	-3.60	-3.59	-3.59	-3.59	-3.58	-3.75	-3.73	-3.71	-3.70	-3.70
U_{diss}^{all}	-1.66	-0.97	-1.11	-1.58	-1.12	-1.06	-1.12	-1.11	-1.16	-1.45	-1.56	-1.29	-1.43

From a thermodynamic perspective, the negative BE values confirm stable adsorption of all clusters on graphene, while the magnitude of BE decreases for higher-energy isomers, consistent with their relative ordering within the ensemble. The BE of Pt_7/G and Pt_8/G isomers is lower than that of Pt_{13} clusters on single and multiple N-doped graphene *[52]*, indicating that N doping increases the exothermicity of Pt–graphene binding. The U_{Diss}^{all} values indicate the tendency of Pt atoms to dissolve in acidic media. The negative values obtained for Pt_7/G and Pt_8/G reflect reduced electrochemical stability compared with larger clusters such as Pt_{55} *[51]*, which possess bulk-like facets. This reduced stability arises from the limited stabilization due to low coordination numbers and the lack of strong interaction between the cluster edges and the graphene substrate. Under electrochemical conditions, LM1 of Pt_8/G and GM of Pt_8/G exhibit the highest stability among their respective ensembles. Notably, the negative U_{Diss}^{all} values do not imply spontaneous fragmentation, since the high BE creates a significant energetic barrier against cluster breakup.

4.3.2. ORR Activity Analysis in the Gas Phase

4.3.2.1. Adsorption Behavior of ORR Intermediates in Gas Phase

Owing to the unique coordination environments and electronic states of heterogeneous sites, identifying the most stable adsorption configurations in the subnano cluster regime is challenging. To understand ORR activity, we analyzed the adsorption-induced interactions of key intermediates (O_2^*, O^*, OH^*, OOH^*, and H_2O^*) across multiple active sites of the stable Pt_7/G and Pt_8/G isomers obtained from global optimization. We have sampled multiple top, bridge and hollow sites considering the heterogeneity of the site on the nanoclusters and the most stable configurations are considered for further investigation. Compared to Pt (111) surfaces and Pt_{79} nanoclusters, adsorption on these subnano clusters exhibits distinct behavior *[53]*, as discussed below.

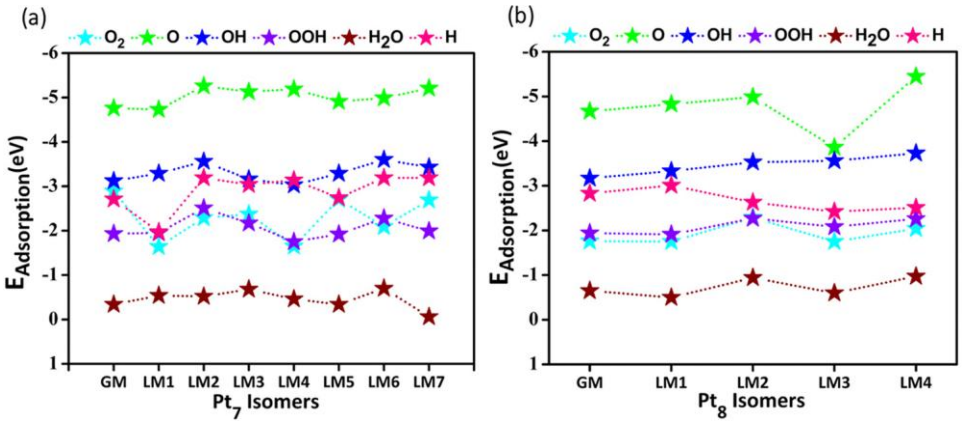


Figure 4.3. Variable trend of adsorption energy of ORR intermediates on global minima (GM) and local minima (LM) of (a) Pt_7/G and (b) Pt_8/G, corresponding to their most stable adsorbate-binding configuration. The adsorption values corresponding to each point in the plot are tabulated in Table 4.2.

Table 4.2. Adsorption energies of ORR intermediate (in eV) correspond to their most stable intermediate-adsorbate configuration in gas-phase. Here t, b, and f represent top, bridge, and face-centered cubic (fcc) sites.

Systems	Isomer	O_2^*	O^*	OH^*	OOH^*	H_2O^*	H^*
Pt_7/G	GM	-2.89 (b)	-4.76 (b)	-3.12 (t)	-1.93 (t)	-0.34 (t)	-2.71 (t)
	LM1	-1.64 (b)	-4.73 (t)	-3.29 (b)	-1.95 (t)	-0.54 (t)	-1.97 (t)
	LM2	-2.29 (b)	-5.26 (b)	-3.56 (t)	-2.51 (t)	-0.52 (t)	-3.19 (t)
	LM3	-2.37 (b)	-5.13 (t)	-3.16 (t)	-2.17 (b)	-0.68 (b)	-3.04 (t)
	LM4	-1.65 (b)	-5.19 (b)	-3.03 (b)	-1.75 (t)	-0.46 (t)	-3.14 (t)
	LM5	-2.71 (b)	-4.91 (b)	-3.29 (b)	-1.92 (t)	-0.34 (t)	-2.74 (b)
	LM6	-2.09 (b)	-4.99 (b)	-3.60 (t)	-2.28 (t)	-0.70 (b)	-3.19 (t)
	LM7	-2.69 (b)	-5.21 (t)	-3.43 (t)	-1.99 (t)	-0.06 (b)	-3.19 (t)
Pt_8/G	GM	-1.76 (b)	-4.67 (b)	-3.17 (b)	-1.94 (b)	-0.65 (t)	-2.83 (t)
	LM1	-1.75 (b)	-4.83 (t)	-3.33 (t)	-1.91 (t)	-0.50 (t)	-3.01 (t)
	LM2	-2.29 (b)	-4.99 (t)	-3.53 (t)	-2.27 (t)	-0.94 (b)	-2.63 (t)
	LM3	-1.75 (b)	-3.86 (b)	-3.56 (b)	-2.08 (b)	-0.60 (b)	-2.42 (t)
	LM4	-2.04 (b)	-5.45 (t)	-3.73 (t)	-2.26 (t)	-0.97 (t)	-2.51 (t)
Bulk Systems	Pt(111)	-0.79 (b)	-4.54 (t)	-2.40 (t)	-1.22 (b)	-0.26 (b)	-2.80 (f)
	Pt_{79}	-1.56 (b)	-5.19 (t)	-2.49 (t)	-1.40 (b)	-0.23 (b)	-2.80 (f)
Bare Nanoclusters	Pt_7	-3.56 (b)	-5.04 (t)	-3.81 (t)	-2.48 (t)	-1.12 (t)	--

O_2^* preferentially binds at bridge sites in the disigma mode, with both oxygen atoms coordinated to Pt atoms (Table 4.2). The GM of Pt_7/G and LM2 of Pt_8/G exhibited the strongest O_2^* adsorption energies among the LMs (Figure 4.3) *[53]*. However, most isomers displayed weaker O_2^* adsorption (~ -2.13 eV) compared to Pt_{13} clusters on defective graphene *[23]*. Among ORR intermediates, O^* shows the strongest adsorption, favoring both top and bridge sites. LM2 of Pt_7/G and LM4 of Pt_8/G demonstrated the highest O^* adsorption energies at bridge and top sites, respectively. Unlike bulk Pt(111), where O^* binds primarily to fcc sites, subnanoclusters favor top and bridge sites. Adsorption of O^* on Pt_7/G and Pt_8/G is more exothermic than on bare or single-vacancy graphene-supported Pd_{13} and MPd_{12} (M = Fe, Co, Ni, Cu, Zn) clusters due to strong binding at low-coordination sites *[54]*. The H^* intermediate prefers top sites on Pt_7/G and Pt_8/G, in contrast to fcc site preference on Pt(111) and Pt_{79} clusters *[53]*. Similarly, OH^*, OOH^*, and H_2O^* intermediates show stronger adsorption on Pt_7/G and Pt_8/G compared to Pt(111) surface. Compared to

bare Pt_7 clusters, most Pt_7/G and Pt_8/G isomers exhibit slightly weaker adsorption, likely due to reduced active site availability *[41]*, yet stronger than on faceted Pt_{79} clusters *[53]*. This enhancement is attributed to the coordinatively unsaturated environment of subnanoclusters.

Table 4.3. d-band center values correspond to different adsorbates on different isomers of Pt_7/G and Pt_8/G clusters.

Prop.	Pt_7/G								Pt_8/G				
	GM	**LM1**	**LM2**	**LM3**	**LM4**	**LM5**	**LM6**	**LM7**	**GM**	**LM1**	**LM2**	**LM3**	**LM4**
O_2^*	-3.56	-3.34	-3.65	-3.56	-3.36	-3.31	-3.29	-3.44	-3.38	-3.43	-3.39	-3.43	-3.55
O^*	-3.65	-3.29	-3.66	-3.44	-3.55	-3.59	-3.50	-3.57	-3.50	-3.42	-3.35	-3.37	-3.49
OH^*	-3.59	-3.37	-3.46	-3.55	-3.53	-3.44	-3.49	-3.49	-3.57	-3.42	-3.42	-3.55	-3.42
OOH^*	-3.64	-3.37	-3.49	-3.52	-3.39	-3.43	-3.36	-3.92	-3.39	-3.39	-3.45	-3.53	-3.42
H_2O^*	-3.67	-3.46	-3.62	-3.56	-3.49	-3.64	-3.44	-3.65	-3.61	-3.60	-3.54	-3.42	-3.61
H^*	-3.51	-3.78	-3.86	-3.94	-3.40	-3.61	-3.53	-3.56	-3.42	-3.33	-3.24	-3.20	-3.28

Notably, no intermediates were stable at hollow sites or directly on Pt atoms bound to the graphene support. To rationalize this behavior, we performed a d-band center (dbc) analysis, a weighted average of the metal d-states during adsorption (Table 4.3) *[55]*. An upshift (downshift) relative to the Fermi level indicates stronger (weaker) adsorption. The dbc trends for Pt_7/G isomers generally oppose those of Pt_8/G, except for LM4 in both cases. For H^* adsorption, the dbc of Pt_7/G isomers consistently exceed that of Pt_8/G. The largest shifts were observed at LM4 of Pt_8/G (−3.28 eV) and LM3 of Pt_7/G (−3.94 eV). Overall, the irregular dbc trends and closely spaced values highlight the structural sensitivity of subnanoclusters toward ORR activity.

4.3.2.2. Free-Energy Profile Diagram in the Gas Phase.

To understand the ORR mechanism, we calculated the reaction free energies and applied the Bell−Evans−Polanyi (BEP) principle *[56]*, which relates more stable adsorption energies to lower activation barriers. Using the most stable adsorption configurations of the intermediates, we analyzed

thermodynamic activity of these clusters at 0 V and 1.23 V, through both the associative and dissociative pathways. The elementary steps for both mechanisms are summarized below:

Associative Pathway:

$$O_2\,(g) + (H^+ + e^-) + * \rightarrow OOH^* \quad \text{(R1)}$$

$$OOH^* + (H^+ + e^-) \rightarrow O^* + H_2O \quad \text{(R2)}$$

$$O^* + (H^+ + e^-) \rightarrow OH^* \quad \text{(R3)}$$

$$OH^* + (H^+ + e^-) \rightarrow H_2O + * \quad \text{(R4)}$$

Dissociative Pathway

$$O_2\,(g) + (H^+ + e^-) + * \rightarrow O^* + O^* + (H^+ + e^-) \quad \text{(R5)}$$

$$O^* + O^* + (H^+ + e^-) \rightarrow O^* + OH^* \quad \text{(R6)}$$

$$O^* + OH^* + (H^+ + e^-) \rightarrow O^* + H_2O \quad \text{(R7)}$$

$$O^* + (H^+ + e^-) \rightarrow OH^* \quad \text{(R8)}$$

$$OH^* + (H^+ + e^-) \rightarrow H_2O + * \quad \text{(R9)}$$

In the associative pathway, the hydrogenation of O_2^* to OOH^* (ΔG_{R2}) represents the most endergonic step at 1.23 V, making it the thermodynamic rate-determining step (RDS) for all isomers except LM7 of Pt_7/G (Table 4.4). In the dissociative pathway, the hydrogenation of OOH^* to O^* with the first H_2O formation (ΔG_{R9}) is the RDS for most isomers, except LM2, LM5, and LM6 of Pt_7/G and LM4 of Pt_8/G, where hydrogenation of $O^* \rightarrow OH^*$, O^*-$O^* \rightarrow O^*{-}OH^*$, $OH^* \rightarrow H_2O$, respectively, becomes the thermodynamic RDS. Comparison of ΔG_{R2} and ΔG_{R7} indicates that the dissociative pathway is preferred for all isomers, supported by the larger negative ΔG values for initial O_2^* dissociation and smaller overpotentials (η) for the dissociative RDS (except LM6 of Pt_7/G). This agrees with Lim et al., who reported that O_2^* dissociation on $Pt_7/defected\_graphene$ occurs at a lower energy barrier than OOH^* formation *[23]*. During optimizations, H_2O_2 intermediates readily dissociated into 2 OH^* species that adsorbed on neighboring sites, facilitated by adjacent active sites capable of breaking O–O bonds. This contrasts with single-atom catalysts, where isolated top sites favor a $2e^-$

pathway producing H_2O_2. Subnanoclusters, instead, promote complete $4e^-$ reduction to H_2O *[57,58]*.

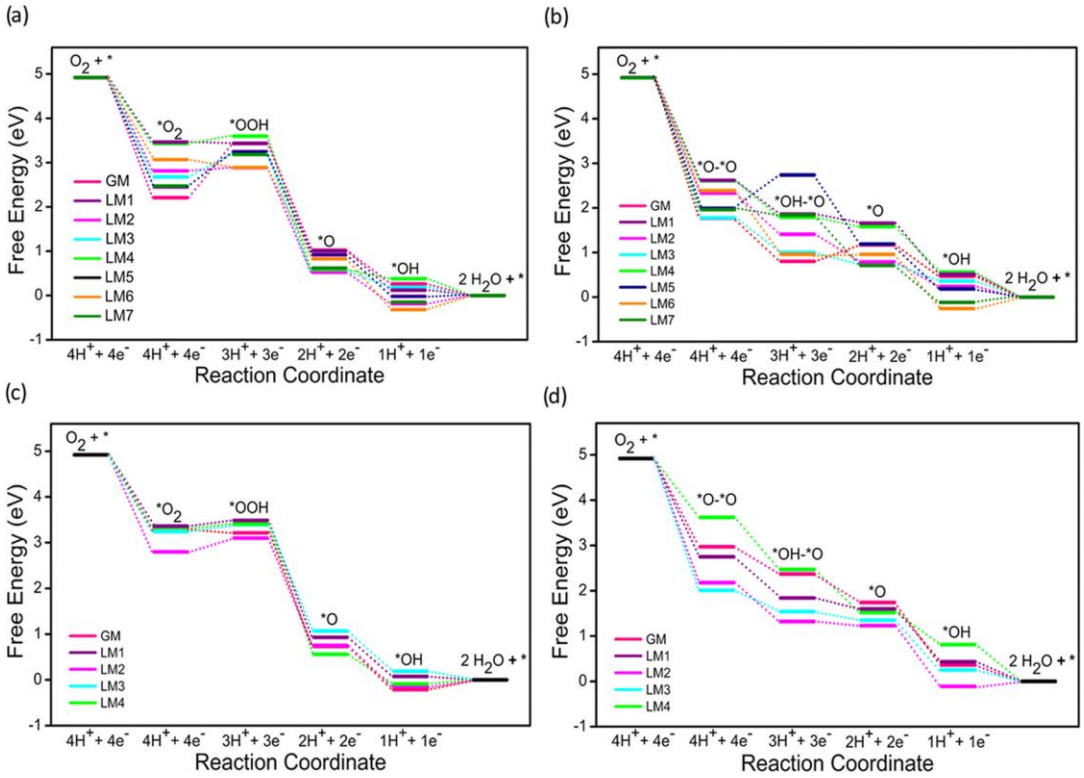


Figure 4.4. Free-energy diagrams of ORR elementary steps for isomers of (a, b) Pt_7/G and (c, d) Pt_8/G at 0 V applied potential via the associative and dissociative pathway, respectively. Here, sign * represents surface-adsorbed species.

Table 4.4. Thermodynamic Rate-Determining Step (RDS) in the Gas Phase for Different Isomers of Pt_7/G and Pt_8/G in the Associative and Dissociative Pathways of ORR. The overpotential values (η) and activation energy barrier (E_a) for each potential thermodynamic RDS are listed.

Substrate		Associative		Dissociative		E_a (eV)
	Isomers	RDS	η (V)	RDS	η (V)	
Pt_7/G	GM	$O_2^* \rightarrow OOH^*$	2.45	$OOH^* \rightarrow O^*$	1.60	0.50
	LM1	$O_2^* \rightarrow OOH^*$	1.21	$OOH^* \rightarrow O^*$	1.05	1.46
	LM2	$O_2^* \rightarrow OOH^*$	1.30	$O^* \rightarrow OH^*$	0.67	0.27
	LM3	$O_2^* \rightarrow OOH^*$	1.76	$OOH^* \rightarrow O^*$	0.95	0.51
	LM4	$O_2^* \rightarrow OOH^*$	1.39	$OOH^* \rightarrow O^*$	1.02	0.93
	LM5	$O_2^* \rightarrow OOH^*$	2.03	$O^*\text{-}O^* \rightarrow O^*\text{-}OH^*$	1.99	0.48
	LM6	$O_2^* \rightarrow OOH^*$	1.06	$OH^* \rightarrow H_2O$	1.49	0.65
	LM7	$OOH^* \rightarrow O^*$	1.94	$OOH^* \rightarrow O^*$	1.35	0.37
Pt_8/G	GM	$O_2^* \rightarrow OOH^*$	1.38	$OOH^* \rightarrow O^*$	1.04	1.10
	LM1	$O_2^* \rightarrow OOH^*$	1.36	$OOH^* \rightarrow O^*$	0.99	0.49
	LM2	$O_2^* \rightarrow OOH^*$	1.53	$OOH^* \rightarrow O^*$	1.14	0.73
	LM3	$O_2^* \rightarrow OOH^*$	1.19	$OOH^* \rightarrow O^*$	0.60	0.70
	LM4	$O_2^* \rightarrow OOH^*$	1.31	$O^* \rightarrow OH^*$	0.52	0.07

Among the dissociative pathway isomers, LM2 of Pt_7/G (η = 0.67 V) and LM4 of Pt_8/G (η = 0.52 V) were the most thermodynamically active, surpassing bulk Pt(111) surfaces (η = 0.78 V) and Pt_{79} nanoclusters (η = 1.00 V) *[59]*. This highlights the enhanced activity of metastable isomers and their potential for reducing Pt loading without compromising ORR performance. Kinetic evaluation of the thermodynamic RDS, based on selective activation barriers (E_a) at 1.23 V using the Langmuir−Hinshelwood mechanism, showed LM2 of Pt_7/G and LM4 of Pt_8/G as the most kinetically favorable isomers (E_a = 0.27 and 0.07 eV, respectively), outperforming GM and other LMs (Table 4.4). Thus, both thermodynamic and kinetic analyses indicate that metastable isomers can surpass GM activity, motivating the quantification of each isomer's contribution to the overall reaction, as described in the following section.

4.3.2.3. Weighted-Average Contribution through Statistical Ensemble Representation of ORR in the Gas Phase

Recent studies by Alexandrova and co-workers highlighted that accurately assessing catalytic activity of subnanoclusters requires considering the cumulative contributions from all individual isomers within the statistical ensemble of the low-energy metastable ensemble (LEME). This approach reflects realistic experimental conditions, where elevated temperatures increase the population of low-energy isomers and allow structural transformations of the GM into LMs. To quantify each isomer's contribution, a normalized rate constant ($k_{weighted}$) was calculated following the method of Sun and Sautet et al. *[60]*:

$$k_{Weighted} = A \exp\left(\frac{-E_a^{LM_i}}{k_B T}\right) \times \exp\left(-\frac{E^{LM_i} - E^{GM}}{k_B T}\right) \quad (4.9)$$

where A is the Arrhenius prefactor (taken as $10^{13}\ s^{-1}$), E_a^{LMi} is the activation barrier of the isomer, E_{LM_i} and E_{GM} are the energies of the isomer and GM, respectively, and T is 300 K. The $k_{weighted}$ factor incorporates both the stability and activity of each isomer. In Figure 4.5, the most negative (or highest positive) $k_{weighted}$ values indicate the isomers contributing most significantly to the overall ORR activity, with the order of contributions as follows:

Pt_7/G : LM2 > GM > LM7 > LM5 > LM3 > LM6 > LM4 > LM1

Pt_8/G : LM4 > LM1 > LM2 > GM > LM3

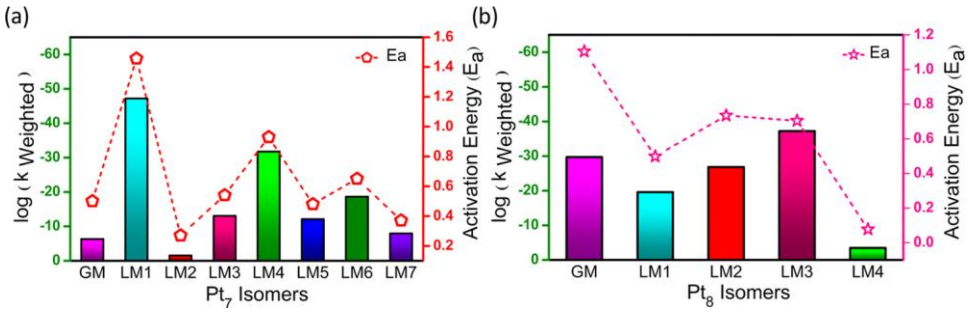


Figure 4.5. Statistical ensemble-averaged ORR activity of (a) Pt_7/G and (b) Pt_8/G isomers. The lowest negative (or the highest positive) value of $k_{weighted}$ represents the maximum contribution toward ORR activity.

The bar graphs represent the logarithm values of the normalized weighted contribution ($k_{weighted}$) of different isomers toward overall ORR activity, and the star marks correspond to the activation energy barrier (E_a) of thermodynamic RDS of ORR for each isomer within our constructed ensemble.

Several metastable isomers exhibited significantly higher activity contributions than the stable GM. For example, LM2 of Pt_7/G showed a 4.12-fold higher activity than GM, while LM4 of Pt_8/G, the highest-energy isomer in the ensemble, displayed an 8.63-fold greater activity. This variation arises from differences in the activation energy (E_a) of the thermodynamic RDS, with isomers having lower E_a contributing more actively without being kinetically trapped. Notably, the activity contributions of bare Pt_7 cluster differed from those of graphene-supported Pt_7/G clusters. For bare Pt_7, LM1 and LM4 exhibited the highest catalytic activity, whereas on graphene, these same isomers showed the lowest contributions. Analysis of the Bader charge distribution revealed that isomers with larger charge separation contributed more to overall activity, except for LM7 of Pt_7/G. This observation aligns with previous experimental and theoretical studies on Pt_4Sn_3/SiO_2 and Pd_6/G, where isomers within 0.4 eV of GM showing substantial charge redistribution during intermediate adsorption had higher activity. However, a similar trend was not observed for Pt_8/G isomers. These findings highlight the importance of considering contributions from LEME rather than relying solely on the GM for accurately describing overall catalytic activity, offering a new perspective for subnano cluster catalyst design.

4.3.2.4. Solvent-Mediated Metastability-Triggered Reactivity from LEME

To explore the behavior of metastable isomers under electrochemical conditions, we investigated the role of solvents in their ORR activity using an implicit solvation model, which captures polarization, electrostatic, and dispersion interactions between solute and solvent within a dielectric

continuum *[63]*. The study included adsorption behavior, reaction free energies of ORR intermediates, and statistical ensemble analysis of different isomers in the solvent medium. For comparison, we also evaluated the periodic Pt(111) surface to distinguish subnano cluster behavior from bulk systems in solution. We observed that the solvent generally stabilizes intermediate-cluster configurations, as indicated by negative solvation energies (Table 4.6), except for H^* on GM and LM6 of Pt_7/G. The extent of stabilization varies among intermediate and isomers. The maximum stabilization occurred for H_2O^* adsorption on LM2 of Pt_7/G ($E_{solvation}$ = −0.59 eV), while the minimum was observed for H^* adsorption on LM2 of Pt_8/G ($E_{solvation}$ = −0.01 eV).

Table 4.6. Solvation Energy ($E_{solvation}$) for Different ORR Intermediates for Different Isomers of Pt_7/G and Pt_8/G Systems in eV. $E_{solvation}$ is expressed as $E_{solvation} = E_{solvent} - E_{gas}$, where $E_{solvent}$ and E_{gas} are the adsorption energies in the solvent and gas-phase medium, respectively. Negative values of $E_{solvation}$ indicate the extent of stabilization of the ORR intermediates in the solvent medium compared to the gas phase.

	Pt_7/G								Pt_8/G				
Prop.	**GM**	**LM1**	**LM2**	**LM3**	**LM4**	**LM5**	**LM6**	**LM7**	**GM**	**LM1**	**LM2**	**LM3**	**LM4**
O_2^*	-0.10	-0.14	-0.16	-0.16	-0.11	-0.40	-0.04	-0.01	-0.20	-0.17	-0.21	-0.08	-0.14
O^*	-0.12	-0.12	-0.14	-0.26	-0.10	-0.13	-0.09	-0.10	-0.14	-0.24	-0.19	-0.20	-0.11
OH^*	-0.23	-0.18	-0.20	-0.20	-0.23	-0.06	-0.18	-0.40	-0.14	-0.19	-0.27	-0.21	-0.20
OOH^*	-0.27	-0.15	-0.32	-0.23	-0.14	-0.19	-0.17	-0.41	-0.35	-0.27	-0.38	-0.28	-0.27
H_2O^*	-0.51	-0.44	-0.59	-0.56	-0.21	-0.60	-0.35	-0.56	-0.63	-0.20	-0.55	-0.43	-0.18
H^*	0.03	-0.06	-0.06	-0.08	-0.03	-0.01	0.01	-0.03	-0.03	-0.04	-0.01	-0.10	-0.03

Table 4.7. Adsorption energies corresponding to the most stable intermediate-adsorbate configurations in a solvent medium. Here t, b, and f represent top, bridge, and face-centered cubic (fcc) sites.

Systems	Isomer	O_2^*	O^*	OH^*	OOH^*	H_2O^*	H^*
Pt_7/G	GM	-2.99 (b)	-4.88 (b)	-3.35 (t)	-2.20 (t)	-0.85 (t)	-2.67 (t)
	LM1	-1.80 (b)	-4.99 (t)	-3.49 (b)	-2.18 (t)	-1.10 (t)	-2.03 (t)
	LM2	-2.43 (b)	-5.38 (b)	-3.74 (t)	-2.66 (t)	-0.96 (t)	-3.26 (t)
	LM3	-2.37 (b)	-5.13 (t)	-3.16 (t)	-2.17 (b)	-0.68 (b)	-3.04 (t)
	LM4	-1.65 (b)	-5.19 (b)	-3.03 (b)	-1.75 (t)	-0.46 (t)	-3.14 (t)
	LM5	-2.71 (b)	-4.91 (b)	-3.29 (b)	-1.92 (t)	-0.34 (t)	-2.74 (b)
	LM6	-2.09 (b)	-4.99 (b)	-3.60 (t)	-2.28 (t)	-0.70 (b)	-3.19 (t)
	LM7	-2.75 (b)	-5.21 (t)	-3.43 (t)	-1.99 (t)	-0.06 (b)	-3.19 (t)
Pt_8/G	GM	-1.96 (b)	-4.81 (b)	-3.31 (b)	-2.29 (b)	-1.18 (t)	-2.86 (t)
	LM1	-1.83 (b)	-5.03 (t)	-3.54 (t)	-2.19 (t)	-0.93 (t)	-3.05 (t)
	LM2	-2.46 (b)	-5.23 (t)	-3.72 (t)	-2.54 (t)	-1.14 (b)	-2.46 (t)
	LM3	-1.89 (b)	-3.97 (b)	-3.76 (b)	-2.35 (b)	-0.78 (b)	-2.53 (t)
	LM4	-2.25 (b)	-5.64 (t)	-3.73 (t)	-2.27 (t)	-1.52 (t)	-2.54 (t)
Bulk Systems	Pt(111)	-0.89 (f)	-4.38 (f)	-2.78 (t)	-1.53 (b)	-0.58 (b)	-2.86 (f)

For most isomers of Pt_7/G and Pt_8/G, the solvation energies of ORR intermediates follow the trend: $H^* < O_2^* \sim O^* < OH^* < OOH^* < H_2O^*$. This can be rationalized by their size, polarity, and hydrogen-bonding ability. H_2O^* exhibits the largest negative solvation energy due to its high polarity and strong hydrogen-bonding interactions with the solvent. OH^* and OOH^* are also polar and capable of hydrogen bonding, resulting in significant stabilization. In contrast, O^* is smaller and less polarized, placing its solvation energy between H and the polar intermediates, while H shows the smallest solvation energy due to its minimal size and polarity. Overall, larger, polarizable intermediates with hydrogen-bonding capability are strongly stabilized, whereas smaller, less polar species experience weaker solvation effects. Compared to the Pt(111) surface in solvent, adsorption of intermediates is generally more favorable on subnanoclusters for most

isomers (Table 4.7) *[29]*. Similarly, adsorption energies on solvent-phase Pt_7/G and Pt_8/G isomers are higher than on bare Pt_7 clusters in the gas phase *[41]*. This solvent-mediated enhancement is expected to significantly influence the ORR mechanism. Accordingly, we calculated the ΔG values at 0 and 1.23 V for both associative and dissociative pathways in the solvent medium and reported the overall activity.

Table 4.8. Thermodynamic Rate-Determining Step (RDS) in the Solvent Medium for Different Isomers of Pt_7/G and Pt_7/G in Associative and Dissociative Pathways of ORR. The overpotential values (η) and activation energy barrier (E_a) for each potential RDS are listed in the table.

Substrate		Associative		Dissociative		E_a (eV)
	Isomers	RDS	η (V)	RDS	η (V)	
Pt_7/G	GM	$O_2^* \rightarrow OOH^*$	1.93	$OOH^* \rightarrow O^*$	1.77	0.47
	LM1	$O_2^* \rightarrow OOH^*$	1.14	$OOH^* \rightarrow O^*$	1.13	1.40
	LM2	$O_2^* \rightarrow OOH^*$	1.29	$OOH^* \rightarrow O^*$	0.83	0.35
	LM3	$O_2^* \rightarrow OOH^*$	1.93	$OOH^* \rightarrow O^*$	1.13	0.63
	LM4	$O^* \rightarrow OH^*$	0.93	$OOH^* \rightarrow O^*$	0.70	0.99
	LM5	$OOH^* \rightarrow O^*$	1.99	$O^*\text{-}O^* \rightarrow O^*\text{-}OH^*$	1.98	0.40
	LM6	$O^* \rightarrow OH^*$	0.99	$OH^* \rightarrow H_2O$	1.65	0.65
	LM7	$OOH^* \rightarrow O^*$	1.99	$OH^* \rightarrow H_2O$	1.35	0.27
Pt_8/G	GM	$O_2^* \rightarrow OOH^*$	1.23	$OOH^* \rightarrow O^*$	1.08	1.21
	LM1	$O_2^* \rightarrow OOH^*$	1.16	$OOH^* \rightarrow O^*$	1.03	0.55
	LM2	$O_2^* \rightarrow OOH^*$	1.42	$OOH^* \rightarrow O^*$	1.25	0.76
	LM3	$O_2^* \rightarrow OOH^*$	1.23	$OOH^* \rightarrow O^*$	0.62	0.67
	LM4	$O_2^* \rightarrow OOH^*$	1.50	$O^* \rightarrow OH^*$	0.50	0.12

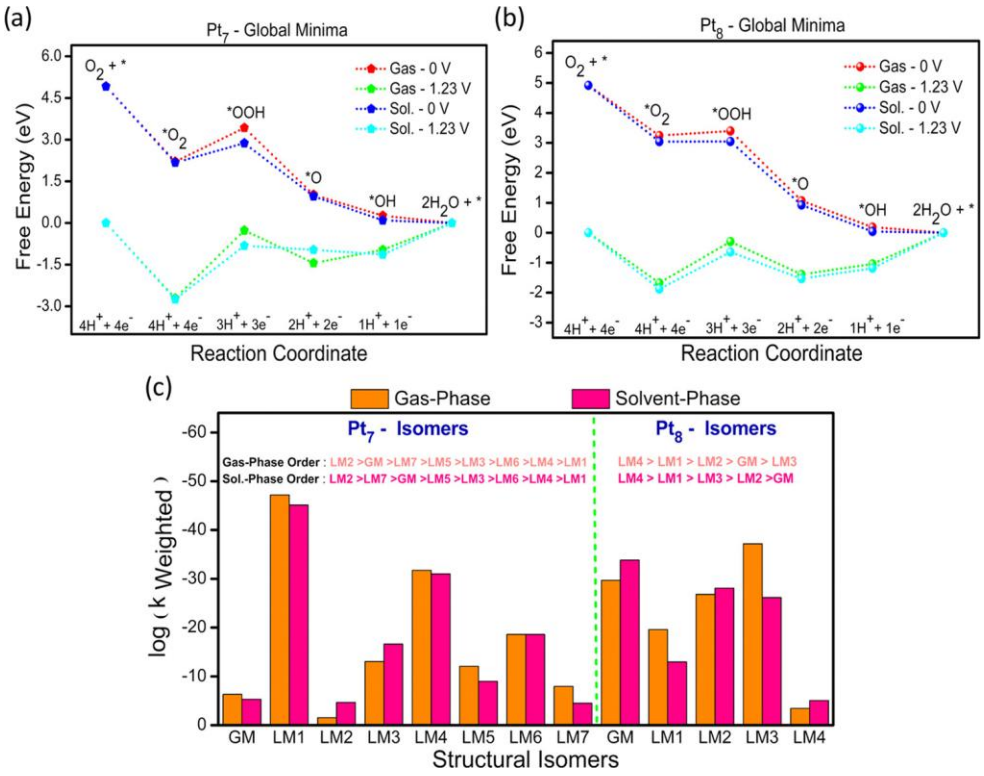


Figure 4.6. (a) Free-energy change (ΔG) in the gas and solvent phases for an associative pathway at 0 and 1.23 V for GM of Pt_7/G. (b) Free-energy change (ΔG) in the gas and solvent phases for an associative pathway at 0 and 1.23 V for GM of Pt_8/G. (c) Comparison between the statistical ensemble-averaged activity of Pt_7/G and Pt_8/G isomers in the gas and solvent phases.

The ΔG values of each reaction step in the solvent medium differed from the gas phase. For GM of Pt_7/G, the hydrogenation of O_2^* to OOH^* at 1.23 V becomes exothermic in the solvent (Figure 4.6a), while the subsequent hydrogenation of OOH^* to O^* and H_2O^* turns endothermic compared to the gas phase. For GM of Pt_8/G, all steps are energetically downhill in the solvent phase (Figure 4.6b). In the solvent, the associative pathway has O_2^* → OOH^* as the thermodynamic RDS, while the dissociative pathway has OOH^* → O^* as the RDS, except for LM4 of Pt_7/G and Pt_8/G, where O^* → OH^* is the RDS in the associative and dissociative pathways, respectively (**Table 4.8**). LM4 of Pt_7/G and Pt_8/G emerge as the most thermodynamically active isomers in solution, with η = 0.70 V and 0.12 V,

respectively. In general, η values increase for the dissociative pathway in the solvent relative to the gas phase due to stronger intermediate adsorption and stabilization via solvent interactions, consistent with previous reports on dispersion and relativistic corrections improving DFT predictions for Pt-based ORR *[64]*. Compared to bulk Pt, solvent-phase subnanoclusters (LM2, LM4 of Pt_7/G; LM3, LM4 of Pt_8/G) show a significant decrease in η *[7,59]*, highlighting the importance of including solvent effects to accurately predict catalytic activity.

Kinetic analysis of the thermodynamic RDS in solution using the Langmuir–Hinshelwood mechanism indicates that LM2 of Pt_7/G and LM1 of Pt_8/G are the most kinetically active isomers, with E_a = 0.35 eV and 0.12 eV, respectively. The variations in thermodynamics and kinetics in solvent compared to gas phase affect the contribution of each isomer in the ensemble. To account for this, the normalized, Boltzmann-weighted rate constant ($k_{weighted}$) was recalculated in the solvent medium, and the order of contribution to overall ORR activity is as follows:

Pt_7/G : LM2 > LM7 > GM > LM5 > LM3 > LM6 > LM4 > LM1

Pt_8/G : LM4 > LM1 > LM3 > LM2 > GM

Interestingly, the active contributions of GM, LM2, and LM3 of Pt_8/G changed in the solvent medium (Figure 4.6c). The most stable GM showed the lowest contribution, whereas the least stable LM4 exhibited the highest contribution to overall ORR activity. Compared to GM, LM1 and LM4 of Pt_8/G showed 2.6-fold and 6.7-fold higher contributions, respectively. For Pt_7/G, the overall trend of isomer contributions remained similar, though the magnitudes varied due to differences in E_a between the gas and solvent phases. Compared to bare Pt_7 clusters, the supported and solvated LMs of Pt_7/G and Pt_8/G displayed significantly higher contributions relative to GM *[41]*, highlighting the pronounced fluxionality and enhanced activity of metastable isomers under realistic conditions. While the exact contributions may depend on experimental factors such as pH and

temperature, these results demonstrate that cluster restructuring combined with solvent effects provides an accurate description of catalytic activity.

4.3.2.5. *Ab initio* Thermodynamic Analysis.

Under experimental conditions for the ORR, the surfaces of Pt nanoclusters tend to develop a high intermediate coverage. Previous investigations have also demonstrated the oxidation tendencies of Pt clusters under ambient ORR conditions *[65]*. This motivated us to explore the energetics of low-energy isomers of Pt_7/G and Pt_8/G under a high coverage of intermediates. Since the *O intermediate represents the strongest binding affinity with each isomer, we conducted an ab initio thermodynamic analysis to examine the stable oxidized phase of Pt_7O_x (x = 1−7) and Pt_8O_x (x = 1−8) for the five most stable isomers. For this purpose, multiple configurations of Pt_7O_x and Pt_8O_x were generated, corresponding to each coverage by adsorbing the *O intermediate at various positions on initially optimized Pt_7/G and Pt_8/G isomers. Subsequently, each configuration was optimized, and the most stable configuration was selected for further analysis.

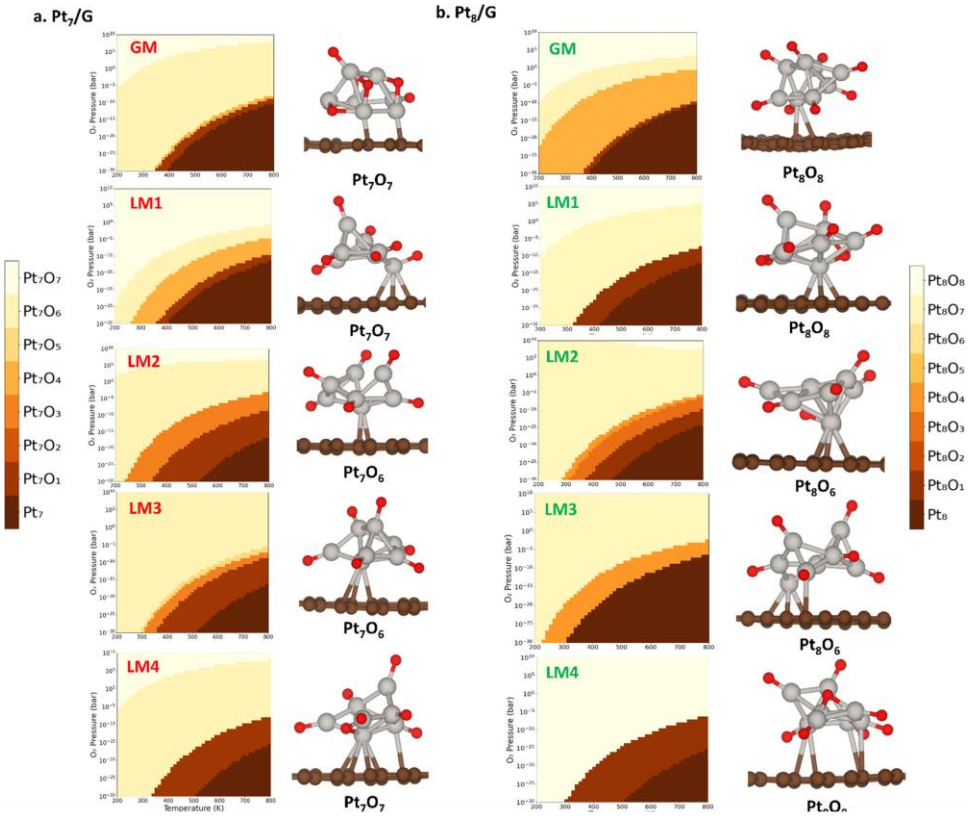
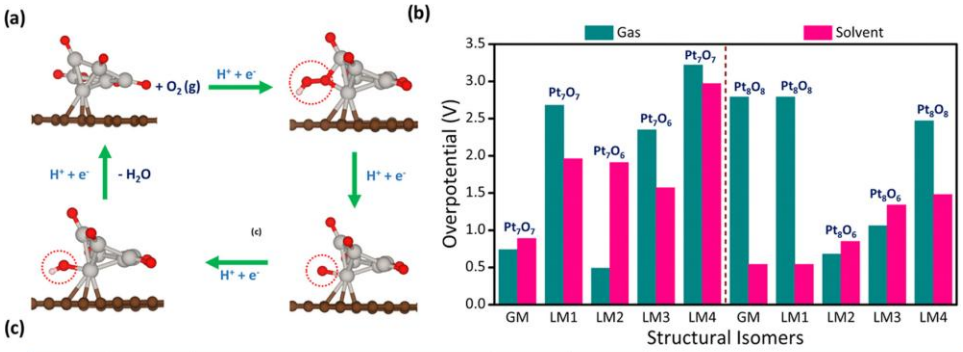


Figure 4.7. *Ab initio* thermodynamic analysis of the *O coverage of low-energy isomers. Second-order phase diagram of (a) Pt_7O_x (x = 1–7) and (b) Pt_8O_x (x = 1–8) stable isomers supported on graphene. For each isomer, the stable configuration of O^* coverage under ambient ORR conditions is also represented.

According to this approach, the chemical potential of gas phase oxygen was treated as a function of temperature and pressure $\mu_o(T, P)$ and is expressed as:

$$\mu_o(T, P) = \frac{1}{2}\left[E_{O_2}\,(T = 0\ K, P^o)\ \mu_{O_2}(T, P^o) + k_B T \ln \frac{P}{P^o}\right] \qquad (4.10)$$

where T is the temperature, P is the oxygen partial pressure, P^o is the standard atmospheric pressure, and k_B is the Boltzmann constant. To construct phase diagrams, the Python Multiscale Thermochemistry Toolbox (pMuTT) was employed, which utilizes the vibrational degrees of freedom of Pt and O atoms at different T and P *[66]*. The phase diagram and the most stable configuration of multiple adsorbates under the ORR conditions for each isomer are represented in Figure 4.7. All the isomers undergo significant structural transformations at high O^* coverage, indicating their fluxionality under reaction conditions. In contrast to our previous investigations, where the distribution of O^* atoms was more stabilized on bridging positions *[22]*, the O^* atoms occupy various top and bridging positions on Pt_7/G and Pt_8/G isomers at high O^* coverage. Interestingly, under ambient ORR conditions (T = 300 K and P_{O_2} = 1 bar), the GM, LM1, and LM4 of Pt_7/G favored the formation of Pt_7O_7 coverage, whereas LM2 and LM3 favored Pt_7O_6 coverage. Similarly, for Pt_8/G, the GM, LM1, and LM4 favored Pt_8O_8 coverage, while LM2 and LM3 favored Pt_8O_6 coverage. In line with the bare Pt_7 cluster, graphene supported Pt_7 and Pt_8 clusters also favor high O^* coverage. Therefore, it can be concluded that the Pt_7/G and Pt_8/G isomers exhibit a greater oxidation tendency under ORR conditions, favoring the formation up to 1 monolayer of Pt oxides.



(c)

Isomers	Stable Oxidized Phase	Pt_7O_x/Graphene				Isomers	Stable Oxidized Phase	Pt_8O_x/Graphene			
		Gas-Phase		Solvent-Phase				Gas-Phase		Solvent-Phase	
		RDS	η (V)	RDS	η (V)			RDS	η (V)	RDS	η (V)
GM	Pt_7O_7	$O_2(g) - OOH^*$	0.74	$OH^* - H_2O$ (l)	0.89	GM	Pt_8O_8	$OOH^* - O^*$	1.09	$OOH^* - O^*$	0.86
LM1	Pt_7O_7	$O_2(g) - OOH^*$	2.68	$O_2(g) - OOH^*$	1.96	LM1	Pt_8O_8	$O_2(g) - OOH^*$	2.79	$O^* - OH^*$	0.36
LM2	Pt_7O_6	$OOH^* - O^*$	0.49	$O_2(g) - OOH^*$	1.91	LM2	Pt_8O_6	$OOH^* - O^*$	0.68	$O_2(g) - O^*$	0.85
LM3	Pt_7O_6	$O_2(g) - OOH^*$	2.35	$OH^* - H_2O$(l)	1.57	LM3	Pt_8O_6	$O^* - OH^*$	1.02	$O_2(g) - OOH^*$	1.39
LM4	Pt_7O_7	$OH^* - H_2O$(l)	3.26	$O_2(g) - OOH^*$	2.97	LM4	Pt_8O_8	$O_2(g) - OOH^*$	2.47	$O_2(g) - OOH^*$	1.48

Figure 4.8. (a) Schematic representation of the ORR mechanism on the stable phase of oxidized Pt clusters; (b) overpotential values (η) for the most stable oxidized isomers (Pt_7O_7, Pt_7O_6, Pt_8O_8, and Pt_8O_6) in the gas and solvent phases; and (c) the thermodynamic rate-determining step (RDS) of ORR along with their overpotential values for Pt_7O_x and Pt_8O_x isomers in the gas and solvent phases.

Subsequently, we extend our investigation to quantify the ORR activity of the stable oxidized phase of Pt isomers in both the gas and solvent phases following the recent study conducted by Ding et al *[67]*. The ΔG calculations were carried out based on the scheme represented in Figure 4.8a, and the elementary steps followed to derive the ORR mechanism are provided in Text S5, Supporting Information. As shown in Figure 4.8b, the irregular trends observed in ΔG and η values reflect the distinct behavior of oxidized Pt_7O_x and Pt_8O_x isomers compared to bare Pt_7/G and Pt_8/G isomers. In the gas phase, LM2 with a stable configuration of Pt_7O_6 and Pt_8O_6 emerged as the thermodynamically active isomer. However, in the solvent phase, GM with a Pt_7O_7 configuration and LM1 with a Pt_8O_8 configuration became the thermodynamically most active isomers, exhibiting the lowest η values. Additionally, in the solvent phase, the stable phase isomers of Pt_7O_x consistently exhibited higher η values compared to

the stable phase isomers of Pt_8O_x (Figure 4.8c). This observation emphasized the change in the scaling relationship under a high coverage of intermediates during ambient ORR conditions, further distinguishing the activity at low and high intermediate coverage.

4.4. CONCLUSIONS

In summary, our study on graphene-supported Pt_7 and Pt_8 subnanoclusters highlights the critical role of solvent-mediated fluxionality and statistical ensemble representation in the oxygen reduction reaction (ORR). Structural exploration revealed numerous metastable isomers within 0.4 eV of the global minimum, reflecting the complex potential energy surface of supported clusters compared to bare clusters. Mechanistic analysis shows a clear preference for the dissociative pathway over the associative pathway, with significantly reduced overpotentials for the thermodynamic rate-determining step (RDS) at the subnano cluster scale compared to bulk Pt(111) surfaces and Pt_{79} clusters. Compared to bare Pt_7 clusters, the root-mean-square displacement (RMSD) of adsorbed ORR intermediates on Pt_7 and Pt_8 supported clusters underscore their fluxional behavior. Statistical ensemble analysis further demonstrates that isomers with lower activation energy barriers and greater cluster-to-support charge transfer contribute most strongly to overall activity. Inclusion of solvent effects leads to notable changes in adsorption energies, thermodynamics, and kinetics, altering the contribution of individual isomers under reaction conditions. Specifically, the overpotential for the dissociative pathway increases in solvent, reflecting stabilization of intermediates, which aligns with experimental observations and complements known dispersion and relativistic effects. Overall, high-energy metastable isomers of supported and solvated clusters play a significant role in ORR due to their unique ensemble and reduced energy barriers in aqueous environments. While these trends depend on cluster size and reaction conditions, solvent-mediated metastability provides a reliable framework to model complex, fluxional electrochemical interfaces accurately.

4.5. References

1. Debe, M. K. (2012), Electrocatalyst Approaches and Challenges for Automotive Fuel Cells, *Nature*, 486, 43–51 (DOI: 10.1038/nature11115)
2. Lubitz, W., Tumas, W. (2007), Hydrogen: An Overview, *Chem. Rev.*, 107, 3900–3903 (DOI: 10.1021/cr050200z)
3. Watanabe, M., Tryk, D. A., Wakisaka, M., Yano, H., Uchida, H. (2012), Overview of Recent Developments in Oxygen Reduction Electrocatalysis, *Electrochim. Acta*, 84, 187–201 (DOI: 10.1016/j.electacta.2012.04.035)
4. Shao, M., Chang, Q., Dodelet, J. P., Chenitz, R. (2016), Recent Advances in Electrocatalysts for Oxygen Reduction Reaction, *Chem. Rev.*, 116, 3594–3657 (DOI: 10.1021/acs.chemrev.5b00462)
5. Viswanathan, V., Hansen, H. A., Rossmeisl, J., Nørskov, J. K. (2012), Universality in Oxygen Reduction Electrocatalysis on Metal Surfaces, *ACS Catal.*, 2, 1654–1660 (DOI: 10.1021/cs300227s)
6. Bing, Y., Liu, H., Zhang, L., Ghosh, D., Zhang, J. (2010), Nanostructured Pt-Alloy Electrocatalysts for PEM Fuel Cell Oxygen Reduction Reaction, *Chem. Soc. Rev.*, 39, 2184–2202 (DOI: 10.1039/b912552c)
7. Kulkarni, A., Siahrostami, S., Patel, A., Nørskov, J. K. (2018), Understanding Catalytic Activity Trends in the Oxygen Reduction Reaction, *Chem. Rev.*, 118, 2302–2312 (DOI: 10.1021/acs.chemrev.7b00488)
8. Du, M., Cui, L., Cao, Y., Bard, A. J. (2015), Mechanoelectrochemical Catalysis of the Effect of Elastic Strain on a Platinum Nanofilm for the ORR Exerted by a Shape Memory Alloy Substrate, *J. Am. Chem. Soc.*, 137, 7397–7403 (DOI: 10.1021/jacs.5b03034)
9. Oezaslan, M., Hasché, F., Strasser, P. (2013), Pt-Based Core-Shell Catalyst Architectures for Oxygen Fuel Cell Electrodes, *J. Phys. Chem. Lett.*, 4, 3273–3291 (DOI: 10.1021/jz4014135)

10. Nair, A. S., Mahata, A., Pathak, B. (2018), Multilayered Platinum Nanotube for Oxygen Reduction in a Fuel Cell Cathode: Origin of Activity and Product Selectivity, *ACS Appl. Energy Mater.*, 1, 3451–3460 (DOI: 10.1021/acsaem.8b00641)
11. Imaoka, T., Kitazawa, H., Chun, W. J., Yamamoto, K. (2015), Finding the Most Catalytically Active Platinum Clusters with Low Atomicity, *Angew. Chem. Int. Ed.*, 54, 9810–9815 (DOI: 10.1002/anie.201504473)
12. Imaoka, T., Kitazawa, H., Chun, W. J., Omura, S., Albrecht, K., Yamamoto, K. (2013), Magic Number Pt13 and Misshapen Pt12 Clusters: Which One Is the Better Catalyst?, *J. Am. Chem. Soc.*, 135, 13089–13095 (DOI: 10.1021/ja405922m)
13. Nair, A. S., Anoop, A., Ahuja, R., Pathak, B. (2021), Role of Atomicity in the Oxygen Reduction Reaction Activity of Platinum Sub Nanometer Clusters: A Global Optimization Study, *J. Comput. Chem.*, 42, 1937–1947 (DOI: 10.1021/acs.jctc.1c01157)
14. Jimenez-Izal, E., Alexandrova, A. N. (2018), Computational Design of Clusters for Catalysis, *Annu. Rev. Phys. Chem.*, 69, 377–400 (DOI: 10.1146/annurev-physchem-050317-014216)
15. Zhai, H., Alexandrova, A. N. (2017), Fluxionality of Catalytic Clusters: When It Matters and How to Address It, *ACS Catal.*, 7, 1905–1911 (DOI: 10.1021/acscatal.6b03243)
16. Wales, D. J., Doye, J. P. K. (1997), Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms, *J. Phys. Chem. A*, 101, 5111–5116 (DOI: 10.1021/jp970984n)
17. Deaven, D. M., Ho, K. M. (1995), Molecular Geometry Optimization with a Genetic Algorithm, *Phys. Rev. Lett.*, 75, 288–291 (DOI: 10.1103/PhysRevLett.75.288)

18. Avendaño-Franco, G., Romero, A. H. (2016), Firefly Algorithm for Structural Search, *J. Chem. Theory Comput.*, 12, 3416–3428 (DOI: 10.1021/acs.jctc.5b01157)
19. Jimenez-Izal, E., Zhai, H., Liu, J. Y., Alexandrova, A. N. (2018), Nanoalloying MgO-Deposited Pt Clusters with Si to Control the Selectivity of Alkane Dehydrogenation, *ACS Catal.*, 8, 8346–8356 (DOI: 10.1021/acscatal.8b02443)
20. Baxter, E. T., Ha, M. A., Cass, A. C., Alexandrova, A. N., Anderson, S. L. (2017), Ethylene Dehydrogenation on $Pt_{4,7,8}$ Clusters on Al_2O_3: Strong Cluster Size Dependence Linked to Preferred Catalyst Morphologies, *ACS Catal.*, 7, 3322–3335 (DOI: 10.1021/acscatal.7b00409)
21. Ha, M. A., Baxter, E. T., Cass, A. C., Anderson, S. L., Alexandrova, A. N. (2017), Boron Switch for Selectivity of Catalytic Dehydrogenation on Size-Selected Pt Clusters on Al_2O_3, *J. Am. Chem. Soc.*, 139, 11568–11575 (DOI: 10.1021/jacs.7b05894)
22. Sharma, R. K., Nair, A. S., Bharadwaj, N., Roy, D., Pathak, B. (2023), Role of Fluxionality and Metastable Isomers in the ORR Activity: A Case Study, *J. Phys. Chem. C*, 127, 1–12 (DOI: 10.1021/acs.jpcc.2c06265)
23. Lim, D. H., Wilcox, J. (2012), Mechanisms of the Oxygen Reduction Reaction on Defective Graphene-Supported Pt Nanoparticles from First-Principles, *J. Phys. Chem. C*, 116, 3653–3660 (DOI: 10.1021/jp210796e)
24. Xu, D., Liu, Y. J., Zhao, J. X., Cai, Q. H., Wang, X. Z. (2014), Theoretical Study of the Deposition of Pt Clusters on Defective Hexagonal Boron Nitride Sheets, *J. Phys. Chem. C*, 118, 17, 9290–9300 (DOI: 10.1021/jp4087943)
25. Hyman, M. P., Medlin, J. W. (2006), Mechanistic Study of the Electrochemical Oxygen Reduction Reaction on Pt(111), *J. Phys. Chem. B*, 110, 31, 15515–15517 (DOI: 10.1021/jp061813y)

26. Jacob, T., Goddard, W. A. (2006), Water Formation on Pt and Pt-Based Alloys: A Theoretical Description, *ChemPhysChem*, 7, 1927–1934 (DOI: 10.1002/cphc.200500613)
27. Janik, M. J., Taylor, C. D., Neurock, M. (2009), First-Principles Analysis of the Initial Electroreduction Steps of Oxygen over Pt(111), *J. Electrochem. Soc.*, 156, B126–B133 (DOI: 10.1149/1.3008005)
28. Nair, A. S., Pathak, B. (2022), Accounting for Dispersion Effects in the DFT Framework of Electrocatalysis: A Hybrid Solvation Model-Based Case Study of the ORR, *J. Phys. Chem. C*, 126, 14, 5893–5902 (DOI: 10.1021/acs.jpcc.1c09207)
29. Sha, Y., Yu, T. H., Liu, Y., Merinov, B. V., Goddard, W. A. (2010), Theoretical Study of Solvent Effects on Platinum-Catalyzed ORR, *J. Phys. Chem. Lett.*, 1, 5, 856–861 (DOI: 10.1021/jz9003153)
30. Munarriz, J., Zhang, Z., Sautet, P., Alexandrova, A. N. (2022), Graphite-Supported Ptn Cluster Electrocatalysts: Major Change of Active Sites as a Function of the Applied Potential, *ACS Catal.*, 12, 23, 14399–14413 (DOI: 10.1021/acscatal.2c04643)
31. Zhang, Z., Zandkarimi, B., Munarriz, J., Dickerson, C. E., Alexandrova, A. N. (2022), Fluxionality of Subnano Clusters Reshapes the Activity Volcano of Electrocatalysis, *ChemCatChem*, 14, 15, e202200345 (DOI: 10.1021/acs.jclett.8b03680)
32. Vajda, S., Pellin, M. J., Greeley, J. P., Marshall, C. L., Curtiss, L. A., Ballentine, G. A., Elam, J. W., Catillon-Mucherie, S., Redfern, P. C., Mehmood, F., Zapol, P. (2009), Subnanometre Platinum Clusters as Highly Active and Selective Catalysts for Oxidative Dehydrogenation of Propane, *Nat. Mater.*, 8, 3, 213–216 (DOI: 10.1038/nmat2384)
33. Poths, P., Hong, Z., Li, G., Anderson, S. L., Alexandrova, A. N. (2022), "Magic" Sinter-Resistant Cluster Sizes of Ptn Supported on Alumina, *J. Phys. Chem. Lett.*, 13, 47, 10723–10730 (DOI: 10.1021/acs.jpclett.2c03114)

34. Kresse, G., Hafner, J. (1994), Ab Initio Molecular-Dynamics Simulation of the Liquid–Metal–Amorphous Semiconductor Transition in Germanium, *Phys. Rev. B*, 49, 20, 14251–14269 (DOI: 10.1103/PhysRevB.49.14251)
35. Perdew, J. P., Chevary, J. A., Vosko, S. H., Jackson, K. A., Pederson, M. R., Singh, D. J., Fiolhais, C. (1992), Applications of the GGA for Exchange and Correlation, *Phys. Rev. B*, 46, 11, 6671–6687 (DOI: 10.1103/PhysRevB.46.6671)
36. Joubert, D. (1999), From Ultrasoft Pseudopotentials to the Projector Augmented-Wave Method, *Phys. Rev. B*, 59, 3, 1758–1775 (DOI: 10.1103/PhysRevB.59.1758)
37. Teter, M. P., Payne, M. C., Allan, D. C. (1989), Solution of Schrödinger's Equation for Large Systems, *Phys. Rev. B*, 40, 18, 12255–12263 (DOI: 10.1103/PhysRevB.40.12255)
38. Grimme, S., Antony, J., Ehrlich, S., Krieg, H. (2010), A Consistent and Accurate Ab Initio Parametrization of DFT-D for the 94 Elements H–Pu, *J. Chem. Phys.*, 132, 15, 154104 (DOI: 10.1063/1.3382344)
39. Henkelman, G., Uberuaga, B. P., Jónsson, H. (2000), Climbing Image Nudged Elastic Band Method for Finding Saddle Points and Minimum Energy Paths, *J. Chem. Phys.*, 113, 22, 9901–9904 (DOI: 10.1063/1.1329672)
40. Nørskov, J. K., Rossmeisl, J., Logadottir, A., Lindqvist, L., Kitchin, J. R., Bligaard, T., Jónsson, H. (2004), Origin of the Overpotential for Oxygen Reduction at a Fuel-Cell Cathode, *J. Phys. Chem. B*, 108, 46, 17886–17892 (DOI: 10.1021/jp047349j)
41. Nair, A. S., Anoop, A., Ahuja, R., Pathak, B. (2022), Relativistic Effects in Platinum Nanocluster Catalysis: A Statistical Ensemble-Based Analysis, *J. Phys. Chem. A*, 126, 8, 1485–1494 (DOI: 10.1021/acs.jpca.1c09981)
42. Zandkarimi, B., Alexandrova, A. N. (2019), Dynamics of Subnanometer Pt Clusters Can Break the Scaling Relationships in

Catalysis, *J. Phys. Chem. Lett.*, 10, 3, 460–467 (DOI: 10.1021/acs.jpclett.8b03680)

43. Tang, W., Sanville, E., Henkelman, G. (2009), A Grid-Based Bader Analysis Algorithm without Lattice Bias, *J. Phys.: Condens. Matter*, 21, 8, 084204 (DOI: 10.1088/0953-8984/21/8/084204)
44. Sanville, E., Kenny, S. D., Smith, R., Henkelman, G. (2007), Improved Grid-Based Algorithm for Bader Charge Allocation, *J. Comput. Chem.*, 28, 5, 899–908 (DOI: 10.1002/jcc.20575)
45. Zhai, H., Alexandrova, A. N. (2016), Ensemble-Average Representation of Pt Clusters in Catalysis, *J. Chem. Theory Comput.*, 12, 12, 6213–6226 (DOI: 10.1021/acs.jctc.6b00994)
46. Verga, L. G., Aarons, J., Sarwar, M., Thompsett, D., Russell, A. E., Skylaris, C. K. (2016), Effect of Graphene Support on Large Pt Nanoparticles, *Phys. Chem. Chem. Phys.*, 18, 48, 33093–33101 (DOI: 10.1039/c6cp07334d)
47. Barrena-Espés, D., Boneta, S., Polo, V., Munárriz, J. (2023), Exploring the Potential Energy Surface of Pt_6 Sub-Nano Clusters over Graphene, *Int. J. Mol. Sci.*, 24, 1, 870 (DOI: 10.3390/ijms24010870)
48. Ramos-Sanchez, G., Balbuena, P. B. (2013), Interactions of Platinum Clusters with a Graphite Substrate, *Phys. Chem. Chem. Phys.*, 15, 28, 11950–11960 (DOI: 10.1039/c3cp51791h)
49. Doye, J. P. K., Wales, D. J. (1999), The Dynamics of Structural Transitions in Sodium Chloride Clusters, *J. Chem. Phys.*, 111, 24, 4800–4811 (DOI: 10.1063/1.480465)
50. Bord, J., Kirchhoff, B., Baldofski, M., Jung, C., Jacob, T. (2023), An Atomistic View of Platinum Cluster Growth on Pristine and Defective Graphene Supports, *Small*, 19, 10, 2207484 (DOI: 10.1002/smll.202207484)
51. Jinnouchi, R., Toyoda, E., Hatanaka, T., Morimoto, Y. (2010), First Principles Calculations on Site-Dependent Dissolution Potentials of

Supported and Unsupported Pt Particles, *J. Phys. Chem. C*, 114, 41, 17854–17863 (DOI: 10.1021/jp106593d)

52. Tian, Y., Liu, Y. J., Zhao, J. X., Ding, Y. H. (2015), High Stability and Superior Catalytic Reactivity of N-Doped Graphene Supporting Pt Nanoparticles for ORR: A DFT Study, *RSC Adv.*, 5, 43, 34265–34274 (DOI: 10.1039/c5ra02585k)
53. Mahata, A., Rawat, K. S., Choudhuri, I., Pathak, B. (2016), Cuboctahedral vs. Octahedral Platinum Nanoclusters: Shape-Dependent Catalytic Activity, *Catal. Sci. Technol.*, 6, 21, 7685–7695 (DOI: 10.1039/c6cy01709f)
54. Liu, X., Meng, C., Han, Y. (2013), Defective Graphene Supported MPd_{12} (M = Fe, Co, Ni, Cu, Zn, Pd) Nanoparticles for ORR: A First-Principles Study, *J. Phys. Chem. C*, 117, 3, 1126–1133 (DOI: 10.1021/jp3090952)
55. Hammer, B., Nørskov, J. K. (1995), Electronic Factors Determining the Reactivity of Metal Surfaces, *Surf. Sci.*, 343, 3, 211–220 (DOI: 10.1016/0039-6028(96)80007-0)
56. Duan, Z., Wang, G. (2011), A First-Principles Study of ORR on a Pt(111) Surface Modified by a Subsurface Transition Metal (M = Ni, Co, or Fe), *Phys. Chem. Chem. Phys.*, 13, 45, 20178–20187 (DOI: 10.1039/c1cp21687b)
57. Yang, S., Kim, J., Tak, Y. J., Soon, A., Lee, H. (2016), Single-Atom Catalyst of Platinum Supported on Titanium Nitride for Selective Electrochemical Reactions, *Angew. Chem. Int. Ed.*, 55, 6, 1822–1826 (DOI: 10.1002/anie.201509241)
58. Wang, Y., Balbuena, P. B. (2005), Potential Energy Surface Profile of ORR on a Pt Cluster: Adsorption and Decomposition of OOH and H_2O_2, *J. Chem. Theory Comput.*, 1, 5, 814–827 (DOI: 10.1021/ct0500794)

59. Nair, A. S., Pathak, B. (2019), Computational Screening for ORR Activity of 3d Transition Metal Based M@Pt Core-Shell Clusters, *J. Phys. Chem. C*, 123, 6, 3013–3022 (DOI: 10.1021/acs.jpcc.8b11483)
60. Sun, G., Sautet, P. (2018), Metastable Structures in Cluster Catalysis from First-Principles: Structural Ensemble and Metastability Triggered Reactivity, *J. Am. Chem. Soc.*, 140, 8, 2816–2827 (DOI: 10.1021/jacs.7b11239)
61. Zandkarimi, B., Gorey, T. J., Li, G., Munarriz, J., Anderson, S. L., Alexandrova, A. N. (2020), Alloying with Sn Suppresses Sintering of Size-Selected Subnano Pt Clusters on SiO_2, *Chem. Mater.*, 32, 19, 8400–8410 (DOI: 10.1021/acs.chemmater.0c02926)
62. Tsunoyama, H., Ohnuma, A., Takahashi, K., Velloth, A., Ehara, M., Ichikuni, N., Tabuchi, M., Nakajima, A. (2019), Enhanced Oxygen Reduction Activity of Platinum Subnanocluster Catalysts through Charge Redistribution, *Chem. Commun.*, 55, 84, 12663–12666 (DOI: 10.1039/c9cc06327g)
63. Mathew, K., Sundararaman, R., Letchworth-Weaver, K., Arias, T. A., Hennig, R. G. (2014), Implicit Solvation Model for DFT Study of Nanocrystal Surfaces and Reaction Pathways, *J. Chem. Phys.*, 140, 8, 084106 (DOI: 10.1063/1.4865107)
64. Nair, A. S., Pathak, B. (2019), Importance of Dispersion and Relativistic Effects for ORR Overpotential Calculation on Pt(111), *arXiv*, 1904.12345
65. Taleblou, M., Camellone, M. F., Fabris, S., Piccinin, S. (2022), Oxidation of Gas-Phase and Supported Pt Nanoclusters: An Ab Initio Investigation, *J. Phys. Chem. C*, 126, 26, 10787–10799 (DOI: 10.1021/acs.jpcc.2c02176)
66. Lym, J., Wittreich, G. R., Vlachos, D. G. (2020), A Python Multiscale Thermochemistry Toolbox (PMuTT), *Comput. Phys. Commun.*, 247, 106864 (DOI: 10.1016/j.cpc.2019.106864)

67. Ding, X., Jia, C., Ma, P., Chen, H., Xue, J., Wang, D., Wang, R., Cao, H., Zuo, M., Zhou, S., Zhang, Z., Zeng, J., Bao, J. (2023), Remote Synergy between Heterogeneous Single Atoms and Clusters for Enhanced Oxygen Evolution, *Nano Lett.*, 23, 8, 3265–3275 (DOI: 10.1021/acs.nanolett.3c00228)

CHAPTER 5

Deciphering Key Descriptors for Non-Scaling Relationships

CHAPTER 5

5.1 Introduction

Electrochemical energy conversion technologies are widely regarded as one of the most promising solutions to address global energy demands and environmental challenges within a sustainable, carbon-neutral framework *[1]*. Among them, fuel cells *[2]* and metal-air batteries *[3]* stand out as potential candidates for clean energy applications. However, their large-scale deployment is hindered by the sluggish kinetics of the oxygen reduction reaction (ORR) at the cathode, which leads to high overpotentials and reduced efficiency. To this end, platinum-based electrocatalysts are valued for their high activity for ORR; however, their widespread commercialization is limited by high cost and vulnerability to poisoning during long-term operation, motivating the search for low-cost alternatives *[4]*.

Subnanometer clusters have recently drawn significant attention as a distinct class of electrocatalysts because of their high atomic utilization and remarkable catalytic properties. Unlike bulk materials, these clusters at finite temperatures are characterized by a relatively flat potential energy surface, which imparts fluxionality and non-Arrhenius behavior *[5]*. Their large fraction of undercoordinated sites contributes to dynamic structural changes, resulting in size-dependent, non-monotonic catalytic activity *[6]*. In this non-scalable regime, cluster structures and electronic distributions are highly complex, giving rise to metastability-driven reactivity. Therefore, accurate modelling of their catalytic performance requires ensemble representations that capture the diversity of accessible states *[7]*. Importantly, the interaction between subnano clusters and their support can locally reshape coordination environments, modify conventional scaling relationships, and create opportunities to overcome the volcano-plot limitations *[8]*. Despite these promises, systematic evaluation of how different sizes of supported subnano clusters impact the catalytic activity remains limited. This gap reflects both the intrinsic complexity of the non-

scalable regime and the challenges of performing high-throughput *ab initio* simulations in a size range where "every atom count".

Machine learning (ML) has emerged as a powerful tool for accelerating catalyst discovery by capturing complex interactions between catalysts and adsorbates, while unravelling the multidimensional structure-property relationships *[9]*. ML algorithms have been applied to a wide range of electrochemical reactions, enabling faster discovery and optimization of catalysts across vast chemical spaces *[10,11]*. Descriptor-driven frameworks have also been used to identify cost-effective Pt-free catalysts, including single-atom and dual-atom systems as well as high-entropy alloys *[12,13]*. While these advances demonstrate the promise of ML in catalyst discovery, their applications to subnanometer regime clusters remain limited. In particular, the factors governing adsorption energies and the catalytic behavior of supported subnano clusters are not fully understood. Additionally, systematic studies of the multifidelity features contributing to the non-scaling relationships arising from intermediate adsorption at undercoordinated active sites are still in their infancy and require in-depth data-driven explorations.

Herein, we present a combined DFT-ML study on the ORR activity of size-selected Pt_n clusters ($n = 7 - 13$) supported on graphene (Pt_n/G), as shown in Figure 5.1.

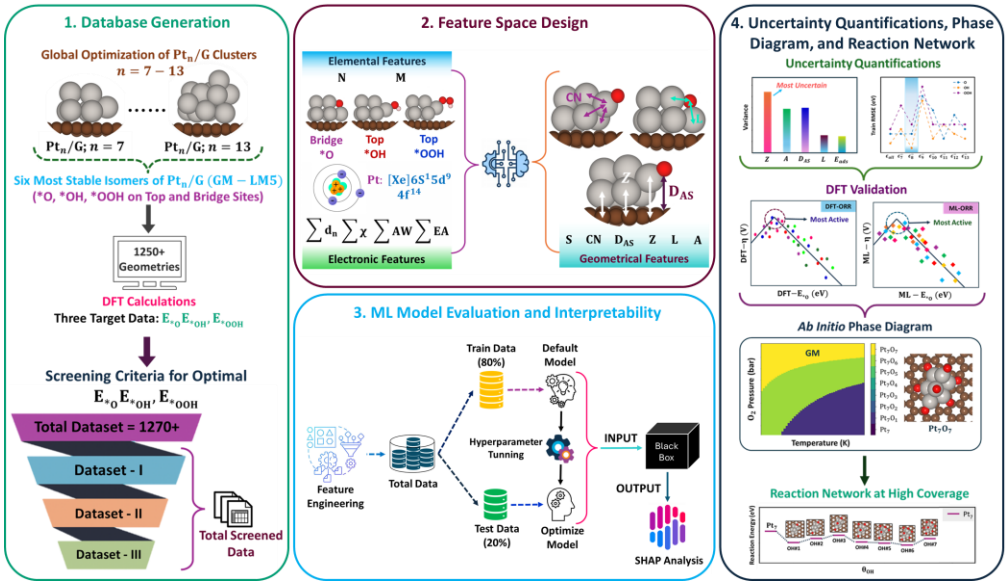


Figure 5.1: Machine learning (ML) framework for catalyst discovery. The workflow consists of four streamlined steps: (1) global optimization of graphene-supported $Pt_n(n = 7 - 13)$ subnano clusters to identify relevant global and local minima, followed by construction of an adsorption energy database for *O, *OH, and *OOH intermediates using DFT calculations; (2) feature space design to encode elemental, geometric, and electronic descriptors that capture the local coordination environment and structure-activity relationships in the subnanometer regime; (3) ML model evaluation and interpretability; and (4) uncertainty quantification, DFT validation, and *ab initio* phase diagrams under higher intermediate coverage and derivation of reaction networks.

To model the electrodes with graphene-supported subnano clusters Pt_n/G $(n = 7 - 13)$, we performed systematic global optimizations to sample the potential energy surface and extract an ensemble of low-energy isomers for catalytic investigations. We demonstrate that the fluxional nature of these clusters leads to strong variations in intermediate interactions, resulting in a complex, non-monotonic structure-activity relationship. Using the Sabatier principle as an optimal screening criterion, we trained ML models to predict ORR activity from non-*ab initio* features that capture the interplay of geometric and electronic environments. A

feature-driven interpretability analysis reveals that the geometric descriptors of heterogeneous active sites play a crucial role in breaking scaling relationships, a finding further validated by the uncertainty principle. The ML predictions were validated, active electrocatalysts were screened, and the ORR mechanism was derived following the adsorbate evolution mechanism. We further examined low-coverage-based active catalysts to construct reaction networks and phase diagrams at higher intermediate coverages. Overall, our combined DFT and ML explorations provide insight into the non-scaling behavior, fluxional dynamics, and coverage-dependent ORR activity of graphene-supported subnano clusters through combined DFT and ML explorations.

5.2 Computational Details

5.2.1 Electronic Structure Calculations

All the plane-wave spin-polarized density functional calculations were conducted using the Vienna *ab initio* simulation package (VASP) *[14]* with the Perdew-Burke-Ernzerhof (PBE) functional within the generalized gradient approximation (GGA) *[15]*. Ion core and valence electron interactions were described using the projector augmented wave (PAW) method *[16]*. For relaxation calculation, kinetic energy cut-offs of 470 eV with energy convergence criteria of 10^{-4} were employed, and geometric relaxation continued until forces on atoms were smaller than 0.02 eV/Å. Gaussian smearing with a sigma value of 0.2 eV was applied, and the Γ-centered (1×1×1) K-points grids were considered for the sampling of the Brillouin zone in congruence with the previous reports. A sufficiently large box with dimensions 18×20×20 $Å^3$ was utilized to optimize the distinct geometries to avoid the possibility of spurious interactions between the adjacent images along each axis.

5.2.2 Global Optimization Method

Initial geometries of Pt_n/G subnano clusters were generated using a parallel global optimization pathway toolkit (PGOPT) *[17]*. PGOPT generates structures based on the bond-length distribution algorithm (BLDA) through

an efficient sampling of configurational space. More details on the working mechanism of BLDA to generate the initial population for global optimization are provided in *[17]*. A total population of 4368 was generated and optimized with DFT, and only unique geometries were extracted, while the duplicates were discarded through a depth-first-search (DFS) similarity measurement.

5.2.3 Structural Design of Pt_n/G Subnano Electrocatalysts

To investigate the size-dependent ORR activity of Pt_n/G $(n = 7 - 13)$ subnano clusters, we first extracted representative structures by exhaustively sampling the energy landscape of each size. Global optimization searches were performed for Pt clusters containing 7-13 atoms on graphene using the Parallel Global Optimization Pathway Toolkit to identify the global minima (GM) and low-energy local minima (LM) configurations *[17]*. Multiple states for each size were generated using the bond-length distribution algorithm approach, and unique geometries were extracted through the depth-first search method. For their local optimization, the plane-wave spin-polarized DFT calculations were employed in the Vienna *ab initio* simulation package (VASP) *[14]* with the Perdew-Burke-Ernzerhof (PBE) functional within the generalized gradient approximation (GGA) *[15]*. The thermodynamic energy distributions of the sampled states, referenced to the GM energy ($\Omega = E_{LM} - E_{GM}$), are shown in Figure 5.2a, with the total number of sampled configurations indicated above each distribution. In total, 4368 unique catalytic states were obtained from independent searches across different cluster sizes. The Boltzmann distribution of all the isomers within Ω = 0.4 eV of the GM, representing the most accessible states for the neutral system, is shown in Figure 5.2b. Side and top views of the GM and low-lying LM structures included in our ensembles for further investigations are presented in Figured 5.2c and 5.3.

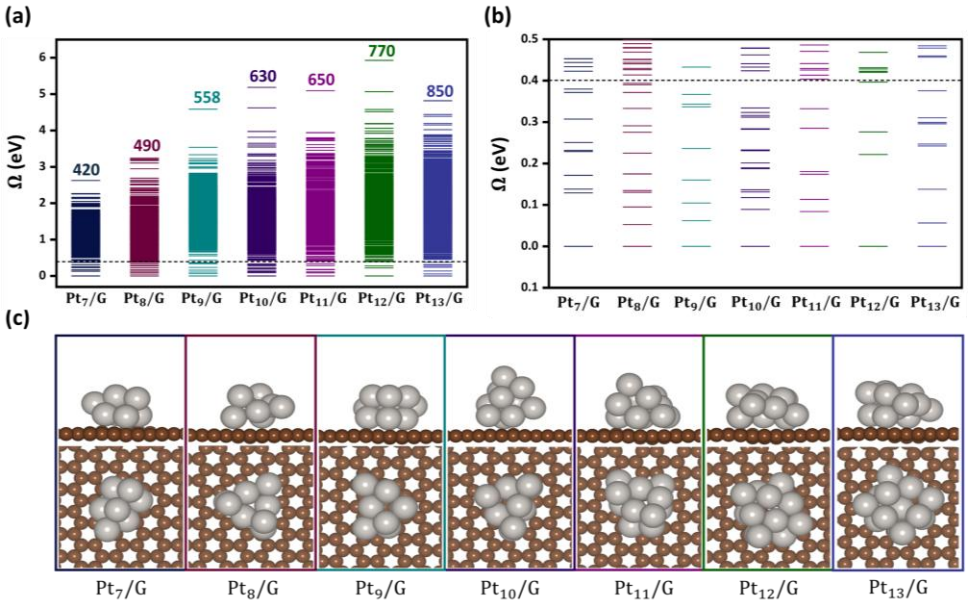
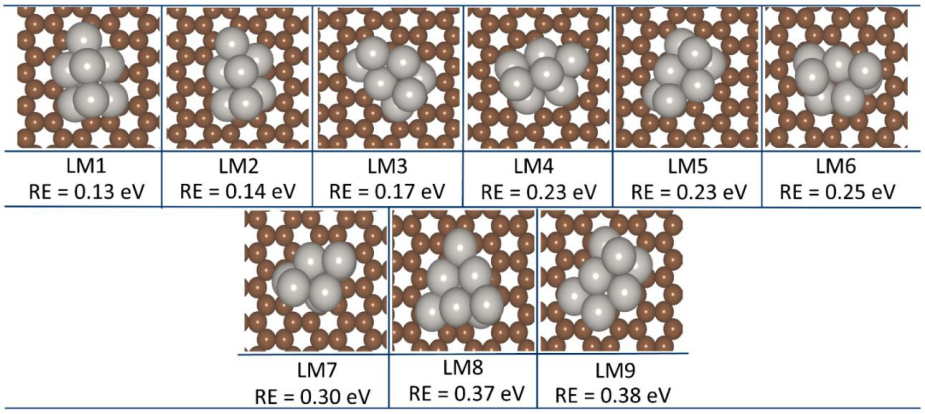
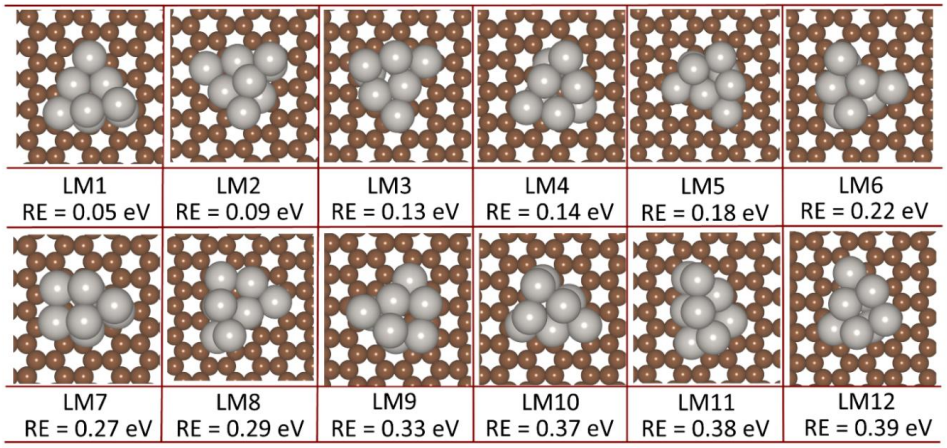


Figure 5.2: Energy distribution from global optimization. (a) Isomeric energy distribution of the global minima (GM) and local minima (LMs) of Pt_n/G $(n = 7 - 13)$ sampled in this study, with each state's energy referenced to that of the GM ($\Omega = E_{LM} - E_{GM}$, in eV). The number of configurations for each size regime is indicated above the corresponding distribution. The horizontal dashed line indicates the Boltzmann cutoff at Ω = 0.4 eV. (b) Boltzmann distribution of all unique structures within 0.4 eV of the GM, representing the most accessible states for the neutral systems. (c) Side and top views of the most stable GM configurations for each cluster size, highlighting the size-dependent structural variations of these subnano clusters. The geometries of the LM states for each size regime, utilized to construct our ensemble representation within a Ω = 0.4 eV difference for the ML-DFT analysis, are provided in Figure 5.3.

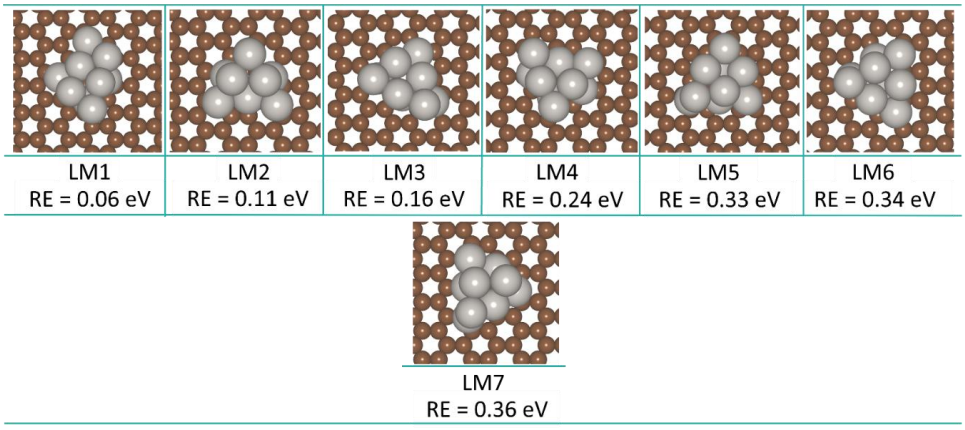
Pt_7/Graphene



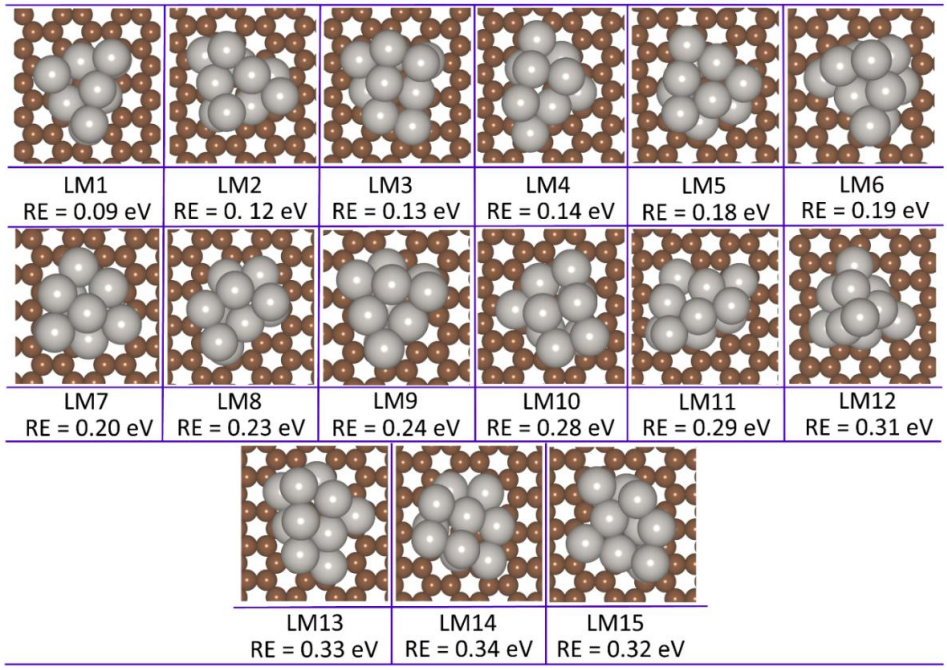
Pt_8/Graphene



Pt_9/Graphene

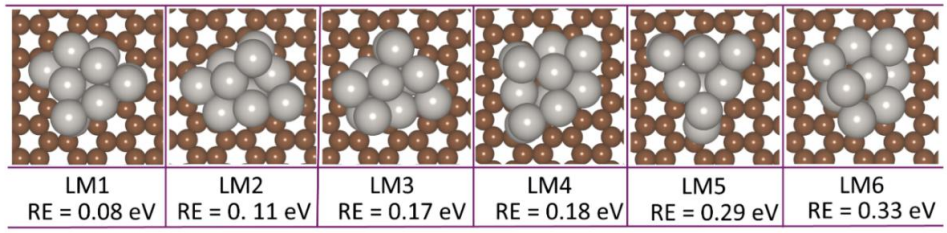


Pt_{10}/Graphene



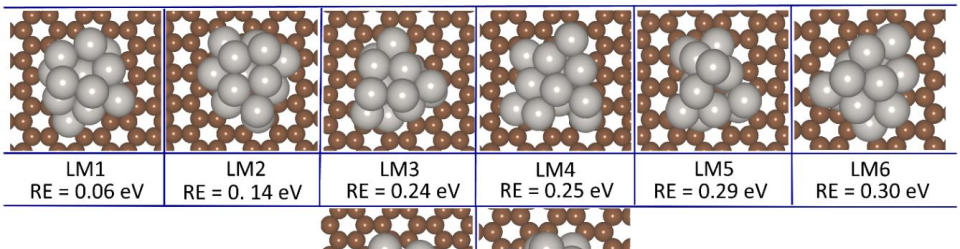
LM1	LM2	LM3	LM4	LM5	LM6
RE = 0.09 eV	RE = 0. 12 eV	RE = 0.13 eV	RE = 0.14 eV	RE = 0.18 eV	RE = 0.19 eV
LM7	**LM8**	**LM9**	**LM10**	**LM11**	**LM12**
RE = 0.20 eV	RE = 0.23 eV	RE = 0.24 eV	RE = 0.28 eV	RE = 0.29 eV	RE = 0.31 eV
LM13	**LM14**	**LM15**			
RE = 0.33 eV	RE = 0.34 eV	RE = 0.32 eV			

Pt_{11}/Graphene



LM1	LM2	LM3	LM4	LM5	LM6
RE = 0.08 eV	RE = 0. 11 eV	RE = 0.17 eV	RE = 0.18 eV	RE = 0.29 eV	RE = 0.33 eV

Pt_{13}/Graphene



LM1	LM2	LM3	LM4	LM5	LM6
RE = 0.06 eV	RE = 0. 14 eV	RE = 0.24 eV	RE = 0.25 eV	RE = 0.29 eV	RE = 0.30 eV
LM7	**LM8**				
RE = 0.31 eV	RE = 0.37 eV				

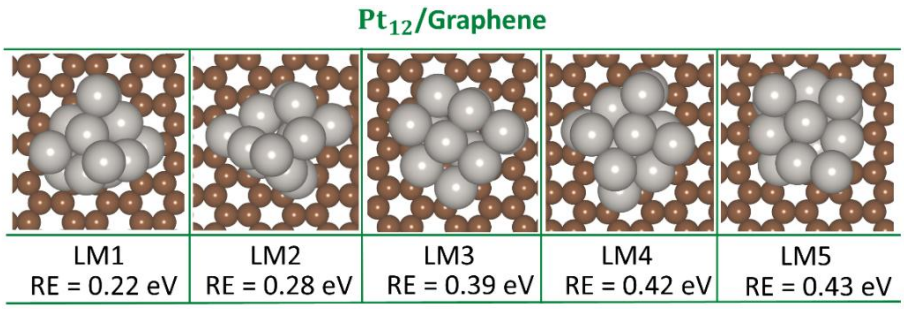


Figure 5.3: Low-lying local minima configurations of Pt_n/G $(n = 7 - 13)$ subnanoclusters and their relative energies (RE) with respect to the global minimum (in eV).

In the Pt_n/G $(n = 7 - 13)$ size regime, subnano cluster shows diverse conformations with a distinct evolution in morphology, symmetry, and stability driven by the interplay of Pt-Pt bonding and Pt-C interactions on the graphene substrate. For Pt_7/G, GM adopts a compact, non-planar C_{3v}-symmetry structure *[18]*, in contrast to the planar or prismatic motifs typically favored by unsupported or oxide-supported clusters *[19,20]*. Similarly, the accessible isomers of Pt_8/G are dominated by bilayer and trilayer non-planar arrangements with shallow energy landscapes, reflecting their high fluxionality and dynamic behavior. The LM1 of Pt_8/G lies only 0.05 eV above the GM, underscoring its high accessibility within the ensemble-derived activity. For Pt_9/G, although planar geometries are preferred for unsupported Pt clusters due to relativistic effects *[21]*, the supported GM shifts to a non-planar geometry, marking the onset of distinct 3D cluster character. As the size increases from Pt_{10}/G to Pt_{12}/G, the GM and LM adopt disordered tetrahedral and octahedral motifs, losing the "magic number" stability observed in vacuum because of strong d-π hybridization and binding with graphene *[22]*. For Pt_{13}/G, the GM becomes a distorted tricapped pentagonal prism or a distorted octahedron, in contrast to the icosahedral or high-symmetry structures typically observed for bare clusters. For further catalytic investigations, we constructed an ensemble of the most stable GM and the first five LMs, which are expected to have the

highest population within Ω = 0.4 eV of the GM, and performed a combined ML and DFT-based ORR mechanistic analysis on them.

5.3 Results and Discussion

5.3.1 Adsorption Energy Landscape of Pt_n/G clusters

In line with the Bell-Evans-Polanyi (BEP) principle, the adsorption energy (E_{ads}) of the ORR intermediate serves as an important descriptor to evaluate the catalyst's performance *[23]*. However, the inherent fluxionality of subnano clusters due to the presence of multiple heterogeneous active sites makes it challenging to identify the most stable adsorption configurations on the local atomic environment. Therefore, to generate our initial E_{ads} database, we optimized multiple configurations of single-intermediate adsorbed on GM and LM isomers, while sampling different top and bridge positions for principal ORR intermediates ($^*O$, $^*OH$, and $^*OOH$). The total number of configurations explored to construct the initial database is provided in Table 5.1. The E_{ads} for each intermediate was computed using the following equation:

$$E_{ads} = E_{Pt_n/G\,(*O,*OH,*OOH)} - (E_{Pt_n/G} + E_{*O,*OH,*OOH}) \qquad (5.1)$$

where $E_{Pt_n/G(*O,*OH,*OOH)}$, $E_{Pt_n/G}$, and $E_{*O,*OH,*OOH}$ represents the total energies of the cluster with $^*O$, $^*OH$, and $^*OOH$, supported cluster and their corresponding intermediates, respectively. The E_{ads} trends for each intermediate are summarized in Figure 5.4a, illustrating the overall energy distribution sampled across various Pt_n/G sizes and heterogeneous sites.

Table 5.1: Total number of configurations sampled for each isomer of Pt_n/G $(n = 7 - 13)$ to identify the most stable adsorption configurations.

Clusters	Sites	GM	LM1	LM2	LM3	LM4	LM5
Pt_7/G	Top	9	9	9	12	9	12
	Bridge	6	6	6	12	6	12
Pt_8/G	Top	15	18	18	15	12	18
	Bridge	11	20	18	12	9	18
Pt_9/G	Top	15	12	15	12	12	15
	Bridge	12	9	15	9	12	12
Pt_{10}/G	Top	18	9	15	18	15	9
	Bridge	15	17	12	15	12	12
Pt_{11}/G	Top	12	15	15	18	18	12
	Bridge	9	12	12	12	15	12

Pt_{12}/G	Top	15	15	18	21	18	21
	Bridge	12	18	28	21	12	21
Pt_{13}/G	Top	24	24	21	21	24	18
	Bridge	28	33	23	21	32	15

Table 5.2: Adsorption energies (E_{ads}) of ORR intermediates (in eV units) corresponding to their most stable intermediate-cluster adsorbed configuration. Here, t, and b represent top, and bridge sites, respectively. The sign * indicates adsorbed intermediates on active sites of isomers.

Clusters	Int.	Adsorption Energy (Eads)					
		GM	LM1	LM2	LM3	LM4	LM5
Pt_7/G	*O	-4.83(t)	4.91(b)	-5.01(t)	-5.06(b)	-4.89(t)	-4.89(b)
	*OH	-3.17(t)	-3.47(t)	-3.37(t)	-3.46(t)	-3.33(t)	-3.47(t)
	*OOH	-1.84(t)	-1.86(t)	-1.94(t)	-2.17(t)	-1.85(t)	-2.09(t)
Pt_8/G	*O	-4.94(b)	-4.96(t)	-5.92(b)	-4.83(t)	-4.99(t)	-5.69(t)
	*OH	-3.61(t)	-3.43(t)	-4.19(b)	-3.24(t)	-3.64(t)	-4.28(t)
	*OOH	-2.22(t)	-2.08(t)	-2.82(t)	-1.66(t)	-2.47(t)	-2.54(t)
Pt_9/G	*O	-4.86(t)	-5.56(b)	-4.95(t)	-4.72(b)	-5.06(b)	-4.97(t)
	*OH	-3.31(t)	-4.04(t)	-3.41(t)	-3.27(t)	-3.47(b)	-3.37(t)
	*OOH	-2.10(t)	-2.84(t)	-2.03(t)	-1.94(t)	-1.97(b)	-2.07(t)
Pt_{10}/G	*O	-5.67(t)	-5.09(t)	-5.54(b)	-5.48(t)	-5.31(b)	-4.86(t)
	*OH	-4.29(t)	-3.65(t)	-3.92(b)	-3.92(t)	-3.65(b)	-3.23(b)
	*OOH	-2.79(t)	-2.09(t)	-2.77(t)	-2.49(t)	-2.27(b)	-1.87(t)
Pt_{11}/G	*O	-4.94(t)	-5.14(b)	-6.16(b)	-5.14(b)	-5.17(t)	-4.85(t)
	*OH	-3.47(t)	-3.28(b)	-4.64(t)	-3.51(b)	-3.53(t)	-3.42(t)
	*OOH	-2.22(t)	-1.89(t)	-3.23(t)	-2.02(t)	-2.23(t)	-1.77(t)
Pt_{12}/G	*O	-4.98(t)	-5.03(t)	-5.18(t)	-5.60(b)	-5.49(b)	-5.06(t)
	*OH	-3.42(t)	-3.26(t)	-3.76(t)	-4.06(b)	-3.81(b)	-3.88(t)
	*OOH	-2.21(t)	-2.12(t)	-2.44(t)	-2.93(b)	-2.57(t)	-2.37(t)
Pt_{13}/G	*O	-5.06(b)	-5.19(b)	-5.46(t)	-5.28(b)	-5.53(t)	-4.93(t)
	*OH	-3.24(t)	-3.38(b)	-3.95(t)	-3.65(t)	-4.05(t)	-3.43(t)
	*OOH	-1.81(t)	-1.99(b)	-2.52(t)	-2.31(b)	-2.62(b)	-2.24(t)

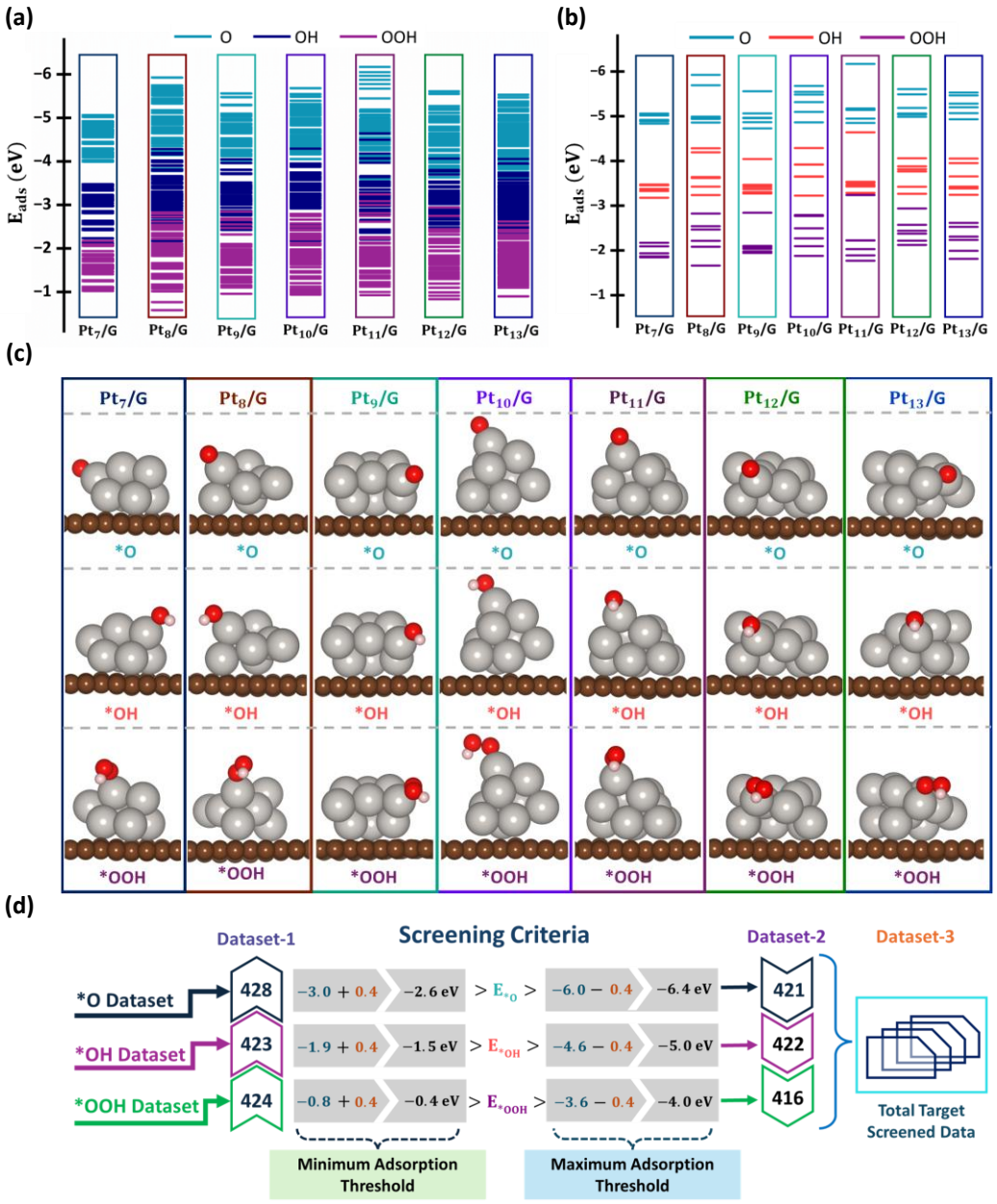


Figure 5.4: Database generation for intermediate adsorption. (a) Adsorption energy distribution (E_{ads}, in eV) of key ORR intermediates (*O, *OH, and *OOH) on Pt_n/G subnano clusters (n = 7-13), sampled across global minima (GM) and low-lying local minima (LM) structures. The variations reflect sensitivity to cluster size, isomeric distribution, and coordination environment. (b) Distribution of E_{ads} for the most stable single-intermediate configurations across GM and LM isomers within 0.4 eV of the GM, based on Boltzmann statistics. (c) Side-view representations of the most favorable

adsorption configurations for $^{*}O$, $^{*}OH$, and $^{*}OOH$ on GM structures of Pt_n/G clusters, where adsorbed species are denoted with an asterisk (*). Platinum, oxygen, and hydrogen atoms are shown in grey, red, and pink, respectively. (d) Schematic representation of the screening strategy used to extract optimal E_{ads} values for $^{*}O$, $^{*}OH$, and $^{*}OOH$ following the Sabatier principle, enabling identification of promising ORR electrocatalysts.

The E_{ads} values for each intermediate lack a consistent trend, with their peak energy distribution spanning distinct regions: −3.22 to −6.11 eV for $^{*}O$, −1.97 to −4.64 eV for $^{*}OH$, and −0.12 to −3.23 eV for $^{*}OOH$. Compared to the bulk Pt(111) surface with E_{*_O} = −4.69 eV, $E_{*_{OH}}$ = −2.63 eV, and $E_{*_{OOH}}$ = −1.64 eV *[24]*, the subnano clusters exhibit a broad range of E_{ads}, with the majority showing stronger binding to the intermediate, attributed to the presence of undercoordinated active sites. Notably, the E_{ads} for Pt_8/G exhibits a wide range of distribution for ORR intermediates across different isomers ($E_{*_O} = -5.92$ eV for LM2; $E_{*_{OH}} = -4.28$ eV for LM5; $E_{*_{OOH}} = -2.82$ for LM2), compared to other Pt_n/G size regimes (Table 5.2). Figure 5.4b represents the energy distribution of the most stable single-intermediate configurations across different isomers in the ensemble ($E_{*_O} = -5.92$ eV for LM2 of Pt_8/G, $E_{*_{OH}} = -4.29$ eV for GM of Pt_{10}/G, and $E_{*_{OOH}} = -2.84$ eV for LM1 of Pt_9/G) (Table 5.2). The side views of the optimized geometries for these configurations on the GMs are shown in Figure 5.4c. In the subnanometer regime, the intermediate preferentially binds to the top and bridge sites, unlike on the Pt(111) surface, where the *O intermediate favors adsorption at the hollow site *[24]*. Within this sampling space, we have curated a data set of 1275 data points with three intermediates (428 for E_{*_O}, 423 for $E_{*_{OH}}$, and 424 datapoints for $E_{*_{OOH}}$), represented as “Dataset-1” in Figure 5.4d.

Owing to their multiple undercoordinated sites, every atom in subnano clusters exhibits a unique electronic structure and coordination geometry, resulting in significant variations in E_{ads}. According to the Sabatier

principle *[25]*, the intermediate should neither bind too strongly ($E_{*O} = -6.11$ eV as in Pt_{11}/G) leading to chemisorption, nor too weakly ($E_{*OOH} = -0.12$ eV as in Pt_{13}/G) leading to physisorption, to achieve high ORR activity (Figure 5.4b). After acquiring Dataset-1, we applied three screening criteria based on the E_{ads} of each intermediate to screen the active electrocatalysts, as shown in Figure 5.4d. The upper and lower limits for E_{*O}, E_{*OH}, and E_{*OOH} were determined through a comprehensive analysis of various systems, including Pt_{1-10} subnano clusters, Pt_{79} nanocluster, and Pt(111) surface *[6,19,24,26,27]*. Additionally, to account for the metastability-triggered reactivity, weighted by the Boltzmann statistics, we extended each energy window by ± 0.4 eV (total extension of 0.8 eV), allowing us to preliminarily exclude the thermodynamically inaccessible states *[6,28-30]*. Following this screening criterion, we curated Dataset-2, comprising 421, 422, and 416 datapoints for E_{*O}, E_{*OH}, and E_{*OOH}, respectively. Therefore, this E_{ads} dataset is uniformly sampled across catalysts with varying sizes, coordination environments, and three intermediates, which is pivotal for developing robust ML models.

5.3.2 Feature Space Design for ML Models.

Integrating both physical and chemical descriptors into ML models is critical for achieving high predictive accuracy. In this context, feature engineering plays an important role in model development, enabling the effective representation of complex local chemical environments while mapping the structure-activity relationships *[9]*. To capture the factors governing intermediate adsorption on active sites of subnano clusters, we constructed a comprehensive set of input descriptors, categorized into three groups: elemental, electronic, and geometric, as summarized in Table 5.3.

Table 5.3: List of Primary Features including elemental, electronic, and geometric descriptors.

Groups	Descriptors	Symbol
Elemental Features	Number of Pt atoms	N
	Number of exposed Pt atom of the cluster	M
Electronic Features	Sum of number of d electrons	Σd_n
	Sum of electronegativity of the cluster	$\Sigma\chi$
	Sum of electron affinity of the cluster	ΣEA
	Sum of atomic weight of cluster	ΣAW
Geometric Features	Adsorption site (top, bridge)	S
	Coordination number	CN
	Perpendicular distance between adsorbate and surface	D_{AS}
	Perpendicular distance between active Pt atom and surface	Z
	Pt-O bond length	L
	Number of Pt atoms binding to the surface	A

Elemental descriptors characterize the physical composition of the catalysts, providing a direct measure of cluster size and the density of the active sites. Electronic descriptors characterize the electronic structure of the cluster, which reflects its ability to donate or accept electrons during adsorption, thereby influencing reactivity. Geometric features define the local chemical environment of the subnano clusters. Features including the perpendicular distance between the adsorbate (D_{AS}) and the active Pt atoms (Z) relative to the surface are especially significant, as they describe the strength of interaction between the Pt atoms, adsorbates, and the surface. Together, these features capture the structural and electronic determinants governing catalytic performance.

Further, to derive a rational and non-redundant feature set for ML investigation, we evaluated both linear and monotonic relationships between input features and the target E_{ads} using Pearson's correlation coefficient (PCC) and Spearman's rank correlation coefficient (SCC) (Figure 5.5) *[31,32]*. To minimise redundancy, we retained only those features with low

feature-feature correlation ($|PCC| < 0.8$ and $|SCC| < 0.8$) and high correlation with the target variable. The target E_{ads} for each intermediate can be expressed as a function of the most relevant features:

$$E_{Pt_n/G}^{*O,*OH,*OOH} = f(\Sigma d_n, S, CN, D_{AS}, Z, L, A) \quad (5.2)$$

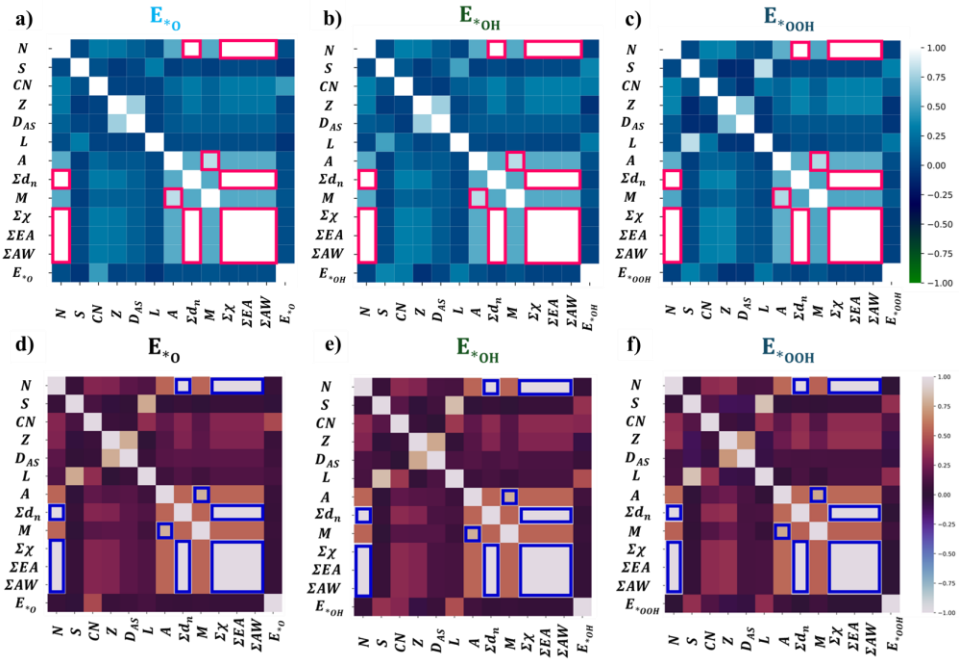


Figure 5.5. (a-c) The Pearson Correlation Coefficient (PCC) and (d-f) Spearman's correlation coefficient (SCC) matrix for E_{*O}, E_{*OH}, and E_{*OOH}, where the colour gradient encodes the strength and direction of the correlations, ranging from strong positive correlation (+1) to strong negative correlation (−1).

5.3.3 Model Training and Performance Evaluation

Following feature engineering, model training was carried out to optimise the parameters of various ML algorithms to minimise the discrepancy between predicted and actual E_{ads}. Dataset-2 was split into training and testing subsets using an 80:20 ratio. Hyperparameter tuning was performed using both grid search and randomised search strategies, each integrated with 5-fold cross-validation to identify optimal models. Model performance was evaluated based on the lowest mean absolute error (MAE), root mean square error (RMSE), and the highest coefficient of determination (R^2)

values. To ensure robust and generalizable prediction across the chemically diverse subnano cluster space, seven regression algorithms were employed: kernel ridge regression (KRR), eXtreme gradient boosting regression (XGBR), random forest regression (RFR), adaptive boosting regression (ABR), extra trees regression (ETR), gradient boosting regression (GBR), and CatBoost regression (CR). All models were implemented using the open-source Scikit library *[33]*. Given the complexity and heterogeneity of the subnanometer catalyst, the combination of kernel-based and ensemble-based methods offers a balanced strategy capable of capturing both linear and non-linear relationships in the data. Model performance metrics are summarized in Figure 5.6.

In Figure 5.6, panels (a-c) show the comparative test performance of various ML models for predicting the E_{ads} of the three key intermediates. Due to variations in the statistical distribution and feature-target relationships across the datasets, different models performed best for different intermediates. The GBR model showed the highest accuracy in predicting E_{*O} and E_{*OOH} with MAE/RMSE values of 0.26/0.33 and 0.28/0.36 eV, respectively, for the test datasets. In contrast, the RFR outperformed other models for E_{*OH} with an MAE/RMSE of 0.21/0.28 eV for the test dataset. Notably, we have achieved this accuracy for E_{*O}, E_{*OH}, and E_{*OOH} values across different heterogeneous sites of Pt_n/G subnano clusters (n=7-13). The predictive reliability of these models is further supported by the parity plots in panels (d-f) of Figure 5.6, which display correlations between DFT-calculated and ML-predicted E_{ads} for both training and testing sets. Based on the above ML model performances, GBR was selected for predicting E_{*O} and E_{*OOH}, while RFR was selected for predicting E_{*OH} in the subsequent E_{ads} evaluations and ORR mechanistic pathway analysis.

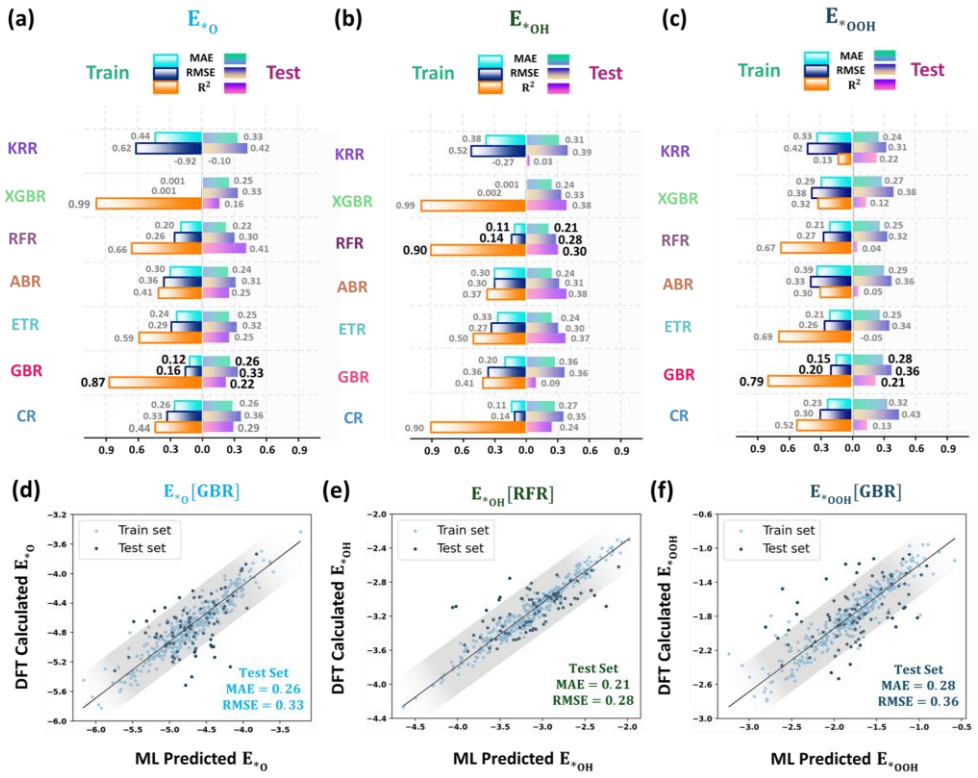


Figure 5.6: ML model training and evaluation. Mean absolute error (MAE), and root-mean-square error (RMSE) for KRR, XGBR, RFR, ABR, ETR, GBR, and CR models using (a) E_{*O}, (b) E_{*OH}, and (c) E_{*OOH} datasets (eV units). The best-performing model for each intermediate is highlighted in bold. Parity plots comparing DFT-calculated and ML-predicted values for (d) E_{*O}, (e) E_{*OH}, and (f) E_{*OOH}, obtained from the respective best-performing models.

5.3.4 Feature-driven Interpretability of Low Scaling Relationships

Note that subnano clusters can readily adapt their shapes and binding sites in response to different adsorbates, which gives rise to non-Arrhenius behavior *[6]*. To examine this, we analyzed the correlations between E_{*O} vs. E_{*OH}, and E_{*OH} vs. E_{*OOH} for both ML and DFT datasets (Figure 5.7). The consistently low R^2 values across different size regimes indicate the breaking of scaling relationships in the subnanometer regime. This non-scaling behavior arises from the high fluxionality of the clusters and the

presence of multiple heterogeneous active sites on their irregular geometries, in contrast to bulk surfaces where $R^2 = 0.91$ *[8]*.

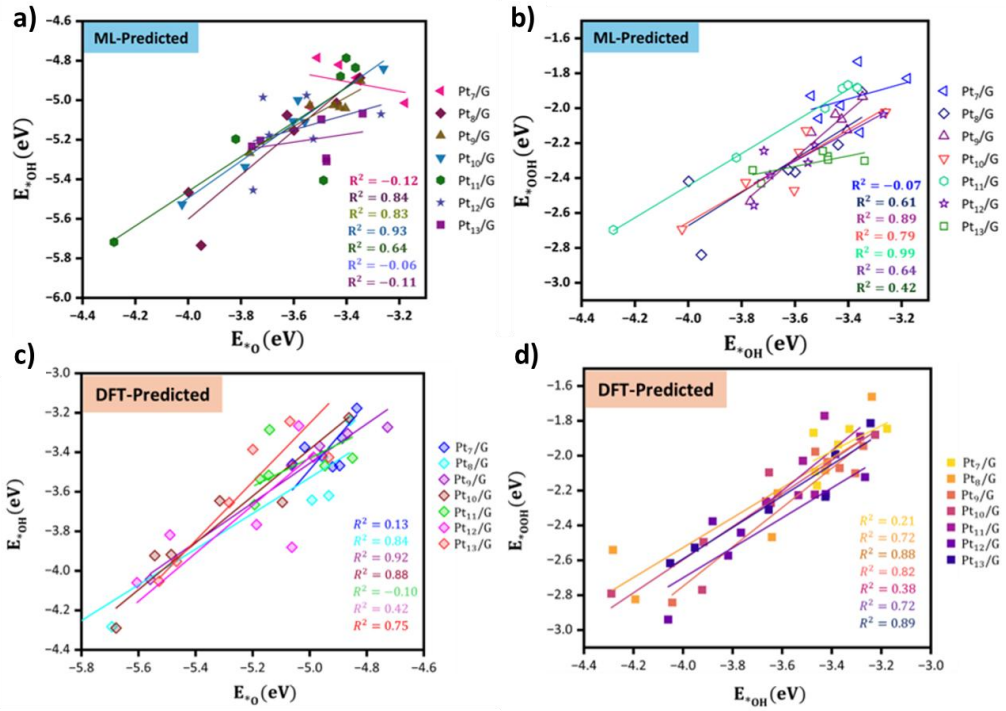


Figure 5.7: Scaling Relationships Analysis for Pt_n/G $(n = 7 - 13)$ subnano clusters. Panel (a) and (b) show the ML-predicted scaling relationships between E_{*O} vs. E_{*OH} and E_{*OH} vs. E_{*OOH}, respectively. The coefficient of determination (R^2) of the best-fitted line is displayed on the right side of each plot. Panels (c) and (d) present the corresponding DFT-predicted scaling relationships for E_{*O} vs. E_{*OH} and E_{*OH} vs. E_{*OOH}, respectively.

To identify the factors responsible for this deviation, we performed a feature-driven uncertainty analysis to identify the size regimes and key geometric descriptors. Figure 5.8 shows that the E_{*O}, E_{*OH}, and E_{*OOH} are spread across distinct and irregular regions, further reflecting this heterogeneity. We then carried out a leave-one-feature-out analysis on Pt_n/G, where each size regime was sequentially excluded, and the model performance (RMSE and MAE) was re-evaluated for E_{*O}, E_{*OH}, and E_{*OOH} (Figure 5.8a-d). Notably, excluding Pt_8/G clusters (denoted as ϵ_8^-)

consistently lowered both training and test errors, whereas removing other sizes had a minimal effect. Unlike bare Pt_8 clusters with planar geometries *[18]*, Pt_8/G isomers are especially heterogeneous, contributing most to the low R^2 values. Their broad and inconsistent descriptor distribution acts as noise, reducing the overall model's robustness.

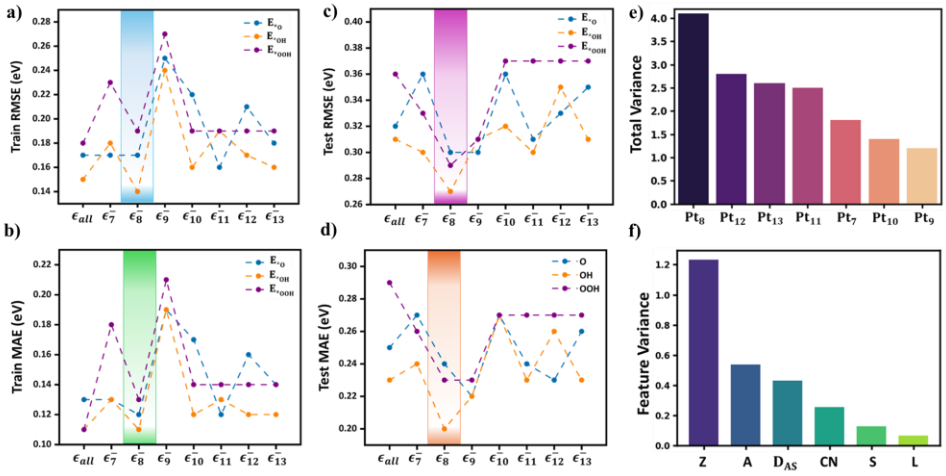


Figure 5.8. Quantitative Evaluation of Uncertainty in the Dataset (a-d) Training and testing RMSE and MAE values of the GBR model upon systematic exclusion of individual cluster sizes. The shaded bar highlights the minimum error observed for different intermediates when Pt_8/G clusters are excluded from Dataset-2 (denoted as ϵ_8^-). (e) Bar plots showing the variance contributions in the GBR model across cluster sizes, based on descriptor-driven cluster behavior within the optimal E_{ads} window (-4.92 < E_{ads} < -1.23 eV). (f) Quantification of geometric feature-based variance in the Pt_8/G isomers dataset.

We further analyzed the combined variance of Pt_n/G clusters within an optimal E_{ads} window (-4.92 < E_{ads} < -1.23 eV) relevant for ORR intermediates (Figure 5.8e). Again, Pt_8/G displayed the highest variance, strongly linked to overfitting and reduced predictive accuracy compared to other cluster sizes. To interpret the underlying cause of this uncertainty, we performed a feature-level variance analysis of Pt_8/G systems (Figure 5.8f). Geometric descriptors, specifically Z, A, and D_{AS}, contribute the most to the variance, followed by CN. Among these, Z emerges as the single most

influential descriptor, reflecting its dual role in encoding cluster height (anchoring of graphene) and electronic coupling with the substrate. This explains the heterogeneity observed in Pt_8 clusters anchoring, spatial Pt atom arrangements, and coordination environments across GM and LM isomers, which collectively degrade model performance.

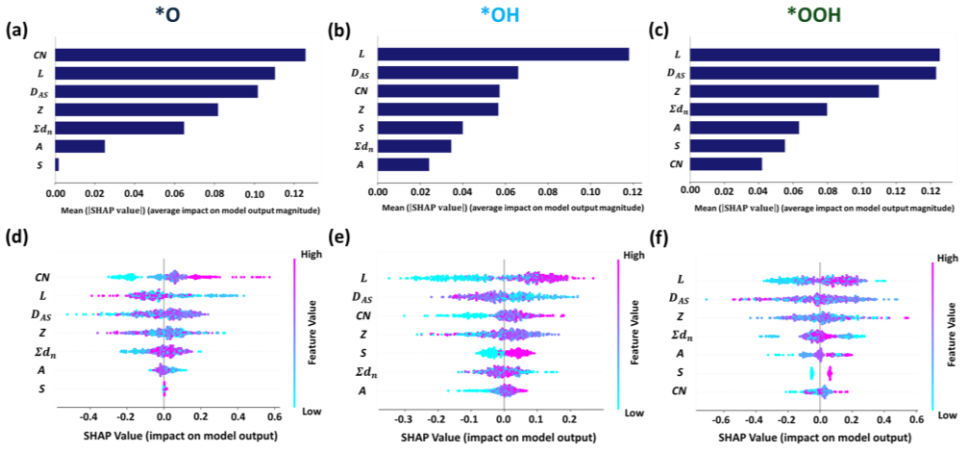


Figure 5.9: SHAP Analysis: Global interpretation of the best-performing models for predicting the adsorption energies of a) E_{*O}, (b) E_{*OH}, and (c) E_{*OOH}. The top row shows bar plots of mean absolute SHAP values, indicating the average importance of each feature in the model's predictions. The bottom row (d–f) presents beeswarm plots for (d) E_{*O}, (e) E_{*OH}, and (f) E_{*OOH}, illustrating how individual feature values affect model outputs. Each point represents a sample, with its SHAP value plotted on the x-axis. The color gradient along each feature axis reflects the actual feature value, ranging from low (cyan) to high (magenta).

Although ML models capture complex non-linear patterns effectively, they often lack interpretability, reinforcing their "black-box" perception. To improve transparency, we applied TreeExplainer-based SHapley Additive exPlanations (SHAP) analysis *[34]*, a cooperative game theory approach that assigns a contribution value to each feature for a given prediction. SHAP analysis provides both global and local interpretability, offering insights into feature importance and model decision pathways for E_{ads}. A detailed SHAP-based global feature attribution analysis is provided in

Figure 5.9. For *O, the coordination number (CN) of the binding site exerts the strongest influence, with higher CN correlating with weaker adsorption due to electron delocalization and reduced orbital overlap. In contrast, *OH and *OOH adsorption are primarily controlled by the bond length (L) and the perpendicular adsorbate-surface distance (D_{AS}), underscoring their sensitivity to local geometric and electronic environments. Specifically, shorter L values lead to stronger binding, whereas larger D_{AS} values weaken adsorption. Interestingly, the effect of CN is intermediate-specific: it enhances *O binding but suppresses *OOH adsorption, suggesting preferential binding at lower-coordinated sites. Collectively, these results demonstrate that adsorption energetics in subnano clusters arise from the interplay between coordination environment, geometric structure, and electronic coupling, with bond length emerging as the most consistent determinant across all intermediates.

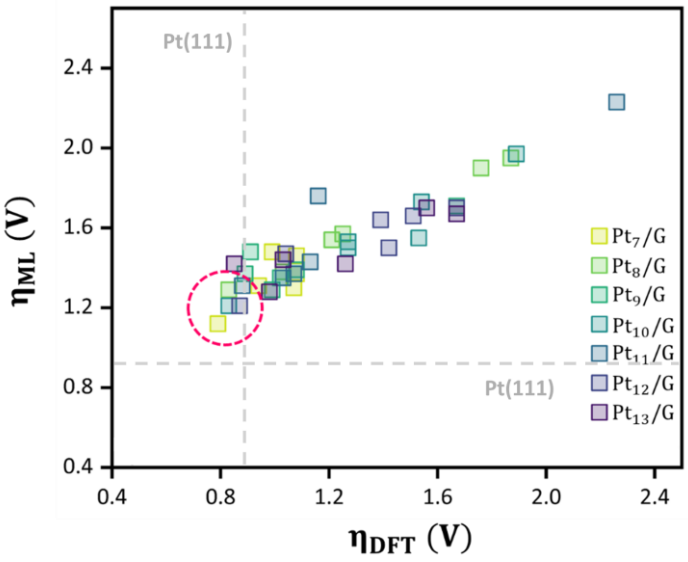


Figure 5.10: Correlation between DFT-calculated and ML-predicted overpotential (η) values for the screened catalysts. The two faded dashed grey lines intersect at η = 0.92 V, corresponding to the benchmark overpotential of Pt(111) (faded grey dashed line) and serve as a reference for identifying top-performing catalysts. The five most catalytically

efficient candidates exhibiting lower overpotentials in both methods are highlighted within the dark pink circle.

5.3.5 Validation of ML-Predicted Adsorption Energy.

To explore the catalytic activity of all the isomers extracted in our ensemble, we considered the most stable intermediate adsorption configuration from the DFT-calculated and their corresponding ML-predicted values from the E_{ads} dataset. Subsequently, we derive the ORR mechanism via an associative pathway to identify the active electrocatalysts, compiled as Dataset-3 in our investigation. We utilized the computational hydrogen electrode (CHE) model proposed by Nørskov and co-workers to calculate the reaction energy at 0 and 1.23 V, following the elementary steps outlined below *[35]*:

$$O_2(g) + (H^+ + e^-) + * \rightarrow OOH^* \quad \textbf{(}\Delta\mathbf{R_1}\textbf{)}$$

$$OOH^* + (H^+ + e^-) \rightarrow O^* + H_2O\,(l) \quad \textbf{(}\Delta\mathbf{R_2}\textbf{)}$$

$$O^* + (H^+ + e^-) \rightarrow OH^* \quad \textbf{(}\Delta\mathbf{R_3}\textbf{)}$$

$$OH^* + (H^+ + e^-) \rightarrow H_2O\,(l) + * \quad \textbf{(}\Delta\mathbf{R_4}\textbf{)}$$

The ORR activity of the subnano clusters in the ensemble is evaluated by the overpotential values (η) for the rate-determining step (RDS) at 1.23 V, using the following equations:

$$\eta_{ORR}^{DFT/ML} = \frac{\max(\Delta R_1, \Delta R_2, \Delta R_3, \Delta R_4)}{e} \quad (5.3)$$

The variable distribution of the DFT-calculated η_{ORR}^{DFT} and ML-predicted η_{ORR}^{ML} values for the RDS (Figure 5.10) underscores the non-monotonic ORR activity in the subnanometer regime with size and isomers. Specifically, the GM of Pt_7/G, LM3 of Pt_8/G, and LM5 of Pt_{10}/G emerged as the active electrocatalysts with lower η_{ORR}^{DFT} values of 0.79 V, 0.83 V, and 0.83 V, respectively, with RDS involving *OH → H_2O formation (ΔR_4) for each catalysts (Table 5.4). In both the DFT and ML datasets (Figure 5.11a,b), these catalysts lie near the apex of the volcano plots. Figure 5.11c-d presents bubble plots highlighting the similarity in the trends of η_{ORR}^{DFT} and η_{ORR}^{ML} values for all catalysts in our ensemble across each size range at 1.23 V.

Compared to bare Pt_7 and Pt_{13} clusters, the supported Pt_7/G and Pt_{13}/G isomers exhibits higher η values in both DFT and ML dataset *[19,29]*.

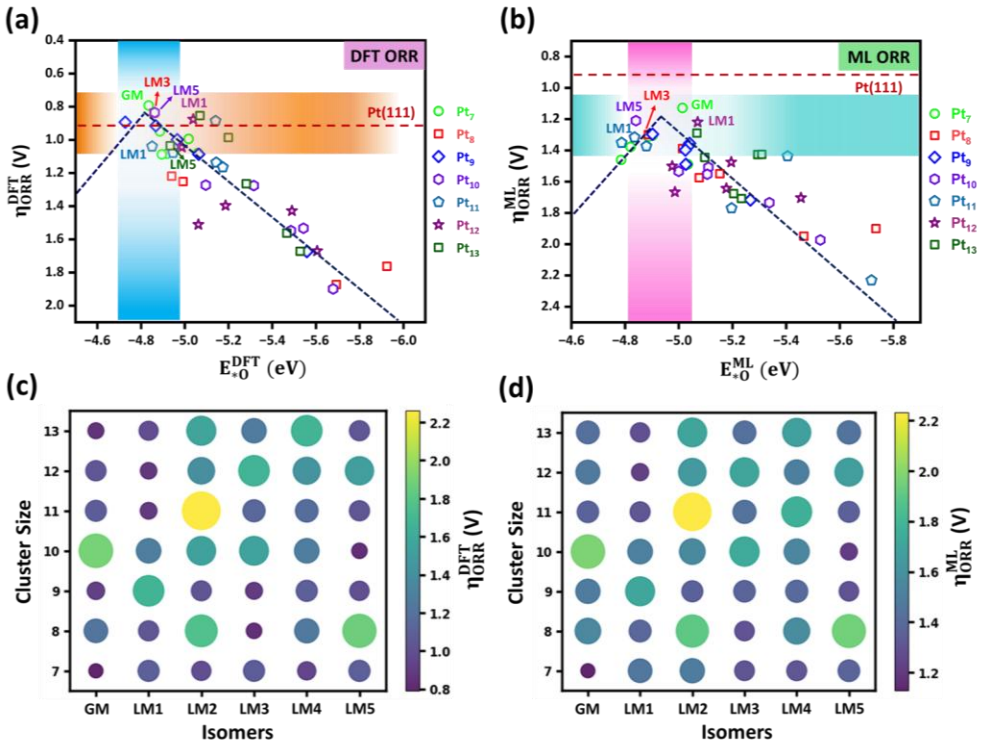


Figure 5.11: ORR volcano plot and activity mapping. Volcano plots showing the ORR performance of subnano clusters based on (a) DFT-calculated overpotentials (η_{ORR}^{DFT}) and (b) ML-predicted overpotential (η_{ORR}^{ML}) for electrocatalysts in Dataset 3. Catalysts near the apex, located at the intersection of the dual-gradient bands, exhibit the lowest η_{ORR}^{DFT} and η_{ORR}^{ML} values, indicating high ORR activity. The horizontal dashed line represents the overpotential of the Pt(111) surface (η = 0.92 V), taken from our previous investigation for comparison. Panels (c) and (d) show bubble plots of DFT-calculated η_{ORR}^{DFT} and ML-predicted η_{ORR}^{ML} values, respectively, across different cluster sizes and isomers.

Table 5.4: Overpotential value (η) of the thermodynamic rate-determining step (RDS) for best catalyst in the DFT and ML dataset within each size-regime. The elementary steps followed to derive the ORR mechanism via an associative pathway at 0 and 1.23 V are as follows:

Catalysts	Active Isomers	η_{ORR}^{DFT}	RDS (DFT)	η_{ORR}^{ML}	RDS (ML)
Pt_7/G	GM	0.79	$OH^* \rightarrow H_2O$ (l)	1.12	$OH^* \rightarrow H_2O$ (l)
Pt_8/G	LM3	0.83	$OH^* \rightarrow H_2O$ (l)	1.29	$OH^* \rightarrow H_2O$ (l)
Pt_9/G	LM3	0.89	$OH^* \rightarrow H_2O$ (l)	1.37	$OH^* \rightarrow H_2O$ (l)
Pt_{10}/G	LM5	0.83	$OH^* \rightarrow H_2O$ (l)	1.21	$OH^* \rightarrow H_2O$ (l)
Pt_{11}/G	LM1	0.88	$OH^* \rightarrow H_2O$ (l)	1.31	$OH^* \rightarrow H_2O$ (l)
Pt_{12}/G	LM1	0.87	$OH^* \rightarrow H_2O$ (l)	1.21	$OH^* \rightarrow H_2O$ (l)
Pt_{13}/G	GM	0.85	$OH^* \rightarrow H_2O$ (l)	1.42	$OH^* \rightarrow H_2O$ (l)

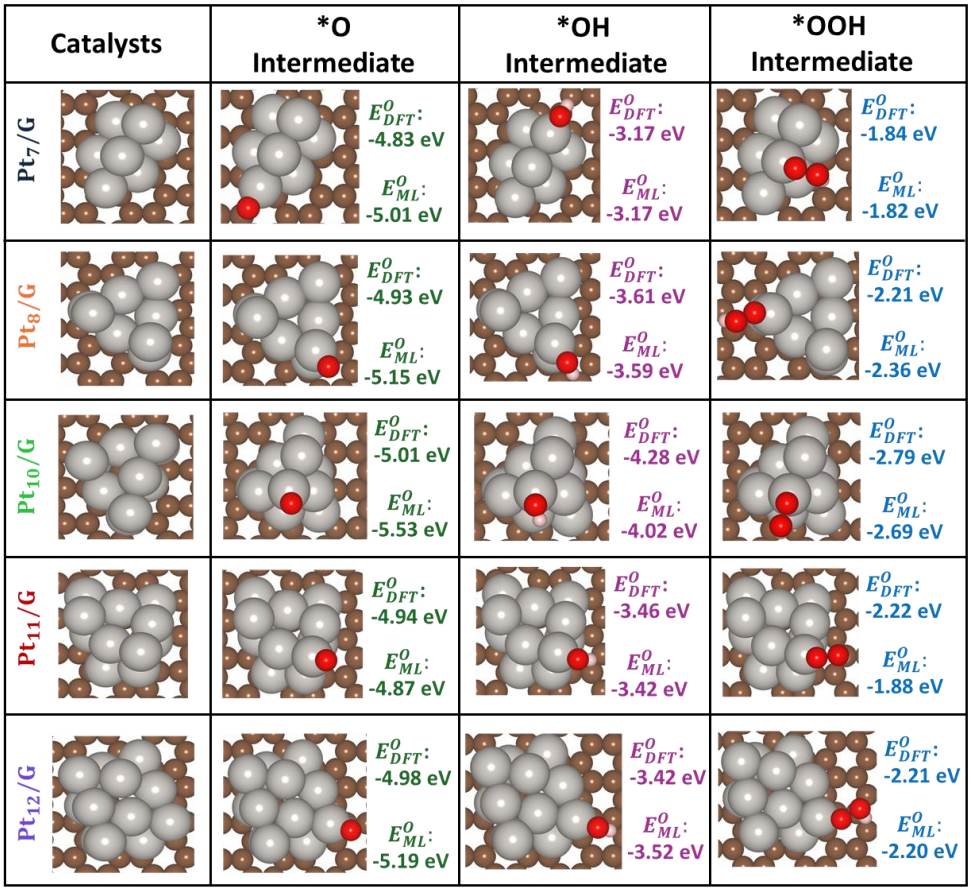


Figure 5.12: DFT-calculated and ML-predicted E_{*O}, E_{*OH}, and E_{*OOH} (in eV) of the five most promising catalysts. The overpotential values (η) for

the thermodynamically rate-determining step (RDS), calculated by combining the E_{*O}, E_{*OH}, and E_{*OOH} values in DFT and ML datasets in Dataset-3 for an associative pathway. The η values of the best catalyst were calculated through the computational hydrogen electrode (CHE) model, tabulated in Table 5.4.

For enhanced ORR performance in fuel cells, the catalyst at the apex of the activity volcano should exhibit near optimal E_{ads} for each intermediate *[36]*. Upon closer examination, the DFT-calculated and ML-predicted E_{*O}, E_{*OH}, and E_{*OOH} values of the best electrocatalysts at the volcano peak fall within an optimal energy range, closely matching the predefined E_{ads} interval defined by the equation:

$$-4.92\ \text{eV} < E_{ads} < -1.82\ \text{eV} \qquad (5.4)$$

Within this specified range, the active subnano clusters exhibit optimal binding affinities with each intermediate and emerge as the most promising candidates (Figure 5.12). This E_{ads} range in the subnanometer regime can therefore be used to directly identify active electrocatalysts, potentially overcoming the conventional scaling relationship for ORR.

5.3.6 *Ab initio* Thermodynamic Analysis

Since multiple cluster configurations are expected to coexist under reaction conditions, we investigated the nature of active isomers at higher intermediate coverage. Under experimental ORR conditions, Pt clusters are prone to oxidation, forming oxide-driven structures at ambient temperatures and pressures relevant to catalysis *[37]*. To capture these effects, we developed an *ab initio* thermodynamic framework to analyse the oxidation thermodynamics of Pt_nO_x (n=7-13, x=1-13). For each size and coverage, we sample the configurational space by generating oxide-based structures, followed by DFT optimisation. In this framework, the chemical potential of gas-phase oxygen was treated as a function of temperature and pressure, $\mu_O(T, P)$, expressed as:

$$\mu_O(T,P) = \frac{1}{2}\left[E_{O_2}(T = 0K, P^0) + \mu_{O_2}(T, P^0) + k_B \ln\left(\frac{P}{P^0}\right)\right] \qquad (5.5)$$

where T is temperature, P is the partial pressure of oxygen, P^0 is the standard atmospheric pressure, and k_B is the Boltzmann constant. Vibrational contributions from Pt and O atoms were included in the free energy calculations. The Python Multiscale Thermochemistry Toolbox (pMuTT) was used to compute temperature- and pressure-dependent thermodynamic corrections from vibrational frequency analysis *[38]*. The resulting phase diagram of *O coverage at ambient ORR conditions are shown in Figure 5.13.

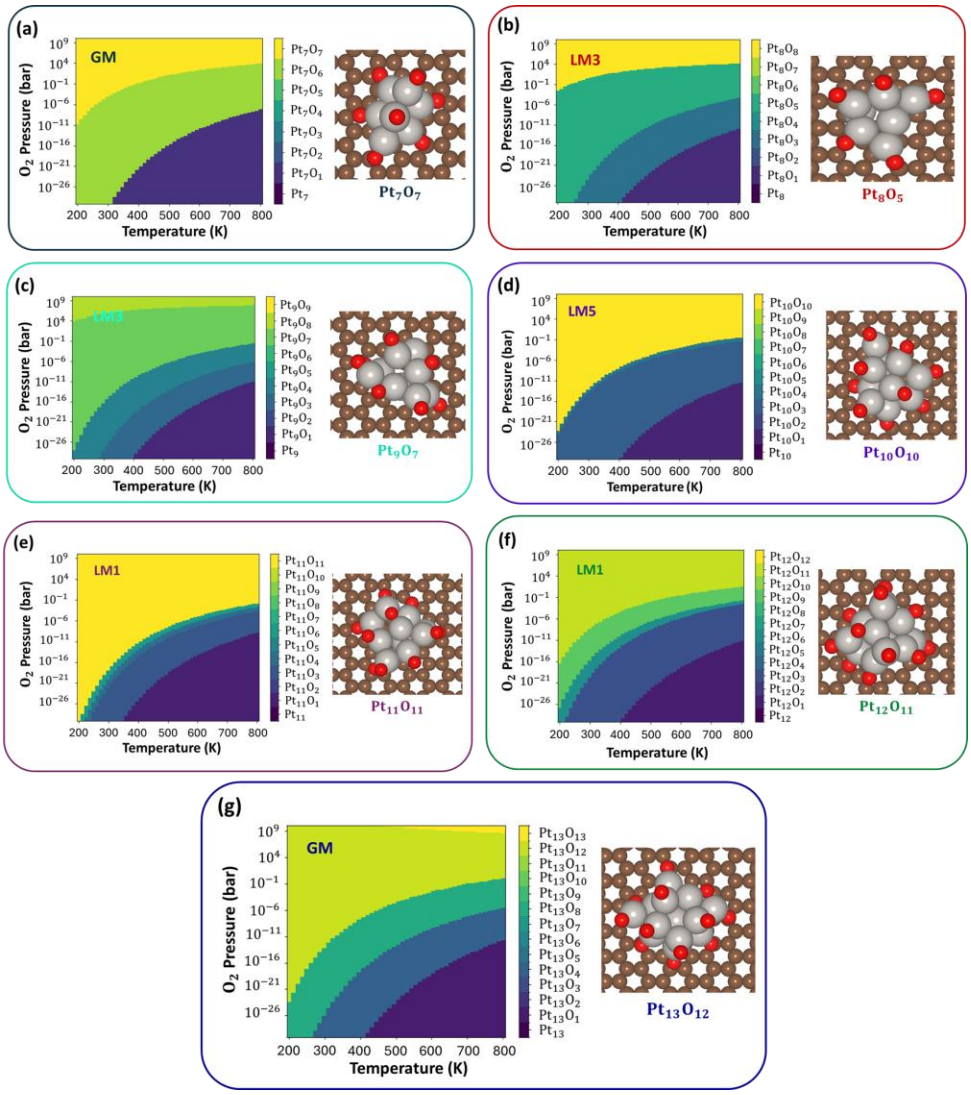


Figure 5.13: *Ab initio* thermodynamic analysis. Phase diagrams for active isomers illustrating the evolution of stable *O coverage configurations. Panels (a-g) correspond to Pt_7O_x (x = 1-7), Pt_8O_x (x = 1-8), Pt_9O_x (x = 1-

9), $Pt_{10}O_x$ (x = 1-10), $Pt_{11}O_x$ (x = 1-11), $Pt_{12}O_x$ (x = 1-12), and $Pt_{13}O_x$ (x = 1-13). For each cluster size, the most stable high-coverage geometries under ambient ORR conditions are shown alongside the corresponding phase diagram.

All isomers undergo significant geometric transformations at the core under high *O coverage, indicating fluxionality under reaction conditions. Unlike the recent report by Zhang *et al.*, the Pt_nO_x clusters in each size regime remain intact *[39]*, without oxygen insertion into the core or sintering of Pt-O motifs, highlighting their stability. At higher coverage, O intermediates predominantly occupy top sites in each configuration, in contrast to the more diverse adsorption sites observed at lower coverage. Under ambient ORR conditions (T = 300 K, P_{O_2} = 1 bar), the isomers tend to form a complete oxygen monolayer, suggesting a strong oxidation tendency. The extent of this coverage may vary with experimental parameters such as electrolyte concentration and pH, potentially leading to even more pronounced structural changes in the subnano clusters. Overall, our results demonstrate that increasing oxygen coverage, coupled with cluster restructuring, drives adsorbate-dependent fluxionality across different sizes and provides a systematic framework to model these complexities.

5.3.7 Ensemble Level Understanding of Reaction Networks.

Since the RDS involves *OH → H_2O formation (ΔR_4) at low intermediate coverage for different isomers, it is important to derive their ORR performance at higher *OH coverage. To account for this, we extend our analysis to include the progressive addition of *OH intermediates to the cluster ensemble. The configurations for higher *OH coverages were derived using Yan et al.'s bond model *[40]*, which allows construction of a reaction network by connecting neighboring structures with successive *OH coverage (e.g., Pt_nOH_x and Pt_nOH_{x+1}), as illustrated in Figure 5.14 *[41]*. The pairwise energetics was then calculated, and the reaction energies for each elementary step were plotted, assuming full monolayer formation.

Within this framework, the reaction barriers are assumed to scale linearly with the reaction energies, following the BEP relationship.

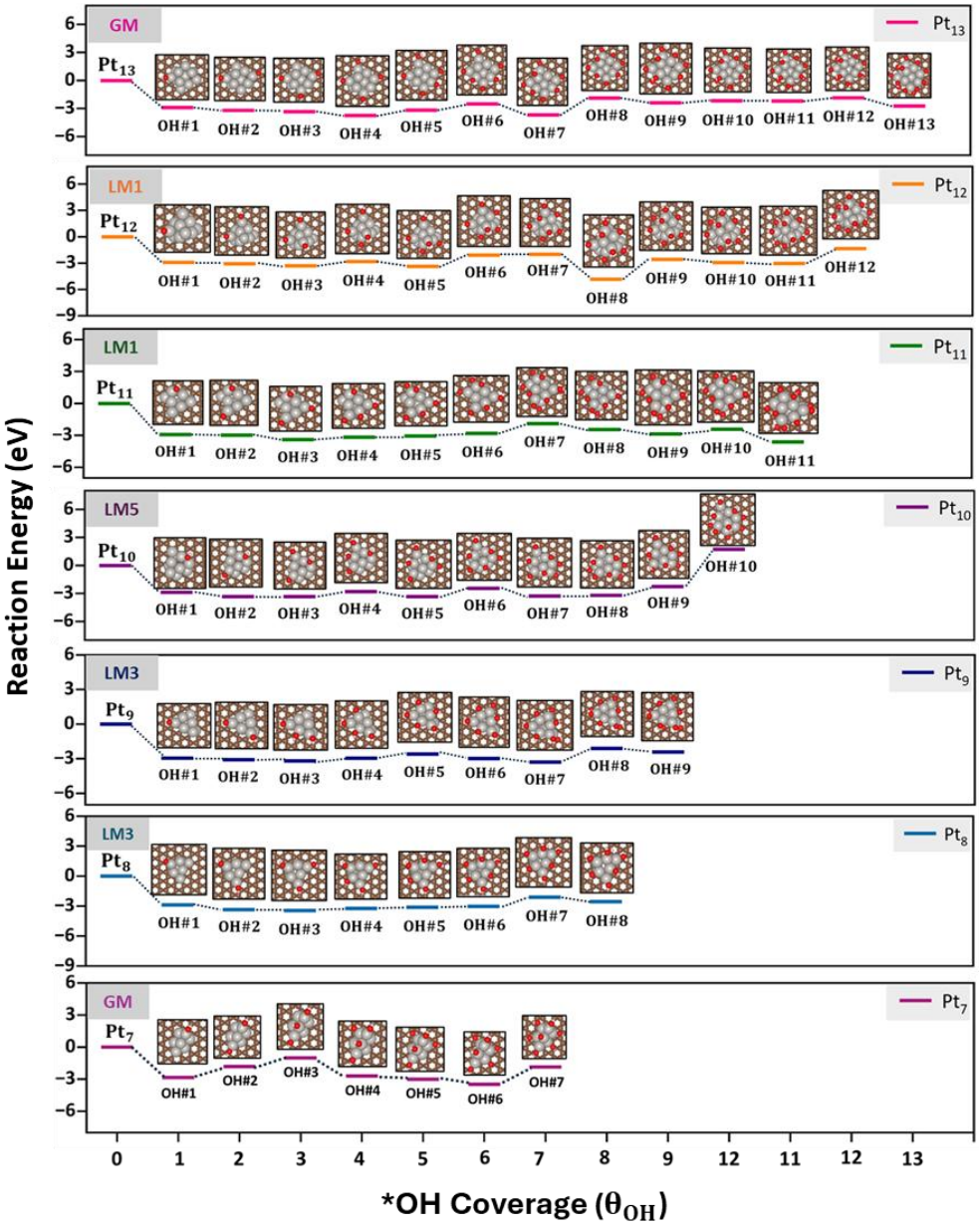


Figure 5.14: Reaction energy diagram at higher *OH coverage (θ_{OH}). Schematics of the hydroxylation process on the active Pt_n/G cluster catalyst leading to complete monolayer *OH coverage: Pt_7OH_x (x = 1-7), Pt_8OH_x (x = 1-8), Pt_9OH_x (x = 1-9), $Pt_{10}OH_x$ (x = 1-10), $Pt_{11}OH_x$ (x = 1-11), $Pt_{12}OH_x$ (x = 1-12), and $Pt_{13}OH_x$ (x = 1-13). For each cluster size, the most stable high-coverage geometries and their reaction networks (e.g.,

Pt_nOH_x and Pt_nOH_{x+1}) are shown alongside the corresponding reaction energy diagram.

Note that for each size regime, adsorption of the first OH#1 is thermodynamically favorable, revealing that the real catalytic interface will be initially covered by *OH. Unlike the irregular reaction energies seen for successive *OH additions on Pt_7/G, the Pt_{8-11}/G isomer show more stable reaction energies, with the most endothermic steps at OH#6 → OH#7, OH#7 → OH#8, OH#9 → OH#10, and OH#6 → OH#7, respectively. For Pt_{10-13}/G, the endothermic step begins with formation of OH#4, OH#4, OH#5, and #OH5, marking the inflexion point corresponding to the onset of saturation, which is often accompanied by structure rearrangements in the clusters and a shift in adsorption from top to bridge sites. Beyond this point, the further *OH addition results in steric hindrance and lateral interactions among neighboring adsorbates, which further hinder stabilization. Consequently, the ORR activity diminishes at high *OH coverages, underscoring the importance of adsorbate-induced fluxionality and coverage effects in determining the catalytic performance of subnano clusters.

5.4 Conclusions

In summary, our combined DFT-ML investigations of Pt_n/G (n = 7-13) identify the geometric factors of supported nanoclusters that govern the breaking of scaling relationships in ORR. The perpendicular distance between the active Pt atom and the graphene surface emerges as the most influential descriptor underlying the observed non-Arrhenius behavior. Furthermore, feature interpretability reveals that Pt_8/G isomers exhibit the highest structural heterogeneity, which introduces noise and reduces model robustness. Using a data-driven framework, we capture the trends in ORR activity from both ML-predicted and DFT-calculated overpotentials within the catalyst ensemble. The trained ML models, based on accessible non-*ab initio* geometric descriptors, achieve reasonable accuracy in predicting *OOH, *O, and *OH adsorption energies (MAE = 0.28, 0.26, and 0.21 eV,

respectively). Thermodynamic analysis indicates that graphene-supported clusters favor complete *O monolayer formation, leading to relatively high overpotentials compared to bare clusters. Overall, this study highlights the structural origins of non-monotonic ORR activity in the subnanometer regime for their fuel cell applications.

5.5. References

1. She, Z. W., Kibsgaard, J., Dickens, C. F., Chorkendorff, I., Nørskov, J. K., Jaramillo, T. F. (2017), Combining theory and experiment in electrocatalysis: Insights into materials design, *Science*, 355, eaad4998 (DOI: 10.1126/science.aad4998)
2. Shi, Z., Li, J., Wang, Y., Liu, S., Zhu, J., Yang, J., Wang, X., Ni, J., Jiang, Z., Zhang, L., Wang, Y., Liu, C., Xing, W., Ge, J. (2023), Customized reaction route for ruthenium oxide towards stabilized water oxidation in high-performance PEM electrolyzers, *Nat. Commun.*, 14, 36380 (DOI: 10.1038/s41467-023-36380-9)
3. Shinde, S. S., Lee, C. H., Jung, J. Y., Wagh, N. K., Kim, S. H., Kim, D. H., Lin, C., Lee, S. U., Lee, J. H. (2019), Unveiling dual-linkage 3D hexa-iminobenzene metal–organic frameworks towards long-lasting reversible Zn–air batteries, *Energy Environ. Sci.*, 12, 727–738 (DOI: 10.1039/c8ee02679c)
4. Shu, Q., Zhang, J., Hu, B., Deng, X., Yuan, J., Ran, R., Zhou, W., Shao, Z. (2022), Rational design of a high-durability Pt-based ORR catalyst supported on Mn/N codoped carbon sheets for PEMFCs, *Energy Fuels*, 36, 1611–1621 (DOI: 10.1021/acs.energyfuels.1c04306)
5. Lavroff, R. H., Morgan, H. W. T., Zhang, Z., Poths, P., Alexandrova, A. N. (2022), Ensemble representation of catalytic interfaces: Soloists, orchestras, and everything in-between, *Chem. Sci.*, 13, 7746–7760 (DOI: 10.1039/d2sc01367c)
6. Zandkarimi, B., Alexandrova, A. N. (2019), Dynamics of subnanometer Pt clusters can break the scaling relationships in

catalysis, *J. Phys. Chem. Lett.*, 10, 460–467 (DOI: 10.1021/acs.jpclett.8b03680)

7. Sun, G., Sautet, P. (2018), Metastable structures in cluster catalysis from first principles: Structural ensemble in reaction conditions and metastability-triggered reactivity, *J. Am. Chem. Soc.*, 140, 2816–2827 (DOI: 10.1021/jacs.7b11239)
8. Zhang, Z., Zandkarimi, B., Munarriz, J., Dickerson, C. E., Alexandrova, A. N. (2022), Fluxionality of subnano clusters reshapes the activity volcano of electrocatalysis, *ChemCatChem*, 14, e202200345 (DOI: 10.1002/cctc.202200345)
9. Mai, H., Le, T. C., Chen, D., Winkler, D. A., Caruso, R. A. (2022), Machine learning for electrocatalyst and photocatalyst design and discovery, *Chem. Rev.*, 122, 13478–13540 (DOI: 10.1021/acs.chemrev.2c00061)
10. Ma, S., Liu, Z.-P. (2020), Machine learning for atomic simulation and activity prediction in heterogeneous catalysis: Current status and future, *ACS Catal.*, 10, 13213–13226 (DOI: 10.1021/acscatal.0c03472)
11. Toyao, T., Maeno, Z., Takakusagi, S., Kamachi, T., Takigawa, I., Shimizu, K.-I. (2020), Machine learning for catalysis informatics: Recent applications and prospects, *ACS Catal.*, 10, 2260–2297 (DOI: 10.1021/acscatal.9b04186)
12. Wang, J., Deng, D., Wu, Q., Liu, M., Wang, Y., Jiang, J., Zheng, X., Zheng, H., Bai, Y., Chen, Y., Xiong, X., Lei, Y. (2023), Insight on atomically dispersed Cu catalysts for electrochemical CO_2 reduction, *ACS Nano*, 17, 18172–18183 (DOI: 10.1021/acsnano.3c07307)
13. Roy, D., Mandal, S. C., Pathak, B. (2021), Machine learning-driven high-throughput screening of alloy-based catalysts for selective CO_2 hydrogenation to methanol, *ACS Appl. Mater. Interfaces*, 13, 56783–56794 (DOI: 10.1021/acsami.1c16696)
14. Kresse, G., Hafner, J. (1994), Ab initio molecular-dynamics simulation of the liquid-metal–amorphous-semiconductor transition in

germanium, *Phys. Rev. B*, 49, 14251–14269 (DOI: 10.1103/PhysRevB.49.14251)

15. Perdew, J. P., Chevary, J. A., Vosko, S. H., Jackson, K. A., Pederson, M. R., Singh, D. J., Fiolhais, C. (1992), Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation, *Phys. Rev. B*, 46, 6671–6687 (DOI: 10.1103/PhysRevB.46.6671)
16. Joubert, D. (1999), From ultrasoft pseudopotentials to the projector augmented-wave method, *Phys. Rev. B*, 59, 1758–1775 (DOI: 10.1103/PhysRevB.59.1758)
17. Zhai, H., Alexandrova, A. N. (2016), Ensemble-average representation of Pt clusters under catalytic conditions via GPU-accelerated global optimization, *J. Chem. Theory Comput.*, 12, 6213–6226 (DOI: 10.1021/acs.jctc.6b00994)
18. Xiao, L., Wang, L. (2004), Structures of platinum clusters: Planar or spherical, *J. Phys. Chem. A*, 108, 8605–8614 (DOI: 10.1021/jp0485035)
19. Nair, A. S., Anoop, A., Ahuja, R., Pathak, B. (2022), Relativistic effects in platinum nanocluster catalysis: A statistical ensemble analysis, *J. Phys. Chem. A*, 126, 1357–1367 (DOI: 10.1021/acs.jpca.1c09981)
20. Baxter, E. T., Ha, M. A., Cass, A. C., Alexandrova, A. N., Anderson, S. L. (2017), Ethylene dehydrogenation on Pt_{4-8} clusters on Al_2O_3, *ACS Catal.*, 7, 3322–3335 (DOI: 10.1021/acscatal.7b00409)
21. Chaves, A. S., Piotrowski, M. J., da Silva, J. L. F. (2017), Evolution of structural and electronic properties of transition-metal clusters, *Phys. Chem. Chem. Phys.*, 19, 15484–15502 (DOI: 10.1039/c7cp02240a)
22. Bord, J., Kirchhoff, B., Baldofski, M., Jung, C., Jacob, T. (2023), Atomistic view of platinum cluster growth on pristine and defective graphene, *Small*, 19, 2207484 (DOI: 10.1002/smll.202207484)
23. Ferrin, P., Simonetti, D., Kandoi, S., Kunkes, E., Dumesic, J. A., Nørskov, J. K., Mavrikakis, M. (2009), Modeling ethanol

decomposition on transition metals, *J. Am. Chem. Soc.*, 131, 5809–5815 (DOI: 10.1021/ja8099322)

24. Nair, A. S., Pathak, B. (2019), Computational screening of ORR activity of 3d transition-metal-based core–shell clusters, *J. Phys. Chem. C*, 123, 3582–3591 (DOI: 10.1021/acs.jpcc.8b11483)
25. Ooka, H., Huang, J., Exner, K. S. (2021), The Sabatier principle in electrocatalysis: Basics, limitations, and extensions, *Front. Energy Res.*, 9, 654460 (DOI: 10.3389/fenrg.2021.654460)
26. Nair, A. S., Mahata, A., Pathak, B. (2018), Multilayered platinum nanotube for ORR in fuel-cell cathodes, *ACS Appl. Energy Mater.*, 1, 2259–2268 (DOI: 10.1021/acsaem.8b00641)
27. Sharma, R. K., Minhas, H., Pathak, B. (2024), Metastability-triggered reactivity of $Pt_{7–8}$ clusters on graphene for ORR, *J. Phys. Chem. C*, 128, 8432–8443 (DOI: 10.1021/acs.jpcc.4c00376)
28. Poths, P., Li, G., Masubuchi, T., Morgan, H. W. T., Zhang, Z., Alexandrova, A. N., Anderson, S. L. (2023), Self-limiting poisoning in ultra-stable sub-nano cluster catalysts, *ACS Catal.*, 13, 958–969 (DOI: 10.1021/acscatal.2c05634)
29. Sharma, R. K., Nair, A. S., Bharadwaj, N., Roy, D., Pathak, B. (2023), Role of fluxionality and metastable isomers in ORR activity, *J. Phys. Chem. C*, 127, 257–268 (DOI: 10.1021/acs.jpcc.2c06265)
30. Zandkarimi, B., Alexandrova, A. N. (2020), Can fluxionality alone cause non-Arrhenius behavior, *J. Phys. Chem. C*, 124, 19856–19865 (DOI: 10.1021/acs.jpcc.0c04136)
31. Li, B., Zhang, P., Liang, S., Ren, G. (2008), Feature extraction and selection for fault diagnosis using wavelet entropy, *Proc. ICSP*, 2008, 2186–2190 (DOI: 10.1109/ICOSP.2008.4697740)
32. Wang, N. N., Dong, J., Deng, Y. H., Zhu, M. F., Wen, M., Yao, Z. J., Lu, A. P., Wang, J. B., Cao, D. S. (2016), Prediction of Caco-2 permeability using boosting and NSGA-II, *J. Chem. Inf. Model.*, 56, 763–773 (DOI: 10.1021/acs.jcim.5b00642)

33. Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011), Scikit-learn: Machine learning in Python, *J. Mach. Learn. Res.*, 12, 2825–2830
34. Lundberg, S. M., Erion, G., Chen, H., et al. (2020), From local explanations to global understanding with explainable AI, *Nat. Mach. Intell.*, 2, 56–67 (DOI: 10.1038/s42256-019-0138-9)
35. Nørskov, J. K., Rossmeisl, J., Logadottir, A., et al. (2004), Origin of the overpotential for oxygen reduction at fuel-cell cathodes, *J. Phys. Chem. B*, 108, 17886–17892 (DOI: 10.1021/jp047349j)
36. Kulkarni, A., Siahrostami, S., Patel, A., Nørskov, J. K. (2018), Understanding catalytic activity trends in ORR, *Chem. Rev.*, 118, 2302–2312 (DOI: 10.1021/acs.chemrev.7b00488)
37. Taleblou, M., Camellone, M. F., Fabris, S., Piccinin, S. (2022), Oxidation of gas-phase and supported Pt nanoclusters, *J. Phys. Chem. C*, 126, 11129–11140 (DOI: 10.1021/acs.jpcc.2c02176)
38. Lym, J., Wittreich, G. R., Vlachos, D. G. (2020), PMuTT: A Python multiscale thermochemistry toolbox, *Comput. Phys. Commun.*, 247, 106864 (DOI: 10.1016/j.cpc.2019.106864)
39. Zhang, Z., Masubuchi, T., Sautet, P., Anderson, S. L., Alexandrova, A. N. (2023), Hydrogen evolution on electrode-supported Pt_n clusters, *Angew. Chem. Int. Ed.*, 62, e202218210 (DOI: 10.1002/anie.202218210)
40. Dean, J., Taylor, M. G., Mpourmpakis, G. (2019), Unfolding adsorption on metal nanoparticles, *Sci. Adv.*, 5, eaax5101 (DOI: 10.1126/sciadv.aax5101)
41. Munarriz, J., Zhang, Z., Sautet, P., Alexandrova, A. N. (2022), Graphite-supported Pt_n cluster electrocatalysts under applied potential, *ACS Catal.*, 12, 14954–14966 (DOI: 10.1021/acscatal.2c04643)

CHAPTER 6

Unraveling Scaling Relationships in Dual-Atom Catalysts

CHAPTER 6

6.1 Introduction

Electrochemical technologies such as fuel cells, water splitting, metal-air batteries, and CO_2 conversion offers great promise for clean energy conversion and storage, driving progress toward sustainable and efficient energy solutions *[1,2]*. Their performance, however, strongly depends on the intrinsic properties and structural design of the electrocatalysts. Therefore, achieving well-defined active sites has long been a key objective in heterogeneous catalysis, yet the structural complexity and dynamic nature of catalysts make this goal challenging *[3]*. The development of single-atom catalysts (SACs), where isolated metal atoms are anchored on conductive supports, has enabled precise control over the active sites for various electrochemical reactions *[4]*. These range from single-intermediate reactions, such as the hydrogen evolution reaction (HER), to multi-intermediate reactions such as the oxygen evolution reaction (OER), oxygen reduction reaction (ORR). This performance is attributed to key advantages of SACs, including their unsaturated coordination environment, maximal metal utilization and relatively uniform active sites. However, when each active sites behaves as a single isolated center, it enforces a fixed adsorption mode. This makes the adsorption energies of multi-step intermediates strongly correlated at that site, leading to scaling relationships that restrict design flexibility and limit further improvements in catalytic activity *[5,6]*. Dual-atom catalysts (DACs) have emerged as another promising class, offering enhanced reactivity through synergistic interactions between paired metal atoms *[7]*. DACs retain the key benefits of SACs while offering added flexibility. The presence of two neighbouring metal atoms introduces synergistic effects, giving dual-metal sites the ability to adjust and optimize the adsorption energies of specific intermediates involved in multi-step reactions *[8]*.

The composition of DACs gives rise to two main features that define their distinct catalytic behaviour. First, the dopant atoms form the active sites, while the more inert host surface can serve as a desorption site for reaction

products. In this configuration, the dopant governs the catalytic activity, and the host helps suppress side reactions by stabilising intermediates or facilitating product removal once they spill over from the active site. This spatial and functional separation enhances selectivity by promoting product desorption and minimising undesired reactions *[9]*. The second key feature is the free-atom-like nature of the dopant's d-states, which can strongly interact with the molecular orbitals of reactants when their energy level aligns, a concept that has been widely discussed *[10]*. More specifically, the d-band centre relative to the Fermi-level has been widely used to explain adsorption trends on pure transition metals and alloys, serving as a key indicator of their electrocatalytic activity *[11]*. However, given the vast number of possible homometallic and heterometallic pairs that can be embedded in the host surface, *in silico* analysis of the d-states for each catalyst becomes prohibitively expensive. This challenge underscores the necessity for developing a data-driven approach that incorporates these key descriptors as features to effectively capture the structure-activity relationship within the local coordination environment of DACs.

Recently, machine learning (ML) has played a central role in optimizing the electrocatalysts performance, while capturing the interactions between catalysts and reaction intermediates to reveal the complex structure-activity relationships *[12-15]*. However, a major challenge is designing ML models with descriptor variables that enhance both predictive power and interpretability *[16,17]*. To address this challenge, quantitative structure-property relationship (QSPR) models have been developed to correlate the structural properties of catalysts with their performance and to identify key parameters such as electronic structure, physicochemical properties, and adsorption energies *[18]*. These correlations offer a pathway for understanding catalytic behavior and guiding the rational design of efficient catalysts. While explicit QSAR equations are limited, ML methods can uncover governing factors more effectively by leveraging large datasets and advanced algorithms, showing strong capability in identifying the

determinants of catalytic efficiency. However, it remains challenging to design ML models encoded with descriptors that accurately capture both adsorption behavior and the catalytic origins of the multistep OER and ORR reactions on transition-metal-based DACs. Most importantly, incorporating d-band characteristics of intermediate adsorption into ML models still depends heavily on DFT calculations, highlighting the need for electronic features that can capture these characteristics reliably without relying on DFT.

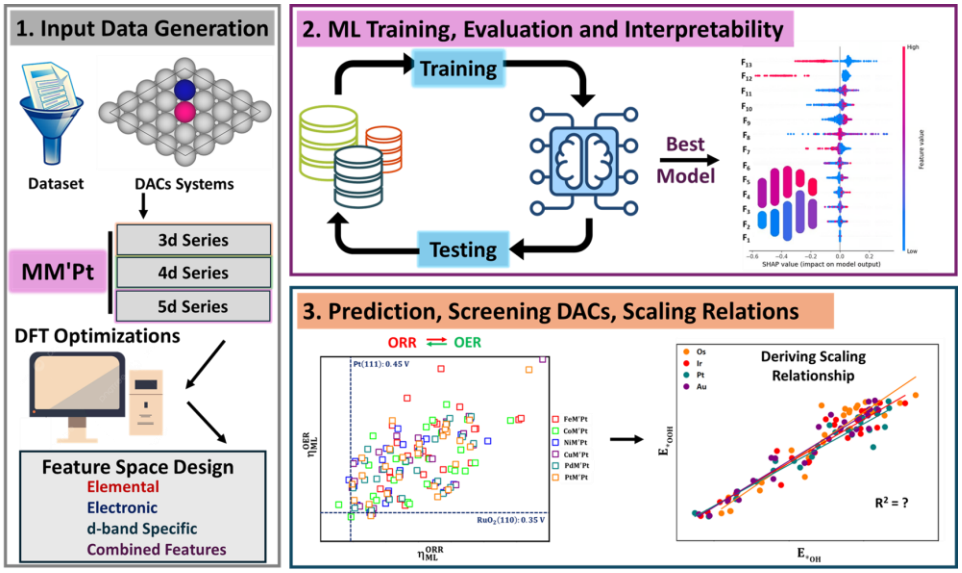


Figure 6.1: Workflow of the machine learning framework used to screen active dual-atom catalysts (DAC) electrocatalysts for the OER/ORR. The process involves: (1) building an adsorption energy database for *OOH, *O, and *OH intermediates on different transition-metal DACs and designing features to capture their local chemical environment, (2) evaluating and interpreting ML models, and (3) predicting, screening, and establishing scaling relationships among key intermediates.

In view of the myriad possibilities, we developed a ML framework to screen late transition-metal-based heterometallic dimers embedded in a Pt (111) surface for OER and ORR activity (Figure 6.1). By encoding the d-band model using elemental numerical values from the solid-state table for the embedded dimers, we aimed to uncover structure-activity relationships and capture the non-monotonic trends observed for both OER and ORR in these

DACs. Using model interpretability technique, we identified the key descriptors that govern the adsorption of reaction intermediates at the active sites. We also examined the scaling behavior between the principal intermediates and emphasized the need for a multicomponent descriptor that better reflects overall reactivity. By validating the ML-predictions with DFT-calculations, we identified bimetallic compositions with optimal OER and ORR activity and established the Sabatier range for adsorption energies linked to the highest performance. The insights from this study support the rational screening of dual-atom catalysts for fuel-cell applications and point to promising directions for developing improved descriptors for ML models in heterogeneous catalysis, guided by scaling relations and Sabatier principles.

6.2 Computational Details

6.2.1 Electronic Calculations: All plane-wave spin-polarized density functional theory (DFT) calculations were performed using the Vienna *ab initio* simulation package (VASP) *[19]* with the Perdew-Burke-Ernzerhof (PBE) exchange-correlation functional based on the generalized gradient approximation (GGA) *[20]*. The interactions between ion cores and valence electrons were treated using the projector augmented wave (PAW) method. During structural relaxations, a kinetic energy cut-off of 500 eV was used, and convergence was achieved when the total energy difference between interactions was below 10^{-4} eV and atomic forces were reduced to less than 0.02 eV/Å. Gaussian smearing with a width of 0.2 eV was applied, and Γ-centered k-point grids of (1×1×1) were adopted for Brillouin zone sampling, consistent with earlier studies. A higher (3 × 3 × 3) *k*-mesh was used to calculate the partial density of state. Bader atomic charges *[21]* were determined using the Henkelman code with the near-grid algorithm refine edge method. All the isosurface values for the charge density difference analysis were set to be 0.0025 eV $Å^{-3}$, where pink and green colors represent the charge depletion and accumulation zones, respectively.

6.2.2 Surface Charging Method: To simulate the electrode-electrolyte interface under varying applied potentials, we employed the Surface Charging (SC) method *[22]*. Within the computational hydrogen electrode (CHE) framework *[23]*, the applied potential influences only the chemical potential of exchanged electrons, while the electronic energy remains independent of the potential and is taken from the neutral system. However, the free energies of intermediates in the neutral case differ from those at U = 0.0 V. To capture the potential-dependent reaction energies of the best OER/ORR electrocatalysts across the range -1.2 to 1.2 V vs SHE, we used the SC approach. This model is based on the linearized Poisson-Boltzmann (PB) equation and involves several key approximations:

1. Constant surface charge density: Assumes the surface charge remains constant, which may not hold under large charge transfers in the Stern layer.
2. Constant electrolyte concentration: Assumes no variations in electrolyte concentration during charging.
3. Linearized PB equation: Assumes the PB equation can be linearized, which becomes less accurate at high surface potentials.
4. Constant dielectric constant: Assumes the solvent's dielectric constant does not vary with time or position.

The electrode potential relative to the standard hydrogen electrode (SHE) was obtained using:

$$U_{SHE} = \frac{\Phi_F - \Phi_{SHE}}{e} \tag{6.1}$$

where U_{SHE} is the potential versus SHE, Φ_F is the Fermi work function, Φ_{SHE} = 4.3 eV is the absolute potential of the SHE, and e is the elementary charge. The Fermi work function is given by:

$$\Phi_F = E_{vac} - E_F \tag{6.2}$$

where E_{vac} and E_F are the vacuum level and the Fermi energy, respectively. Finally, the relationship between the electrode potential and total energy follows a quadratic form:

$$g(x) = ax^2 + bx + c \quad (6.3)$$

6.3 Results and Discussion

6.3.1 Engineering DACs Chemical Space

In our investigation of active DAC electrocatalysts, we first model the structures using a 4 × 4 × 1 supercell of the Pt (111) surface. The divacancy site was created by substituting two Pt atoms in the surface lattice with two different transition metals (MM′Pt), as shown in Figure 6.2a. To evaluate the effect of each metal site on the catalytic performance separately, DACs composed of two different metal atoms are analysed as two separate instances, where M and M′ are defined as the centre metal atoms and the assistance metal atom, respectively. Based on the high activity and recent synthesis strategies reported for late transition metals *[8]*, we select Fe, Co, Ni, Cu, Pd, and Pt as candidates for M, and screen a range of 3d, 4d, and 5d transition metals for M′. Metals such as Tc, Cd, La, and Hg were excluded due to concerns related to radioactivity or toxicity *[24]*. In total, we extracted 156 electrocatalysts (6×26) electrocatalysts, consisting of 6 homometallic and 150 heterometallic DACs, for subsequent electrocatalytic investigations. All these structures were optimized using plane-wave spin-polarised DFT calculations in Vienna *ab initio* simulation package (VASP) with the Perdew-Burke-Ernzerhof (PBE) functional within the generalized gradient approximation (GGA) *[19,20]*. Note that during geometric optimization, the dopant atoms in all DACs systems remained in-plane without significant bulging or segregation, confirming the structural stability of the investigated surfaces.

6.3.2 Adsorption Energy Landscape on DACs

According to the Bell-Evans-Polanyi (BEP) principle, the adsorption energy (E_{ads}) of key intermediates is a crucial descriptor for evaluating catalytic performance *[25,26]*. Therefore, to accurately investigate the catalytic performance of the active sites on the catalyst surface, E_{ads} is selected as the target variables in this work to derive the ORR and OER mechanism (Figure 6.2b). To generate our initial database, we sampled

single-intermediate configurations across the DACs by adsorbing intermediates on different M′ sites while keeping M fixed. The E_{ads} of the *OOH, *O, and *OH intermediates (hereon represented as $E_{*_{OOH}}$, E_{*_O} and $E_{*_{OH}}$) at the M′ site of the DACs is computed as:

$$E_{ads} = E_{MM'Pt(*_{OOH},*_O,*_{OH})} - (E_{MM'Pt} + E_{*_{OOH},*_O,*_{OH}}) \quad (6.4)$$

where $E_{MM'Pt(*_{OOH},*_O,*_{OH})}$, $E_{MM'Pt}$, and $E_{*_{OOH},*_O,*_{OH}}$ represent the total energies of the DACs with *OOH, *O, and *OH at site M′, MM′Pt, and the corresponding intermediates in the gas phase, respectively.

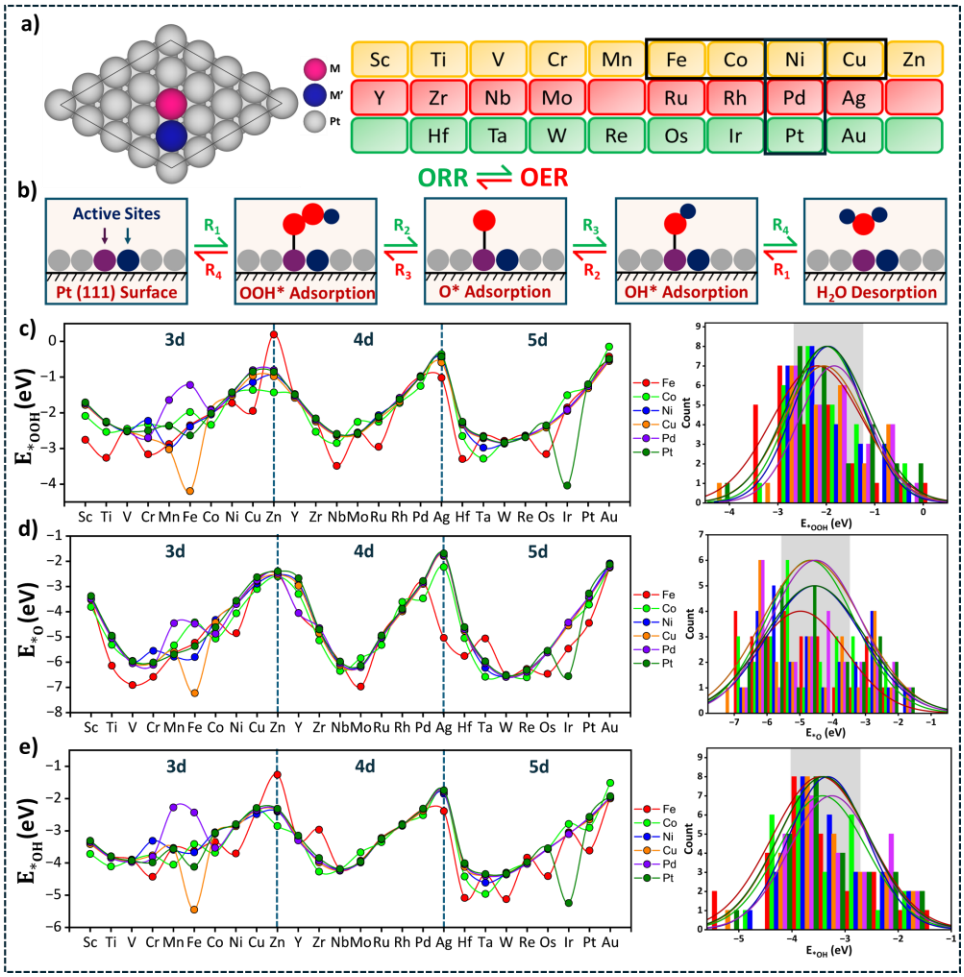


Figure 6.2: Input Dataset Curation for ML Model Evaluation. (a) Schematic of the heterometallic dual-atom catalyst (DAC) systems studied on the Pt (111) surface, featuring different transition metals (TMs) at the active sites. A cropped periodic table highlights the selected elements, where M atoms (M = Fe, Co, Ni, Cu, Pd, and Pt) are fixed and outlined with

black boxes, while M′ spans across the 3d, 4d, and 5d series. Empty blocks indicate TMs not included in this study. (b) Reaction scheme for ORR and OER demonstrating the end-on adsorption configurations of the key intermediates (*OOH, *O, and *OH) on DAC active sites. (c-e) Adsorption energy (E_{ads}) profiles of *OOH, *O, and *OH across the MM′Pt combinations, with solid lines connecting the markers acting as a visual guide. The alongside density plot illustrates the distribution of adsorption energies across all evaluated DACs, with the peak regions highlighted in grey shade.

The calculated values and trend of E_{ads} for each intermediate across different DACs is summarised in Figure 6.2c-e, with four interesting observations emerging within the diverse distributions of the E_{ads} dataset: (i) the E_{ads} values for each intermediate exhibit a lack of consistent trend, with the density of distributions falling within different energy ranges. The peak density distribution of E_{*OOH}, E_{*O}, E_{*OH} ranges from -1.3 to -2.8 eV, -3.6 to -5.6 eV, and -2.8 to -4.1 eV, respectively. (ii) Compared to the bulk Pt(111) surface with E_{*OOH} = -1.64 eV, E_{*O} = -4.69 eV, and E_{*OH} = -2.63 eV *[27]*, these DACs exhibit a broader range of E_{ads}, with majority showing stronger adsorption towards the key intermediates. (iii) Interestingly, few catalysts exhibit significant negative E_{ads} values (E_{*O}= -7.23 eV for CuFePt), indicating an over-binding of intermediates with active sites, potentially leading to the poisoning of the catalyst. (iv) Transitioning from the 3d to 5d TM series, we predominantly observed a periodic *U-shaped* pattern for each intermediate, with E_{ads} increasing from v^1 to v^{4-5} (metal with 3-6 valence electrons) and then decreasing from v^6 to v^{10}, with a few exceptions. Within this chemical space, we have curated a comprehensive dataset of 156 DACs with three ORR intermediates, resulting in 468 E_{ads} datapoints (156 × 3) for subsequent investigations.

6.3.3 Feature Design Space for DACs

Prior to the construction of ML models, mining effective features that uniquely define the local environment and accurately reveal the structure-

activity relationship is crucial *[12,18]*. To encode the electronic characteristics of catalytic centres influencing the electron transfer during the intermediate adsorption, we classified the extracted features into four different categories: (i) elemental, (ii) electronic, (iii) d-band specific features, and (iv) combined features, as tabulated in Table 6.1. To distinguish between the M and M′ metals, features pertaining to both metals were included. For instance, the atomic number of M and M′ is denoted as Z_1 and Z_2, respectively.

Table 6.1. List of Features Including Elemental, Electronic, d-band Specific Indicators, and Combined Features. Features pertaining to M and M′ elements for DACs Alloys are distinguished by subscripts "1" and "2", respectively.

Category	Feature	M	M'
Elemental	Atomic number	Z_1	Z_2
	Atomic weight	A_1	A_2
	Atomic radii	R_1	R_2
Electronic	Pauling's electronegativity	PE_1	PE_2
	Total valence electrons	V_1	V_2
	Valence d electrons	V_1^d	V_2^d
	Number of unfilled d electrons	U_1^d	U_2^d
	First ionization potential	IP_1	IP_2
	Sum of polarizability	SP_1	SP_2
d-band Specific	Idealized d-band filling	Idf_1	Idf_2
	d-band center	ϵ_1	ϵ_2
	Coupling matrix	V_1^{ad}	V_2^{ad}
Combined Features	Average atomic number	$\overline{Z}$	
	Average atomic weight	$\overline{A}$	
	Average atomic radii	$\overline{R}$	
	Average Pauling's electronegativity	$\overline{PE}$	
	Average total valence electrons	$\overline{V}$	
	Average valence d electrons	$\overline{V^d}$	
	Average number of unfilled d electrons	$\overline{U^d}$	
	Average first ionization potential	$\overline{IP}$	
	Average sum of polarizability	$\overline{SP}$	

Elemental features describe the physical properties of transition metals involved in adsorption, whereas the electronic features reflect the donation and acceptance capabilities of the active sites. Considering the significant

role of the electronic density distributions of d-states on TM site atoms, Hammer and Nørskov introduced a physics-inspired d-band theory to conceptualize the adsorption of transition-metal systems *[28]*. According to this theory, the contribution of metal sp states is approximated as constant, while the variations in E_{ads} are predominantly governed by the metal d states, thereby establishing a link between the electronic properties of TM and E_{ads}. For this purpose, the most common descriptors used are d-band centre (ε_d) and d-band filling (d_f) associated with the metal center. In addition to the ε_d, and d_f, recent studies have also underscored the effectiveness of reactivity descriptors like d-band centre plus half the d bandwidth (ε_d^w), and the maximum of the d-band Hilbert-transform (ε_u), which explicitly accounts for the higher-order characteristics of the d-band moments *[29]*. Although these features have the potential to accelerate the data-driven exploration of catalytic activity of bulk surfaces, their generalizability across the DACs has yet to be explored. Moreover, many prototypical electronic descriptors associated with d-band characteristics require expensive self-consistent quantum calculations.

To address this limitation, we attempted to encode the d-band model using elemental-specific numeric values provided in the solid-state table for surfaces and categorized them as "d-band specific", as shown in Table 6.1 *[30]*. This includes the coupling matrix (V_{ad}^2), idealized d band filling (I_{df}), and d-band center with respect to size ($\overline{\epsilon_d}$), where $\overline{\epsilon_d} = \frac{d\epsilon_d}{d \ln S_{BW}}$, and S_{BW} represents the Bulk Wigner-Seitz radius. Notably, these features govern the inherent characteristics of d-band electronic states and can be easily substituted without resorting to expensive DFT calculations. Additionally, we have incorporated the average features as "combined features" to account for the influence of different TM atoms constituting the dual-active sites. The linear and monotonic correlations between features and target E_{ads} properties were evaluated using the Pearson correlation coefficient (PCC) *[31]*. To minimize redundancy and simplify the input dataset, we retained feature pairs exhibiting |PCC| < 0.9, while ensuring that the

features with the highest contribution to the target variable were retained for the ML analysis. Overall, the E_{ads} of ORR intermediates at the local coordination environment of DACs can be expressed as a function of:

$$E_{TM_n}^{*O,*OH,*OOH} = f\left(IP_2, SP_1, Idf_1, Idf_2, \epsilon_1, \epsilon_2, V_1^{ad}, V_2^{ad}, \overline{R}, \overline{PE}, \overline{U^d}, \overline{IP}\right) \quad (6.5)$$

Ultimately, the final dataset captures the inherent complexity of characterising E_{ads} across diverse DAC sites, highlighting the need for a data-driven framework to address this combinatorial challenge.

6.3.4 Machine Learning Training and Evaluation

Following feature engineering, model training involves identifying suitable parameters for the mathematical formulation of the respective model to minimise the gap between the model's validation results on the specified dataset and real predictions *[12]*. We employed seven different ML algorithms, namely catboost regression (CR), gradient boosting regression (GBR), extra tree regression (ETR), adaboost regression (ABR), random forest regression (RFR), and extreme Gradient Boosting Regression (XGBR), kernel ridge regression (KRR), as available in the Scikit-learn open-source library (Figure 6.3A) *[32]*. The entire dataset was divided into training and testing sets in an 80:20 ratio, and models were optimised utilising the randomised searchCV method and evaluated with standard metrics (MAE/RMSE and R^2 score), with hyperparameters for each model optimized utilizing the RandomizedSearchCV method. Consequently, the best models were employed to predict the E_{*OOH}, E_{*O}, E_{*OH} values for all the electrocatalysts present in our dataset.

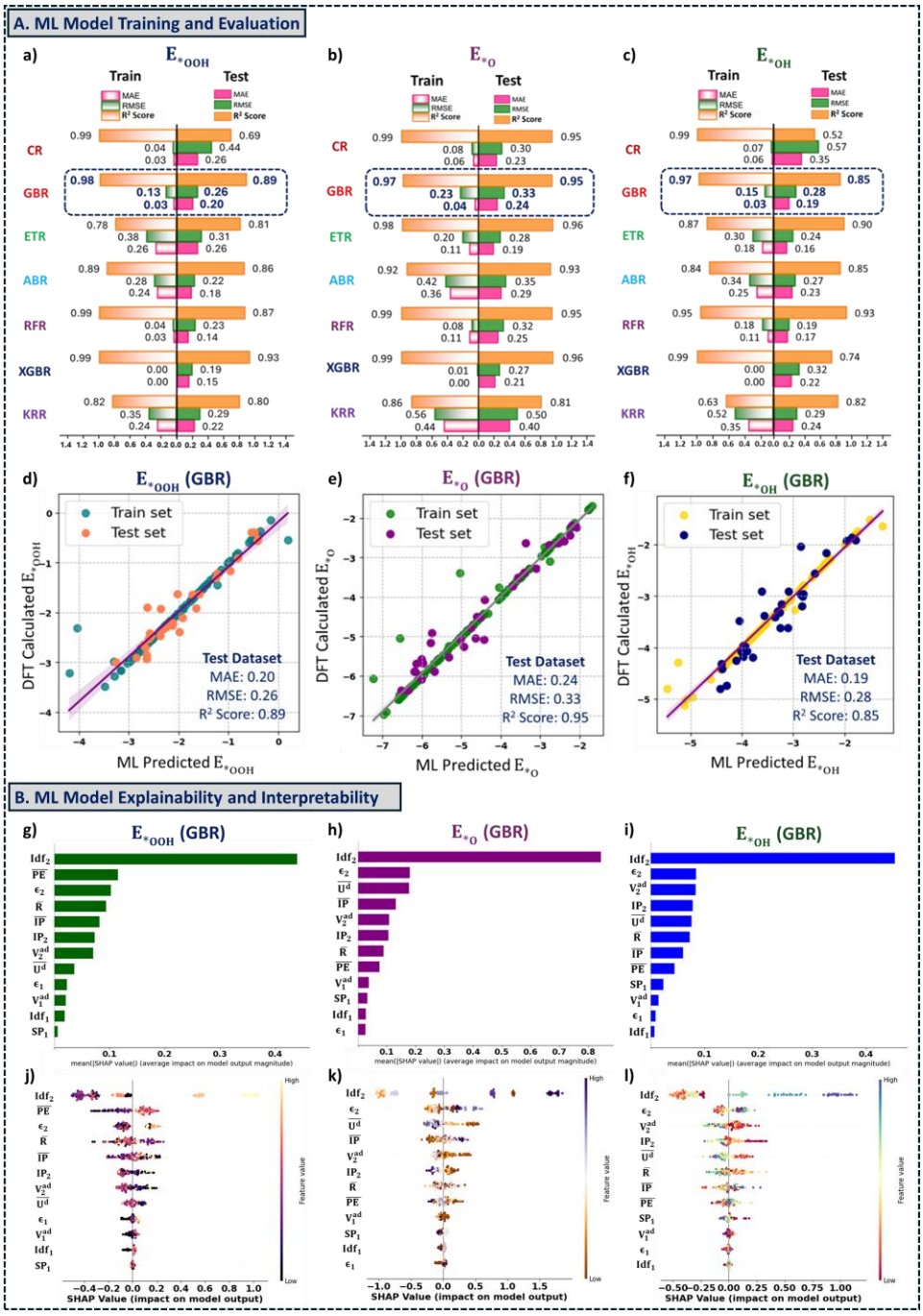


Figure 6.3: (A) ML Models Training and evaluation: Mean absolute error (MAE), root-mean-square error (RMSE), and coefficient of determination (R^2) for CR, GBR, ETR, ABR, RFR, XGBR, and KRR models using (a) E_{*OOH}, (b) E_{*O}, and (c) E_{*OH} datasets (in eV). The best-performing models for each intermediate are outlined with a rectangular dotted box, with their performance metrics marked in bold. Parity plots comparing DFT-

calculated and ML-predicted (d) E_{*OOH}, (e) E_{*O}, and (f) E_{*OH} values are shown for the best-performing GBR model. (B) ML Models Explainability and Interpretability: Global interpretation of the best-performing models for extracted features in predicting (g) E_{*OOH}, (h) E_{*O}, and (i) E_{*OH}. Beeswarm plot that displays an information-dense summary of how the most important features affect the prediction of (j) E_{*OOH}, (k) E_{*O}, (l) and E_{*OH}. A solitary dot on the feature row represents each datapoint with its SHAP value on the x-axis. The color gradient along the y-axis represents the feature values, ranging from low to high.

Figure 6.3a-c represents the final test scores of various ML models for each intermediate after hyperparameter tuning. Among them, the GBR model achieved the best performance in predicting E_{*OOH}, E_{*O}, E_{*OH}, with low MAE: RMSE values of 0.20: 0.26 eV, 0.24: 0.33 eV and 0.19: 0.28 eV along with high R^2 scores of 0.89, 0.95, and 0.85 for test datasets, respectively. Furthermore, the parity plots (Figure 6.3d-f) show a linear correlation between DFT-calculated and ML-predicted E_{ads} values, demonstrating the prediction accuracy of the best-performing models for different intermediates across both training and test datasets. Similarly, the GBR model was chosen to predict E_{*OOH}, E_{*O}, and E_{*OH} values to screen the most effective catalysts in our further investigations.

6.3.5 Machine Learning Explainability and Interpretability

ML algorithms capture complex relationships through ensembles of decision trees, but this often makes them “black boxes”, limiting their interpretability. To address this, we employed explainable AI using Shapley Additive exPlanations (SHAP) analysis *[33]*, a game theory-based approach that interprets ML model outputs by assigning each feature a value representing its contribution to a specific prediction. This analysis reveals how individual features influence the prediction of E_{*OOH}, E_{*O}, E_{*OH}, as represented in Figure 6.3B, providing both local and global insights into feature importance.

In SHAP analysis, the features I_{df_2} and ϵ_2 consistently ranked among the top three in the bar plots, highlighting the strong influence of d-states on intermediate adsorption at DACs (Figure 6.3g-i). For each intermediate, higher values of I_{df} showed a positive correlation with E_{ads} (Figure 6.3j-l), and the same trend was observed for ϵ_2. In the solid-state table, the I_{df_2} and ϵ_2 vary in tandem, with higher positive values observed from left to right across the series *[30]*. These trends are consistent with d-band theory, where increase in I_{df_2} and ϵ_2 lead to repulsive interactions and weaker metal-adsorbate coupling, thereby lowering E_{ads}. Additional key features emerged for specific intermediates: $\overline{PE}$ for E_{*OOH}, $\overline{U^d}$ for E_{*O}, and V_2^{ad} for E_{*OH}. Low $\overline{PE}$ values drive E_{*OOH} towards negative ranges, while low $\overline{U^d}$ and V_2^{ad} shifted E_{ads} toward positive ranges for E_{*O} and E_{*OH}, indicating their inverse relationship with adsorption strength. Overall, these results suggest that both d-states of TMs primarily govern the adsorption behavior of intermediates within the local chemical environment of DACs.

6.3.6 Screening and Validating Active Electrocatalysts

To validate the accuracy of our best-performing GBR model, we predicted the E_{*OOH}, E_{*O}, and E_{*OH} values for all catalysts included in our dataset. To identify the active electrocatalysts, we combined the DFT-calculated and ML-predicted E_{*OOH}, E_{*O}, and E_{*OH} values and derived an ORR mechanism via the adsorbate evolution mechanism (AEM) for comparison. The reaction energies were then evaluated using the computational hydrogen electrode (CHE) approach introduced by Nørskov and coworkers *[23]*, at both 0 V and 1.23 V, based on the following elementary steps:

In an acidic medium, the OER is considered a four-step process as follows:

$$* + H_2O\,(l) \rightarrow OH^* + (H^+ + e^-) \qquad (\Delta \mathbf{R_1})$$

$$OH^* \rightarrow O^* + (H^+ + e^-) \qquad (\Delta \mathbf{R_2})$$

$$O^* + H_2O\,(l) \rightarrow OOH^* + (H^+ + e^-) \qquad (\Delta \mathbf{R_3})$$

$$OOH^* \rightarrow O_2(g) + (H^+ + e^-) + * \qquad (\Delta \mathbf{R_4})$$

And the 4e$^-$ ORR is the reverse reaction of the OER, with a four-step process as follows:

$$O_2(g) + (H^+ + e^-) + * \rightarrow OOH^* \quad \mathbf{(\Delta R_5)}$$

$$OOH^* + (H^+ + e^-) \rightarrow O^* + H_2O\ (l) \quad \mathbf{(\Delta R_6)}$$

$$O^* + (H^+ + e^-) \rightarrow OH^* \quad \mathbf{(\Delta R_7)}$$

$$OH^* + (H^+ + e^-) \rightarrow H_2O\ (l) + * \quad \mathbf{(\Delta R_8)}$$

The catalytic activity of the DACs was evaluated by calculating the overpotential values (η) of the rate-determining step (RDS) at 1.23 V, using the equation:

$$\eta_{OER}^{ML/DFT} = \frac{\max(\Delta R_1, \Delta R_2, \Delta R_3, \Delta R_4)}{e} \quad (6.6)$$

$$\eta_{ORR}^{ML/DFT} = \frac{\max(\Delta R_5, \Delta R_6, \Delta R_7, \Delta R_8)}{e} \quad (6.7)$$

In fuel cells, OER corresponds to the charging process, while ORR is linked to the discharging process. As the state-of-the-art catalytic structures, the IrO_2 (110) and Pt(111) surfaces were chosen as benchmarks for OER and ORR, with η_{OER}= 0.35, and η_{ORR}= 0.45, respectively.

Figure 6.4A illustrates the trends of ML-predicted and DFT-calculated OER and ORR activities for the DACs alloys, obtained from the best-performing GBR models. For OER, CoPdPt, CoCuPt, and PtRhPt were predicted as the most active electrocatalysts with a low overpotential of η_{ML}^{OER}: η_{DFT}^{OER} = 0.37:0.36 V, η_{ML}^{OER}: η_{DFT}^{OER} = 0.38:0.55 V, and η_{ML}^{OER}: η_{DFT}^{OER} = 0.38:0.38 V, with RDS involving *OH → *O, *OH → *O, and *O → *OOH, respectively *[23,24]*. These η_{ML}^{OER} values are comparable to that of RuO_2(110) (η_{OER}= 0.35 V). In contrast, FeVPt, CoRePt, and FeNbPt were predicted as the worst OER electrocatalysts with higher overpotential for RDS as follows: η_{ML}^{OER}: η_{DFT}^{OER} = 2.57:2.57 V, η_{ML}^{OER}: η_{DFT}^{OER} = 2.04:2.04 V, and η_{ML}^{OER}: η_{DFT}^{OER} = 2.00:2.00 V, with RDS involving *O → *OOH, *O → *OOH, and *OOH → O_2(g), respectively.

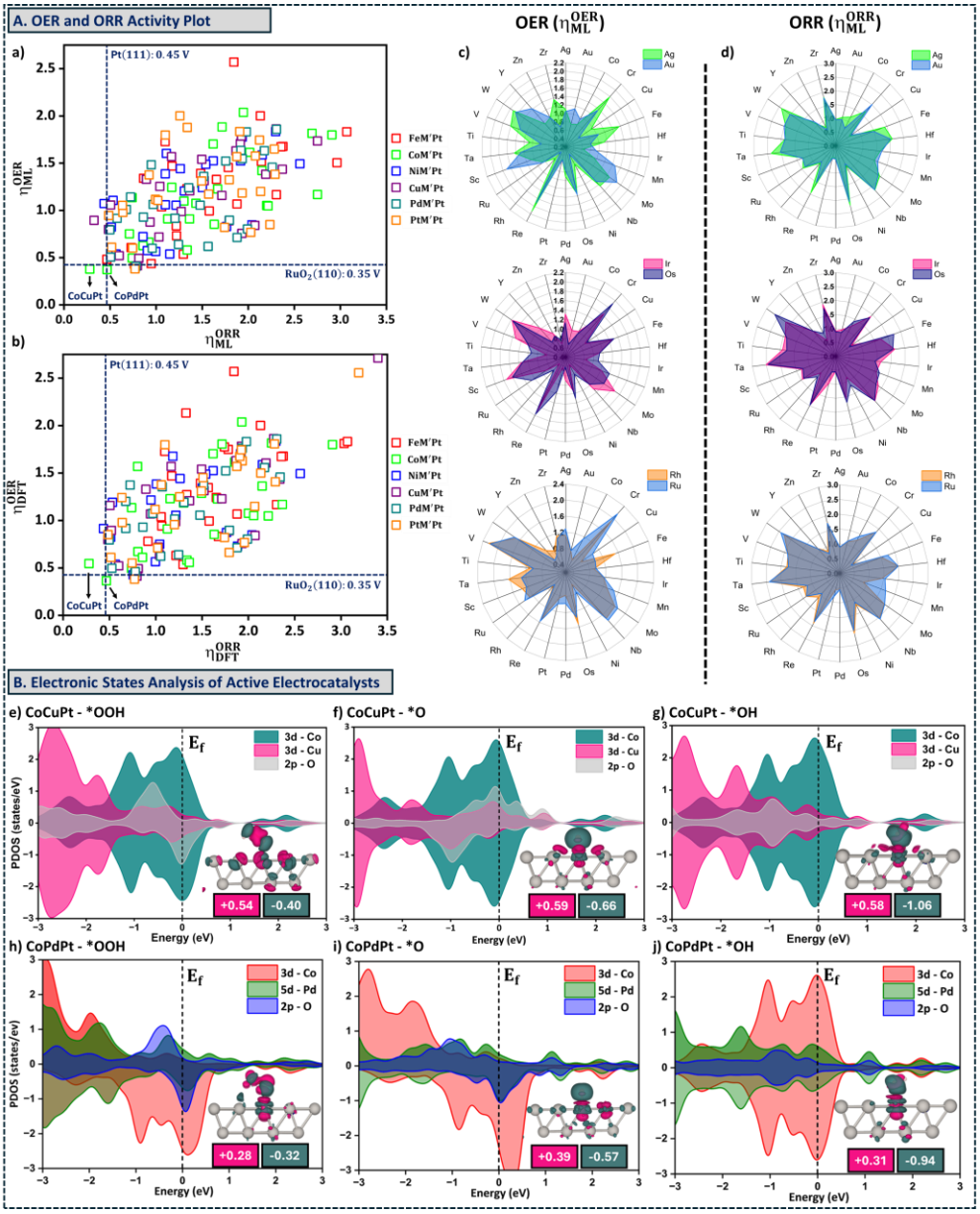


Figure 6.4: (A) OER and ORR activity plot. Distribution of overpotentials for OER and ORR activity derived from (a) ML-predicted (η_{ML}^{OER}, η_{ML}^{ORR}), and (b) DFT-calculated (η_{DFT}^{OER}, η_{DFT}^{ORR}) screening. The intersection of the horizontal and vertical dashed lines indicates the low-overpotential regions, serving as a reference point to identify active bifunctional electrocatalysts from the ML and DFT dataset. Radar plots represent the predicted (c) η_{ML}^{OER} and (d) η_{ML}^{ORR} values from the trained GBR model for late-transition metal-based DACs (M = Ag, Au, Ir, Os, Rh, and Ru) included as the unseen

dataset. (B) Electronic States Analysis. Partial density of states (PDOS) for key OER/ORR intermediates of (e-g) CoCuPt and (h-j) CoPdPt, with Fermi level represented by E_f. Insets represent the charge density difference (CDD) plots of the cropped surface motifs of the active electrocatalysts, where pink and green denote the regions of charge accumulation and depletion, respectively (isosurface = $0.0025\ eVÅ^{-3}$). The numeric values correspond to the charge transfer estimated from Bader charge analysis, reported in |e| units.

For ORR, CoCuPt, CuCuPt, and NiCuPt were predicted as the active electrocatalysts with lower overpotential values for RDS as follows: η_{ML}^{ORR}: η_{DFT}^{ORR} = 0.28:0.28 V, η_{ML}^{ORR}: η_{DFT}^{ORR} = 0.33:0.52 V, and η_{ML}^{ORR}: η_{DFT}^{ORR} = 0.44:0.44 V, with RDS involving *OH → H_2O(l), O_2(g) → *OOH, and *OH → H_2O(l), respectively. These η_{ML}^{ORR} values are lower compared to Pt (111) surface (η_{ORR}= 0.45 V), representing high ORR activity of these DACs for ORR process. In contrast, FeWPt, FeHfPt, and CoTaPt were predicted as the worst OER electrocatalysts with higher overpotential for RDS as follows: η_{ML}^{ORR}: η_{DFT}^{ORR} = 3.07:3.07 V, η_{ML}^{ORR}: η_{DFT}^{ORR} = 2.96:3.03 V, and η_{ML}^{ORR}: η_{DFT}^{ORR} = 2.91:2.91 V, all with RDS involving *OH → H_2O(l).

For efficient OER and ORR in fuel cells, the most effective catalysts are located near the apex of the activity volcano, characterised by E_{ads} close to the optimum for each reaction intermediate *[35]*. The DFT-calculated and ML-predicted E_{*O}, E_{*OH}, and E_{*OOH} for the best OER and ORR catalysts show strong agreement, validating the ability of our trained GBR model to capture the non-monotonic activity trend across DACs (Figure 6.5). Importantly, E_{ads} for these active catalysts fall within the optimal adsorption energy range (-4.92 eV < E_{ads} < -0.92 eV), where DACs achieve the best balance in adsorption strength and outperform benchmark surfaces such as RuO_2(110) and Pt(111) *[23,24]*.

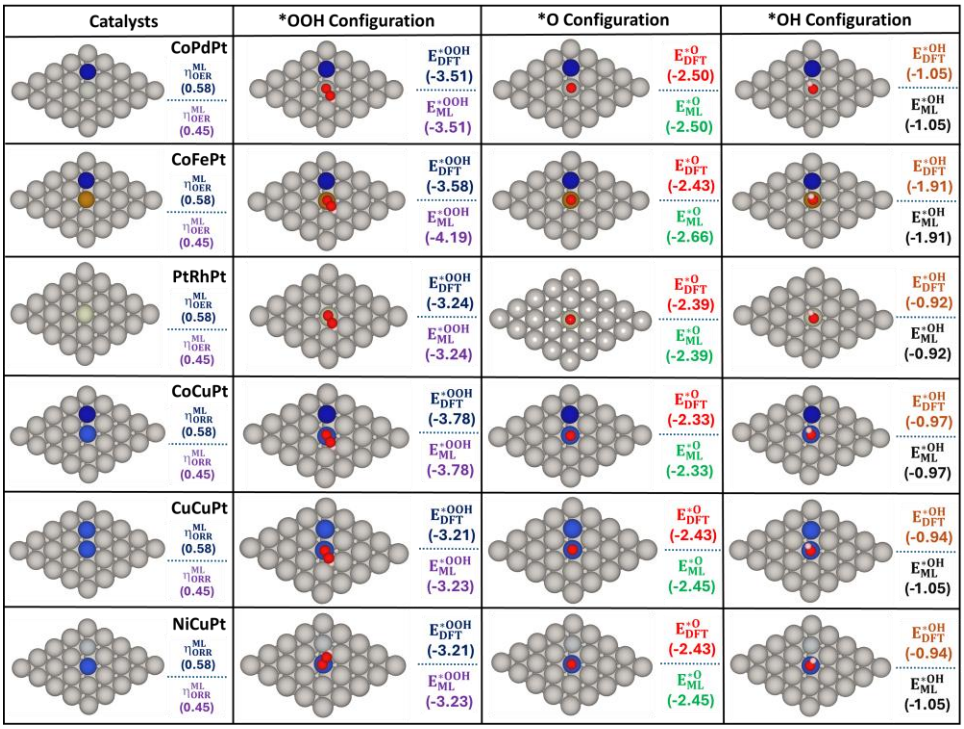


Figure 6.5: ML-predicted and DFT-calculated adsorption energies of the best three electrocatalysts for the OER and ORR processes. The overpotential values (η) for the thermodynamically rate-determining step (RDS) were determined by combining the adsorption energies of *OOH, *O, and *OH intermediates obtained from both DFT and ML datasets, following the associative pathway.

After validating, we utilized the best-performing GBR model to predict the OER/ORR activity of the unseen RuM′Pt, RhM′Pt, AgM′Pt, OsM′Pt, IrM′Pt and AuM′Pt, with the trends of η^{OER}_{ML}, and η^{ORR}_{ML} values for these catalysts are represented in Figure 6.4c-d. To identify an active bifunctional electrocatalyst that efficiently facilitates both reactions, we computed the widely accepted descriptor, $\eta_{Bi} = \eta^{OER}_{ML} + \eta^{ORR}_{ML}$, represented near the intersection of the Pt(111) and RuO_2 lines in Figure 6.4a-b. Among the catalysts, CoCuPt emerged as the best bifunctional electrocatalyst with a low η_{Bi} = 0.66 V, outperforming the benchmark Ru/C systems (η_{Bi} = 1.01 V). Additionally, CoPdPt emerged as an active bifunctional electrocatalyst

with lower η_{Bi} values of 0.84 V. The reaction energy diagram for OER and ORR at 0 V and 1.23 V for these electrocatalysts is provided in panels a and b of Figure 6.6. Further, we examined the electronic structures of CoCuPt and CoPdPt to understand the origins of their higher OER and ORR activity, respectively (Figure 6.4B). The partial density of states (PDOS) analysis reveals the presence of frontier Co (3d) and Cu (3d) states in CoCuPt, and Co (3d) and Pd (5d) states in CoPdPt near the Fermi level (E_f), highlighting the synergistic interaction between the two constituent metal atoms in driving the overall catalysis in DACs. In addition, the narrow distribution of O (2p) states near the E_f, together with a moderate charge transfer from the surface to the intermediates, indicates optimal adsorption strength, which accounts for their enhanced catalytic activity.

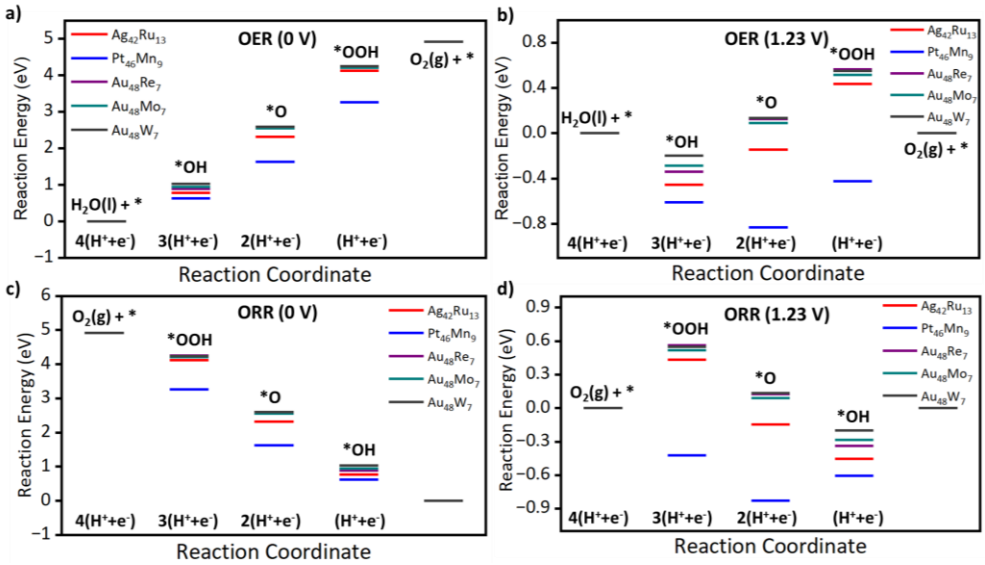


Figure 6.6: Reaction energy diagrams for OER and ORR mechanisms of the top five bifunctional electrocatalysts at 0 V and 1.23 V, respectively.

6.3.7 Scaling Relationship for OER/ORR Activity.

Scaling relationships in catalysis describe simple linear correlations between the thermodynamic properties of chemically related species across different catalytic surfaces *[6,8,36]*. Although useful, these relationships impose intrinsic limits on the maximum catalytic activity. Previous studies have shown that single-atom catalysts (SACs) and subnano clusters can

break these scaling constraints, enhancing performance *[37,38]*. More recently, DACs, particularly those with heteronuclear active sites, have demonstrated the ability to surpass SACs while also mitigating the scaling limitation *[8]*. In our study, we examined the correlation between OER and ORR for our DACs, as presented in Figure 6.7. To derive the meaningful descriptors for the OER/ORR activity of DACs in our database, we streamline our scaling investigation into two different analysis: (1) correlation between ML-predicted E_{ads} (E_{*O} *vs.* E_{*OH}, and E_{*OH} *vs.* E_{*OOH}) (Figure 6.7a-f), and (2) ML-predicted E_{*O} and $\eta_{OER}^{ML}/\eta_{ORR}^{ML}$ (Figure 6.7g-l).

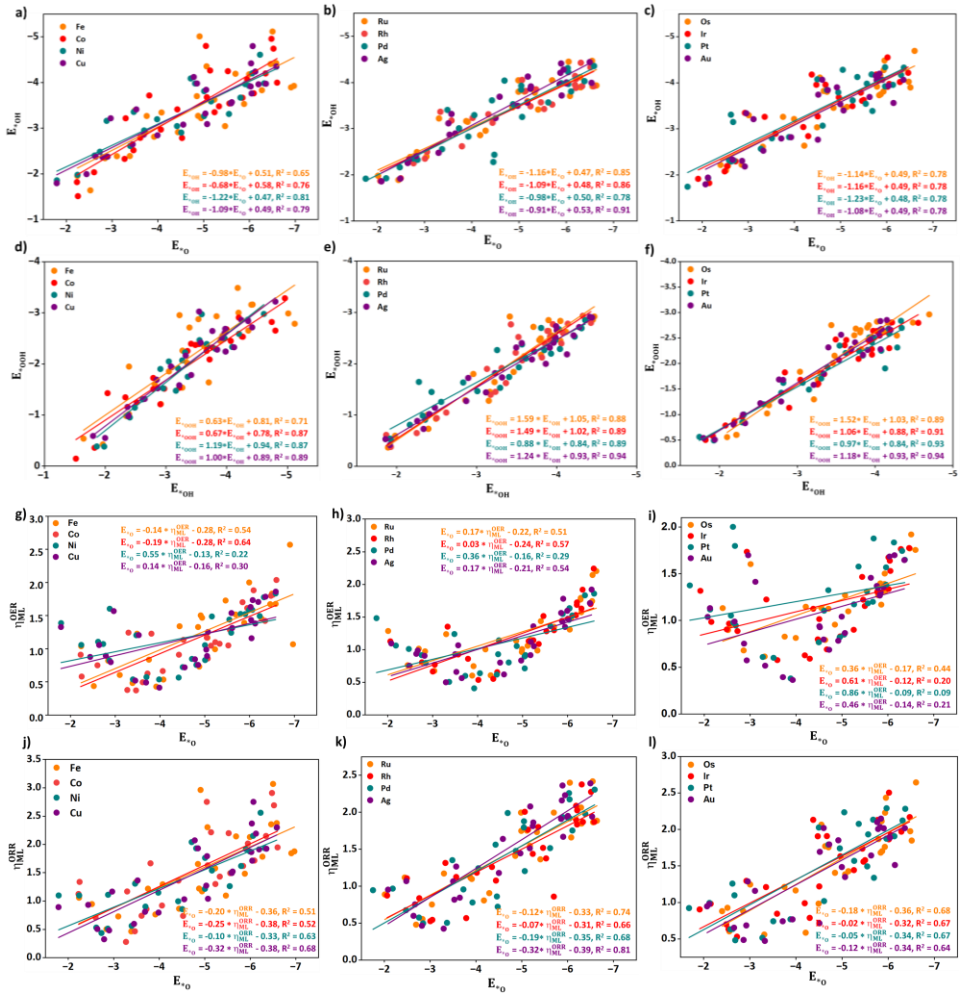


Figure 6.7: Correlations between intermediates and OER/ORR activity. The scaling relationship between the (a-c) E_{*O} vs E_{*OH}, and (d-f) E_{*OH} vs E_{*OOH} of MM′Pt. Relationship between (g-i) E_{*O} vs η_{ML}^{ORR}, and (j-l) E_{*O} vs η_{ML}^{OER} of MM′Pt.

Interestingly, a strong linear correlation is observed between E_{*_O} and $E_{*_{OH}}$ for RuM′Pt (0.85), RhM′Pt (0.86), and AgM′Pt (0.91). In contrast, the 3d-based DACs (FeM′Pt, CoM′Pt, NiM′Pt, and CuM′Pt) and 5d-based DACs (OsM′Pt, IrM′Pt, PtM′Pt, and AuM′Pt) show weaker correlations, compared to the $R^2 = 0.91$ for bulk surfaces. For E_{*OH} and E_{*OOH}, the correlation remains consistently strong across most DACs, except FeM′Pt (R^2 = 0.71). In scaling relationship, the slope of the fitted line reflects the optimal electron density contribution from the catalyts to the bound intermediates, oxygen in this case and the intercept represent the surface geometries *[36]*. However, the computed slopes for E_{*_O} *vs.* $E_{*_{OH}}$, and $E_{*_{OH}}$ *vs.* $E_{*_{OOH}}$ deviate from the expected values of 0.5 and 1.0, respectively. This deviation indicates a reduced electron contribution to the bound oxygen atom in most MM′Pt systems. Overall, these results highlight the limitations of the effective medium theorem in describing E_{ads} for DACs.

Further, we examined the scaling relationships between the ML-predicted E_{ads} and the overall OER/ORR activity to uncover the underlying structure-activity relationships for DACs. For this purpose, we analyzed the correlation between the ML-predicted E_{*_O} and the $\eta_{OER}^{ML}/\eta_{ORR}^{ML}$, selecting E_{*_O} as the descriptor owing to its stronger adsorption tendency and its frequent involvement in the RDS across most electrocatalysts (Figure 6.7g-l). For OER, a weak correlation is observed, with R^2 values ranging from 0.09 to 0.64, indicating a slight improvement in R^2 values for ORR activity. This deviation highlights that the activity of MM′Pt-based DACs cannot be captured by a single descriptor, unlike bulk systems, where scaling relations often reduce catalyst performance to a single dominant adsorption energy. Instead, DAC's activity emerges as a multi-component descriptor requiring contributions from E_{*_O}, $E_{*_{OH}}$, and $E_{*_{OOH}}$ to be considered simultaneously. Consequently, the weak scaling observed here reflects a breaking of conventional scaling relations and points to $\eta_{OER}^{ML}/\eta_{ORR}^{ML}$ as a multi-component descriptor. This is particularly relevant for constructing

accurate activity volcano plots for DACs, where deviations from bulk scaling open pathways for surpassing conventional catalytic limitations.

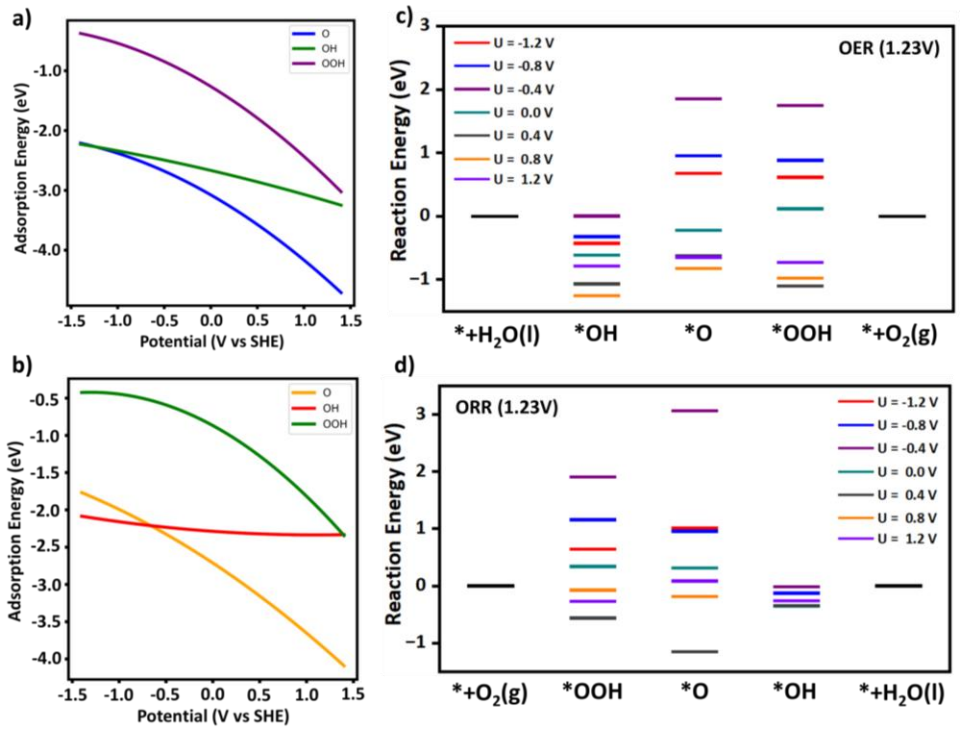


Figure 6.8: Potential-dependent activity. Adsorption energies of *OOH, *O, and *OH intermediates for (a) CoPdPt and (b) CoCuPt as a function of applied potential (vs. SHE). Reaction energy diagrams for (c) OER and (d) ORR of the best performing electrocatalyst at 1.23 V.

6.3.8 Potential Dependent Activity

The above catalysts are screened using the CHE model developed by Nørskov and colleagues, which maintains a constant electron count during the reaction *[23,29,40]*. These simulations employ a constant charge model (CCM), where the total number of electrons remains fixed and the applied potential is implicitly incorporated into the reaction free energy, thereby maintaining system neutrality. However, this method overlooks the dynamic electron exchange with electrodes during the reaction, which can cause inaccuracies by neglecting charge effects. To address these limitations, we use the constant potential method (CPM) for the best OER and ORR electrocatalysts in our ML dataset to more accurately capture their behaviour under realistic conditions. In CPM, the number of electrons

varies with changes in electrode potential, enabling grand canonical DFT calculations where the electrode can exchange electrons with an external reservoir to maintain a constant potential *[22]*.

Table 6.2: Potential dependent η_{OER}, and η_{ORR} of the best electrocatalysts (vs SHE) at 1.23 V.

Potential (vs SHE)	CoPdPt (OER)	CoCuPt (ORR)
	η_{OER}(V)	η_{ORR}(V)
U = -1.2 V	1.10	0.64
U = -0.8 V	1.27	1.16
U = -0.4 V	1.85	1.90
U = 0.0 V	0.34	0.35
U = 0.4 V	1.09	0.81
U = 0.8 V	0.97	0.27
U = 1.2 V	0.73	0.26

Figure 6.8a-b shows the sensitivity of the E_{ads} as a function of the applied potential when explicitly accounting for potential via surface charging, with all energy points well fitted to quadratic relationships. For CoPdPt, E_{*_O}, $E_{*_{OH}}$, and $E_{*_{OOH}}$ show weaker adsorption at negative potentials due to electrostatic repulsion with the negatively charged surface, while at positive potentials, E_{ads} becomes more negative, indicating stronger adsorption. However, for CoCuPt, the $E_{*_{OH}}$ remains nearly constant, showing a clear crossover with E_{*_O} at U = -0.5 V, consistent with the previous investigations on nanoclusters *[41]*. Furthermore, we derive the U-dependent reaction energy diagram via the AEM mechanism at 1.23 V (Figures 6.8c-d). Notably, the reaction energy does not linearly correlate with the electrode potential, with a sudden boost in the thermodynamic η_{OER}, and η_{ORR} as the electrode potential goes from 0.0 V to $\pm$0.4 V (Table 6.2). This demonstrates that tuning the external potential can modulate electrocatalytic activity by controlling the flow of electrons during the charging of the catalyst surface.

6.3.9 Conclusion

In summary, we developed an ML framework to investigate the structure-activity relationships governing the OER and ORR performance of late-transition-metal based heterometallic dual-atom catalysts ($MM'Pt$) embedded in the Pt(111) surface. Using feature engineering, we represent a new set of non-ab-initio descriptors derived from solid-state data that can capture the adsorption behavior of key intermediates without relying on DFT-derived electronic features. With these easily accessible inputs, the trained ML models achieve MAEs of 0.20, 0.24, and 0.19 eV for *OOH, *O, and *OH adsorption energies. The dopant atom's d-band filling and d-band center strongly dictate the adsorption on local environment, providing a practical route to approximate d-state characteristics while avoiding expensive calculations. Among the systems screened, CoPdPt stands out as a bifunctional catalyst with a low combined overpotential of 0.84 V. Most importantly, we represent the non-scaling behaviour of DACs, indicating deviation from the Sabatier principle and representing optimal activity compared to bulk RuO_2 and Pt(111) surface, suggesting routes to enhance activity beyond the volcano apex. Further, grand canonical DFT calculations further show that the activity varies non-linearly with applied potential, with a steep increase in overpotential between applied potential of 0.0 and $\pm$0.4 V. Overall, this study introduces a set of descriptors that effectively captures the catalytic trends in DACs and highlights the need to identify such fundamental, low-cost features to accelerate catalyst discovery through ML-driven screening.

6.4 References

1. She, Z. W., Kibsgaard, J., Dickens, C. F., Chorkendorff, I., Nørskov, J. K., Jaramillo, T. F. (2017), Combining theory and experiment in electrocatalysis: Insights into materials design, *Science*, 355, eaad4998 (DOI: 10.1126/science.aad4998)

2. Lewis, N. S. (2016), Research opportunities to advance solar energy utilization, *Science*, 351, aad1920 (DOI: 10.1126/science.aad1920)
3. Kress, P. L., Zhang, S., Wang, Y., Çınar, V., Friend, C. M., Sykes, E. H. C., Montemore, M. M. (2023), A priori design of dual-atom alloy sites and experimental demonstration of ethanol dehydrogenation and dehydration on PtCrAg, *J. Am. Chem. Soc.*, 145, 8711–8722 (DOI: 10.1021/jacs.2c13577)
4. Liang, X., Fu, N., Yao, S., Li, Z., Li, Y. (2022), The progress and outlook of metal single-atom-site catalysis, *J. Am. Chem. Soc.*, 144, 10252–10273 (DOI: 10.1021/jacs.1c12642)
5. Exner, K. S. (2021), Why the breaking of the OOH versus OH scaling relation might cause decreased electrocatalytic activity, *Chem Catalysis*, 1, 258–260 (DOI: 10.1016/j.checat.2021.06.011)
6. Pérez-Ramírez, J., López, N. (2019), Strategies to break linear scaling relationships, *Nat. Catal.*, 2, 971–976 (DOI: 10.1038/s41929-019-0376-6)
7. Pérez-Ramírez, J., López, N. (2019), Strategies to break linear scaling relationships, *Nat. Catal.*, 2, 971–976 (DOI: 10.1038/s41929-019-0376-6)
8. Fang, C., Zhou, J., Zhang, L., Wan, W., Ding, Y., Sun, X. (2023), Synergy of dual-atom catalysts deviated from the scaling relationship for oxygen evolution reaction, *Nat. Commun.*, 14, 40177 (DOI: 10.1038/s41467-023-40177-1)
9. Li, Y., Wei, B., Zhu, M., Chen, J., Jiang, Q., Yang, B., Hou, Y., Lei, L., Li, Z., Zhang, R., Lu, Y. (2021), Synergistic effect of atomically dispersed Ni–Zn pair sites for enhanced CO_2 electroreduction, *Adv. Mater.*, 33, 2102212 (DOI: 10.1002/adma.202102212)
10. Berger, F., Michaelides, A. (2025), When are dopant d-states free-atom-like? Periodic trends and confinement effects in single-atom alloys, *J. Am. Chem. Soc.*, 147, 11234–11245 (DOI: 10.1021/jacs.5c07771)

11. Saini, S., Halldin Stenlid, J., Abild-Pedersen, F. (2022), Electronic structure factors and the importance of adsorbate effects in chemisorption on surface alloys, *npj Comput. Mater.*, 8, 156 (DOI: 10.1038/s41524-022-00846-z)
12. Mai, H., Le, T. C., Chen, D., Winkler, D. A., Caruso, R. A. (2022), Machine learning for electrocatalyst and photocatalyst design and discovery, *Chem. Rev.*, 122, 13478–13540 (DOI: 10.1021/acs.chemrev.2c00061)
13. Kim, M., Yeo, B. C., Park, Y., Lee, H. M., Han, S. S., Kim, D. (2020), Artificial intelligence to accelerate the discovery of N_2 electroreduction catalysts, *Chem. Mater.*, 32, 709–720 (DOI: 10.1021/acs.chemmater.9b03686)
14. Sharma, R. K., Minhas, H., Pathak, B. (2025), Machine learning-guided discovery of alloy nanoclusters: Steering morphology-based activity and selectivity relationships in bifunctional electrocatalysts, *ACS Appl. Mater. Interfaces*, 17, 11234–11246 (DOI: 10.1021/acsami.5c07198)
15. Sharma, R. K., Jena, M. K., Minhas, H., Pathak, B. (2024), Machine-learning-assisted screening of nanocluster electrocatalysts: Mapping and reshaping the activity volcano for the oxygen reduction reaction, *ACS Appl. Mater. Interfaces*, 16, 63589–63601 (DOI: 10.1021/acsami.4c14076)
16. Guo, C., Zhang, F., Han, X., Zhang, L., Hou, Q., Gong, L., Wang, J., Xia, Z., Hao, J., Xie, K. (2023), Intrinsic descriptor-guided noble metal cathode design for Li–CO_2 batteries, *Adv. Mater.*, 35, 2302325 (DOI: 10.1002/adma.202302325)
17. Zhu, Q., Gu, Y., Wang, X., Gu, Y., Ma, J. (2024), Synergistic effect between metal and sulfur vacancy to boost CO_2 reduction efficiency, *JACS Au*, 4, 1–12 (DOI: 10.1021/jacsau.3c00558)

18. Zhu, Q., Gu, Y., Ma, J. (2025), Digital descriptors in predicting catalytic reaction efficiency and selectivity, *J. Phys. Chem. Lett.*, 16, 2357–2368 (DOI: 10.1021/acs.jpclett.4c03733)
19. Kresse, G., Hafner, J. (1994), Ab initio molecular-dynamics simulation of the liquid-metal–amorphous-semiconductor transition in germanium, *Phys. Rev. B*, 49, 14251–14269 (DOI: 10.1103/PhysRevB.49.14251)
20. Perdew, J. P., Chevary, J. A., Vosko, S. H., Jackson, K. A., Pederson, M. R., Singh, D. J., Fiolhais, C. (1992), Applications of the generalized gradient approximation, *Phys. Rev. B*, 46, 6671–6687 (DOI: 10.1103/PhysRevB.46.6671)
21. Tang, W., Sanville, E., Henkelman, G. (2009), A grid-based Bader analysis algorithm without lattice bias, *J. Phys.: Condens. Matter*, 21, 084204 (DOI: 10.1088/0953-8984/21/8/084204)
22. Cao, H., Lv, X., Qian, S., Li, J., Wang, Y. G. (2024), Constant potential thermodynamic integration for electrochemical reactions, *J. Phys. Chem. Lett.*, 15, 2156–2163 (DOI: 10.1021/acs.jpclett.3c03318)
23. Nørskov, J. K., Rossmeisl, J., Logadottir, A., Lindqvist, L., Kitchin, J. R., Bligaard, T., Jónsson, H. (2004), Origin of the overpotential for oxygen reduction at fuel-cell cathodes, *J. Phys. Chem. B*, 108, 17886–17892 (DOI: 10.1021/jp047349j)
24. Zhang, S., Sykes, E. H. C., Montemore, M. M. (2022), Tuning reactivity in trimetallic dual-atom alloys, *Chem. Sci.*, 13, 14321–14331 (DOI: 10.1039/d2sc03650a)
25. Ferrin, P., Simonetti, D., Kandoi, S., Kunkes, E., Dumesic, J. A., Nørskov, J. K., Mavrikakis, M. (2009), Modeling ethanol decomposition on transition metals, *J. Am. Chem. Soc.*, 131, 5809–5815 (DOI: 10.1021/ja8099322)
26. Bligaard, T., Nørskov, J. K., Dahl, S., Matthiesen, J., Christensen, C. H., Sehested, J. (2004), The Brønsted–Evans–Polanyi relation and

volcano curves, *J. Catal.*, 224, 206–217 (DOI: 10.1016/j.jcat.2004.02.034)

27. Nair, A. S., Pathak, B. (2019), Computational screening for ORR activity of 3d transition-metal-based core–shell clusters, *J. Phys. Chem. C*, 123, 3582–3591 (DOI: 10.1021/acs.jpcc.8b11483)
28. Xin, H., Linic, S. (2010), Exceptions to the d-band model of chemisorption, *J. Chem. Phys.*, 132, 221101 (DOI: 10.1063/1.3437609)
29. Jäger, M. O. J., Ranawat, Y. S., Canova, F. F., Morooka, E. V., Foster, A. S. (2020), Machine-learning-aided screening of hydrogen adsorption, *ACS Comb. Sci.*, 22, 832–840 (DOI: 10.1021/acscombsci.0c00102)
30. Ruban, A., Hammer, B., Stoltze, P., Skriver, H. L., Nørskov, J. K. (1997), Surface electronic structure and reactivity of transition and noble metals, *J. Mol. Catal. A: Chem.*, 115, 421–429 (DOI: 10.1016/S1381-1169(96)00348-2)
31. Li, B., Zhang, P., Liang, S., Ren, G. (2008), Feature extraction and selection using wavelet entropy, *Proc. ICSP*, 2008, 2186–2190 (DOI: 10.1109/ICOSP.2008.4697740)
32. Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011), Scikit-learn: Machine learning in Python, *J. Mach. Learn. Res.*, 12, 2825–2830
33. Lundberg, S. M., Erion, G., Chen, H., et al. (2020), Explainable AI for tree-based models, *Nat. Mach. Intell.*, 2, 56–67 (DOI: 10.1038/s42256-019-0138-9)
34. Rossmeisl, J., Qu, Z. W., Zhu, H., Kroes, G. J., Nørskov, J. K. (2007), Electrolysis of water on oxide surfaces, *J. Electroanal. Chem.*, 607, 83–89 (DOI: 10.1016/j.jelechem.2006.11.008)
35. Kulkarni, A., Siahrostami, S., Patel, A., Nørskov, J. K. (2018), Understanding catalytic activity trends in ORR, *Chem. Rev.*, 118, 2302–2312 (DOI: 10.1021/acs.chemrev.7b00488)

36. Montemore, M. M., Medlin, J. W. (2014), Scaling relations between adsorption energies, *Catal. Sci. Technol.*, 4, 3748–3761 (DOI: 10.1039/c4cy00335g)
37. Li, L., Yuan, K., Chen, Y. (2022), Breaking scaling relationship limits: Single- to dual-atom catalysts, *Acc. Mater. Res.*, 3, 451–463 (DOI: 10.1021/accountsmr.1c00264)
38. Zandkarimi, B., Alexandrova, A. N. (2019), Dynamics of subnanometer Pt clusters and scaling relations, *J. Phys. Chem. Lett.*, 10, 460–467 (DOI: 10.1021/acs.jpclett.8b03680)
39. Calle-Vallejo, F., Loffreda, D., Koper, M. T. M., Sautet, P. (2015), Structural sensitivity in adsorption-energy scaling relations, *Nat. Chem.*, 7, 403–410 (DOI: 10.1038/nchem.2226)
40. Zhu, J., Hu, L., Zhao, P., Lee, L. Y. S., Wong, K. Y. (2020), Recent advances in electrocatalytic hydrogen evolution, *Chem. Rev.*, 120, 851–918 (DOI: 10.1021/acs.chemrev.9b00248)
41. Munarriz, J., Zhang, Z., Sautet, P., Alexandrova, A. N. (2022), Graphite-supported Pt_n cluster electrocatalysts under applied potential, *ACS Catal.*, 12, 14954–14966 (DOI: 10.1021/acscatal.2c04643)

CHAPTER 7

Steering Morphology Based Activity and Selectivity Relationships

Chapter 7

Section: A

7.1 Introduction

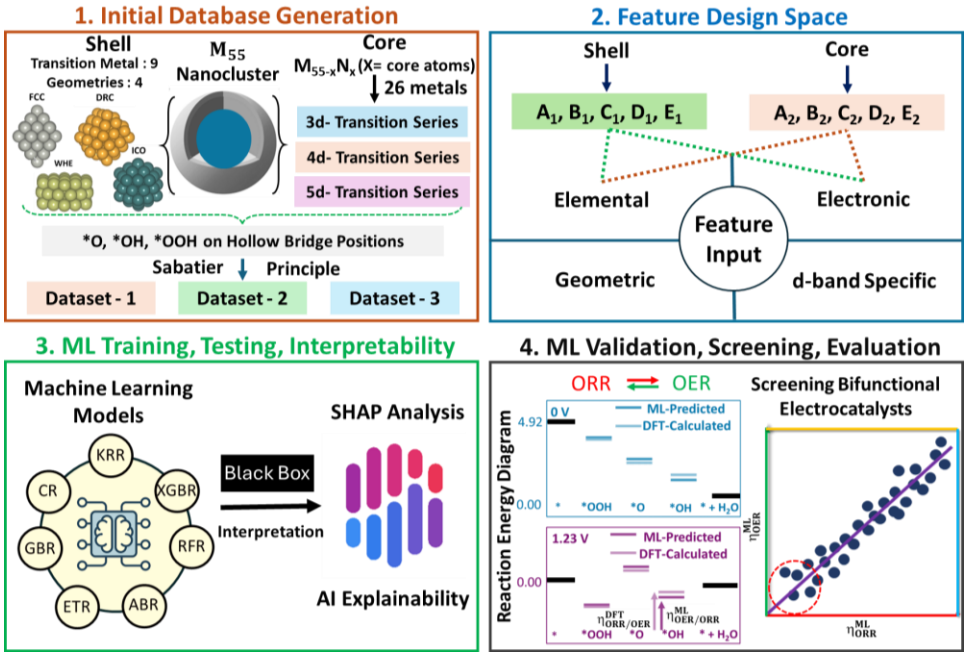
The progressive energy and environmental crisis raise imperative demands for the development of sustainable and efficient energy processes to meet global energy needs *[1]*. Among the most promising technologies are metal-air batteries *[2]*, regenerative fuel cells *[3]*, and proton-exchange membrane fuel cells (PEMFCs) *[4]*, which offer low greenhouse gas emissions, high energy density, and efficient energy conversion. Central to these devices are electrocatalysts, which facilitate the conversion of chemical energy into electrical energy. In particular, the oxygen evolution reaction (OER) and oxygen reduction reaction (ORR) are key 4-electron transfer processes, with OER functioning as the anode reaction in electrolyzers and ORR as the cathode reaction in fuel cells *[5]*. However, both OER and ORR reactions are limited by their sluggish kinetics, high overpotential, and multistep proton-coupled electron-transfer (PCET) reactions, further restricting their widespread commercialization *[6,7]*. Therefore, optimizing bifunctional electrocatalysts that facilitate both OER and ORR is crucial for scaling up industrial energy production. Currently, commercial ORR and OER catalysts, primarily based on precious metals like Pt and IrO_2, have limited potential for enhancement through increased catalyst loading or surface roughening to expose more active sites *[8,9]*. Therefore, there is a strong impetus to improve intrinsic catalytic activity by exploring other classes of materials, such as nanoclusters and their alloys *[10,11]*. Alloy nanoclusters, composed of two or more metals, enhance catalytic activity through synergistic effects while providing a cost-effective alternative to precious metals like Pt and Pd by reducing their usage in electrocatalysis. Amid these alternatives, core-shell nanoclusters have garnered growing attention for their ability to optimize catalytic activity through careful selection of constituent metals and fine-tuning of surface state properties *[12]*.

Core-shell nanoclusters, composed of a low-cost metal core and a noble metal shell, are designed to sustain electrochemical conditions while

optimizing catalytic performance, offering versatility due to diverse thermodynamically feasible metal combinations. Previous studies on core-shell catalysts for ORR focused on extended surface alloys (ESAs), where "Pt-skin" surfaces formed by leaching and "Pt-skeleton" surfaces formed by annealing exhibited high activity *[13,14]*. Greeley demonstrates enhanced ORR activity in Pt_3M (M= Ti, Fe, Co, Ni) alloy with Pt overlayers, attributed to weaker Pt-O bonds reducing oxygen site blocking *[15]*. Henkelman and colleagues correlated O-binding energy with catalytic activity in ~1.5 nm clusters, while binary and ternary Pt alloys with early transition metals (e.g., Ti, Cr) showed promising performance *[16]*. Despite progress, systematic explorations of 3d, 4d, and 5d transition metal-based core-shell clusters remain limited, highlighting the need for comprehensive evaluations of energetics and overpotential for OER/ORR. Additionally, the vast chemical space of core-shell electrocatalysts also necessitates data-driven approaches to enable accurate high-throughput predictions of catalytic performance.

Recently, machine learning (ML) has become a powerful tool for accelerating catalyst discovery by disentangling the complex catalyst-intermediate interaction in reactions like nitrogen reduction reaction (NRR) *[17]*, hydrogen evolution reaction (HER) *[18]*, CO_2 reduction reaction (CO_2RR) *[19]* and ORR/OER *[20,21]*. However, all these studies focused on bulk-based systems. For core-shell nanoclusters, Mao et al. employed DFT-based high-throughput screening to evaluate the HER activity of icosahedral $Cu_{55-n}M_n$ (M=Co, Ni, Ru, and Rh) nanoclusters, identifying $Cu_{49}Ni_6$ as the most active catalysts *[22]*. Similarly, Xu et al. and Jäger et al. evaluated icosahedral core-shell nanoclusters as HER-active catalysts, highlighting the importance of surface electronic structure *[23,24]*. Notably, these studies often consider a single geometry of the nanocluster, neglecting the activity derived from the multiple coexisting structurally diverse isomers for the magic-sized nanoclusters. Furthermore, the fundamental determinants of adsorption and the catalytic origins of

multistep OER/ORR with varying elemental composition are yet to be thoroughly investigated.



Scheme 7.1: Schematic illustration of the ML framework for screening active bifunctional electrocatalysts for OER/ORR in core-shell TM_{55} nanoclusters with an icosahedron (ICO), face-centered cubic (FCC), wheel-type (WHE), and distorted-reduced core (DRC) geometries. The framework involves four streamlined processes: (1) Generation of adsorption energy database for *O, *OH, and *OOH intermediates across nanoclusters with various geometries and compositions, (2) Designing feature space to encode structure-activity relationships of the local chemical environment of core-shell TM_{55} nanoclusters, with subscripts 1 and 2 referring to features of the shell and core atoms, respectively, (3) Evaluating and interpreting the ML model, and (4) Validation and screening of electrocatalysts using DFT and ML predictions, based on reaction energies and overpotential at 0 V and 1.23 V, where the horizontal black line represents the neutral catalyst. The framework highlights identifying bifunctional electrocatalysts with DFT-calculated $(\eta_{OER}^{DFT}, \eta_{ORR}^{DFT})$ and ML-predicted $(\eta_{OER}^{ML}, \eta_{ORR}^{ML})$ overpotential values for OER and ORR.

Herein, we present a machine-learning framework for the structural screening of 234 electrocatalysts, consisting of 9 unary and 225 binary core-shell nanoclusters derived from 55-atom transition metal-based structures, exhibiting multiple geometries for OER, ORR, and bifunctional activity, as illustrated in Figure 7.1. By distinguishing the feature space of shell and core atoms, we define the single-cluster catalyst identity of these electrocatalysts, elucidating the intricate relationship between intermediate adsorption strength, catalytic activity, and elemental composition. Using ML algorithms, we predict the adsorption energies of key intermediates (*O, *OH, and *OOH) across a vast chemical space to identify active catalysts for OER, ORR, and bifunctional applications. The versatility of our ML approach is demonstrated by mapping the reaction energy changes and reshaping the activity volcano, transitioning from RuO_2(IrO_2)/Pt (111) to late-transition Au/Ag-based nanoclusters for OER/ORR. Furthermore, we address the selectivity challenge between the competing two-electron and four-electron ORR pathways, facilitating the classification of catalysts into these dominant mechanisms. The trends observed here provide valuable insights for screening bimetallic catalysts suitable for fuel cell applications.

7.2 Computational Details

All plane-wave spin-polarized density functional theory (DFT) calculations were performed using the Vienna *ab initio* simulation package (VASP) *[25]* with the Perdew-Burke-Ernzerhof (PBE) exchange-correlation functional based on the generalized gradient approximation (GGA) *[26]*. The interactions between ion cores and valence electrons were treated using the projector augmented wave (PAW) method *[27]*. During structural relaxations, a kinetic energy cut-off of 500 eV was used, and convergence was achieved when the total energy difference between interactions was below 10^{-4} eV and atomic forces were reduced to less than 0.02 eV/Å. Gaussian smearing with a width of 0.2 eV was applied, and Γ-centered k-point grids of (1×1×1) were adopted for Brillouin zone

sampling, consistent with earlier studies *[28,29]*. The Pt (111) and RuO_2 surfaces are modelled using 4 × 4 × 1 and 4 × 2 × 1 supercells, respectively, comprising four atomic layers for Pt and six layers for RuO_2. A vacuum spacing of 20 Å is introduced along the z-direction to eliminate spurious interaction between periodic images.

For the RuO_2 (110) surface, a Hubbard U correction of U = 3.2 eV is applied to better account for the on-site Coulomb interactions and accurately describes the electronic structure of the oxide system. To employ the implicit solvation method, we used the VASPsol package that integrates solvation into the VASP within a self-consistent continuum model to simulate the reaction in aqueous electrolyte solution *[30]*. We set a relative permittivity of 78.4 for the aqueous electrolyte and a Debye screening length of 3.0 Å, corresponding to an ionic force of 1 M at 298 K. Note that we have employed the PBE functional for all DFT calculations. While the DFT+U approach is valuable for systems with strong electron correlations, such as transition metal oxides *[31]*, its applicability to metallic or alloy nanoclusters is limited and may result in overestimated overpotentials or lattice parameters *[32]*. Furthermore, achieving accurate predictions requires system-specific benchmarking of U_{eff} values, which is computationally expensive and beyond the scope of the present work.

7.3 Methods

7.3.1 Structure Design of TM_{55} Nanoclusters

Among metal clusters, 55-atom nanoclusters have gained significant interest due to their "magic size", offering significant stability and making them ideal for catalytic applications *[33]*. When composed of late-transition metals, these clusters exhibit tunable electronic properties alongside improved catalytic performance for ORR/OER. Using DFT calculations, we curated a database of 234 electrocatalysts by optimizing the most stable geometries of late-transition metal TM_{55} (TM = 3d: Co, Ni, Cu; 4d: Rh, Pd, Ag; 5d: Ir, Pt, and Au) identified in previous studies, using the VASP code

[33,34]. Often considered reliable putative global minima configurations (pGMCs), these geometries were selected from a diverse pool of structural candidates. Each structure is characterized by a unique atomic arrangement specific to its respective TM, including high-symmetry icosahedron (ICO) for Co_{55}, Ni_{55}, Cu_{55}, and Ag_{55}; bulk-like face-centered cubic (FCC) for Rh_{55}, and Pd_{55}; hexagonal close-packed wheel-type (WHE) for Ir_{55}; and distorted-reduced core (DRC) for Pt_{55} and Au_{55}, as illustrated in Figure 7.2a.

Compared to Pt (111), the surface atoms of these TM_{55} geometries exhibit diverse coordination numbers (CN) ranging from 4 to 12, highlighting a broader range of coordination environments (Figure 7.2b) *[35]*. Among these, 3d and 5d TM_{55} nanoclusters display relatively saturated coordination environments, with the majority of surface atoms having CN values of 8-9 for 3d clusters and 7 for 5d clusters. Interestingly, these pGMCs are further distinguished by distinct structural motifs, featuring various core-shell arrangements with core atom occupancies of 7, 9, 10, and 13 for different TM_{55} nanoclusters (Figure 7.2c). Such variations in core composition may significantly influence the structural, electronic, and catalytic properties of the clusters, driven by the interplay of size effects, lattice strain, electronic factors, and chemical interactions between the core and shell elements *[36]*.

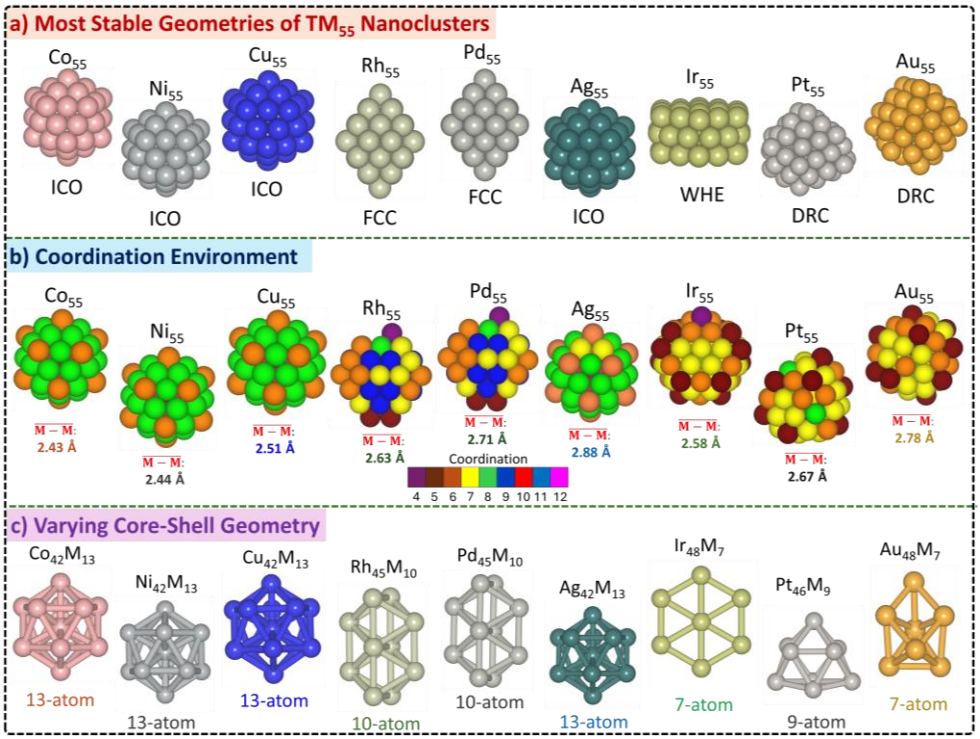


Figure 7.2: (a) Most stable geometries of the putative global minimum DFT-PBE configurations (pGMC) of TM_{55} nanoclusters (TM = 3d: Co, Ni, Cu; 4d: Rh, Pd, Ag; 5d: Ir, Pt, and Au), (b) Coordination environments of different structural motifs, with each atom colored by its coordination number (CN) ranging from 4 to 12, and the average metal-metal bond distance $\overline{M-M}$ in Å, and (c) Variations in core-shell geometries of TM_{55} nanoclusters with core atom occupancies (M) of 7, 9, 10 and 13 for different TM_{55} nanoclusters.

7.3.2 Core-Shell Engineering

Compared to pure clusters, core-shell nanoclusters experience tensile or compressive strain due to lattice mismatch between core and shell metals, which alters their electronic structure and influences catalytic properties *[37]*. To design the core-shell catalyst model, we systematically replaced the core atoms of TM_{55} nanoclusters (TM = Co, Ni, Cu, Rh, Pd, Ag, Ir, Pt, and Au) with 3d, 4d, and 5d transition metals, creating bimetallic compositions with one element in the shell and another in the core. This process resulted in the following compositions: $Ag_{42}M_{13}$, $Au_{48}M_7$,

$Co_{42}M_{13}$, $Cu_{42}M_{13}$, $Ir_{48}M_7$, $Ni_{42}M_{13}$, $Pd_{45}M_{10}$, $Pt_{46}M_9$, and $Rh_{45}M_{10}$ (Figure 7.3a). Metals such as Tc, Cd, La, and Hg were excluded due to their radioactivity and toxicity *[38]*. In total, we extracted a dataset of 234 electrocatalysts, comprising 9 unary and 225 core-shell TM_{55} nanoclusters, for subsequent electrocatalytic investigations. Notably, no surface segregation or core melting was observed during optimization, confirming the stability of these alloys. The average metal-metal bond length ($\overline{M-M}$) in these bimetallic nanoclusters followed a non-linear trend: $Ag_{42}M_{13}$ > $Au_{48}M_7$ > $Pd_{45}M_{10}$ > $Pt_{46}M_9$ > $Rh_{45}M_{10}$ > $Ir_{48}M_7$ > $Cu_{42}M_{13}$ > $Ni_{42}M_{13}$ ≈ $Co_{42}M_{13}$. This irregularity arises from a combination of orbital hybridization, and atomic size mismatch.

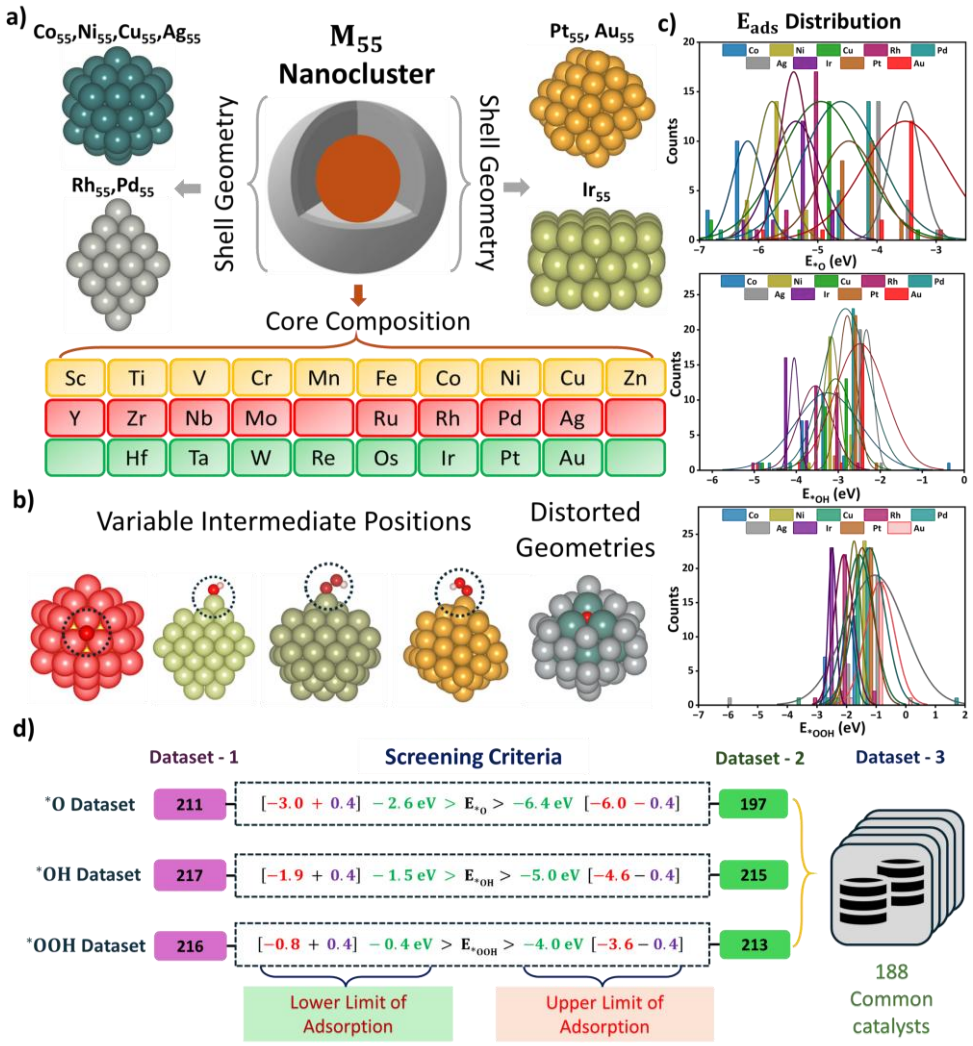


Figure 7.3: (a) Schematic illustration of the elemental composition of the core and shell of the closed-shell TM_{55} (TM = Co, Ni, Cu, Rh, Pd, Ag, Ir, Pt, and Au) across different structural motifs (ICO, FCC, WHE, and DRC), where the empty blocks denote the excluded transition metals from the core, (b) Distinct hollow and top positions considered for single-intermediate adsorption. Note that the distorted geometries representing surface segregation were removed from our dataset, (c) Range of adsorption energy (E_{ads}) distribution for *O, *OH, and *OOH intermediates, demonstrating variations specific to different core-shell TM_{55} nanoclusters alloys, and (d) Screening criteria for E_{ads} of *O, *OH, and *OOH following the Sabatier principle to identify active OER, ORR, and bifunctional electrocatalysts.

7.4 Results and Discussion

7.4.1 Adsorption Energy Landscape

After extracting the stable bimetallic core-shell geometries, we explore the bifunctional activity on their surfaces. In line with the Bell-Evans-Polanyi (BEP) principle, the adsorption energy (E_{ads}) of an intermediate serve as a key descriptor for evaluating a catalyst's performance and predicting its activity *[39,40]*. Therefore, to investigate the catalytic properties of our nanoclusters, we select the E_{ads} of the principle OER/ORR intermediate (*O, *OH, and *OOH), hereon represented as E_{*O}, E_{*OH}, and E_{*OOH}, as the target variable. However, given the inherent fluxional nature originating from the multiple heterogeneous sites on the nanocluster surface, we apply the bond-centric model of Yan et al. to describe the stability of binding sites, justifying that chemically unsaturated sites generally exhibit stronger adsorbate binding *[41]*. Additionally, to generate our initial E_{ads} database, we follow our previous investigation to identify the stable binding sites for *O, *OH, and *OOH intermediates on TM_{55} bimetallic NC alloys (Figure 7.3b) *[42]*. The E_{ads} of the intermediates ($^{*}O$, $^{*}OH$, and $^{*}OOH$) at the most stable site of the clusters are then computed as:

$$E_{ads} = E_{NC+Intermediate} - (E_{NC} + E_{Intermediate}) \quad (7.1)$$

where $E_{NC+Intermediate}$, E_{NC} , and $E_{Intermediate}$ represent the total energies of the cluster with $^*O$, $^*OH$, and $^*OOH$, bare nanoclusters, and the corresponding intermediates in the gas phase, respectively.

During intermediate adsorption, we observed surface segregation in several nanoclusters, where core atoms migrated to the surface, leading to significant deformation of the overall geometries (Figure 7.3b). After excluding these geometries, the density distribution of E_{ads} for the remaining geometries is summarized in Figure 7.3c, highlighting the variations of E_{*O}, E_{*OH}, and E_{*OOH} across 234 electrocatalysts. The E_{ads} values display a broad distribution without a consistent trend, indicating a diverse range of adsorption energies. The peak density distribution of E_{*O}, E_{*OH}, and E_{*OOH} ranges from -3.4 to -6.4 eV, from -2.5 to -4.3 eV, and from -0.9 to -2.6 eV, respectively. Compared to the bulk Pt (111) surface with E_{*O}= -4.69 eV, E_{*OH}= -2.63 eV, and E_{*OOH}= -1.64 eV *[42]*, these alloys exhibit a broad range of E_{ads}, with the majority showing stronger binding to the intermediate due to the presence of undercoordinated active sites (Figure 7.3c). Within this chemical space, we have compiled a comprehensive dataset of 234 electrocatalysts with three OER/ORR intermediates, resulting in 644 E_{ads} datapoints (211 for $^*O$, 217 for $^*OH$, and 216 for $^*OOH$), labelled as Dataset-1 in Figure 7.3d.

7.4.2 Screening Optimal E_{ads} Dataset

After acquiring our initial E_{ads} dataset (Dataset-1), the next step involves extracting the optimal E_{*O}, E_{*OH}, and E_{*OOH} values. In accordance with the Sabatier principle *[43]*, the interaction between the intermediate and the active site of the cluster should neither lead to strong chemisorption (E_{*O}= -6.85 eV as in $Pd_{45}Pt_{10}$) and weak physisorption (E_{*OH}= -0.19 eV as in $Co_{42}Fe_{13}$). Such interactions can either lead to poisoning or deactivation, reducing the overall catalytic efficiency. Therefore, to screen active bifunctional electrocatalysts from Dataset-1, we apply three criteria based on the E_{ads} values of each intermediate to eliminate inactive catalysts, as

shown in Figure 7.3d. The lower and upper extremes for E_{*_O}, $E_{*_{OH}}$, and $E_{*_{OOH}}$ were selected based on previously reported E_{ads} values for intermediates across different surfaces, including bulk surfaces, bulk nanoclusters, and subnano regime *[28,42,44-47]*. To further enhance the robustness of the dataset, we extended these limits by 0.4 eV on both ends (total of 0.8 eV), weighted by Boltzmann statistics to account for metastability-induced reactivity under experimental conditions *[28,29,48,49]*. By applying these screening criteria, we extracted our Dataset- 2 encompassing 197, 215, and 213 data points for E_{*_O}, $E_{*_{OH}}$, and $E_{*_{OOH}}$, respectively (Figure 7.3d). We highlight that this E_{ads} dataset was uniformly sampled across a wide range of chemical elements (9 shell and 26 core compositions) and three intermediates, which is crucial for developing a robust ML model.

7.4.3 Feature Design for Machine Learning

Incorporating physical and chemical properties into ML models has proven to enhance prediction accuracy. Therefore, feature engineering is a crucial step in building robust ML models while encoding the local chemical environment and quantifying the structure-activity relationships *[50]*. To effectively capture the electronic and geometric properties during intermediate adsorption on active sites of bimetallic NC alloys, we extracted a comprehensive set of features, categorized as elemental, electronic, geometric, and d-band-specific features, as detailed in Table 7.1. However, to distinguish between the core and shell compositions, features for both metals were included. For instance, the number of valence electrons for shell and core atoms is denoted as N_1 and N_2, respectively.

Table 7.1: List of Primary Features including Elemental, Electronic, Geometric, and d-band Specific Descriptors. Features Pertaining to the Shell and Core of bimetallic NCs alloys are distinguished by subscripts 1 and 2, respectively.

Feature	Shell	Core
Atomic Number	Z_1	Z_2
Atomic Weight	W_1	W_2
Sum of Atomic Weight	ΣW_1	ΣW_2
Sum of Period	ΣP_1	ΣP_2
Sum of Group	ΣG_1	ΣG_2
Atomic Radius	R_1	R_2
Covalent Radius	CR_1	CR_2
Pauling Electronegativity	PE_1	PE_2
Number of Valence Electrons	N_1	N_2
Total Number of Valence Electrons	ΣN_1	ΣN_2
Valence d Electrons	VE_1	VE_2
Total Number of Valence d Electrons	ΣVE_1	ΣVE_2
Polarizability of Atoms	P_1	P_2
Idealized d-band Filling (Size-dependent)	Idf_1	Idf_2
Size-Dependent d-band Center	ϵd_1	ϵd_2
Coupling Matrix	Vad_1^2	Vad_2^2
Difference in Pauling Electronegativity	DPE	
Site of Adsorption	S	
Coordination Number of Binding Site	CN	

Elemental features (Z_1, Z_2, W_1, W_2, ΣW_1, ΣW_2, ΣG_1, ΣG_2, R_1, R_2, CR_1, and CR_2) describe the physical properties of transition metals involved in adsorption, whereas the electronic features (ΣP_1, ΣP_2, PE_1, PE_2, N_1, N_2, ΣN_1, ΣN_2, VE_1, VE_2, ΣVE_1, ΣVE_2, P_1, and P_2) reflect the donation and acceptance capabilities of the active site. Geometric features, such as the active site of adsorption (S) and coordination number (CN), represent the local chemical environment. Additionally, summation features were included to account for geometric differences and variations in core-shell occupancy among the bimetallic NC alloys investigated.

The d-band model proposed by Hammer and Nørskov has been particularly successful in correlating systematic perturbations in the electronic structures of adsorbates, thereby explaining chemisorption trends across various transition metal surfaces *[51]*. Advanced descriptors such as the d-band center (ϵ_d), d-band filling (d_f), d-band center plus half the d-bandwidth (e_d^w), and the maximum of the d-band Hilbert-transform (e_u)

have been used in previous studies to link the electronic properties of transition metal surfaces with E_{ads} *[52,53]*. While these descriptors show potential for accelerating data-driven exploration of catalytic activity on bulk surfaces, their applicability to nanoclusters remains largely unexplored. Moreover, many standard electronic descriptors tied to d-band characteristics often require self-consistent quantum mechanism calculations. To overcome this limitation, we encode the d-band model using elemental-specific numerical values obtained from the Solid-State Table, classifying them as "d-band-specific" features (Table 7.1) *[54]*. This set included the coupling matrix (Vad_1^2, Vad_2^2), idealized d-band filling (Idf_1, Idf_2) and d-band center with respect to size (ϵd_1,ϵd_2), where $\epsilon_d = \frac{d\epsilon_d}{d \ln S_{BW}}$, and S_{BW} represents the Bulk Wigner-Seitz radius. These descriptors effectively capture the intrinsic properties of d-states, enabling their use without the need for costly DFT calculations.

To derive a rational feature set, the linear and monotonic relationship between features and target properties were evaluated using the Pearson correlation coefficient (PCC) and Spearman's rank correlation coefficient (SCC) *[55,56]*. To streamline the input dataset and avoid redundancy, pairwise features with low correlation (|PCC| < 0.9, |SCC| < 0.9) and maximum contribution towards the output were allowed to coexist for our ML analysis. As a result, the final set of independent features were chosen for the ML analysis. Using these features, the E_{ads} for each intermediate on various bimetallic nanocluster alloys can be expressed by the equation:

$$E_{*O,*OH,*OOH} = f(\Sigma W_2, CR_1, CR_2, PE_1, PE_2, DPE, \Sigma N_1, \Sigma N_2, \Sigma VE_1, VE_2, P_1, CN, Idf_1, \epsilon d_1, \epsilon d_2, Vad_1^2, Vad_2^2) \quad (7.2)$$

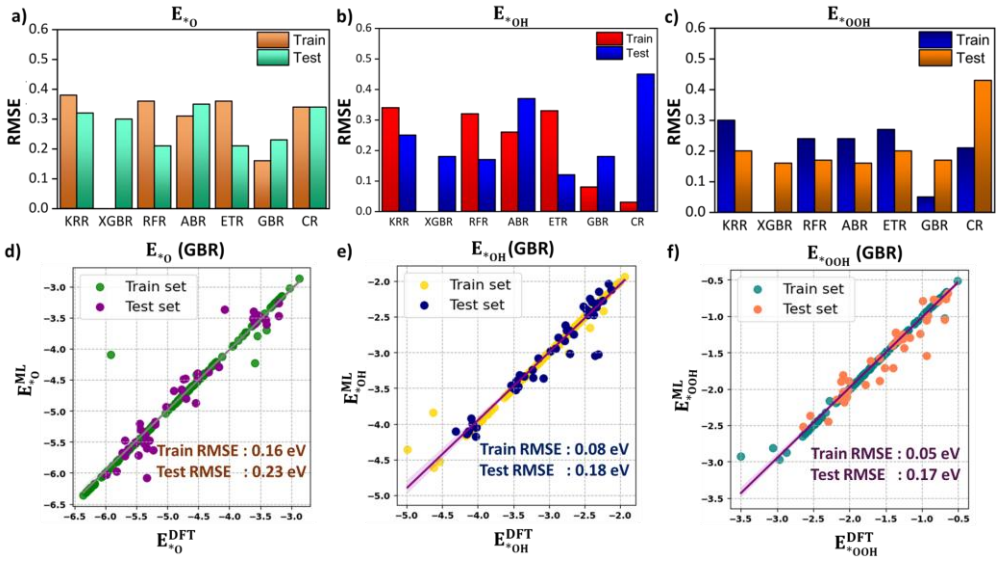


Figure 7.4: Performance evaluation of machine learning (ML) models: root mean squared error (RMSE in eV) for KRR, XGBR, RFR, ABR, ETR, GBR, and CR models for **(a)** E_{*O}, **(b)** E_{*OH}, and **(c)** E_{*OOH} for training and testing datasets. Parity plots comparing DFT-calculated and ML-predicted values for **(d)** E_{*O}, **(e)** E_{*OH}, and **(f)** E_{*OOH} using the best-performing GBR model.

7.4.4 Model Training and Evaluation

Following feature engineering, model training focuses on optimizing parameters for the selected mathematical model to minimize discrepancies between the model's validation performance and predictive accuracy. The entire Dataset-2 was divided into training (80%) and testing (20%) sets, with 5-fold cross-validation employed to ensure well-representative samples instead of relying solely on random splitting. Subsequently, seven ML models were used: kernel ridge regression (KRR), extreme gradient boosting regression (XGBR), random forest regression (RFR), adaboost regression (ABR), extra trees regression (ETR), gradient boosting regression (GBR), and catboost regression (CR), as implemented in Scikit-learn. Both ensemble and kernel-based approaches were employed to capture complex E_{ads} trends across the large chemical space. The models

were optimized using RandomSearchCV and evaluated using RMSE, and R^2 metrics, with results represented in Figure 7.4.

After training the regression models for the intermediate dataset, GBR emerged as the best predictor for E_{*O}, E_{*OH}, and E_{*OOH} values (panels a-c of Figure 7.4). In contrast, other models exhibited poor performance, exhibiting significant deviations between predicted values and true values, along with instances of overfitting and underfitting. Furthermore, the linear correlation between DFT-calculated and ML-predicted E_{ads} values, as shown in the parity plots (panels d-f of Figure 7.4), highlight the prediction accuracy of the GBR model for both training and testing data sets. In GBR, weak learners are constructed sequentially, with each new learner aiming to reduce the errors of the cumulative model from previous iterations *[58]*. Thus, the GBR model was selected to screen active bifunctional catalysts in the subsequent investigation.

7.4.5 ML Interpretability with SHAP: ML algorithms excel in capturing complex relationships within datasets but are often considered "black box" due to their lack of transparency. To improve interpretability, we use SHapley Additive exPlanations (SHAP) analysis, a model-agnostic technique based on cooperative game theory, which quantifies each feature's contribution to the model's predictions *[59]*. Figure 7.5 displays the SHAP summary and beeswarm plots, highlighting the hierarchical ranking of global feature importance in descending order and summarizing the impact of these key features on the predictions.

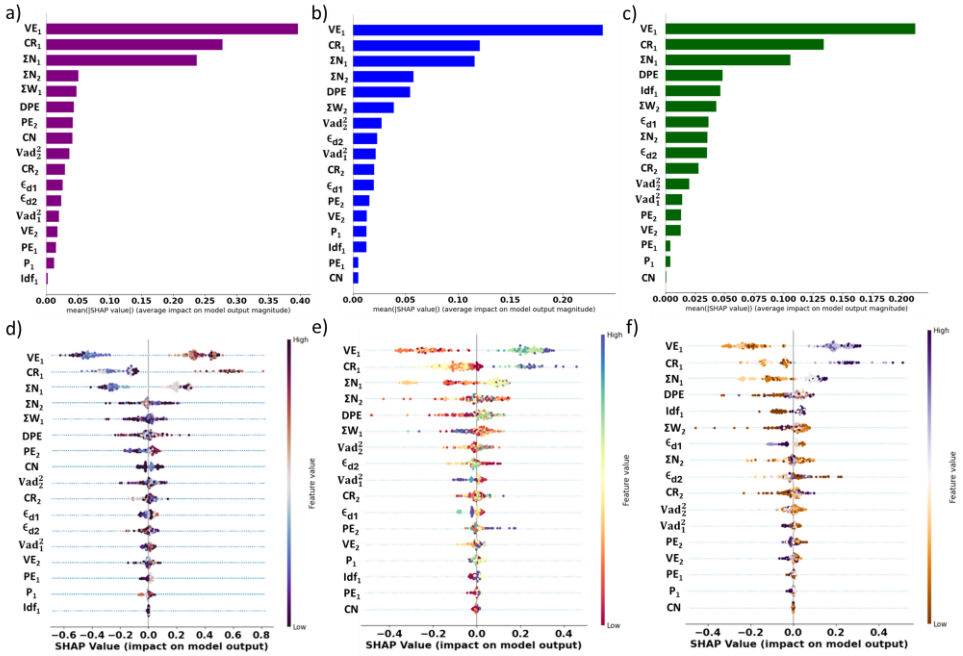


Figure 7.5: SHAP Analysis: Global interpretation of the best-performing ML models for extracted features in predicting (a) E_{*_O}, (b) $E_{*_{OH}}$, and (c) $E_{*_{OOH}}$. The Beeswarm plot displays an information-dense summary of how the most important features affect the predictions of (d) E_{*_O}, (e) $E_{*_{OH}}$, and (f) $E_{*_{OOH}}$. A solitary dot on the feature row represents each data point with its SHAP value on the x-axis. The color gradient along the y-axis represents the feature values, ranging from low to high.

Interestingly, VE_1, CR_1, and ΣN_1 consistently ranked among the top three features in the bar plots, underscoring the pivotal role of valence d-electrons (electronic features) in intermediate adsorption on bimetallic nanoparticle alloys (panels a-c of Figure 7.5). Additionally, higher VE_1 and ΣN_1 values positively correlate with E_{ads}, suggesting that increased valence electrons enhance interatomic repulsion with adsorbed intermediates, thereby reducing the adsorption strength (panels d-f of Figure 7.5). Similarly, higher values of CR_1 exhibit a positive correlation with the E_{ads}, attributed to the repulsion between the incoming adsorbate and the delocalized metallic radii of transition metals. Notably, ΣN_1 and ΣN_2 also exert a non-negligible impact on E_{ads}, where ΣN_1 predominantly drives adsorption trends, and

ΣN_2 impacts both lower and higher impacts both lower and higher extremes of E_{ads}, reflecting a nuanced interplay. This underscores the synergistic interaction between the electronic properties of core and shell atoms, where polarization between these regions modulates adsorption strength, capturing the intricate interdependence of local atomic environments on catalytic behavior. Unlike subnano clusters where the d-band specific features (Idf_1,Idf_2,ϵd_1,ϵd_2,Vad_1^2,Vad_2^2) significantly influence E_{ads} *[60]*, they play a less dominant role in the core-shell TM_{55} nanoclusters. This distinction underscores that while subnano clusters resemble single-atom catalysts (SACs) behavior, the catalytic properties of core-shell TM_{55} magic nanoclusters align with single-cluster catalyst, where intermediate adsorption is predominantly governed by the electronic and elemental characteristics of the ensemble active site.

7.4.6 Predicting Active Electrocatalysts: To validate the accuracy of our best-performing GBR model, we predicted the E_{*_O}, $E_{*_{OH}}$, and $E_{*_{OOH}}$ values of all the catalysts falling within our defined Sabatier range. Subsequently, we screened 188 nanoclusters with common E_{*_O}, $E_{*_{OH}}$, and $E_{*_{OOH}}$ values from ML-predicted datasets, and derived the OER and ORR mechanisms, compiled as Dataset-3. For this, we utilized the computational hydrogen electrode (CHE) model proposed by Nørskov and co-workers and calculated the reaction energy at 0 and 1.23 V following the elementary steps outlined below *[61]*:

In an acidic medium, the OER is considered a four-step process as follows:

$$* + H_2O\,(l) \rightarrow OH^* + (H^+ + e^-) \qquad (\Delta \mathbf{R_1})$$

$$OH^* \rightarrow O^* + (H^+ + e^-) \qquad (\Delta \mathbf{R_2})$$

$$O^* + H_2O\,(l) \rightarrow OOH^* + (H^+ + e^-) \qquad (\Delta \mathbf{R_3})$$

$$OOH^* \rightarrow O_2(g) + (H^+ + e^-) + * \qquad (\Delta \mathbf{R_4})$$

And the $4e^-$ ORR is the reverse reaction of the OER with four-step as follows:

$$O_2(g) + (H^+ + e^-) + * \rightarrow OOH^* \qquad (\Delta \mathbf{R_5})$$

$$OOH^* + (H^+ + e^-) \rightarrow O^* + H_2O\,(l) \qquad (\Delta \mathbf{R_6})$$

$$O^* + (H^+ + e^-) \rightarrow OH^* \qquad (\Delta R_7)$$

$$OH^* + (H^+ + e^-) \rightarrow H_2O\,(l) + * \qquad (\Delta R_8)$$

The catalytic activity of the nanocluster alloys was evaluated by calculating the overpotential values (η) of the rate-determining step (RDS) at 1.23 V, using the equation:

$$\eta_{OER} = \frac{\max(\Delta R_1, \Delta R_2, \Delta R_3, \Delta R_4)}{e} \qquad (7.3)$$

$$\eta_{ORR} = \frac{\max(\Delta R_5, \Delta R_6, \Delta R_7, \Delta R_8)}{e} \qquad (7.4)$$

7.4.7 Screening OER/ORR Electrocatalysts: Figure 7.6 illustrates the irregular trends in OER and ORR activity observed with variations in the core-shell compositions of bimetallic NC alloys. In fuel cells, OER corresponds to the charging process, while ORR is linked to the discharging process. For the OER, $Au_{48}W_7$, $Pt_{46}Mn_9$, and $Au_{48}Mo_7$ were predicted as the active electrocatalysts (Figure 7.6) with low η_{OER}^{ML} values of 0.33, 0.40, and 0.42 V, respectively, with RDS involving $OH^* \rightarrow O^*$ (ΔR_2) for $Au_{48}W_7$, and $O^* \rightarrow OOH^*$ (ΔR_3) formation for $Pt_{46}Mn_9$ and $Au_{48}Mo_7$. These η_{OER}^{ML} values are lower (or comparable) to traditional active OER catalysts, such as RuO_2 (η_{OER} = 0.42 V) and IrO_2 (η_{OER} = 0.56 V). Interestingly, $Au_{48}W_7$ also exhibits superior OER activity (η_{OER}^{ML} = 0.33 V) compared to previously theoretically reported carbon-based single-atom catalysts like N/C-coordinated graphene (Co-doped), C_3N_4 (Ni-doped), graphdiyne (Co-doped), covalent organic framework (Cu-doped), and metal–organic framework (Co-doped), which show η_{OER} values of 0.46 V, 0.96 V, 0.64 V, 0.69 V, and 0.29 V, respectively *[63]*. In contrast, $Ir_{48}Hf_7$, $Co_{42}V_{13}$, and Au_{55} are predicted as inactive OER electrocatalysts with significantly higher η_{OER}^{ML} values of 2.64 V, 2.66 V, and 3.35 V respectively, with the RDS involving $O^* \rightarrow OOH^*$ (ΔR_3) for $Ir_{48}Hf_7$, and $Co_{42}Fe_{13}$ and $OH^* \rightarrow O^*$ (ΔR_2) for Au_{55}.

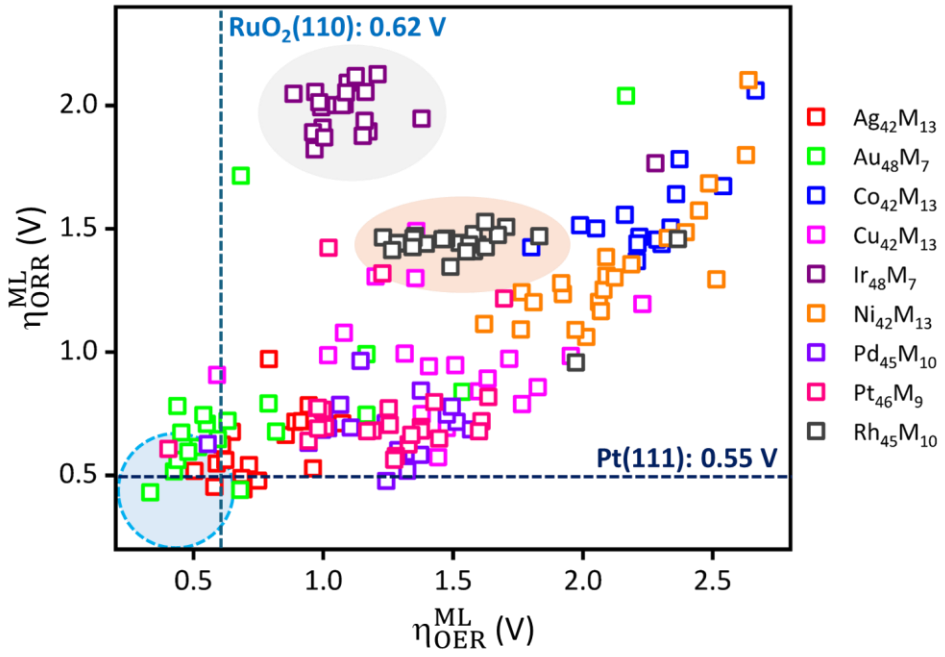


Figure 7.6: Distribution plot of overpotential for catalysts in OER (η_{OER}^{ML}) and ORR (η_{ORR}^{ML}), derived from ML-based screening. The intersection of the dark and light blue lines indicates catalysts with low η_{OER}^{ML} and η_{ORR}^{ML}, serving as a reference point for identifying active bifunctional electrocatalysts, highlighted within the blue circle. The blue horizontal and vertical lines indicate our theoretically computed overpotential values for ORR (0.55 V) and OER (0.62 V), corresponding to the Pt (111) and RuO_2 (110) surfaces, respectively.

For the ORR, $Au_{48}W_7$, $Au_{48}Ni_7$, $Ag_{42}Re_{13}$, and $Ag_{42}Ru_{13}$ were predicted as active electrocatalysts (Figure 7.6) with lower η_{ORR}^{ML} values of 0.43, 0.44, 0.44, and 0.45 V, respectively, with the RDS involving $O_{2(g)} \rightarrow OOH^*$ (ΔR_5) for $Au_{48}W_7$, and $Au_{48}Ni_7$, and $OH^* \rightarrow H_2O$ (l) (ΔR_8) for $Ag_{42}Re_{13}$ and $Ag_{42}Ru_{13}$. In contrast, $Ir_{48}Pt_7$, Au_{55}, and $Co_{42}Fe_{13}$ are categorized as poor electrocatalysts with higher η_{ORR}^{ML} values of 2.13 V, 2.39 V, and 4.60 V, respectively. For majority of the electrocatalysts, the ORR predominantly involves the formation of H_2O (l) via the $OH^* \rightarrow H_2O$ (l) (ΔR_8) formation; however, for the OER, the RDS for most electrocatalysts is the $O^* \rightarrow OOH^*$ (ΔR_5) formation. Interestingly, the active

electrocatalysts are predominantly associated with Au and Ag shell compositions, reshaping the OER/ORR activity volcano plots by shifting the apex from benchmarked systems like RuO_2 and Pt(111) surface to $Au_{48}M_7$ and $Ag_{42}M_{13}$ based nanoclusters exhibiting DRC and ICO geometry *[64]*. Additionally, we computed the OER and ORR overpotential of the top-performing electrocatalysts under implicit solvation conditions and found that the Au-based nanoclusters consistently exhibit low overpotential and exhibit low overpotentials and superior catalytic performance (Table 7.2). Apart from these clusters, $Pt_{46}Mn_9$ and $Pd_{45}W_{10}$ bimetallic NC alloys have also emerged as active catalysts for OER and ORR activity, with overpotential values η_{OER}^{ML} = 0.40 V and η_{ORR}^{ML} = 0.48 V, respectively.

Table 7.2: Computed OER and ORR overpotential values (in V units) for the rate-determining step (RDS) at 1.23 V for the top five electrocatalysts under gas-phase and implicit solvent-phase conditions.

Catalysts	OER		Catalysts	ORR	
	Gas	Solvent		Gas	Solvent
$Au_{48}W_7$	0.33	0.66	$Au_{48}W_7$	0.43	0.64
$Au_{48}Mo_7$	0.42	0.38	$Au_{48}Ni_7$	0.44	0.49
$Au_{48}Ag_7$	0.44	0.62	$Ag_{42}Re_{13}$	0.44	0.84
$Au_{48}Re_7$	0.44	0.58	$Ag_{42}Ru_{13}$	0.45	0.82
$Pt_{46}Mn_9$	0.40	1.49	$Pd_{45}W_{10}$	0.48	0.85

7.4.8 Screening Bifunctional Electrocatalysts: To identify an active bifunctional electrocatalyst that efficiently facilitates both reactions, we computed the widely accepted descriptor, $\eta_{Bi} = \eta_{OER}^{ML} + \eta_{ORR}^{ML}$ *[65]*, represented near the intersection of the Pt(111) and RuO_2 lines in Figure 7.6. Among the catalysts, $Au_{48}W_7$ emerged as the best bifunctional electrocatalyst with a low η_{Bi} = 0.76 V, outperforming the benchmark Ru/C systems (η_{Bi} = 1.01 V) *[66]*. Additionally, $Ag_{42}Ru_{13}$, $Pt_{46}Mn_9$, $Au_{48}Re_7$, and $Au_{48}Mo_7$ emerged as active bifunctional electrocatalysts with lower η_{Bi} values of 1.03, 1.00, 1.00, and 0.94 V, respectively. The reaction energy diagram for OER and ORR at 0 V and 1.23 V for these electrocatalysts is

provided in panels a and b of Figure 7.7. In contrast, $Ni_{42}Sc_{13}$ and $Co_{42}V_{13}$ were identified as inactive bifunctional electrocatalysts, exhibiting significantly higher η_{Bi} = 4.42 V and 4.72 V, respectively. Additionally, $Ir_{48}M_7$, $Co_{42}M_{13}$, $Rh_{45}M_{10}$, and $Ni_{42}M_{13}$ based-electrocatalysts were identified as inactive bifunctional electrocatalysts with significantly higher η_{OER}^{ML} and η_{ORR}^{ML} values, rendering them unsuitable for both OER and ORR.

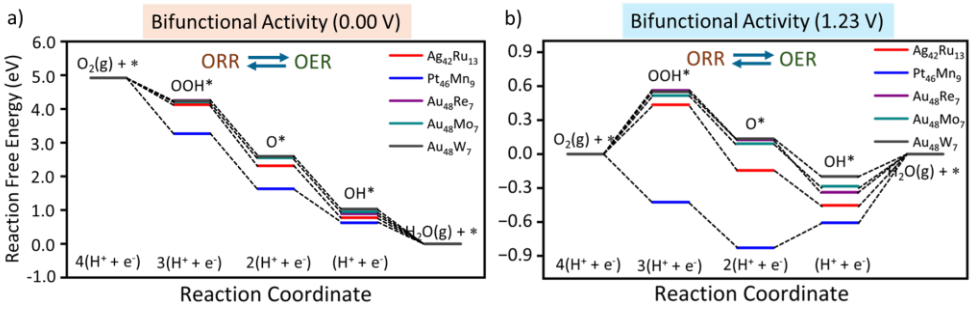


Figure 7.7: Reaction energy diagrams for OER and ORR mechanisms of the top five bifunctional electrocatalysts at (a) 0 V and (b) 1.23 V, respectively.

For enhanced OER and ORR performance in fuel cells, the catalysts at the apex of the activity volcano should exhibit near optimal E_{ads} for each intermediate. The DFT-calculated and ML-predicted E_{*O}, E_{*OH}, and E_{*OOH} for the top five bifunctional catalysts align closely, highlighting the accuracy of the GBR model in capturing non-monotonic ORR activity across nanoclusters (Figure 7.8). Notably, the E_{*O}, E_{*OH}, and E_{*OOH} values of the active electrocatalyst fall within the optimal E_{ads} interval window of $-4.92 < E_{ads} < 0.00$ eV, where nanoclusters demonstrate optimum binding affinity and outperform the $RuO_2(110)$ and Pt(111) surface. This E_{ads} range offers a pathway to identify active electrocatalysts that could disrupt conventional scaling relationships and volcano plots for the OER/ORR, potentially replacing the Ru and Ir/Pt and Pd elements at the peak.

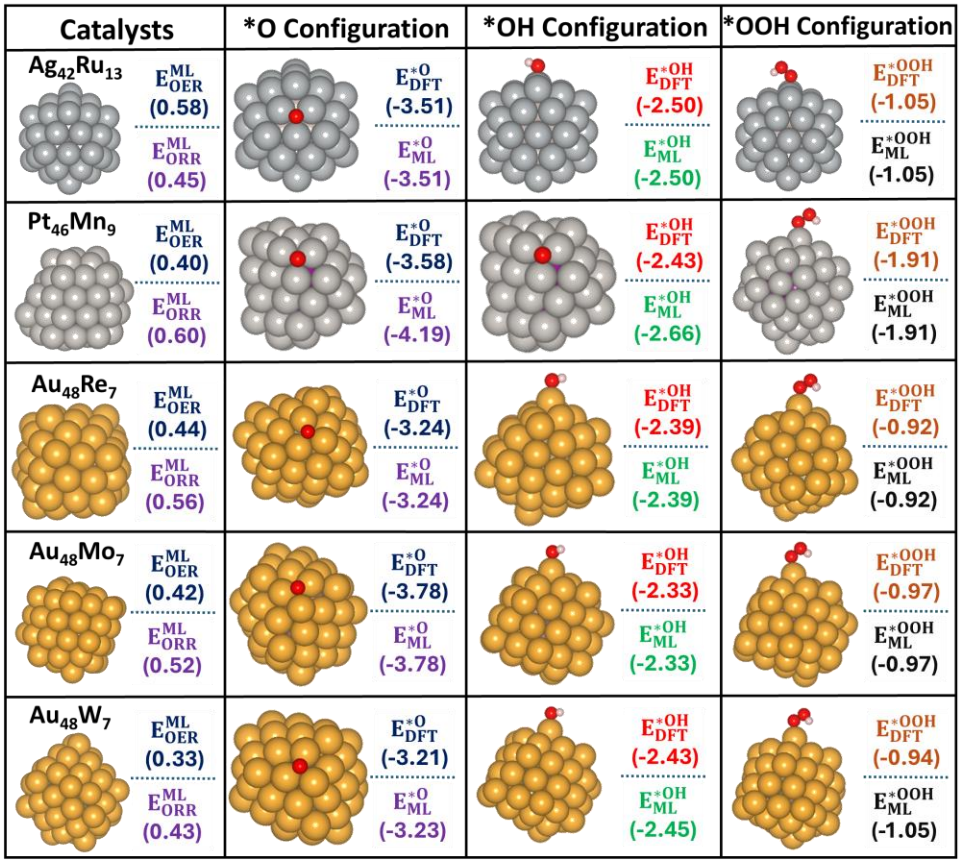


Figure 7.8: DFT-calculated and ML-predicted E_{*O}, E_{*OH}, and E_{*OOH} (in eV) for the top five bifunctional electrocatalysts. The ML-derived overpotential values for OER (η_{OER}^{ML}) and ORR (η_{ORR}^{ML}), determined from the thermodynamically rate-determining step (RDS), were derived using the adsorption energies from the ML dataset (Dataset-3).

Note that we employed the standard PBE functional for all DFT calculations. Although the DFT+U method effectively corrects self-interaction errors in transition metal systems with localized d-electrons, its accuracy relies on selecting appropriate system-specific U_{eff} values. However, due to the lack of experimental data for our nanoclusters, determining reliable U_{eff} values for different transition metals remains challenging. Further, to assess the impact of the Hubbard correction, we evaluated η_{ORR} on bulk Pt (111) surface by varying U_{eff} from 2 to 6 eV (Table 7.3). The calculated η_{ORR} using PBE functionals (0.55 V) and moderate U_{eff} values (~0.55 V) align well with the reported value of 0.45 V

by Nørskov et al. *[61]*, while U_{eff} = 6 eV leads to significant deviation. More importantly, although η_{ORR} remain similar, the reaction energies of individual ORR steps vary considerably with changing U_{eff} (Figure 7.9), contrasting with a recent report on pure nanoclusters where OH adsorption was largely unaffected *[67]*.

Table 7.3: Comparison of overpotential values η_{ORR} with varying U_{eff} values for bulk Pt (111) surface.

U_{eff} (eV)	η_{ORR} (V)
PBE	0.55
2	0.54
3	0.57
4	0.54
5	0.54
6	2.36

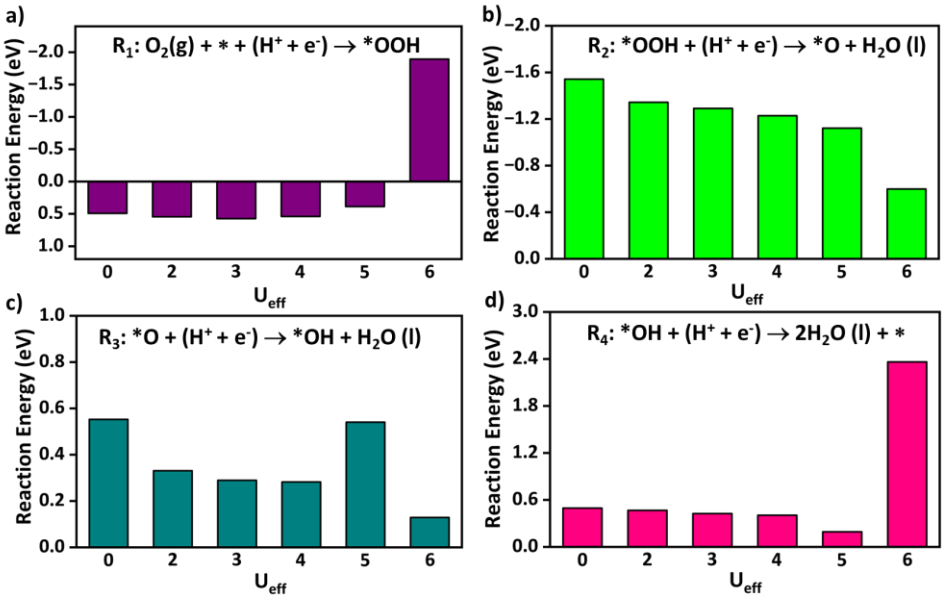


Figure 7.9: Reaction energies (in eV) for the elementary steps of the ORR on the bulk Pt (111) surface as a function of U_{eff} (ranging from 0 to 6 eV) at 1.23 V (equilibrium potential). Adsorbed species are denoted with an asterisk (∗), while species without the asterisk represent the bare Pt (111) surface.

Table 7.4: Comparison of OER and ORR overpotential values for the rate-determining step (RDS) at 1.23 V for the top five electrocatalysts with DFT and DFT+U with $U_{eff} = 2.0$ eV.

Catalysts	OER		Catalysts	ORR	
	DFT	DFT + U		DFT	DFT + U
$Au_{48}W_7$	0.33	0.55	$Au_{48}W_7$	0.43	0.59
$Au_{48}Mo_7$	0.40	2.54	$Au_{48}Ni_7$	0.44	1.88
$Au_{48}Ag_7$	0.42	2.06	$Ag_{42}Re_{13}$	0.44	0.68
$Au_{48}Re_7$	0.44	0.48	$Ag_{42}Ru_{13}$	0.45	0.93
$Pt_{46}Mn_9$	0.44	0.53	$Pd_{45}W_{10}$	0.48	1.26

We also tested U_{eff} = 2.0 eV, following Piotrowski et al, and compared DFT + U and standard PBE results (Table 7.4). We found that applying U_{eff} tends to overestimate OER and ORR overpotentials at the RDS for most of the top-performing bifunctional electrocatalysts. Given the sensitivity to U_{eff}, the absence of benchmarking data, and the good agreement of PBE results with previous reports *[61,68]*, we resorted to the standard DFT. This ensures consistent and reliable trends across all calculations without introducing uncertainties from arbitrary U selection.

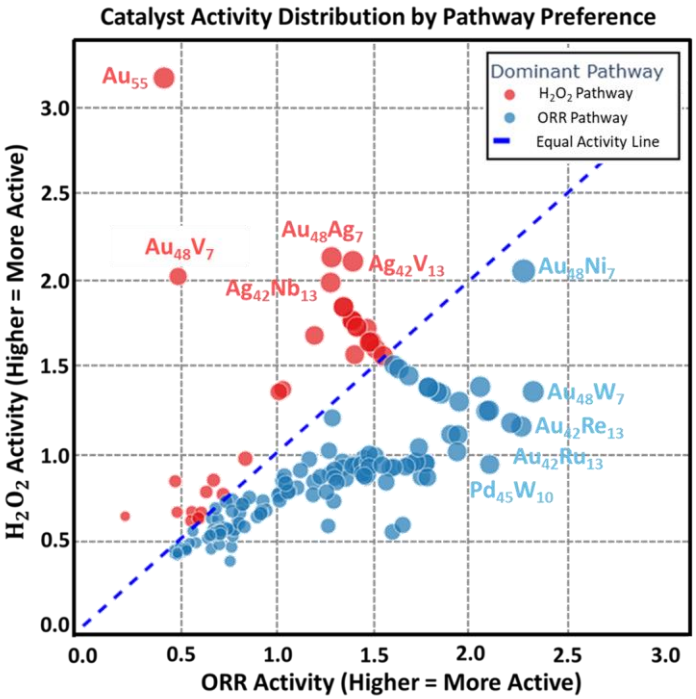


Figure 7.10: Selectivity plot illustrates pathway preference for ORR (η_{ORR}^{ML}) and H_2O_2 ($\eta_{H_2O_2}^{ML}$) formation, classifying catalysts into two dominant pathways. Catalyst sizes are scaled by activity strength, annotated with key properties, and compared against a diagonal benchmark line indicating equal activity.

7.4.9 Steering Reaction Pathway Selectivity

Note that the two-electron process leading to H_2O_2 formation is thermodynamically less favored due to its lower equilibrium potential but is kinetically preferred over the four-electron pathway to H_2O, as it involves fewer proton-electron transfers and avoids the constraints of scaling relations *[69,70]*. To address the selectivity challenge between these two competing reaction pathways, we normalize the key performance metrics of ML-predicted ORR (η_{ORR}^{ML}) and hydrogen peroxide ($\eta_{H_2O_2}^{ML}$) activity indicators, enabling the classification of catalysts into two dominant pathways (Figure 7.10). Among 188 bimetallic NC alloys, 34 favor the H_2O_2 pathway and 154 favour H_2O pathway, with Au_{55} ($\eta_{H_2O_2}^{ML}$ = 0.32 V) and $Au_{48}Ni_7$ (η_{ORR}^{ML}= 0.44 V) demonstrating superior activity for their respective pathways. Upon deeper investigation, catalysts such as Au_{55}, $Au_{48}V_7$, $Au_{48}Ag_7$, $Ag_{42}Nb_{13}$, and $Ag_{42}V_{13}$, which favor the H_2O_2 formation, exhibit lower $E_{*_{OOH}}$ values compared to those such as $Au_{48}Ni_7$, $Au_{48}W_7$, $Au_{42}Re_{13}$, $Au_{42}Ru_{13}$, and $Pd_{45}W_{10}$ which prefers H_2O formation.

7.5 Conclusion

In summary, we present a combined DFT-ML approach to evaluate the ORR and OER performance of 234 core-shell TM_{55} nanoclusters alloys composed of 9 shell metals and 26 core metals, exhibiting different morphologies. Our trained ML model accurately maps the shift in the apex of the OER and ORR activity volcano from RuO_2 and Pt to Au/Ag-based nanoclusters, which predominantly exhibit different geometries. This study established the single-cluster catalyst identity of TM_{55} nanoclusters, demonstrating that adsorption at local active sites is governed by the

electronic and elemental characteristics of the metal atoms rather than d-states. Among the screened candidates, our ML model predicts $Au_{48}W_7$ as the most promising bifunctional electrocatalyst, exhibiting a significantly lower overpotential (η_{Bi} = 0.76 V) compared to the benchmark Ru/C system (η_{Bi} = 1.01 V). Additionally, we address the selectivity challenge between the two-electron and four-electron ORR pathways, showing that catalysts with lower $E_{*_{OOH}}$ values favor H_2O_2 formation over H_2O. Unlike previous studies, which consider icosahedral geometry as the global minima state for TM_{55} nanoclusters, our ML model highlights the significant contributions and superior activity of coexisting low-energy isomers with different geometry. This data-driven approach offers an effective strategy for screening bimetallic TM_{55} alloys, paving the way for the rational design of efficient bifunctional cluster catalysts.

Section: B

7.6 Introduction

Traditional ML models rely on human-engineered descriptors, such as group, period, electronegativity, ionic radius, partial charge, and d-band characteristics, to numerically represent atomic environments *[71]*. Advanced descriptors-based frameworks, including atom-centered symmetry functions (ACSF) *[72]*, smooth overlap of atomic positions (SOAP) *[73]*, and Coulomb matrices *[74]*, attempt to preserve rotational and translational invariance while encoding chemical and structural information. Despite their interpretability, these handcrafted features are constrained by human intuition and may overlook latent structure-property relationships.

To overcome these limitations, deep learning (DL) approaches have gained prominence for their ability to autonomously learn hierarchical feature representations directly from raw structural inputs. Graph neural networks (GNNs) have emerged as state-of-the-art (SOTA) architectures for material representation, leveraging the natural graph topology of atomic structures, where atoms are nodes and bonds are edges *[75-77]*. Unlike fixed

descriptors, GNNs employ trainable embeddings that evolve through message-passing and neural updates during training, enabling them to capture both local coordination environments and global topological information. Since the introduction of models like CGCNN *[78]* and SchNet *[79]* in 2018, GNN variants such as MEGNet *[80]*, iCGCNN *[81]*, and ALIGNN *[82]* have demonstrated excellent performance across molecules, crystalline solids, and surfaces. These models differ in their convolutional operators, edge/node feature encodings, and readout schemes. Nonetheless, their application to nanoclusters remains nascent due to challenges associated with irregular, non-periodic geometries, geometric invariance, and graph construction methods.

Herein, we evaluate the performance of three SOTA GNN architectures for predicting adsorption energies (E_{ads}) of key ORR intermediates (*O, *OH, and *OOH) on bimetallic TM_{55} nanocluster alloys, exhibiting diverse morphologies and core-shell compositions (Figure 7.11). Our curated DFT dataset spans a large E_{ads} value from -0.26 to -8.68 eV, ensuring diversity while training the powerful GNN models. Each GNN is systematically optimized for our non-crystalline dataset through parameter tuning and scalability analysis. By coupling GNN-based predictions with DFT validation, we identify three active nanocluster electrocatalysts $Pd_{45}Sc_{10}$, $Pd_{45}Nb_{10}$, $Au_{48}Cr_7$, exhibiting low theoretical overpotentials and enhanced ORR activity. To gain mechanistic insights, we correlate catalytic performance with physics-based descriptors such as the coupling matrix (V_{ad}^2) and the d-band center (ϵ_d). Statistical analysis of these quantities reveals optimal descriptor ranges associated with low E_{ads} values, supporting rapid, descriptor-informed screening. Further analysis of the projected density of states (PDOS) underscores the role of d-p orbital hybridization in modulating adsorption strength and electronic properties. This integrated DFT-GNN framework offers a robust pathway for accelerating the discovery of efficient ORR electrocatalysts by combining

predictive accuracy, chemical interpretability, and computational scalability.

7.7 Computational Details

All the plane-wave spin-polarized density functional calculations were conducted using the Vienna *ab initio* simulation package (VASP) *[34]* with the Perdew–Burke–Ernzerhof (PBE) *[83]* functional within the generalized gradient approximation (GGA) *[84]*. Ion core and valence electron interactions were described using the projector augmented wave (PAW) method. For relaxation calculations, kinetic energy cut-offs of 500 eV with energy convergence criteria of 10^{-4} eV were employed, and geometric relaxation continued until forces on atoms were smaller than 0.02 eV/Å. Gaussian smearing with a σ value of 0.2 eV was applied, and the Γ-centered (1 × 1 × 1) K-point grids were considered for the sampling of the Brillouin zone in congruence with the previous reports. A higher (3 × 3 × 3) *k*-mesh was used to calculate the density of states (DOS). A sufficiently large box with dimensions 20 × 20 × 20 $Å^3$ was utilized to optimize the distinct geometries to avoid the possibility of spurious interaction between the adjacent images along each axis. Further, crystal orbital Hamilton population (COHP) *[85]* analysis was employed using the LOBSTER package.

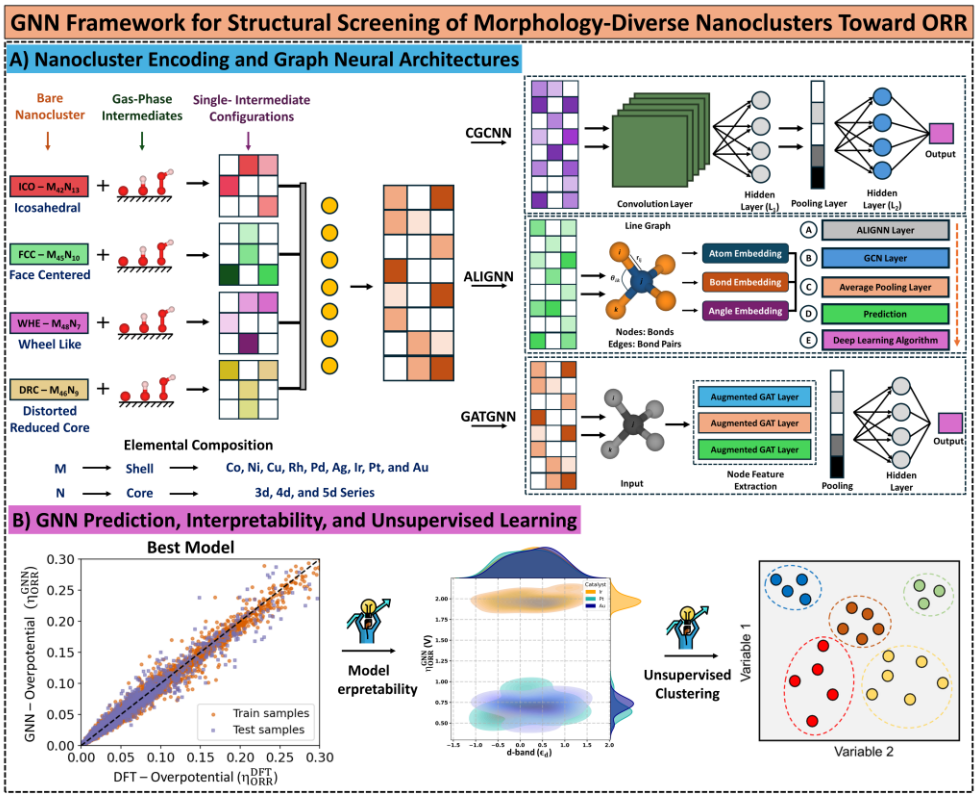


Figure 7.11. Schematic illustration of the GNN workflow. The process involves: (A) encoding adsorption of a single intermediate on bare bimetallic TM_{55} nanoclusters alloys into three GNN architectures - crystal graph convolution neural network (CGCNN), atomistic line graph neural network (ALIGNN), and deeper global attention graph neural network (DeeperGATGNN), to predict the adsorption energy (E_{ads}) of *OOH, *O, and *OH intermediates. (B) Prediction of E_{ads} from the best GNN model comparing the GNN-predicted overpotential values (η_{ORR}^{GNN}) and DFT-calculated overpotential values (η_{ORR}^{DFT}) for ORR. The model interpretability includes deriving the complex-relationships between the d-band descriptors and η_{ORR}^{GNN}. Statistical analysis including unsupervised clustering-based analysis to derive the morphology-based ORR activity.

7.8 Results and Discussion

7.8.1 Data Analysis for Deep Learning Training. Initially, we select TM_{55} nanoclusters as a parent system to investigate their morphology-dependent ORR activity while encoding them in the graph networks. Using

DFT calculations, we created our initial database of 55-atom nanoclusters (NCs), which represent a distinct class of 'magic-sized' clusters known for their enhanced stability during experimental synthesis *[33].* All the plane-wave spin-polarized DFT calculations were conducted using the Vienna *ab initio* simulation package (VASP) with the Perdew-Burke-Ernzerhof (PBE) functional within the generalized gradient approximation (GGA). Owing to their closed-shell configurations, these clusters exhibit a bulk-like environment at their core, while a subset of surface atoms simultaneously displays pronounced surface effects. Previous investigations have often identified high-symmetry icosahedral (ICO) motifs as stable geometries due to their dense atomic packing and reduced surface energy *[33]*. However, the fluxional behaviour at key active sites introduces structural sensitivity, arising from the interplay between morphological and electronic factors. This results in varying geometries of putative global minima (pGMs) for different elements based on NCs. Therefore, to capture the structure sensitivity of the reaction across different ensemble sites, we considered the diverse geometries exhibited by late transition metal (TM) 55-atom NCs. To explore their structure-activity relationship, we considered 3d (Co, Ni, Cu), 4d (Rh, Pd, Ag), and 5d (Ir, Pt, Au) series-based nanoclusters exhibiting distinct geometries: high-symmetry icosahedral (ICO) structures for Co_{55}, Ni_{55}, Cu_{55}, and Ag_{55}; bulk-like face-centered (FCC) structures for Rh_{55}, and Pd_{55}; a hexagonal closed-packed wheel-like (WHE) structure for Ir_{55}; and distorted reduced-core (DRC) structures for Pt_{55} and Au_{55}, as illustrated in Figure 7.12a.

Interestingly, the geometries of different TM_{55} NCs exhibit distinct core atom arrangements, ranging from 7 atoms in the WHE structure, 9 atoms in the DRC, 10 atoms in the FCC, to 13 atoms in the ICO configurations *[33]*. To design core-shell NCs, the core atoms of these TM_{55} clusters were systematically substituted with 3d, 4d, and 5d TMs, resulting in bimetallic compositions with 9 TM atoms in the shell and 26 TM in the core, as illustrated in Figure 7.12b. This strategy yielded a diverse dataset of

bimetallic alloy nanoclusters with compositions as $Co_{42}M_{13}$, $Ni_{42}M_{13}$, $Cu_{42}M_{13}$, $Rh_{45}M_{10}$, $Pd_{45}M_{10}$, $Ag_{42}M_{13}$, $Ir_{48}M_7$, $Pt_{46}M_9$, and $Au_{48}M_7$, which were selected for further investigation of their ORR activity. Elements such as Tc, Cd, La, and Hg were excluded from the core due to their radioactivity and toxicity concerns *[38]*. In total, we initially constructed 234 electrocatalysts, including 9 monometallic and 225 bimetallic alloy nanoclusters for subsequent evaluation of their electrocatalytic performance. Notably, initial structural optimizations revealed no signs of surface segregation or core melting, highlighting the intrinsic thermodynamic stability of the selected alloy configurations.

To further evaluate the ORR activity of these bimetallic electrocatalysts, we calculate the adsorption energy (E_{ads}) of key ORR intermediates, which serve as essential descriptors for catalytic performance in accordance with the Bell-Evans-Polanyi (BEP) principle *[39,40]*. Specifically, we considered the adsorption energies of $^*OOH$, $^*O$, and $^*OH$, denoted as E_{*_O}, $E_{*_{OH}}$, and $E_{*_{OOH}}$, as the target variables. Given the inherent fluxionality of nanoclusters and the presence of low-symmetry adsorption sites, we employed a bond-centric coordination model to characterize local atomic environments and assess site-specific adsorption behaviour *[41]*. This approach effectively captures the role of under-coordinated surface atoms, known for their enhanced binding affinity due to their unsaturation, and offers a physically meaningful descriptor of adsorption stability in morphologically complex systems. Additionally, building on our previous investigations, we identified the thermodynamically most favorable adsorption.

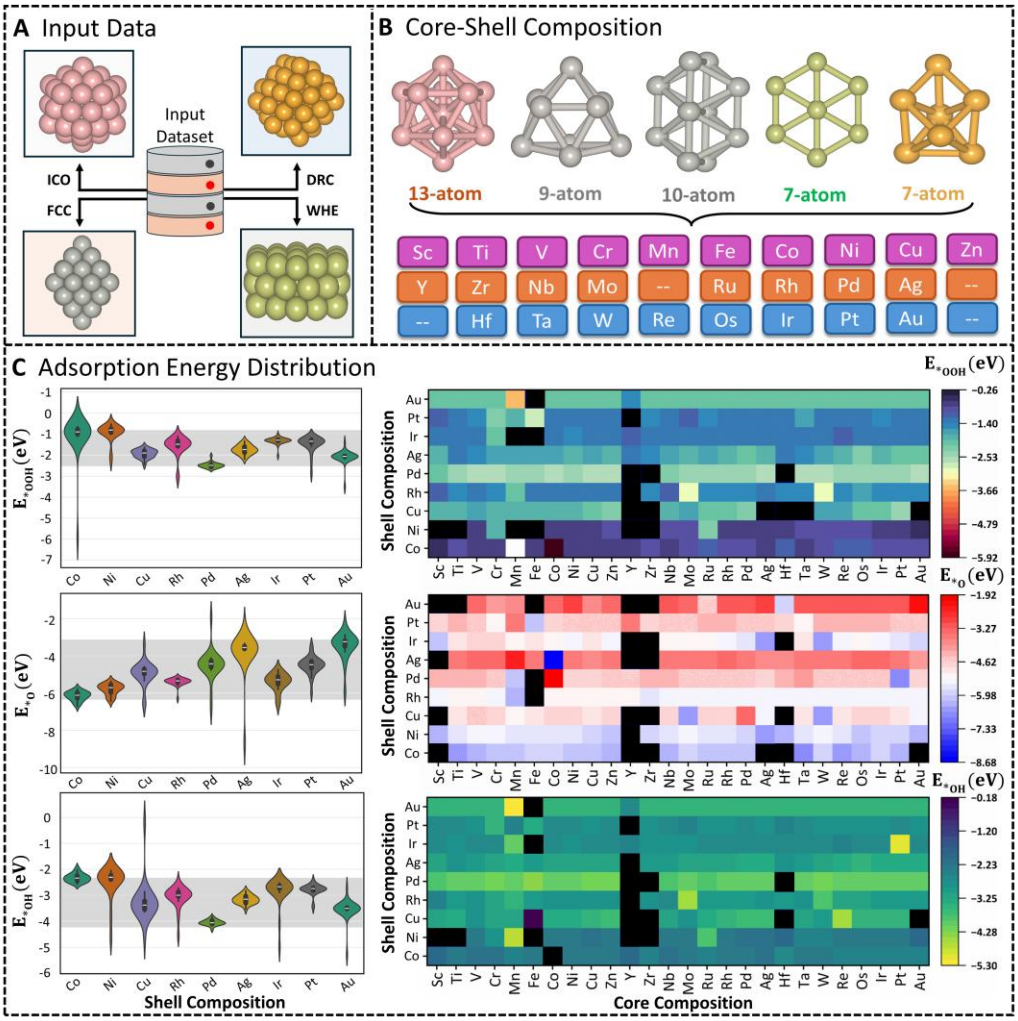


Figure 7.12: Curation of the Input Dataset for GNN Models Evaluation. (a) Schematic illustration of a closed-shell TM_{55} nanoclusters (TM = Co, Ni, Cu, Rh, Pd, Ag, Ir, Pt, and Au), featuring diverse geometries including ICO, FCC, WHE, and DRC motifs. (b) Geometric arrangements and elemental composition at the core of the TM_{55} nanoclusters, with empty cells indicating core-composition excluded from this study. (c) Violin plots depicting the distribution of E_{*O}, E_{*OH}, and E_{*OOH} across various core-shell TM_{55} combinations, in eV units. The distributions are estimated using Gaussian kernel density estimation (KDE), and the accompanying heatmaps summarize the compositional overview of adsorption energies (E_{ads}), with a color gradient representing the range from highest to lowest values. Note that the blacked-out entries indicate the data points excluded

from the analysis due to significant structural distortions following intermediate adsorption.

adsorption sites for ORR intermediates across a series of TM_{55} bimetallic nanoclusters. The E_{ads} values at these preferred sites for intermediates are computed as follows:

$$E_{ads} = E_{NC+Intermediate} - (E_{NC} + E_{Intermediate}) \quad (7.5)$$

where $E_{NC+Intermediate}$, E_{NC}, and $E_{Intermediate}$ represent the total energies of the cluster with $^*O$, $^*OH$, and $^*OOH$, bare TM_{55} nanoclusters, and the corresponding intermediates in the gas phase, respectively. During intermediate adsorption, several nanoclusters exhibited surface segregation, wherein core atoms migrated/exposed toward the surface, leading to pronounced structural distortions; these unstable geometries were excluded from further analysis.

The energy distribution of E_{ads} for the remaining stable configurations is summarized in Figure 7.12c, showcasing the variation of E_{*OOH}, E_{*O}, and E_{*OH} across the evaluated electrocatalysts. Three key observations emerged from the distribution of E_{ads} across the dataset: (1) The E_{ads} values span a broad range with no uniform trend, highlighting the structural and compositional diversity of the nanoclusters. The highest density of E_{ads} is observed in the following intervals: -0.9 to -2.6 eV for $^*OOH$, -3.4 to -6.4 eV for $^*O$, -2.5 to -4.3 eV for $^*OH$. (2) When compared to E_{ads} on the Pt(111) surface (E_{*OOH} = -1.64 eV, E_{*O} = -4.69 eV, and E_{*OH} = -2.63 eV) *[42]*, a majority of the nanoclusters demonstrate stronger binding to ORR intermediates, attributed to the presence of undercoordinated active sites on the surfaces. And (3) Notably, a subset of nanoclusters exhibits strong negative E_{ads} values, indicating over-binding of intermediates. This overbinding deviates from the Sabatier optimum *[43]*, and can lead to surface poisoning, ultimately lowering the catalytic efficiency in ORR. We highlight that this dataset is uniformly sampled across 26 elements, encompassing 9 monometallic and 225 bimetallic [234*3] TM_{55}

nanoclusters, three distinct adsorbates ($^*OOH$, $^*O$, and $^*OH$) and a wide range of E_{ads} (from -0.26 to -8.68 eV). This diversity ensures a representative dataset, essential for training reliable and generalizable GNN models.

7.8.2 Graph Representation and Deep Learning Framework. In the following section, the optimized non-Euclidean geometries are encoded as graph-based descriptors and used in regression tasks employing different SOTA GNN architectures: CGCNN *[78]*, ALIGNN *[82]*, and DeeperGATGNN *[86]*. During training, these models autonomously extract and learn structural representation directly from the crystal graphs, while primarily utilizing atomic features and interatomic distances to construct node and edge-level descriptors. By combining feature extraction and property prediction into a unified, end-to-end learning framework, GNNs infer target material properties directly from the topological and geometric configurations of atomic structures. Their distinct message-passing strategies enable effective capture of both local and global structural information, making them well-suited for data with complex, non-Euclidean geometry.

Specifically, CGCNN applies atom-wise message passing based on neighboring atoms and interatomic distances. These messages are processed through gated and message-processing layers, followed by average pooling of atom-level features, which are then passed through fully connected layers for property prediction. ALIGNN extends this approach with a dual-graph framework, performing edge-gated convolutions on both the atomic graph and its line graph, where bonds become nodes and edges represent bond pairs sharing a common atom. This enables simultaneous updates of node and edge features while effectively capturing angular dependencies. DeeperGATGNN further advances these architectures by stacking multiple Augmented Graph Attention (AGAT) layers, each integrated with additive skip connections and Differentiable Group Normalization (DGN). A global attention pooling mechanism, followed by

a series of fully connected layers, enables deep, hierarchical representation learning, significantly enhancing predictive performance.

7.8.3 Optimization and Scalability Comparison: To ensure a fair evaluation and meaningful comparison among SOTA GNN models, we evaluated the performance of CGCNN, ALIGNN, and DeeperGATGNN on our nanocluster datasets using identical data splits, consistent cross-validation protocols, and model-specific hyperparameter optimization. Following the benchmarking methodology of Fung et al. *[71]*, each model was initially trained for 2500 epochs using default hyperparameters to monitor convergence behavior and assess overfitting tendencies. This pre-screening phase provided insight into the learning dynamics and early stopping points unique to each architecture, as shown in Figure 7.13a-c. The training stabilized at different epochs for each model: CGCNN around 500, ALIGNN near 300, and DeeperGATGNN close to 500, reflecting their distinct optimization landscapes. CGCNN exhibited a relatively noisy loss profile, suggesting sensitivity to learning rate and batch size, likely due to its gated message-passing mechanism that relies solely on atomic neighbors and bond distances. In contrast, ALIGNN showed smooth and stable convergence, attributed to its dual-graph framework that incorporates angular information via edge-gated convolutions on both atomistic and line graphs. DeeperGATGNN also demonstrated stable training behaviour, enabled by its deep architecture comprising stacked AGAT layers, additive skip connections, and DGN, which together facilitate multi-hop message passing and robust hierarchical representation learning.

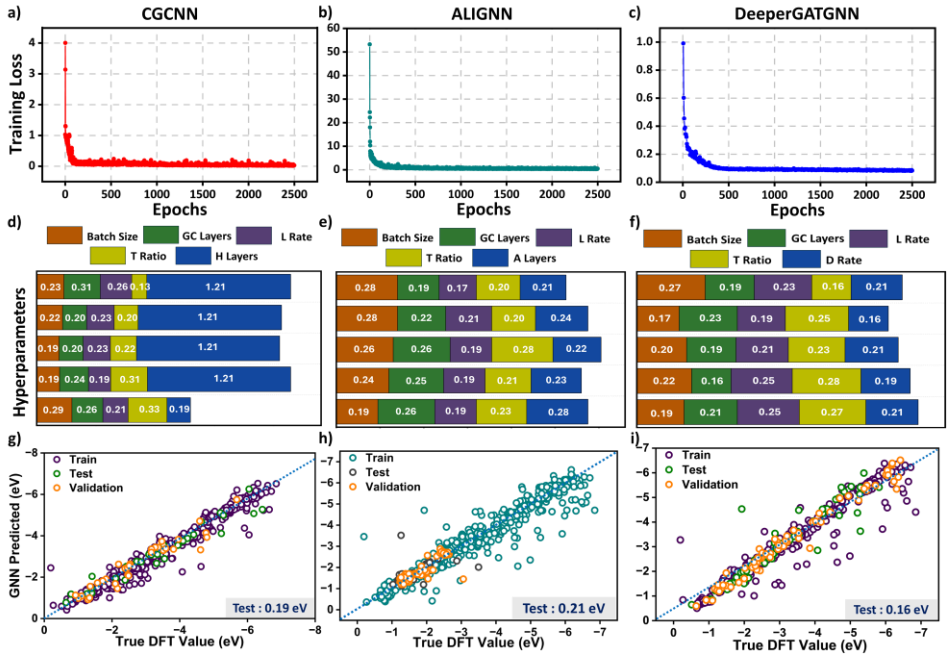


Figure 7.13: Comparative Evaluation of GNN models: CGCNN, ALIGNN, and DeeperGATGNN on nanocluster datasets. (a-c) Training loss curves over 2500 epochs, illustrating convergence behaviour and learning stability across different architectures. (d-f) Scalability comparison analysis across model-specific parameters, where overall lower MAE values (in eV) reflect higher predictive accuracy and robustness. (g-i) Parity plots comparing predicted and DFT-calculated adsorption energies for training, testing, and validation sets, with test set mean absolute error (MAE) values are annotated to highlight each model's generalization performance.

To evaluate the scalability and performance sensitivity of each GNN architecture, we conducted a systematic hyperparameter optimization study using 5-fold cross-validation on our dataset (Figure 7.14). Each model was initially tested using default settings, followed by targeted tuning of key model-specific parameters. First, we examined the effect of batch size on prediction accuracy by conducting three parallel runs for each model with batch sizes ranging from 8 to 40 in increments of 8. As shown in Figure 7.14, CGCNN and ALIGNN achieved optimal performance with a batch size of 24, while DeeperGATGNN reached its lowest mean absolute error (MAE) at a batch size of 32. Second, we assessed the influence of graph

convolution (GC) depth on model performance by progressively increasing the number of GC layers in steps of 5. CGCNN exhibited declining predictive accuracy with increasing depth, with the MAE rising to 0.31 eV at 25 GC layers, indicating limited scalability. In contrast, ALIGNN and DeeperGATGNN demonstrated strong depth scalability, both achieving a minimum MAE of 0.19 eV at GC layers. Third, we assessed the impact of learning rate (LR) on model performance by incrementally increasing it from 0.001 to 0.005 in steps of 0.001. Both ALIGNN and DeeperGATGNN demonstrated improved performance at higher LR, achieving their lowest MAE at LR = 0.005, and consistently outperforming CGCNN across all tested values.

Fourth, we evaluated the effect of training data size on model generalizability by varying the training set ratio from 0.50 to 0.90 in increments of 0.10. All three models achieved their lowest MAE when trained on 90% of the data, indicating strong scalability and the capacity to effectively learn from larger datasets. Finally, we fine-tuned architecture-specific parameters: the number of hidden layers (H Layers) for CGCNN, ALIGNN layers (A Layers) for ALIGNN, and dropout rates (D Rates) for DeeperGATGNN. For CGCNN, increasing the number of hidden layers from 5 to 25 led to a sharp deterioration in performance, with MAE rising from 0.19 to 1.21 eV, suggesting high sensitivity to network depth. In contrast, ALIGNN and DeeperGATGNN exhibited relative robustness, with changes in A Layers and D rates producing only minor variations in performance. A comparative overview of model scalability across these parameters is presented in Figure 7.14d-f, where the configuration yielding the lowest MAE is considered the most predictive. To ensure methodological rigor, we adopted the optimal hyperparameter settings corresponding to the lowest MAE for each model, and generated parity plots for the training, validation, and test sets (Figure 7.13c). All three GNN architectures performed reasonably well on our nanocluster dataset. However, CGCNN and DeeperGATGNN clearly outperform ALIGNN,

achieving substantially lower test MAEs of 0.19 eV and 0.16 eV, which indicates their superior ability to capture the underlying structure-property relationships in nanoclusters. Based on this comprehensive analysis, DeeperGATGNN was selected for downstream evaluation to identify the most active electrocatalyst, due to its parsimonious architecture, strong scalability, and compatibility with our dataset.

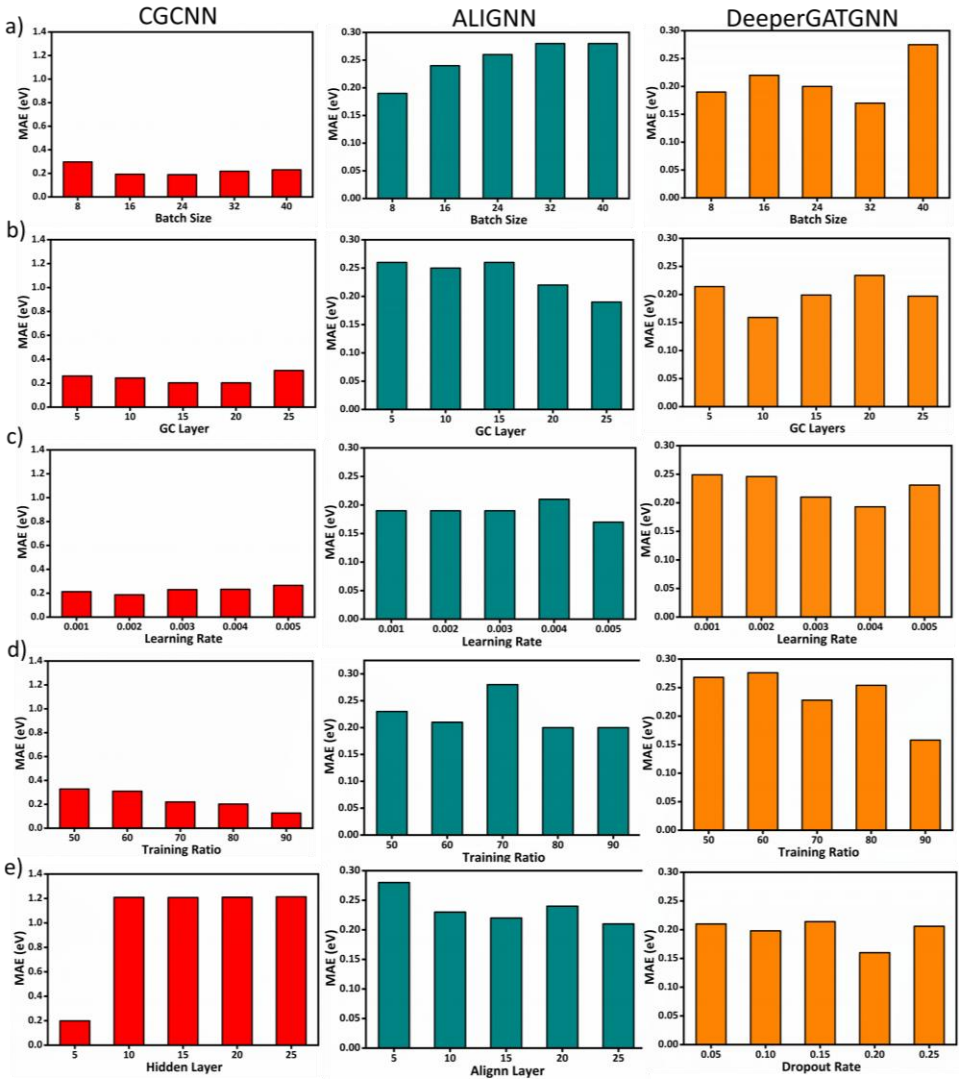


Figure 7.14: Parameter study of GNN models. (a) MAE vs Batch Size, (b) MAE vs GC Layers, (c) MAE vs Learning Rate, (d) MAE vs Training Ratio, and (e) MAE vs Model Specific parameters, for CGCNN, ALIGNN, and DeeperGATGNN models. All the experiments are done using a minimum of epochs and 5-fold cross-validation on the overall datasets.

7.8.4 Activity Comparison and Impact of the d-band Centre Analysis. We now employ our trained DeeperGATGNN model to predict the E_{*_O}, $E_{*_{OH}}$, and $E_{*_{OOH}}$ values for all catalysts in our dataset, enabling the construction of the full ORR mechanism via the associative pathway for all the catalysts in our training dataset. To identify active electrocatalysts, we apply the computational hydrogen electrode (CHE) model proposed by Nørskov and coworkers, calculating the reaction free energies at both 0 V and 1.23 V based on the following elementary steps *[61]*:

$$O_{2(g)} + (H^+ + e^-) + * \rightarrow OOH^* \quad \textbf{(R1)}$$

$$OOH^* + (H^+ + e^-) \rightarrow O^* + H_2O\,(l) \quad \textbf{(R2)}$$

$$O^* + (H^+ + e^-) \rightarrow OH^* \quad \textbf{(R3)}$$

$$OH^* + (H^+ + e^-) \rightarrow H_2O\,(l) + * \quad \textbf{(R4)}$$

Further, we compute the activity of each nanocluster alloy by calculating the overpotential values (η) of the rate-determining step (RDS) at 1.23 V, using the equation:

$$\eta_{ORR}^{GNN} = \frac{\max(\Delta R_1, \Delta R_2, \Delta R_3, \Delta R_4)}{e} \quad (7.6)$$

The overpotential values (η_{ORR}^{GNN}) for the rate-determining step (RDS) at 1.23 V for each catalyst were calculated using CHE model. Figure 7.15a-c highlights the non-monotonic trend in ORR activity observed as a function of core-shell composition in the bimetallic nanoclusters. Interestingly, $Pd_{45}Sc_{10}$, $Pd_{45}Nb_{10}$, $Au_{48}Cr_7$, were identified as the active catalysts for ORR with their low η values for RDS as follows: $\eta_{ORR}^{DFT} : \eta_{ORR}^{GNN} = 0.60\ V : 0.41\ V$, $\eta_{ORR}^{DFT} : \eta_{ORR}^{GNN} = 0.73\ V : 0.60\ V$, and $\eta_{ORR}^{DFT} : \eta_{ORR}^{GNN} = 0.79\ V : 0.47\ V$, exhibiting FCC, FCC, and DRC based geometries. While the majority of nanoclusters exhibit the last proton-electron transfer step ($^*OH$ to H_2O, denoted as R_4) as the RDS, $Ag_{42}M_{13}$ and $Au_{48}M_7$-based clusters primarily feature the $^*OOH$ formation from $O_2(g)$ as the RDS. In contrast, $Ir_{48}Au_7$, $Ir_{48}Cr_7$, $Ir_{48}Zn_7$ exhibit notably high η values for RDS as follows:

$\eta_{ORR}^{DFT} : \eta_{ORR}^{GNN} = 2.02\ V : 2.13\ V$, $\eta_{ORR}^{DFT} : \eta_{ORR}^{GNN} = 2.11\ V : 2.07\ V$, and $\eta_{ORR}^{DFT} : \eta_{ORR}^{GNN} = 2.08\ V : 2.07\ V$.

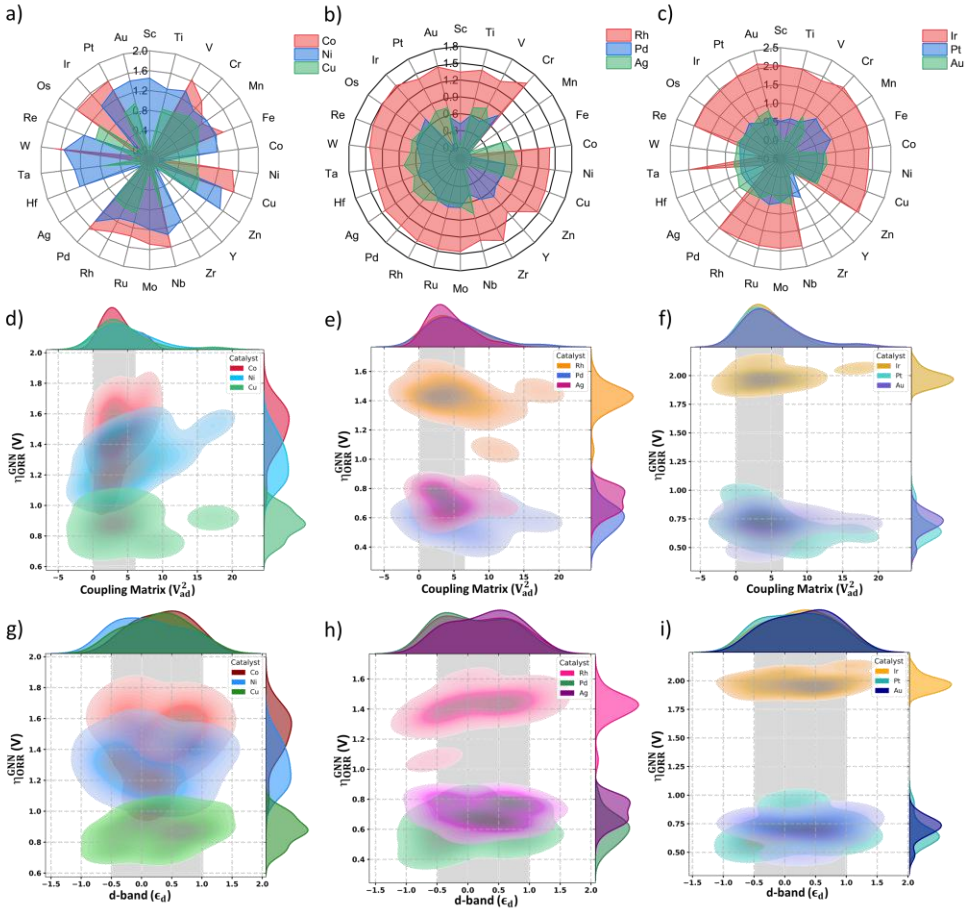


Figure 7.15: Activity comparison for transition metal clusters. Radar plots showing variations in DeeperGATGNN-predicted ORR overpotentials (η_{ORR}^{GNN}) at 1.23 V for different core-shell compositions: (a) 3d: $Co_{42}M_{13}$, $Ni_{42}M_{13}$, $Cu_{42}M_{13}$, (b) 4d: $Rh_{45}M_{10}$, $Pd_{45}M_{10}$, $Ag_{42}M_{13}$, (c) 5d: $Ir_{48}M_7$, $Pt_{46}M_9$, and $Au_{48}M_7$. Elements labelled in the legends and circles indicate shell and core compositions, while vertical lines show the range of predicted η_{ORR}^{GNN}. Note that nanocluster (NC) electrocatalysts excluded from the dataset are marked by η_{ORR}^{GNN} = 0.00 V. Kernel density estimation (KDE) plots illustrating the correlation between electronic descriptors and η_{ORR}^{GNN}: Coupling matrix vs η_{ORR}^{GNN} for (d) 3d: $Co_{42}M_{13}$, $Ni_{42}M_{13}$, $Cu_{42}M_{13}$, (e) 4d: $Rh_{45}M_{10}$, $Pd_{45}M_{10}$, $Ag_{42}M_{13}$, and (f) 5d: $Ir_{48}M_7$, $Pt_{46}M_9$, and $Au_{48}M_7$; d-

band center vs η_{ORR}^{GNN} for the same metal groups are represented in (g-i). KDE contours represent smoothed density distributions, highlighting regions of high data concentration in descriptor-overpotential space. Marginal histograms display distribution trends of individual variables, offering insight into underlying structure-activity relationships.

Note that the node- and edge-level embeddings in the DeeperGATGNN model utilize higher-order neighbour information, enabling the model to learn from extended structural environments and basic elemental-based features. However, the catalytic activity of TM alloys is fundamentally governed by their electronic structure, particularly d-state characteristics. This limits the physical interpretability of GNNs, which often lack explicit connections to physics-based descriptors. To bridge this gap, we integrate concepts from the d-band model proposed by Hammer and Nørskov, which has been widely successful in correlating catalytic activity with the electronic structure of TM surfaces. While extracting d-band-related descriptors such as the coupling matrix and d-band center typically requires self-consistent quantum mechanical calculations, which are computationally expensive. To circumvent these limitations, we utilize element-specific numerical values derived from solid-state surface tables that encode key d-band characteristics of each element without requiring ab initio calculations. Specifically, we employ the coupling matrix (V_{ad}^2) and the d-band center ($\epsilon_d = \frac{d\epsilon_d}{d \ln S_{BW}}$), where S_{BW} is the Wigner-Seitz radius. These descriptors capture intrinsic d-state properties across the periodic table, enabling us to efficiently relate electronic structure features to ORR activity.

To establish the relationship between these descriptors and catalytic performance, we visualize the correlation between η_{ORR}^{GNN} and the descriptors V_{ad}^2 and ϵ_d using kernel density estimation (KDE) plots for our metal alloy compositions, represented in Figure 7.15d-i. The KDE plots provide both joint probability distributions (2D contours) and marginal densities (1D

ridges) across three nanoclusters of transition metal families: 3d (Co, Ni, Cu), 4d (Rh, Pd, Ag), and 5d (Ir, Pt, Au). The top row displays η_{ORR}^{GNN} vs V_{ad}^2 (plots 7.15d-f), while the bottom row shows η_{ORR}^{GNN} vs ϵ_d (plots 7.15g-i). In the 3d series, Co and Ni exhibit broad distributions over moderate to high coupling strengths ($V_{ad}^2 \approx$ 5-15) and ϵ_d in the range of 0-1.5 eV. Their predicted h_{ORR}^{GNN} span 0.8–1.4 V, suggesting relatively strong E_{ads}. In contrast, Cu lies in the low V_{ad}^2 regime (< 5) and has a more negative ϵ_d (~ -1.0 eV), resulting in lower η_{ORR}^{GNN} (0.6 – 0.8 V), consistent with more optimal binding energies and enhanced ORR activity. For the 4d metals (plots 7.15e,i), Rh and Pd span similar V_{ad}^2 ranges (~5-15), but exhibit different catalytic profiles. Rh shows a KDE peak at η_{ORR}^{GNN} ~ 0.7 V, while Pd skew towards higher values (~1.2-1.6 V), indicating stronger binding and less favorable desorption kinetics. Ag appears distinctly in the low-coupling, more negative ϵ_d regime, with sharp KDE peaks at low η_{ORR}^{GNN}, reaffirming its experimentally known catalytic activity.

In the 5d metals (plots 7.15f,h), Pt and Ir are concentrated at high V_{ad}^2 (~10-20). Pt exhibits a bimodal KDE, with one lobe at η_{ORR}^{GNN} ~ 0.8 V, indicative of partial optimization. However, Ir shows a high-density KDE at η_{ORR}^{GNN} ~ 1.6 – 2.0 V, reflecting excessive binding strength and overpotential penalties. Au stands out with a KDE peak at low η_{ORR}^{GNN} values (~ 0.6 – 0.8 V), despite only moderate coupling. This behaviour stems from its deep ϵ_d (< -0.5 eV), which weakens adsorbate interaction and facilitates optimal catalytic energetics. Focusing on the KDE analysis in the bottom row (plots 7.15g-i), a clear trend emerges: more negative ϵ_d strongly correlate with lower η_{ORR}^{GNN}, supporting the principles of the d-band theory and the Sabatier principle. Cu, Ag, and Au consistently show KDE maxima at ϵ_d < -0.5 eV and η_{ORR}^{GNN} ~ 0.6-0.8 V, validating their favourable ORR activity profiles. In contrast, Pd and Ir, characterized by shallower ϵ_d values (> 0 eV), exhibit KDE peaks at higher overpotentials (~ 1.2–2.0 V), indicating overbinding of intermediates and reduced catalytic performance. Overall, KDE analysis

reveals that catalysts with the lowest η_{ORR}^{GNN} value (~0.6 – 0.8 V) consistently fall within an optimal V_{ad}^{2} range of 2-8 and ϵ_d range of -1.2 to -0.4 eV. These descriptor ranges correspond to moderate metal-adsorbate hybridization strengths and stabilization, which together facilitate optimal adsorption energies for key ORR intermediates.

7.8.5 Statistical Data Analysis and Prediction. Further, we employ statistical tools to group catalyst materials based on the predicted η_{ORR}^{GNN} and computed η_{ORR}^{DFT} values to understand the variation of morphology towards the ORR activity. Initially, we distribute the catalysts based on predicted η_{ORR}^{GNN} and computed η_{ORR}^{DFT} values in the scatter plot, as represented in Figure 7.16a (with respect to Pt). The bimetallic nanoclusters such as $Ag_{42}M_{13}$, $Au_{48}M_7$, and $Pd_{45}M_{10}$ are positioned in the low η_{ORR}^{GNN} and η_{ORR}^{DFT} regime, indicating their high suitability for the ORR. In contrast, clusters like $Ni_{42}M_{13}$, $Co_{42}M_{13}$ and $Ir_{48}M_7$ fall into the high η_{ORR}^{GNN} and η_{ORR}^{DFT} range, rendering them less favourable for ORR catalysis. Additionally, we expand our dataset to include $Os_{45}M_{10}$, $Re_{42}M_{13}$, $Ru_{45}M_{10}$, and $Zn_{46}M_9$ - based bimetallic NC alloys. For these, the E_{ads} of key ORR intermediates were predicted using the DeeperGATGNN model, initialized with optimized model checkpoints, while computing the predicted η_{ORR}^{GNN} values for the same.

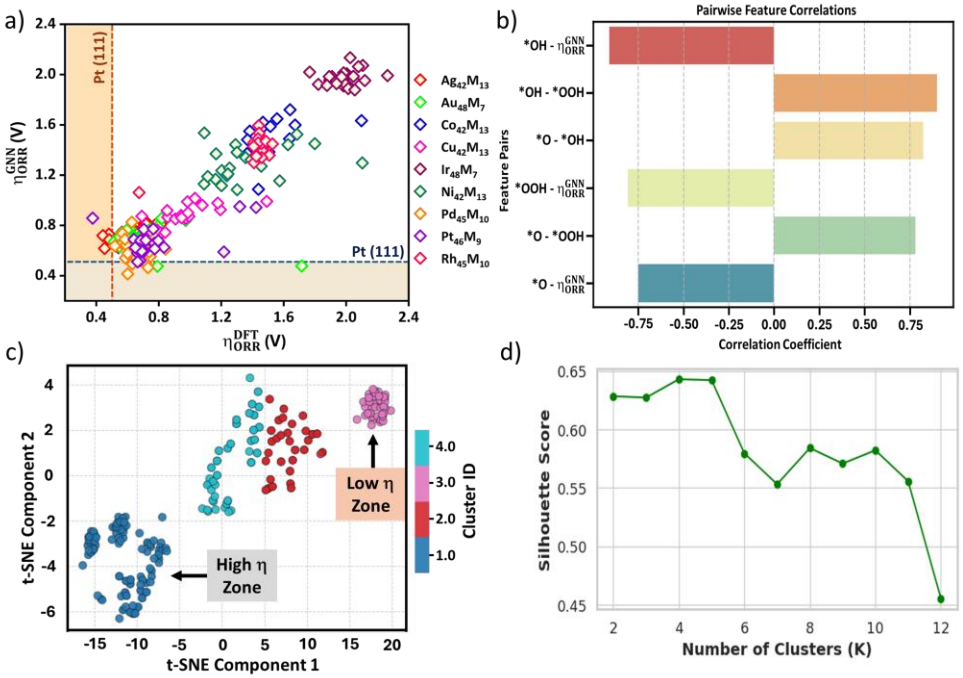


Figure 7.16: Statistical analysis of ORR activity. (a) Scatter plot comparing GNN-predicted ORR overpotentials η_{ORR}^{GNN} with DFT-computed values η_{ORR}^{DFT} for various core-shell nanocluster configurations. Each marker represents a unique nanocluster, and the orange and blue dashed lines indicate the benchmark overpotentials of Pt(111) from DFT and GNN, respectively. The shaded orange region highlights catalysts outperforming Pt(111) by both methods. (b) Pearson correlation coefficients showing relationships among input features and their correlation with E_{ads} and η_{ORR}^{DFT}. (c) t-SNE projection of NC alloys coloured by Cluster ID, illustrating distinct zones corresponding to high and low-overpotential (η) catalysts based on E_{ads} and η_{ORR}^{DFT}. (d) Silhouette score analysis used to determine the optimal number of K-means clusters applied to t-SNE reduced data. The maximum score (~0.65) is achieved at K = 4, indicating the optimal clustering configuration.

Notably, the scaling relationship in heterogeneous catalysis establishes a linear correlation between the easily computable thermodynamic properties and catalytic activity, facilitating descriptor-based designing. To probe this phenomenon for ORR on TM_{55} nanoclusters, we evaluate the Pearson

correlation coefficient (PCC) *[55]* in both input-input (intermediate-intermediate) and input-output (intermediate-activity) correlations using E_{ads} and η_{ORR}^{GNN} as a principal component for ORR (Figure 7.16b). Interestingly, we observed a breaking of the scaling relationship between E_{*O} vs E_{*OH} and E_{*O} vs E_{*OOH} with $R^2 \sim 0.75$, suggesting no single variable can be utilized as an independent descriptor for predicting overall activity across these nanoclusters (Figure 7.16a). To further elucidate the influence of compositional diversity on predicted η_{ORR}^{GNN}, we employ the t-distributed stochastic neighbour embedding (t-SNE) approach *[55]*, an unsupervised nonlinear dimensionality reduction technique based on the Eads. The resulting low-dimensional projection qualitatively reveals clusters of catalysts with similar characteristics, with catalysts positioned closely in the embedding space exhibiting similar reactivity descriptors and local neighbourhood similarities (Figure 7.16c). To uncover the latent chemical groupings, we applied K-means clustering to the t-SNE embedded space. The optimal number of clusters (K = 4) was selected based on the elbow method, which corresponds to the highest silhouette score of 0.64 at n_{init} = 400, indicating distinct clusters formation (Figure 7.16d). This provides an important comparison of the four catalyst clusters (Clusters 1-4) based on their composition, geometry, E_{ads} ranges and predicted η_{ORR}^{GNN}, with reference to our previous investigations.

Cluster 1 primarily consists of ICO geometry-based $Co_{42}M_{13}$, $Ni_{42}M_{13}$, $Rh_{45}M_{10}$, $Re_{42}M_{13}$, and $Ir_{48}M_7$, where the *O intermediates exhibit strong physisorption (~ -5 to -6.5 eV), resulting in very high predicted h_{ORR}^{GNN} ranging from 1.06 to 3.56 V. Cluster 2 included FCC geometry-based $Pd_{45}M_{10}$, $Os_{45}M_{10}$, and $Ru_{45}M_{10}$, with E_{ads} values for ORR intermediates falling near the optimal range (-1.00 to -4.50 eV), and comparatively lower η_{ORR}^{GNN} values between 0.41 and 0.71 V. Cluster 3 is solely composed of $Ag_{42}M_{13}$, $Au_{48}M_7$ nanoclusters, showing E_{ads} values close to the Sabatier optimum (-0.60 to -0.35 eV), along with low η_{ORR}^{GNN}, suggesting favourable

ORR activity. Finally, Cluster 4 includes $Cu_{42}M_{13}$, $Pt_{46}M_9$, and $Zn_{46}M_9$ based nanoclusters, where most E_{ads} values lie within the Sabatier optimal window (-1.23 to -4.2 eV), coupled with low predicted η_{ORR}^{GNN} ranging from 0.64 to 1.01 V.

7.8.6 Electronic Properties Analysis of Active Catalysts. To gain deeper insights into the metal-adsorbate interactions, we selected *OH intermediate involved in the RDS for $Pd_{45}Sc_{10}$, and $Pd_{45}Nb_{10}$ exhibiting the highest ORR activity and analyzed their electronic structures through partial density of states (PDOS) and crystal orbital Hamilton population (COHP) analysis *[85]*, as shown in Figure 7.17. The PDOS analysis indicates a stronger hybridization between Pd (4d) and O (2p) states near the Fermi level (E_f) indicating optimal binding interaction between metal and OH intermediate (Figure 7.17a-b). In the context of d-band theory *[87]*, the position of the d-band center ($\overline{\epsilon_d^{Pd}}$) serves as a key descriptor for the adsorption strength: a $\overline{\epsilon_d^{Pd}}$ with respect to the E_f corresponds to stronger metal-adsorption coupling due to greater orbital overlap. Both Pd-based alloys exhibit similar distant values for $\overline{\epsilon_d^{Pd}}$ relative to E_f ($Pd_{45}Sc_{10}: -2.08$ eV; $Pd_{45}Nb_{10}: -2.00$ eV), indicating optimal binding strength rather than strong chemisorption.

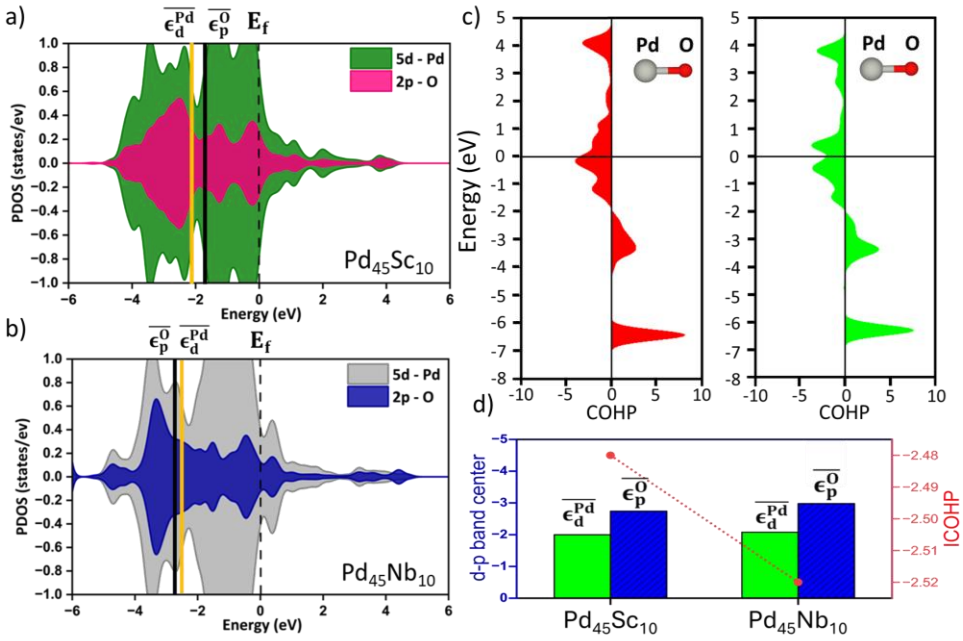


Figure 7.17: Electronic structure analysis: Partial density of states (PDOS) analysis of (a) $Pd_{45}Sc_{10}$, and (b) $Pd_{45}Nb_{10}$, during *OH adsorption involved in the RDS at 1.23 V. The Fermi level, d-band center, and p-band center are represented by E_f, $\overline{\epsilon_d^{Pd}}$, and $\overline{\epsilon_p^{O}}$, respectively. (c) Crystal orbital Hamilton population (COHP) analysis of the Pd-O bond of $Pd_{45}Sc_{10}$ (filled in red), and $Pd_{45}Nb_{10}$ (filled in green). (d) Comparison of $\overline{\epsilon_d^{Pd}}$, $\overline{\epsilon_p^{O}}$ and integrated crystal orbital Hamilton population (ICOHP) of $Pd_{45}Sc_{10}$, and $Pd_{45}Nb_{10}$ electrocatalysts.

COHP analysis further resolves the bonding and antibonding characteristics of the metal–oxygen (M–O) interactions (Figure 7.17c). In this framework, bonding states are represented by positive COHP values (right of the axis), while antibonding states appear as negative values (left of the axis). The valence states of Pd contribute predominantly to bonding interactions, while antibonding states are located near the E_f in the valence band and extend into the conduction band, under −OH adsorption conditions. These antibonding states, when partially occupied, indicate a balanced metal–oxygen bond that is neither too strong nor too weak—an essential requirement for higher catalytic activity. To quantitatively assess the M–O

bond strength, the integrated COHP (ICOHP) values were computed, reflecting the overall orbital overlap between the metal atoms and adsorbed oxygen species (Figure 7.17d). The Pd-based alloys present a more favorable balance between bond strength and reactivity, consistent with their superior OER activity. Collectively, these results demonstrate that the upshift in ϵ_d^{Pd}, coupled with moderate ICOHP values and optimized d–p hybridization, promotes effective charge redistribution and stabilizes oxygen-containing intermediates, thereby enhancing catalytic performance.

7.8.7 Conclusion

In summary, we present a combined DFT and GNN framework to accelerate the screening of late-transition metal-based TM_{55} nanocluster alloys for ORR. By constructing a representative dataset of adsorption energies (E_{ads}) for key ORR intermediates, spanning a wide range from 0.26 to -8.68 eV, we evaluate the performance of three state-of-the-art GNN models on nanoclusters with varying compositions and morphologies. Our findings underscore the importance of GNN model optimization, emphasizing the need to fine-tune baseline models to enhance sensitivity to the underlying dataset. Among the evaluated models, DeeperGATGNN achieved the highest predictive performance, with a mean absolute error of 0.27 eV. Through a comparative scalability analysis, we demonstrate the advantages of SOTA GNNs in capturing complex, compositionally diverse datasets, owing to their learned rather than explicitly defined representations. We further report the non-monotonic ORR activity trends by incorporating physics-based d-band descriptors, such as the coupling matrix (V_{ad}^2) and the d-band center and identify their optimal regions in alignment with the d-band theory. Additionally, we employ an unsupervised machine learning approach to identify the breakdown of linear scaling relationships and to derive optimal E_{ads} cutoffs for each intermediate. Finally, we analyze the electronic structures of key intermediates using PDOS and COHP analysis to assess the role of d-p

orbital hybridization in enhancing charge distribution and intermediate stabilization.

7.9 References

1. Wang, Y., Wang, D., Li, Y. (2021), A fundamental comprehension and recent progress in advanced Pt-based ORR nanocatalysts, *SmartMat.*, 2, e1023 (DOI: 10.1002/smm2.1023)
2. Zhang, L., Roling, L. T., Wang, X., Vara, M., Chi, M., Liu, J., Choi, S. I., Park, J., Herron, J. A., Xie, Z., Mavrikakis, M., Xia, Y. (2015), Platinum-based nanocages with subnanometer-thick walls and well-defined, controllable facets, *Science*, 349, 412–416 (DOI: 10.1126/science.aab0801)
3. Gao, R., Wang, J., Huang, Z. F., Zhang, R., Wang, W., Pan, L., Zhang, J., Zhu, W., Zhang, X., Shi, C., Lim, J., Zou, J. J. (2021), Pt/Fe_2O_3 with Pt–Fe pair sites as a catalyst for oxygen reduction with ultralow Pt loading, *Nat. Energy*, 6, 614–623 (DOI: 10.1038/s41560-021-00826-5)
4. Wang, Y. J., Zhao, N., Fang, B., Li, H., Bi, X. T., Wang, H. (2015), Carbon-supported Pt-based alloy electrocatalysts for the oxygen reduction reaction in polymer electrolyte membrane fuel cells: particle size, shape, and composition manipulation and their impact on activity, *Chem. Rev.*, 115, 3433–3467 (DOI: 10.1021/cr500519c)
5. Tyo, E. C., Vajda, S. (2015), Catalysis by clusters with precise numbers of atoms, *Nat. Nanotechnol.*, 10, 577–588 (DOI: 10.1038/nnano.2015.140)
6. Ferrando, R., Jellinek, J., Johnston, R. L. (2008), Nanoalloys: From theory to applications of alloy clusters and nanoparticles, *Chem. Rev.*, 108, 845–910 (DOI: 10.1021/cr040090g)
7. Yang, H. (2011), Platinum-based electrocatalysts with core–shell nanostructures, *Angew. Chem. Int. Ed.*, 50, 2674–2676 (DOI: 10.1002/anie.201005868)
8. Stamenkovic, V. R., Fowler, B., Mun, B. S., Wang, G., Ross, P. N., Lucas, C. A., Markovic, N. M. (2007), Improved oxygen reduction

activity on $Pt_3Ni(111)$ via increased surface site availability, *Science*, 315, 493–497 (DOI: 10.1126/science.1135941)

9. Stamenković, V., Schmidt, T. J., Ross, P. N., Marković, N. M. (2002), Surface composition effects in electrocatalysis: Kinetics of oxygen reduction on well-defined Pt_3Ni and Pt_3Co alloy surfaces, *J. Phys. Chem. B*, 106, 11970–11979 (DOI: 10.1021/jp021182h)
10. Greeley, J., Stephens, I. E. L., Bondarenko, A. S., Johansson, T. P., Hansen, H. A., Jaramillo, T. F., Rossmeisl, J., Chorkendorff, I., Nørskov, J. K. (2009), Alloys of platinum and early transition metals as oxygen reduction electrocatalysts, *Nat. Chem.*, 1, 552–556 (DOI: 10.1038/nchem.367)
11. Corona, B., Howard, M., Zhang, L., Henkelman, G. (2016), Computational screening of core@shell nanoparticles for the hydrogen evolution and oxygen reduction reactions, *J. Chem. Phys.*, 145, 244708 (DOI: 10.1063/1.4972579)
12. Kim, M., Yeo, B. C., Park, Y., Lee, H. M., Han, S. S., Kim, D. (2020), Artificial intelligence to accelerate the discovery of N_2 electroreduction catalysts, *Chem. Mater.*, 32, 709–720 (DOI: 10.1021/acs.chemmater.9b03686)
13. Pandit, N. K., Roy, D., Mandal, S. C., Pathak, B. (2022), Rational designing of bimetallic/trimetallic hydrogen evolution reaction catalysts using supervised machine learning, *J. Phys. Chem. Lett.*, 13, 7468–7476 (DOI: 10.1021/acs.jpclett.2c01401)
14. Roy, D., Mandal, S. C., Pathak, B. (2022), Machine learning assisted exploration of high entropy alloy-based catalysts for selective CO_2 reduction to methanol, *J. Phys. Chem. Lett.*, 13, 5890–5897 (DOI: 10.1021/acs.jpclett.2c00929)
15. Ye, Q., Yi, X., Wang, C. Z., Zhang, T., Liu, Y., Lin, S., Fan, H. J. (2024), Data-driven screening of pivotal subunits in edge-anchored single-atom catalysts for oxygen reactions, *Adv. Funct. Mater.*, 34, 2400107 (DOI: 10.1002/adfm.202400107)

16. Wu, L., Guo, T., Li, T. (2022), Data-driven high-throughput rational design of double-atom catalysts for oxygen evolution and reduction, *Adv. Funct. Mater.*, 32, 2203439 (DOI: 10.1002/adfm.202203439)
17. Mao, X., Wang, L., Xu, Y., Wang, P., Li, Y., Zhao, J. (2021), Computational high-throughput screening of alloy nanoclusters for electrocatalytic hydrogen evolution, *NPJ Comput. Mater.*, 7, 78 (DOI: 10.1038/s41524-021-00514-8)
18. Xu, H., Zhu, J. Z., Zou, C., Zhang, F., Ming, D., Guan, D., Ma, L. (2023), Theoretical design of core–shell 3d-metal nanoclusters for efficient hydrogen-evolving reaction, *Energy Fuels*, 37, 16435–16444 (DOI: 10.1021/acs.energyfuels.3c02448)
19. Jäger, M. O. J., Ranawat, Y. S., Canova, F. F., Morooka, E. V., Foster, A. S. (2020), Efficient machine-learning-aided screening of hydrogen adsorption on bimetallic nanoclusters, *ACS Comb. Sci.*, 22, 699–707 (DOI: 10.1021/acscombsci.0c00102)
20. Wien, T. U., Hauptstrage, W. (1994), *Ab initio* molecular-dynamics simulation of the liquid-metal–amorphous-semiconductor transition in germanium, *Phys. Rev. B*, 49, 1705–1715
21. Perdew, J. P., Chevary, J. A., Vosko, S. H., Jackson, K. A., Pederson, M. R., Singh, D. J., Fiolhais, C. (1992), Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation, *Phys. Rev. B*, 46, 6671–6687 (DOI: 10.1103/PhysRevB.46.6671)
22. Joubert, D. (1999), From ultrasoft pseudopotentials to the projector augmented-wave method, *Phys. Rev. B*, 59, 1758–1775 (DOI: 10.1103/PhysRevB.59.1758)
23. Zandkarimi, B., Alexandrova, A. N. (2019), Dynamics of subnanometer Pt clusters can break the scaling relationships in catalysis, *J. Phys. Chem. Lett.*, 10, 460–467 (DOI: 10.1021/acs.jpclett.8b03680)

24. Sharma, R. K., Nair, A. S., Bharadwaj, N., Roy, D., Pathak, B. (2023), Role of fluxionality and metastable isomers in the ORR activity: A case study, *J. Phys. Chem. C*, 127, 85–94 (DOI: 10.1021/acs.jpcc.2c06265)
25. Mathew, K., Sundararaman, R., Letchworth-Weaver, K., Arias, T. A., Hennig, R. G. (2014), Implicit solvation model for density-functional study of nanocrystal surfaces and reaction pathways, *J. Chem. Phys.*, 140, 084106 (DOI: 10.1063/1.4865107)
26. Capdevila-Cortada, M., Łodziana, Z., López, N. (2016), Performance of DFT+U approaches in the study of catalytic materials, *ACS Catal.*, 6, 8370–8379 (DOI: 10.1021/acscatal.6b01907)
27. Mishra, A. K., Roldan, A., De Leeuw, N. H. (2016), CuO surfaces and CO_2 activation: A dispersion-corrected DFT+U study, *J. Phys. Chem. C*, 120, 2198–2214 (DOI: 10.1021/acs.jpcc.5b10431)
28. Piotrowski, M. J., Ungureanu, C. G., Tereshchuk, P., Batista, K. E. A., Chaves, A. S., Guedes-Sobrinho, D., Da Silva, J. L. F. (2016), Theoretical study of the structural, energetic, and electronic properties of 55-atom metal nanoclusters: A DFT investigation within van der Waals corrections, spin–orbit coupling, and PBE+U of 42 metal systems, *J. Phys. Chem. C*, 120, 28844–28856 (DOI: 10.1021/acs.jpcc.6b10404)
29. Kresse, G., Hafner, J. (1994), *Ab initio* molecular-dynamics simulation of the liquid-metal–amorphous-semiconductor transition in germanium, *Phys. Rev. B*, 49, 14251–14269 (DOI: 10.1103/PhysRevB.49.14251)
30. Calle-Vallejo, F., Martínez, J. I., García-Lastra, J. M., Sautet, P., Loffreda, D. (2014), Fast prediction of adsorption properties for platinum nanocatalysts with generalized coordination numbers, *Angew. Chem. Int. Ed.*, 53, 8316–8319 (DOI: 10.1002/anie.201402958)
31. Zhu, B., Lu, J., Sakaki, S. (2021), Catalysis of core–shell nanoparticle M@Pt (M = Co and Ni) for oxygen reduction reaction and its electronic

structure in comparison to Pt nanoparticle, *J. Catal.*, 397, 147–158 (DOI: 10.1016/j.jcat.2021.02.031)

32. Kim, S., Kim, M. C., Yeo, B. C., Han, S. S. (2023), High-throughput design of bimetallic core–shell catalysts for the electrochemical nitrogen reduction reaction, *J. Mater. Chem. A*, 11, 24874–24884 (DOI: 10.1039/d3ta05408j)
33. Zhang, L., Guo, X., Zhang, S., Huang, S. (2022), Building up the "genome" of bi-atom catalysts toward efficient HER/OER/ORR, *J. Mater. Chem. A*, 10, 11231–11244 (DOI: 10.1039/d2ta02050e)
34. Bligaard, T., Nørskov, J. K., Dahl, S., Matthiesen, J., Christensen, C. H., Sehested, J. (2004), The Brønsted–Evans–Polanyi relation and the volcano curve in heterogeneous catalysis, *J. Catal.*, 224, 206–217 (DOI: 10.1016/j.jcat.2004.02.034)
35. Ferrin, P., Simonetti, D., Kandoi, S., Kunkes, E., Dumesic, J. A., Nørskov, J. K., Mavrikakis, M. (2009), Modeling ethanol decomposition on transition metals: A combined application of scaling and Brønsted–Evans–Polanyi relations, *J. Am. Chem. Soc.*, 131, 5809–5815 (DOI: 10.1021/ja8099322)
36. Dean, J., Taylor, M. G., Mpourmpakis, G. (2019), Unfolding adsorption on metal nanoparticles: Connecting stability with catalysis, *Sci. Adv.*, 5, eaax5101 (DOI: 10.1126/sciadv.aax5101)
37. Nair, A. S., Pathak, B. (2019), Computational screening for ORR activity of 3d transition metal-based M@Pt core–shell clusters, *J. Phys. Chem. C*, 123, 3584–3594 (DOI: 10.1021/acs.jpcc.8b11483)
38. Ooka, H., Huang, J., Exner, K. S. (2021), The Sabatier principle in electrocatalysis: Basics, limitations, and extensions, *Front. Energy Res.*, 9, 654460 (DOI: 10.3389/fenrg.2021.654460)
39. Sharma, R. K., Minhas, H., Pathak, B. (2024), Investigating the metastability-triggered reactivity of $Pt_{7,8}$ clusters on graphene: Unraveling statistical ensemble representation for ORR in gas and

implicit solvent phases, *J. Phys. Chem. C*, 128, 8992–9003 (DOI: 10.1021/acs.jpcc.4c00376)

40. Nair, A. S., Anoop, A., Ahuja, R., Pathak, B. (2021), Role of atomicity in the oxygen reduction reaction activity of platinum sub-nanometer clusters: A global optimization study, *J. Comput. Chem.*, 42, 2055–2065 (DOI: 10.1002/jcc.26725)
41. Nair, A. S., Anoop, A., Ahuja, R., Pathak, B. (2022), Relativistic effects in platinum nanocluster catalysis: A statistical ensemble-based analysis, *J. Phys. Chem. A*, 126, 1407–1416 (DOI: 10.1021/acs.jpca.1c09981)
42. Nair, A. S., Mahata, A., Pathak, B. (2018), Multilayered platinum nanotube for oxygen reduction in a fuel cell cathode: Origin of activity and product selectivity, *ACS Appl. Energy Mater.*, 1, 4090–4099 (DOI: 10.1021/acsaem.8b00641)
43. Nair, A. S., Mahata, A., Pathak, B. (2018), Multilayered platinum nanotube for oxygen reduction in a fuel cell cathode: Origin of activity and product selectivity, *ACS Appl. Energy Mater.*, 1, 4090–4099 (DOI: 10.1021/acsaem.8b00641)
44. Zandkarimi, B., Alexandrova, A. N. (2020), Can fluxionality of subnanometer cluster catalysts solely cause non-Arrhenius behavior in catalysis?, *J. Phys. Chem. C*, 124, 19914–19923 (DOI: 10.1021/acs.jpcc.0c04136)
45. Mai, H., Le, T. C., Chen, D., Winkler, D. A., Caruso, R. A. (2022), Machine learning for electrocatalyst and photocatalyst design and discovery, *Chem. Rev.*, 122, 13478–13515 (DOI: 10.1021/acs.chemrev.2c00061)
46. Hammer, B., Nørskov, J. K. (2000), Theoretical surface science and catalysis: Calculations and concepts, *Adv. Catal.*, 45, 71–129 (DOI: 10.1016/S0360-0564(02)45013-4)
47. Xin, H., Linic, S. (2010), Exceptions to the d-band model of chemisorption on metal surfaces: The dominant role of repulsion

between adsorbate states and metal d-states, *J. Chem. Phys.*, 132, 221101 (DOI: 10.1063/1.3437609)

48. Vojvodic, A., Nørskov, J. K., Abild-Pedersen, F. (2014), Electronic structure effects in transition metal surface chemistry, *Top. Catal.*, 57, 25–32 (DOI: 10.1007/s11244-013-0159-2)
49. Ruban, A., Hammer, B., Stoltze, P., Skriver, H. L., Nørskov, J. K. (1997), Surface electronic structure and reactivity of transition and noble metals, *J. Mol. Catal. A: Chem.*, 115, 421–429 (DOI: 10.1016/S1381-1169(96)00348-2)
50. Li, B., Zhang, P., Liang, S., Ren, G. (2008), Feature extraction and selection for fault diagnosis of gear using wavelet entropy and mutual information, *Proc. Int. Conf. Signal Process. (ICSP)*, 115–118 (DOI: 10.1109/ICOSP.2008.4697740)
51. Wang, N. N., Dong, J., Deng, Y. H., Zhu, M. F., Wen, M., Yao, Z. J., Lu, A. P., Wang, J. B., Cao, D. S. (2016), ADME properties evaluation in drug discovery: Prediction of Caco-2 cell permeability using a combination of NSGA-II and boosting, *J. Chem. Inf. Model.*, 56, 763–773 (DOI: 10.1021/acs.jcim.5b00642)
52. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., VanderPlas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É. (2011), Scikit-learn: Machine learning in Python, *J. Mach. Learn. Res.*, 12, 2825–2830
53. Natekin, A., Knoll, A. (2013), Gradient boosting machines, a tutorial, *Front. Neurorobot.*, 7, 21 (DOI: 10.3389/fnbot.2013.00021)
54. Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S. I. (2020), From local explanations to global understanding with explainable AI for trees, *Nat. Mach. Intell.*, 2, 56–67 (DOI: 10.1038/s42256-019-0138-9)
55. Sharma, R. K., Minhas, H., Pathak, B. (2024), High-throughput screening of bifunctional catalysts for oxygen evolution/reduction

reaction at the subnanometer regime, *Nanoscale*, 16, 16452–16463 (DOI: 10.1039/d4nr02787f)

56. Nørskov, J. K., Rossmeisl, J., Logadottir, A., Lindqvist, L., Kitchin, J. R., Bligaard, T., Jónsson, H. (2004), Origin of the overpotential for oxygen reduction at a fuel-cell cathode, *J. Phys. Chem. B*, 108, 17886–17892 (DOI: 10.1021/jp047349j)
57. Rossmeisl, J., Qu, Z. W., Zhu, H., Kroes, G. J., Nørskov, J. K. (2007), Electrolysis of water on oxide surfaces, *J. Electroanal. Chem.*, 607, 83–89 (DOI: 10.1016/j.jelechem.2006.11.008)
58. Wang, Y., Huang, X., Fu, H., Shang, J. (2022), Theoretically revealing the activity origin of the hydrogen evolution reaction on carbon-based single-atom catalysts and identifying ideal catalysts for water splitting, *J. Mater. Chem. A*, 10, 23631–23642 (DOI: 10.1039/d2ta07167c)
59. Zhang, Z., Zandkarimi, B., Munarriz, J., Dickerson, C. E., Alexandrova, A. N. (2022), Fluxionality of subnano clusters reshapes the activity volcano of electrocatalysis, *ChemCatChem*, 14, e202200345 (DOI: 10.1002/cctc.202200345)
60. Zhang, Z., Zandkarimi, B., Munarriz, J., Dickerson, C. E., Alexandrova, A. N. (2022), Fluxionality of subnano clusters reshapes the activity volcano of electrocatalysis, *ChemCatChem*, 14, e202200345 (DOI: 10.1002/cctc.202200345)
61. Prabu, M., Ketpang, K., Shanmugam, S. (2014), Hierarchical nanostructured $NiCo_2O_4$ as an efficient bifunctional non-precious metal catalyst for rechargeable zinc–air batteries, *Nanoscale*, 6, 3173–3181 (DOI: 10.1039/c3nr05835b)
62. Kan, D., Wang, D., Cheng, Y., Lian, R., Sun, B., Chen, K., Huo, W., Wang, Y., Chen, G., Wei, Y. (2021), Designing efficient bifunctional ORR/OER Pt single-atom catalysts based on O-terminated MXenes by first-principles calculations, *ACS Appl. Mater. Interfaces*, 13, 52645–52655 (DOI: 10.1021/acsami.1c12893)

63. Man, I. C., Su, H. Y., Calle-Vallejo, F., Hansen, H. A., Martínez, J. I., Inoglu, N. G., Kitchin, J., Jaramillo, T. F., Nørskov, J. K., Rossmeisl, J. (2011), Universality in oxygen evolution electrocatalysis on oxide surfaces, *ChemCatChem*, 3, 1159–1165 (DOI: 10.1002/cctc.201000397)
64. Exner, K. S. (2023), Steering selectivity in the four-electron and two-electron oxygen reduction reactions: On the importance of the volcano slope, *ACS Phys. Chem. Au*, 3, 125–134 (DOI: 10.1021/acsphyschemau.2c00054)
65. Exner, K. S. (2020), Does a thermoneutral electrocatalyst correspond to the apex of a volcano plot for a simple two-electron process?, *Angew. Chem. Int. Ed.*, 59, 10236–10240 (DOI: 10.1002/anie.202003688)
66. Fung, V., Zhang, J., Juarez, E., Sumpter, B. G. (2021), Benchmarking graph neural networks for materials chemistry, *NPJ Comput. Mater.*, 7, 84 (DOI: 10.1038/s41524-021-00554-0)
67. Behler, J. (2011), Atom-centered symmetry functions for constructing high-dimensional neural network potentials, *J. Chem. Phys.*, 134, 074106 (DOI: 10.1063/1.3553717)
68. De, S., Bartók, A. P., Csányi, G., Ceriotti, M. (2016), Comparing molecules and solids across structural and alchemical space, *Phys. Chem. Chem. Phys.*, 18, 13754–13769 (DOI: 10.1039/c6cp00415f)
69. Rupp, M., Tkatchenko, A., Müller, K. R., von Lilienfeld, O. A. (2012), Fast and accurate modeling of molecular atomization energies with machine learning, *Phys. Rev. Lett.*, 108, 058301 (DOI: 10.1103/PhysRevLett.108.058301)
70. Rupp, M., Tkatchenko, A., Müller, K. R., von Lilienfeld, O. A. (2012), Fast and accurate modeling of molecular atomization energies with machine learning, *Phys. Rev. Lett.*, 108, 058301 (DOI: 10.1103/PhysRevLett.108.058301)

71. Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., Sun, M. (2020), Graph neural networks: A review of methods and applications, *AI Open*, 1, 57–81 (DOI: 10.1016/j.aiopen.2021.01.001)
72. Battaglia, P. W., Hamrick, J. B., Bapst, V., Sanchez-Gonzalez, A., Zambaldi, V., Malinowski, M., Tacchetti, A., Raposo, D., Santoro, A., Faulkner, R., Gulcehre, C., Song, F., Ballard, A., Gilmer, J., Dahl, G., Vaswani, A., Allen, K., Nash, C., Langston, V., Dyer, C., Heess, N., Wierstra, D., Kohli, P., Botvinick, M., Vinyals, O., Li, Y., Pascanu, R. (2018), Relational inductive biases, deep learning, and graph networks, *arXiv preprint*, arXiv:1806.01261
73. Xie, T., Grossman, J. C. (2018), Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties, *Phys. Rev. Lett.*, 120, 145301 (DOI: 10.1103/PhysRevLett.120.145301)
74. Schütt, K. T., Sauceda, H. E., Kindermans, P. J., Tkatchenko, A., Müller, K. R. (2018), SchNet: A deep learning architecture for molecules and materials, *J. Chem. Phys.*, 148, 241722 (DOI: 10.1063/1.5019779)
75. Chen, C., Ye, W., Zuo, Y., Zheng, C., Ong, S. P. (2019), Graph networks as a universal machine learning framework for molecules and crystals, *Chem. Mater.*, 31, 3564–3572 (DOI: 10.1021/acs.chemmater.9b01294)
76. Chen, C., Ye, W., Zuo, Y., Zheng, C., Ong, S. P. (2019), Graph networks as a universal machine learning framework for molecules and crystals, *Chem. Mater.*, 31, 3564–3572 (DOI: 10.1021/acs.chemmater.9b01294)
77. Choudhary, K., DeCost, B. (2021), Atomistic line graph neural network for improved materials property predictions, *NPJ Comput. Mater.*, 7, 185 (DOI: 10.1038/s41524-021-00650-1)
78. Blöchl, P. E. (1994), Projector augmented-wave method, *Phys. Rev. B*, 50, 17953–17979 (DOI: 10.1103/PhysRevB.50.17953)

79. Perdew, J. P., Burke, K., Ernzerhof, M. (1996), Generalized gradient approximation made simple, *Phys. Rev. Lett.*, 77, 3865–3868 (DOI: 10.1103/PhysRevLett.77.3865)
80. Deringer, V. L., Tchougréeff, A. L., Dronskowski, R. (2011), Crystal orbital Hamilton population (COHP) analysis as projected from plane-wave basis sets, *J. Phys. Chem. A*, 115, 5461–5466 (DOI: 10.1021/jp202489s)
81. Omee, S. S., Louis, S. Y., Fu, N., Wei, L., Dey, S., Dong, R., Li, Q., Hu, J. (2022), Scalable deeper graph neural networks for high-performance materials property prediction, *Patterns*, 3, 100491 (DOI: 10.1016/j.patter.2022.100491)
82. Saini, S., Halldin Stenlid, J., Abild-Pedersen, F. (2022), Electronic structure factors and the importance of adsorbate effects in chemisorption on surface alloys, *NPJ Comput. Mater.*, 8, 165 (DOI: 10.1038/s41524-022-00846-z)

CHAPTER 8

Future Perspectives

8.1 Scope for Future Work

The future scope of this work lies in advancing a unified framework to capture the dynamic ON and OFF states of ligand-protected gold subnanoclusters under realistic electrochemical conditions. The proposed direction emphasizes the development of accelerated dynamics simulations driven by machine-learning interatomic potentials to resolve time-dependent structural and electronic transformations that are inaccessible through static first-principles approaches. A key focus will be on establishing a reliable training protocol capable of describing ligand passivation and depassivation events, including bond dissociation and reformation processes that govern catalytic switching behavior. This will enable a detailed understanding of the true nature of ligand-based metal–passivation networks and the dynamic evolution of core–shell architectures across different cluster sizes.

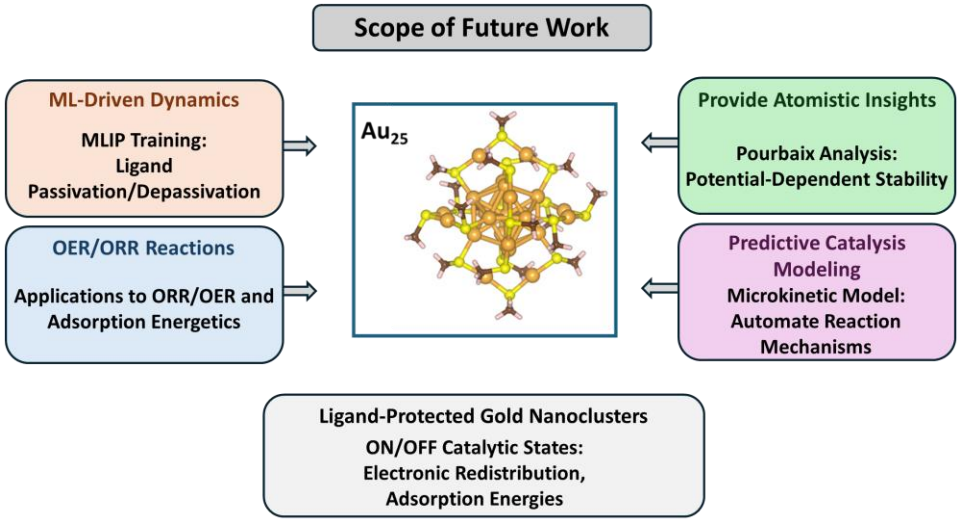


Figure 8.1: Schematic illustration of the proposed framework for exploring ON/OFF catalytic states in ligand-protected gold nanoclusters. The scheme highlights ML-driven accelerated dynamics for capturing ligand passivation and depassivation, atomistic insights into electronic redistribution and structural evolution, potential- and pH-dependent stability analysis via Pourbaix diagrams, and automated microkinetic modeling. Together, these elements aim to establish a predictive

understanding of structure–dynamics–activity relationships and extend applications toward ORR/OER catalysis.

The methodology will be extended to quantify how these dynamic states influence adsorption energetics and reaction pathways relevant to oxygen reduction and oxygen evolution reactions. Higher intermediate coverages and potential-dependent stability will be systematically explored using Pourbaix diagram analysis to bridge atomistic insights with electrochemical environments. Finally, the integration of automated microkinetic modeling will allow direct translation of atomistic dynamics into reaction rates and activity trends, offering a predictive link between structural fluxionality and catalytic performance. Collectively, this future work aims to establish a transferable and data-driven paradigm for understanding and designing subnanometer electrocatalysts beyond static structure–activity relationships.