Karlsruher Institut für Technologie

Simulation of Catalytic Ethylene Conversions in Metal-Organic Frameworks and Development of Machine-Learning Anharmonic Free Energy Methods

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN
(Dr. rer. nat.)

von der KIT-Fakultät für Chemie und Biowissenschaften
des Karlsruher Instituts für Technologie (KIT)

genehmigte
Dissertation

von

M.Sc. Thanh-Nam Huynh

aus Hoainhon (Vietnam)

1. Referent: Prof. Dr. Felix Studt
2. Referentin: apl. Prof. Dr. Karin Fink
Tag der mündlichen Prüfung: 18. Dezember 2025

Abstract

A transition toward a circular carbon economy, in which carbon is continually recycled rather than emitted, represents a promising route to mitigate carbon emissions, particularly in energy and chemical sectors. Such a transformation relies on continued advances in catalyst developments, which underpins more than 85% of current chemical and fuel production processes. Single-atom catalysts (SACs), featuring isolated and atomically dispersed active metal centers, offer exceptional activity, selectivity, and atom efficiency, thereby reducing material costs and maximizing catalytic utilization. Metal–organic frameworks (MOFs) are ideal porous supports for SACs, owing to their well-defined crystalline architectures, tunable pore structures, and improved chemical and thermal stability. Consequently, MOF-supported SACs have emerged as a rapidly growing class of catalysts and a model platform for probing fundamental catalytic phenomena at the atomic scale.

Computational catalysis provides molecular insight into active site structures, reaction mechanisms, and material design principles, offering perspectives often inaccessible to experimental approaches. Catalysis in MOFs, like in other heterogeneous systems, inherently spans multiple length and time scales, necessitating carefully benchmarked multiscale methodologies to achieve accurate thermodynamic and kinetic descriptions. A hierarchical, bottom-up strategy, linking quantum-mechanical, molecular-dynamics, and kinetic models, offers a robust framework to bridge these scales. This thesis addresses two central challenges in this direction: obtaining accurate reaction energetics and thermodynamic properties.

The first part investigates the activity of two Ni-based active-site motifs for ethylene dimerization via the Cossee–Arlman (linear insertion) mechanism. The more active configuration is identified as a $Ni^{II}O_2$-type site, which provides enhanced accessibility for ethylene coordination. The C–C coupling step is found to be rate-determining, yielding butene as the predominant product. The NU-1000 framework facilitates chain growth preferentially within its smaller trigonal pores as compared to the hexagonal mesopores. Truncated cluster models reproduce periodic energetics within $4\,\mathrm{kJ\,mol^{-1}}$, indicating these models can be used to provide computational efficiency without compromising accuracy. Free energy profile corrected to the TPSSh level predicts turnover frequencies (TOFs) in close agreement with experiment. Insights from the Ni-based system are extended to the Cr-NU-1000 catalyst, revealing a tetrahedral Cr^{III} active site responsible for ethylene polymerization. The polymerization follows a Cossee–Arlman pathway with insertion barriers as low as $50\,\mathrm{kJ\,mol^{-1}}$ and chain-termination barriers exceeding these by more than $30\,\mathrm{kJ\,mol^{-1}}$, consistent with experimental observations.

In the second part, cost-efficient machine-learned force fields (MLFFs) are implemented to accelerate phase space sampling in two rigorous free energy estimation frameworks: λ-thermodynamic integration (λ-TI) and the intermediate hard-sphere model (IHSM). This approach bridges the time-scale limitations of conventional DFT-MD while preserving DFT-level accuracy. Complementary statistical techniques, including free energy perturbation (FEP), machine learning thermodynamic perturbation theory (MLPT), Bennett's acceptance ratio (BAR), and the phase-space overlap index, I_w, are used to assess MLFF-derived results and correct them to the desired DFT accuracy. The robustness of these MLFF-aided workflows is meticulously validated, and the workflows were applied to diverse adsorption phenomena, namely the chemisorption of oxygenated species, and the physisorption and dissociative adsorption of methane on Pt(111). The atomic oxygen adsorption exhibited nearly harmonic behavior and was reasonably described by the HA, whereas OH and OOH adsorption displayed strong anharmononicity, with the MLFF-TI method recovering up to 10% of the entropy loss estimated under HA. For CH_4 adsorption, the influence of density functionals were evaluated. BEEF-vdW, which tends to underestimate adsorption strength, predicted higher anharmonic contributions relative to PBE-D3(BJ). At high temperatures or within the physisorption regime, where the adsorbate became more mobile, the impact of anharmonicity was found to be nearly independent of the employed density functionals. Using the MLFF-IHSM approach, the experimental adsorption entropy for CH_4 physisorption was accurately reproduced at both theory levels, while the adsorption entropies from both levels converged at 700 K for dissociative adsorption process. This part highlights the limitations of the harmonic approximation and emphasize the necessity of employing more accurate methods that can reliably capture anharmonic effects to improve the estimation of adsorption free energies and their derived thermodynamics and kinetics. Finally, a proof-of-concept demonstration shows that DFT-accurate anharmonic free energies can be achieved using a pretrained foundation MLFF model with only 200 additional DFT single-point evaluations, paving the way for transferable and scalable applications to MOF systems using these general-purpose models.

Zusammenfassung

Ein Übergang hin zu einer Kreislaufwirtschaft, in der Kohlenstoff kontinuierlich recycelt statt emittiert wird, stellt einen vielversprechenden Ansatz zur Reduktion von CO_2-Emissionen dar, insbesondere in der Energie- und Chemieindustrie. Eine solche Transformation erfordert kontinuierliche Fortschritte in der Katalysatorentwicklung, die über 85% der derzeitigen chemischen und Brennstoffproduktionsprozesse ermöglicht. Einzelatomkatalysatoren (SACs), die durch isolierte, atomar dispergierte aktive Metallzentren charakterisiert sind, zeichnen sich durch herausragende Aktivität, Selektivität und Atomeffizienz aus. Dadurch lassen sich Materialkosten senken und die katalytische Nutzung der eingesetzten Metalle maximieren. Metall-organische Gerüstverbindungen (MOFs) bieten sich aufgrund ihrer wohldefinierten kristallinen Architektur, ihrer einstellbaren Porenstrukturen sowie ihrer hohen chemischen und thermischen Stabilität als ideale poröse Trägermaterialien für SACs an. MOF-gestützte SACs haben sich somit zu einer rasch wachsenden Klasse moderner Katalysatoren und zu einer Modellplattform zur Untersuchung fundamentaler katalytischer Prozesse auf atomarer Skala entwickelt.

Die computergestützte Katalyse liefert tiefgehende molekulare Einblicke in die Struktur aktiver Zentren, in Reaktionsmechanismen und in grundlegende Prinzipien des Materialdesigns. Dadurch eröffnet sie Perspektiven, die experimentellen Methoden häufig nicht zugänglich sind. Wie auch andere heterogene Systeme umfasst die Katalyse in MOFs intrinsisch mehrere Längen- und Zeitskalen. Eine präzise Beschreibung der thermodynamischen und kinetischen Eigenschaften erfordert daher sorgfältig validierte multiskalige Ansätze. Eine hierarchische Bottom-up-Strategie, welche quantenmechanische Berechnungen, molekulardynamische Simulationen und kinetische Modellierungen miteinander verknüpft, bietet hierfür einen leistungsfähigen Rahmen. Diese Arbeit adressiert zwei zentrale Herausforderungen in diesem Kontext: die genaue Bestimmung von Reaktionsenergien und die präzise Ermittlung thermodynamischer Eigenschaften.

Im ersten Teil wird die katalytische Aktivität zweier Ni-basierter aktiver Zentren für die Ethylen-Dimerisierung über den Cossee–Arlman-Mechanismus (lineare Insertion) untersucht. Als aktivere Konfiguration wird ein $Ni^{II}O_2$-artiges Zentrum identifiziert, das eine verbesserte Zugänglichkeit für die Koordination von Ethylen bietet. Der C–C-Kupplungsschritt erweist sich als geschwindigkeitsbestimmender Schritt und führt zu Buten als Hauptprodukt. Das NU-1000-Gerüst begünstigt das Kettenwachstum bevorzugt in den kleineren trigonal geformten Poren im Vergleich zu den hexagonalen Mesoporen. Cluster-Modelle reproduzieren die periodischen Energien innerhalb von $4\,\mathrm{kJ\,mol^{-1}}$, was belegt, dass diese Modelle eine deutliche Rechenzeiteinsparung ermöglichen, ohne an Genauigkeit einzubüßen. Das auf TPSSh-Niveau korrigierte Freie-Energie-Profil

liefert Turnover-Frequenzen (TOFs), die in guter Übereinstimmung mit experimentellen Daten stehen. Die Erkenntnisse aus dem Ni-basierten System werden anschließend auf den Cr-NU-1000-Katalysator übertragen, wobei ein tetraedrisches Cr^{III}-Zentrum als aktives Zentrum für die Ethylen-Polymerisation identifiziert wird. Die Polymerisation verläuft über einen Cossee–Arlman-Mechanismus mit Insertionsbarrieren von lediglich etwa 50 kJ mol^{-1} und Kettenabbruchbarrieren, die diese Werte um mehr als 30 kJ mol^{-1} übersteigen – im Einklang mit den experimentellen Beobachtungen.

Im zweiten Teil werden kosteneffiziente maschinell gelernte Kraftfelder (MLFFs) implementiert, um die Phasenraumabtastung in zwei gründlichen Methoden der Freie-Energie-Bestimmung zu beschleunigen: der λ-thermodynamischen Integration (λ-TI) und dem Intermediate Hard-Sphere Model (IHSM). Dieser Ansatz überwindet die Zeitskalenlimitierungen konventioneller DFT-Molekulardynamik (DFT-MD), während die Genauigkeit auf DFT-Niveau erhalten bleibt. Zur Bewertung und Korrektur der MLFF-basierten Ergebnisse auf DFT-Genauigkeit werden ergänzende statistische Methoden eingesetzt, darunter Freie-Energie-Perturbation (FEP), die Machine-Learning-basierte Thermodynamische Perturbationstheorie (MLPT), das Bennett's Acceptance Ratio (BAR) sowie der Phasenraum-Überlappungsindex I_w. Die Robustheit dieser MLFF-gestützten Workflows wird umfassend validiert und auf verschiedene Adsorptionsphänomene angewendet, darunter die Chemisorption oxygenierter Spezies sowie die Physisorption und dissoziative Adsorption von Methan auf Pt(111). Die Adsorption von atomarem Sauerstoff zeigte ein nahezu harmonisches Verhalten und konnte durch die harmonische Näherung (HA) angemessen beschrieben werden, während die Adsorption von OH und OOH eine ausgeprägte Anharmonizität aufwies. Mit der MLFF-TI-Methode konnte bis zu 10 % des unter der HA geschätzten Entropieverlustes zurückgewonnen werden. Für die Adsorption von CH_4 wurde der Einfluss verschiedener Dichtefunktionale untersucht. BEEF-vdW, das tendenziell die Adsorptionsstärke unterschätzt, sagte höhere anharmonische Beiträge im Vergleich zu PBE-D3(BJ) voraus. Bei hohen Temperaturen oder im Physisorptionsregime, in dem das Adsorbat eine höhere Mobilität aufweist, erwies sich der Einfluss der Anharmonizität als nahezu unabhängig vom verwendeten Dichtefunktional. Mit dem MLFF-IHSM-Ansatz konnte die experimentell bestimmte Adsorptionsentropie für die Physisorption von CH_4 auf beiden Theorieebenen genau reproduziert werden, während sich die Adsorptionsentropien beider Ansätze bei etwa 700 K für den dissoziativen Adsorptionsprozess annäherten. Dieser Abschnitt verdeutlicht die Einschränkungen der harmonischen Näherung und betont die Notwendigkeit, genauere Methoden einzusetzen, die Anharmonizität zuverlässig erfassen können, um die Bestimmung von freien Adsorptionsenergien sowie der daraus abgeleiteten thermodynamischen und kinetischen Größen zu verbessern. Abschließend zeigt eine Proof-of-Concept-Studie, dass DFT-genaue anharmonische freie Energien mithilfe eines vortrainierten MLFF-Basismodells mit lediglich etwa 200 zusätzlichen DFT-berechnungeinzelner Strukturen erzielt werden können. Damit wird ein Weg zu übertragbaren und skalierbaren Anwendungen auf MOF-Systeme eröffnet, die auf allgemein einsetzbaren, maschinell gelernten Kraftfeldmodellen basieren.

Contents

Part I.

Introduction

1. Motivation

1.1. Context

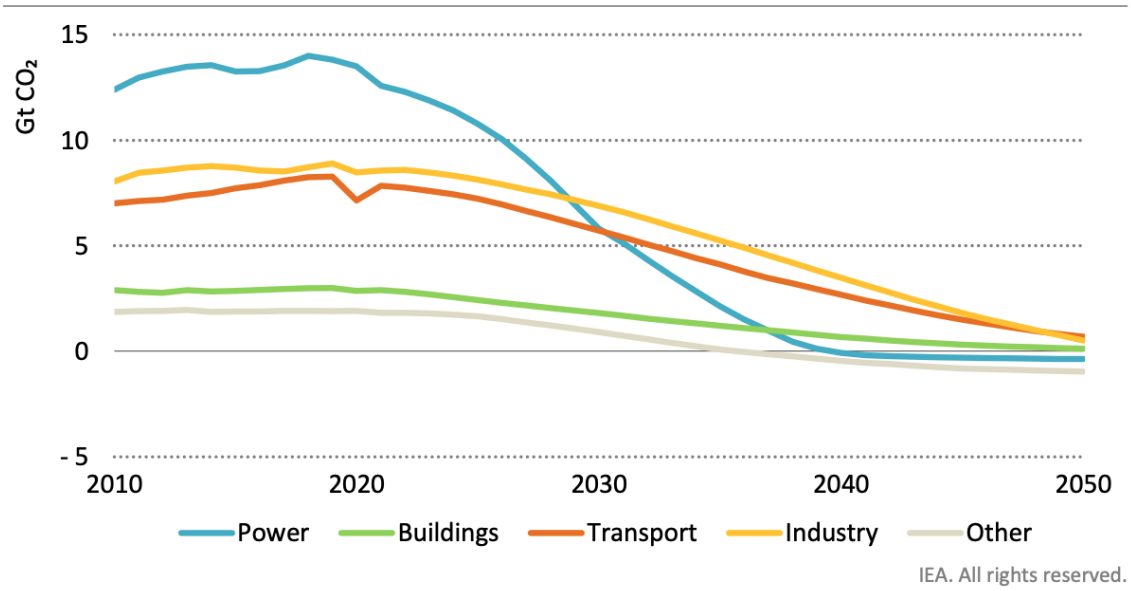


Figure 1.1.: Projection of CO_2 emission by sector in the Net-Zero Emissions by 2050 Scenario. Reprinted with permission from [1]. Copyright 2021 IEA (CC-BY)

The global energy and chemical sectors are undergoing a profound transformation driven by the urgent need to mitigate climate change and transition toward sustainable carbon utilization. For over a century, fossil resources, which supply fuels, energy, and essential materials, have underpinned modern society. However, their continued and expanding use has led to escalating greenhouse gas emissions and environmental degradation, not to mention their unsustainable long-term capacity reduction. To address these challenges, tremendous efforts have been committed to achieve a net zero carbon dioxide emissions by 2050 in energy sectors, aiming to limit the global temperature rise to 1.5° above pre-industrial levels[1, 2]. While progress in renewable electricity and green hydrogen has been rapid, certain sectors, such as transport and heavy industry, remain particularly challenging to decarbonize due to their high energy density requirements and reliance on carbon-based feedstocks (Figure 1.1)[1]. For example, aviation accounted for approximately 2.5% of total global CO_2 emissions from fossil fuel combustion in 2023[3], yet it continues to be one of the hardest sectors to abate. A promising route to mitigate these persistent

emissions is the development of a circular carbon economy, in which carbon is continually recycled rather than released.

Catalysis lies at the heart of these transformations, as it underpins more than 85% of all chemical production processes[4, 5]. Catalysts determine not only the rate but also the selectivity of chemical reactions, thereby directly influencing both carbon and energy efficiency. Therefore, advancing a sustainable circular economy requires a fundamental understand of catalyst active structures and their catalytic mechanisms, identification of performance limitations, and the rational design of improved catalytic systems.

In recent decades, the field has witnessed the convergence of new materials, innovative concepts, and advanced computational paradigms. Single-atom catalysts (SACs) have emerged as a frontier in atom-efficient catalysis, bridging homogeneous and heterogeneous regimes by offering site precision and structural robustness[6–8]. Their well-defined active sites not only maximize the utilization of (often precious) metals but also enable fundamental studies of catalytic mechanisms. Meanwhile, metal-organic frameworks (MOFs), with their modular architectures and tunable chemical environments, offer unprecedented opportunities for designing catalysts[9, 10]. Their crystalline structures and exceptionally high surface areas provides ideal platforms for hosting atomically dispersed metal species. The integration of SACs within MOFs thus combines the precision of single-atom catalysis with the structural versatility of MOFs, opening new avenues for both fundamental understanding and practical catalytic applications. These developments coincide with the rise of machine learning approaches[11], which enable accelerated and more accurate simulations, allowing for deeper insights into the reaction mechanisms and the dynamic behavior of active sites. Collectively, these emerging directions hold great promise for overcoming current limitations in catalyst design and advancing sustainable catalysis.

1.2. The Ethylene Economics

Ethylene (C_2H_4) is the cornerstone of the modern chemical industry and one of the most produced organic compounds worldwide, with an annual global output around 200 million tonnes[12, 13]. Its exceptional versatility as a platform molecule stems from the reactive C=C bond, which can be transformed into a broad range of higher-value products through catalytic routes. Ethylene serves as a primary feedstock for the production of polyethylene (PE), polyvinylethylene (PVC), ethylene dioxide, ethylene dichloride, and a variety of higher α-olefins[14]. The "ethylene economy" is often viewed as a central pillar of the global petrochemical sector and as a potential bridge toward more sustainable chemical manufacturing. Currently, ethylene is produced either by the energy-intensive steam cracking of naphtha from crude oil or the inexpensive cracking of ethane from shale gas. Recent advances in carbon recycling and alternative feedstock utilization have opened pathways for sustainable ethylene production, including methanol-to-olefins (MTO) process[15], CO_2 hydrogenation[16, 17], and electrochemical CO_2-to-ethylene conversion[18]. These technologies demonstrate how ethylene can serve as a key intermediate in a circular economy, linking renewable or waste-derived carbon sources to fuels and chemicals.

Ethylene oligomerization and polymerization represent two of the most economically and scientifically important transformations in catalytic chemistry. The oligomerization of ethylene into linear α-olefins (LAOs, C_4-C_{20}) provides key intermediates for detergents, lubricants, surfactants, plasticizers, and hydrocarbon fuels[19–21]. These higher olefins also serve as precursors for sustainable aviation fuels (SAFs) after hydroprocessing[22, 23]. Industrially, LAO production have relied predominantly on homogeneous transition-metal catalysts, especially complexes of nickel[19], titanium[24], or zirconium[25]. To improve the process economics and environmental sustainability, extensive effort has been devoted to developing heterogeneous catalysts for ethylene oligomerization, for example by incorporating active Ni center onto solid supports, such as sillica, alumina, or zeolites[26–37].

Meanwhile, ethylene polymerization produces polyethylene, which is by far the most widely used polymer material. Industrial polymerization relies primarily on three classes of catalysts: Phillips chromium catalysts [38, 39], Ziegler–Natta titanium catalysts [40, 41], and metallocene catalysts [40]. Among these, the heterogeneous Phillips catalyst remains one of the most successful and widely applied, accounting for nearly half of global polyethylene production [38, 39]. Despite its long industrial history, the structure and electronic nature of the chromium active site and the detailed polymerization mechanism remain incompletely understood. Among several proposed pathways, the Cossee–Arlman linear insertion and metallacycle mechanisms (Figure 1.2) are generally regarded as the most plausible, though their relative importance likely depends on the catalyst's local environment and oxidation state.

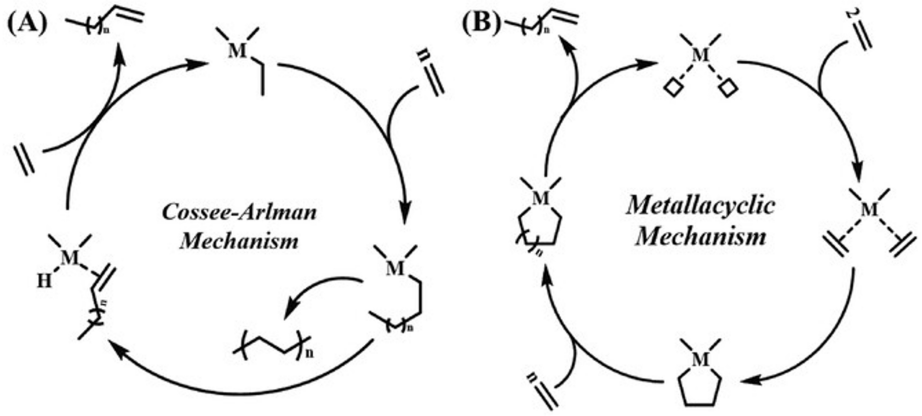


Figure 1.2.: The mechanisms proposed for catalytic ethylene oligomerization and polymerization. a) Cossee-Arlman and b) metallacyle mechanisms. Reprinted with permission from [42]. Copyright 2022 Wiley.

Understanding the fundamental chemistry of ethylene conversion holds both technological and scientific significance. Insights gained from studying these transformation in well-defined catalytic systems can accelerate the rational design of next-generation catalysts that combine high activity, selectivity, and stability with improved atom efficiency and sustainability.

1.3. The Emergence of Metal-Organic Frameworks in Catalysis

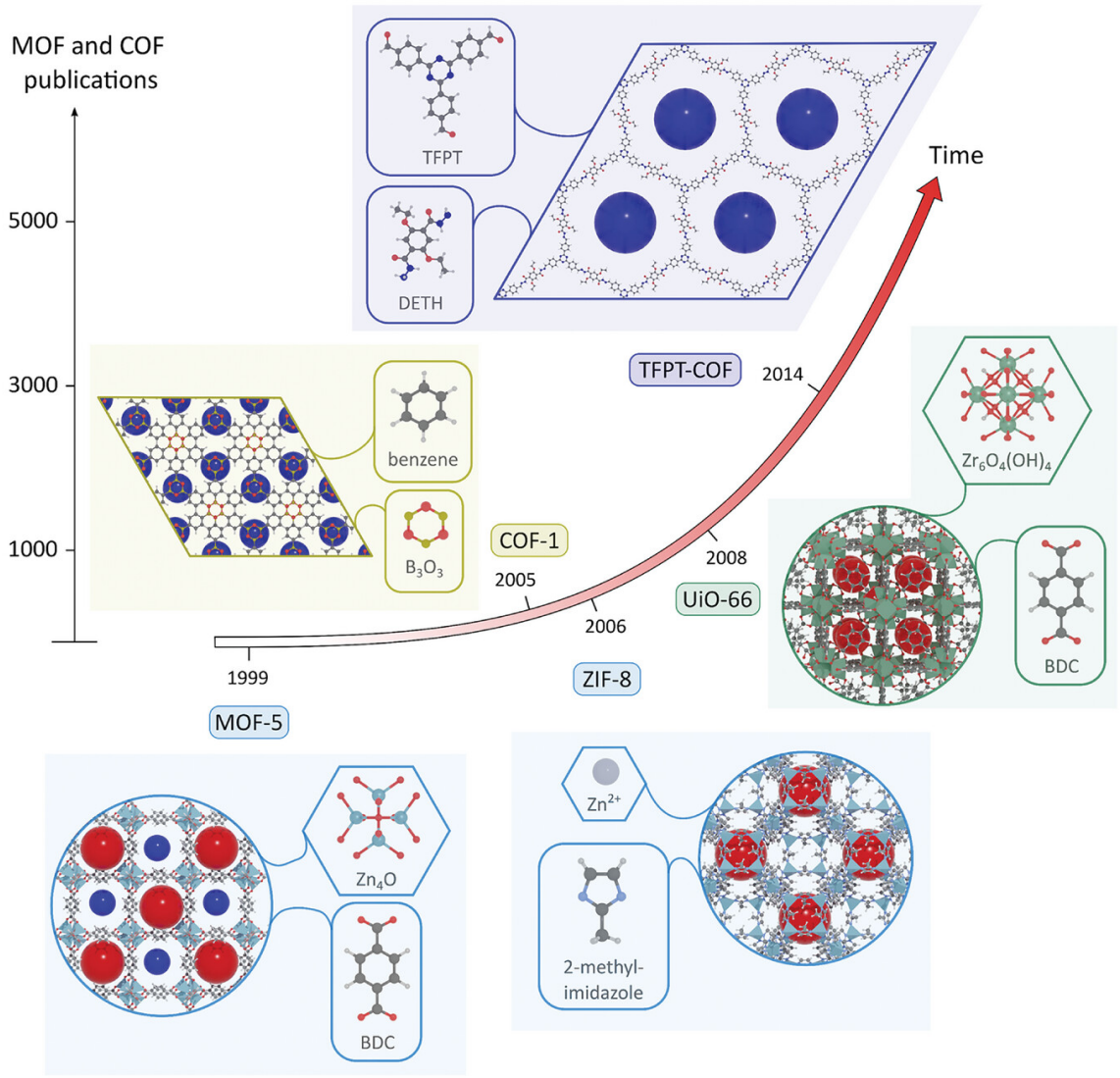


Figure 1.3.: The increase in MOF and COF publications over the year, together with some three-dimension (pore)structure of some representative MOFs (below the arrow) and COFs (above the arrow). Reprinted with permission from [43]. Copyright 2024 Wiley

In the past three decades, metal–organic frameworks (MOFs) have emerged as a versatile class of crystalline porous materials that bridge the structural precision of molecular chemistry with the robustness of extended solids. Figure 1.3 illustrates the development timeline of MOFs and covalent organic frameworks (COFs), another member of the reticular materials family, showing the exponential growth in publications on these two subclasses of reticular materials over the years. Their modular architectures, comprising metal nodes and organic linkers, offer an immense chemical space for structural and functional tunability, enabling the rational design of frameworks with tailored pore environments,

topology, and reactivity for specific catalytic transformations. Owing to their high surface areas, well-defined active sites, and adjustable microenvironments, MOFs can integrate the advantages of both homogeneous and heterogeneous catalysts, providing high activity, selectivity, and recyclability within a single platform. These unique attributes have made them particularly attractive as catalysts *per se*[44–48] or as solid supports for catalytic active species[49–52].

Among the diverse families of MOFs, zirconium-based frameworks have garnered particular attention due to their exceptional chemical and thermal robustness, which stems from the strong Zr–O coordination bonds [53–57]. This inherent stability under catalytic and hydrothermal conditions makes Zr-based MOFs, such as UiO-66 [$Zr_6(\mu_3{-}O)_4(\mu_3{-}OH)_4(BDC)_6$][58], MOF-808 [$Zr_6(\mu_3{-}O)_4(\mu_3{-}OH)_4(OH)_6(H_2O)_6(BTC)_2$][59], and NU-1000 [$Zr_6(\mu_3{-}O)_4(\mu_3{-}OH)_4(OH)_4(H_2O)_4(TBAPy)_2$] [60] (BDC = 1,4-benzenedicarboxylate, BTC = 1,3,5-benzenetricarboxylate, and TBAPy = 4,4′,4″,4‴-(pyrene-1,3,6,8-tetrayl)tetrabenzoate), ideal candidates for practical heterogeneous catalysis. Figure 1.4 illustrates the structures of these representative Zr-based MOFs.

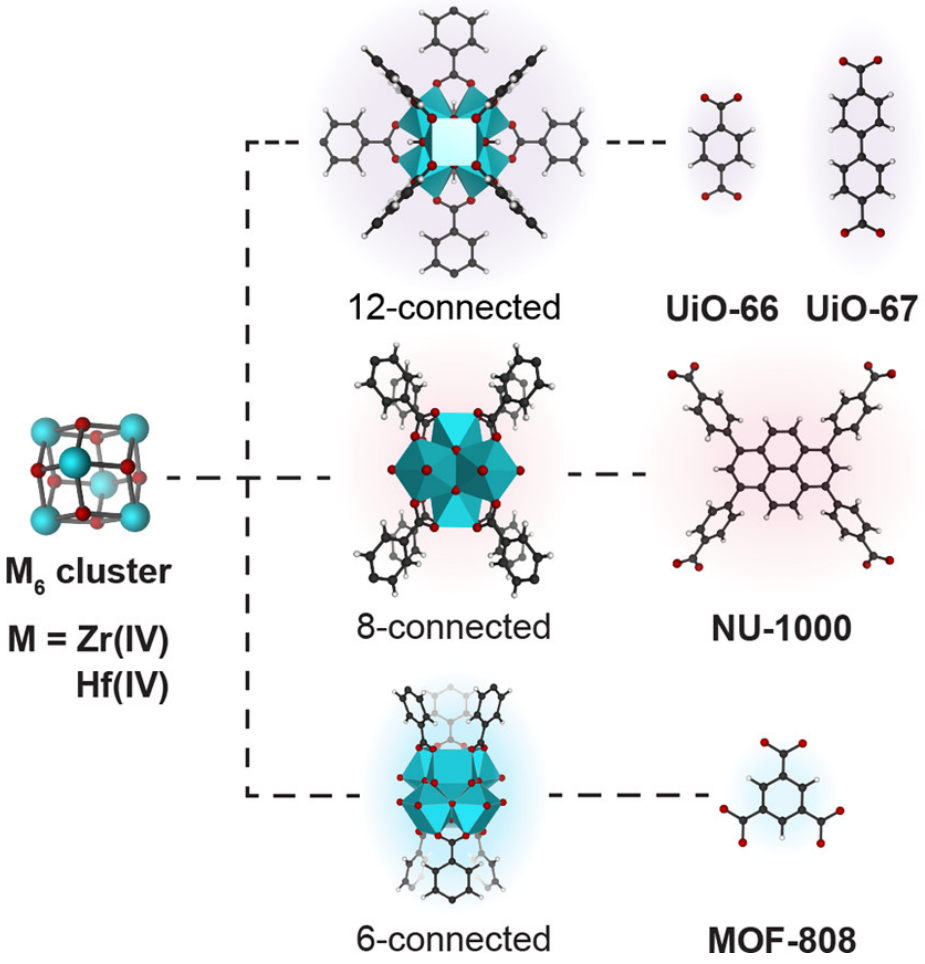


Figure 1.4.: Representative Zr-based MOFs. UiO-66: ditopic BDC ligands create a 12-connected SBU, NU-1000: tetratopic TBAPy ligands form 8-connected SBUs, and MOF-808: triptoic BTC ligands for a 6-connected SBU. Reprinted from Ref. [61] with permission. Copyright 2020 American Chemical Society.

The $Zr_6(\mu_3{-}O)_4(\mu_3{-}OH)_4$-based MOFs have demonstrated catalytic activity in a wide range of reactions, including butene hydrogenation and isomerization[62, 63], the degradation of warfare agents[64–72], biomass valorization[73–79], and other Lewis acid-catalyzed transformations[80–83]. Beyond their intrinsic activity, numerous strategies have been employed to enhance or diversify the catalytic functions of these frameworks. Such modifications include partial dehydration of Zr nodes to expose coordinatively unsaturated sites[63], defect engineering to create open metal sites and increase porosity[80–84], and linker functionalization to introduce additional reactive moieties[65, 85], among others.

The $Zr_6(\mu_3{-}O)_4(\mu_3{-}OH)_4$ nodes typically contain terminal hydroxyl groups that can serve as anchoring points for catalytically active metal species. In highly connected frameworks such as UiO-66, which features 12-fold coordination, defect engineering, by removing one or more linkers, is often necessary to create accessible sites for metal incorporation. Metal atoms can then be introduced via atomic layer deposition, solvent-assisted deposition, or impregnation techniques. A broad range of metals has been successfully deposited onto Zr-based MOFs such as NU-1000, UiO-66, and MOF-808, including Ir[86], Rh[87, 88], Pd[89], Pt[90], Au[91, 92], Ni[93–97], Co[98, 99], Cr[100–103], Cu[99, 104–110], Zn[111], and V[112]. These single-atom catalysts (SACs) have exhibited excellent activity and selectivity across diverse reactions, including CO oxidation[104, 106, 108], CO_2 reduction and hydrogenation[111], methane oxidation[110], oxidative dehydrogenation[98, 99, 109, 112], hydrogenation[90, 113], isomerization, hydroformylation[87], and biomass conversion[107]. Moreover, SACs supported on Zr-based MOFs have shown significant potential for olefin oligomerization[93, 94, 96, 97, 114, 115] and polymerization[100–102], with Ni- and Cr-based SACs demonstrating particularly high performance in these transformations.

Recent advances in experimental characterization have greatly deepened our understanding of single-metal sites in MOF-based catalysts[51, 116]. Techniques such as single-crystal X-ray diffraction (SC-XRD), *in situ* infrared spectroscopy, X-ray absorption spectroscopy (XANES/EXAFS), X-ray photoelectron spectroscopy (XPS), and aberration-corrected high-angle annular dark-field scanning transmission electron microscopy (HAADF-STEM) enable identification of atomic coordination environments, oxidation states, and dynamic structural evolution under reaction conditions. Complementary *operando* and time-resolved methods now provide direct insight into intermediate states and active-site transformations, offering a more complete picture of catalytic processes. In parallel, theoretical modeling, particularly density functional theory (DFT) and machine-learning-accelerated simulations, plays a crucial role in interpreting experimental data, elucidating reaction mechanisms, and guiding the rational design of next-generation single-atom catalysts.

1.4. Challenges in Modeling MOF Catalysis

Modeling catalysis in MOFs, particularly those incorporating transition metal SACs, presents significant challenges. The interdependence of model construction and the

choice of electronic structure method gives rise to a inherently multifaceted and multiscale problem. Furthermore, under operating conditions, both the framework and the guest species interacting with it undergo dynamic structural and electronic changes, depending sensitively on external variables, such as temperature and pressure. Consequently, the active sites may evolve across multiple geometrical configurations, spin states, and local environments. Capturing this complexity requires high-accuracy theoretical approaches that bridge static and dynamic representations of the catalyst, integrating first-principles accuracy with statistical sampling. The following sections outline how this multiscale nature shapes the choice of atomistic models, the selection of electronic structure methods, and the treatment of thermochemical effects necessary construct accurate catalytic free energy landscapes.

1.4.1. Representative Models

The first challenge in computational heterogeneous catalysis lies in constructing an accurate and representative model of the catalytic system that can adequately capture the elementary steps of the catalytic process. An optimal computational model must reflect both the chemical nature of the active site and the steric confinement imposed by the surrounding pore environment, while maintaining a balance between computational cost and model complexity to ensure tractable simulations. In the case of MOF catalysis, the model choice also reflects the researcher's conceptual perspective, whether the MOF is viewed as a molecule or an extended solid[61]. This raises a fundamental modeling dilemma: should the system be treated as a periodic solid, which provides a complete electronic structure description but is typically constrained to the less accurate generalized gradient approximation (GGA) functionals, or as a truncated cluster model, which enables the use of higher-level methods at the expense of a full representation of the extended framework? The answer to this question determines both the computational software and the level of electronic structure theory employed.

Cluster models have become the method of choice for representing catalytic structures when the focus lies on reactions occurring at isolated active sites[43, 117]. Yang et al. investigated Ir-based SACs on NU-1000 and UiO-66 for ethylene hydrogenation and dimerization using benzoate and acetate cluster models, where the linkers were presented by the respective benzoate and acerate ligands[113]. Similar benzoate model was also employed to explain the destruction mechanism of the nerve agent simulant dimethyl 4-nitrophenyl phosphate and the highly toxic chemical warfare agent GD (O-pinacolyl methylphosphonofluoridate) catalyzed by NU-1000 [68]. Bernales et al. elucidated the mechanism of ethylene dimerization on Ni-AIM-NU-1000 and Co-AIM-NU-1000 catalysts using formate clusters, where organic linkers were simplified to formate groups to reduce computational cost[114]. Their results revealed that the Cossee–Arlman mechanism was the most thermodynamically favorable pathway. Formate cluster models were employed to investigate butene dimerization on UiO-66-supported catalysts[95, 97], whereas acetate models have been used to study CO oxidation on Cu/UiO-66[104, 106, 108]. Owing to its computational inexpensiveness, truncated models have also been adopted for virtual

screening of catalysts. Ortuño et al. conducted a computational screening of first-row transition-metal SACs on NU-1000 for acceptorless alcohol dehydrogenation, identifying Ni- and Co-based systems as the most promising[118]. The screening was performed using a benzoate model of NU-1000. Ye et al. extended this approach by screening a broader range of transition-metal-based SACs on the same framework using a formate cluster model for ethylene dimerization, demonstrating that Ni^{II}-NU-1000 exhibits the highest activity, while Cr^{II}, Co^{III}, and Pd^{II} analogues are also potential candidates[119]. It is worth noting that smaller, further truncated models, derived from the formate cluster, have been employed to enable multireference calculations, to further improve the accuracy for systems with near-degenerate electronic structures. In the work of Bernales et al.[114], for example, a minimal model comprising two Zr atoms with only their first coordination spheres was adopted to perform CASSCF/CASPT2 calculations. These high-level computations attributed the superior activity of Ni catalyst to stabilizing hybridization between its empty *d* orbital and the 2*p* orbitals of the reactive carbons in the C-C coupling transition state.

Extended periodic models have also been employed, though to a lesser extent, in studies of Zr-based MOF catalysis. These models have been particularly useful for investigating catalytic processes occurring at open metal site and structural defects in UiO-66-based frameworks. Earlier studies employed periodic model to support the proposed mechanisms of thermal activation and defect formation[81, 82]. Hajek et al.[120] examined aldol condensation reactions in UiO-66 and UiO−66−NH_2 using both periodic and cluster models. They observed that confinement effects in the periodic systems stabilized adsorbed intermediates, resulting in energy shifts of approximately $20-30\,\mathrm{kJ\,mol^{-1}}$, primarily affecting the barriers associated with product desorption. The same group later employed a periodic model to study Fischer esterification on these catalysts, revealing that the presence of amine groups significantly reduced the reaction barriers[121]. For NU-1000, Planas and co-workers applied both periodic and cluster models to elucidate the proton topology within the framework[122]. Similarly, Ye and co-workers performed a periodic DFT study comparing ethylene dimerization on isolated Ni atoms and Ni_4 clusters in NU-1000, finding that both systems exhibited comparable catalytic activity[115].

When investigating dynamic processes, especially under industrially relevant conditions, the periodic representation is the model of choice. For example, periodic simulations were essential to probe the influence of nanoconfinement and hydrophilic adsorption sites in Zr-based MOFs on the dynamic hydrogen-bonded network of water[123].

The widespread use of cluster models is largely justified by the assumption that the local electronic structure dominates reactivity and that steric effects from the pore environment are minimal. Nevertheless, this assumption demands further validation, particularly under reaction conditions where dispersion interactions and confinement effects may play significant roles. Conflicting findings have been reported: for example, Vandichel et al.[82] concluded that cluster models were insufficient for modeling aldol condensations[82] and Oppenauer oxidation[80], whereas Le et al.[124] found that benzoate cluster models performed reasonably well compared with periodic models. To summarize, we cite Mancuso et al.[61]: *"the ideal model would be one that treats the MOF as a periodic solid, but*

can readily apply high levels of theory (typically limited to molecular software packages) to important components in the material."

1.4.2. Electronic Structure Methods

The second major challenge in modeling reactions within MOF catalysts concerns the selection of an appropriate electronic structure method capable of capturing the intricate electronic structure of the framework and active site. So far, density functional theory (DFT) has served as the primary "workhorse" due to its favorable balance between computational cost and accuracy[117, 125, 126]. The main limitation of DFT, however, is that the exact form of exchange-correlation (XC) functional is unknown. Consequently, the exact ground-state energy or density of a system cannot be computed. As a results, approximations to the XC functional must be made, leading to the development of a large library of density functional approximations. So far, most studies on SACs in Zr_6O_8-based MOFs have relied on generalized gradient approximation (GGA) and meta-GGA functionals. Cluster calculations are most commonly performed at M06-L level using mixing double- and triple-ζ basis sets[68–70, 83, 94, 95, 97, 113, 114, 118, 119, 122], occasionally with Grimme's D3 dispersion correction[104, 106, 108]. Periodic calculations, on the other hand, have predominantly employed the PBE functional, either with D2[127], with D3 dispersion corrections[80, 82, 115, 120, 121], or without any dispersion consideration[80, 122]. While these GGA and meta-GGA methods enable rapid exploration of catalytic reaction landscapes and high-throughput screening of candidate catalysts, they suffer from well-known deficiencies, including self-interaction errors and poor descriptions of strong correlation effects and highly localized *d*-electrons[61, 128], features that can be present in some MOF-based SACs. Hybrid functionals, incorporating a fraction of Hartree–Fock exchange, often provide improved electronic descriptions, especially for spin-polarized or semiconducting MOFs, albeit at significantly higher computational cost. However, the functionals in this rung, such as B3LYP, M06, or range-separated ω-B97X-D, have been used less frequently[69, 80, 120].

It is highly valuable to perform systematic functional assessments to identify the most suitable XC functional for a given catalytic reaction or catalyst. DFT results can be benchmarked either against experimental data or against higher level methods. Whereas comprehensive DFT validations for zeolite catalysis and other heterogeneous reactions are already available[129, 130], such benchmarks remain scarce for MOF-based systems. To make matters worst, a recent benchmark of 22 density functionals against experimental activation barriers for ethylene hydrogenation by Dohrmann *et al.*[131] revealed that DFT generally fails to reproduce the qualitative experimental trends for this low-barrier reaction. The authors, however, identified a subset of functionals showing reasonable accuracy, including M06-L, PBE0, and TPSSh. Le et al.[124] examined the impact of different functionals on the conversion of methyl levulinate to γ-valerolactone over defective UiO-66 and suggested the M06 functional as the most reliable method. Given the scarcity of such functional benchmarks for Zr-based MOF catalysis, systematic evaluations of this kind remain highly desirable.

On the other hand, further extending the accessible time and length scales in MOF catalysis remains challenging as even GGA-level DFT calculations are computationally demanding. More efficient approaches to approximating the potential energy surface (PES) are therefore required. While physics-based classical force fields (FFs), such as UFF4MOF[132–134] or MOF-FF[135], have long served as the primary tools for modeling sorptions and mass transport in MOFs, they are inherently unable to describe chemical reactions, which are essential for understanding catalytic processes under realistic conditions. Recent advances in machine learned force fields (MLFFs), also known as machine learned interaction potentials (MLIPs), trained from consistent sets of first-principles data without predefined functional forms, have opened new avenues for investigating the dynamics of MOF catalysis[136–139]. However, their applicability is currently limited by the need for large and computationally expensive DFT training datasets. Consequently, the development of MLFF architectures with improved data efficiency is necessary. Alternatively, emerging general purpose *foundation models*, such as MACE-MP-0[140] and UMA[141], also hold great promise for accelerating MOF simulations, though further refinement is still required.

1.5. Accounting for Thermal Effects

Another unresolved challenge in modeling catalytic reactions is how to properly account for thermal effects. In principle, accurately capturing these effects requires exhaustive sampling of the relevant potential energy surface (PES) at the experimental operating temperature to construct the partition function, from which thermodynamic quantities such as enthalpy, entropy, and Gibbs free energy are derived. In computational catalysis, several analytical models have traditionally been employed in static DFT studies to approximate thermochemical effects. Among these, the harmonic approximation (HA) has long been the standard approach. This method, however, is notorious for its susceptibility to high uncertainty. From a conceptual standpoint, HA assumes that all degrees of freedom, including adsorbate–surface interactions, can be represented as harmonic vibrations, allowing analytical construction of the partition function. The adsorbate under HA model represents the localized limit in translation and rotation, in contrast to the highly mobile adsorbate in the 2-dimensional (2D) gas model (also referred to as 2D free translator (FT)) and free rotor (FR) model. This assumption generally leads to underestimation of entropy for adsorbates exhibiting strongly anharmonic modes (*e.g.,* hindered translations, rotations, or framework soft modes). Moreover, the HA partition function depends exponentially on inverse of vibrational frequencies (in quasi-classical expression), meaning that low-frequency modes contribute significantly to partition function, entropy and free energy. Unfortunately, the numerical noise inherent in finite-difference calculations and the limitations of DFT become notable in this low-frequency region. As will be shown later, underestimation of these frequencies can lead to substantial overestimation of entropy and a corresponding underestimation of free energy. Thus, the uncertainty associated with the harmonic approximation can exhibit in either direction.

Several modifications have been proposed to address low-frequency limitations of HA. Truhlar and coworkers introduced the quasi-rigid-rotor harmonic oscillator (QRRHO) model, which raises all frequencies below a certain threshold (typically 100 cm^{-1}) to that value before constructing partition functions[142, 143]. However, the physical justification for this correction remains unclear. Grimme later proposed a more elegant way, also named QRRHO, to treat the modes with frequencies below a certain threshold as free rotors, providing a smooth interpolation between the harmonic and the free-rotor regimes[144]. This model allows more reliable evaluation of vibrational entropies and free energies. More recently, Otlyotov and Minenkov extended the Grimme's QRRHO by applying the interpolation approach to thermal corrections to internal energy[145]. Although these HA variants, initially introduced for molecular systems, mitigate the uncertainty in entropy and free energy, they still fail to address the underestimation of adsorption entropy.

To improve upon the HA model, numerous sophisticated methods have been further developed to describe adsorption on solid surfaces more accurately. Campbell and coworkers introduced the hindered translator (HT) and hinder rotor (HR) models, in which the PESs of adsorbate's in-plane translation and rotation are approximated by cosine functions with amplitudes and frequencies determined by the energy barrier and periodicity of corresponding mode[146]. This model provides a consistent bridge between the localized HA and mobile 2D FT (and FR) limits. To apply this model, the mobility of adsorbate must be known, and corresponding barriers and periodicities must be determined *a priori*. The Campbell's HTHR model was later simplified through additional approximations to reduce the number of required input parameters[147, 148]. Alternatively, Sauer's group pursued a different strategy to capture anharmonicity for adsorbate confined within zeolite cavities. They constructed the partition functions for each low-frequency surface-adsorbate modes by solving 1D Schrödinger equations along those modes and summing over all possible states[149–151]. A conceptually similar approach was later reported by Waitt et al. to evaluate translational entropies of adsorbates on Pt and Au surfaces[152].

Further progress has been achieved through approaches that perform direct samplings of surrogate translational and/or rotational PESs constructed from DFT data. Jørgensen and Grönbeck introduced the complete potential energy sampling (CPES) method to evaluate the translational entropy of CO and O on Pt(111). In a subsequent study, the same group employed Monte Carlo simulations for PES samplings to investigate the adsorption of N_2 and CH_4 in chabazite and MFI frameworks[154]. Bajpai and colleagues also adopted an approach similar to CPES, using a more sophisticated method to construct the underlying PES and applied it to the adsorption of several atomic species on Au and Pt surfaces[155]. Blöndal et al. developed the Monte Carlo phase (configuration) space integration method, in which the direct sampling is carried out via Monte Carlo integration of the PES. The method was successfully applied to the adsorption of H, CH_3OH, and CH_3 on Ni(111), and CO on Pt(111)[156–158].

Alternatively, molecular dynamics (MD) simulations can be employed to sample the potential energy surface (PES) and capture full system dynamics, including anharmonic effects. The quasi-harmonic approximation (QHA), based on the vibrational densities of states computed from MD trajectories, has been shown to improve the accuracy of

free energy calculations for ethanol adsorption in H-ZSM-5 zeolite[159]. Although QHA partially recovers anharmonic contributions, it fails to account for effects such as hindered translation[160]. Another MD-based technique, thermodynamic integration (TI)[161–163], can be used to compute the free energy difference between any two thermodynamic states. Amsler and co-workers developed several variants of the λ-TI method that enable recovery of anharmonic contributions to adsorption and activation free energies[160, 164]. They demonstrated efficient sampling of the DFT-derived PES and achieved converged adsorption free energies at reduced computational cost by employing carefully chosen internal coordinates. Very recently, Plessow proposed a complementary strategy based on free energy perturbation (FEP) theory to compute anharmonic adsorption free energies of weakly interacting reactants in zeolite cavities relative to the gas phase[165]. By introducing an intermediate hard sphere model (IHSM), the method enhances the phase space overlap between separated reference and interacting adsorbed systems. Nevertheless, for more complex systems, achieving convergence using λ-TI or IHSM at the DFT level remains a formidable computational challenge.

To address this challenge, Jinnouchi and co-workers introduced an efficient hybrid approach that combines machine learning (ML) with TI to compute adsorption and hydration free energies[166, 167]. In their workflow, MLFFs served as an efficient surrogate model, enabling accurate free energy calculations at a fraction of the DFT cost. However, the choice of reference state in their TI implementation introduced convergence difficulties, as the phase spaces of the reference and target systems, corresponding to species in vacuum and in solution, respectively, differed significantly. To overcome this limitation, they employed a more elaborate two-stage TI scheme along with the introduction of a model potential. For evaluating the anharmonic contributions to adsorption free energies, however, Amsler's λ-TI approaches remain particularly advantageous[160].

It is evident that the use of MLFFs offers a promising solution to alleviate the computational bottlenecks associated with sampling in λ-TI methods. Owing to their high efficiency and accuracy, MLFFs have increasingly been employed as reliable surrogate models for free energy calculations across diverse systems[166, 168–170].

1.6. The Scope of this Work

This thesis seeks to tackle the major challenges in computational catalysis within MOFs and heterogeneous systems at large, as outlined in the preceding discussion. Particular interest is placed on the adsorption and conversion of hydrocarbon, which serve as model reactions to bridge spatial and temporal scales relevant to catalysis. The thesis is organized into two main chapters, each addressing distinct but complementary aspects of the problem.

Chapter 3 presents a detailed theoretical investigations of ethylene oligomerization and polymerization catalyzed by single-atom catalysts (SACs) supported on NU-1000. In the first part, the reaction mechanism of ethylene dimerization on Ni-NU-1000 is investigated,

followed by thorough comparisons in model representations and density functional performance within a highly accurate hierarchical cluster framework. The study is then extended to tetramerization, applying the validated computational approach to evaluate the influence of NU-1000's pore structure on the reaction energetics and mechanism. In the second part, insights gained from the Ni-based system are transferred to elucidate the active site structure and the ethylene polymerization mechanism of the Cr-NU-1000 catalyst.

Chapter 4 focuses on overcoming limitations of the harmonic approximation by incorporating machine learned force fields (MLFFs) into free energy estimation methods, thereby accelerating necessary molecular dynamics (MD) simulations while retaining the high accuracy of the target DFT levels. Two complementary workflow, namely, MLFF-aided λ-thermodynamic integration (MLFF-TI) and MLFF-aided intermediate hard sphere model (MLFF-IHSM), are developed and applied to a variety of adsorption phenomena. These include the chemisorption of oxygenated species, the physisorption and dissociative adsorption of methane on the Pt(111) surface. The origin and magnitude of anharmonic effects, as well as the influence of temperature and density functional choice on the anharmonic free energies, are analyzed. The robustness of the developed workflows is meticulously validated, culminating in a proof-of-concept demonstration using a universal MLFF. This achievement paves the way for application of such MLFFs to MOF systems, where training a tailored force field remains computationally prohibitive.

2. Theory

This chapter provides preliminary but essential descriptions of the theoretical foundation on which the present work lies. Such a foundation can be divided into two domains, namely, electronic structure theory and statistical mechanics and thermodynamics. The division was made largely based on the scope within which the respective theory has been developed. In particular, electronic structure theory is concerned with describing electronic motions in a given microscopic state. Whereas, statistic mechanics and thermodynamics deal with observations at macroscopic systems and explain them by bridging the micro-macro scale gap by making use of statistics. The invocation of both theoretical domains facilitates an improved description of simulations closer to experimentally operating conditions. This improvement is of particular importance and interest in the field of computational catalysis.

2.1. Electronic Structure Theory

This section begins with a recapitulation of quantum chemistry, the solid foundation for most electronic structure methods that have been developed and employed in the modern computational chemistry research. Subsection 2.1.2 introduces two common wavefunction-based methods to solve the Schrödinger equation. Subsection 2.1.3 then presents density functional theory (DFT), which reformulates the problem in terms of the electron density, offering a more practical balance between accuracy and efficiency for large systems. Finally, Subsection 2.1.4 discusses the recent emergence of machine-learned force fields (MLFFs), which leverage data-driven approaches to approximate the potential energy surface (PES) with near *ab initio* accuracy at a fraction of the computational expense.

2.1.1. Quantum Chemistry

In quantum mechanics (QM), the total energy, E, of a quantum system can be computed by solving the time-independent Schrödinger equation:

$$\hat{\mathbf{H}}\boldsymbol{\Psi}(\mathbf{r},\mathbf{R}) = \mathrm{E}\boldsymbol{\Psi}(\mathbf{r},\mathbf{R}) \tag{2.1}$$

where $\Psi(\mathbf{r}, \mathbf{R})$ denotes the system's wavefunction, which depends on the coordinates of N electrons ($\mathbf{r}$) and M nuclei ($\mathbf{R}$). The Hamiltonian operator, $\hat{\mathbf{H}}$ is expressed in atomic units as:

$$\begin{aligned}\hat{\mathbf{H}} &= -\sum_{i=1}^{N}\frac{1}{2}\nabla_i^2 - \sum_{A=1}^{M}\frac{1}{2M_A}\nabla_A^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{|\mathrm{R}_A - \mathrm{r}_i|} + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{|\mathrm{r}_i - \mathrm{r}_j|} + \sum_{A=1}^{M}\sum_{B>A}^{M}\frac{Z_A Z_B}{|\mathrm{R}_A - \mathrm{R}_B|} \\ &= \hat{\mathbf{T}}_\mathrm{e} + \hat{\mathbf{T}}_\mathrm{n} + \hat{\mathbf{V}}_\mathrm{ne} + \hat{\mathbf{V}}_\mathrm{ee} + \hat{\mathbf{V}}_\mathrm{nn}. \end{aligned} \tag{2.2}$$

The first two terms correspond to the kinetic energies of the electrons and nuclei. The third term describes the attractive Coulomb interactions between electrons and nuclei, while the fourth and fifth terms present electron-electron and nucleus-nucleus repulsions, respectively.

Under the Born-Oppenheimer approximation, which assumes that nuclei remain relatively stationary relative to the much faster electronic motions, the electronic Schrödinger equation takes the form:

$$\hat{\mathbf{H}}_\mathrm{elec}\Psi_\mathrm{elec}(\mathbf{r}, \mathbf{R}) = E_\mathrm{elec}\Psi_\mathrm{elec}(\mathbf{r}, \mathbf{R}) \tag{2.3}$$

where $\hat{\mathbf{H}}_\mathrm{elec}$ denotes the electronic Hamiltonian, consisting of the first, third, and fourth terms on the right hand side of Eq. (2.2), and $\Psi_\mathrm{elec}(\mathbf{r}, \mathbf{R})$ is the electronic wavefunction with the nuclear coordinates $\mathbf{R}$ becoming fixed parameters. Within this framework, the Born–Oppenheimer approximation naturally introduces the concept of a potential energy surface (PES), on which the stationary points, such as minima and transition states, define the key nuclear geometries along a reaction pathway. Sampling of PESs is fundamental to computational chemistry, particularly in computational heterogeneous catalysis. Nevertheless, the electronic Schrödinger equation cannot be solved exactly for systems more complex than hydrogenic species, and thus various approximation methods must be employed. The following subsections will introduce several of these methods.

2.1.2. Wavefunction-based Methods

Slater determinant An electronic wavefunction (as in Eq. (2.3)) must satisfy the antisymmetry condition, which requires that the wavefunction changes its sign upon the interchange of any two electrons. This condition is fulfilled by adopting the Slater determinant *ansatz*, Φ_SD, where the columns correspond to single-electron wavefunctions (*i.e.,* spin-orbitals) and the rows represent electron coordinates. In a molecular system, the one-electron wavefunctions are referred to as molecular orbitals (MOs). The Slater determinant for a general system of N electrons occupying N spin-orbitals is given by:

$$\Phi_\mathrm{SD} = \frac{1}{\sqrt{N!}}\begin{vmatrix} \phi_1(1) & \phi_2(1) & \cdots & \phi_N(1) \\ \phi_1(2) & \phi_2(2) & \cdots & \phi_N(2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(N) & \phi_2(N) & \cdots & \phi_N(N) \end{vmatrix} \tag{2.4}$$

where $1/\sqrt{N!}$ is a normalization factor. The spin-orbitals ϕ_i are constructed as products of a spatial orbital and a spin function (α or β) and are required to be orthonormal ($\langle\phi_i|\phi_j\rangle = \delta_{ij}$).

Hartree Fock theory The Hartree-Fock (HF) theory forms the foundation of modern *ab initio* electronic structure theory. The method approximate the many-electron wavefunction by a single Slater determinant. The ground-state energy of a given system is determined variationally via:

$$E = \left\langle \Phi \left| \bar{\mathbf{H}} \right| \Phi \right\rangle = \sum_{i}^{N} \langle \phi_i | \, \hat{\mathbf{h}}_i \, | \phi_i \rangle + \sum_{i}^{N} \sum_{j>i}^{N} \left(\left\langle \phi_j \left| \hat{\mathbf{J}}_i \right| \phi_j \right\rangle - \left\langle \phi_j \left| \hat{\mathbf{K}}_i \right| \phi_j \right\rangle \right) + \hat{\mathbf{V}}_{\mathrm{nn}} \tag{2.5}$$

where Φ is the trial determinant. $\hat{\mathbf{h}}_i = -\frac{1}{2}\nabla_i^2 - \sum_A^M \frac{Z_A}{|\mathrm{R}_A - \mathrm{r}_i|}$ is the one-electron operator describing the kinetic energy and nuclear attraction of electron i. The operator $\hat{\mathbf{J}}_i$ and $\hat{\mathbf{K}}_i$ are two-electron operators, which accounts for Coulomb repulsion and exchange interaction, respectively. Minimization of the total electronic energy with respect to variations in the spin-orbitals, subject to orthonormality constraints, leads to the Hartree-Fock equations. This constrained optimization is performed using Lagrange multipliers and yields an effective one-electron operator known as the Fock operator,

$$\hat{\mathbf{F}}_i = \hat{\mathbf{h}}_i + \sum_{j=1}^{N} \left(\hat{\mathbf{J}}_j - \hat{\mathbf{K}}_j \right) \tag{2.6}$$

which incorporates the average electrostatic potential due to all other electrons. The HF equations take the form of coupled eigenvalue equations:

$$\hat{\mathbf{F}}\phi_i = \varepsilon_i \phi_i \tag{2.7}$$

where ε_i are orbital energies and ϕ_i are the canonical molecular orbitals. Because the Fock operator depends on the occupied orbitals themselves, the equations must be solved self-consistently, leading to the so-called self-consistent field (SCF) procedure.

Coupled Cluster The fundamental principle of coupled cluster (CC) theory is the exponential *ansatz* of CC wavefunction:

$$\mathbf{\Psi}_{\mathrm{CC}} = e^{\hat{\mathbf{T}}} \mathbf{\Phi}_0 \tag{2.8}$$

where $\mathbf{\Phi}_0$ denotes the reference wavefunction, typically obtained from a Hartree-Fock calculation, and $\hat{\mathbf{T}}$ is the cluster (excitation) operator, defined as:

$$\hat{\mathbf{T}} = \hat{\mathbf{T}}_1 + \hat{\mathbf{T}}_2 + \hat{\mathbf{T}}_3 + \cdots + \hat{\mathbf{T}}_N. \tag{2.9}$$

Here the $\hat{\mathbf{T}}_i$ denotes the level of excitation, each of which generates all possible determinants containing i excitations when acting on HF reference wavefunction. For example, the single and double excitations act as follows:

$$\hat{\mathbf{T}}_1\mathbf{\Phi}_0 = \sum_{i}^{occ}\sum_{a}^{vir} t_i^a \mathbf{\Phi}_i^a \tag{2.10}$$

and

$$\hat{\mathbf{T}}_2\mathbf{\Phi}_0 = \sum_{i<j}^{occ}\sum_{a<b}^{vir} t_{ij}^{ab} \mathbf{\Phi}_{ij}^{ab} \tag{2.11}$$

where $\mathbf{\Phi}_i^a$ and $\mathbf{\Phi}_{ij}^{ab}$ are singly and doubly excited determinants, respectively, and t are the corresponding amplitudes. The exponential operator $e^{\hat{\mathbf{T}}}$ can be expanded as a Taylor series:

$$e^{\hat{\mathbf{T}}} = \mathbf{1} + \hat{\mathbf{T}} + \frac{1}{2}\hat{\mathbf{T}}^2 + \frac{1}{6}\hat{\mathbf{T}}^3 + \cdots = \sum_{k=0}^{\infty}\frac{1}{k!}\hat{\mathbf{T}}^k. \tag{2.12}$$

Substituting Eq. (2.9) into Eq. (2.12) yields:

$$e^{\hat{\mathbf{T}}} = \mathbf{1} + \hat{\mathbf{T}}_1 + \left(\hat{\mathbf{T}}_2 + \frac{1}{2}\hat{\mathbf{T}}_1^2\right) + \left(\hat{\mathbf{T}}_3 + \hat{\mathbf{T}}_2\hat{\mathbf{T}}_1 + \frac{1}{6}\hat{\mathbf{T}}_1^3\right) + \ldots \tag{2.13}$$

The time-independent Schrödinger equation for the CC wavefunction can be written as:

$$\hat{\mathbf{H}}e^{\hat{\mathbf{T}}}\mathbf{\Phi}_0 = E_{\mathrm{CC}}e^{\hat{\mathbf{T}}}\mathbf{\Phi}_0 \tag{2.14}$$

In standard implementations of CC theory, the Schrödinger equation is then projected on the reference determinant by left-multiplying with $\mathbf{\Phi}_0^*$, leading to:

$$E_{\mathrm{CC}} = \left\langle \mathbf{\Phi}_0 \middle| \hat{\mathbf{H}}e^{\hat{\mathbf{T}}} \middle| \mathbf{\Phi}_0 \right\rangle \tag{2.15}$$

Expanding the exponential term and recognizing that $\hat{\mathbf{H}}$ consists of only one- and two-electron terms, Eq. (2.15) becomes:

$$E_{\mathrm{CC}} = E_0 + \sum_{i<j}^{occ}\sum_{a<b}^{vir}\left(t_{ij}^{ab} + t_i^a t_j^b - t_i^b t_j^a\right)\left(\left\langle\phi_i\phi_j|\phi_a\phi_b\right\rangle - \left\langle\phi_i\phi_j|\phi_b\phi_a\right\rangle\right). \tag{2.16}$$

Thus, the CC correlation energy can be determined entirely by the singles and doubles and the two-electron integrals.

In practical calculations, the cluster operator must be truncated to include only excitations up to a specific order. The widely used CC with single and double excitations (CCSD) method employs $\hat{\mathbf{T}} = \hat{\mathbf{T}}_1 + \hat{\mathbf{T}}_2$ and its computational cost scales approximately as N_{basis}^6. Higher-order effects, for example, due to connected triples, can be determined perturbatively and supplemented to the CCSD results, yielding the CCSD(T) method, which is often regarded as the "gold standard" method.

However, the CCSD(T) approach remains computationally demanding, if not infeasible, for large systems, especially those containing transition metal centers. The DLPNO-CCSD(T) method is a highly efficient, local correlation approximation to the canonical CCSD(T) approach, designed to achieve near-linear scaling with system size. It exploits the locality of electron correlation by constructing pair-specific natural orbital domains, allowing only the most significant electron pairs to be treated explicitly. This leads to a substantial reduction in computational cost while maintaining excellent accuracy, typically recovering over 99.8% of the CCSD(T) correlation energy[171].

2.1.3. Density Functional Theory

The Hohenberg-Kohn theorems The Hohenberg–Kohn (HK) theorems establish the theoretical foundation of density functional theory (DFT) by demonstrating that the ground-state properties of a many-electron system are uniquely determined by its electron density, rather than by the many-electron wavefunction[172]. The first Hohenberg–Kohn theorem states that the external potential $V_{\text{ext}}(\mathbf{r})$, and hence the energy of the system, are uniquely determined (up to an additive constant) by the ground-state electron density $\rho_0(\mathbf{r})$. Consequently, the total energy of a system can be expressed as a unique functional of the electron density, $E[\rho]$. The second Hohenberg–Kohn theorem establishes a variational principle for this energy functional: among all possible electron densities $\rho(\mathbf{r})$ that integrate to the correct number of electrons, the ground-state density $\rho_0(\mathbf{r})$ minimizes the total energy functional,

$$E_{\text{HK}}[\rho] = \left\langle \Psi\left[\rho\right] \middle| \hat{\mathbf{H}} \middle| \Psi\left[\rho\right] \right\rangle = F_{\text{HK}}[\rho] + \int V_{\text{ext}}(\mathbf{r})\rho(\mathbf{r})\, d\mathbf{r}, \tag{2.17}$$

where $\hat{\mathbf{H}} = \hat{\mathbf{T}} + \hat{\mathbf{V}}_{\text{ee}} + \hat{\mathbf{V}}_{\text{ext}}$, and $F_{\text{HK}}[\rho]$ is a universal functional incorporating the kinetic energy and electron–electron interaction, independent of the external potential.

Together, these theorems transform the complex many-body Schrödinger equation, dependent on $3N$ electronic coordinates, into a problem involving only the three spatial variables of the electron density.

The Kohn-Sham approach The Kohn–Sham (KS) formulation provides a practical framework for applying density functional theory (DFT) by mapping the complex, interacting many-electron system onto an equivalent auxiliary system of non-interacting electrons that reproduces the same ground-state electron density.

In the KS framework, the total electronic energy functional is written as

$$E[\rho] = T_s[\rho] + E_{\text{ext}}[\rho] + J[\rho] + E_{\text{xc}}[\rho], \tag{2.18}$$

where $T_s[\rho]$ is the kinetic energy of the non-interacting reference system, $E_{\text{ext}}[\rho]$ represents the electron–nuclear interaction energy, $J[\rho]$ is the classical Coulomb (Hartree) energy, and $E_{\text{xc}}[\rho]$ is the exchange–correlation functional. The latter contains all the

many-body effects missing from the non-interacting model, including the difference between the exact and non-interacting kinetic energies, as well as exchange and correlation contributions beyond the mean-field approximation.

Minimization of $E[\rho]$ with respect to the electron density, under the constraint of orbital orthonormality, leads to the Kohn–Sham equations, a set of self-consistent one-electron equations of the form

$$\left[-\frac{1}{2}\nabla^2 + V_{\text{eff}}(\mathbf{r})\right]\phi_i(\mathbf{r}) = \varepsilon_i\phi_i(\mathbf{r}), \tag{2.19}$$

where $\phi_i(\mathbf{r})$ are the KS orbitals and ε_i their associated orbital energies. The effective potential, $V_{\text{eff}}(\mathbf{r})$, comprises three contributions,

$$V_{\text{eff}}(\mathbf{r}) = V_{\text{ext}}(\mathbf{r}) + V_{\text{H}}(\mathbf{r}) + V_{\text{xc}}(\mathbf{r}), \tag{2.20}$$

with V_{ext} being the external potential (typically electron–nucleus attraction), $V_{\text{H}}(\mathbf{r}) = \int \frac{\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}'$ the Hartree potential, and $V_{\text{xc}}(\mathbf{r}) = \delta E_{\text{xc}}[\rho]/\delta\rho(\mathbf{r})$ the exchange–correlation potential.

The electron density is reconstructed from the KS orbitals as $\rho(\mathbf{r}) = \sum_i^{\text{occ}} |\phi_i(\mathbf{r})|^2$. Because both V_{H} and V_{xc} depend on the density, the KS equations must be solved iteratively until self-consistency is achieved.

While the KS formalism provides an exact reformulation of the many-body problem in principle, its accuracy in practice depends entirely on the approximation chosen for $E_{\text{xc}}[\rho]$. This term embodies the quantum mechanical effects of exchange, correlation, and residual kinetic energy, and its precise form remains unknown. The development of accurate exchange–correlation functionals is therefore the central challenge and ongoing focus of DFT research.

2.1.4. Machine Learned Force Fields

DFT offers an excellent compromise between accuracy and computational cost; however, its inherent cubic scaling, $O(N^3)$, limits its applications to systems with modest temporal and spatial dimensions. Recent progress in machine learned force fields (MLFFs), also referred to as machine learning interatomic potentials (MLIPs), has largely overcome these constraints, transforming molecular simulations across chemistry, biology, materials science, physics, and related fields. This advancement stems from the ability of MLFFs to achieve near-DFT accuracy in predicting energies and forces while maintaining computational efficiency comparable to that of traditional physics-based force fields. Among the various approaches to construct MLFFs, Gaussian process regression (GPR) and artificial neural networks are the two most widely adopted methodologies. The following subsection outlines the theoretical framework of the MLFF implemented in VASP, which employs GPR as its foundation (see Ref. [173–175]). Readers seeking a more comprehensive discussion of neural network models, other GPR-based MLFF architectures, and their practical applications are referred to several excellent review articles[11, 176–187].

Overview The machine-learned force field (MLFF) developed by Jinnouchi et al.[173–175] and implemented in VASP[188–191] constructs a force field by combining the flexibility of GPR with the statistical rigor of Bayesian inference. The total potential energy of a system is expressed as a sum of atomic contributions, each depending solely on the local atomic environment within a predefined cutoff radius R_{cut}:

$$\mathcal{V} = \sum_i \mathcal{V}_i. \tag{2.21}$$

Each atomic energy $\mathcal{V}_i$ is modeled as a functional of a set of descriptors that encode both radial and angular features of the surrounding atoms, representing the atom's local environment.

Local Environment Representation To characterize local atomic environments, each atom i is associated with a continuous atomic density,

$$\rho_i(\mathbf{r}) = \sum_j f_{\text{cut}}(r_{ij})\, g(\mathbf{r} - \mathbf{r}_{ij}), \tag{2.22}$$

where $f_{\text{cut}}(r_{ij})$ is a cutoff function that smoothly decays to zero beyond R_{cut} and $g(\mathbf{r})$ is a Gaussian smoothing function:

$$g(\mathbf{r}) = \frac{1}{\sqrt{2\sigma_{\text{atoms}}\pi}} \exp\left(-\frac{|\mathbf{r}|^2}{2\sigma_{\text{atom}}^2}\right). \tag{2.23}$$

The density, $\rho(\mathbf{r})$, can be interpreted as the probability density of finding atom j at position $\mathbf{r}$ around the atom i within a radius R_{cut}. $\mathbf{r}_i$, $\mathbf{r}_{ij}$ and r_{ij} denotes the position vector of i, the direction vector from atom i to atom j, and the distance between the two atoms, respectively. Two types of $\rho(r)$-dependent descriptors are introduced (Figure 2.1) to ensure rotational invariance, namely the radial (Eq. (2.24)) and angular distribution functions (Eq. (2.25)).

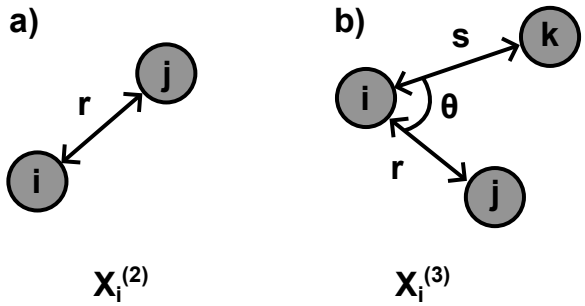


Figure 2.1.: a) Radial and b) angular descriptors. Reproduced from Ref. [173] with permission. Copyright 2019 American Physical Society.

$$\rho_i^{(2)}(r) = \frac{1}{4\pi} \int \rho_i(r\hat{\mathbf{r}})\, d\hat{\mathbf{r}}, \tag{2.24}$$

$$\rho_i^{(3)}(r,s,\theta) = \iint \delta\left(\hat{\mathbf{r}}\cdot\hat{\mathbf{s}} - \cos\theta\right)\rho_i(r\hat{\mathbf{r}})\rho_i^*(s\hat{\mathbf{s}})\,d\hat{\mathbf{r}}d\hat{\mathbf{s}}, \tag{2.25}$$

The density can be expanded in a basis of orthogonal radial functions $\chi_{nl}(r)$ and spherical harmonics $Y_{lm}(\hat{r})$:

$$\rho_i(\mathbf{r}) = \sum_{nlm} c^i_{nlm}\chi_{nl}(r)Y_{lm}(\hat{r}), \tag{2.26}$$

where the expansion coefficients c^i_{nlm} quantify the distribution of neighboring atoms in the local environment, n, l, and m denote radial numbers, angular and magnetic quantum numbers, respectively. Using the above equation, the radial and angular descriptors can be rewritten as:

$$\rho_i^{(2)}(r) = \frac{1}{\sqrt{4\pi}}\sum_{i=1}^{N_R^0} c^i_{n00}\chi_{nl}(r), \tag{2.27}$$

$$\rho_i^{(3)}(r,s,\theta) = \sum_{l=1}^{L_{max}}\sum_{n=1}^{N_R^l}\sum_{v=1}^{N_R^l}\sqrt{\frac{2l+1}{2}}p^i_{nvl}\chi_{nl}(r)\chi_{vl}(s)P_l(\cos\theta) \tag{2.28}$$

where the χ_{vl} and P_l denote the normalized spherical Bessel functions and Legendre polynomials of order l, respectively. The p^i_{nvl} is the expansion coefficients of $\rho_i^{(3)}$ with respect to the orthonormal radial and angular basis functions:

$$p^i_{nvl} = \sqrt{\frac{8\pi^2}{2l+1}}\sum_{m=-l}^{l} c^i_{nlm}c^{i*}_{vlm}. \tag{2.29}$$

Atomic Energy Model Each local atomic energy is expressed as a functional of these descriptors:

$$\mathcal{V}_i = F[\rho_i^{(2)}, \rho_i^{(3)}], \tag{2.30}$$

and is approximated by expansion over a finite set of reference atomic environments $\{\mathbf{X}_{i_B}\}$, converted from a set of N_B local reference structures:

$$F[\rho_i] = \sum_{i_B=1}^{N_B} w_{i_B}K(\mathbf{X}_i, \mathbf{X}_{i_B}). \tag{2.31}$$

Here, $\mathbf{X}_i$ collects all descriptor coefficients of the local environment, w_{i_B} are the regression weights, and $K(\mathbf{X}_i, \mathbf{X}_{i_B})$ quantifies similarity between atomic environments.

The kernel combines a linear (pairwise) and a nonlinear (many-body) term,

$$K(\mathbf{X}_i, \mathbf{X}_{i_B}) = \left[\beta(\mathbf{X}_i^{(2)}\cdot\mathbf{X}_{i_B}^{(2)}) + (1-\beta)(\hat{\mathbf{X}}_i^{(3)}\cdot\hat{\mathbf{X}}_{i_B}^{(3)})\right]^{\zeta}, \tag{2.32}$$

where $\hat{\mathbf{X}}_i^{(3)}$ denotes the normalized angular descriptor vector, and β and ζ are tunable parameters to control the relative weighting of the radial and angular terms and sharpness of the kernel, respectively.

Bayesian Regression Framework The parameters $\mathbf{w} = \{w_{i_B}\}$ are determined by a Bayesian linear regression procedure. Given reference data $\mathbf{Y}$ (including energies, forces, and stress tensors), the model assumes a linear relationship:

$$\mathbf{Y} = \Phi\mathbf{w} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma_v^2 I), \tag{2.33}$$

where the design matrix Φ collects kernel evaluations and their derivatives, and a Gaussian prior $\mathbf{w} \sim \mathcal{N}(0, \sigma_w^2 I)$ imposes regularization.

Bayesian inference yields the posterior distribution of the regression weights:

$$p(\mathbf{w}|\mathbf{Y}) = \mathcal{N}(\bar{\mathbf{w}}, \Sigma), \tag{2.34}$$

with

$$\bar{\mathbf{w}} = \frac{1}{\sigma_v^2}\Sigma\boldsymbol{\Phi}^T\mathbf{Y}, \qquad \Sigma^{-1} = \frac{1}{\sigma_w^2}\mathbf{I} + \frac{1}{\sigma_v^2}\boldsymbol{\Phi}^T\boldsymbol{\Phi}. \tag{2.35}$$

The mean vector $\bar{\mathbf{w}}$ corresponds to the optimal regression coefficients, while Σ encodes the uncertainty in these parameters.

Predictions for a new configuration are obtained from:

$$p(\mathbf{y}|\mathbf{Y}) = \mathcal{N}(\phi\bar{\mathbf{w}}, \sigma), \quad \sigma = \sigma_v^2\mathbf{I} + \phi^T\Sigma\phi. \tag{2.36}$$

The predictive mean yields the model's estimated energies and forces, while the covariance term quantifies the uncertainty of the prediction. The latter serves as the key criterion for determining when additional DFT calculations are required during on-the-fly learning.

2.2. Statistical Mechanics and Thermodynamics

2.2.1. Partition Function and Thermodynamic Functions

Partition function The partition function provides a measure of the total number of accessible microscopic states of a system and serves as the foundation for connecting microscopic properties to macroscopic thermodynamic observables. The canonical partition function for a system with discrete quantum states is defined as:

$$Q(N, V, T) = \sum_i g_i \exp\left(-\frac{\varepsilon_i}{k_\mathrm{B}T}\right), \tag{2.37}$$

where g_i is the degeneracy and ε_i is the energy of the i-th quantum state. At sufficiently high temperatures, where quantum effects become negligible, the system can be described by the classical partition function:

$$Q(N, V, T) = M \int \exp\left(\frac{-\mathcal{H}}{k_\mathrm{B}T}\right) d\boldsymbol{p}\, d\boldsymbol{q}, \tag{2.38}$$

where $\mathcal{H}$ is the classical Hamiltonian of the system, k_B is the Boltzmann constant, M is a normalization factor.

Thermodynamic State Functions Once the partition function is known, all important thermodynamic state functions can be derived from it. The relationships for quantities of primary interest in this work, namely the internal energy (U), Helmholtz free energy (A), enthalpy (H), entropy (S), and Gibbs free energy (G), are given below.

$$U = k_\mathrm{B}T^2 \left(\frac{\partial \ln Q}{\partial T}\right)_V \tag{2.39}$$

$$A = -k_\mathrm{B}T \ln Q \tag{2.40}$$

$$H = k_\mathrm{B}T^2 \left(\frac{\partial \ln Q}{\partial T}\right)_V + k_\mathrm{B}TV \left(\frac{\partial \ln Q}{\partial V}\right)_T \tag{2.41}$$

$$S = k_\mathrm{B}T \left(\frac{\partial \ln Q}{\partial T}\right)_V + k_\mathrm{B} \ln Q \tag{2.42}$$

$$G = H - TS = k_\mathrm{B}TV \left(\frac{\partial \ln Q}{\partial V}\right)_T - k_\mathrm{B}T \ln Q. \tag{2.43}$$

Separation of degrees of freedom Following the ideal gas approximation, as is commonly applied for molecular systems, the problem of determining the ensemble partition function Q can be reduced to finding the *molecular* partition function q. This treatment can also be extended to periodically extended systems considered in this work. The problem can be further simplified by assuming the electronic, vibrational, rotational, and translational degrees of freedom are separable. Consequently, the molecular partition function can be expressed as the product of the corresponding individual contributions, each of which can be evaluated analytically.

$$q = q_\mathrm{trans} \cdot q_\mathrm{rot} \cdot q_\mathrm{vib} \cdot q_\mathrm{elec} \tag{2.44}$$

Translational partition function The translational partition function q_trans describes the motion of a particle with mass m in D dimensions, where $D = 3$ for a gas and $D = 2$ for a free translator on a surface. It depends on the characteristic length L (volume of a gas, area of a surface) and the thermal de Broglie wavelength Λ.

$$q_\mathrm{trans} = \frac{L^D}{\Lambda^D} \quad \text{with } \Lambda = \frac{h}{\sqrt{2\pi m k_\mathrm{B}T}} \tag{2.45}$$

Rotational partition function The rotational partition function depends on moments of inertia, $\mathbf{I}$, rotational symmetry number, σ, and temperature T. The analytical expressions for linear and nonlinear molecules are given in Eq. (2.46) and (2.47), respectively.

$$q_\mathrm{rot} = \frac{8\pi^2 I k_\mathrm{B}T}{h^2 \sigma} \tag{2.46}$$

$$q_\mathrm{rot} = \frac{\sqrt{\pi}}{\sigma} \left(\frac{8\pi^2 k_\mathrm{B}T}{h^2}\right)^{3/2} \sqrt{I_1 I_2 I_3} \tag{2.47}$$

Vibrational partition function Each vibrational mode i with angular frequency ω_i can be approximated as either a *quantum* or *classical* harmonic oscillator, with the distinction becoming negligible when $k_\mathrm{B}T \gg \hbar\omega_i$. The vibrational partition function for the i-th mode is given by the quantum and classical expressions in Eqs. (2.48) or (2.49), respectively.

$$q_{\mathrm{vib},i} = \frac{\exp\left(\frac{-\hbar\omega_i}{2k_\mathrm{B}T}\right)}{1-\exp\left(\frac{-\hbar\omega_i}{k_\mathrm{B}T}\right)} \tag{2.48}$$

$$q_{\mathrm{vib},i} = \frac{k_\mathrm{B}T}{\hbar\omega_i} \tag{2.49}$$

The total vibrational partition function is obtained as the product of all individual vibrational modes:

$$q_\mathrm{vib} = \prod_i q_{\mathrm{vib},i} \tag{2.50}$$

Electronic partition function The electronic partition function accounts for the occupation of electronic quantum states, given by:

$$q_\mathrm{elec} = g_0 \exp\left(-\frac{\varepsilon_0}{k_\mathrm{B}T}\right) \tag{2.51}$$

where g_0 is the multiplicity corresponding to the ground state with energy ε_0.

2.2.2. Transition State Theory

Potential energy surface – the E-representation Within the Born–Oppenheimer approximation, a chemical reaction can be visualized as the motion of nuclei on a potential energy surface (PES). Stationary points on this surface, such as local minima and first-order saddle points, represent possible intermediates or transition structures that the reaction may pass through on its way from reactants to products. The energy difference between any two points determines the relative likelihood of moving between them, according to the Boltzmann distribution.

Rate constant – the k-representation Consider a bimolecular reaction of the type A + B $\longrightarrow$ C + D. The rate of reaction can be expressed as:

$$\frac{\mathrm{d[C]}}{\mathrm{d}t} = \frac{\mathrm{d[D]}}{\mathrm{d}t} = -\frac{\mathrm{d[A]}}{\mathrm{d}t} = -\frac{\mathrm{d[B]}}{\mathrm{d}t} = k\mathrm{[A][B]}, \tag{2.52}$$

where k is the rate constant, which can often be determined experimentally through kinetic measurements.

Bridging the two representations Transition state theory (TST) provides a semi-classical framework that links the reaction rate constant of a rare event to the underlying potential energy surface, thereby bridging the E- and k-representations. In TST, the PES is conceptually divided into two distinct regions: the reactant region and the product region. The $(3N-7)$-dimensional hyperplane that separates these regions is defined as the transition state (TS). At thermal equilibrium, the system is assumed to follow a Boltzmann energy distribution across all accessible quantum states along the reaction coordinate. Because the transition state is in equilibrium with the reactants, the rate constant can be expressed by the Eyring equation:

$$k = \frac{k_\mathrm{B}T}{h} \exp\left(-\frac{\Delta G^\ddagger}{k_\mathrm{B}T}\right), \tag{2.53}$$

where $\Delta G^\ddagger$ is the Gibbs free energy of activation, defined as the free energy difference between the transition state and the reactants. In catalytic cycles, the Eyring equation is used to evaluate the rate constants of the individual elementary steps.

Energetic span model Building upon the TST, the energetic span model enables the calculation of the turnover frequency (TOF) of catalytic cycles from their computed energy profiles. The TOF is governed by two key points on the energy landscape: the TOF-determining transition state (TDTS) and the TOF-determining intermediate (TDI). The free energy difference between these two points, referred to as the *energetic span* (δG), represents the apparent activation free energy of the overall catalytic cycle. The TOF can thus be expressed by an equation analogous to the Eyring equation:

$$\mathrm{TOF} = \frac{k_\mathrm{B}T}{h} \exp\left(-\frac{\delta G}{k_\mathrm{B}T}\right). \tag{2.54}$$

2.2.3. Molecular Dynamics

Phase space The (classical) dynamics of an N-particle system can be fully described by specifying a set of $6N$ time-dependent functions, $\{\boldsymbol{q}_1(t), \ldots, \boldsymbol{q}_N(t), \boldsymbol{p}_1(t), \ldots, \boldsymbol{p}_N(t)\}$, where $\boldsymbol{q}_i(t)$ and $\boldsymbol{p}_i(t)$ denote the position and momenta vectors of particle i, respectively. At any given instant of time t, the complete state of the system is therefore determined by $6N$ numerical values, which together define the *microscopic state* of the system. These $6N$ numbers can alternatively be viewed as the coordinates of a single point in a $6N$-dimensional space, known as the *phase space*, and represented as:

$$\mathrm{x} = (\boldsymbol{q}_1, \ldots, \boldsymbol{q}_N, \boldsymbol{p}_1, \ldots, \boldsymbol{p}_N). \tag{2.55}$$

By integrating the equations of motion, one obtain the time evolution of this point

$$\mathrm{x}(t) = (\boldsymbol{q}_1(t), \ldots, \boldsymbol{q}_N(t), \boldsymbol{p}_1(t), \ldots, \boldsymbol{p}_N(t)) \equiv \mathrm{x}_t, \tag{2.56}$$

which traces a *trajectory* through phase space.

Ensemble An *ensemble* refers to a statistical collection of systems that share the same microscopic interactions and common macroscopic properties, such as total energy, volume, and number of particles. The canonical (NVT) ensemble, in particular, describes a closed system that can exchange heat with a thermal reservoir while maintaining fixed composition, volume, and temperature. Once an ensemble is defined, macroscopic observables are obtained by ensemble averaging, that is, by taking statistical averages over all systems that belong to the ensemble.

Molecular Dynamics simulations Molecular Dynamics (MD) is a numerical technique used to solve the Newtonian equations of motion (Eq. (2.57)) for a many-body system, thereby describing its time evolution. At each time step, information about the positions and momenta, which together define a point in phase space (Eq. (2.55)), is required to propagate the system to the next step. MD simulations thus enable efficient sampling of the phase space. The Hamiltonian governing the time evolution of the system may be classical, *ab initio*, or machine learning–based in nature.

$$m_i \ddot{\boldsymbol{r}}_i = \boldsymbol{F}_i \tag{2.57}$$

Ergodic hypothesis The *ergodic hypothesis*, which underpins the MD method, states that ensemble averages are equivalent to time averages in the limit of infinite simulation time. That is, the ensemble average of an observable O can be determined as:

$$\langle O \rangle = \lim_{t\to\infty} \frac{1}{t} \int_0^t O\left(\boldsymbol{q}(t), \boldsymbol{p}(t)\right)\, dt = \lim_{M\to\infty} \frac{1}{M} \sum_{n=1}^{M} O_n. \tag{2.58}$$

Systems subject to holonomic constraints In MD simulations, it is often convenient to impose a set of constraints to the system of interest. A constraint is referred to as *holonomic* if it depends only on positions of the particles and, possibly, on time. Time-independent holonomic constraints are considered here, and can be expressed as N_c conditions of the form:

$$\sigma_k\left(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_N\right) = 0 \quad k = 1, \ldots, N_c \tag{2.59}$$

In this case, the equations of motion are written as:

$$m_i \ddot{\boldsymbol{r}}_i = \boldsymbol{F}_i + \sum_{k=1}^{N_c} \lambda_k \nabla_i \sigma_k, \tag{2.60}$$

where λ_k denotes a set of Lagrange multipliers introduced to enforce the constraints. This set of multipliers is determined on-the-fly during constrained MD simulations, typically through an iterative procedure such as the SHAKE algorithm..

2.2.4. λ-path Thermodynamic Integration Methods

Thermodynamic Integrations in Internal Coordinates The rotationally and translationally invariant internal coordinates $\boldsymbol{q} = \boldsymbol{q}(\boldsymbol{x})$ can be used to define the harmonic reference in the λ-path thermodynamic integration (λ-TI). The anharmonic free energy of a system, A_1 is expressed as:

$$A_1 = A_{0,\boldsymbol{x}} + \Delta A_{0,\boldsymbol{x}\to 0,\boldsymbol{q}} + \Delta A_{0,\boldsymbol{q}\to 1} \tag{2.61}$$

where $A_{0,\boldsymbol{x}}$ is the free energy of the harmonic reference in Cartesian coordinates $\boldsymbol{x}$, $\Delta A_{0,\boldsymbol{x}\to 0,\boldsymbol{q}}$ is the free energy difference associated with the transformation from the system harmonic in $\boldsymbol{x}$ to that harmonic in $\boldsymbol{q}$, and $\Delta A_{0,\boldsymbol{q}\to 1}$ is the contribution from the transformation of the system harmonic in $\boldsymbol{q}$ to the fully anharmonic system 1. The $A_{0,\boldsymbol{x}}$ is computed from a harmonic force field as:

$$A_{0,\boldsymbol{x}} = A_{el}(\boldsymbol{x}_0) - k_\mathrm{B}T \sum_{i=1}^{N_{vib}} \ln \frac{k_\mathrm{B}T}{\hbar\omega_i} \tag{2.62}$$

where $A_{el}(\boldsymbol{x}_0)$ is the electronic free energy for the minimum energy configuration of the reference system, N_{vib} denotes the number of vibrational modes i with real angular frequencies ω_i obtained via the finite difference method, and $\hbar$ is the reduced Planck's constant. The second term in the right-hand side of Eq. (2.61) is evaluated as:

$$\Delta A_{0,\boldsymbol{x}\to 0,\boldsymbol{q}} = \int_0^1 \left\langle \mathcal{H}_{0,\boldsymbol{q}} - \mathcal{H}_{0,\boldsymbol{x}} \right\rangle_\lambda \, d\lambda \tag{2.63}$$

where the Hamiltonian $\mathcal{H}_\lambda$ is defined as:

$$\mathcal{H}_\lambda = \lambda\mathcal{H}_{0,\boldsymbol{q}} + (1-\lambda)\mathcal{H}_{0,\boldsymbol{x}} \tag{2.64}$$

The Hamiltonian $\mathcal{H}_{0,\boldsymbol{x}}$ and $\mathcal{H}_{0,\boldsymbol{q}}$ differ only in their respective potential energies $\mathcal{V}_{0,\boldsymbol{x}}(\boldsymbol{x})$ and $\mathcal{V}_{0,1}(\boldsymbol{q})$, which are given by Eqs. (2.65) and (2.66), respectively.

$$\mathcal{V}_{0,\boldsymbol{x}}(\boldsymbol{x}) = \mathcal{V}_{0,\boldsymbol{x}}(\boldsymbol{x}_0) + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}_0)^\mathrm{T}\underline{\mathbf{H}}^{\mathbf{x}}(\boldsymbol{x}-\boldsymbol{x}_0) \tag{2.65}$$

$$\mathcal{V}_{0,\boldsymbol{q}}(\boldsymbol{q}) = \mathcal{V}_{0,\boldsymbol{q}}(\boldsymbol{q}_0) + \frac{1}{2}(\boldsymbol{q}-\boldsymbol{q}_0)^\mathrm{T}\underline{\mathbf{H}}^{\mathbf{q}}(\boldsymbol{q}-\boldsymbol{q}_0) \tag{2.66}$$

Here, $\boldsymbol{x}_0$ and $\boldsymbol{q}_0$ denote the Cartesian and internal coordinates at the potential energy minimum. The relation between the two Hessian matrices in $\boldsymbol{x}$ and $\boldsymbol{q}$ is given by:

$$\underline{\boldsymbol{H}}^{\boldsymbol{q}} = \underline{\boldsymbol{A}}^T_{\mathbf{x}_0}\underline{\boldsymbol{H}}^{\boldsymbol{x}}\underline{\boldsymbol{A}}_{\mathbf{x}_0} \quad \text{and} \quad \underline{\boldsymbol{H}}^{\boldsymbol{x}} = \underline{\boldsymbol{B}}^T_{\mathbf{x}_0}\underline{\boldsymbol{H}}^{\boldsymbol{q}}\underline{\boldsymbol{B}}_{\mathbf{x}_0}, \tag{2.67}$$

where $\underline{\boldsymbol{A}}$ is the Moore-Penrose pseudoinverse of the Wilson B-matrix:

$$\underline{\boldsymbol{B}}_{ij} = \frac{\partial q_i}{\partial x_j}. \tag{2.68}$$

Similarly, the $\Delta A_{0,\boldsymbol{q}\to 1}$ contribution is obtained as:

$$\Delta A_{0,1\to 1} = \int_0^1 \left\langle \mathcal{H}_1 - \mathcal{H}_{0,1} \right\rangle_\lambda \, d\lambda, \tag{2.69}$$

where $\mathcal{H}_1$ is the Hamiltonian of the fully interacting system (*e.g.*, a DFT Hamiltonian). In this case, the Hamiltonian $\mathcal{H}_\lambda$ is defined to be:

$$\mathcal{H}_\lambda = \lambda\mathcal{H}_1 + (1-\lambda)\mathcal{H}_{0,\boldsymbol{q}}. \tag{2.70}$$

Harmonic references with holonomic constraints When holonomic constraints are introduced into a harmonic reference system, the corresponding constrained degrees of freedom must be projected out of the Hessian matrix to eliminate their contribution to the system's free energy. To achieve this, a square projector is constructed using Eq. (2.71) and then modified to account for the constraints according to Eq. (2.72).

$$\underline{\boldsymbol{P}} = \underline{\boldsymbol{B}\boldsymbol{B}}^T \tag{2.71}$$

$$\underline{\boldsymbol{P}'} = \underline{\boldsymbol{P}} - \underline{\boldsymbol{P}\boldsymbol{C}} \left(\underline{\boldsymbol{C}\boldsymbol{P}\boldsymbol{C}}\right)^{-1} \underline{\boldsymbol{C}\boldsymbol{P}} \tag{2.72}$$

Here, $\underline{\boldsymbol{C}}$ denotes a diagonal matrix in which the entries corresponding to the constrained coordinates in $\underline{\boldsymbol{B}}$ are set to 1. while all others are 0. The constraint-inclusive modified Wilson B-matrix $\underline{\boldsymbol{B}}_c$ is then obtained as:

$$\underline{\boldsymbol{B}}_c = \underline{\boldsymbol{A}\boldsymbol{P}'} \tag{2.73}$$

The matrices $\underline{\boldsymbol{B}}_c$ and its Moore-Penrose pseudoinverse $\underline{\boldsymbol{A}}_c$ are subsequently employed in initializing the Hessian matrices in Cartesian and internal coordinates (see Eq. (2.67)). The free energy difference between the harmonic reference system and the fully interacting system with the coordinate(s) $\xi(\boldsymbol{x})$ being constrained at ξ' is computed as:

$$\Delta A_{0,\boldsymbol{x}\to 1}(\xi') = \int_0^1 d\lambda \frac{1}{\langle Z^{-1/2}\rangle_{\lambda,\xi'}} \langle Z^{-1/2}(\mathcal{V}_1 - \mathcal{V}_{0,\boldsymbol{x}})\rangle_{\lambda,\xi'} \tag{2.74}$$

where Z represents the inverse mass metric tensor, defined as:

$$Z = \sum_{i=1}^{N} \frac{1}{m_i} \sum_{\mu=x,y,z} \left(\frac{\partial \xi}{\partial x_{i,\mu}}\right)^2. \tag{2.75}$$

Part II.

Contents

3. Single-Atom Catalysts Supported on Metal-Organic Frameworks

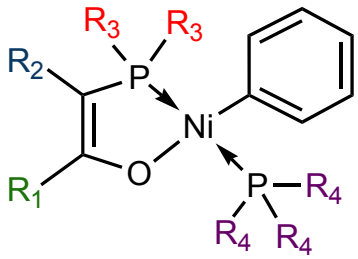
Scope. This chapter presents computational studies on the oligomerization and polymerization of ethylene catalyzed by transition metal single atom catalysts (SACs) supported on metal-organic frameworks (MOFs) based on the density function theory (DFT). Section 3.1 investigates the ethylene dimerization and oligomerization on Ni single site catalysts supported on zirconia-based MOFs. The corresponding results for the former reaction are to be published with the co-first-author Nikita Matsokin, who conducted post-Hartree-Fock calculations as part of his Ph.D. thesis, and co-authors Felix Studt, Karin Fink, and Dmitry I. Sharapa as advisors. The findings for the oligomerization reaction contribute to a joint multiscale kinetics study lead by Aleksandr Avdoshin. Section 3.2 elucidates the nature of catalytic active sites and reaction mechanisms of the ethylene polymerization on chromium-supported NU-1000 catalysts. This work is to be published with the co-first-author Nikita Matsokin, who will correct the DFT results using DLPNO-CCSD(T) and CASSCF methods, and co-authors Dmitry I. Sharapa, Karin Fink, and Felix Studt as advisors.

3.1. Ethylene Oligomerization on Ni-supported MOF Catalysts

3.1.1. Overview and Simulation Details

The immobilization of Ni-based homogeneous catalysts, such as SHOP catalysts (Scheme 3.1), ethylene oligomerization onto solid supports has gained substantial attention in recent decades. Such a heterogenization approach addresses the long-standing separation issues and associated increased operational costs, which have plagued homogeneous systems. Furthermore, well-dispersed distributions of the catalytic active sites on supports enhance atom efficiency, and proper selections of porous structure aid in achieving selectivity toward desired product spectra. Alongside traditional supports like amorphous silica[36], alumina or zeolites[192–194], metal-organic frameworks (MOFs) have increasingly been proven to be a versatile and promising class of supports (in addition to their intrinsic catalytic activity). MOFs offer well-defined crystalline structures with remarkably high surface areas and considerable structural diversity. In addition to their intrinsic properties,

their highly amendable nature towards structural modifications facilitates the introduction of catalytically active environments, often in the form of single-atom catalysts (SACs).



Scheme 3.1.: SHOP catalysts for ethylene oligomerization. Reprinted with permission from [195]. Copyright 2023 The Royal Society of Chemistry.

On the experimental side, Ni-based SACs supported in MOFs have demonstrated catalytic activities comparable to their homogeneous counterparts, along with excellent selectivity, particularly toward butenes[195]. Various strategies have been employed to introduce Ni SACs into MOF structures. One approach involves anchoring Ni complexes into the organic linkers of MOF scaffolds. Ethylene dimerization has been reported on Ni SACs anchored in (Fe)MIL-101[196], MixMOFs[197], NU-1000[198], UiO-67(bpydc)[199], UiO-66[200], MIL-125[201], PCN-701[202], achieving turnover frequencies (TOFs) exceeding 10 000 h^{-1} and butene selectivities above 90%. Alternatively, Ni atoms can be doped into the Zn-based nodes of ZIF-8[203] and MOF-5[204], yielding impressive TOFs of 1 000 000 h^{-1} and 352 000 h^{-1}, respectively. Ion-exchanged Ni SACs on Zn-based MFU-4l[205] and CFA-1[206] have also shown high activity and selectivity for ethylene dimerization. In addition, Ni can be deposited onto the surface of MOF metal nodes. Examples include Ni-NU-1000[93, 207] and Ni/UiO-66[94], which exhibit not only high activity but also remarkable catalytic stability.

Computational approaches are powerful tools for gaining a deeper understanding of catalyst structure and reaction mechanisms, thereby aiding the interpretation of experimental results and guiding catalyst design. The work of Chen et al. was supported by DFT calculations at the PBE-D3 level using a cluster model. The authors identified the square-planar coordination environment of the Ni active sites as the key factor responsible for its enhanced catalytic activity[203]. Ortega and coworkers conducted a DFT study at the PBE0 level to rationalize the difference in experimental activity between Ni-exchanged MFU-4l and CFA-1, as reported by Dincă's group[205, 206]. The lower activity of the latter catalyst was attributed to steric hindrance and the donor strength of the biphenyl linkers[208]. The work of Yeh and colleagues on ethylene dimerization over Ni/UiO-66 was also complemented by DFT results at the meta-GGA M06-L level, employing a formate cluster model. By considering the effects of an additional physisorbed ethylene molecule, they successfully reproduced the experimentally observed apparent activation energy of the reaction[94]. Song et al. performed DFT calculations at GGA-BP86 level on truncated model containing two Zr_6O_8 nodes connected by a linker bearing a α-diimine Ni species to explain the activity of diimine-Ni@PCN-701. They found that the enhanced performance of the SAC compared to its molecular counterpart originates from the reduced barriers in the Cossee-Arlman pathway, due to the electron-withdrawing effects of the Zr nodes[202].

For Ni in NU-1000, several computational studies have examined different aspects of the dimerization reaction. Bernales et al. characterized Ni and Co SACs on NU-1000 using a formate model, revealing that Ni exhibits lower activation barriers and higher activity and Co on a preferred Cossee-Arlman mechanism. Interestingly, the findings inspired subsequent experiments that confirmed the theoretical predictions[114]. Later, Ye and coworkers demonstrated that the catalytic performance of Ni in NU-1000 were similar whether the active site was a mononuclear Ni or a Ni-oxo cluster[115]. In a separate study, Ye screened a series of transition-metal SACs on NU-1000 for the ethylene dimerization and identified Ni as the most promising candidate[119]. This study also showed that steric effects from linker representations and their electronic properties in cluster models had negligible qualitative influence on the reaction mechanism[119]. However, Liu et al. demonstrated that, quantitatively, these structural and electronic factors played a crucial role in enhancing butene selectivity, thereby supporting their experimental observations[207].

The aforementioned computational studies show considerable variation in their choice of density functionals. The models employed were predominantly finite, truncated representations, in which, the influence of long-range interactions and potential disruptions to the electronic structure has not been carefully evaluated. For the Ni-NU-1000 catalytic system, such effects are expected to be even more significant due to its complex pore architecture. Additionally, although the catalyst has been examined in several studies, no clear consensus has been reached regarding the structure of Ni active site. In this chapter, we systematically investigate the ethylene oligomerization up to tetramers on the Ni^{II}-NU-1000 catalyst, considering multiple theoretical aspects, including model performance, catalyst structures, reaction mechanisms, and the influence of pore effects. The free energy data shown in Figures 3.3- 3.5 and in the relevant discussion were obtained within a hierarchical cluster correction scheme that were performed by **Nikita Matsokin** from the group of **Prof. Dr. Karin Fink** at the **Institute of Nanotechnology (INT), KIT**. The corrections was done based on my periodic PBE-D3(BJ) electronic energies and optimized geometries, and the resulting the data are included here, with permission, to ensure the most accurate energetics for comparison with experimental turnover frequencies. The insights gained from this benchmark also guided the selection of functionals employed in Section 3.2.

3.1.1.1. Simulation Details

For the periodic calculations, we employed a unit cell of NU-1000 with dimensions of 39.90 Å×39.90 Å×16.50 Å, with $\alpha = \beta = 90°$ and $\gamma = 120°$. These structural parameters are in good agreement with the experimental crystallographic values reported by Planas et al.[122]. The active site was modeled by depositing a Ni atom onto one face of the three Zr_6O_8 nodes with staggered mixed proton topology[122] (Figure 3.1a). The periodic calculations were performed at the Perdew-Burke-Ernzerhof (PBE) functional[210] with Grimme's D3 dispersion correction[211, 212] using the Vienna *Ab Initio* Simulation Package (VASP) code [188–191]. Plane-wave basis set and standard PAW potentials were employed with an energy cutoff of 520 eV. The Brillouin zone was sampled at the Γ-point. Electronic

energy was converged to 10^{-6} eV and forces were to 0.01 eV Å^{-1} for all the intermediates and transition states. Exceptions are for TS3-4, where a threshold in forces of 0.02 eV Å^{-1} was used. During geometry optimizations, all atoms were allowed to fully relax.

The cluster models were constructed by truncating the periodic NU-1000 structure around the active site. Two types of cluster models were considered: a formate model, in which the organic linkers were replaced with formate groups , and a benzoate model, where benzoate groups served as linker representations ((Figure 3.1b,c)). For the latter model, two approaches were tested: (i) assuming rigid linkers by fixing the positions of all linker carbon atoms, or (ii) allowing limited flexibility by constraining only the *para*-carbon atoms of the phenyl rings to preserve the *pseudo*-lattice of the crystalline framework. In the formate model, the carbon atoms of the formate ligands were kept fixed during geometry optimization. In both cases, all frozen atoms were held to their positions from the optimized periodic NU-1000 structure. Electronic structure calculations on the cluster models were performed with the ORCA 5.0.3 program package[213] using the PBE-D3(BJ)/def2-TZVP level of theory. Additionally, the effective core potentials were applied to Zr atoms.

To improve the accuracy of electronic energy, the energy results were corrected to the TPSSh level using a hierachical cluster approach. The TPSSh functional has been proved to yield energy results as accurate as DLPNO-CCSD(T) (see Figure 3.3). To this end, formate cluster models of the catalytic active site were extracted from their periodic structures with danging bonds saturated by hydrogen atoms, whose positions were optimized at PBE-D3(BJ)/def2-TZVP level. Single-point energy calculations were carried out for these cluster models at TPSSh-D3(BJ)/def2-QZVPP. The TPSSh-corrected energy, E^{TPSSh} is calculated *via*:

$$E^{\mathrm{TPSSh}} = E_{pbc}^{\text{PBE-D3(BJ)}} + [E_{cluster}^{\mathrm{TPSSh}} - E_{cluster}^{\text{PBE-D3(BJ)}}] \tag{3.1}$$

where $E_{pbc}^{\text{PBE-D3(BJ)}}$ is the energy for periodic model at PBE-D3(BJ) level, $E_{cluster}^{\mathrm{TPSSh}}$ and $E_{cluster}^{\text{PBE-D3(BJ)}}$ are the energies calculated for the formate model at TPSSh and PBE-D3(BJ) levels, respectively. All cluster calculations for this correction process were performed by **Nikita Matsokin** in the ORCA 5.0.4[213].

Partial Hessian matrices for the portion of the periodic structure used in the formate model were computed by finite difference method with atomic displacements of 0.01 Å at the PBE-D3(BJ) level. The vibrational data were applied to compute the thermochemical properties for the periodic model and in the correction scheme. Analytical frequency calculations were performed for the finite cluster models to achieve their thermochemical properties. All eigenvalues were inspected to confirm the minimum and first-order saddle-point nature of intermediates and transition state structures, respectively. The only exceptions were intermediate **7** and **9**, oriented toward the hexagonal pore, which each exhibited a single imaginary mode associated with a flat rotational potential energy surface.

The standard harmonic oscillator approximation was applied for solid-state structures. For each stationary point i (intermediate or transition state), the Gibbs free energy at temperature T is obtained as:

$$G_i(T) = E_i^{\mathrm{TPSSh}} - k_{\mathrm{B}} T \ln Q_{pbc,i}^{PBE-D3(BJ)}(T) \tag{3.2}$$

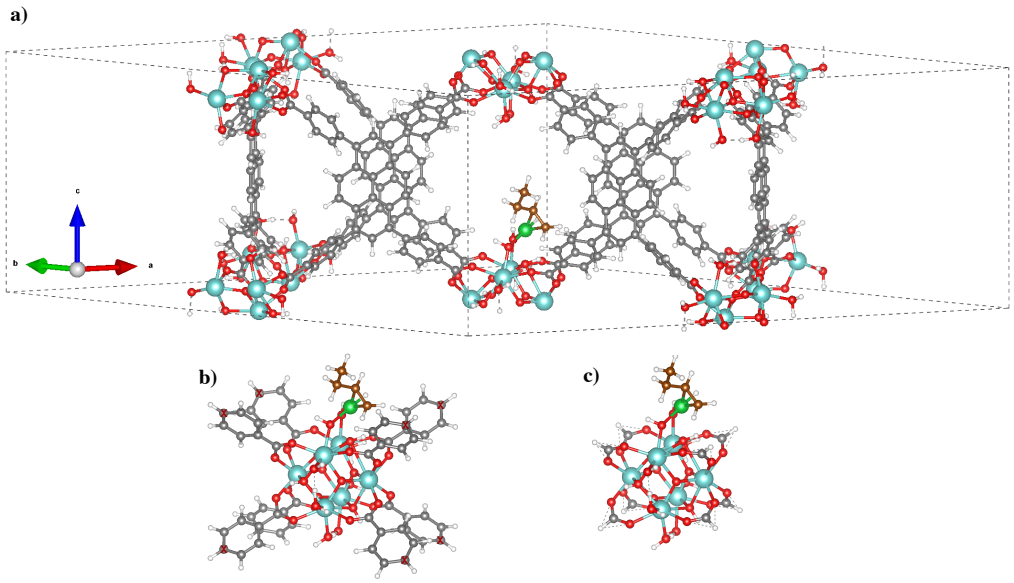


Figure 3.1.: The three models used to simulate the catalytic performance of Ni^{II}-NU-1000. a) Visualization of the MOF structure and the unit cell used for periodic model consisting of three Zr_6O_8 nodes and one Ni active site, calculated in VASP; b) the benzoate model (the red crosses denote the *para*-carbon atoms being fixed in the *pseudo*-rigid variant); c) the formate model calculated using ORCA. In the visualization, different chemical elements are represented by specific colors: Zr (cyan), Ni (green), O (red), framework C (gray); reactive C (brown), and H (white).

where E_i^{TPSSh} is the TPSSh-corrected electronic energy from Eq. (3.1), k_B is the Boltzmann constant, and $Q_{pbc,i}^{PBE-D3(BJ)}(T)$ is the partition function obtained from the partial Hessian matrix at PBE-D3(BJ) level. The partition function is given as:

$$Q(T) = \prod_{j=1}^{\mathrm{N_{vib}}} q_{\mathrm{HO,\,j}} = \prod_{j=1}^{\mathrm{N_{vib}}} \frac{e^{\frac{\hbar\omega_j}{2k_\mathrm{B}T}}}{1 - e^{\frac{\hbar\omega_j}{k_\mathrm{B}T}}} \tag{3.3}$$

where $\hbar = \frac{h}{2\pi}$ is the reduced Planck constant, ω_j is the angular frequency of vibrational mode j, $\mathrm{N_{vib}}$ is the total number of vibrational modes. The imaginary modes in the aforementioned intermediate **7** and **9** were treated as free rotations of a propyl and butyl group, respectively, along the corresponding C-C axis. The partition function in these cases is calculated as:

$$Q(T) = q_{\mathrm{1D\text{-}FR}} \times \prod_{j=1}^{\mathrm{N_{vib}}-1} q_{\mathrm{HO,\,j}} = \frac{(8\pi^3 I k_\mathrm{B} T)^{1/2}}{\sigma h} \times \prod_{j=1}^{\mathrm{N_{vib}}-1} \frac{e^{\frac{\hbar\omega_j}{2k_\mathrm{B}T}}}{1 - e^{\frac{\hbar\omega_j}{k_\mathrm{B}T}}} \tag{3.4}$$

where h is the Planck constant, σ is the rotational symmetry ($\sigma = 3$ for propyl and butyl rotations herein), I is the moment of inertia of the rotation.

For gas-phase molecules, an ideal gas, rigid-rotor harmonic oscillator approximation was used.

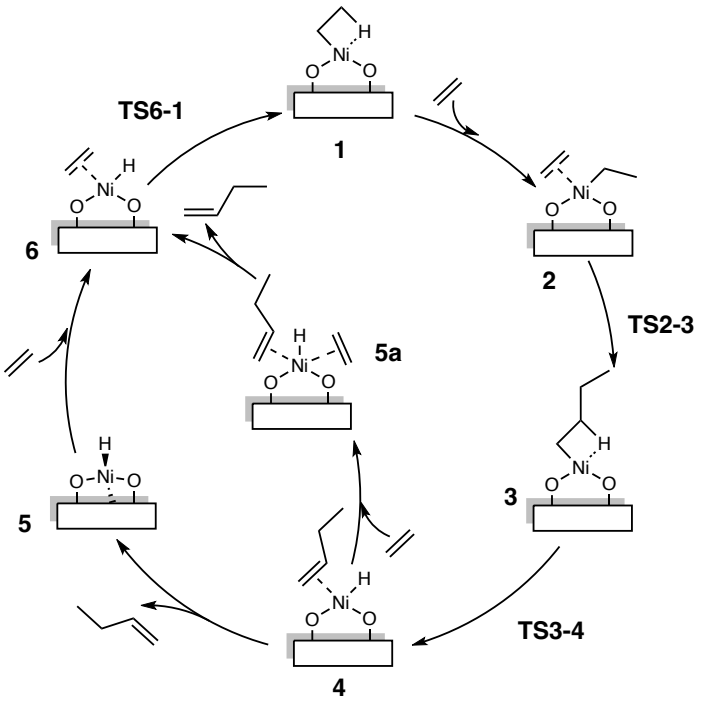
3.1.2. Results and Discussion

Ethylene dimerization mechanism We initially focused our investigation on elucidating the reaction mechanism of ethylene dimerization on the Ni^{II}-NU-1000 catalyst. A detailed Cossee-Arlman mechanism, illustrated in Scheme 3.2, was considered. This mechanism, established for a variety of Ni-based homogeneous catalysts[20], has also been suggested for Ni-based heterogeneous catalysts[205, 214], including Ni-AIM-NU-1000[93, 114, 115, 209] and Ni-UiO-66[94]. We noted that multiple structural motifs of the Ni active site on the Zr_6O_8node have been reported to be located at different steps along the oligomerization reaction. These motifs may act as distinct catalytic sites with varying activity, underscoring the importance of evaluating the reaction mechanism on each motif.

This study examined two stable configurations of the Ni^{II}-NU-1000 active site. In both motifs, the Ni center adopts a square-planar geometry, but they differ in the number of available coordination sites. The first motif, referred to as $Ni^{II}O_2$-type, features Ni bound to two terminal hydroxo ligands of the Zr_6O_8 node, leaving two vacant sites for coordinations. In contrast, the $Ni^{II}O_3$-type involves the Ni atom coordinated to an additional bridging hydroxo ligand, reducing the number of accessible sites to one if the square-planar geometry is to be preserved. A schematic depiction of these motifs is shown in Figure 3.2a, along with their relative electronic energies at PBE-D3(BJ)/def2-TZVP level. Gibbs free energy profiles (Figure 3.2b) indicate that $Ni^{II}O_2$-type sites are more catalytically active than their $Ni^{II}O_3$-type counterparts. The higher activity arises from the greater accessibility of coordination sites and the ability to maintain a square-planar environment throughout the cycle, resulting in reduced activation barriers for both ethylene insertion and β-hydride elimination steps.

Figure 3.2c illustrates key stationary structures involved in the ethylene dimerization on the $Ni^{II}O_2$-type active site. On this motif, the cycle initiates with a [Ni-ethyl] species (**1**) stabilized by a Ni-H agostic interaction. Adsorption of an ethylene molecule produces a square-planar π-complex (**2**), with an adsorption free energy of $-47\,\mathrm{kJ\,mol^{-1}}$, significantly more stable than a five-coordinate $Ni^{II}O_3$-type counterpart (**S2**), which was reported previously[114, 115, 209]. The subsequent C-C coupling proceeds through square-planar transition state **TS2-3** with an activation free energy of $84\,\mathrm{kJ\,mol^{-1}}$, yielding a [Ni-butyl] intermediate (**3**). This intermediate is stabilized by a β-agostic hydrogen bridge, which facilitates the subsequent β-hydride elimination step (**TS3-4**) with an activation free energy of only $33\,\mathrm{kJ\,mol^{-1}}$. Product release can occur either by direct desorption of 1-butene (**4** $\rightarrow$ **5**) or via substitution by an addition ethylene molecule (**4** $\rightarrow$ **5a**), which is thermodynamically favored by $23\,\mathrm{kJ\,mol^{-1}}$ under ambient conditions. However, this preference is temperature-dependent, as the direct desorption pathway dominates at elevated temperatures (>380 K). Additionally, it is expected that at higher ethylene pressures, the substitution pathway will be further favored. The catalytic cycle closes with facile regeneration of the active [Ni-ethyl] species (**6** $\rightarrow$ **1**) via ethylene hydrogenation, requiring only $6\,\mathrm{kJ\,mol^{-1}}$.

In contrast, the $Ni^{II}O_3$-type active site exhibits reduced activity due to unfavorable coordination enviroments. The coordination of ethylene to the [Ni-ethyl] species (**S1**) is endogonic, generating a trigonal bipyramidal complex that is $38\,\mathrm{kJ\,mol^{-1}}$ less stable than



Scheme 3.2.: The schematic illustration of the Cossee-Arlman mechanism for ethylene dimerization on Ni^{II}-NU-1000. The presentations of NU-1000's Zr_6O_8 node and organic linkers are omitted for clarity.

the square-planar **2**. Similar instability is observed in later intermediates, such as [H-Ni-butyl]. The C-C coupling and β-hydride elimination steps proceed through transition states with substantially higher activation free energies of 101 kJ mol^{-1} and 78 kJ mol^{-1}, respectively. These increased barriers reflect the tendency of the trigonal bipyramidal structures to distort toward square-planar geometries. The product desorption is comparatively favorable. We note that a stable [H-Ni-ethylene] species of this motif could not be located, and therefore, the regeneration of the active site was assumed to occur analogously to the $Ni^{II}O_2$-type cycle.

According to the energetic span model[215], the intermediate **2** and transition state **TS2-3** act as the turnover-determining intermediate (TDI) and turnover-determining transition state (TDTS), respectively, in the $Ni^{II}O_2$-type cycle, while **S1** and **TS-S2-S3** serve as the respective TDI and TDTS for the $Ni^{II}O_3$-type cycle. Thus, in both cases, the ethylene insertion step is identified as the TOF-determining step of the reaction.

Models performance Most of theoretical studies on SACs supported on MOFs have relied on cluster models to represent the catalytic relevant sites, as this approach allows higher-level methods to be applied while keeping the computational cost manageable. Nevertheless, how well such models perform in capturing the long-range effects of the periodic framework remains to be assessed, since such relevant benchmarking is still scarce in the literature. In this work, we benchmark the commonly used cluster models of the Ni^{II}-NU-1000 system against a fully periodic reference.

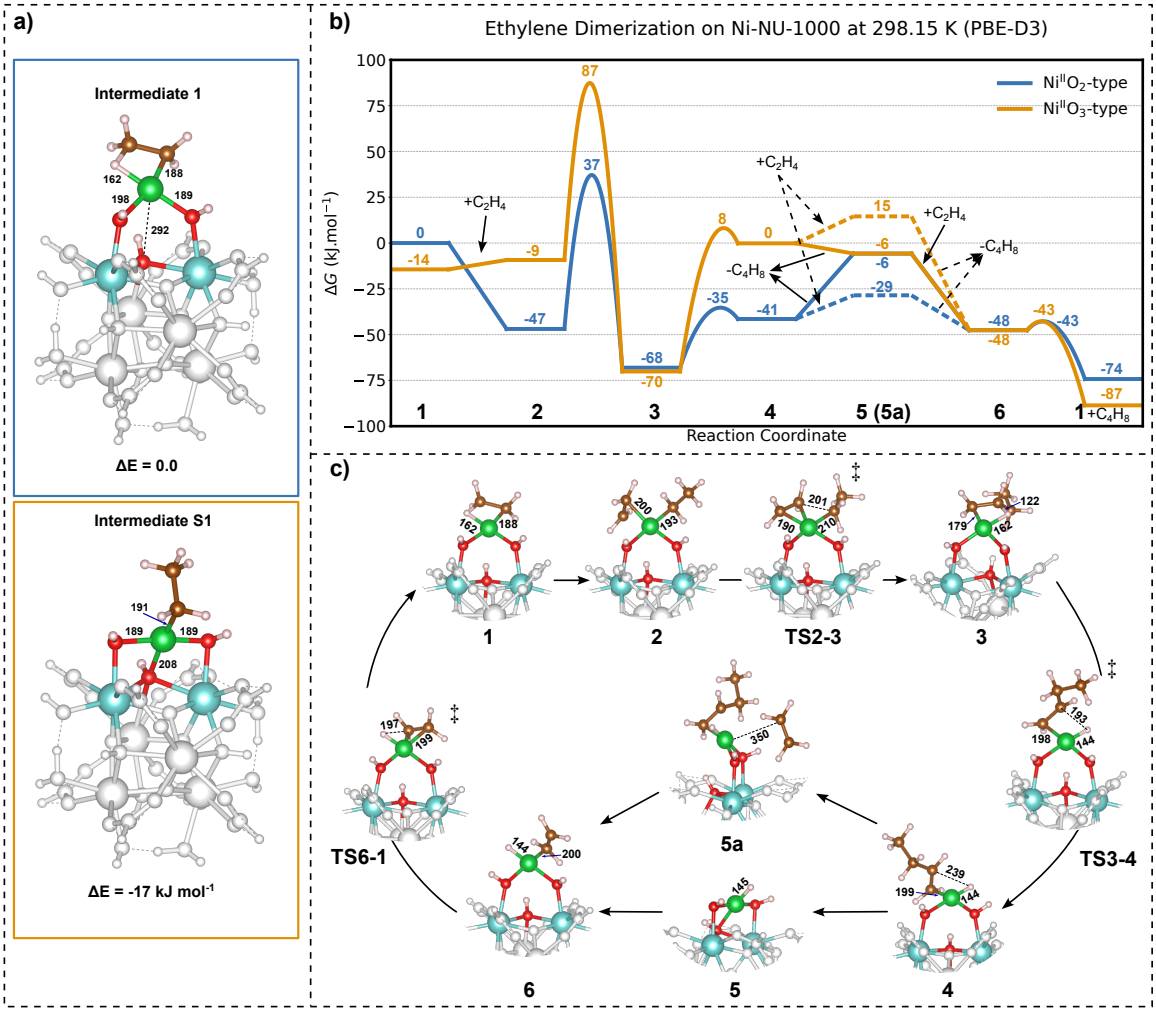


Figure 3.2.: a) The two configurations of the [Ni-ethyl] structure: intermediate **1** (left) at the $Ni^{II}O_2$-type active site and intermediate **S1** (right) at the $Ni^{II}O_3$-type active site, together with their relative electronic energies. b) Gibbs free energy profiles of ethylene dimerization on the two active sites at 298.15 K and 1 bar. c) Key stationary structures involved in the ethylene dimerization on the $Ni^{II}O_2$-type active site calculated at PBE-D3(BJ)/def2-TZVP level using formate cluster model. Representative chemical elements are represented by specific colors: Zr (cyan), Ni (green), O (red), C (brown), H (light pink). The bond length and interatomic distances are given in pm.

For this purpose, we adopted a periodic reference model consisting of three metal nodes, with the Ni active site located on one of them (Figure 3.1a). This model retains the full steric environment. Two cluster variants were considered, namely the benzoate model (Figure 3.1b) and the formate model (Figure 3.1c). To preserve the framework rigidity, the carbon atoms of these ligands were fixed to their positions in optimized periodic NU-1000 structure. The benzoate model partially accounts for the steric influences from bulky linkers, while the formate model neglects them completely. A third *pseudo*-rigid benzoate model was also considered, in which only the *para*-carbon atoms of the phenyl rings were fixed. This approach allows rotational flexibility of the linkers while maintaining the *pseudo*-lattice of the framework[61], thus mimicking linker motion in the extended system. The cluster model calculations were carried out at the PBE-D3(BJ)/def2-TZVP,

while similar method with an energy cutoff of 520 eV was used for the periodic calculations (as discussed in Section 3.1.1.1).

To assess the performance of these models, we compared the activation energies of the following five steps/states that are critical to the ethylene dimerization on Ni^{II}-NU-1000:

1. C-C coupling

$$\text{[Et-Ni-ethyl]} \longrightarrow \text{[Ni-butyl]}, \quad \Delta\Delta E_1^{\ddagger} = \Delta E(\textbf{TS2-3}) - \Delta E(\textbf{2}). \tag{3.5}$$

2. β-H elimination

$$\text{[Ni-butyl]} \longrightarrow \text{[H-Ni-But]}, \quad \Delta\Delta E_2^{\ddagger} = \Delta E(\textbf{TS3-4}) - \Delta E(\textbf{3}). \tag{3.6}$$

3. Direct but-1-ene desorption

$$\text{[Ni-butyl]} \longrightarrow \text{[Ni-H]} + C_4H_{8\textbf{(g)}}, \quad \Delta\Delta E_3^{\ddagger} = \Delta E(\textbf{5}) + \Delta E(C_4H_8) - \Delta E(\textbf{3}). \tag{3.7}$$

4. Substitutive but-1-ene desorption

$$\text{[Ni-butyl]}{+}C_2H_{4\textbf{(g)}} \longrightarrow \text{[H-Ni-But]}{+}C_2H_{4\textbf{(ph)}}, \quad \Delta\Delta E_4^{\ddagger} = \Delta E(\textbf{5a}){-}\Delta E(\textbf{3}){-}\Delta E(C_2H_4). \tag{3.8}$$

5. Ethylene hydrogenation

$$\text{[H-Ni-Et]} \longrightarrow \text{[Ni-ethyl]}, \quad \Delta\Delta E_5^{\ddagger} = \Delta E(\textbf{TS6-1}) - \Delta E(\textbf{6}). \tag{3.9}$$

where (**g**) and (**ph**) denote gas-phase and physisorbed molecules, respectively. The benchmarks were carried out on the $Ni^{II}O_2$-type active site. Table 3.1 summarizes the activation energies predicted at the PBE-D3(BJ) level using the different models. Overall, all truncated models reproduce the periodic mechanism with mean absolute deviations (MADs) within 4 kJ mol^{-1}, which reflects the localized nature of the electronic structure in Ni^{II}-NU-1000. The largest discrepancies are found for the β-hydride abstraction and direct desorption in the formate model, which underestimates both barriers by 10 kJ mol^{-1} and 8 kJ mol^{-1}, respectively. This underestimation originates from the model's failure to capture dispersion interactions between the Ni-butyl fragment and the ligand environment in intermediate **3**. Consistently, adsorbed intermediates predicted by the formate model are generally less stable than those of the periodic model, especially for sterically demanding structures. Including benzene rings in the benzoate models improves steric representation and barrier prediction. The rigid benzoate model predicts barriers within 5 kJ mol^{-1} of the periodic results, with a MAD as low as 2.5 kJ mol^{-1}. Introducing flexibility in the *pseudo*-rigid benzoate model does not significantly improve the accuracy. Nevertheless, both benzoate models outperform the formate model due to their enhanced ability to capture dispersion interactions.

Table 3.1.: Electronic energy barriers in kJ mol^{-1} at PBE-D3(BJ) theory level for the most important mechanistic states calculated from Eqs. (3.5)-(3.9) using different simulation models.

Reaction	pbc	*pseudo*-rigid benzoate	rigid benzoate	formate
C-C coupling	82	79	78	83
β-H elimination	40	44	45	30
ethylene hydrogenation	8	12	9	4
direct desorption	129	123	128	121
substitutive desorption	-5	-2	-3	-1

Corrected free energy profile To improve the accuracy of energy predictions, Sauer and co-workers proposed a hierarchical cluster approach to correct periodic DFT results with higher-level methods, typically MP2 or CCSD(T) variants[216–220]. This approach has been commonly applied in computational zeolite catalysis. Following this concept, we employed an analogous correction scheme for our periodic PBE-D3(BJ) calculations.

Because MOF structures are generally larger and more complex than zeolites, performing full DLPNO-CCSD(T) corrections is computationally prohibitive; instead, it is more practical to employ density functionals (DFs) at higher rungs. Within the hierarchical cluster approach, various DFs have been benchmarked against DLPNO-CCSD(T) reference to identify the optimal balance between accuracy and efficiency. As shown in Figure 3.3, among the tested functionals, TPSSh-D3(BJ) provided the closest agreement with DLPNO-CCSD(T), followed by M06-L-D3(0) and B3LYP-D3(BJ). Accordingly, the free energy values reported hereafter in this section were corrected to the TPSSh-D3(BJ) level within the hierarchical cluster scheme.

Figure 3.4 presents the Gibbs free energy profile at 298.15 K for ethylene dimerization on Ni^{II}-NU-1000 obtained at TPSSh-D3(BJ) level within the hierarchical cluster approach. The ethylene insertion step is the rate-determining step, with an activation barrier of 77 kJ mol^{-1}. The β-hydride elimination barrier is moderate at 41 kJ mol^{-1}, whereas the hydrogenation barrier is easily overcome. The product desorption is expected to proceed mainly through substitution by an physisorbed ethylene molecule, while direct desorption is about 15 kJ mol^{-1} less favorable.

Experimental kinetics measurements reported a turnover frequency (TOF) of 252 h^{-1} for ethylene dimerization on Ni-NU-1000 at 318 K and 2 bar[93]. Using the energetic span model[215] and the Gibbs free energy profile in Figure 3.4, we estimated a TOF of 612 h^{-1} at 298.15 K and 1 bar. This prediction is of the same order of magnitude as the experimental value and improves upon earlier theoretical estimates by three orders of magnitude[115].

Further oligomerization and framework effects With a reliable computational protocol in place, we extended our investigation to ethylene oligomerization up to tetramers on the Ni^{II}-NU-1000 catalyst. Here, we focused on how the framework environment shapes the reaction pathway and associated energetics. In particular, we compared the energy profiles for chain growth directed toward the hexagonal versus the trigonal pore of NU-1000.

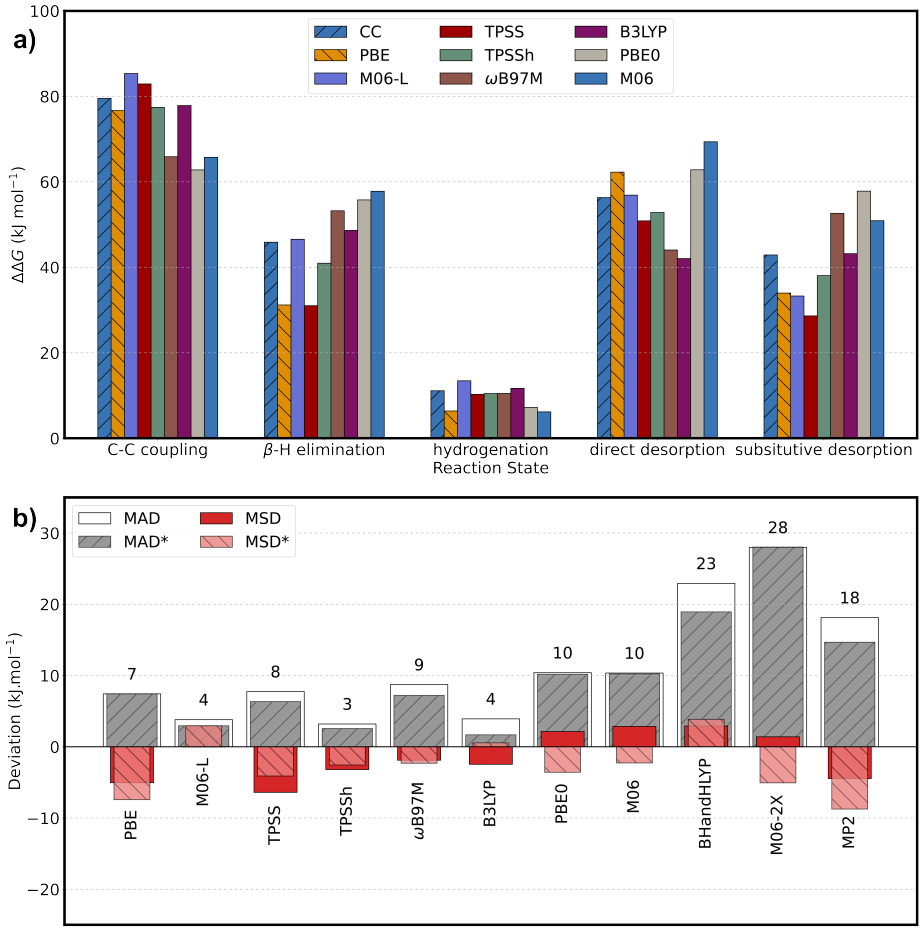


Figure 3.3.: a) Comparison of activation free energies in kJ mol^{-1} of important mechanistic states estimated by several selected DFs within the hierarchical cluster approach. b) Accuracy of MP2 and DFs in terms of mean absolute deviation (MAD) and mean signed deviation (MSD) with respect to the coupled cluster reference. The MAD* and MSD* consider only the TS-involved barriers, whereas MAD and MSD represent the deviations for the five barriers given in Eqs. (3.5)- (3.9). The number above each bar denotes the corresponding MAD (in kJ mol^{-1}). All DFT calculations were performed with the D3 dispersion correction. The benchmark was performed by Matsokin based on our results at the periodic PBE-D3(BJ) level.

Figure 3.5 compares the activation free energies for di-, tri-, and tetramerization of ethylene on Ni^{II}-NU-1000, with the growing chain oriented toward either the hexagonal (32 Å) or trigonal (12 Å) pore of the framework. In both orientations, the ethylene insertion remains the rate-determining step across all oligomerization reactions. However, when the alkyl chain extends toward the hexagonal pore, the insertion barrier increases with chain length, whereas it remains essentially constant for growth into the trigonal pore. By contrast, carbon chain length has little effect on direct olefin desorption in the hexagonal pore but significantly raises the direct desorption barrier in the trigonal pore, making product release more difficult in this more confined pore. These effects can be attributed to stabilizing dispersion interactions between the growing chain and the organic linkers in the smaller pore. Meanwhile, the β-hydride abstraction and substitutive desorption steps are largely unaffected by either chain length or orientation. As a result, longer chains

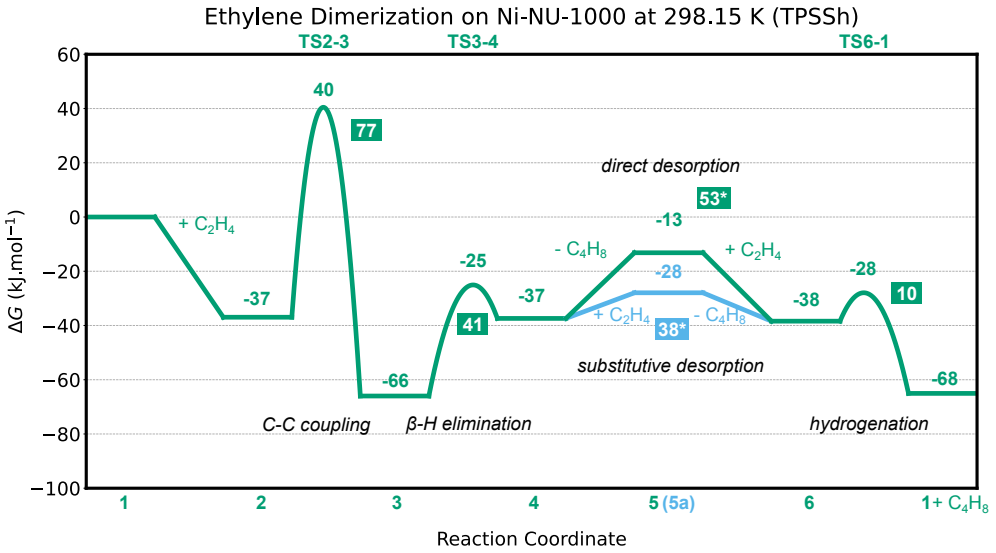


Figure 3.4.: The Gibbs free energy profile at 298.15 K for ethylene dimerization reaction on Ni^{II}-NU-1000 obtained at TPSSh-D3(BJ) level within the hierarchical cluster approach. This profile is the most accurate result predicted by examined DFs with respect to the coupled cluster reference. The framed values are the activation free energy for each reaction state, in which the asterisk denotes an effective desorption barrier, defined in eqs (3.7) and (3.8).

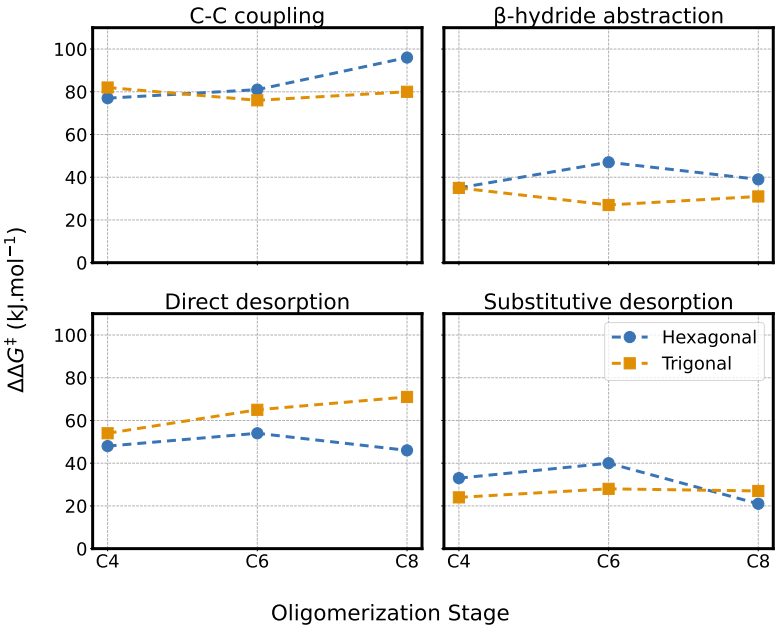


Figure 3.5.: Comparison in activation free energies (in kJ mol^{-1}) at 298.15 K and 1 bar for ethylene oligomerization on Ni^{II}-NU-1000 with the growing carbon chain oriented toward the hexagonal and trigonal pore. The values are obtained at TPSSh-D3(BJ) level within the hierarchical cluster approach. The desorption barriers are defined in eqs (3.7) and (3.8) with E replaced by G.

preferentially grow into NU-1000's trigonal pores. At low pressures, the oligomer products trapped in these pores may re-adsorb at the active site and undergo further chain growth, leading to the formation of polymeric products, as observed experimentally[93]. The preference for the trigonal pores was also predicted theoretically by Avdoshin et al.[221], who further showed that light olefins, such as ethylene, displace the heavier ones in these pores at high pressures[221]. Based on this, we speculate that operating under higher ethylene pressures could facilitate product desorption and mitigate catalyst deactivation.

3.2. Ethylene Polymerization on Cr-supported MOF Catalysts

3.2.1. Overview and Simulation Details

Chromium-based catalysts have long served as industrial benchmarks in olefin polymerization, most notably exemplified by the Phillips catalyst (Cr/SiO_2), which accounts for nearly half of global polyethylene (PE) production[38, 39]. Despite its industrial success, the Phillips system remains scientifically elusive. The precise nature of its active site(s) and the mechanistic details of its activity and selectivity continue to be debated after more than six decades of research[222]. These challenges primarily arise from the amorphous nature of the silica support, the heterogeneity of metal centers, and the difficulty of characterizing reactive intermediates under operando conditions. To address these limitations, metal-organic frameworks (MOFs) have emerged as ideal platforms for studying structure-function relationships in heterogeneous catalysis at molecular level, especially as they allow for the controlled incorporation of isolated transition-metal active sites.

Building on this advantage, Goetjen and coworkers have synthesized a MOF-supported chromium catalyst, namely Cr-SIM-NU-1000 with SIM denoting **S**olvothermal deposition **i**n **M**OFs (hereafter referred to as Cr-NU-1000 for simplicity)[100–102]. They demonstrated that Cr-NU-1000, when activated by diethylaluminum chloride (DEAC), exhibited remarkable activity for ethylene polymerization. The resulting PE was shown to have relatively uniform molecular weights, indicative of single-atom catalysis. Using a combination of crystallographic and spectroscopic techniques, the authors identified a [Cr-ethyl] species as the active site, with chromium in the oxidation state III, and established that the polymerization proceeds *via* a linear-chain insertion (Cossee-Arlman) mechanism.

However, despite these experimental advances, the detailed geometric and electronic struture of the active site and the precise reaction pathway remain incompletely understood. In particular, questions persist regarding the role of chromium oxidation state and the energetics of competing mechanisms, including the metallacyle pathway proposed in earlier studies of chromium-based catalysts. In this context, this chapter presents a DFT investigation aimed at elucidating the mechanistic aspects of ethylene polymerization on Cr-NU-1000. Two competing pathways, namely the Cossee-Arlman insertion and the metallacyle ring-expansion mechanisms, are thoroughly examined for Cr in oxidation states I, II, III, and IV.

3.2.1.1. Simulation Details

A formate model, similar to that used for Ni-NU-1000 and described in Section 3.1.1.1, was employed in this study. The geometry optimizations were performed following the same procedure described in Section 3.1.1.1, using ORCA 5.0.3[213], for all possible spin states of chromium. For catalytic reactions proceeding *via* metallacycle mechanisms which involve intermediates with different electronic ground states, the minimum energy crossing points (MECPs) along these pathways were identified using the approach of Harvey and coworkers[223]. Single-point energy calculations were carried out with the three best-performing density methods identified for Ni^{II}-NU-1000, namely TPSSh-D3(BJ), M06-L-D3(0), and B3LYP-D3(BJ), in combination with the def2-TZVPP basis set to improve the accuracy of electronic energies. The vibrational information at PBE-D3(BJ)/def2-TZVP level was used to estimate Gibbs free energies. The standard harmonic oscillator was applied for stationary structures, while the ideal-gas, rigid-rotor harmonic oscillator model was used for gas-phase molecules.

3.2.2. Results and Discussion

The Cr loading in the experimental Cr-SIM-NU-1000 was reported to range between 1.4 Cr to 2.5 Cr atoms per Zr_6O_8 node[100–102]. In this study, we employed a simplified cluster model containing a single Cr atom, coordinated to the metal node through aquo, hydroxo, and μ_3-O ligands. The exact coordination varies with the oxidation state of chromium (see Figure 3.6 for representative structures). As demonstrated in Section 3.1.2 for the analogous Ni-NU-1000 system, the steric effects from the NU-1000's organic linkers play negligible impacts on the energetics of the Cossee-Arlman pathway. Hence, the formate model is expected to provide a sufficiently accurate representation of the present system. Our conclusions are primarily based on free energy diagrams calculated at TPSSh-D3(BJ)/def2-TZVPP//PBE-D3(BJ)/def2-TZVP theory level, covering ethylene dimerization and extending up to the formation of hexyl species to compare chain termination and propagation. TPSSh was selected based on the benchmarking results in Section 3.1. Additional corrections were performed using M06-L-D3(0) and B3LYP-D3(BJ), next best performing functionals relative to DLPNO-CCSD(T), and their results will be discussed in the last part of this section.

The linear insertion Cossee-Arlman and the ring expansion metallacycle mechanisms are often proposed for ethylene oligomerization and polymerization on homo- and heterogeneous chromium catalysts. Therefore, we compared the thermodynamic results of the two mechanistic pathways for each oxidation state, and between different oxidation states of chromium and validated them by experimental observations[101, 102].

Cossee-Arlman mechanism The Cossee-Arlman mechanism on Cr-NU-1000 is illustrated schematically in Figure 3.7a and is analogous to that for Ni-NU-1000. The Cossee-Arlman mechanism on Cr^{II} catalysts is shown to favor oligomerization, likely yielding products

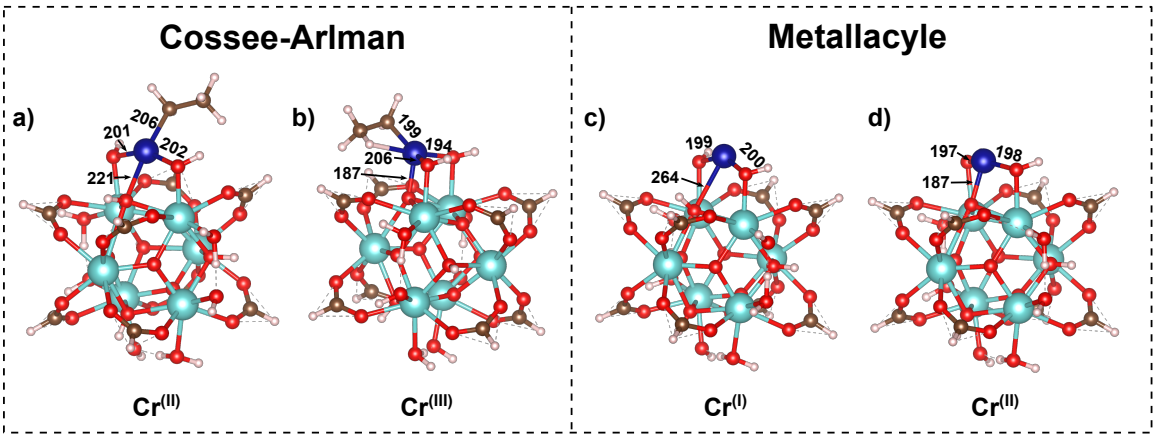


Figure 3.6.: Schematic representations of starting structures in various polymerization pathways proposed for Cr-NU-1000 catalyst. Representative chemical elements are represented by specific colors: Zr (cyan), Cr (dark blue), O (red), C (brown), H (light pink). The bond length and interatomic distances are given in pm.

that follow a Flory-Schulz distribution. This conclusion is drawn from the corresponding free energy diagrams (Figure 3.7b, c), in which the quintet state represents the ground-state spin multiplicity for all intermediates and transition states (TSs). Similar to the Ni system discussed in Section 3.1, two motifs of Cr^{II} sites were found, with structures containing $Cr^{II}O_3$-type motif lying along the more stable pathway. In this pathway, the first C-C coupling step requires an effective activation free energy of 69 kJ mol^{-1}, whereas the chain-termination steps require 89 kJ mol^{-1} to 100 kJ mol^{-1} at a pressure of 1 bar. However, it was found that the $Cr^{II}O_3$-type [Cr-butyl] structure can readily transform into a square planar $Cr^{II}O_2$-type structure featuring a β-agostic hydrogen interaction, from which the elimination is energetically more affordable. The effective barrier for β-hydrogen elimination is 75 kJ mol^{-1}, comparable to a 70 kJ mol^{-1} barrier for the subsequent C-C coupling step. Thus, these elementary reactions are kinetically competitive and the resulting products are expected to be an oligomer mixture. Increasing the ethylene pressure to 40 bar lowers the chain-growth barriers to their intrinsic values (≈60 kJ mol^{-1}), which are approximately 10 kJ mol^{-1} lower than that of chain termination. Therefore, high pressures are predicted to shift the product distribution toward heavier oligomers, but are unlikely to yield PE as a single dominant product. Product desorption does not play a significantly role within the pressure range examined.

On the other hand, the computed Gibbs free energy diagrams for Cr^{III}-NU-1000 (Figure 3.7d, e) clearly indicate its strong tendency to catalyze ethylene polymerization. The most stable structures involved in these diagrams were found in quartet spin states. The activation free energy for carbon insertion (chain-growth) are about 50 kJ mol^{-1}, compared to approximately ≈ 80 kJ mol^{-1} for chain termination. Such a difference of more than 30 kJ mol^{-1} demonstrates a clear kinetic preference for continual chain growth toward polymerization. This trend remains consistent at both 1 and 40 bar. Furthermore, as shown in the Figure 3.6b, the [Cr-ethyl] species exhibit a tetrahedral coordination environment, in excellent agreement with the crystallographically resolved structure[101, 102]. We note that a trigonal bipyramidal geometry was identified for Cr center with additional

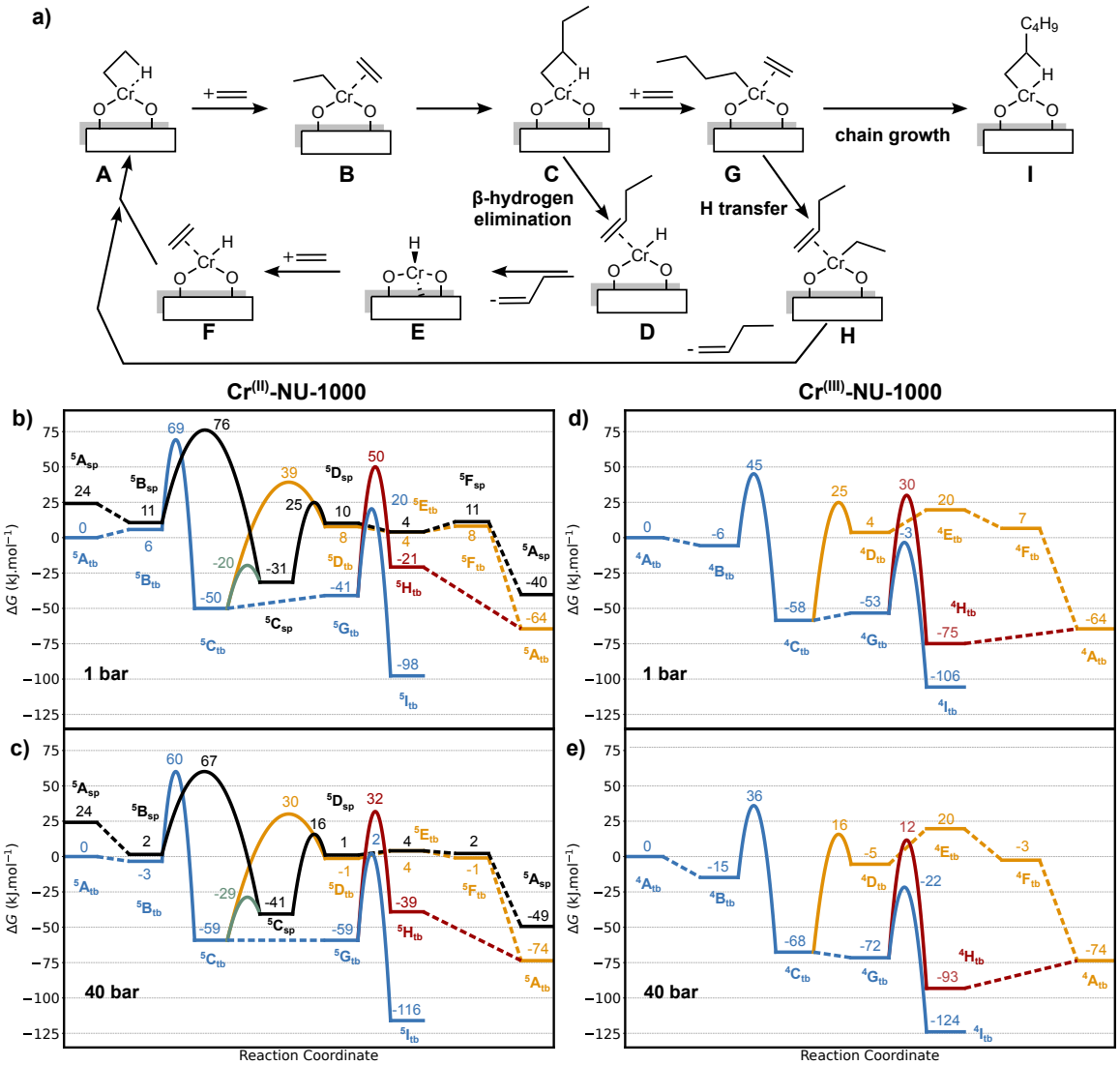


Figure 3.7.: Cossee-Arlman mechanism on Cr-NU-1000. a) The schematic illustration of the Cossee-Arlman mechanism for ethylene polymerization on Cr-NU-1000. The presentation of NU-1000's Zr_6O_8 node and organic linkers are omitted for clarity. b-e) Calculated Gibbs free energy diagrams for the ethylene polymerization at 298.15 K in $Cr^{(II)}$-NU-1000 (b, c) and $Cr^{(III)}$-NU-1000 (d, e) at pressure of 1 bar (b, d) and 40 bar (c, e). The blue line represents the chain growth pathway, orange and red lines represent the chain termination *via* β-hydrogen elimination and hydrogen transfer, respectively. The black lines in panels b and c show dimerization pathway for the square planar Cr^{II} active sites. The intermediates are denoted as ${}^{M}X_s$, where M is the spin multiplicity, X is the respective structure in the mechanism, and s indicates the geometry of the active site: tb = trigonal bipyramidal and sq = square planar.

coordination. Based on the findings regarding the framework effects of NU-1000 for chain growth in Ni catalyst (Section 3.1.2), we speculate that the polymeric carbon chains preferentially extend toward the triangular pores. This orientation likely hinders polymer removal from the framework, consistent with experimental observations[101].

Metallacyle mechanism The metallacyle mechanism is given in Figure 3.8a. The catalytic cycle begins with the coordination of two ethylene molecules to the unsaturated chromium species **M-A**, forming a [Cr-(ethylene)$_2$] complex **M-C**. This is followed by an oxidative C-C coupling step, which produces the **M-D** intermediate with a chromacyclopentane ring. From this point, the metallacyclic ring may either expand through additional ethylene coordination and insertion, for example, to form chromacycloheptane species **M-I** or undergo ring opening to yield α-olefin products. The generation of α-olefin can take place via a stepwise β-hydrogen transfer to the Cr center followed by reductive elimination or through a concerted intramolecular β-hydrogen transfer to the α'-carbon of the metallacycle. We found that the concerted pathways is energetically favored even for 1-butene formation in both Cr catalysts examined, and therefore only this pathway is depicted in the Gibbs free energy diagrams (Figure 3.8b, c). Throughout the entire catalytic cycle, the oxidation state of the chromium center alternates between $Cr^{(n)}$ and $Cr^{(n+2)}$.

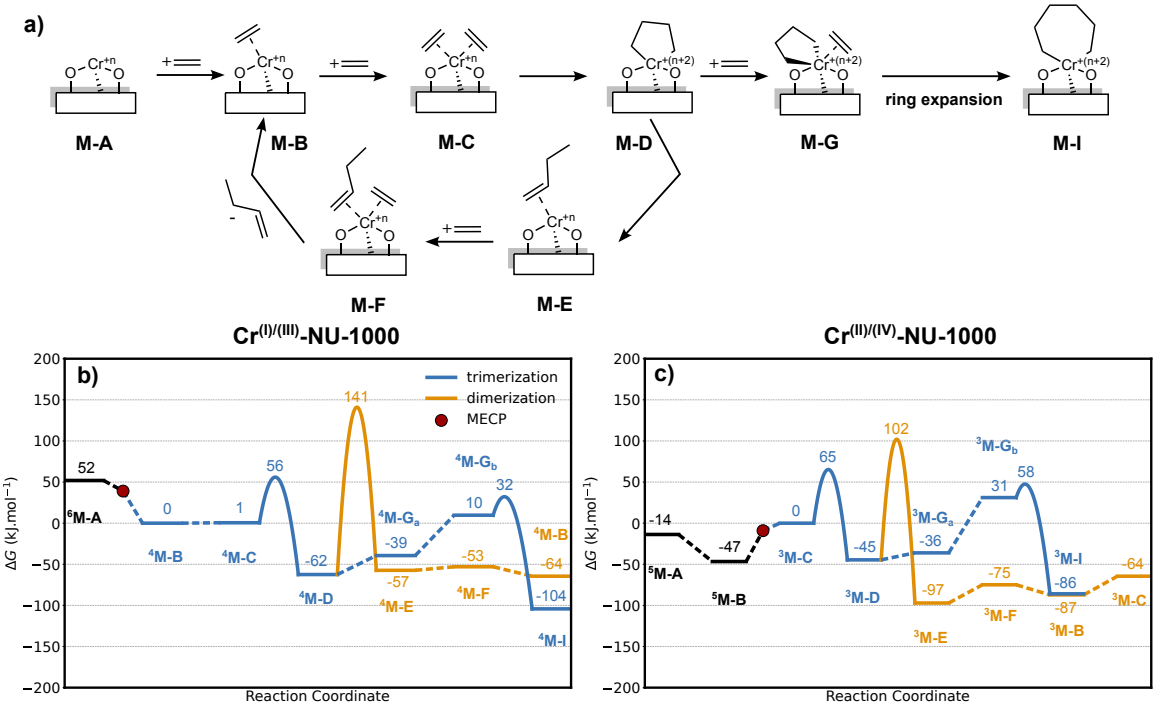


Figure 3.8.: Metallacycle mechanism on Cr-NU-1000. a) The schematic illustration of the metallacycle mechanism for ethylene polymerization on Cr-NU-1000. The presentation of NU-1000's Zr_6O_8 node and organic linkers are omitted for clarity. b, c) Calculated Gibbs free energy diagrams for the ethylene polymerization at 298.15 K in $Cr^{(I)/(III)}$-NU-1000 (b) and $Cr^{(II)/(IV)}$-NU-1000 (c) at the reference pressure of 1 bar. The zero-energy reference was chosen to be the first structure after the minimum-energy crossing points (MECPs). The intermediates are denoted as ^{M}X, where M is the spin multiplicity and X is the respective structure in the mechanism.

The undercoordinated **M-A** intermediate in the $Cr^{(I/III)}$ cycle (Figure 3.8b) features a sextet ground state, whereas the remainder of the pathway lies on the quartet potential energy surface. The activation free energy for the oxidative addition step is comparable to that of the ethylene insertion step in the Cossee-Arlman pathway of Cr^{III}-NU-1000 catalyst (56 vs. 51 kJ mol^{-1}), suggesting that this step can readily proceed. The ring expansion through β-hydrogen transfer, however, requires a free energy barrier of 204 kJ mol^{-1},

which is prohibitively high. The alternative ring-expansion pathway demands a barrier of 95 kJ mol^{-1}, nearly twice that of ring formation and too large to be surmounted at the reaction temperature. The increased barrier is attributed to the energetic penalty associated with rotation and bending of the chromacycle in the transition state $\mathbf{TS_{M\text{-}G\rightarrow M\text{-}I}}$ (Figure 3.9d), relative to the more stable configuration in **M-D** (Figure 3.9c). As the ring size increases, this penalty is expected to rise further. To verify this trend, the subsequent ring-expansion and ring-opening steps were examined. The formation of a nine-membered ring was found to require a barrier of 120 kJ mol^{-1}, while the ring-opening step exhibited a comparable barrier to 126 kJ mol^{-1}. The remarkable decrease in β-hydrogen transfer barrier can be explained by the reduced structural distortion in the transition state of the seven-membered ring (Figure 3.9b) compared to that of the five-membered analogue (Figure 3.9a). This tendency to favor ring opening while disfavoring further ring expansion with increasing ring size is consistent with previous studies, which propose that Cr$^{(I/III)}$ catalysts promote selective ethylene oligomerization via a metallacycle mechanism[86].

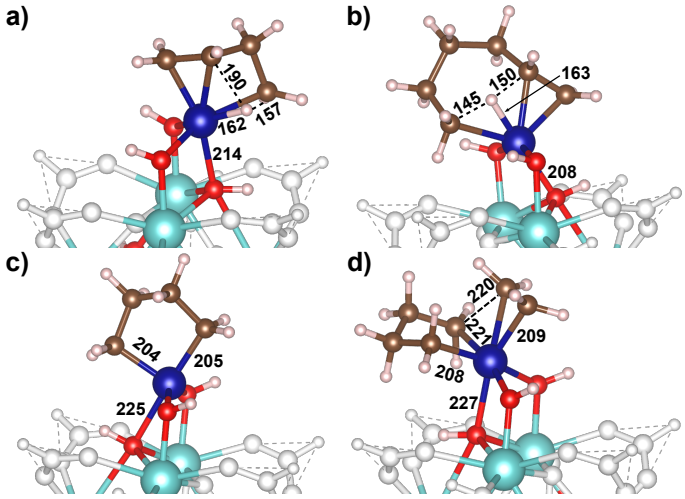


Figure 3.9.: Schematic illustrations of some key intermediates and transition states involved in the metallacyle pathway for Cr$^{(I)/(III)}$-NU-1000: a) concerted five-membered ring opening TS ($\mathbf{TS_{M\text{-}D\rightarrow M\text{-}E}}$), b) concerted seven-membered ring opening TS ($\mathbf{TS_{M\text{-}I\rightarrow M\text{-}J}}$), c) five-membered ring intermediate (**M-D**), d) seven-membered ring formation TS ($\mathbf{TS_{M\text{-}G\rightarrow M\text{-}I}}$).

The Cr$^{(II/IV)}$ cycle follows a sequence of steps similar to that of the Cr$^{(I/III)}$ pathway. In this case, the ground-state spin multiplicities of the first two intermediates are quintet, switching to a triplet state for the subsequent steps of the reaction. Unlike the Cr$^{(I/III)}$ cycle, the oxidative addition step exhibits a significantly higher effective activation barrier of 112 kJ mol^{-1}, whereas the ring-expansion barrier remains comparable at 92 kJ mol^{-1}. The β-hydrogen transfer step requires 147 kJ mol^{-1}, which, although lower than the corresponding barrier in the Cr$^{(I/III)}$ cycle, is still too high to be kinetically accessible under reaction temperature.

Given these substantial free energy barriers, both catalytic cycles are expected to exhibit negligible activity toward ethylene oligomerization or polymerization via the metallacycle mechanism. It should be noted that steric hinderance imposed by the organic linkers,

which was not accounted in the employed formate model, could further increase these barriers.

In summary, our theoretical findings indicate that the ethylene polymerization on the Cr-NU-1000 catalyst proceeds via a Cossee-Arlman mechanism at tetrahedral Cr^{III} active sites. This conclusion strongly supports the experimental observations reported for this system[100–102].

Influence of density functionals The above conclusion has been drawn using the TPSSh functional, selected based on the benchmarking results for Ni-NU-1000 catalyst discussed in Section 3.1. Here, we further examine whether this conclusion depends on the choice of density functional. Table 3.2 summarizes the key effective activation barriers, as computed using PBE, M06-L, TPSSh, and B3LYP functionals (in combination with D3 dispersion correction).

Table 3.2.: Calculated reaction barriers in kJ mol^{-1} at 298.15 K and 1 bar for various steps in the ethylene dimerization pathway on Cr-based sites using different density functionals. D3(BJ) correction was applied for PBE, TPSSh, and B3LYP calculations, while D3(0) was applied for M06-L calculations.

Reaction step	PBE	M06-L	TPSSh	B3LYP
Cr^{II} – Cossee–Arlman				
C–C coupling (dimer)	65	85	76	91
β-H elimination	75	75	75	78
H transfer	69	107	100	125
C–C coupling (trimer)	56	77	70	87
Cr^{III} – Cossee–Arlman				
C–C coupling (dimer)	40	47	51	75
β-H elimination	89	80	83	95
H transfer	70	89	88	122
C–C coupling (trimer)	42	51	55	84
Cr^{I}/Cr^{III} – Metallacycle				
Oxidative C–C coupling	47	56	56	82
H transfer (C_4)	177	185	204	223
Ring expansion (C_6)	80	75	95	122
H transfer (C_6)	104	126	126	153
Ring expansion (C_8)	107	114	120	141
Cr^{II}/Cr^{IV} – Metallacycle				
Oxidative C–C coupling	68	130	112	151
H transfer	138	150	147	158
Ring expansion	84	88	92	113

Among these methods, GGA-PBE functional greatly underestimates nearly all reaction barriers, while B3LYP systematically predicts higher values, resulting in discrepancies of up to ≈80 kJ mol^{-1}. The M06-L and TPSSh functionals show the best agreement, particularly for Cossee-Arlman pathways. Despite quantitative variations, all four functionals yield

qualitatively consistent trends regarding the catalytic activity across different oxidation states and mechanisms. Specifically, the Cossee-Arlman mechanism on Cr^{III} catalyst is consistently predicted to be the most plausible pathway for polyethylene formation, whereas the same mechanism on the Cr^{II} analogue leads to ethylene oligomerization. In contrast, the metallacyle pathways exhibit considerably higher barriers than the Cossee-Arlman kinetically relevant barriers. Furthermore, the energetic preference for ring opening over ring expansion in larger chromacyclic intermediates of the $Cr^{(I/III)}$ cycle is reproduced by all tested functionals.

3.3. Conclusions

This chapter investigates ethylene oligomerization on Ni–NU-1000 and ethylene polymerization on Cr–NU-1000 single-atom catalysts (SACs) by DFT methods. A systematic exploration of computational and mechanistic aspects was first performed for the Ni-based system, including benchmarkings across periodic and cluster models, density functionals, and evaluation of framework effects. The insights obtained from this analysis were then applied to the Cr-based catalyst to gain deeper understanding of its structural and mechanistic behavior.

For Ni–NU-1000, two square-planar motifs were identified: a more open $Ni^{II}O_2$ site and a more saturated $Ni^{II}O_3$ site. The Cossee–Arlman pathway operates on both sites, but the $Ni^{II}O_2$ motif is significantly more active due to ability to maintain square-planar coordination throughout the catalytic cycle. Ethylene insertion is the turnover-determining stage for both motifs, with notably lower barrier on $Ni^{II}O_2$. The β-hydrogen elimination is facilitated by a β-agostic hydrogen interaction, resulting in preferential butene formation. Product release is favored by substitution with physisorbed ethylene under ambient conditions, while direct desorption becomes competitive at higher temperature and low ethylene pressure. An assessment of truncated models showed that both formate and benzoate models reproduce the periodic results within MADs of $4\,kJ\,mol^{-1}$, with the benzoate cluster models performing slightly better. The TPSSh free energy profile predicts TOF of the correct order of magnitude and resolves the three-order overestimation observed in previous theoretical studies. Extension from dimerization to oligomerization up to C_8 revealed pronounced framework effects. Chains oriented into the trigonal pores experience stabilizing dispersion interactions that maintain nearly constant insertion barriers across chain lengths and penalizes direct desorption, promoting further growth at low ethylene pressure. In contrast, growth toward the hexagonal pores leads to increasing insertion barriers but facilitates easier desorption.

For Cr-NU-1000 systems, we compared Cossee-Arlman and metallacycle mechanisms across Cr oxidation states I–IV and spin multiplicities. Cossee-Arlman mechanism on tetrahedral Cr^{III} sites shows low insertion barriers ($\approx 50\,kJ\,mol^{-1}$) and chain-termination barriers that are $>30\,kJ\,mol^{-1}$ higher, indicating a strong preference for polymer growth. This finding is consistent with the experimentally observed [Cr–ethyl] active site and

polyethylene formation. In contrast, Cossee-Arlman mechanism on Cr^{II} favors oligomerization, where chain-growth and chain-termination steps are energetically competitive due to the flexible Cr active sites that interconverts between trigonal bipyramidal and square planar geometries. Higher ethylene pressure lowers the effective insertion barriers for both catalysts but is unlikely to shift the Cr^{II} system to polymerization. Metallacyle pathways are kinetically unfavorable for both $Cr^{(I/III)}$ and $Cr^{(II/IV)}$ cycles. While oxidative C–C coupling can be accessible in the former, subsequent ring expansion or β-H transfer requires prohibitive barriers in both cycles (typically $\geq 90\,\mathrm{kJ\,mol^{-1}}$ to $200\,\mathrm{kJ\,mol^{-1}}$). Across all functionals, PBE underestimates barriers, B3LYP gives higher barriers, and M06-L and TPSSh yield consistent results. Crucially, all four functionals preserve the same qualitative conclusions.

Overall, this chapter demonstrates that theoretical simulations can successfully reproduce and rationalize the experimental observation in complex SACs supported on MOF systems. These agreements were achieved through a comprehensive evaluation of different theoretical aspects: model selection, electronic structure methods, catalytic mechanisms, active site geometries and their electronic properties. The results not only provide molecular-level insight into the catalytic ethylene conversion on MOF-based SACs but also serve as a methodological benchmark for future computational studies of heterogeneous catalysis in MOFs.

4. Anharmonic Corrections to the Adsorption Free Energy

Scope. This chapter presents the incorporation of machine learned force fields (MLFFs) into the thermodynamic λ-path integration (λ-TI), developed by Amsler and coworkers[160, 164, 224], and the intermediate hard sphere model (IHSM), developed by Pleßow[165], to compute accurately the adsorption free energy of several molecule and reactive species on transition metal surface. Section 4.1 explains in details the introduction of MLFFs into the two workflows and how the accuracy of the results is preserved at DFT level. The MLFF-aided λ-TI (MLFF-TI) workflow is applied to study the adsorption of three oxygenated species on Pt(111) surface in Section 4.2. Parts of Section 4.1 and 4.2 have been published and are adapted and reprinted with permission from [225]. Copyright 2025 American Chemical Society. The work was published with co-authors Bassim Mounssef Jr., who jointly developed the MLFF-TI workflow, Dmitry I. Sharapa, Felix Studt, and Tomáš Bučko in advisor roles. Section 4.3 employed the MLFF-aided IHSM (MLFF-IHSM) and MLFF-TI workflows to investigate the role of density functional on describing the physisorption and dissociative adsorption of CH_4 on Pt(111) surface. This work is to be published with the co-authors Dmitry I. Sharapa, Felix Studt, and Tomáš Bučko as advisors.

4.1. Method Development and Theory

4.1.1. Overview

Adsorption represents one of the most, if not the most, important processes in heterogeneous catalysis. Such an elementary step results from a subtle balance between enthalpy and entropy of adsorption, with the former often being favorable while the latter being opposing. Having an accurate estimation of these contributions, and therefore, free energy, is of utmost importance to gain deeper understanding of the catalytic processes on solid surfaces, particularly at finite temperatures.

The use of molecular dynamics (MD) simulations for studying heterogeneous catalysis has become more popular in recent years. This is partially due to the fact that static models cannot provide results that are straightforward to explain and a dynamic model is therefore needed. On the other hand, the trend is driven by the advancements in computing infrastructure and algorithm, and more recently, the burgeoning adoption of machine learned force fields (MLFFs), also known as machine learned interatomic

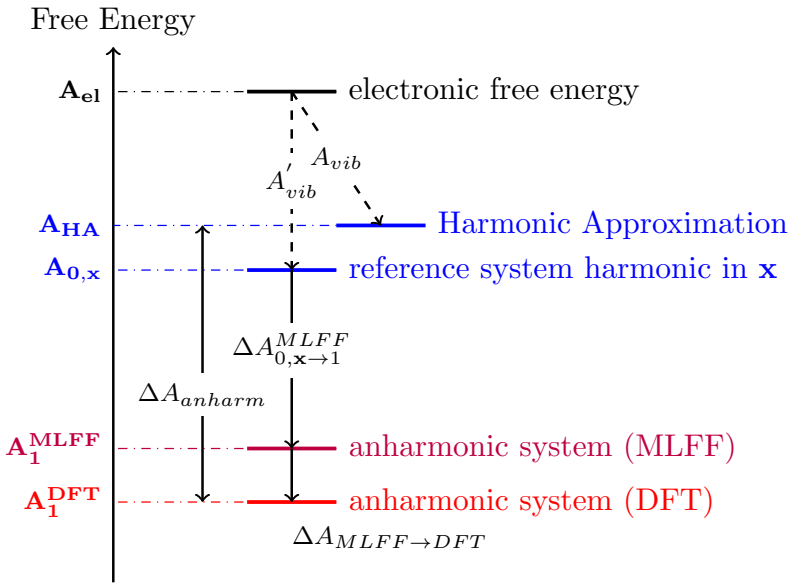
potentials (MLIPs). However, computing absolute free energies using MD simulations is inherently challenging. Typically, one can only calculate free energy difference between defined states.

The work of Amsler et al.[160, 164, 224] has established a rigorous framework for incorporating anharmonic effects, enabling accurate evaluation of adsorption and activation free energies. The anharmonic modes of a given system exhibiting along a fully interacting trajectory driven by an accurate electronic structure method, often DFT, can now be fully and accurately captured. Some of these modes, however, can only exhibit at a sufficiently long time interval, *e.g.,* subnanoseconds, thus requiring MD trajectories that are orders of magnitudes larger. Performing such lengthened MD simulations with common DFT methods is practically impossible.

A further limitation of Amsler's λ-TI methodology arises for physisorbed species, where constrained TI must be used with, at least, the surface-molecule distance being fixed. Consequently, very long MD trajectories are required to obtain reliable probability density distributions for this highly mobile translational mode.

Recently, Plessow proposed a strategy based on free energy perturbation (FEP) theory to compute anharmonic adsorption free energies of weakly interacting reactants in zeolite cavities relative to the gas phase[165]. By introducing an intermediate hard sphere model (IHSM), the method enhances the phase space overlap between separated and interacting systems, thereby overcoming the TI sampling issue. This newly developed IHSM method serves as an ideal complement to the existing TI-based approaches, effectively broadening the methodological toolkit for modeling adsorption thermodynamics. Nonetheless, similar to the original λ-TI method, this FEP-based approach still relies on multiple long DFT-MD simulations, which limits its broader applicability in large systems.

In this section, we introduce MLFFs as surrogate models for DFT, employed to perform the MD simulations required in the two anharmonic free energy approaches described above. The use of MLFFs drastically accelerates the calculations, enabling MD trajectories that would otherwise be computationally infeasible at the DFT level, thereby improving statistical sampling and convergence. Nevertheless, the MLFF-derived free energies are only as accurate as the employed MLFF accuracy. Deviations from free energy values of the reference DFT method on which the MLFF was trained are expected, though their magnitude is not straightforwardly assessable. It is therefore desirable to ensure DFT accuracy of the results predicted by MLFFs. In this section, we demonstrates the application of various free energy calculation methods as complementary tools to guarantee DFT accuracy.



Scheme 4.1.: Relations between various free energy terms involved in the MLFF-TI workflow. The true harmonic free energy (A_{HA}) and free energy of the reference harmonic in Cartesian coordinates ($A_{0,\boldsymbol{x}}$) are obtained analytically from the respective Hessian matrices. Reprinted with permission from [225]. Copyright 2025 American Chemical Society.

4.1.2. The Machine Learned Force Field-aided Thermodynamic Integration (MLFF-TI) Workflow

In the thermodynamic integration (TI) method[161], the free energy of a target system 1 is expressed relative to a chosen reference system 0 as:

$$A_1 = A_0 + \Delta A_{0\to 1}, \tag{4.1}$$

where $\Delta A_{0\to 1}$ represents the Helmholtz free energy difference between the two systems. Various reference systems have been employed depending on the system and property of interest[167, 226–232]. In the λ-TI method developed by Amsler et al.[160, 164, 224], a harmonic system was selected as a reference system, resulting in the term $\Delta A_{0\to 1}$ being interpreted as the anharmonic correction to the free energy estimated by the harmonic oscillator approximation (HA) if the reference A_0 is exactly the standard HA free energy. In practice, one can choose any arbitrary harmonic reference. The Eq. (4.1) can be rewritten in terms of a general harmonic reference in Cartesian coordinates as:

$$A_1 = A_{0,\boldsymbol{x}} + \Delta A_{0,\boldsymbol{x}\to 1}, \tag{4.2}$$

where $A_{0,\boldsymbol{x}}$ is the free energy of an arbitrary reference system that is harmonic in Cartesian coordinates, $\Delta A_{0,\boldsymbol{x}\to 1}$ is the free energy difference between the anharmonic and the harmonic reference systems, defined as:

$$A_{0,\boldsymbol{x}\to 1} = -k_{\mathrm{B}}T \ln\left[\frac{Q_1}{Q_{0,\boldsymbol{x}}}\right]. \tag{4.3}$$

Here, the $Q_\alpha = M \int \exp\left(-\mathcal{H}_\alpha / k_\mathrm{B}T\right)\, d\boldsymbol{p}\, d\boldsymbol{q}$ is the partition function of the system α described by the Hamiltonian $\mathcal{H}$, M is a normalization constant, T is the temperature, and k_B is the Boltzmann constant. Since the MD simulations are inherently semiclassical, the first term on the right hand side of Eq. (4.2) is computed via the classical free energy expression:

$$A_{0,\boldsymbol{x}} = A_{el}(\boldsymbol{x}_0) - k_\mathrm{B}T \sum_{i=1}^{N_{vib}} \ln \frac{k_\mathrm{B}T}{\hbar\omega_i}, \tag{4.4}$$

where $A_{el}(\boldsymbol{x}_0)$ is the electronic free energy for the minimum energy configuration of the reference system, N_{vib} denotes the number of vibrational modes i with real angular frequencies ω_i obtained via the finite difference method, and $\hbar$ is the reduced Planck's constant. It should be noted that the classical anharmonic contribution can be used to correct the free energy obtained from quantum harmonic partition function as it is straightforward that from the Pitzer–Gwinn approximation[146, 233, 234]:

$$Q_{1,\text{quantum}} = Q_{1,\text{classical}} \times \frac{Q_{0,\text{ quantum}}}{Q_{0,\text{ classical}}}, \tag{4.5}$$

one can obtain:

$$A_{1,\text{quantum}} = A_{0,\text{quantum}} + \Delta A_{0\to 1,\text{classical}}. \tag{4.6}$$

where $Q_{0,\text{quantum}}$, $Q_{1,\text{quantum}}$ denote the quantum mechanical partition functions of the harmonic and anharmonic systems, respectively, while $Q_{0,\text{classical}}$, and $Q_{1,\text{classical}}$ represent their corresponding classical counterparts. The terms $A_{1,\text{quantum}}$ and $A_{0,\text{quantum}}$ refer to the quantum mechanical Helmholtz free energies of the harmonic and anharmonic systems, respectively, and $\Delta A_{0\to 1,\text{classical}} \equiv \Delta A_{0\to 1}$ is the Helmholtz free energy difference on the right hand side of Eq. (4.1).

The electronic free energy is calculated as[235]:

$$A_\mathrm{el}\left(\boldsymbol{x}_0\right) = -k_\mathrm{B}T \ln\left[g_0 \exp\left(\frac{-\mathcal{V}_{0,\boldsymbol{x}}\left(\boldsymbol{x}_0\right)}{k_\mathrm{B}T}\right)\right] \tag{4.7}$$

with g_0 being the spin multiplicity of the ground state. The second term of Eq. (4.2), $\Delta A_{0,\boldsymbol{x}\to 1}$, can be calculated via λ-TI using straightforward unconstrained MDs as:

$$\Delta A_{0,\boldsymbol{x}\to 1} = \int_0^1 \left\langle \mathcal{V}_1 - \mathcal{V}_0 \right\rangle_\lambda \, d\lambda, \tag{4.8}$$

where λ is the coupling parameter describing the transformation path from the harmonic reference state ($\lambda = 0$) to the fully interacting state ($\lambda = 1$), $\mathcal{V}_0$ and $\mathcal{V}_1$ being the potential energies of a configuration evaluated by harmonic and anharmonic Hamiltonians, respectively. The harmonic potential energy is computed from a harmonic force field as:

$$\mathcal{V}_0(\boldsymbol{x}) = \mathcal{V}_0(\boldsymbol{x}_0) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}_0)^\mathrm{T} \underline{\mathbf{H}}^\mathrm{x} (\boldsymbol{x} - \boldsymbol{x}_0), \tag{4.9}$$

where $\underline{\mathbf{H}}^{\mathbf{x}}_{ij} = \frac{\partial^2 \mathcal{V}_{0,\boldsymbol{x}}(\boldsymbol{x})}{\partial x_i \partial x_j}|_{\boldsymbol{x}=\boldsymbol{x}_0}$ is the Hessian matrix determined for an arbitrary reference $\boldsymbol{x_0}$. The $\mathcal{V}_1$ is predicted by an MLFF. The $\langle \cdots \rangle_\lambda$ is the ensemble average of the quantity over the configuration space at state λ, described by the Hamiltonian:

$$\mathcal{H}_\lambda = \lambda \mathcal{H}_1 + (1-\lambda)\mathcal{H}_0. \tag{4.10}$$

Alternatively but equivalently, constrained λ-TI[164] could be employed, which is a preferable choice for systems with convergence issues. In this method, holonomic constraints are introduced to control degrees of freedom heavily affecting the sampling convergence. The contributions of the modes are, then, decoupled from the $\Delta A_{0\to1}$:

$$\Delta A_{0,\boldsymbol{x}\to1} = \Delta A_{0\to1}(\xi') - k_\mathrm{B}T \ln\left[\frac{P_0(\xi')}{P_1(\xi')}\right]. \tag{4.11}$$

The term, $\Delta A_{0,\boldsymbol{x}\to1}$ is defined as the free energy difference between the harmonic reference system and the fully interacting system with $\xi(\boldsymbol{x})$ being constrained at ξ'. This term can be computed as:

$$\Delta A_{0,\boldsymbol{x}\to1}(\xi') = \int_0^1 d\lambda \frac{1}{\langle Z^{-1/2}\rangle_{\lambda,\xi'}} \langle Z^{-1/2}(\mathcal{V}_1 - \mathcal{V}_{0,\boldsymbol{x}})\rangle_{\lambda,\xi'}. \tag{4.12}$$

where $\langle \cdots \rangle_{\lambda,\xi'}$ is the ensemble average of the quantity over the configuration space at state λ with $\xi(\boldsymbol{x}) = \xi'$, Z is the inverse mass metric tensor, defined as:

$$Z = \sum_{i=1}^{N} \frac{1}{m_i} \sum_{\mu=x,y,z} \left(\frac{\partial \xi}{\partial x_{i,\mu}}\right)^2 \tag{4.13}$$

where m_1 is the mass of atom i and μ is the Cartesian component of the position vector $\boldsymbol{x}$. The term $P_{0,\boldsymbol{x}}(\xi')$ and $P_{0,\boldsymbol{x}}(\xi')$ are the probabilities of $\xi(\boldsymbol{x})$ being at ξ' in the NVT ensembles governed by harmonic force field and MLFF, respectively. These terms can be obtained from probability density analyses of the two respective unconstrained trajectories.

Note that the Eq. (4.11) can easily be generalized to multiple independent degrees of freedom:

$$\Delta A_{0,\boldsymbol{x}\to1} = \Delta A_{0\to1}(\xi'_1, \xi'_2, \ldots, \xi'_n) - k_\mathrm{B}T \ln \frac{P_0(\xi'_1, \ldots, \xi'_n)}{P_1(\xi'_1, \ldots, \xi'_n)} \tag{4.14}$$

and can serve as the basis for the approximation:

$$\begin{aligned} \Delta A_{0,\boldsymbol{x}\to1} &\approx \Delta A_{0\to1}(\xi'_1, \xi'_2, \ldots, \xi'_n) - k_\mathrm{B}T \ln \frac{\prod_{i=1}^n P_0(\xi'_i)}{\prod_{i=1}^n P_1(\xi'_i)} \\ &= \Delta A_{0\to1}(\xi'_1, \xi'_2, \ldots, \xi'_n) - \sum_{i=1}^{n} k_\mathrm{B}T \ln \frac{P_0(\xi'_i)}{P_1(\xi'_i)} \end{aligned} \tag{4.15}$$

with n being the number of constrained degrees of freedom. We further discuss the validity of this approximation in Section A.2 of the Appendix. The relations between different free energy terms involved in this MLFF-TI workflow are depicted in Scheme 4.1.

The developed MLFF-TI workflow offers two major practical advantages over the original λ-TI methods introduced by Amsler et al.[160, 164, 224]. First, MLFFs enable nanosecond-scale, near-DFT-accuracy MD simulations, on the order of 5×10^6 MD steps, at a fraction of the DFT computational cost, thereby providing highly reliable statistical convergence that would otherwise be computationally infeasible with pure DFT. Second, while the original approach required carefully designed internal coordinates for each system, the present workflow greatly simplifies the choice of coordinates for defining the harmonic reference. Even for problematic degrees of motion, a few well-chosen geometric constraints are sufficient to improve convergence while adding minimal coordinate complexity. Furthermore, as an inherent properties of the λ-TI formalism, the total free energy, A_1, is independent of the choice of the harmonic reference. Consequently, the reference can be entirely generated using MLFF, meaning that all quantities relevant to the harmonic part, *e.g.*, $A_{el}(\boldsymbol{x}_0)$, ω_i, and $\underline{\mathbf{H}}^{\mathbf{x}}$, are computed consistently at the MLFF level, further reducing computational cost. More importantly, this approach establishes a conceptual framework for future free energy calculations in which every steps can be performed using a cost-effective, pretrained universal MLFF. The resulting free energy estimates can then be systematically corrected to a desired DFT-level accuracy via free energy perturbation theory (Section 4.1.3) or machine learning thermodynamic perturbation (Section 4.1.5) using only a limited number of DFT evaluations. The robustness of the MLFF-TI workflow is demonstrated in Section 4.1.8.

4.1.3. Free Energy Perturbation Theory

Free energy perturbation (FEP) theory is a widely used and powerful method for estimating free energy difference between an unperturbed reference system and a perturbed or target system, which is often the system of interest. The Hamiltonian of the target system, $\mathcal{H}_1$, can be expressed as the Hamiltonian of the reference system, $\mathcal{H}_0$, plus a perturbation term, $\Delta\mathcal{H}$, that is:

$$\mathcal{H}_1 = \mathcal{H}_0 + \Delta\mathcal{H}. \tag{4.16}$$

The Helmholtz free energy difference, ΔA, between the two systems can then be evaluated as[161, 236]:

$$\Delta A = -k_\mathrm{B}T \ln \left\langle \exp\left(-\frac{\Delta\mathcal{H}}{k_\mathrm{B}T}\right)\right\rangle_0 . \tag{4.17}$$

Because the kinetic term contribute equally to both systems, it cancels out and the Eq. (4.17) becomes:

$$\Delta A = -k_\mathrm{B}T \ln \left\langle \exp\left(-\frac{\Delta\mathcal{V}}{k_\mathrm{B}T}\right)\right\rangle_0 = -k_\mathrm{B}T \ln \left\langle \exp\left(-\frac{\mathcal{V}_1 - \mathcal{V}_0}{k_\mathrm{B}T}\right)\right\rangle_0 , \tag{4.18}$$

where $\Delta\mathcal{V} = \mathcal{V}_1 - \mathcal{V}_0$ is the potential energy difference between the target system and the reference system. The Helmholtz free energy difference, ΔA, from Eq. (4.18) can be determined by sampling only the ensemble of the reference state 0 and the calculation is so-called forward FEP. Alternatively, we can obtain the backward FEP expression by interchanging the roles of the two systems:

$$\Delta A = +k_\mathrm{B}T \ln \left\langle \exp\left(+\frac{\Delta\mathcal{V}}{k_\mathrm{B}T}\right)\right\rangle_1 = +k_\mathrm{B}T \ln \left\langle \exp\left(+\frac{\mathcal{V}_1 - \mathcal{V}_0}{k_\mathrm{B}T}\right)\right\rangle_1 , \tag{4.19}$$

Here, the ensemble average is evaluated over the equilibrium configurations of the target system 1.

The MLFF-TI workflow discussed in Section 4.1.2 provides free energy estimates that are accurate up to the level of the employed MLFF level. The resulting free energies depend not only on the accuracy of the MLFF in reproducing DFT energies and forces but also on the phase space it samples compared to that obtained from DFT. Consequently, it is plausible that different MLFFs, even when trained on the same DFT level, may yield slightly different free energy values, making direct comparison between systems prone to errors. To recover DFT-level accuracy, the forward FEP approach can be applied to compute the free energy difference between MLFF and DFT:

$$\Delta A_{\mathrm{MLFF}\to\mathrm{DFT}} = -k_{\mathrm{B}}T \ln \left\langle \exp\left(-\frac{\mathcal{V}_{\mathrm{DFT}} - \mathcal{V}_{\mathrm{MLFF}}}{k_{\mathrm{B}}T}\right)\right\rangle_{\mathrm{MLFF}}, \tag{4.20}$$

where $\mathcal{V}_{\mathrm{DFT}}$ and $\mathcal{V}_{\mathrm{MLFF}}$ are the potential energies computed at the DFT and MLFF levels, respectively, and $\langle \cdots \rangle_{\mathrm{MLFF}}$ is the ensemble average over the configurations sampled from the MLFF trajectories. This expression can be interpreted as the free energy needed to shift the system from MLFF PES to that of DFT. Conversely, the reverse transformation can be evaluated as:

$$\Delta A_{\mathrm{DFT}\to\mathrm{MLFF}} = -\Delta A_{\mathrm{MLFF}\to\mathrm{DFT}} = +k_{\mathrm{B}}T \ln \left\langle \exp\left(-\frac{\mathcal{V}_{\mathrm{DFT}} - \mathcal{V}_{\mathrm{MLFF}}}{k_{\mathrm{B}}T}\right)\right\rangle_{\mathrm{MLFF}}, \tag{4.21}$$

which would be of use in Section 4.1.7. That section also makes intensive use of the FEP formalism, which will be outlined in greater detail later.

4.1.4. Evaluation of Phase Space Overlap: the $\mathrm{I_w}$ Index

The expressions for free energy differences (Eqs. (4.18) and (4.19)) are formally exact for any perturbation. However, this does not guarantee that the results obtained are necessarily accurate in practice. The criterion for a reliable perturbation is a sufficiently overlapping between the phase spaces of the two systems, which requires the low-energy tail of the $\Delta\mathcal{V}$ distribution to be sufficiently sampled[161, 237]. To evaluate such an phase space overlap, we used the I_w index, proposed by Bučko et al.[238]. The index is defined as the minimal proportion of configurations, sorted in the ascending order of potential energy difference, which collectively contribute at least 50% of the Boltzmann weight, $w_i = \exp \frac{-\Delta\mathcal{V}_i}{k_{\mathrm{B}}T}$. By definition, I_w ranges from 0 (no overlap) to 0.5 (perfect overlap), serving as a simple yet informative diagnostic for evaluating FEP convergence. It has been shown that an I_w value of 0.03 could already be able to yield reliable perturbations[239]. A detailed discussion of this diagnostic is presented in Section 4.1.8.3.

4.1.5. Machine Learning Thermodynamic Perturbation

The MLFFs used in this chapter were trained specifically at the DFT level of interest, and therefore, the phase space overlap, as relected in the I_w index, is significant. As a result,

performing DFT single point (SP) calculation on only 100-200 uncorrelated configurations from the production trajectories is sufficient to obtain reliable free energy correction.

With the growing development and popularity of universal MLFFs, it is highly desirable to integrate these models into the MLFF-TI framework to reduce the computational cost associated with MLFF training. However, this approach introduces an overlap issue as the phase spaces sampled by universal MLFFs are often deviate from those of the target DFT level employing more rigorous computational settings. This discrepancy arises because most training data for universal models were generated using looser convergence criteria to accelerate dataset construction (with the exception for OMol dataset[240]). The insufficient overlap between MLFF and DFT phase spaces renders the correction in Eq. (4.20) when applied through brute-force FEP using a limited number of configurations, susceptible to large statistical uncertainty. Improving the statistics of the perturbation require tens to hundreds of thousand SP calculations, thereby negating the computational efficiency gained from using MLFFs. To address this challenge, we applied the machine learning thermodynamic perturbation approach[238, 239, 241, 242], which was initially proposed to obtain enthalpies and free energies at RPA level from the production trajectories at PBE-D2 level. Specifically, a Δ-ML model is trained on DFT energies of tens of configurations extracted from the universal MLFF product run and was used to predict the potential energy difference for as much as the whole production trajectories (*i.e.,* millions of steps). We recommend to refer to Ref. [238, 239, 241, 242] for details of the Δ-ML model. The Helmholtz free energy difference between the universal MLFF and the target DFT level is computed as:

$$\Delta A_{\mathrm{MLFF}\rightarrow\mathrm{DFT}} = -k_{\mathrm{B}}T \ln \left\langle \exp\left(-\frac{\Delta\tilde{\mathcal{V}}}{k_{\mathrm{B}}T}\right)\right\rangle_{\mathrm{MLFF}}, \tag{4.22}$$

where $\Delta\tilde{\mathcal{V}}$ is predicted using the trained Δ-ML model. An exemplary free energy estimation for O@Pt system at 400 K was carried out using the `GemNet-OC-S2EFS-OC20+OC22` model[243] in `fairchem` package[244] and is discussed in Section 4.1.8.4.

4.1.6. Bennett's Acceptance Ratio

When FEP data are collected for both reference and target ensembles, the free energy difference can be evaluated using either forward (Eq. (4.18)) or backward (Eq. (4.19)) formulation. Although these two expressions are formally equivalent, their convergence behavior may differ, leading to a preferred perturbation direction[161, 237], which can be estimated by the I_w index described earlier. To address this issue, Bennett proposed an more efficient and statistically robust approach for computing free energy differences, later known as the Bennett's acceptance ratio (BAR) method[245]. In the BAR formalism[245, 246], the free energy difference is given by:

$$\Delta A_{0\rightarrow 1} = k_{\mathrm{B}}T \left(\ln \frac{\sum_{i=1}^{N_1} f\left(-\Delta\mathcal{V} + C\right)}{\sum_{i=1}^{N_0} f\left(\Delta\mathcal{V} - C\right)} - \ln \frac{N_1}{N_0} \right) + C \tag{4.23}$$

where N_0 and N_1 are the sample sizes for the forward and backward FEP calculations, respectively, f is the Fermi function,

$$f(x) = \frac{1}{1 + \exp\left(\frac{x}{k_B T}\right)}. \tag{4.24}$$

The term C is defined as:

$$C = k_B T \ln \frac{Q_0 N_1}{Q_1 N_0}. \tag{4.25}$$

and can be determined iteratively based on the condition:

$$\sum_{i=1}^{N_1} f\left(-\Delta\mathcal{V} + C\right) = \sum_{i=1}^{N_0} f\left(\Delta\mathcal{V} - C\right). \tag{4.26}$$

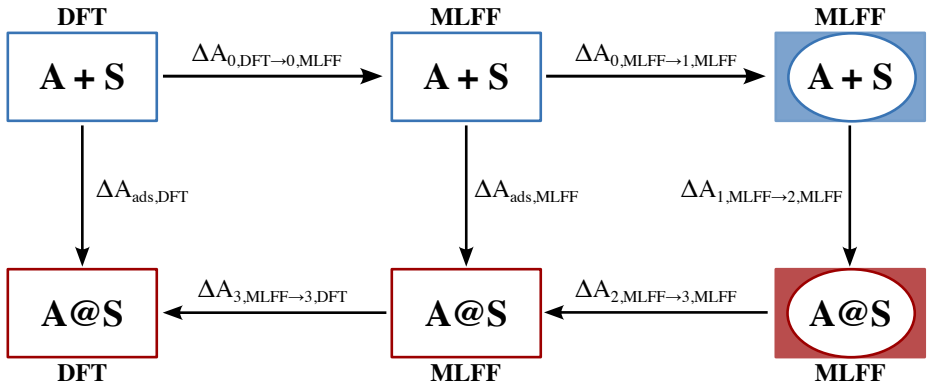
The BAR estimate of the free energy difference is then obtained as:

$$\Delta A_{0\to1}^{\mathrm{BAR}} = C - k_B T \ln \frac{N_1}{N_0}. \tag{4.27}$$

The BAR method provides a highly accurate and statistically efficient way to compute free energy differences when trajectories for both states are available, such as for the intermediate hard sphere model discussed in Section 4.1.7. Furthermore, the multistate BAR (MBAR) approach[247] can, in principle, serve as an alternative to the MLFF-TI method described previously, offering a means to reduce the number of sampled λ points required for convergence.

4.1.7. The Machine Learned Force Field-aided Intermediate Hard Sphere Method (MLFF-IHSM)

While MLFF-TI method described in Section 4.1.2 is highly effective for calculating free energies of adsorption of intermediate species and chemisorbed molecules with varying degrees of anharmonicity, physisorbed molecules remain practically challenging. This difficulty arises because the constrained TI must be used with at least the surface-molecule being fixed, and thus, long MD trajectories are required to obtain reliable probability distributions for this constrained coordinate. Recently, Plessow developed a new approch based on FEP theory to compute anharmonic adsorption free energies of weakly bound reactants in zeolite cages relative to the gas phase[165]. To enhance the overlap between separated and interacting systems, an intermediate hard sphere model (IHSM) was introduced, thereby circumventing the TI sampling issue. However, like the original λ-TI method[160, 164, 224], this FEP-based approach still requires multiple long DFT-based MD simulations, which limits its broader applicability in large systems. To overcome this limitation, MLFFs can also be employed in place of DFT to drastically reduce the computational burden, while maintaining DFT accuracy. This section introduces the resulting



Scheme 4.2.: Schematic representation of the thermodynamic pathway used to calculate free energy of adsorption, $\Delta A_{\mathrm{ads,DFT}}$, in the MLFF-IHSM approach. **A + S** and **A@S** represent the separated and interacting systems of adsorbate and substrate, respectively. Shaded boxes with a white oval represent the IHSM states.

MLFF-aided IHSM (MLFF-IHSM) method (Scheme 4.2); for detailed methodological aspects, readers are referred to Ref. [165].

Central to the calculation of adsorption free energy in the IHSM approach is the interaction energy between adsorbate and the solid substrate. This energy is defined as the difference in the Hamiltonian between the noninteracting reference ("sep") system, consisting of the clean substrate and the gas-phase adsorbate, and the interacting ("int") system. Because all MD simulations in the MLFF-IHSM approach are (partly) driven by MLFFs, the interaction energy, $\mathcal{V}_{\mathrm{int}}$, is expressed as:

$$\Delta\mathcal{H}_{\mathrm{MLFF}} = \mathcal{H}_{\mathrm{int,MLFF}} - \mathcal{H}_{\mathrm{sep,MLFF}} \tag{4.28}$$

$$= \mathcal{V}_{\mathrm{int}}. \tag{4.29}$$

Each simulation used to evaluate $\Delta A_{\mathrm{ads,MLFF}}$ follows a general Hamiltonian:

$$\mathcal{H} = \mathcal{H}_{\mathrm{sep,MLFF}} + \mathcal{V}, \tag{4.30}$$

where $\mathcal{V}$ varies depending on the state, as defined in Eqs. (4.31)-(4.34). Here, $\mathcal{V} = 0$ corresponds to the noninteracting state, and $\mathcal{V} = \mathcal{V}_{\mathrm{int}}$ to the fully interacting state.

$$\mathcal{V}_{0,\mathrm{MLFF}} = 0 \tag{4.31}$$

$$\mathcal{V}_{1,\mathrm{MLFF}} = \mathcal{V}_{\mathrm{IHSM}} \tag{4.32}$$

$$\mathcal{V}_{2,\mathrm{MLFF}} = \mathcal{V}_{\mathrm{int}} + \mathcal{V}_{\mathrm{IHSM}} \tag{4.33}$$

$$\mathcal{V}_{3,\mathrm{MLFF}} = \mathcal{V}_{\mathrm{int}} \tag{4.34}$$

The IHSM is defined as:

$$\mathcal{V}_{\mathrm{IHSM}} = \sum_{i=1}^{N_{\mathrm{substrate}}} \sum_{j=1}^{N_{\mathrm{adsorbate}}} \mathcal{V}_{\mathrm{IHSM}}^{ij}(r_{ij}) \tag{4.35}$$

where the pairwise potential is:

$$\mathcal{V}_{\mathrm{IHSM}}^{ij}(r_{ij}) = \begin{cases} 0 & \text{if } r_{ij} > r_{ij}^{\mathrm{hard}} \\ \infty & \text{else} \end{cases}. \tag{4.36}$$

Here, r_{ij} denotes the pairwise atomic distance, and r_{ij}^{hard} represents the hard-sphere cutoff radius .

For adsorption on solid surfaces that are modeled as slabs, the IHSM can alternatively be defined as the confinement potential restricting the adsorbate within a vacuum layer between h_{min} and h_{max}:

$$\mathcal{V}_{\text{IHSM}} = \sum_{i=1}^{N_{\text{adsorbate}}} \mathcal{V}_{\text{IHSM}}^{i}(z_i) \tag{4.37}$$

$$\mathcal{V}_{\text{IHSM}}^{i}(z_i) = \begin{cases} 0 & \text{if } h_{\text{min}} < z_i < h_{\text{max}} \\ \infty & \text{else} \end{cases} . \tag{4.38}$$

where z_i is the position of the atom i along the z direction.

The Helmholtz adsorption free energy at MLFF level is computed as:

$$\Delta A_{\text{ads,MLFF}} = \Delta A_{0,\text{MLFF}\rightarrow 1,\text{MLFF}} + \Delta A_{1,\text{MLFF}\rightarrow 2,\text{MLFF}} + \Delta A_{2,\text{MLFF}\rightarrow 3,\text{MLFF}}. \tag{4.39}$$

The first contribution is evaluated using the forward FEP expression:

$$\Delta A_{0,\text{MLFF}\rightarrow 1,\text{MLFF}} = -k_{\text{B}}T \ln \left\langle \exp\left(-\frac{\mathcal{V}_{\text{IHSM}}}{k_{\text{B}}T}\right)\right\rangle_0 \tag{4.40}$$

$$= -k_{\text{B}}T \ln\left(\frac{v_1}{v_0}\right) \tag{4.41}$$

Determining this term requires pairing configurations of the gas-phase adsorbate and the clean substrate obtained from the corresponding separate MDs. A procedure similar to Ref. [165] is employed, in which every 100^{th} MD steps of the two 4×10^6-step trajectories are paired to each other. For each pair, 10 random positions in the simulation cell are selected for the adsorbate. This pairing generates a trajectory of 4×10^5 pair configurations for evaluating the potential. The ensemble average of the exponential term is corresponds to the fraction of pairs that do not violate the IHSM. Hence, this term can be interpreted as the work required to reduce the accessible volume from the full simulation cell v_0 to teh confined region v_1 (Eq. (4.41)). Using the IHSM definition from Eqs. (4.37) and (4.38), the volume v_1 can be obtained straightforwardly.

The second term in Eq. (4.39) is computed either by forward (Eq. (4.42) or backward FEP (Eq. (4.43), or more reliably via BAR method (Eqs. (4.44) and (4.44)).

$$\Delta A_{1,\text{MLFF}\rightarrow 2,\text{MLFF}}^{\text{FEP-fw}} = -k_{\text{B}}T \ln \left\langle \exp\left(-\frac{\mathcal{V}_{\text{int}}}{k_{\text{B}}T}\right)\right\rangle_1 \tag{4.42}$$

$$\Delta A_{1,\text{MLFF}\rightarrow 2,\text{MLFF}}^{\text{FEP-bw}} = +k_{\text{B}}T \ln \left\langle \exp\left(+\frac{\mathcal{V}_{\text{int}}}{k_{\text{B}}T}\right)\right\rangle_2 \tag{4.43}$$

$$\Delta A_{1,\text{MLFF}\rightarrow 2,\text{MLFF}}^{\text{BAR}} = C - k_{\text{B}}T \ln\left(\frac{N_2}{N_1}\right) \tag{4.44}$$

where C is determined iteratively by solving the following equation:

$$\sum_{i=1}^{N_2} f(-\mathcal{V}_{\text{int}}^{i} + C) = \sum_{i=1}^{N_1} f(\mathcal{V}_{\text{int}}^{i} - C). \tag{4.45}$$

Here, N_1 and N_2 are the numbers of structures from trajectories of state 1 and 2, respectively, that are employed in the perturbations, $f(x)$ is the Fermi function (Eq. (4.24)).

The final contribution is obtained from backward FEP:

$$\Delta A_{2,\text{MLFF}\to 3,\text{MLFF}} = +k_{\text{B}}T \ln \left\langle \exp\left(-\frac{\mathcal{V}_{\text{IHSM}}}{k_{\text{B}}T}\right)\right\rangle_3 \tag{4.46}$$

This term, as opposed to $\Delta A_{0,\text{MLFF}\to 1,\text{MLFF}}$, can be understood as the work released as the accessible volume expands in the interacting system.

The overall $\Delta A_{\text{ads,MLFF}}$ is obtained from multiple MD trajectories (5×10^6 steps), ensuring statistical reliability. The adsorption free energy can then be corrected to DFT accuracy following the Scheme 4.2:

$$\Delta A_{\text{ads,DFT}} = \Delta A_{0,\text{DFT}\to 0,\text{MLFF}} + \Delta A_{\text{ads,MLFF}} + \Delta A_{3,\text{MLFF}\to 3,\text{DFT}}. \tag{4.47}$$

Here, the first correction term is evaluated separately for gas-phase adsorbate and clean substrate using Eq. (4.21), while the last term is obtained via Eq. (4.20).

4.1.8. Robustness of the MLFF-TI Workflow

The incorporation MLFFs enables the generations of ns-scale MD trajectories, thereby greatly reducing the statistical uncertainty in the calculated results. In this section, we further examine the performance and robustness of the MLFF-TI approach. In particular, we assess the reference-independence of the λ-TI formalism to: (i) verify the accuracy of surface anharmonicity calculations, and (ii) demonstrate the use of MLFF-based harmonic references as a means to minimize the number of DFT evaluations required. Finally, we discuss the reliability of the FEP approach and evalute its consistency using the overlap diagnostic index I_w. The O@Pt system at 300 K, described in Section 4.2, serves as the test case for this analysis. The two sets of MLFFs used in this section, namely **MLFF-400** and **MLFF-1K**, were trained primarily on MD trajectories at 400 K and 1000 K, respectively. Details of the training can be found in Section A.1 of the Appendix. Readers are encouraged to consult subsection 4.2.2.1 for details on the technical computations of free energy terms discussed below. The results presented in this subsection substantiate the discussion provided in subsection 4.2.2.3.

4.1.8.1. Consistency of MLFF-TI results accounting for surface vibrational modes

It is important to note that the entropy contributions of surface atoms are often inaccurately captured within the HA. This shortcoming primarily arises from the anharmonic behavior of soft vibrational modes and the inherent limitations of DFT methods in describing them, leading to considerable uncertainty in HA entropy evaluation. Nevertheless, surface entropy plays a decisive role in many catalytic processes, particularly those occurring on defective surfaces, supported catalysts, or zeolitic systems[248].

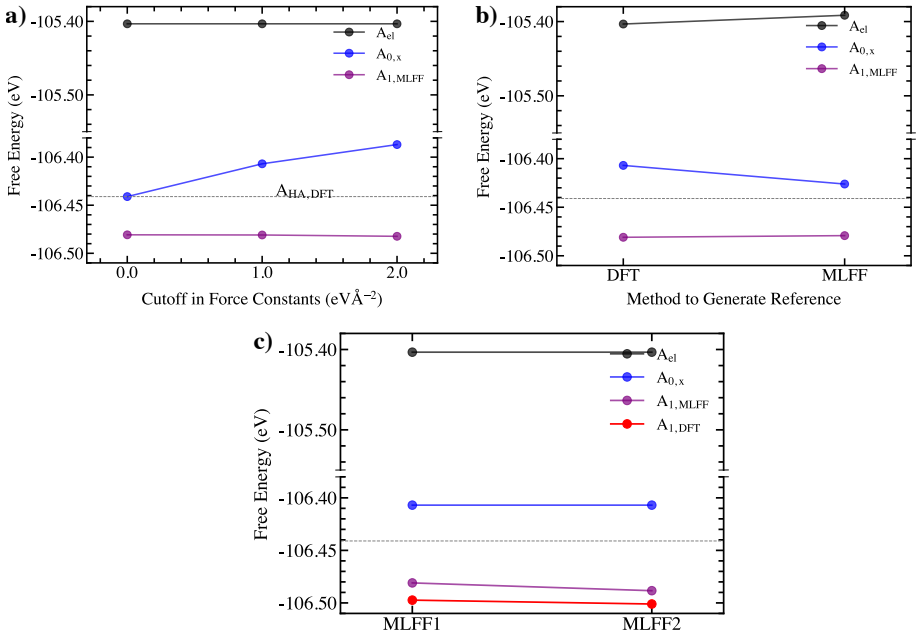


Figure 4.1.: Demonstrations of the robustness of the MLFF-TI approach in O@Pt (111) system. a) Anharmonic correction for different harmonic references defined by different minimum limits in Hessian matrix's eigenvalues showing the convergence of the anharmonic free energies $A_{1,\mathrm{MLFF}}$. This implies the decoupling of free energy estimated by λ-TI from the uncertainty of harmonic reference in evaluating low-frequency vibrational free energy. b) Demonstrations of the anharmonic free energies $A_{1,\mathrm{MLFF}}$ being independent of whether the harmonic reference is generated by DFT or MLFF. c) Demonstration of the total free energy, $A_{1,\mathrm{DFT}}$ being independent of the employed MLFFs thanks to FEP. The error bars of $A_{1,\mathrm{MLFF}}$ and $A_{1,\mathrm{DFT}}$ are negligible within the visual representation. Reprinted with permission from [225]. Copyright 2025 American Chemical Society.

The λ-TI method offers an effective means to overcome the deficiencies of the HA as it evaluates the entropic effects by sampling the full PES rather than only the vicinity of the minimum. Consequently, the computed free energy becomes independent of the chosen harmonic reference, effectively decoupling anharmonic contributions from reference-related uncertainties. This property is clearly illustrated in Figure 4.1a for the O@Pt(111) system. Raising the minimum limits in force constants from 0.0 (*i.e.,* no limit) to 2.0 eVÅ^{-2} shifts the soft-mode frequencies from 19.5 cm^{-1} to 53.6 cm^{-1}. As a result, the entropic contributions of these modes decrease, reducing the corresponding uncertainty in the HA

estimation. The concept underpins the correction scheme proposed by Ribeiro et al.[143], in which, vibrational frequencies below a certain threshold (typically 50 cm^{-1} or 100 cm^{-1}) are shifted upward to that value prior to entropy evaluation by HA. Such a pragmatic approach has been widely adopted across various systems, including adsorption studies. In the present case, this correction results in a 54 meV increase in the HA free energy $A_{0,\boldsymbol{x}}$, from −106.441 eV to −106.387 eV (Table 4.1). However, this variation is fully compensated by the corresponding $\Delta A_{0\rightarrow 1}$ term, yielding nearly identical anharmonic free energies of -106.481 ± 0.001 across all references. Thus, irrespective of the chosen harmonic reference, the λ-TI method consistently delivers reliable and reference-independent estimates of anharmonic free energies that include surface vibrational contributions.

Table 4.1.: Free energies in eV of the O@Pt(111) system calculated using different references defined by different minimum limits in Hessian matrix' eigenvalues as well as being generated by DFT and MLFF. The MLFF400-OPt was used for all calculations.

Method	C_{limit} (eV Å^{-2})	A_{el}	$A_{0,\boldsymbol{x}}$	$\Delta A^{MLFF}_{0,\boldsymbol{x}\rightarrow 1}$	A^{MLFF}_{1}
DFT	0	-105.403	-106.441	-0.040 (0.001)	-106.481 (0.001)
DFT	1	-105.403	-106.407	-0.074 (0.001)	-106.481 (0.001)
DFT	2	-105.403	-106.387	-0.095 (0.001)	-106.482 (0.001)
MLFF400-OPt	1	-105.392	-106.426	-0.053 (0.001)	-106.479 (0.001)

4.1.8.2. The use of MLFF-generated harmonic reference

Because the λ-TI methods are inherently reference-independent, the total free energy at the MLFF level, A^{MLFF}_{1}, depends solely on the PES predicted by the corresponding MLFF. When the MLFF-PES is sufficiently continuous for Hessian matrix generation, it is feasible to construct the harmonic reference directly at the MLFF level for subsequent TI calculations. This approach eliminates the need for additional DFT-based geometry optimizations and frequency analyses, substantially reducing the computational cost, particularly for large systems such as metal–organic frameworks (MOFs) that contain hundreds of atoms per simulation cell. To assess this strategy, we applied it to the O@Pt system using the harmonic reference generated by the **MLFF400–OPt** model, with the results shown in Figure 4.1b. The MLFF slightly overestimates the electronic free energy of the reference by 11 meV, yielding a value of −105.392 eV, and underestimates the vibrational frequencies, which leads to a corresponding underestimation of the harmonic free energy by 19 meV. Nevertheless, after applying the TI correction, the total free energies, A^{MLFF}_{1}, obtained using DFT- and MLFF-derived references converge to nearly identical values of -106.481 ± 0.001 and -106.479 ± 0.001 eV, respectively. This confirms that, while the harmonic vibrational free energy can be mildly affected by MLFF training errors, the anharmonic free energy obtained through MLFF–TI remains robust and insensitive to the choice of reference.

4.1.8.3. Reliability of FEP calculations and phase space overlap

Having established the consistency of A_1^{MLFF} within a given MLFF, we next examine the reliability of the FEP method and the I_w index as an indicator of phase space overlap. To access this, two distinct MLFFs were trained for each of the three systems, namely Pt, O@Pt, and OH@Pt (details of the MLFFs are given in the Appendix), and subsequently employed for TI and FEP calculations. As anticipated, different MLFFs produced slightly varied total free energies, A_1^{MLFF}. Among them, the **MLFF400** models yielded values in closest agreement with DFT benchmarks (see Figure4.1 and Table 4.2). For the clean Pt surface, **MLFF1K–Pt** slightly overestimates the total free energy at -102.207 ± 0.001 eV, whereas **MLFF400-Pt** gives a more accurate value of -102.224 ± 0.001 eV Similarly, for the O@Pt system, **MLFF400-OPt** and **MLFF1K-OPt** predict free energies of -106.481 ± 0.001 and -106.488 ± 0.001 eV, respectively, differing by only 7 meV. For OH@Pt, **MLFF400-OHPt** reproduces the reference DFT value most accurately, giving a total free energy of approximately -100.620 ± 0.001 eV.

Table 4.2.: Free energies in eV of the O@Pt(111) system calculated using different MLFFs. The same reference were used for each system.

MLFF	A_1^{MLFF}	$\Delta A_{MLFF \rightarrow DFT}$	I_w	A_1^{DFT}
MLFF400-Pt	-102.224 (0.001)	0.002 (0.001)	0.355	-102.222 (0.002)
MLFF1K-Pt	-102.207 (0.001)	-0.012 (0.001)	0.345	-102.219 (0.002)
MLFF400-OPt	-106.481 (0.001)	-0.016 (0.001)	0.345	-106.497 (0.002)
MLFF1K-OPt	-106.488 (0.001)	-0.013 (0.001)	0.325	-106.501 (0.002)
MLFF400-OHPt	-110.620 (0.001)	0.001 (0.001)	0.265	-110.619 (0.003)
MLFF1K-OHPt	-110.600 (0.002)	-0.014 (0.001)	0.250	-110.613 (0.003)

Regardless of the specific MLFF employed, the total free energies corrected to the DFT level remain effectively identical after applying the FEP correction. The deviations among the predicted A_1^{DFT} values, ranging from 3 to 6 meV, fall well within the statistical uncertainty of the calculations.

The reliability of these FEP corrections arises from the strong overlap between the phase spaces sampled by the MLFF and DFT methods. This is reflected in the high I_w indices, which lie between 0.250 and 0.355 (Table 4.2). It is noteworthy that even an I_w value as low as 0.03 can already yield reliable perturbative corrections[239]. Moreover, the I_w indices obtained from the **MLFF400** calculations consistently exceed those from the **MLFF1K** models, as expected since **MLFF400** was specifically trained to capture the relevant regions of the phase space.

4.1.8.4. TI with pretrained universal MLFFs

As outlined in Section 4.1.2, the MLFF-TI approach provides a conceptual foundation for conducting free energy calculations with cost-efficient, pretrained universal MLFFs, minimizing the reliance on expensive electronic structure computations. To assess this

feasibility, we applied the method to the O@Pt system at 400 K using the `GemNet-OC-S2EFS-OC20+OC22` model[243]. The model is a graph neural network trained on the OC20[244] and OC22[249] datasets, which together contain tens of millions of DFT calculations performed at RPBE and PBE levels of theory. The MLFF-TI simulations were carried out in Python using ASE-based `HarmonicCalculator`, originally developed by Amsler[224] and further refined in this work, and the `OCPCalculator` from the `fairchem` package. For comparison, an additional MLFF was trained at the RPBE level and used to perform standard MLFF-TI calculations with VASP as a reference. The results are summarized in Table 4.3. The λ-TI calculations for both methods employed the same DFT harmonic reference. Therefore, we focus here on the free energy difference between that state and the RPBE-driven fully interacting state, corresponding to the term (**1**) + (**2**).

Table 4.3.: Comparison of MLFF-TI results obtained using the universal MLFF `GemNet-OC-S2EFS-OC20+OC22` and a tailored MLFF trained at the RPBE level. The MLFF results were corrected to RPBE accuracy using either brute-force FEP or MLPT. N_{config} denotes the number of configurations used for the correction.

	`GemNet-OC-S2EFS-OC20+OC22`		RPBE-MLFF
	brute-force FEP	MLPT	brute-force FEP
$\Delta A_{0\to1}^{\text{MLFF}}$ (eV) (**1**)	0.370	0.370	-0.116
$\Delta A_{\text{MLFF}\to\text{DFT}}$ (eV) (**2**)	-0.522	-0.485	-0.005
(**1**) + (**2**)	-0.152	-0.115	-0.121
σ^{FEP} (eV)	0.016	0.006	0.001
I_w	0.010	0.038	0.310
N_{config}	200	80000	200

The tailored RPBE-MLFF yields a highly accurate free energy difference of −0.116 eV, differing by only 0.005 eV from the DFT value of −0.121 eV. In contrast, the `GemNet-OC-S2EFS-OC20+OC22` universal MLFF markedly overestimates this difference, predicting 0.370 eV. Applying brute-force FEP to 200 configurations drastically reduces the free energy difference to −0.152 eV, yet a deviation of −0.031 eV from the DFT benchmark remains. This correction is not statistically reliable, as evidenced by a standard deviation of 0.016 eV and a low overlap indix of $I_w = 0.01$, indicating poor phase-space overlap between universal MLFF and the RPBE reference.

To address this limitation, we employed the MLPT method, described in Section 4.1.5. A Δ-ML model was trained using the same 200 configurations from the brute-force FEP dataset, following the same procedure described in Ref. [241]. This model was then used to predict the $\Delta\mathcal{V}$ values for every tenth configuration of the fully interacting trajectory obtained from the `GemNet-OC-S2EFS-OC20+OC22` calculations, excluding the first 2×10^5 steps for thermal equilibration. The MLPT approach yields a notable improvement, with the FEP uncertainty, σ^{FEP}, reduced to 0.006 eV and the overlap index I_w increased nearly fourfold. For $N_{\text{config}} = 80\,000$, an I_w value of 0.038 implies that over 3000 configurations contribute to the correction. Consequently, the final DFT free energy difference arrives at −0.115 eV, deviating by only 0.006 eV from the reference, consistent with the estimated MLPT uncertainty.

The results in this section demonstrate that DFT-accurate anharmonic free energies can be achieved from universal MLFFs using MLPT approaches, requiring only minimal additional DFT evaluations. It is worth noting that recent studies have shown that the Δ-ML architecture and training set selection strategy can be further optimized[238], potentially reducing the number of required DFT reference points to just a few dozen. This observation likely extends to the MLFF-IHSM method, since the corrections from MLFF to DFT accuracy remain the key step, and the MLPT approach retains its corrective power. Altogether, the robustness of the machine learning aided methods presented in this chapter open the door to practical and reliable free energy estimations using MLFFs, especially the universal *foundation* models[140, 141], which are rapidly emerging as an effective tool in computational catalysis.

4.2. Adsorption Free Energies of Oxygenated Species on Pt(111) Surface - Temperature Effects

4.2.1. Overview

While the thermodynamic properties of gas-phase molecular adsorption can be measured experimentally with high certainty[250], such data for adsorbed reaction intermediates remain scarce[251] and therefore often rely on computational characterization using periodic density functional theory (DFT)[252–254]. However, the accuracy of periodic DFT in predicting adsorbate free energies remains limited[117, 255], hindering its full potential as a quantitative tool for fundamental catalysis research and catalyst design. The free energies of adsorbed intermediates (and transition states) involved in the elementary reaction steps between reactants and products are the critical quantities that determine the rate and selectivity of one catalytic surface relative to another. Therefore, developing reliable methods for accurate computation of adsorption free energies is of utmost importance.

Improving the enthalpic contribution to the free energy often involves accounting for van der Waals interactions[256] and applying corrections to more accurate, yet computationally demanding, levels of theory such as Coupled Cluster (CC) or the Random Phase Approximation (RPA), either partially[257, 258] or fully[259]. In contrast, obtaining accurate entropic contributions remains challenging due to the inherent limitations of the commonly used harmonic approximation (HA). Section 1.5 presents state-of-the-art approaches for capturing anharmonic effects and thereby improving upon HA-based thermochemical estimates.

To efficiently incorporate anharmonic corrections beyond the HA, we introduce the machine-learned force field–aided λ-thermodynamic integration (MLFF-TI) workflow. By replacing computationally expensive DFT-based molecular dynamics (MD) simulations with MLFF-driven ones, the workflow enables extensive phase-space sampling at a fraction of the cost while maintaining DFT-level accuracy. This approach provides a robust, and reliable framework for computing anharmonic free energies in realistic catalytic systems.

In the following sections, we apply the MLFF-TI workflow to investigate the adsorption of progressively more complex oxygenated species, atomic oxygen (O), hydroxyl (OH), and hydroperoxyl (OOH), on Pt(111). These adsorption processes represent key elementary steps in heterogeneous catalytic reactions crucial for sustainable energy conversion, such as the electrocatalytic oxygen reduction reaction (ORR)[260–267] and CO oxidation[268]. Our results reveal distinct anharmonic behaviors for different adsorbates, highlighting the inadequacy of a single harmonic model for describing all intermediates along a reaction pathway. Furthermore, we examine the temperature dependence of adsorption free energies and entropies to elucidate how anharmonicity evolves with thermal excitation.

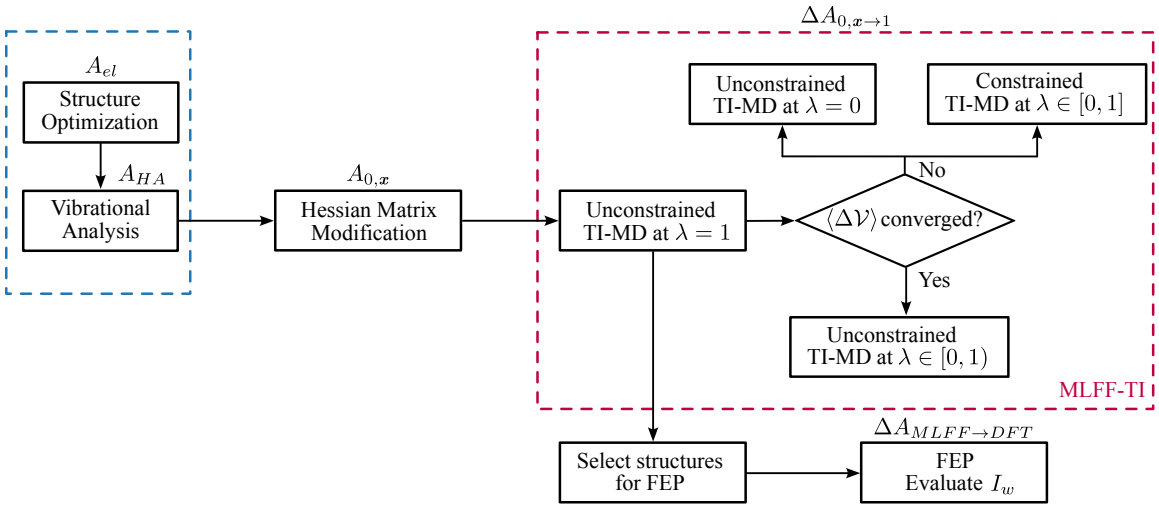
4.2.1.1. Simulation Details

Periodic DFT calculations were conducted using the Vienna Ab Initio Simulation Package (VASP), version 6.4.1[188–191], which includes a fully integrated implementation of thermodynamic integration (TI)[269]. The Kohn–Sham equations were solved using the projector augmented-wave (PAW) method of Blöchl, as adapted by Kresse and Joubert, within a plane-wave basis set employing standard PAW potentials. The BEEF-vdW functional[270] was used for both geometry optimizations and molecular dynamics (MD) simulations, with a plane-wave cutoff energy of 450 eV and a Monkhorst–Pack k-point grid of $3 \times 3 \times 1$ for Brillouin zone sampling. Electronic self-consistent field (SCF) cycles and structural relaxations were converged to 10^{-6} eV and 0.001 eV Å^{-1}, respectively. All calculations were non–spin-polarized.

A 3×3 four-layer slab was used to represent the Pt(111) surface, separated by a 14 Å vacuum gap to eliminate spurious self interactions. The supercell dimensions were fixed for all calculations. Adsorption free energies were evaluated at a coverage of one adsorbate per unit cell (1/9 monolayer), referenced to the ideal gas phase at 101 325 Pa for H_2 and H_2O.

Full Hessian matrices were obtained for all systems using a centered finite-difference scheme. For the λ-path TI simulations, MLFF-driven MD was performed in the NVT ensemble, using the Andersen thermostat[271] with a collision probability of 0.05, chosen for its simplicity and efficient thermalization. Following prior studies[160, 164], hydrogen atoms were modeled as tritium ($m_H = 3$ u) to allow for a larger time step of 0.5 fs. The same electronic-structure parameters were applied consistently in both geometry optimizations and MD simulations.

A schematic overview of the MLFF–TI workflow is illustrated in Scheme 4.3. The procedure begins with the optimized geometry at $\boldsymbol{x}0$ and the corresponding Hessian matrix, which provide the electronic and harmonic free energy contributions, A_{el} and A_{HA}, respectively. To construct the harmonic potential in Eq. (4.9), the Hessian can either be used directly or modified and transformed to improve the stability and convergence of the TI calculations, particularly near the harmonic limit ($\lambda \to 0$). For instance, soft vibrational modes can be regularized by increasing small eigenvalues below a chosen minimum threshold. The



Scheme 4.3.: The workflow for computing anharmonic corrections *via* MLFF-TI in VASP. The terms above some step denote the free energy that can be computed at that stage. The blue dashed box indicates the typical steps within harmonic oscillator approximation, while the red dashed box corresponds to the MLFF-aided TI procedure. Reprinted with permission from [225]. Copyright 2025 American Chemical Society.

harmonic reference free energy, $A_{0,\boldsymbol{x}}$, is then evaluated from the resulting (possibly modified) Hessian matrix, $\underline{\mathbf{H}}^{\mathbf{x}}$. Subsequently, the anharmonic correction, $A^{\mathrm{MLFF}}_{0,\boldsymbol{x}\to 1}$, is computed through either a constrained or unconstrained TI procedure. The selection between the two schemes depends on the convergence behavior of the average potential difference term, $\langle \mathcal{V}_1 - \mathcal{V}_0 \rangle$, obtained from an unconstrained trajectory at $\lambda = 1$. If this quantity exhibits good convergence, the remaining TI calculations at $\lambda < 1$ may proceed with unconstrained MD. Otherwise, anharmonic modes are identified, and appropriate geometric constraints are introduced to ensure stable sampling during TI-MD simulations. In addition, an unconstrained MD simulation at $\lambda = 0$ is performed to provide the probability distributions required for evaluating Eq.(4.11) or Eq.(4.15). Finally, a subset of uncorrelated configurations is extracted from the fully interacting trajectory at $\lambda = 1$ for DFT single-point energy evaluations, which are used to apply the FEP correction.

For simplicity, lattice expansion effects in periodic systems are neglected, allowing the Gibbs free energy (G) to be approximated by the Helmholtz free energy (A). Therefore, unless stated otherwise, all reported free energies are to be interpreted as Gibbs free energies.

4.2.2. Results and Discussion

In this section, we demonstrate the influence of anharmonic effects on the adsorption free energies of oxygenated adsorbates on Pt(111) surface. Specifically, in subsection 4.2.2.1, we present a detailed analysis on the anharmonic correction results obtained for four systems, namely, clean Pt, atomic oxygen at Pt (O@Pt), hydroxyl group at Pt (OH@Pt), and hydroperoxyl group at Pt (OOH@Pt) at 300 K. We compare the corrected Gibbs free energies for these adsorptions against the results from two HA approaches to highlight the

impact of anharmonicity in subsection 4.2.2.2. Subsection 4.2.2.3 discusses the robustness of the MLFF-TI workflow. We then examine the role of temperature on adsorption free energy and entropy in subsection 4.2.2.4. All results were obtained using the proposed MLFF-TI method (as introduced in the Methodology section). The results reported at 300 K in the first three subsections were obtained using **MLFF-400** set of MLFFs, while the **MLFF-1K** set was employed to achieve the results in the last subsection. All the employed MLFFs are shown to be excellent at reproducing the DFT results at the temperature of interest. Details of the training and evaluations can be found in Section A.1 of the Appendix.

4.2.2.1. Anharmonicity in the different systems at 300 K

Bare Pt(111) system We begin by examining the anharmonic contributions to the free energy of the pristine Pt(111) surface, which, together with the gas-phase molecules, serves as the reference for adsorption free energy calculations. In this system, the anharmonicity arises primarily from the *soft* vibrational modes of the surface atoms, characterized by frequencies below 200 cm^{-1} of all surface vibrations, as revealed by DFT vibrational analysis. The anharmonic behavior of the surface is visualized in Figure 4.2a,b, which present the partial radial distribution functions (RDFs) for two representative pairs of Pt surface atoms, obtained from 2.5 ns MD trajectories driven by harmonic and ML force fields. As shown in Figure 4.2a, the RDF corresponding to in-plane Pt–Pt vibrations exhibits only minor deviations between the harmonic and anharmonic simulations. In contrast, Figure 4.2b demonstrates a pronounced broadening of the interlayer (breathing) mode distribution, with a distinct tail extending toward larger interatomic distances in the anharmonic trajectory. This observation indicates that the out-of-plane breathing motions are the dominant contributors to the anharmonic effects observed on the clean Pt(111) surface.

Table 4.4.: Free energies (in eV) of the investigated systems at 300 K and pressure of 101 325 Pa, compared to the free energy obtained from harmonic approximation G_{HA}^{DFT}. $G_{0,\boldsymbol{x}}$: free energy of the chosen harmonic reference; G_1^{MLFF}: the system's anharmonic free energy (corrected by MLFF-TI); G_1^{DFT}: the system's anharmonic free energy corrected by free energy perturbation theory; ΔG_{anharm}^{DFT}: anharmonic correction at DFT level. The corresponding uncertainties are shown in the parentheses.

System	$G_{0,\boldsymbol{x}}$	G_1^{MLFF}	G_1^{DFT}	G_{HA}^{DFT}	ΔG_{anharm}^{DFT}
H_2O[a]	-	-	-13.116	-13.116	-
H_2[a]	-	-	-7.412	-7.412	-
Pt(111)	-102.135	-102.224(0.002)	-102.222(0.002)	-102.168	-0.054(0.002)
O@Pt(111)	-106.407	-106.481(0.001)	-106.497(0.002)	-106.441	-0.056(0.002)
OH@Pt(111)	-110.412	-110.662(0.001)	-110.661(0.003)	-110.518	-0.144(0.003)
OOH@Pt(111)	-113.124	-113.366(0.001)	-113.370(0.002)	-113.244	-0.126(0.002)

[a]Harmonic oscillator/ideal gas/linear rigid rotor approximation.

To evaluate the anharmonic contribution of the system, an unconstrained λ-TI procedure was carried out following Eqs.(4.8)–(4.10). Seven integration points were selected along the λ path (λ = 0.00, 0.25, 0.50, 0.75, 0.90, 0.95 and 1.00), and for each point, a 2.5 ns

MLFF-based MD simulation was performed at 300 K. This dense grid and the extended trajectory length ensured a smooth and numerically stable λ-integration path, yielding well-converged ensemble averages of $\langle \Delta \mathcal{V} \rangle_\lambda$ (Figure4.2c). The averaged potential energy differences decrease nonlinearly from approximately -0.03 to -0.27 eV, showing a gradual decline at small λ values followed by a sharper drop near $\lambda = 1$. As anticipated, the statistical uncertainties grow in the high-λ region, where anharmonic effects dominate. Integration of the resulting λ-dependent curve according to Eq.(4.8) yields an MLFF-level free energy change of $\Delta G_{0,x\to 1} = -0.089 \pm 0.002$ eV.

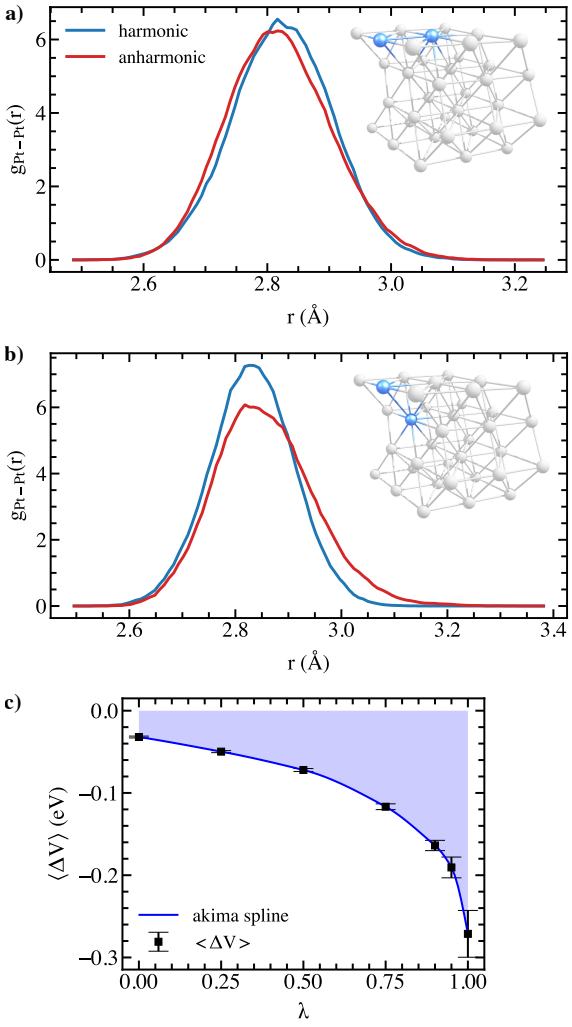


Figure 4.2.: Anharmonic effects in clean Pt(111) surface. Partial radial distribution functions ($g_{Pt-Pt}(r)$) for a pair of (a) in-plane and (b) interlayer surface Pt atoms (shown as blue spheres in the insets) showing the different behaviors of Pt surface in harmonic and anharmonic regimes. The results were computed over two corresponding 2.5 ns MD runs at 300 K driven by the potential harmonic in Cartesian coordinates and by MLFF potential; c) The λ-path representing averaged energy differences for the bare Pt(111) surface system obtained from unconstrained MLFF MD runs along seven λ points at 300 K. The obtained curve is smooth and integrable, over which the integration provides the anharmonic correction $\Delta A_{0\to 1}$. The blue solid curve, serving as visual guide, are the Akima splines connecting the computed points. Reprinted with permission from [225]. Copyright 2025 American Chemical Society.

We note that the harmonic reference system in this study was constructed by adjusting all Hessian eigenvalues below 1 eV Å^{-2} to that threshold, while keeping their corresponding eigenvectors unchanged to prevent numerical instabilities at low λ values (*cf.* Section2.1.2 in Ref.[160]). From the resulting modified Hessian, the vibrational harmonic free energy of the reference state, $G_{0,x}$, was determined to be −102.135 eV. After incorporating anharmonic corrections through MLFF-based TI, the total free energy at MLFF accuracy was refined to −102.224 ± 0.002 eV.

To obtain the corresponding DFT-level free energy, a forward FEP was performed from MLFF to DFT. This perturbation produced a small free energy correction of 0.002 ± 0.001 eV, with a high overlap index $I_w = 0.355$, confirming the excellent consistency between the phase spaces explored by MLFF and DFT. This high overlap is expected, as the phase space at 300 K is relatively simple and the trained MLFF reproduces DFT energies and forces with high fidelity (see RMSD values in Table A.1 in the Appendix). Consequently, the final DFT-accurate free energy, G_1^{DFT}, is −102.222 ± 0.002 eV.

When compared with the harmonic approximation value, $G_{HA}^{DFT} = -102.168$ eV, derived from the unmodified Hessian, the anharmonic effects lower the free energy by approximately 54 meV,corresponding to about 5% of the total HA vibrational free energy.

O@Pt(111) system The reference configuration for the O@Pt(111) system corresponds to an O atom adsorbed at the fcc hollow site, the most stable binding position, consistent with previous reports[272–275]. Upon adsorption, the three translational degrees of freedom of the O atom are transformed into vibrational modes with moderate frequencies (361, 363, and 415 cm^{-1}), indicating that the HA adequately captures the adsorbate's motion. Indeed, as shown in Figure 4.3a, the dynamics of the O atom in the fully interacting regime closely follow the harmonic behavior, with the partial RDFs of the O–Pt distances being nearly identical in both regimes. This confirms that the O adsorbate behaves predominantly harmonically at 300 K. Conversely, the anharmonic RDFs shown in Figure 4.3b highlight the persistent anharmonic effects in the interlayer vibrational modes of the surface atoms, analogous to those found in the clean Pt surface. Thus, the anharmonic contributions in this system are expected to originate mainly from the surface vibrations rather than from the adsorbate itself.

Similarly to the previous system, an unconstrained λ-TI calculation was carried out, yielding a smooth and well-behaved $\langle \Delta \mathcal{V} \rangle_\lambda$ curve that is readily integrable, as shown in Figure 4.3b. As before, due to the imposed minimal eigenvalue limit of 1 eV Å^2 in the Hessian matrix, the modified harmonic reference gives a harmonic free energy $G_{0,x}$ of −106.407 eV, compared to the value for the unmodified reference G_{HA}^{DFT} of −106.441 eV. Integration along the λ-path produces a $\Delta G_{0,x\to 1}$ of −0.081(1) eV, decreasing to the free energy of the system fully interacting at the MLFF level to −106.481(1) eV.

The forward FEP correction introduced a further correction of −0.016(1) eV, lowering the total free energy to −106.497(2) eV. The larger magnitude of this correction compared to the clean surface system originates from the higher RMSD in the MLFF energies (0.018 eV versus 0.010 eV, see Table A.1 and A.2 in the Appendix), though the correction still remains

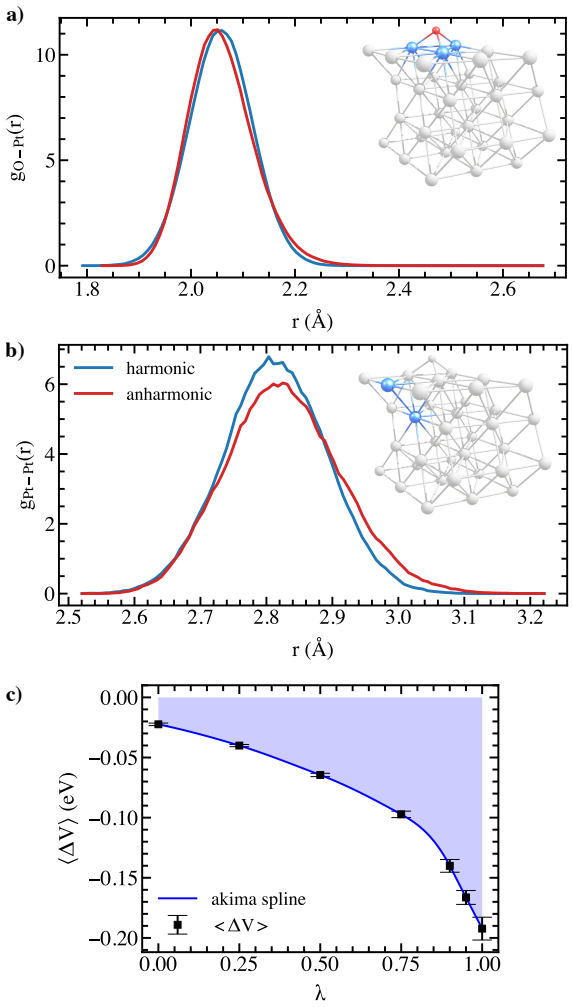


Figure 4.3.: Anharmonic effects in O@Pt(111) system. a) The partial RDFs ($g_{O-Pt}(r)$ and $g_{Pt-Pt}(r)$) for the O atom and the three adjacent Pt atoms (in the inset shown as red and blue spheres, respectively) showing virtually identical behaviors of O adsorbate on Pt(111) surface. The results were obtained using two MD simulations with similar settings used to construct Figure 4.2a; b) The partial RDFs for a pair of interlayer surface Pt atoms (as blue spheres in the inset) showing the anharmonic character of the associated breathing modes. c) The λ-path representing averaged energy differences for the O@Pt(111) system obtained from MLFF MD runs along seven λ points at 300 K. Reprinted with permission from [225]. Copyright 2025 American Chemical Society.

within the corresponding RMSD range. Importantly, despite this slightly larger deviation, the high I_w index of 0.345 confirms strong phase space overlap, ensuring the reliability of the perturbation. As demonstrated in Section 4.1.8, while such deviations in free energy can be further minimized by improving the MLFF accuracy, the final corrected values remain effectively unchanged as long as the phase space overlap is sufficiently large (i.e., I_w is high).

Finally, the anharmonic correction to the free energy for this system is estimated to be approximately −56 meV, which is almost identical to that of the clean Pt surface. This

similarity confirms that the primary source of anharmonicity originates from surface vibrations, as previously discussed. Overall, the anharmonic correction accounts for about 6% of the total vibrational free energy contribution.

OH@Pt(111) system Figure 4.4a illustrates the atop binding configuration of the OH adsorbate on Pt(111), which was used as the reference structure for the TI calculations of this system. It is worth noting that previous studies have reported conflicting results regarding whether the most stable adsorption site of OH on Pt(111) is the bridge or atop position[261, 273, 275]. Our BEEF-vdW calculations indicate that the hydroxyl species preferentially adsorbs at the atop site, which is 0.03 eV more stable than the bridge configuration. Nevertheless, the choice of the harmonic reference structure does not influence the final computed free energy of the system.

A vibrational analysis of OH adsorbate on Pt surface provides partial insight into its pronounced anharmonic nature. Among the six degrees of freedom of the adsorbate, three correspond to vibrational modes with moderate to high frequencies, namely, out-of-plane hindered translation, out-of-plane hindered rotation, and OH stretching vibration at 472, 594, and 2265 cm^{-1}, respectively. In contrast, the remaining in-plane modes, including one hindered rotation and two hindered translations, are characterized by soft vibrations with frequencies of 59, 61, and 108 cm^{-1}. These low-frequency modes are expected to be a source of anharmonicity[158], in addition to the surface motions, as discussed in the preceding sections.

The anharmonic character of the system is further evidenced by the distinct behaviors of the OH adsorbate in the harmonic and fully interacting regimes. Under the harmonic force field, the OH species remains rigidly bound to a single Pt atom corresponding to the reference configuration. Conversely, in the fully interacting regime, the adsorbate becomes highly mobile across the Pt surface at 300 K, as demonstrated by the O-atom heatmap in Figure 4.4b. The OH species frequently hops between adjacent Pt atoms via bridge sites, which indicates a considerable increase in entropy. Interestingly, despite the relatively similar energies of the atop and bridge adsorption sites, the bridge positions are visited far less frequently. Additionally, a "helicopter-like" rotational motion of the OH group about the Pt–O axis is observed (Figure 4.4g).

The hopping and rotational motions of the OH adsorbate poses a practical problem for applying unconstrained TI method in Cartesian coordinates. These large displacements from the reference configuration cause significant fluctuations in the harmonic potential energy $\mathcal{V}_{0,\boldsymbol{x}}$, which in turn results in a very slow convergence of the $\langle \mathcal{V}_1 - \mathcal{V}_{0,\boldsymbol{x}} \rangle$ term at λ values approaching 1. To overcome this issue, we demonstrate the use of the constrained TI method, in which, two holonomic constraints were imposed to suppress the problematic motions, and their contributions to the free energy was subsequently recovered through histogram analyses. Specifically, the hopping motion of the adsorbate was constrained by fixing the distance between the O atom and the Pt atom to which the OH is bonded ($\xi_1 = R_{\mathrm{O-Pt}}$, Figure 4.4d) at its equilibrium value, $\xi_1' = 2.002$ Å. In addition, a second internal coordinate, defined as a symmetric linear combination of three dihedral angles

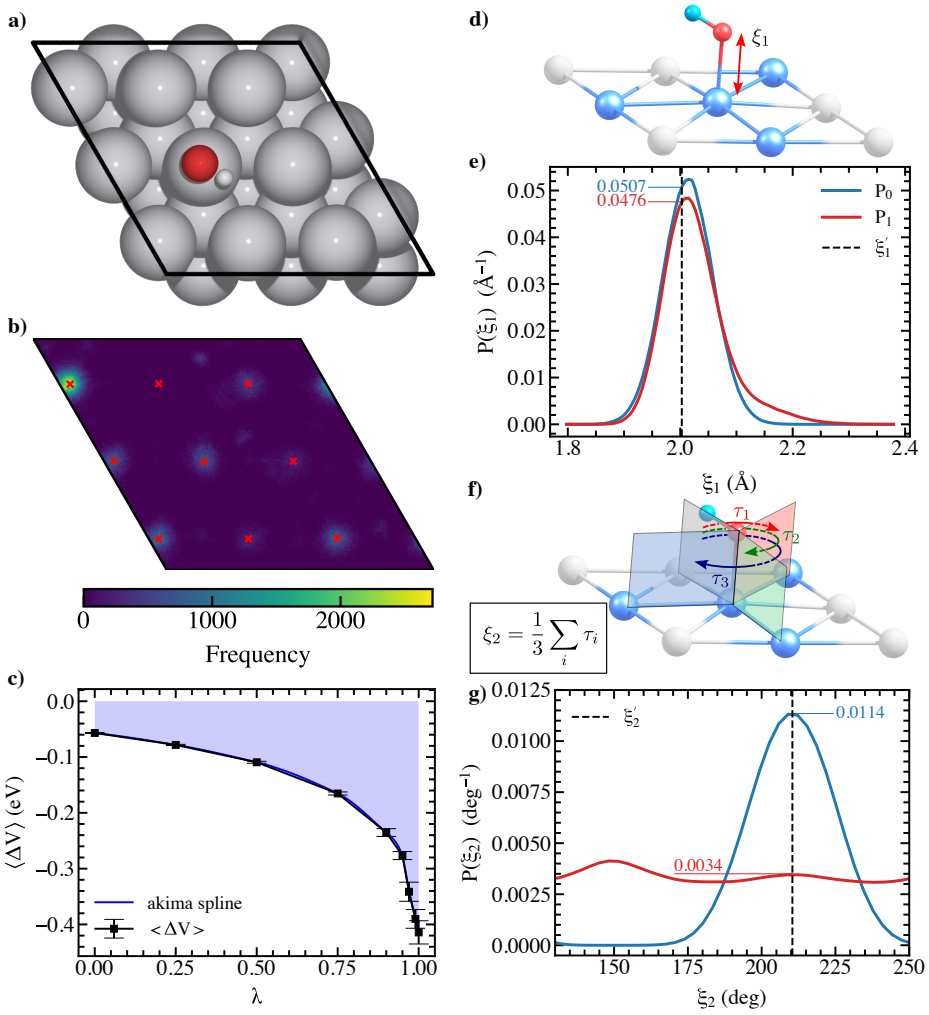


Figure 4.4.: Anharmonic effects in OH@Pt(111) system. a) The reference structure for the OH@Pt(111) system; b) Heatmap showing the mobility of OH adsorbate on Pt(111) surface at 300 K (red crosses depicting the equilibrium positions of the top most Pt atoms). The result is computed over a 5 ns unconstrained MLFF MD trajectories; c) The λ-path representing averaged energy differences for the OH@Pt(111) system obtained from *constrained* MLFF MD runs along nine λ points at 300 K. d-g) The schemes demonstrating the two constraints imposed on the OH@Pt(111) system for constrained TI calculations: the distance between the O atom and its nearest Pt atom, ξ_1, (d) and the linear combination of three torsional angles, ξ_2, (f) and their corresponding probability density distributions of ξ_1 (e) and ξ_2 (g), respectively, for the unconstrained systems driven by harmonic force field (blue solid) and by MLFF (red solid). The dashed lines mark the positions of reference state.

formed by the OH group and neighboring surface Pt atoms ($\xi_2 = \frac{1}{3}(\tau_1 + \tau_2 + \tau_3)$, Figure 4.4f), was constrained at $\xi_2' = 210.4°$ to effectively eliminate the self-rotation of the OH group. The choice of the latter descriptor was inspired by earlier TI studies of ethane hindered rotation and the substitution reaction of CH_3Cl with Cl^- anion[160, 164].

MLFF MD simulations incorporating these two constraints were conducted for nine λ points, with two additional high-λ values (0.97 and 0.99) added to better capture the sharp decline in $\langle \Delta \mathcal{V} \rangle_\lambda$ near the fully interacting limit. Each trajectory was propagated for 2.5 ns.

Furthermore, two unconstrained MLFF MD simulations of 5 ns each were performed to evaluate the probability distributions $P(\xi_1)$ and $P(\xi_2)$.

Table 4.5.: Free energy components contributing to the free energy differences $\Delta G_{0,\boldsymbol{x}\to 1}$ in eV computed by the constrained TI method at 300 K at MLFF level for OH@Pt(111) and OOH@Pt(111) systems. The corresponding uncertainties are shown in the parentheses. See text for the definitions of ξ_i.

System	$\Delta G_{0,\boldsymbol{x}\to 1}(\xi')$	$-k_\mathrm{B}T \ln \frac{P_{0,\boldsymbol{x}}(\xi_i)}{P_1(\xi_i)}$		$\Delta G_{0,\boldsymbol{x}\to 1}(\xi)$
		ξ_1	ξ_2	
OH@Pt	-0.132(0.001)	-0.058(0)	-0.060(0)	-0.250(0.001)
OOH@Pt	-0.144(0.001)	-0.070(0)	-0.028(0)	-0.242(0.001)

The individual free energy contributions to the total $\Delta G_{0,\boldsymbol{x}\to 1}$ are summarized in Table 4.5. By applying the constrained TI approach, a smooth and readily integrable λ-path was achieved (Figure 4.4c). Integration along this curve yields a free energy difference between the two constrained systems, $\Delta G_{0,\boldsymbol{x}\to 1}(\xi')$, of -0.132 ± 0.001 eV. As illustrated in Figure 4.4b, the OH adsorbate visits all surface Pt atoms during the simulation, suggesting that over sufficiently long timescales, the probability distribution of $R_{\mathrm{O-Pt}}$ is identical on every surface atoms. Consequently, a permutation factor of 1/9 was applied to the probability of the reference state in the fully interacting system, meaning that the probability of finding the adsorbate bound to the central Pt atom at $R_{\mathrm{O-Pt}} = 2.002$ Å is one-ninth of $P_1(\xi_1')$. The probability distributions of $R_{\mathrm{O-Pt}}$, obtained from 5 ns unconstrained MD trajectories driven by harmonic force field and MLFF, are shown in Figure 4.4e. At the equilibrium distance, $P_1(\xi_1') \approx 0.0476\text{Å}^{-1}$ and $P_0(\xi_1') \approx 0.0507\text{Å}^{-1}$. The contribution from this degree of free is then computed as $-k_\mathrm{B}T \ln\left(\frac{P_0(\xi_1')}{(1/9)P_1(\xi_1')}\right) = -0.058$ eV.

Figure 4.4g presents the probability distributions for the second constrained coordinate, ξ_2. At the anharmonic state, $P_1(\xi_2)$ reveals frequent rotational motion of the OH group around the axis perpendicular to the surface, characterized by two distinct peaks separated by approximately 120°. Although DFT predicts the reference minimum at ξ_2 =210.4° (Figure4.4a) to be 2 meV more stable, the MLFF trajectory samples the alternative minimum more often. This minor discrepancy is acceptable, as the corresponding energy difference is much smaller than the MLFF's RMSD in energy. Moreover, this small bias does not significantly affect the free energy, and it can be effectively corrected via FEP. The probability densities at the reference configuration are $P_1(\xi_2') = 0.0034\text{deg}^{-1}$ and $P_0(\xi_2') = 0.0114\text{deg}^{-1}$. Considering the threefold rotational symmetry of the OH adsorbate, $P_1(\xi_2')$ is multiplied by a factor of 1/3, contributing to the free energy an amount of approximately -0.060 eV, comparable to the translational component.

Summing these components yields a total $\Delta G_{0,\boldsymbol{x}\to 1}$ of $-0.250(1)$ eV (Eq.(4.15)). Combining this with the harmonic reference contribution, $G_{0,\boldsymbol{x}} = -110.412$eV, results in a total MLFF free energy, $G_1^{\mathrm{MLFF}} = -110.662 \pm 0.001$eV. The subsequent FEP correction introduces only a negligible shift of 0.001 ± 0.002eV, giving a DFT-level total free energy of $G_1^{\mathrm{DFT}} = -110.661 \pm 0.003$eV. The small FEP correction and relatively high I_w index of 0.265 confirm the strong phase space overlap between MLFF and DFT, validating the reliability of the

MLFF predictions. The resulting anharmonic correction amounts to −143 meV, accounting for over 14% of the vibrational free energy estimated from HA.

OOH@Pt(111) system The reference configuration for the OOH@Pt(111) system, corresponding to the end-on adsorption geometry, is illustrated in Figure 4.5a. This structure represents the most stable adsorption mode and is consistent with previous reports in the literature[273, 276, 277]. Among the vibrational modes of the adsorbate, only three exhibit low frequencies: two hindered in-plane translational modes at 41 cm^{-1} and 49 cm^{-1}, and one hindered in-plane rotational mode around the axis perpendicular to the Pt surface passing through the hydrogenated O atom at 103 cm^{-1}. These soft modes are expected to display strong anharmonic behavior and to make substantial contributions to the overall anharmonic correction.

An initial MLFF-MD trajectory confirms that the adsorption of OOH species on the Pt surface exhibits pronounced anharmonic behavior. The anharmonicity is primarily manifested in the rotational motion of the OOH group around the Pt–O bond (as illustrated in Figure 4.5b) and in the flipping motion of the OH fragment. These motions complicate convergence in the TI procedure due to the inadequate representation of rotational degrees of freedom in Cartesian coordinates, as discussed previously. Interestingly, no hopping events of the adsorbate were observed along the trajectory, contrasting with what could be inferred from the vibrational analysis.

To address these convergence issues, the constrained TI method once again proved advantageous. Specifically, a linear combination of three torsional angles, similar to the approach employed in the preceding section, was used to constrain the rotation of the OOH species, while the flipping motion of the OH group was controlled via the Pt–O–O–H dihedral angle. The constraint coordinates were fixed at $\xi_1 = 186.9°$ and $\xi_2 = 93.6°$, as shown in Figure 4.5d,f.

As with the OH@Pt system, constrained TI calculations were performed using a nine-point λ grid, resulting in a smooth and well-converged λ-path for the $\langle \cdots \rangle_\lambda$ terms (Figure 4.5c). Integration along this path yielded a free energy difference $\Delta G_{0,x\to1}(\xi') = -0.144 \pm 0.001$ eV. The probability distributions for the two rotational degrees of freedom (Figure 4.5e,g) further reveal their strong anharmonic characteristics. In particular, the nearly six-fold symmetry of the OOH in-plane rotation on Pt(111) is reflected by two peaks in the P_1 distribution. This rotation occurs even more frequently than in the OH system, as evidenced by the broader distribution at the fully interacting state and the lower probability density near the reference configuration. Applying a symmetry factor of 1/3, the corresponding free energy contribution from this degree of freedom is −0.070 eV. Additionally, the two peaks separated by approximately 180° in Figure 4.5g confirm the rotation of the OH group around the O–O bond. Here, $P_1(\xi_2') = 0.0309\,deg^{-1}$ and $P_0(\xi_2') = 0.0923\,deg^{-1}$, leading to a contribution of −0.028 eV to the overall anharmonic correction. Combining these contributions gives a total $\Delta G_{0,x\to1}$ of −0.242 ± 0.001 eV.

Adding the harmonic free energy term $\Delta G_{0,x} = -113.124$ eV results in $G_1^{MLFF} = -113.366 \pm 0.001$ eV. The FEP correction contributes an additional −0.004±0.001 eV, giving a final value

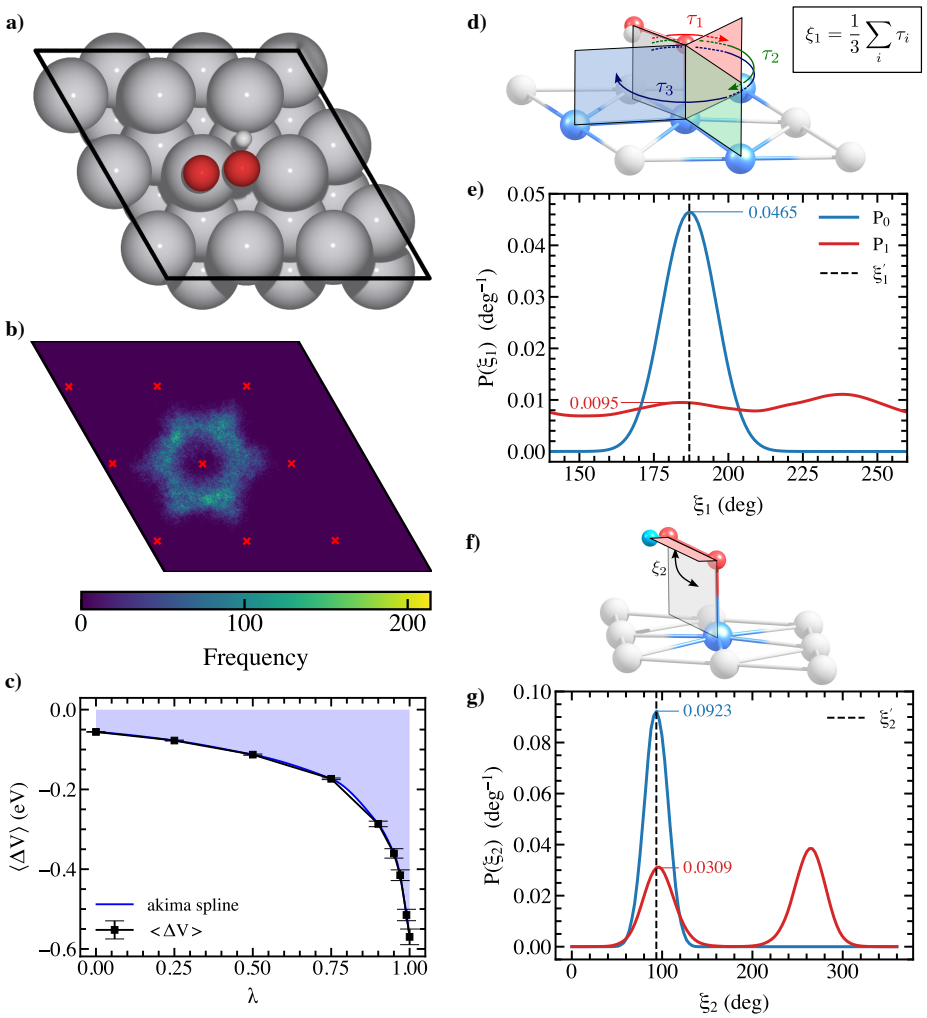


Figure 4.5.: Anharmonic effects in OOH@Pt(111) system. a) The reference structure for the OOH@Pt(111) system; b) Heatmap illustrating the rotation of OOH adsorbate on Pt(111) surface at 300 K (red crosses depicting the equilibrium positions of the top most Pt atoms). The result is computed over a 5 ns unconstrained MLFF MD trajectories; c) The λ-path representing averaged energy differences for the OOH@Pt(111) system obtained from *constrained* MLFF MD runs along nine λ points at 300 K. d, f) The scheme demonstrating the two constraints imposed on the OOH@Pt(111) system for constrained TI calculations. d: the linear combination of three torsional angles (ξ_1) and f: the torsional angle of Pt-O-O-H (ξ_2); e, g) Probability distributions of ξ_1 and ξ_2, respectively, for the unconstrained systems driven by harmonic force field (blue solid) and by MLFF (red solid). The dashed lines mark the positions of reference state. Reprinted with permission from [225]. Copyright 2025 American Chemical Society.

of $G_1^{DFT} = -113.370 \pm 0.002$ eV. The I_w index for this perturbation is reasonably high (0.27), confirming the reliability of the FEP correction. The total anharmonic correction amounts to −126 meV, representing more than 12% of the vibrational free energy contribution estimated within the HA approximation.

4.2.2.2. Free energy of adsorption and role of anharmonic effects

With the accurate anharmonic free energies determined, we next investigate how anharmonicity influences the adsorption of various oxygenated species on the Pt(111) surface. To assess their adsorption free energies, we examine in this sub-section the following reactions, which are central to the ORR mechanism:

$$H_2O + Pt^* \rightarrow O^*Pt + H_2 \tag{4.48}$$

$$H_2O + Pt^* \rightarrow OH^*Pt + \frac{1}{2}H_2 \tag{4.49}$$

$$2H_2O + Pt^* \rightarrow OOH^*Pt + \frac{3}{2}H_2. \tag{4.50}$$

Assuming that the pV terms for the adsorbate and substrate systems effectively cancel, the Gibbs free energies of adsorption can be calculated following Ref. [160]:

$$\Delta G_1 = G_{1,A@S} + aG_{0,H_2} - G_{1,S} - bG_{0,H_2O}, \tag{4.51}$$

where $G_{1,A@S}$ and $G_{1,S}$ represent the anharmonic Gibbs free energies of the adsorbed and clean surface systems, respectively, while G_{0,H_2} and G_{0,H_2O} correspond to the Gibbs free energies of the gas-phase H_2 and H_2O molecules. These gas-phase energies are evaluated using the ideal gas, rigid rotor, and harmonic oscillator approximations, and are weighted by the stoichiometric coefficients a and b. The resulting anharmonic adsorption free energies are compared to their harmonic counterparts in Table 4.6. It is important to note that the anharmonic correction to the adsorption free energy, $\Delta\Delta G_{\text{anharm}}$, originates from the difference between the individual anharmonic corrections, $\Delta G^{\text{DFT}}_{\text{anharm}}$, of the adsorbed and clean surface systems (as reported in Table (4.2)), as illustrated in Figure 4.6.

Table 4.6.: Comparison of the enthalpy, entropic contribution and free energy of adsorptions in eV for three oxygenated species on Pt(111) at 300 K estimated by classical HA and TI methods. Additional data from partial harmonic approximation are given in the parentheses.

	ΔH	$-T\Delta S^{HA}$	$-T\Delta S^{TI}$	ΔG^{HA}	ΔG^{TI}	$\Delta\Delta G_{anharm}$	$\Delta\Delta S$ (%)
O	1.350	0.081	0.079	1.431 (1.450)	1.429 ± 0.004	-0.002	2.7
OH	0.874	0.229	0.139	1.103 (1.118)	1.013 ± 0.005	-0.090	39.3
OOH	3.770	0.311	0.239	4.081 (4.092)	4.009 ± 0.005	-0.072	23.1

For the adsorption of atomic O on Pt, the TI adsorption free energy is 1.429±0.004 eV, which is, within statistical uncertainty, indistinguishable from the full harmonic approximation (FHA) result. This close agreement arises from the mutual cancellation of anharmonic contributions between the adsorbed and clean surface systems. A difference of −21 meV is observed between the TI and partial harmonic approximation (PHA) results, in which the surface atoms are kept fixed. This discrepancy originates from the differing vibrational contributions of surface atoms in the clean and adsorbed configurations, a trend already evident in the FHA comparison. The obtained correction is in reasonable agreement with the reported value of approximately −29 meV at 300 K by Jørgensen and Grönbeck [153] for

O adsorption on Pt(111) at 0.25 monolayer coverage. It is worth noting that although their study primarily targeted the translational free energy of the O adsorbate, contributions from Pt surface atoms were implicitly included, since their z-positions were allowed to relax during sampling.

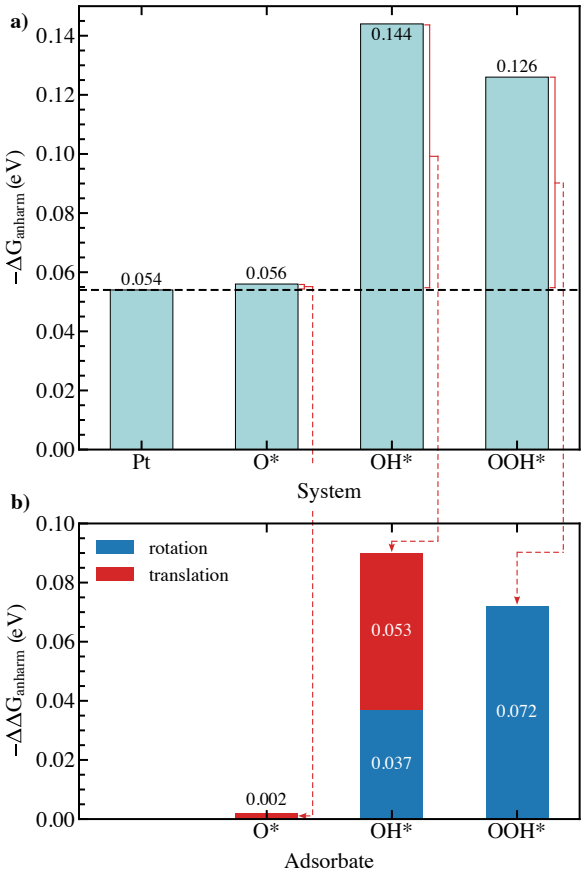


Figure 4.6.: a) The anharmonic corrections to free energy of the investigated systems; b) the resulting anharmonic corrections to the free energy of adsorption for the three adsorbates and the contributions of anharmonic modes. The black dashed horizontal line presents the anharmonic correction to free energy of Pt system. Reprinted with permission from [225]. Copyright 2025 American Chemical Society.

The adsorption free energy of OH obtained from TI is calculated to be 1.013 ± 0.005 eV. Anharmonic effects reduce this value by 90 and 105 meV relative to the predictions from FHA and PHA, respectively. This substantial correction arises from the pronounced anharmonicity of the adsorbed system, primarily due to the increased entropy associated with the hindered translational and rotational motions of the adsorbate (*i.e.*, a smaller entropy loss upon adsorption), with the translational modes contributing most significantly (Figure 4.6b). The anharmonic correction derived from TI closely matches that predicted by the two-dimensional hindered-translator one-dimensional free-rotor model for the adsorbate (see Section A.3 of the Appendix). Overall, the increased entropy accounts for nearly 39% of the total entropic contribution to the adsorption free energy, meaning that the TI calculations recover more than 39% of the entropy loss that is otherwise overestimated by HA.

Similarly, the Gibbs free energy of adsorption for the OOH species obtained via TI is 4.009 ± 0.005 eV, which is lower by 72 and 83 meV compared to the FHA and PHA results, respectively. As previously discussed, the anharmonic contributions originate mainly from the in-plane rotation of the OOH moiety and the flipping motion of the OH fragment. The entropy recovered for this adsorption corresponds to roughly 23% of the entropy loss predicted by HA.

4.2.2.3. Discussion on the MLFF-TI results at 300 K

The MLFF-TI workflow has been successfully applied to correct anharmonic contributions to adsorption free energies on transition metal surfaces. The examined adsorbates demonstrate progressively increasing complexity in their degrees of freedom, revealing a clear correlation between anharmonicity and adsorbate size. While the adsorption of monatomic O exhibits negligible anharmonic corrections, these effects become increasingly considerable for the diatomic OH and triatomic OOH species. This trend arises from the greater number of anharmonic motions with increasing molecular size, such as translational and rotational modes for OH, and multiple internal rotations for OOH. Consequently, we infer that larger adsorbates, including hydrocarbons and their derivatives, are likely to display even stronger anharmonic behavior due to, for example, extensive internal methyl rotations. As the significance of internal motions grows, simplified analytical models (e.g., hindered rotors) that treat adsorbates as rigid entities become insufficient. In such cases, the MLFF-TI approach becomes indispensable for accurately capturing anharmonic effects and determining reliable adsorption free energies.

These results also highlight the importance of accounting for anharmonicity even at relatively low temperatures, such as 300 K, emphasizing the need for a robust and broadly applicable computational framework. The MLFF-TI workflow provides such a framework, combining efficiency and accuracy while allowing flexible implementation through its treatment of coordinate systems, harmonic references, and the use of FEP to mitigate MLFF-related systematic errors.

In surface modeling, it is standard practice to constrain the lower layers of the slab to represent the bulk. Under this setup, Cartesian harmonic potentials (Eq. (4.9)) remain appropriate, as the system as a whole does not undergo translations or rotations. However, soft adsorbate motions, such as internal rotations and surface diffusion, can lead to poor convergence. We have demonstrated that applying MLFFs together with a few simple holonomic constraints, typically expressed as internal coordinates, effectively stabilizes these calculations while keeping the coordinate system simple.

A key strength of the TI formalism is that the final corrected free energy is independent of the chosen harmonic reference configuration[160]. As shown in subsection 4.1.8.1, the total MLFF-level free energy, A_1^{MLFF}, depends solely on the MLFF-predicted PES, regardless of variations in the minimum eigenvalue limit of the Hessian matrix. Moreover, subsection 4.1.8.2 demonstrates that the same accuracy is retained even when using an MLFF-generated reference, eliminating the need for costly DFT-based optimizations

and finite-difference calculations, which is a particularly valuable advantage for large systems.

While different MLFFs can produce slightly varying free energy values, the FEP method and its diagnostic I_w index serve as essential tools for achieving DFT-level accuracy. As discussed in subsection 4.1.8.3, the application of FEP ensures consistency between MLFFs trained on different phase space regions, as evidenced by high I_w values indicating strong overlap with DFT. Interestingly, these results suggest that extremely precise MLFFs are not always necessary, especially for systems with flat PESs, since phase space overlap, rather than raw fitting accuracy, is the critical factor. Furthermore, as demonstrated in subsection 4.1.8.4, DFT-accurate anharmonic free energies can be achieved even from a universal pretrained MLFF (a so-called *foundation model*). When overlap between the universal MLFF and DFT phase spaces is insufficient, it can be effectively improved through the machine learning thermodynamic perturbation theory (MLPT) method, described in Section 4.1.5.

In summary, the MLFF-TI framework provides a robust, efficient, and conceptually simple means to account for anharmonic effects in free energy evaluations. With the growing accessibility and popularity of MLFFs, particularly universal models, this approach is expected to become a routine practice for obtaining accurate adsorption free energies and achieving better alignment with experimental observations.

4.2.2.4. Temperature effects on adsorption free energy and entropy

Subsection 4.2.2.1 established the MLFF-TI procedure for calculating DFT-accurate anharmonic free energies for the adsorption of oxygenated species on Pt(111) and evaluated the anharmonic contributions at 300 K. In that analysis, the adsorbates were assumed to be highly localized at the reference configurations. Consequently, the more flexible OH and OOH species exhibited notably strong anharmonic contributions compared to the HA results. As illustrated in Figure 4.6, a substantial portion of the free energy (and entropy) recovered for OH adsorption originates from differences in the treatment of translational motion between localized and mobile regimes. In the latter case, the translational entropy of the adsorbate depends explicitly on the accessible surface area.

In the following subsection, we extend our analysis to examine the temperature dependence of adsorption free energies and entropies for these adsorbates. To enable consistent comparison across temperatures, we adopt a new harmonic reference, whose partition function differs from that of the localized model by a multiplicative factor M ($M = 9$ in this study), denoting the number of equivalent adsorption sites:

$$Q_{\mathrm{HA}} = M \prod_{i=1}^{N_{vib}} \frac{k_{\mathrm{B}}T}{\hbar\omega_i} \tag{4.52}$$

and the harmonic free energy of this reference becomes:

$$A_{0,\boldsymbol{x}} = A_{el}(\boldsymbol{x}_0) - k_{\mathrm{B}}T \sum_{i=1}^{N_{vib}} \ln \frac{k_{\mathrm{B}}T}{\hbar\omega_i} - k_{\mathrm{B}}T \ln(M)\,. \tag{4.53}$$

This formulation assumes that all adsorption sites are equally probable, implying entropy as well as free energy of adsorption is a function of surface area. The adoption of this reference is motivated by two main considerations. First, in most industrially relevant processes, the TOF typically ranges from 1×10^{-2} to 1×10^{2} s^{-1}, corresponding to effective free energy barriers of 0.6 –0.9 eV at 298.15 K[5]. Diffusion barriers for adsorbates are generally smaller, for instance, 0.56 eV for atomic O on Pt(111), and thus diffusion events are expected to occur frequently before the TOF-determining step. As shown in Section A.4 of the Appendix, the O adsorbate can explore several adsorption sites within 1 ms even at 300 K. Second, the partition function in Eq. (4.52) represents the low-temperature limit of the hindered translator model[146], which interpolates between the 2D gas and HA limits. Using this formulation ensures a consistent baseline for comparing results obtained from different models. Adopting this HA reference, our anharmonic corrections are equivalent to those per periodic primitive unit cell area in Ref. [156–158], or per adsorption site in Ref. [153, 155].

An important implication of adopting this reference is that the 1/9 permutation factor previously applied to the probability of R_{O-Pt} for the anharmonic OH@Pt system now cancels out, since it applies equally to both harmonic and anharmonic probabilities.

O adsorption on Pt(111) Figure 4.7 shows the temperature-dependent anharmonicity of systems involved in atomic oxygen adsorption, covering a temperature range from 200 K to 800 K. With increasing temperature, the adsorbed O atom becomes progressively more mobile, as reflected by the broadened partial RDFs at elevated temperatures. At high temperatures, the RDFs exhibit extended tails up to 3.8 Å, signifying frequent hopping events of the O atom between neighboring adsorption sites. Indeed, as shown in Figure A.2 in the Appendix, such hopping is explicitly observed in a 2 ns MD trajectory already at 600 K, in agreement with predictions from transition state theory predictions for diffusion rates.

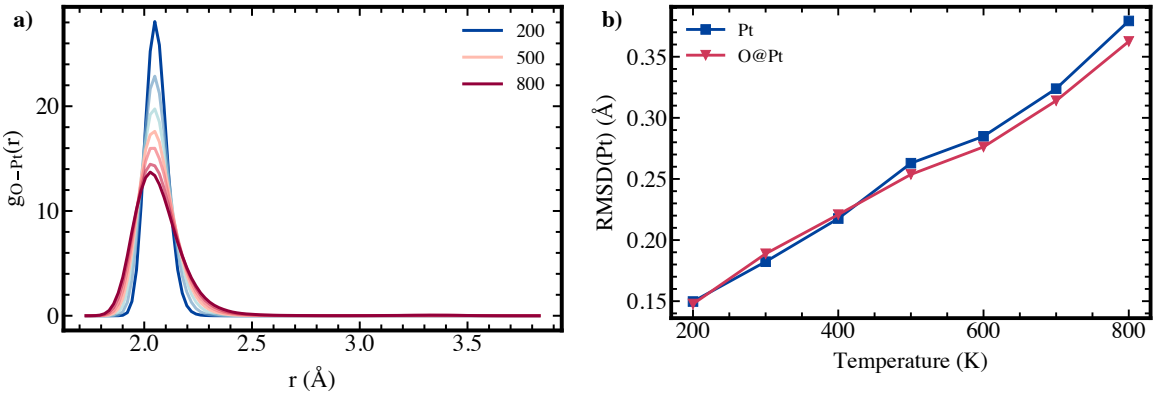


Figure 4.7.: a) Partial RDFs ($g_{O-Pt}(r)$) for the adsorbed O and three adjacent Pt atoms with increasing temperatures, demonstrating its increased anharmonicity with temperature. b) Root mean squared deviations (RMSDs) in atomic positions of surface Pt atoms in the clean Pt(111) and O@Pt(111) system as function of increasing temperature.

Interestingly, while the adsorbed O atom exhibits pronounced anharmonicity with increasing temperature, the surface Pt atoms in the O@Pt(111) system exhibit weaker anharmonic behavior than those on the clean surface. This trend is evidenced by the smaller RMSDs of Pt atomic positions from their equilibrium configurations, particularly above 500 K (Figure 4.7b). The reduced surface mobility can be attributed to the strong O–Pt interaction, which stabilizes the local lattice and suppresses thermal fluctuations of the surface atoms, even at the high-temperature regime where the clean surface becomes markedly more dynamic.

$$\frac{1}{2}O_2 + Pt^* \rightarrow O^*Pt \tag{4.54}$$

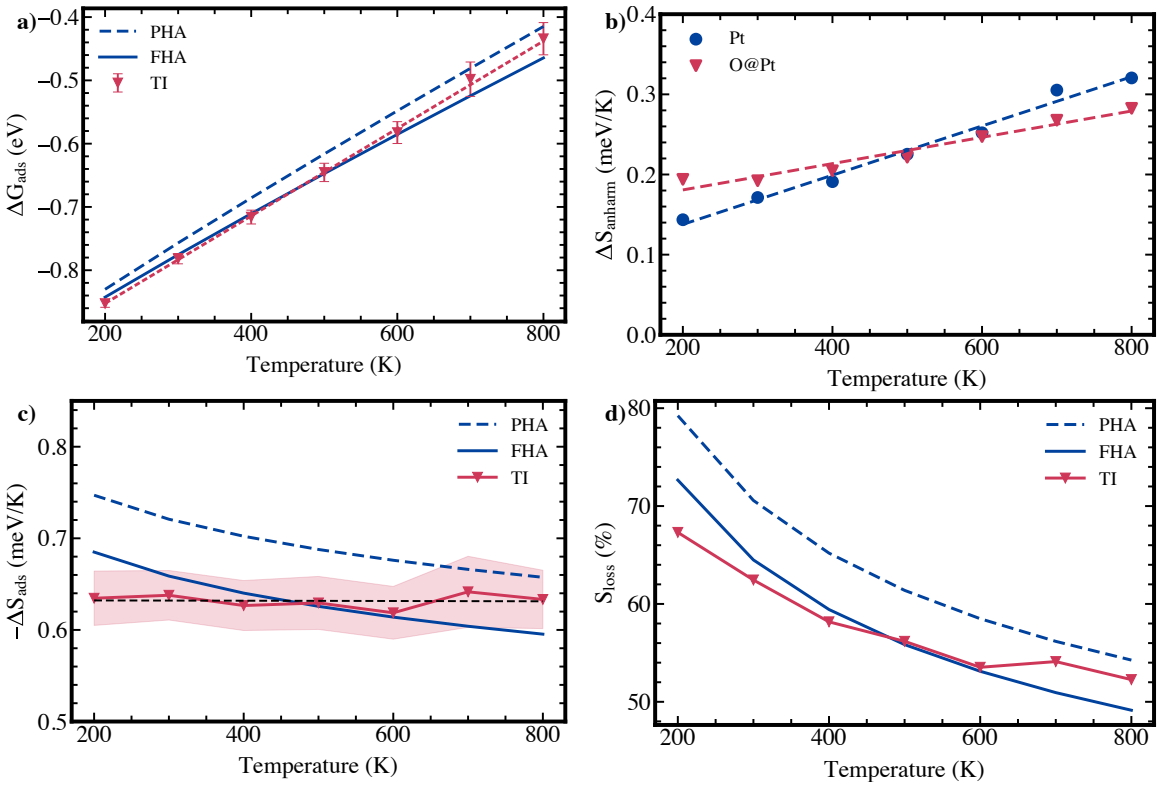


Figure 4.8.: Anharmonic effects of O adsorption as a function of temperature. a) Gibbs free energy of O adsorption on Pt(111) predicted by different methods, namely TI, full harmonic approximation (FHA), and partial harmonic approximation (PHA), as a function of temperature. b) Anharmonic corrections, ΔS_{anharm}, to entropy of clean Pt(111) surface and the adsorbed O@Pt(111) system as functions of temperature. c, d) Adsorption entropy, ΔS_{ads}, and entropy loss upon adsorption, S_{loss}, (in percentage) predicted by the three methods, TI, FHA, and PHA, as functions of temperature. Here, the S_{loss} is defined as the absolute value of the ratio of the adsorption entropy to half of the entropy of O_2 gas at 1 bar and the corresponding temperature. The shaded areas represent the uncertainty in adsorption entropy, resulted from the uncertainty in adsorption free energy from the TI method.

The adsorption free energy of atomic oxygen (Eq. (4.54)) as a function of temperature is shown in Figure 4.8a. Consistent with the results at 300 K, the HA models perform reasonably well across the temperature range. Above 500 K, however, the anharmonic adsorption free energy becomes more positive than those predicted by the FHA, approaching

the PHA values. Because the anharmonicity of the gas-phase molecule is not considered, this trend stems from differences in anharmonic entropy between the adsorbed and clean surfaces. As shown in Figure 4.8b, the anharmonic correction to the entropy of the clean Pt surface exceeds that of the adsorbed system beyond 500 K, leading to a negative anharmonic correction to the adsorption entropy, defined as the difference in anharmonic corrections between the adsorbed and clean systems. This effect directly reflects the reduced surface mobility in the former as discussed above. Interestingly, the adsorption entropy, ΔS_{ads}, given by TI remains nearly constant at $-0.65\,\text{meV}\,\text{K}^{-1}$ with increasing temperature, whereas HA models predict a gradual decline (Figure 4.8c). Such overestimation of adsorption entropy by the HA models has also been reported previously by Sprowl et al.[146].

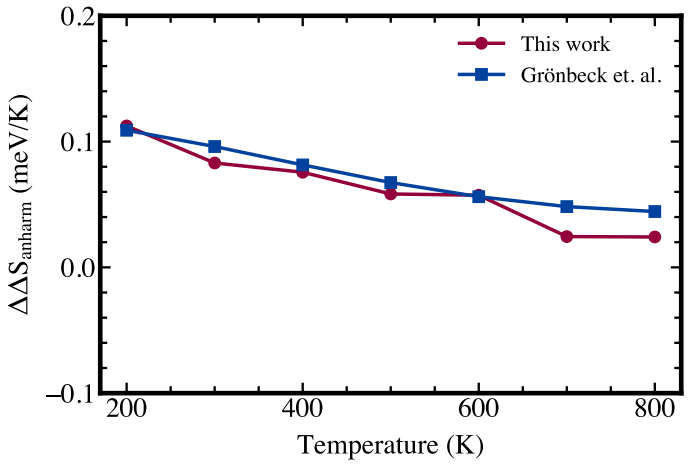


Figure 4.9.: Comparison in anharmonic corrections to the adsorption entropy of O on Pt(111) as a function of temperature, predicted using the MLFF-TI and CPES methods from Ref. [153].

Figure 4.9 compares our computed anharmonic corrections with those reported by Jørgensen and Grönbeck[153]. Because their approach did not include surface contributions, we define our anharmonic correction to adsorption entropy, $\Delta\Delta S_{\text{anharm}}$ as the difference in ΔS_{ads} between TI and PHA, while theirs corresponds to the difference in O entropy between CPES and HA. Our results closely match their reported values, showing only minor deviations above 600 K, likely due to the more rapidly increasing anharmonicity of the clean surface at higher temperatures.

OH adsorption on Pt(111) Figure 4.10 demonstrates that the HA approaches substantially overestimate the Gibbs free energy of OH adsorption (Eq. (4.55)). The MLFF-TI method predicts adsorption free energies that are 0.031 −0.053 kJ mol^{-1} lower than those obtained from the FHA, and 0.039 −0.077 kJ mol^{-1} lower than those from the PHA. These discrepancies primarily arise from the internal rotation of the OH species about the Pt–O axis, as discussed in subsection 4.2.2.1. It is worth noting that the differences would increase to 0.069 −0.166 kJ mol^{-1} and 0.077 −0.191 kJ mol^{-1}, respectively, if local HA references were employed.

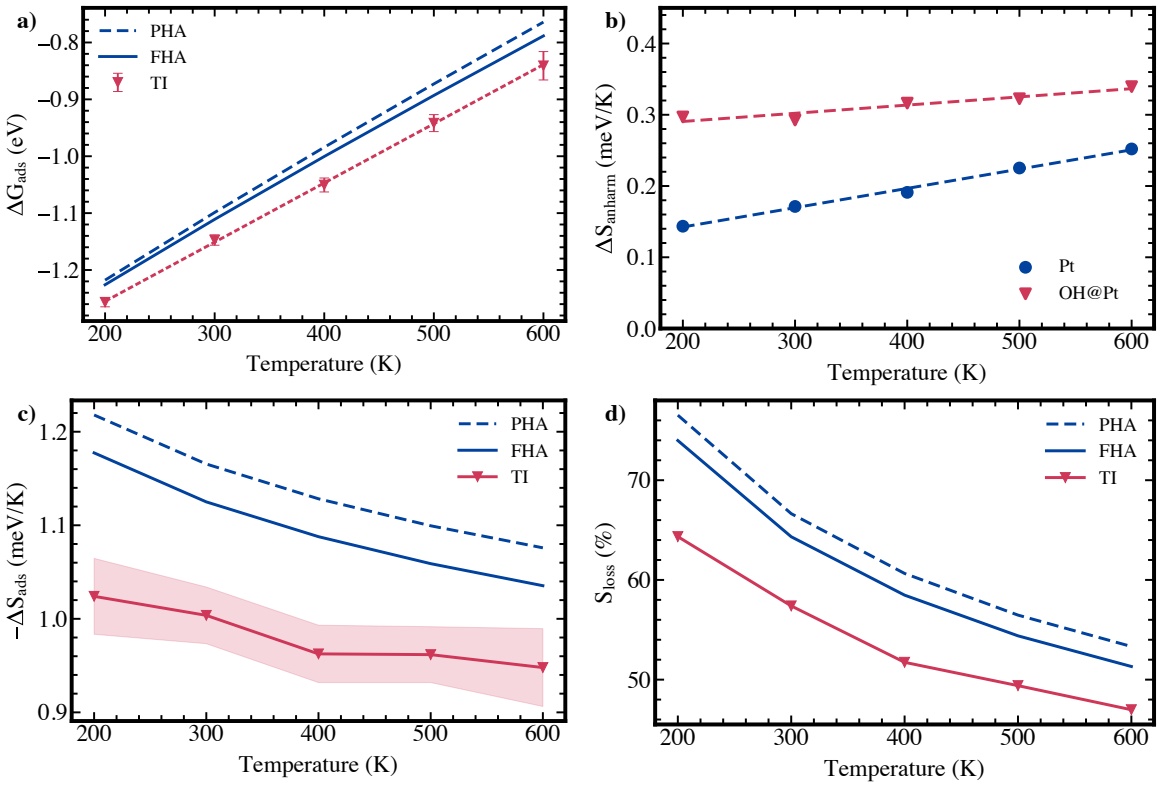


Figure 4.10.: Anharmonic effects of OH adsorption as a function of temperature. a) Gibbs free energy of OH adsorption on Pt(111) predicted by different methods, namely TI, full harmonic approximation (FHA), and partial harmonic approximation (PHA), as a function of temperature. b) Anharmonic corrections, ΔS_{anharm}, to entropy of clean Pt(111) surface and the adsorbed OH@Pt(111) system as functions of temperature. c, d) Adsorption entropy, ΔS_{ads}, and entropy loss upon adsorption, S_{loss}, (in percentage) predicted by the three methods, TI, FHA, and PHA, as functions of temperature. Here, the S_{loss} is defined as the absolute value of the ratio of the adsorption entropy to half of the total entropy of O_2 and H_2 gases at 1 bar and the corresponding temperature. The shaded areas represent the uncertainty in adsorption entropy, resulted from the uncertainty in adsorption free energy from the TI method.

$$\frac{1}{2}O_2 + \frac{1}{2}H_2 + Pt^* \rightarrow OH^*Pt \tag{4.55}$$

A closer examination reveals that the adsorbed system exhibits significantly greater anharmonicity than the clean surface (Figure 4.10b). As temperature rises, the anharmonic contribution of the clean surface grows more rapidly, which can again be attributed to the reduced mobility of surface atoms in the adsorbed system caused by strong OH–Pt interactions. The adsorption entropy increases slightly from $-1.1\,\text{meV}\,\text{K}^{-1}$ at 200 K to $-0.7\,\text{meV}\,\text{K}^{-1}$ at 600 K, corresponding to an entropy loss upon adsorption ranging from 64% to 47%, generally 4-10% lower than HA predictions ((Figure 4.10c, d). Although the HA models approach the TI results with increasing temperatures, larger numerical values of temperatures amplify the entropic contribution to the free energy, leading to increased overestimation seen in Figure 4.10a.

OOH adsorption on Pt(111) The PES of the OOH@Pt(111) system is relatively shallow along the O–O dissociative reaction coordinate, as the OOH species readily decomposes into the more stable co-adsorbed O and OH species on the Pt surface. This behavior arises from the low dissociation barrier of the hydroperoxyl radical on Pt (0.166 eV). Indeed, spontaneous O-O bond cleavage was observed in the trajectories of the fully interacting system at 300 and 400 K. This introduces additional challenges for sampling the PES of the adsorbed state, as dissociation may occur before convergence of $\langle V_1 - V_{0,\boldsymbol{x}} \rangle_\lambda$ in the fully interacting regime. Accordingly, in the MLFF-TI calculations at these temperatures, in addition to the two constraints previously discussed in subsection 4.2.2.1, an extra constraint was applied to the O–O distance to prevent bond cleavage. For probability distribution analyses, multiple unconstrained MLFF–MD trajectories were performed for the fully interacting systems, each terminated immediately before the O–O barrier was crossed, to ensure the statistics of the samplings.

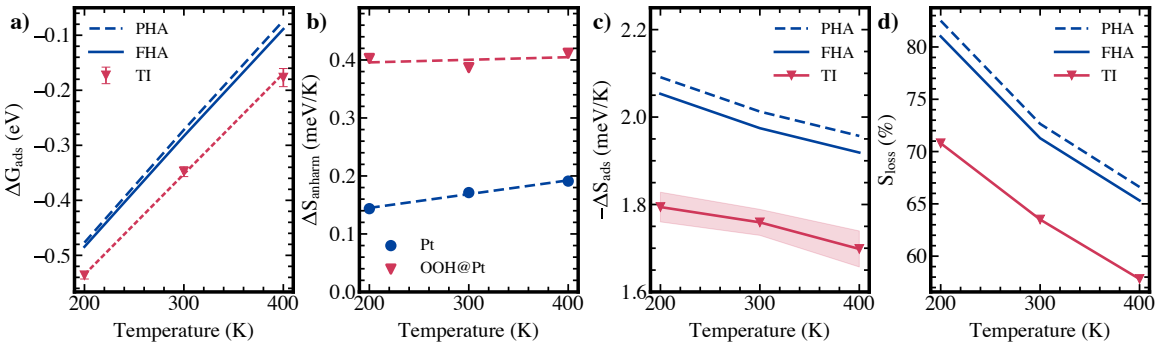


Figure 4.11.: Anharmonic effects of OOH adsorption as a function of temperature. a) Gibbs free energy of OOH adsorption on Pt(111) predicted by different methods, namely TI, full harmonic approximation (FHA), and partial harmonic approximation (PHA), as a function of temperature. b) Anharmonic corrections, ΔS_{anharm}, to entropy of clean Pt(111) surface and the adsorbed OOH@Pt(111) system as functions of temperature. c, d) Adsorption entropy, ΔS_{ads}, and entropy loss upon adsorption, S_{loss}, (in percentage) predicted by the three methods, TI, FHA, and PHA, as functions of temperature. Here, the S_{loss} is defined as the absolute value of the ratio of the adsorption entropy to total entropy of O_2 and half of H_2 gases at 1 bar and the corresponding temperature. The shaded areas represent the uncertainty in adsorption entropy, resulted from the uncertainty in adsorption free energy from the TI method.

$$O_2 + \frac{1}{2}H_2 + Pt^* \rightarrow OOH^*Pt \tag{4.56}$$

The Eq. (4.56) was used to derive the free energy and entropy for the adsorption of OOH species. The HA approaches again overestimate the Gibbs free energy of OOH adsorption (Figure 4.11a). The overestimations are 0.052 to 0.088 eV for FHA and 0.059 to 0.103 eV for PHA, respectively. These amounts are more than 0.020 eV higher than for OH adsorption across the same temperature range. The enhanced anharmonic contributions arise from greater number of anharmonic modes involved in the OOH species, as observed at 300 K. Notably, Figure 4.11b shows that anharmonic correction to entropy for the OOH@Pt(111) system is nearly temperature-independent, indicating that the relevant configurational space was already well-sampled at temperature as low as 200 K. The adsorption entropy

predicted by the TI method increases slightly from -1.8 to $-1.7\,\mathrm{meV\,K^{-1}}$, which is approximately $0.25\,\mathrm{meV\,K^{-1}}$ and $0.3\,\mathrm{meV\,K^{-1}}$ less negative than FHA and PHA predictions, respectively. The recovered anharmonic entropy accounts for more than 20% of the entropy lost upon adsorption. We note in passing that, for this adsorption, a localized HA reference may be more appropriate, since the OOH species is more prone to dissociation than to surface hopping.

Discussion on the temperature effects It has been shown thus far that the temperature dependence of anharmonic thermodynamic quantities, such as the free energy and entropy of adsorption, is strongly adsorbate-specific. The adsorption of O on Pt(111) remains largely harmonic across the studied temperature range, whereas the adsorption of OH exhibits pronounced anharmonicity. In contrast, OOH adsorption occurs on a shallow PES and may no longer correspond to a minimum in the free energy surface at high temperatures. Therefore, applying a single approximate model to all adsorbed intermediates along a reaction pathway can lead to significant errors. The use of MLFF-TI method, which explicitly and accurately captures system anharmonicity, is thus highly necessary. As demonstrated, MLFF-TI recovers over 20% of the entropy loss upon adsorption, providing a more realistic thermodynamic description.

The entropic contributions of the metal surface are often neglected under the HA, or more precisely, assumed to cancel between the adsorbed and clean surfaces. This simplification stems from fact that the low-frequency vibrational modes associated with surface atoms introduce high uncertainty in HA estimations. Many efforts to incorporate anharmonic effects also relied on this assumption. The MLFF-TI workflow, however, enables a direct and accurate treatment of surface atom contributions (see subsection 4.1.8.1), allowing this assumption to be tested explicitly. As demonstrated across all adsorption cases, the FHA predictions, which includes surface vibrational modes, consistently align more closely with the TI results than the PHA estimates, which neglect them. Furthermore, incorporating surface contributions reveals an interesting temperature-dependent interplay of anharmonicity. As temperature increases, adsorbed species exhibit enhanced anharmonic behavior, while simultaneously suppressing the anharmonicity of the underlying surface due to their attractive interactions. As a result, in certain cases, the clean surface become more anharmonic than the adsorbed system.

4.3. Dissociative Adsorption of Methane on Pt(111) Surface - Role of Density Functionals

4.3.1. Overview

The adsorption and dissociation of methane on transition metal surfaces have been subjects of extensive research for several decades because of their high industrial relevance, particularly in processes such as steam reforming[4]. Consequently, this reaction has

been investigated in depth using both experimental and theoretical approaches[278–289]. While experimental studies have provided highly accurate thermodynamic and kinetic data[286, 290], achieving comparable accuracy through theoretical calculations remains a significant challenge[291–293].

It has been demonstrated that employing higher-level electronic structure methods, such as the Random Phase Approximation (RPA) or coupled-cluster (CC), can substantially improve the accuracy of computed electronic energies. For example, Sheldon et al.[257, 259] showed that the adsorption energy of methane on Pt(111) is highly sensitive to the choice of density functional (DF), and that the experimental adsorption energy is well reproduced at the RPA level of theory. Similar correction schemes, such as the hierarchical cluster approach proposed by Sauer and co-workers[216–220] and the autoSKZCAM framework[258], have been widely applied to a variety of catalytic reactions and systems, showing promising agreement with experimental observations. In these schemes, thermal effects are typically assumed to be comparable across different levels of theory, and the vibrational spectra, computed at a lower level, are often used to approximate thermochemical contributions.

The primary objective of the present work is to evaluate the validity of this assumption. Specifically, we examine the influence of two widely used density functionals, BEEF-vdW and PBE-D3, on the computed adsorption free energy and entropy of methane on Pt(111). The analysis is conducted for two distinct adsorption modes of methane: physisorption and dissociative chemisorption, each evaluated using both harmonic and anharmonic models. To accurately capture anharmonic effects, we employ the newly developed MLFF–TI and MLFF–IHSM workflows, described in Section 4.1, for the estimation of anharmonic free energies.

4.3.1.1. Simulation Details

The DFT, MLFF-TI, and MLFF-MD simulations were performed using the same procedure described in subsection 4.2.1.1 at two levels of theory, namely BEEF-vdW and PBE-D3(BJ) (hereafter referred to simply as PBE-D3). The training procedure and details of the MLFFs employed in this study are given in Section A.1.2 of the Appendix. Here, we introduce the details of the MLFF-aided intermediate hard sphere model (MLFF-IHSM).

Following to Ref. [165], the Gibbs free energy of adsorption, ΔG_{ads}, referenced to a pressure of $p_0 = 1$ bar, is expressed as:

$$\Delta G_{\text{ads}} = \Delta A_{\text{ads}} - k_{\text{B}} T \ln\left(\frac{p_0}{p}\right), \tag{4.57}$$

where p is the pressure corresponding to the volumn per molecule of the simulation cell (1366.7 Å^3, which equals approximately 30.3 bar, 50.5 bar, and 70.7 bar at 300, 500, and 700 K, respectively.

To compute the $\Delta A_{0,\text{MLFF}\rightarrow 1,\text{MLFF}}$ via Eq. (4.40), a procedure similar to that in Ref. [165] was used. Every 100^{th} MLFF-MD step from two 4×10^6-step trajectories of the gas-phase

CH_4 and Pt(111) surface were paired. For each pair, 10 random positions of the adsorbate within the simulation cell were selected. This pairing generated a trajectory of 4×10^5 pair configurations, which was divided into two subsets: (1) the accepted trajectory, containing configurations that satisfied the hard sphere condition, and (2) the rejected trajectory, containing those that violated it. The corresponding free energy contribution, $\Delta A_{0,\mathrm{MLFF}\to 1,\mathrm{MLFF}}$, was then computed as:

$$\Delta A_{0,\mathrm{MLFF}\to 1,\mathrm{MLFF}} = -k_\mathrm{B}T \ln\left(\frac{N^{\mathrm{sep}}_{\mathrm{accept}}}{N^{\mathrm{sep}}}\right) \tag{4.58}$$

where $N^{\mathrm{sep}}_{\mathrm{accept}}$ is the number of configurations in the accepted trajectory, $N^{\mathrm{sep}} = 4 \times 10^5$ is the total number of paired configurations, obtained from the separated trajectory.

The configurations from the accepted trajectory were evaluated using the MLFF trained for the CH_4@Pt system to obtain the interaction energy, $\mathcal{V}_{\mathrm{int}}$, from which the second contribution, $\Delta A^{\mathrm{FEP-fw}}_{1,\mathrm{MLFF}\to 2,\mathrm{MLFF}}$, was determined using Eq. (4.42).

Conversely, a 5×10^4-step trajectory of CH_4@Pt, sampled every 100 steps of a 5×10^6-step MLFF-MD trajectory of this system, was employed to evaluate the IHSM potential. Similarly, this trajectory was also divided into two subsets: the accepted and rejected interacting trajectories. The $\Delta A_{2,\mathrm{MLFF}\to 3,\mathrm{MLFF}}$ was computed as:

$$\Delta A_{2,\mathrm{MLFF}\to 3,\mathrm{MLFF}} = +k_\mathrm{B}T \ln\left(\frac{N^{\mathrm{int}}_{\mathrm{accept}}}{N^{\mathrm{int}}}\right), \tag{4.59}$$

where $N^{\mathrm{int}}_{\mathrm{accept}}$ and $N^{\mathrm{int}} = 5 \times 10^4$ represent the number of accepted and total configurations, respectively, for the interacting trajectories.

The configurations from the accepted interacting trajectory were used to determine $\mathcal{V}_{\mathrm{int}}$ by separating the adsorbate and substrate systems and re-evaluating their energies using their respective MLFFs. From the resulting $\mathcal{V}_{\mathrm{int}}$, the backward FEP and BAR evaluations of $\Delta A_{1,\mathrm{MLFF}\to 2,\mathrm{MLFF}}$ were obtained using Eqs. (4.43) and (4.44).

Finally, a set of 200 configurations was randomly selected from the MLFF-MD trajectory of each system and re-evaluated at the corresponding DFT level to obtain the DFT–MLFF correction using Eqs.(4.21) and (4.20). The final adsorption free energy at the DFT level was then determined according to Eq.(4.47).

4.3.2. Results and Discussion

4.3.2.1. Physisorption of methane on Pt(111)

We begin with the calculation of anharmonic free energies, $\Delta G^{\mathrm{anharm}}_{\mathrm{ads}}$, for the physisorption of CH_4 on Pt(111) surface using the MLFF-IHSM workflow, described in subsection 4.1.7. The hard-sphere radii were defined to be twice the minimum atomic distances for Pt-C

and Pt-H obtained from a fully interacting MLFF-MD trajectory at 300 K and at PBE-D3 theory level, corresponding to 5.2 Å and 3.6 Å, respectively. The influence of this choice is analyzed in detail in Section A.6 of Appendix.

Table 4.7.: Contributions to adsorption free energies, $\Delta G_{\text{ads}}^{\text{anharm}}$, in eV, for the physisorption of CH_4 on Pt(111) surface at different temperatures for the BEEF-vdW and PBE-D3 levels of theory. The $\Delta G_{\text{ads}}^{\text{anharm}}$ values are computed as the sum of all contributions, with $\Delta A_{1,\text{MLFF}\to2,\text{MLFF}}^{\text{BAR}}$ used as $\Delta A_{1,\text{MLFF}\to2,\text{MLFF}}$. A reference pressure of 1 bar is used for $\Delta G_{\text{ads}}^{\text{anharm}}$

Quantity	**BEEF-vdW**			**PBE-D3**		
	300 K	500 K	700 K	300 K	500 K	700 K
$\Delta A_{0,\text{DFT}\to0,\text{MLFF}}(CH_4)$	0.0002	-0.0001	-0.0001	0.0002	-0.0002	-0.0005
$\Delta A_{0,\text{DFT}\to0,\text{MLFF}}(\text{Pt})$	0.0124	0.0114	0.0105	0.0146	0.0055	0.0057
$\Delta A_{0,\text{MLFF}\to1,\text{MLFF}}$	0.0390	0.0655	0.0921	0.0382	0.0641	0.0901
$\Delta A_{1,\text{MLFF}\to2,\text{MLFF}}^{\text{fw}}$	-0.0646	-0.0591	-0.0569	-0.1679	-0.1619	-0.1551
$\Delta A_{1,\text{MLFF}\to2,\text{MLFF}}^{\text{bw}}$	-0.0648	-0.0606	-0.0589	-0.1684	-0.1614	-0.1577
$\Delta A_{1,\text{MLFF}\to2,\text{MLFF}}^{\text{BAR}}$	-0.0646	-0.0593	-0.0571	-0.1680	-0.1618	-0.1553
$\Delta A_{2,\text{MLFF}\to3,\text{MLFF}}$	-0.0295	-0.0337	-0.0451	-0.0578	-0.0530	-0.0723
$\Delta A_{3,\text{MLFF}\to3,\text{DFT}}(CH_4@\text{Pt})$	-0.0357	-0.0187	-0.0116	0.0213	0.0370	0.0398
$-k_{\text{B}}T\ln(p_0/p)$	0.0882	0.1690	0.2569	0.0882	0.1690	0.2569
$\Delta G_{\text{ads}}^{\text{anharm}}$	0.0010	0.1341	0.2456	-0.0633	0.0606	0.1644

Table 4.7 presents the computed contributions to $\Delta G_{\text{ads}}^{\text{anharm}}$. Since the term $\Delta A_{0,\text{MLFF}\to1,\text{MLFF}}$ arises from the reduction in accessible volume, its values are consistently positive and independent of the chosen DFT method. In contrast, the $\Delta A_{1,\text{MLFF}\to2,\text{MLFF}}$ term, which represents adsorbate–surface interactions within the hard-sphere region, dominates the overall free energy $\Delta G_{\text{ads}}^{\text{anharm}}$ and shows only minor temperature dependence (variations of approximately 0.01 eV), consistent with the findings of Plessow[165]. This contribution, however, is highly sensitive to the choice of density functional (DF). The BEEF-vdW functional predicts values about 0.1 eV higher than those obtained with PBE-D3 (approximately −0.06 eV vs. −0.16 eV). The $\Delta A_{2,\text{MLFF}\to3,\text{MLFF}}$ term depends on both temperature and the employed DF, with BEEF-vdW again yielding slightly larger values (less negative). The correction terms, $\Delta A_{0,\text{DFT}\to0,\text{MLFF}}(CH_4)$, $\Delta A_{0,\text{DFT}\to0,\text{MLFF}}(\text{Pt})$, and $\Delta A_{3,\text{MLFF}\to3,\text{DFT}}(CH_4@\text{Pt})$, reflect the intrinsic accuracy of the corresponding MLFFs. These corrections are generally negative for BEEF-vdW and positive for PBE-D3. Overall, the resulting adsorption free energies indicate that the methane physisorption is endergonic under most conditions, except for the PBE-D3 case at 300 K, where the process is 0.06 eV exergonic.

We next compare the Gibbs free energy of adsorption computed using the MLFF-IHSM workflow with those obtained from analytical approximations. Table 4.8 lists the ΔG_{ads} values derived from MLFF-IHSM alongside estimates from the localized HA (HA_1), mobilized HA (HA_2), and the 2-dimensional free-translator/3-dimensional free-rotor (2FT3FR) models. Interestingly, while the HA models at PBE-D3 level consistently overestimate the adsorption free energies, as expected, with deviations between 0.11 to 0.17 eV for HA_1 and approximately 0.04 eV for HA_2 across the temperature range, they instead underestimate

Table 4.8.: Comparison of adsorption free energies, in eV, at different temperatures for the BEEF-vdW and PBE-D3 levels of theory, estimated using different models. HA_1 denotes the standard localized HA model; HA_2 represents the mobilized HA accounting for the number of equivalent adsorption sites (Eq. (4.53)), and 2FT3FR refers to the 2-dimensional free-translator/3-dimensional free-rotor model.

Quantity	BEEF-vdW			PBE-D3		
	300 K	500 K	700 K	300 K	500 K	700 K
HA_1	-0.0315	-0.0062	0.0049	0.0460	0.1974	0.3346
HA_2	-0.0883	-0.1008	-0.1276	-0.0108	0.1027	0.2020
2FT3FR	0.0088	0.1160	0.2267	-0.0207	0.1411	0.3066
MLFF-IHSM	0.0100	0.1341	0.2456	-0.0633	0.0606	0.1644

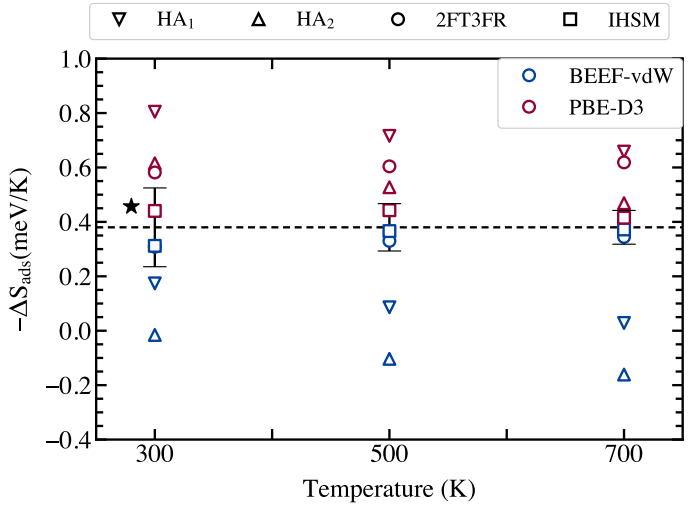


Figure 4.12.: Entropy of CH_4 adsorption on Pt(111) estimated by different models, namely, localized HA (HA_1), mobilized HA (HA_2), 2-dimensional free-translator/3-dimensional free-rotor (2FT3FR), and IHSM. The value are computed at the BEEF-vdW and PBE-D3 levels of theory and at various temperatures. The dashed line denotes the experimentally determined adsorption entropy from Ref. [250] with error bars corresponding deviations of 0.04 eV in $-T\Delta S_{ads}$ term at the respective temperatures. The black star presents the value estimated using hindered-translator hindered-rotor (HTHR) from Ref. [146]. Both experimental and HTHR values were determined at 63 K.

ΔG_{ads} at BEEF-vdW, yielding adsorption free energies that are 0.04 to 0.37 eV more favorable, despite the weaker interaction strength predicted by this functional, as discussed previously. An examination of the vibrational spectra of the CH_4@Pt system evaluated at the two levels reveals that BEEF-vdW yields consistently lower frequencies for modes associated with CH_4, particularly for the "helicopter-like" hindered rotational and in-plane hindered translational modes, which exhibit frequencies of as low as 1.9 and 6.2 cm^{-1}, respectively. These two highly uncertain low-frequency modes contribute disproportionately to the erroneous underestimation of the free energy. This finding underscores a key limitation of the HA for weakly bound systems that its errors can occur in either direction. The 2FT3FR model slightly underestimates adsorption free energies at the BEEF-vdW level but overestimates them at PBE-D3. We consider the latter behavior to be more reasonable, since the 2FT3FR model is often regarded as adsorption transition state [250], and its

predicted ΔG_{ads} values should exceed the true anharmonic free energy of adsorption, $\Delta G_{\text{ads}}^{\text{anharm}}$.

Figure 4.12 illustrates the adsorption entropy (ΔS_{ads}) estimated using the different methods discussed above. For the derivation of the entropy terms, harmonic enthalpies (internal energies) were employed across all models. It is noteworthy that the ΔS_{ads} values obtained from the MLFF–TI method show excellent agreement with experimental data[250], with deviations remaining within chemical accuracy, regardless of the chosen DF. Our predictions also closely align with those from the hindered-translator/hindered-rotor (HTHR) model[146]. A clear trend can be observed as the PBE-D3 functional, which predicts stronger adsorption energies, tends to underestimate the adsorption entropy (that is, it predicts greater entropy loss upon adsorption), whereas BEEF-vdW tends to overestimate the adsorption entropy. Nevertheless, the anharmonic ΔS_{ads} values obtained from the two functionals converge at higher temperatures, indicating that thermal effects reduce the functional dependence of the entropy estimates.

4.3.2.2. Dissociative adsorption of methane on Pt(111)

We now investigate the influence of the two DFs, BEEF-vdW and PBE-D3, on predicting the dissociative adsorption of CH_4 on Pt(111) surface, as described by Eq. (4.60).

$$CH_4\ (g) + 2Pt^* \rightarrow CH_3{}^*Pt + H^*Pt \tag{4.60}$$

Figure 4.13 illustrates the reference structures for the relevant CH_3@Pt(111) and H@Pt(111) systems predicted by the two DFs. Both functionals yield similar optimized geometries for the CH_3@Pt(111) systems, with the CH_3 species occupying an on-top adsorption site. However, the Pt-C bond length predicted by BEEF-vdW is 0.01 Å longer than that obtained at the PBE-D3 level, suggesting slightly weaker interactions. On the other hand, the two DFs differ in their predictions of the preferred adsorption site for atomic hydrogen. BEEF-vdW predicts the on-top site to be the most stable structure, with a Pt-H distance of 1.555 Å, whereas PBE-D3 predicts the fcc hollow site to be the most favorable, with Pt-H distances of 1.865 Å.

The adsorption free energies estimated using HA reveal that the values ΔG_{ads} obtained at the BEEF-vdW level are consistently higher than those from PBE-D3 level (Figure 4.14a). BEEF-vdW predicts the dissociative adsorption process to be endergonic across the entire temperature range of 200 −800 K, while PBE-D3 predicts it to be exergonic below 400 K and endergonic above that temperature. Figure 4.14b and c show that this pronounced discrepancy in ΔG_{ads} arises primarily from the difference in enthalpic contributions, which depend strongly on the accuracy of the employed functional in predicting electronic energies. For the dissociative adsorption of CH_4 on Pt(111), BEEF-vdW overestimates the adsorption enthalpy by more than 0.5 eV relative to PBE-D3. Since experimental enthalpy of this reaction was determined to be between −0.08 eV[286] and −0.16 eV[294], the results from the two functionals can serve as upper and lower limits. Interestingly, the entropic contributions remain nearly identical across the temperature range, with a maximum deviation of only 0.014 eV at 700 K. We found that the BEEF-vdW systematically overestimates

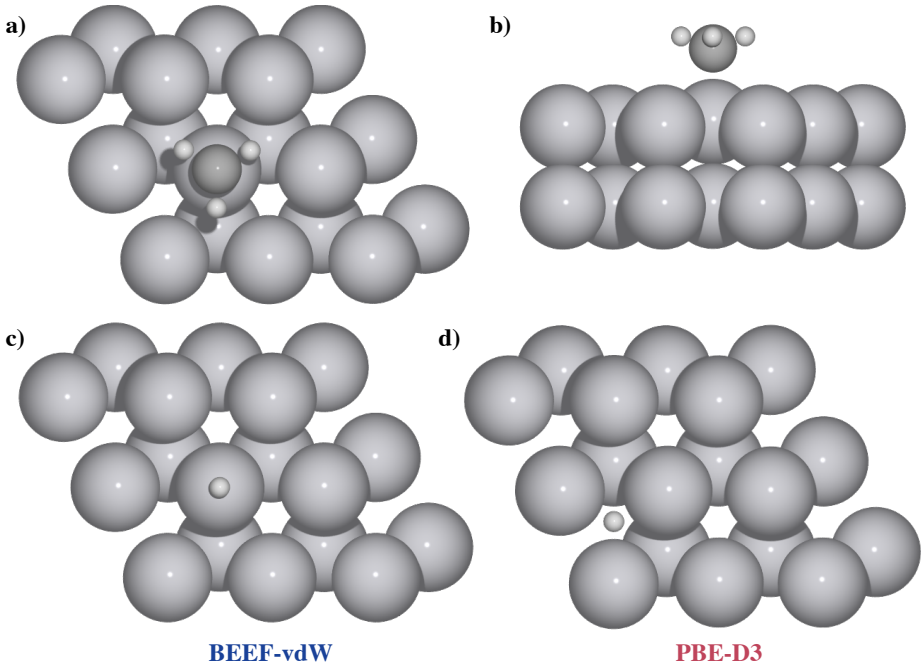


Figure 4.13.: a, b) Top and side views of reference structure for CH_3@Pt(111) predicted by both the BEEF-vdW and PBE-D3 functionals; c, d) The top views of the reference structure for H@Pt(111) predicted by the BEEF-vdW and PBE-D3 functionals, respectively. The two bottom layers are omitted for visual clarity.

the entropies of both the clean surface and adsorbed systems compared to PBE-D3 (see Figure A.3 in the Appendix). However, these overestimations largely cancel out when computing the overall adsorption entropy. In contrast to the case of CH_4 physisorption, the entropy of this dissociative chemisorption can, within the HA framework, be considered largely independent of the chosen functional. Therefore, vibrational spectra computed using a lower-cost functional can be reliably employed to estimate entropic contributions for higher-level methods. This finding supports the implementation of correction schemes where the electronic energy of part or all of the system is refined using higher level methods, such as the hierarchical cluster approach discussed in Section 3.1.2. However, caution should be exercised when physisorbed intermediates are involved.

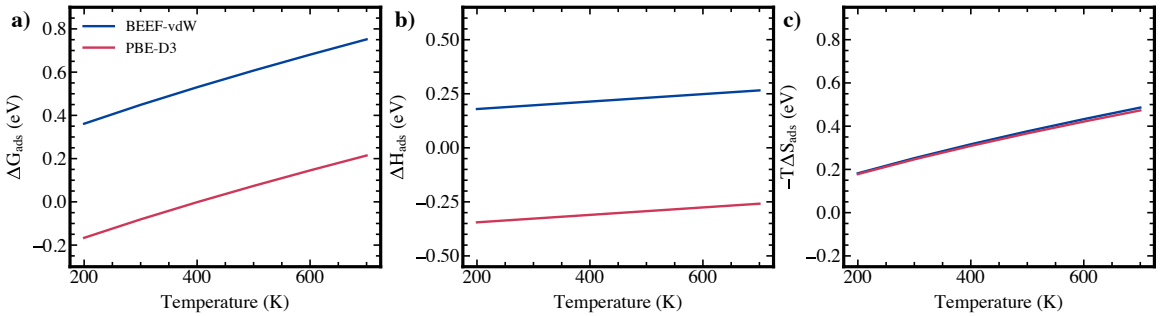


Figure 4.14.: Impact of density functionals on dissociative adsorption CH_4 on Pt(111) surface under HA on a) adsorption free energy, b) enthalpy, and c) entropic contribution as functions of temperature.

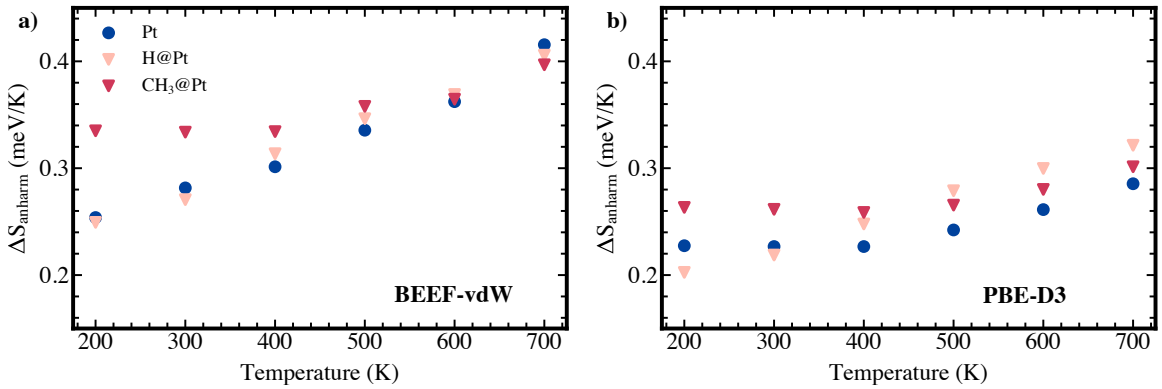


Figure 4.15.: The anharmonic corrections to entropy for the three individual systems relevant to the dissociative adsorption of CH_4 on Pt(111) surface as functions of temperature, obtained using the MLFF-TI workflow at a) BEEF-vdW and b) PBE-D3 levels of theory.

We now address the question of whether this observation remains valid in the anharmonic regime, using the MLFF-TI workflow. Figure 4.15 illustrates the anharmonic correction to entropy, ΔS_{anharm}, for the three individual systems as a function of temperature, obtained with the two functionals. The anharmonic effects are found to be more pronounced at the BEEF-vdW level than at the PBE-D3 level across all examined systems, likely reflecting the weaker interactions predicted by BEEF-vdW. Similar behavior was also observed previously for H@Cu(111) system[156]. The anharmonic contributions range from 0.25 −0.42 meV K^{-1} and 0.20 −0.32 meV K^{-1} for BEEF-vdW and PBE-D3, respectively. At the BEEF-vdW level, the ΔS_{anharm} values of CH_3@Pt system are the largest at low temperatures, while those of H@Pt and clean Pt are essentially identical. The anharmonicity of the two adsorbed systems arises mainly from the translation and self-rotation of the methyl species around the Pt-C axis in the former, and from hydrogen atom's translation in the latter (in addition to the anharmonicity of the surface atoms). It should be noted, however, that mobilized HA models were applied for both CH_3@Pt and H@Pt systems; using localized HA models would result in larger anharmonic contributions for both. The differences in ΔS_{anharm} among the three systems diminish at high temperatures, due to faster increase in anharmonicity for the H@Pt and clean Pt systems. A similar trend is observed at the PBE-D3 level, where anharmonic correction of the CH_3@Pt system is initially the largest but increases more slowly with temperature. The adsorbed hydrogen atom exhibits the least anharmonicity at 200 K, but becomes the most anharmonic after 500 K, while the clean Pt surface shows relatively small anharmonic effects that increase at a rate comparable to that of CH_3@Pt system.

Figure 4.16a and b show the accurate anharmonic free energy and its entropic contribution for the studied reaction, in comparison with the HA results. The MLFF-TI method reduces the ΔG_{ads} by 0.06 −0.13 eV at the BEEF-vdW level and by 0.02 −0.11 eV at the PBE-D3 level. The anharmonic corrections predicted by BEEF-vdW are consistently larger (more negative) than those obtained with PBE-D3, by approximately 0.02 −0.03 eV. These differences correspond to notable increases in adsorption entropy, as shown in Figure 4.16c. A more pronounced recovered entropy of about 0.3 meV K^{-1} is observed for BEEF-vdW at 200 K, decreasing gradually to approximately 0.2 meV K^{-1} at 800 K. Meanwhile, the

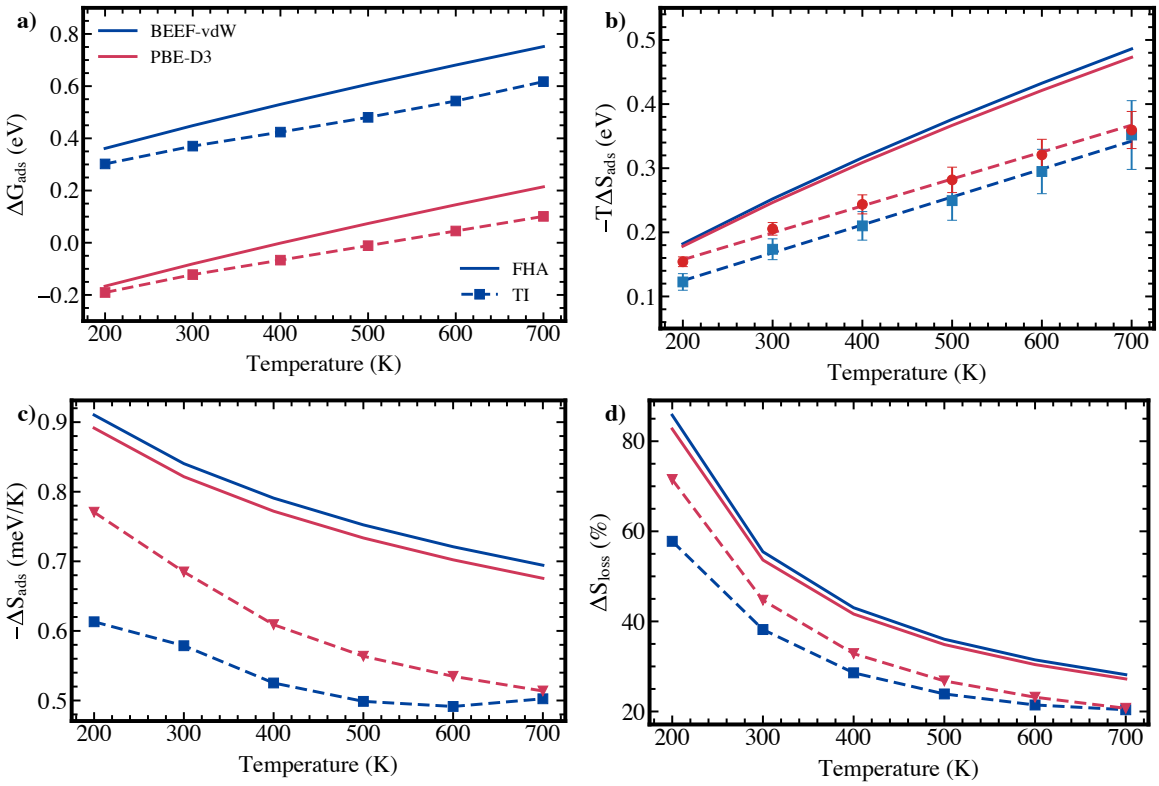


Figure 4.16.: Anharmonic effects of dissociative adsorption of CH_4 on Pt(111) surface as a function of temperature and functional. a) Gibbs free energy of dissociative adsorption of CH_4 on Pt(111) surface predicted by the MLFF-TI (TI) and full harmonic approximation (FHA). b) Entropic contribution, $-T\Delta S_{ads}$, to the free energy for dissociative adsorption of CH_4 on Pt(111) surface predicted by TI and FHA. c, d) Adsorption entropy, ΔS_{ads}, and entropy loss upon adsorption, S_{loss}, (in percentage) predicted by the TI and FHA methods. Here, the S_{loss} is defined as the absolute value of the ratio of the adsorption entropy to total entropy of CH_4 gas at 1 bar and the corresponding temperature. The uncertainties in adsorption entropy and entropy loss are omitted for simplicity.

MLFF-TI method at the PBE-D3 level recovers ΔS_{ads} by an increasing amount, from 0.1 to 0.15 meV K^{-1}, as temperature increases. Notably, the anharmonic adsorption entropies predicted from the two functionals converge at higher temperatures (around 700 K). Overall, the MLFF-TI method recovers between 8% and 25% of the entropy loss upon adsorption across the investigated temperature range.

Finally, to evaluate the impact of anharmonicity on the equilibrium constant for the dissociative adsorption of methane on Pt: CH_4 (g) + Pt^* ⇌ $CH_3{}^*Pt$ + H^*Pt, we compute $K(T)$ as:

$$K(T) = \exp\left(\frac{-\Delta G_{ads}}{k_B T}\right) \tag{4.61}$$

The results with the BEEF-vdW functional exhibit a typical trend for an endothermic process, whereas PBE-D3 predicts an exothermic behavior. The increased adsorption entropies resulting from anharmonic corrections in both levels lead to higher equilibrium constants, suggesting greater adsorbate coverage at a given temperature. For BEEF-vdW, anharmonicity enhances the equilibrium constant by a factor of approximately 31 at 200 K,

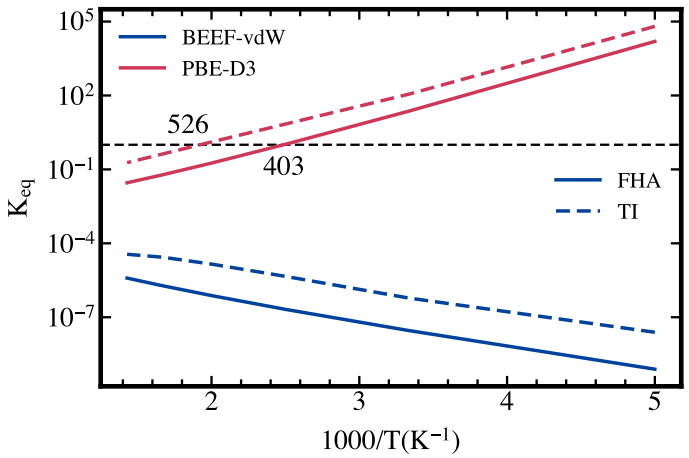


Figure 4.17.: The van't Hoff plot showing the equilibrium constant for the dissociative adsorption of CH_4 on Pt(111) predicted by the MLFF-TI (TI) and full harmonic approximation (FHA) at BEEF-vdW and PBE-D3 levels.

decreasing to a factor of about 9 at 700 K. At the PBE-D3 level, the improvement remains relatively consistent, ranging between factors of 4 and 6 across the entire temperature range. Enhancements of similar factors were also reported for the methane dissociation on Ni(111)[158]. Anharmonic effects further influence the temperature at which the equilibrium shifts to favor (or disfavor) adsorption. At the PBE-D3 level, the HA predicts this temperature to be 403 K, while the MLFF-TI method raises it by more than 100 K to around 526 K, highlighting the significant thermodynamic impact of anharmonicity.

4.4. Conclusions

In this chapter, we introduced and validated two efficient methods for calculating anharmonic free energies in periodically extended systems, both integrating machine-learned force fields (MLFFs). MLFFs significantly reduced the computational cost associated with long DFT-based molecular dynamics (MD) simulations in the workflows of thermodynamic λ-integration (λ-TI) and the intermediate hard-sphere model (IHSM), making these methods practical for routine free energy evaluations in computational catalysis. By enabling nanosecond-scale MD trajectories, MLFFs greatly enhanced sampling efficiency and statistical reliability. To ensure DFT-level accuracy, a combination of complementary statistical approaches, including free-energy perturbation (FEP), machine-learning thermodynamic perturbation theory (MLPT), Bennett's acceptance ratio (BAR), and the phase-space overlap index (I_w), was employed to evaluate and refine MLFF-derived results.

The developed MLFF-aided frameworks proved to be both robust and computationally efficient. Within the MLFF-aided λ-TI (MLFF-TI) approach, we demonstrated that the resulting free energy is independent of the choice of arbitrary harmonic reference. Leveraging this property, we discussed strategies to (i) improve statistical convergence using appropriately modified harmonic references, (ii) reduce computational expense by employing

MLFF-generated references, and (iii) simplify TI setup through Cartesian coordinate–based references. A proof-of-concept demonstration showed that a pretrained, general-purpose MLFF could successfully reproduce results obtained using system-specific MLFFs, further emphasizing the generalizability of the approach.

The MLFF-TI workflow was then applied to investigate the adsorption of oxygenated species, namely, O, OH, and OOH, on Pt(111), highlighting the situations where the harmonic approximation (HA) fails. At 300 K, the atomic oxygen adsorbate behaved nearly harmonically, while both OH and OOH exhibited strong anharmonicity. For OH adsorption, anharmonic effects originated from in-plane translation and helicopter-like rotation around the Pt–O axis, whereas for OOH, they arose from rotations around both the Pt–O and O–O axes of the OOH and OH groups, respectively. With increasing temperature, O adsorption remained largely harmonic, while OH adsorption became progressively more anharmonic. In contrast, OOH adsorption occurred on a shallow potential energy surface and might no longer represent a free energy minimum at elevated temperatures. These findings demonstrate that applying a single approximate model to all intermediates along a reaction pathway can lead to substantial errors. Moreover, the anharmonicity of surface atoms was found to play an increasingly important role with rising temperature.

We further examined the effects of two widely used density functionals, BEEF-vdW and PBE-D3, on adsorption entropies for both physisorbed and dissociatively adsorbed methane on Pt(111). For methane physisorption, BEEF-vdW overestimated, whereas PBE-D3 underestimated the adsorption entropy within HA, underscoring the large uncertainty of this approximation for weakly bound systems. Using the MLFF-IHSM approach, however, both functionals successfully reproduced the experimental adsorption entropy. In the case of dissociative chemisorption, the two functionals yielded comparable HA entropies due to systematic error cancellation. However, in the anharmonic regime, evaluated via MLFF-TI, significant functional-dependent deviations emerged: BEEF-vdW, which predicts weaker adsorbate–surface interactions, gave larger anharmonic corrections at lower temperatures, leading to higher (less negative) adsorption entropies. The PBE-D3 functional also showed noticeable anharmonic contributions, though to a smaller extent. At elevated temperatures, where adsorbates became more mobile, the adsorption entropies predicted by both functionals converged once again. These results indicate that both functionals tend to give consistent entropic contributions for strongly bound systems within HA, and for weakly bound systems when anharmonic effects are properly accounted for.

Overall, the studies presented in this chapter highlight the limitations of the harmonic approximation and emphasize the necessity of employing more accurate methods that can reliably capture anharmonic effects to improve the estimation of adsorption free energies and their derived thermodynamics and kinetics. The developed MLFF-TI and MLFF-IHSM frameworks demonstrate a clear path forward toward efficient and accurate anharmonic free energy calculations for complex catalytic systems.

Part III.

Concluding Remarks

The first part of this work conducted an in-depth investigation into ethylene oligomerization and polymerization on two transition metal-based single-atom catalysts (SACs) supported on the Zr_6O_8-based metal-organic framework (MOF) NU-1000. Different theoretical aspects of the first reaction on a Ni-NU-1000 catalyst were systematically examined. The electronic structure of the catalyst was found to be highly localized, as reflected in the excellent agreement between truncated cluster models and the full periodic representation. Negligible deviations due to the missing linkers confirmed that these simplified models remain reliable for simulating catalytic processes. The local character of electronic structure enabled a hierarchical cluster approach that refined DFT electronic energies: a formate cluster truncated from the plane-wave PBE-D3(BJ)-optimized periodic system was corrected at higher electronic structure levels, while long-range steric effects were preserved. Together with functional benchmarking from our collaborators, where TPSSh, M06-L, and B3LYP combined with D3 dispersion corrections showed the closest agreement with coupled-cluster reference data, these results validated the widely use of cluster model and M06-L theory level in computational studies of catalysis within Zr_6O_8-based MOFs.

Mechanistic analysis demonstrated that Ni centers could adopt distinct configurations, with the more accessible $Ni^{II}O_2$-type species exhibiting higher catalytic activity. In both cases, ethylene insertion was identified as the TOF-determining step, while a β-agostic interaction played a decisive role in facilitating hydrogen elimination, thus rationalizing the selectivity toward dimer formation. Remarkably, the TOF values derived from TPSSh-corrected free energy profiles within the energetic span model closely reproduced experimental measurements, underscoring the robustness of the theoretical framework. Additionally, dispersion interactions between the growing oligomer chains and the MOF linkers were found to stabilize the system, favoring chain growth within the smaller trigonal pores over the hexagonal mesopores.

Insights obtained from the Ni-based catalyst were further extended to the Cr-NU-1000 system. Among various oxidation states explored, the quartet-state Cr^{III} species emerged as the active site for ethylene polymerization. The [Cr^{III}-ethyl] intermediate adopted a tetrahedral geometry that transformed into a trigonal bipyramidal configuration upon ethylene coordination. Polymerization followed a linear insertion Cossee–Arlman mechanism, exhibiting a remarkably low free energy barrier of 50 kJ mol^{-1} at 298.15 K and 1 bar, while chain termination through β-hydrogen elimination or chain transfer was about 30 kJ mol^{-1} less favorable. In contrast, the Cossee–Arlman mechanism on the Cr^{II} catalyst favored oligomerization, attributed to the dynamic conversion of the active site from a trigonal bipyramidal to a square-planar geometry that promotes β-hydrogen elimination. Alternative metallacycle pathways on $Cr^{(I)/(III)}$ and $Cr^{(II)/(IV)}$ catalysts were found to require prohibitively high activation barriers, rendering them inactive under typical conditions. This work provides theoretical support for experimental findings derived from crystallographic and spectroscopic data. Predictions obtained using PBE, TPSSh, M06-L, B3LYP consistently captured qualitative reaction trends across oxidation states and mechanisms.

This study provides highly accurate energetic data derived from the hierarchical cluster approach, along with benchmarked results that offer significant value to the computa-

tional MOF catalysis community. These validated energies form a robust foundation for integrating atomistic insights with macroscopic descriptors such as mass diffusion and pore capacity data, enabling the development of kinetic models that more closely reflect experiments. Within such models, temperature and pressure effects can be systematically tuned to tailor product selectivity toward desired oligomer or polymer distributions, thereby linking molecular-level understanding with process-level optimization. Moreover, incorporating uncertainty quantification, based on the deviations identified in the benchmark assessments, would allow propagation of DFT errors into the kinetic framework, offering a quantitative measure of model reliability and predictive confidence, an aspect often overlooked in existing kinetic modeling studies. The finding that active sites in these SACs exhibited structural flexibility further underscores the need for multiscale modeling strategies. Future studies should therefore employ scale-bridging techniques, such as molecular dynamics (MD), Monte Carlo (MC), and other enhanced sampling approaches, to capture the complex dynamic behavior of catalytic centers and reactions under realistic operating conditions. The development of machine learned force fields (MLFFs) is expected to play a pivotal role in enabling such *operando* simulations, efficiently expanding the accessible time and length scales in MOF-based catalysis.

The second part of this work tackled the long-standing limitations of the harmonic oscillator approximation (HA) by developing efficient workflows to calculate accurate anharmonic free energy for periodically extended systems. The computational challenges associated with thermodynamic λ-integration (λ-TI) and intermediate hard sphere model (IHSM), primarily due to expensive DFT-based phase space samplings, were effectively overcome through the incorporation of MLFFs. The use of MLFFs enabled nanosecond-scale trajectories, achieving extensive sampling that would be impractical with DFT alone, thereby substantially improving convergence and reliability of the results. Complementary statistical techniques, including free energy perturbation (FEP), machine learning thermodynamic perturbation theory (MLPT), Bennett's acceptance ratio (BAR), and the phase space overlap index, I_w, were used to assess MLFF-derived results and correct them to the desired DFT accuracy. This approach successfully bridged the time-scale limitations of conventional DFT-based free energy calculation approaches while preserving DFT-level accuracy.

The developed workflows were applied to a diverse range of adsorption phenomena, from the chemisorption of oxygenated species to the physisorption and dissociative adsorption of methane on Pt(111). The atomic oxygen adsorption exhibited nearly harmonic behavior and was reasonably described by the HA, whereas OH and OOH adsorption displayed strong anharmononicity, with the MLFF-TI method recovering up to 10% of the entropy loss estimated under HA. Anharmonic effects of the clean surface intensified more rapidly with temperature compared to the adsorbed surfaces, leading to a reduction in the anharmonic correction to adsorption free energy at elevated temperatures. The partial harmonic approximation (PHA), which considers only adsorbate-related degrees of freedom, was shown to systematically overestimate the entropy loss upon adsorption and consequently the adsorption free energy. For CH_4 adsorption, the influence of density functionals were evaluated. BEEF-vdW, which tended to underestimate adsorption strength, predicted higher anharmonic contributions relative to PBE-D3(BJ). At high temperatures or within the physisorption regime, where the adsorbate became more mobile, the impact of anhar-

monicity was found to be nearly independent of the employed density functionals. Using the MLFF-IHSM approach, the experimental adsorption entropy for CH_4 physisorption was accurately reproduced at both theory levels, while the adsorption entropies from both levels converged at 700 K for dissociative adsorption process.

The robustness, efficiency, and generality of these MLFF-aided workflows were intensively validated. A key strength of the λ-TI formalism is the independence of the final corrected free energy from the chosen harmonic reference. This property was demonstrated by modifying Hessian matrices to define the harmonic reference to enhance the sampling in harmonic regime and by using MLFF-generated references to eliminate the need for DFT geometry optimization and finite-difference calculations. The FEP and the diagnostic I_w index proved to be powerful and reliable tools for correcting MLFF-level results to DFT accuracy. Furthermore, the use of straightforward Cartesian coordinates, with minimal complexity introduced by a few simple holonomic constraints, typically expressed as internal coordinates, simplified the setup of MLFF-TI simulations, overcoming the difficulty of defining complex curvilinear coordinates in the previous λ-TI implementations. A proof-of-concept demonstration employing a pretrained foundation MLFF model within the MLFF-TI workflow successfully reproduced the DFT-accurate anharmonic free energy with only 200 additional DFT evaluations. Such efficient corrections from MLFF to DFT using perturbation techniques can be readily extended to the MLFF-IHSM method, suggesting that any MLFF-level free energy calculation can, in principle, be refined to achieve DFT accuracy. While demonstrated here for adsorption processes, the MLFF-TI framework can be directly applied to compute activation free energies of elementary reactions, as originally demonstrated in its development.

Looking forward, the realization of a future where universal foundation models are routinely used for accurate free enegy evaluations will depend on further advances in Δ-machine learning architectures and optimized training selection strategies to minimize the number of DFT calculations required. Once such efficient corrective schemes are established, it will become feasible to achieve anharmonic free energies at the most accurate electronic structure levels, including random phase approximation (RPA) and coupled cluster methods such as CCSD(T). Extending these models to MOF systems will open the door to applying MLFF-TI and MLFF-IHSM workflows to key catalytic steps, enabling a deeper understanding of anharmonicity and its impact on thermodynamic quantities such as free energy and entropy. These refined thermodynamic insights, when coupled with multiscale kinetic modeling as outlined previously, will serve to bridge the persistent gap between theory and experiment in complex catalytic materials.

Overall, this work addressed distinct but complementary dimensions of the multiscale challenge in computational catalysis in MOF and other heterogeneous systems. By integrating accurate thermodynamic modeling, efficient machine-learning acceleration, rigorous benchmarking, and comprehensive mechanistic exploration, the present study contributes to a deeper molecular-level understanding of heterogeneous catalytic processes in MOFs and related systems, effectively bridging spatial and temporal scales and advancing theoretical predictions toward experimental observations.

Bibliography

(1) International Energy Agency, *Net Zero by 2050: A Roadmap for the Global Energy Sector*; OECD: 2021.

(2) Calvin, K. et al., *IPCC, 2023: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland.* Arias, P. et al., Eds., 2023.

(3) (IEA), I. E. A. Aviation — IEA, `https://www.iea.org/energy-system/transport/aviation`, Last updated 16 January 2025, IEA, Paris, 2025.

(4) Chorkendorff, I.; Niemantsverdriet, J. W., *Concepts of Modern Catalysis and Kinetics*; Wiley: 2003.

(5) Hagen, J., *Industrial catalysis*, 3rd ed.; Wiley-VCH Verlag: Weinheim, Germany, 2015.

(6) Yang, X.-F.; Wang, A.; Qiao, B.; Li, J.; Liu, J.; Zhang, T. *Accounts of Chemical Research* **2013**, *46*, 1740–1748.

(7) Qiao, B.; Wang, A.; Yang, X.; Allard, L. F.; Jiang, Z.; Cui, Y.; Liu, J.; Li, J.; Zhang, T. *Nature Chemistry* **2011**, *3*, 634–641.

(8) Wang, A.; Li, J.; Zhang, T. *Nature Reviews Chemistry* **2018**, *2*, 65–81.

(9) Yaghi, O. M.; Kalmutzki, M. J.; Diercks, C. S. Index, 2019.

(10) Nobel Prize Outreach Advanced Information, `https://www.nobelprize.org/prizes/chemistry/2025/advanced-information/`, Accessed: 2025-10-20, 2025.

(11) Unke, O. T.; Chmiela, S.; Sauceda, H. E.; Gastegger, M.; Poltavsky, I.; Schütt, K. T.; Tkatchenko, A.; Müller, K.-R. *Chemical Reviews* **2021**, *121*, 10142–10186.

(12) Zimmermann, H.; Walzl, R. Ethylene, 2009.

(13) (IEA), I. E. A. *The Future of Petrochemicals*, Licence: CC BY 4.0; Paris: IEA, 2018.

(14) Geyer, R.; Jambeck, J. R.; Law, K. L. *Science Advances* **2017**, *3*, e1700782.

(15) Tian, P.; Wei, Y.; Ye, M.; Liu, Z. *ACS Catalysis* **2015**, *5*, 1922–1938.

(16) Li, Z.; Wang, J.; Qu, Y.; Liu, H.; Tang, C.; Miao, S.; Feng, Z.; An, H.; Li, C. *ACS Catalysis* **2017**, *7*, 8544–8548.

(17) Gao, P.; Dang, S.; Li, S.; Bu, X.; Liu, Z.; Qiu, M.; Yang, C.; Wang, H.; Zhong, L.; Han, Y.; Liu, Q.; Wei, W.; Sun, Y. *ACS Catalysis* **2017**, *8*, 571–578.

(18) De Luna, P.; Hahn, C.; Higgins, D.; Jaffer, S. A.; Jaramillo, T. F.; Sargent, E. H. *Science* **2019**, *364*, eaav3506.

(19) Keim, W. *Angewandte Chemie International Edition* **2013**, *52*, 12492–12496.

(20) Finiels, A.; Fajula, F.; Hulea, V. *Catal. Sci. Technol.* **2014**, *4*, 2412–2426.

(21) Britovsek, G. J. P.; Malinowski, R.; McGuinness, D. S.; Nobbs, J. D.; Tomov, A. K.; Wadsley, A. W.; Young, C. T. *ACS Catalysis* **2015**, *5*, 6922–6925.

(22) Wang, W.-C.; Tao, L. *Renewable and Sustainable Energy Reviews* **2016**, *53*, 801–822.

(23) Xie, S.; Li, Z.; Luo, S.; Zhang, W. *Renewable and Sustainable Energy Reviews* **2024**, *192*, 114240.

(24) Lancer, A. W. *Journal of Macromolecular Science: Part A - Chemistry* **1970**, *4*, 775–787.

(25) Meiswinkel, A.; Woehl, A.; Mueller, W.; Boelt, H. Developing linear-alpha-olefins technology. From laboratory to a commercial plant, 2011.

(26) Bonneviot, L.; Olivier, D.; Che, M. *Journal of Molecular Catalysis* **1983**, *21*, 415–430.

(27) Peuckert, M.; Keim, W. *Journal of Molecular Catalysis* **1984**, *22*, 289–295.

(28) Elev, I. *Journal of Catalysis* **1984**, *89*, 470–477.

(29) Nesterov, G. A.; Zakharov, V. A.; Fink, G.; Fenzl, W. *Journal of Molecular Catalysis* **1991**, *66*, 367–372.

(30) Braca, G.; Di Girolamo, M.; Galletti, A. R.; Sbrana, G.; Brunelli, M.; Bertolini, G. *Journal of Molecular Catalysis* **1992**, *74*, 421–431.

(31) Espinoza, R.; Korf, C.; Nicolaides, C.; Snel, R. *Applied Catalysis* **1987**, *29*, 175–184.

(32) Braca, G.; Raspolli Galletti, A.; Di Girolamo, M.; Sbrana, G.; Silla, R.; Ferrarini, P. *Journal of Molecular Catalysis A: Chemical* **1995**, *96*, 203–213.

(33) Nkosi, B.; Ng, F.; Rempel, G. *Applied Catalysis A: General* **1997**, *158*, 225–241.

(34) Zhang, Q.; Kantcheva, M.; Dalla Lana, I. G. *Industrial & Engineering Chemistry Research* **1997**, *36*, 3433–3438.

(35) Cai, T. *Catalysis Today* **1999**, *51*, 153–160.

(36) Preishuber-Pflugl, P.; Brookhart, M. *Macromolecules* **2002**, *35*, 6074–6076.

(37) Hulea, V.; Fajula, F. *Journal of Catalysis* **2004**, *225*, 213–222.

(38) McDaniel, M. P. In *Advances in Catalysis*; Elsevier: 2010, pp 123–606.

(39) McDaniel, M. *MRS Bulletin* **2013**, *38*, 234–238.

(40) Peacock, A., *Handbook of Polyethylene: Structures: Properties, and Applications*; CRC Press: 2000.

(41) Claverie, J. P.; Schaper, F. *MRS Bulletin* **2013**, *38*, 213–218.

(42) Zhang, C.-L.; Zhou, T.; Li, Y.-Q.; Lu, X.; Guan, Y.-B.; Cao, Y.-C.; Cao, G.-P. *Small* **2022**, *19*, 2205898.

(43) Temmerman, W.; Goeminne, R.; Rawat, K. S.; Van Speybroeck, V. *Advanced Materials* **2024**, 2412005.

(44) Lee, J.; Farha, O. K.; Roberts, J.; Scheidt, K. A.; Nguyen, S. T.; Hupp, J. T. *Chemical Society Reviews* **2009**, *38*, 1450.

(45) Corma, A.; García, H.; Llabrés i Xamena, F. X. *Chemical Reviews* **2010**, *110*, 4606–4655.

(46) Yang, D.; Gates, B. C. *ACS Catalysis* **2019**, *9*, 1779–1798.

(47) Hall, J. N.; Bollini, P. *Reaction Chemistry & Engineering* **2019**, *4*, 207–222.

(48) Bavykina, A.; Kolobov, N.; Khan, I. S.; Bau, J. A.; Ramirez, A.; Gascon, J. *Chemical Reviews* **2020**, *120*, 8468–8535.

(49) Rogge, S. M. J.; Bavykina, A.; Hajek, J.; Garcia, H.; Olivos-Suarez, A. I.; Sepúlveda-Escribano, A.; Vimont, A.; Clet, G.; Bazin, P.; Kapteijn, F.; Daturi, M.; Ramos-Fernandez, E. V.; Llabrés i Xamena, F. X.; Van Speybroeck, V.; Gascon, J. *Chemical Society Reviews* **2017**, *46*, 3134–3184.

(50) Huang, H.; Shen, K.; Chen, F.; Li, Y. *ACS Catalysis* **2020**, *10*, 6579–6586.

(51) Wei, Y.-S.; Zhang, M.; Zou, R.; Xu, Q. *Chemical Reviews* **2020**, *120*, 12089–12174.

(52) Abdel-Mageed, A. M.; Rungtaweevoranit, B. *Catalysis Today* **2024**, *439*, 114786.

(53) Kim, M.; Cohen, S. M. *CrystEngComm* **2012**, *14*, 4096–4104.

(54) Bai, Y.; Dou, Y.; Xie, L.-H.; Rutledge, W.; Li, J.-R.; Zhou, H.-C. *Chemical Society Reviews* **2016**, *45*, 2327–2367.

(55) Yuan, S.; Qin, J.-S.; Lollar, C. T.; Zhou, H.-C. *ACS Central Science* **2018**, *4*, 440–450.

(56) Ding, M.; Cai, X.; Jiang, H.-L. *Chemical Science* **2019**, *10*, 10209–10230.

(57) Healy, C.; Patil, K. M.; Wilson, B. H.; Hermanspahn, L.; Harvey-Reid, N. C.; Howard, B. I.; Kleinjan, C.; Kolien, J.; Payet, F.; Telfer, S. G.; Kruger, P. E.; Bennett, T. D. *Coordination Chemistry Reviews* **2020**, *419*, 213388.

(58) Cavka, J. H.; Jakobsen, S.; Olsbye, U.; Guillou, N.; Lamberti, C.; Bordiga, S.; Lillerud, K. P. *Journal of the American Chemical Society* **2008**, *130*, 13850–13851.

(59) Furukawa, H.; Gándara, F.; Zhang, Y.-B.; Jiang, J.; Queen, W. L.; Hudson, M. R.; Yaghi, O. M. *Journal of the American Chemical Society* **2014**, *136*, 4369–4381.

(60) Mondloch, J. E.; Bury, W.; Fairen-Jimenez, D.; Kwon, S.; DeMarco, E. J.; Weston, M. H.; Sarjeant, A. A.; Nguyen, S. T.; Stair, P. C.; Snurr, R. Q.; Farha, O. K.; Hupp, J. T. *Journal of the American Chemical Society* **2013**, *135*, 10294–10297.

(61) Mancuso, J. L.; Mroz, A. M.; Le, K. N.; Hendon, C. H. *Chemical Reviews* **2020**, *120*, 8641–8715.

(62) Hicks, K. E.; Rosen, A. S.; Syed, Z. H.; Snurr, R. Q.; Farha, O. K.; Notestein, J. M. *ACS Catalysis* **2020**, *10*, 14959–14970.

(63) Hicks, K. E.; Wolek, A. T. Y.; Farha, O. K.; Notestein, J. M. *ACS Catalysis* **2022**, *12*, 13671–13680.

(64) Katz, M. J.; Mondloch, J. E.; Totten, R. K.; Park, J. K.; Nguyen, S. T.; Farha, O. K.; Hupp, J. T. *Angewandte Chemie International Edition* **2013**, *53*, 497–501.

(65) Katz, M. J.; Moon, S.-Y.; Mondloch, J. E.; Beyzavi, M. H.; Stephenson, C. J.; Hupp, J. T.; Farha, O. K. *Chemical Science* **2015**, *6*, 2286–2291.

(66) Katz, M. J.; Klet, R. C.; Moon, S.-Y.; Mondloch, J. E.; Hupp, J. T.; Farha, O. K. *ACS Catalysis* **2015**, *5*, 4637–4642.

(67) Moon, S.-Y.; Liu, Y.; Hupp, J. T.; Farha, O. K. *Angewandte Chemie International Edition* **2015**, *54*, 6795–6799.

(68) Mondloch, J. E.; Katz, M. J.; Isley III, W. C.; Ghosh, P.; Liao, P.; Bury, W.; Wagner, G. W.; Hall, M. G.; DeCoste, J. B.; Peterson, G. W.; Snurr, R. Q.; Cramer, C. J.; Hupp, J. T.; Farha, O. K. *Nature Materials* **2015**, *14*, 512–516.

(69) Troya, D. *The Journal of Physical Chemistry C* **2016**, *120*, 29312–29323.

(70) Islamoglu, T.; Ortuño, M. A.; Proussaloglou, E.; Howarth, A. J.; Vermeulen, N. A.; Atilgan, A.; Asiri, A. M.; Cramer, C. J.; Farha, O. K. *Angewandte Chemie International Edition* **2018**, *57*, 1949–1953.

(71) Liao, Y.; Sheridan, T. R.; Liu, J.; Lu, Z.; Ma, K.; Yang, H.; Farha, O. K.; Hupp, J. T. *ACS Catalysis* **2023**, *14*, 437–448.

(72) Gibbons, B.; Johnson, E. M.; Javed, M. K.; Yang, X.; Morris, A. J. *ACS Applied Materials & Interfaces* **2024**, *16*, 52703–52711.

(73) Cirujano, F.; Corma, A.; Llabrés i Xamena, F. *Chemical Engineering Science* **2015**, *124*, 52–60.

(74) Cirujano, F.; Corma, A.; Llabrés i Xamena, F. *Catalysis Today* **2015**, *257*, 213–220.

(75) Valekar, A. H.; Cho, K.-H.; Chitale, S. K.; Hong, D.-Y.; Cha, G.-Y.; Lee, U.-H.; Hwang, D. W.; Serre, C.; Chang, J.-S.; Hwang, Y. K. *Green Chemistry* **2016**, *18*, 4542–4552.

(76) Kuwahara, Y.; Kango, H.; Yamashita, H. *ACS Sustainable Chemistry & Engineering* **2016**, *5*, 1141–1152.

(77) Rojas-Buzo, S.; García-García, P.; Corma, A. *ChemSusChem* **2017**, *11*, 432–438.

(78) Ouyang, W.; Zhao, D.; Wang, Y.; Balu, A. M.; Len, C.; Luque, R. *ACS Sustainable Chemistry & Engineering* **2018**, *6*, 6746–6752.

(79) Ma, Y.; Shan, Z.; Wang, X.; Zhao, W.; Gao, M.; Wang, K.; Ning, L.; Liu, X.; Yu, H. *ACS Applied Nano Materials* **2023**, *6*, 11772–11779.

(80) Vermoortele, F.; Vandichel, M.; Van de Voorde, B.; Ameloot, R.; Waroquier, M.; Van Speybroeck, V.; De Vos, D. E. *Angewandte Chemie International Edition* **2012**, *51*, 4887–4890.

(81) Vermoortele, F.; Bueken, B.; Le Bars, G.; Van de Voorde, B.; Vandichel, M.; Houthoofd, K.; Vimont, A.; Daturi, M.; Waroquier, M.; Van Speybroeck, V.; Kirschhock, C.; De Vos, D. E. *Journal of the American Chemical Society* **2013**, *135*, 11465–11468.

(82) Vandichel, M.; Hajek, J.; Vermoortele, F.; Waroquier, M.; De Vos, D. E.; Van Speybroeck, V. *CrystEngComm* **2015**, *17*, 395–406.

(83) Zhao, R.; Scott, T. R.; Schmid, J.; Löbbert, L.; Bermejo-Deval, R.; Liu, Y.; Gagliardi, L.; Neurock, M.; Lercher, J. A. *Journal of Catalysis* **2025**, *448*, 116204.

(84) Feng, Q.; Yang, C.; Chang, L.; Li, G.; Tang, Z. *Nano Research* **2024**, *17*, 8653–8670.

(85) Khosroshahi, N.; Doaee, S.; Safarifard, V.; Rostamnia, S. *Heliyon* **2024**, *10*, e31254.

(86) Yang, Y.; Liu, Z.; Cheng, R.; He, X.; Liu, B. *Organometallics* **2014**, *33*, 2599–2607.

(87) Samanta, P.; Solé-Daura, A.; Rajapaksha, R.; Wisser, F. M.; Meunier, F.; Schuurman, Y.; Sassoye, C.; Mellot-Draznieks, C.; Canivet, J. *ACS Catalysis* **2023**, *13*, 4193–4204.

(88) Siradze, S.; Poissonnier, J.; Håkonsen, S. F.; Frøseth, M.; Thorshaug, K.; Heyn, R. H.; Thybaut, J. W. *Chemical Engineering Journal* **2025**, *516*, 163794.

(89) Wu, M.-R.; Chen, J.-C.; Yeh, J.-Y.; Krisbiantoro, P. A.; Li, Y.-P.; Wu, K. C.-W. *Applied Catalysis B: Environment and Energy* **2026**, *382*, 125952.

(90) Kim, I. S. et al. *Angewandte Chemie* **2018**, *130*, 921–925.

(91) Ning, L.; Liao, S.; Liu, X.; Guo, P.; Zhang, Z.; Zhang, H.; Tong, X. *Journal of Catalysis* **2018**, *364*, 1–13.

(92) He, M.-Q.; Chang, X.-Y.; Li, H.-W.; Wu, Y. *International Journal of Molecular Sciences* **2024**, *25*, 6779.

(93) Li, Z.; Schweitzer, N. M.; League, A. B.; Bernales, V.; Peters, A. W.; Getsoian, A. B.; Wang, T. C.; Miller, J. T.; Vjunov, A.; Fulton, J. L.; Lercher, J. A.; Cramer, C. J.; Gagliardi, L.; Hupp, J. T.; Farha, O. K. *Journal of the American Chemical Society* **2016**, *138*, 1977–1982.

(94) Yeh, B.; Vicchio, S. P.; Chheda, S.; Zheng, J.; Schmid, J.; Löbbert, L.; Bermejo-Deval, R.; Gutiérrez, O. Y.; Lercher, J. A.; Lu, C. C.; Neurock, M.; Getman, R. B.; Gagliardi, L.; Bhan, A. *Journal of the American Chemical Society* **2021**, *143*, 20274–20280.

(95) Löbbert, L.; Chheda, S.; Zheng, J.; Khetrapal, N.; Schmid, J.; Zhao, R.; Gaggioli, C. A.; Camaioni, D. M.; Bermejo-Deval, R.; Gutiérrez, O. Y.; Liu, Y.; Siepmann, J. I.; Neurock, M.; Gagliardi, L.; Lercher, J. A. *Journal of the American Chemical Society* **2023**, *145*, 1407–1422.

(96) Yeh, B.; Chheda, S.; Zheng, J.; Schmid, J.; Löbbert, L.; Bermejo-Deval, R.; Gutiérrez, O. Y.; Lercher, J. A.; Gagliardi, L.; Bhan, A. *Catalysis Science & Technology* **2023**, *13*, 4213–4222.

(97) Zheng, J. et al. *Journal of Catalysis* **2022**, *413*, 176–183.

(98) Li, Z.; Peters, A. W.; Bernales, V.; Ortuño, M. A.; Schweitzer, N. M.; DeStefano, M. R.; Gallington, L. C.; Platero-Prats, A. E.; Chapman, K. W.; Cramer, C. J.; Gagliardi, L.; Hupp, J. T.; Farha, O. K. *ACS Central Science* **2016**, *3*, 31–38.

(99) Kusonjariyakun, N.; Santatiwongchai, J.; Meeprasert, J.; Chotpatiwetchkul, W.; Maihom, T.; Bureekaew, S.; Impeng, S. *Surfaces and Interfaces* **2024**, *55*, 105379.

(100) Goetjen, T. A.; Zhang, X.; Liu, J.; Hupp, J. T.; Farha, O. K. *ACS Sustainable Chemistry & Engineering* **2019**, *7*, 2553–2557.

(101) Goetjen, T. A.; Knapp, J. G.; Syed, Z. H.; Hackler, R. A.; Zhang, X.; Delferro, M.; Hupp, J. T.; Farha, O. K. *Catalysis Science & Technology* **2022**, *12*, 1619–1627.

(102) Goetjen, T. A.; Ferrandon, M. S.; Kropf, A. J.; Lamb, J. V.; Delferro, M.; Hupp, J. T.; Farha, O. K. *The Journal of Physical Chemistry C* **2022**, *126*, 20388–20394.

(103) Samadi, H.; Taghizadeh, M.; Habibpoor, H. *Microporous and Mesoporous Materials* **2025**, *398*, 113815.

(104) Abdel-Mageed, A. M.; Rungtaweevoranit, B.; Parlinska-Wojtan, M.; Pei, X.; Yaghi, O. M.; Behm, R. J. *Journal of the American Chemical Society* **2019**, *141*, 5201–5210.

(105) Yang, Y.; Zhang, X.; Kanchanakungwankul, S.; Lu, Z.; Noh, H.; Syed, Z. H.; Farha, O. K.; Truhlar, D. G.; Hupp, J. T. *Journal of the American Chemical Society* **2020**, *142*, 21169–21177.

(106) Abdel-Mageed, A. M.; Rungtaweevoranit, B.; Impeng, S.; Bansmann, J.; Rabeah, J.; Chen, S.; Häring, T.; Namuangrak, S.; Faungnawakij, K.; Brückner, A.; Behm, R. J. *Angewandte Chemie International Edition* **2023**, *62*, e202301920.

(107) Wu, Y.; Wang, H.; Peng, J.; Zhang, J.; Ding, M. *Chemical Engineering Journal* **2023**, *454*, 140156.

(108) Impeng, S.; Salaya-Gerónimo, E.; Kunkel, B.; Bartling, S.; Faungnawakij, K.; Rungtaweevoranit, B.; Abdel-Mageed, A. M. *Journal of Materials Chemistry A* **2024**, *12*, 3084–3095.

(109) Makhmutov, D.; Rungtaweevoranit, B.; Ahmed, A. A.; Faungnawakij, K.; Al-Yusufi, M.; Salaya, E.; Wohlrab, S.; Armbruster, U.; Abdel-Mageed, A. M. *Applied Surface Science* **2025**, *688*, 162323.

(110) Chauhan, M.; Rana, B.; Gupta, P.; Kalita, R.; Choudhary, A.; Manna, K. *Applied Catalysis B: Environment and Energy* **2025**, *377*, 125474.

(111) Zhang, J.; An, B.; Li, Z.; Cao, Y.; Dai, Y.; Wang, W.; Zeng, L.; Lin, W.; Wang, C. *Journal of the American Chemical Society* **2021**, *143*, 8829–8837.

(112) Nguyen, H. G. T.; Schweitzer, N. M.; Chang, C.-Y.; Drake, T. L.; So, M. C.; Stair, P. C.; Farha, O. K.; Hupp, J. T.; Nguyen, S. T. *ACS Catalysis* **2014**, *4*, 2496–2500.

(113) Yang, D.; Odoh, S. O.; Wang, T. C.; Farha, O. K.; Hupp, J. T.; Cramer, C. J.; Gagliardi, L.; Gates, B. C. *Journal of the American Chemical Society* **2015**, *137*, 7391–7396.

(114) Bernales, V.; League, A. B.; Li, Z.; Schweitzer, N. M.; Peters, A. W.; Carlson, R. K.; Hupp, J. T.; Cramer, C. J.; Farha, O. K.; Gagliardi, L. *The Journal of Physical Chemistry C* **2016**, *120*, 23576–23583.

(115) Ye, J.; Gagliardi, L.; Cramer, C. J.; Truhlar, D. G. *Journal of Catalysis* **2017**, *354*, 278–286.

(116) Sarma, B. B.; Maurer, F.; Doronkin, D. E.; Grunwaldt, J.-D. *Chemical Reviews* **2022**, *123*, 379–444.

(117) Sauer, J. *Journal of Catalysis* **2024**, *433*, 115482.

(118) Ortuño, M. A.; Bernales, V.; Gagliardi, L.; Cramer, C. J. *The Journal of Physical Chemistry C* **2016**, *120*, 24697–24705.

(119) Ye, J.; Gagliardi, L.; Cramer, C. J.; Truhlar, D. G. *Journal of Catalysis* **2018**, *360*, 160–167.

(120) Hajek, J.; Vandichel, M.; Van de Voorde, B.; Bueken, B.; De Vos, D.; Waroquier, M.; Van Speybroeck, V. *Journal of Catalysis* **2015**, *331*, 1–12.

(121) Caratelli, C.; Hajek, J.; Cirujano, F. G.; Waroquier, M.; Llabrés i Xamena, F. X.; Van Speybroeck, V. *Journal of Catalysis* **2017**, *352*, 401–414.

(122) Planas, N.; Mondloch, J. E.; Tussupbayev, S.; Borycz, J.; Gagliardi, L.; Hupp, J. T.; Farha, O. K.; Cramer, C. J. *The Journal of Physical Chemistry Letters* **2014**, *5*, 3716–3723.

(123) Lamaire, A.; Wieme, J.; Vandenhaute, S.; Goeminne, R.; Rogge, S. M. J.; Van Speybroeck, V. *Nature Communications* **2024**, *15*, 9997.

(124) Le, T.-H. T.; Ferro-Costas, D.; Fernández-Ramos, A.; Ortuño, M. A. *The Journal of Physical Chemistry C* **2024**, *128*, 1049–1057.

(125) Bursch, M.; Mewes, J.-M.; Hansen, A.; Grimme, S. *Angewandte Chemie International Edition* **2022**, *61*, e202205735.

(126) Teale, A. M. et al. *Physical Chemistry Chemical Physics* **2022**, *24*, 28700–28781.

(127) Ortuño, M. A.; Rellán-Piñeiro, M.; Luque, R. *ACS Sustainable Chemistry & Engineering* **2022**, *10*, 3567–3573.

(128) Mardirossian, N.; Head-Gordon, M. *Molecular Physics* **2017**, *115*, 2315–2372.

(129) Goncalves, T. J.; Plessow, P. N.; Studt, F. *ChemCatChem* **2019**, *11*, 4368–4376.

(130) Plessow, P. N.; Studt, F. *The Journal of Physical Chemistry Letters* **2020**, *11*, 4305–4310.

(131) Dohrmann, N.; King, D. S.; Gaggioli, C. A.; Gagliardi, L. *The Journal of Physical Chemistry C* **2023**, *127*, 16891–16900.

(132) Addicoat, M. A.; Vankova, N.; Akter, I. F.; Heine, T. *Journal of Chemical Theory and Computation* **2014**, *10*, 880–891.

(133) Coupry, D. E.; Addicoat, M. A.; Heine, T. *Journal of Chemical Theory and Computation* **2016**, *12*, 5215–5225.

(134) Yang, Y.; Ibikunle, I. A.; Sava Gallis, D. F.; Sholl, D. S. *ACS Applied Materials & Interfaces* **2022**, *14*, 54101–54110.

(135) Bureekaew, S.; Amirjalayer, S.; Tafipolsky, M.; Spickermann, C.; Roy, T. K.; Schmid, R. *physica status solidi (b)* **2013**, *250*, 1128–1141.

(136) Eckhoff, M.; Behler, J. *Journal of Chemical Theory and Computation* **2019**, *15*, 3793–3809.

(137) Achar, S. K.; Wardzala, J. J.; Bernasconi, L.; Zhang, L.; Johnson, J. K. *Journal of Chemical Theory and Computation* **2022**, *18*, 3593–3606.

(138) Yu, Y.; Zhang, W.; Mei, D. *The Journal of Physical Chemistry C* **2022**, *126*, 1204–1214.

(139) Vandenhaute, S.; Cools-Ceuppens, M.; DeKeyser, S.; Verstraelen, T.; Van Speybroeck, V. *npj Computational Materials* **2023**, *9*, 19.

(140) Batatia, I. et al. A foundation model for atomistic materials chemistry, 2024.

(141) Wood, B. M. et al. UMA: A Family of Universal Models for Atoms, 2025.

(142) Zhao, Y.; Truhlar, D. G. *Physical Chemistry Chemical Physics* **2008**, *10*, 2813.

(143) Ribeiro, R. F.; Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *The Journal of Physical Chemistry B* **2011**, *115*, 14556–14562.

(144) Grimme, S. *Chemistry – A European Journal* **2012**, *18*, 9955–9964.

(145) Otlyotov, A. A.; Minenkov, Y. *Journal of Computational Chemistry* **2023**, *44*, 1807–1816.

(146) Sprowl, L. H.; Campbell, C. T.; Árnadóttir, L. *The Journal of Physical Chemistry C* **2016**, *120*, 9719–9731.

(147) Ye, J.; Truhlar, D. G. *Journal of the American Chemical Society* **2022**, *144*, 12850–12860.

(148) Lei, Y.; Liu, L.; Zhang, E. *Journal of Chemical Theory and Computation* **2024**, *21*, 46–57.

(149) Piccini, G.; Sauer, J. *Journal of Chemical Theory and Computation* **2013**, *9*, 5038–5045.

(150) Piccini, G.; Sauer, J. *Journal of Chemical Theory and Computation* **2014**, *10*, 2479–2487.

(151) Piccini, G.; Alessio, M.; Sauer, J.; Zhi, Y.; Liu, Y.; Kolvenbach, R.; Jentys, A.; Lercher, J. A. *The Journal of Physical Chemistry C* **2015**, *119*, 6128–6137.

(152) Waitt, C.; Miles, A. R.; Schneider, W. F. *The Journal of Physical Chemistry C* **2021**, *125*, 20331–20342.

(153) Jørgensen, M.; Grönbeck, H. *The Journal of Physical Chemistry C* **2017**, *121*, 7199–7207.

(154) Jørgensen, M.; Chen, L.; Grönbeck, H. *The Journal of Physical Chemistry C* **2018**, *122*, 20351–20357.

(155) Bajpai, A.; Mehta, P.; Frey, K.; Lehmer, A. M.; Schneider, W. F. *ACS Catalysis* **2018**, *8*, 1945–1954.

(156) Blöndal, K.; Sargsyan, K.; Bross, D. H.; Ruscic, B.; Goldsmith, C. F. *The Journal of Physical Chemistry C* **2021**, *125*, 20249–20260.

(157) Blöndal, K.; Sargsyan, K.; Bross, D. H.; Ruscic, B.; Goldsmith, C. F. *ACS Catalysis* **2022**, *13*, 19–32.

(158) Blöndal, K.; Badger, K.; Sargsyan, K.; Bross, D. H.; Ruscic, B.; Goldsmith, C. F. *Physical Chemistry Chemical Physics* **2024**, *26*, 17265–17273.

(159) Alexopoulos, K.; Lee, M.-S.; Liu, Y.; Zhi, Y.; Liu, Y.; Reyniers, M.-F.; Marin, G. B.; Glezakou, V.-A.; Rousseau, R.; Lercher, J. A. *The Journal of Physical Chemistry C* **2016**, *120*, 7172–7182.

(160) Amsler, J.; Plessow, P. N.; Studt, F.; Bučko, T. *Journal of Chemical Theory and Computation* **2021**, *17*, 1155–1169.

(161) Chipot, C.; Pohorille, A., *Free energy calculations*; Chipot, C., Pohorille, A., Eds.; Springer series in chemical physics; Springer: Berlin, Germany, 2007.

(162) Straatsma, T. P.; Berendsen, H. J. C. *The Journal of Chemical Physics* **1988**, *89*, 5876–5886.

(163) De Oliveira, C. A. F.; Hamelberg, D.; McCammon, J. A. *Journal of Chemical Theory and Computation* **2008**, *4*, 1516–1525.

(164) Amsler, J.; Plessow, P. N.; Studt, F.; Bučko, T. *Journal of Chemical Theory and Computation* **2023**, *19*, 2455–2468.

(165) Plessow, P. N. *Journal of Chemical Theory and Computation* **2025**, *21*, 8518–8532.

(166) Jinnouchi, R.; Karsai, F.; Kresse, G. *Physical Review B* **2020**, *101*, 060201.

(167) Jinnouchi, R.; Karsai, F.; Verdi, C.; Kresse, G. *The Journal of Chemical Physics* **2021**, *154*, 094107.

(168) Grabowski, B.; Ikeda, Y.; Srinivasan, P.; Körmann, F.; Freysoldt, C.; Duff, A. I.; Shapeev, A.; Neugebauer, J. *npj Computational Materials* **2019**, *5*, 80.

(169) Cheng, B.; Engel, E. A.; Behler, J.; Dellago, C.; Ceriotti, M. *Proceedings of the National Academy of Sciences* **2019**, *116*, 1110–1115.

(170) Jaykhedkar, N.; Bystrický, R.; Sýkora, M.; Bučko, T. *The Journal of Chemical Physics* **2024**, *160*, 014710.

(171) Liakos, D. G.; Sparta, M.; Kesharwani, M. K.; Martin, J. M. L.; Neese, F. *Journal of Chemical Theory and Computation* **2015**, *11*, 1525–1539.

(172) Hohenberg, P.; Kohn, W. *Physical Review* **1964**, *136*, B864–B871.

(173) Jinnouchi, R.; Karsai, F.; Kresse, G. *Physical Review B* **2019**, *100*, 014105.

(174) Jinnouchi, R.; Lahnsteiner, J.; Karsai, F.; Kresse, G.; Bokdam, M. *Physical Review Letters* **2019**, *122*, 225701.

(175) Jinnouchi, R.; Miwa, K.; Karsai, F.; Kresse, G.; Asahi, R. *The Journal of Physical Chemistry Letters* **2020**, *11*, 6946–6955.

(176) Deringer, V. L.; Caro, M. A.; Csányi, G. *Advanced Materials* **2019**, *31*, 1902765.

(177) Noé, F.; Tkatchenko, A.; Müller, K.-R.; Clementi, C. *Annual Review of Physical Chemistry* **2020**, *71*, 361–390.

(178) Zuo, Y.; Chen, C.; Li, X.; Deng, Z.; Chen, Y.; Behler, J.; Csányi, G.; Shapeev, A. V.; Thompson, A. P.; Wood, M. A.; Ong, S. P. *The Journal of Physical Chemistry A* **2020**, *124*, 731–745.

(179) Behler, J. *Chemical Reviews* **2021**, *121*, 10037–10072.

(180) Deringer, V. L.; Bartók, A. P.; Bernstein, N.; Wilkins, D. M.; Ceriotti, M.; Csányi, G. *Chemical Reviews* **2021**, *121*, 10073–10141.

(181) Mishin, Y. *Acta Materialia* **2021**, *214*, 116980.

(182) Ko, T. W.; Ong, S. P. *Nature Computational Science* **2023**, *3*, 998–1000.

(183) Li, H.; Jiao, Y.; Davey, K.; Qiao, S.-Z. *Angewandte Chemie International Edition* **2023**, *62*, e202216383.

(184) Wan, K.; He, J.; Shi, X. *Advanced Materials* **2023**, *36*, 2305758.

(185) Duignan, T. T. *ACS Physical Chemistry Au* **2024**, *4*, 232–241.

(186) Tang, D.; Ketkaew, R.; Luber, S. *Chemistry – A European Journal* **2024**, *30*, e202401148.

(187) Jacobs, R. et al. *Current Opinion in Solid State and Materials Science* **2025**, *35*, 101214.

(188) Kresse, G.; Hafner, J. *Physical Review B* **1993**, *47*, 558–561.

(189) Kresse, G.; Hafner, J. *Physical Review B* **1994**, *49*, 14251–14269.

(190) Kresse, G.; Furthmüller, J. *Computational Materials Science* **1996**, *6*, 15–50.

(191) Kresse, G.; Furthmüller, J. *Physical Review B* **1996**, *54*, 11169–11186.

(192) Lallemand, M.; Rusu, O. A.; Dumitriu, E.; Finiels, A.; Fajula, F.; Hulea, V. *Applied Catalysis A: General* **2008**, *338*, 37–43.

(193) Lallemand, M.; Finiels, A.; Fajula, F.; Hulea, V. *Chemical Engineering Journal* **2011**, *172*, 1078–1082.

(194) Martínez, A.; Arribas, M. A.; Concepción, P.; Moussa, S. *Applied Catalysis A: General* **2013**, *467*, 509–518.

(195) Rajapaksha, R.; Samanta, P.; Quadrelli, E. A.; Canivet, J. *Chemical Society Reviews* **2023**, *52*, 8059–8076.

(196) Canivet, J.; Aguado, S.; Schuurman, Y.; Farrusseng, D. *Journal of the American Chemical Society* **2013**, *135*, 4195–4198.

(197) Liu, B.; Jie, S.; Bu, Z.; Li, B.-G. *RSC Adv.* **2014**, *4*, 62343–62346.

(198) Madrahimov, S. T.; Gallagher, J. R.; Zhang, G.; Meinhart, Z.; Garibay, S. J.; Delferro, M.; Miller, J. T.; Farha, O. K.; Hupp, J. T.; Nguyen, S. T. *ACS Catalysis* **2015**, *5*, 6713–6718.

(199) Gonzalez, M. I.; Oktawiec, J.; Long, J. R. *Faraday Discussions* **2017**, *201*, 351–367.

(200) Zhang, N.; Jin, R.; Mao, G.; Tan, J.; Chen, L.; Li, C.; Wang, J. *Inorganica Chimica Acta* **2022**, *531*, 120674.

(201) Tan, J.; Zhang, N.; Wang, L.; Chen, L.; Li, C.; Wang, J. *Journal of Coordination Chemistry* **2023**, *76*, 456–466.

(202) Song, L.; Chen, L.; Sun, J.; Wang, L.; Li, M.; Cai, Z. *ACS Catalysis* **2025**, *15*, 12590–12597.

(203) Chen, C.; Alalouni, M. R.; Dong, X.; Cao, Z.; Cheng, Q.; Zheng, L.; Meng, L.; Guan, C.; Liu, L.; Abou-Hamad, E.; Wang, J.; Shi, Z.; Huang, K.-W.; Cavallo, L.; Han, Y. *Journal of the American Chemical Society* **2021**, *143*, 7144–7153.

(204) Chen, C.; Meng, L.; Alalouni, M. R.; Dong, X.; Wu, Z.-P.; Zuo, S.; Zhang, H. *Small* **2023**, *19*, 2301235.

(205) Metzger, E. D.; Brozek, C. K.; Comito, R. J.; Dincă, M. *ACS Central Science* **2016**, *2*, 148–153.

(206) Metzger, E. D.; Comito, R. J.; Wu, Z.; Zhang, G.; Dubey, R. C.; Xu, W.; Miller, J. T.; Dincă, M. *ACS Sustainable Chemistry & Engineering* **2019**, *7*, 6654–6661.

(207) Liu, J.; Ye, J.; Li, Z.; Otake, K.-i.; Liao, Y.; Peters, A. W.; Noh, H.; Truhlar, D. G.; Gagliardi, L.; Cramer, C. J.; Farha, O. K.; Hupp, J. T. *Journal of the American Chemical Society* **2018**, *140*, 11174–11178.

(208) Ortega, D. E.; Matute, R. A. *Catalysis Science & Technology* **2021**, *11*, 2422–2432.

(209) Bernales, V.; League, A. B.; Li, Z.; Schweitzer, N. M.; Peters, A. W.; Carlson, R. K.; Hupp, J. T.; Cramer, C. J.; Farha, O. K.; Gagliardi, L. *The Journal of Physical Chemistry C* **2017**, *121*, 11975–11975.

(210) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Physical Review Letters* **1996**, *77*, 3865–3868.

(211) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *The Journal of Chemical Physics* **2010**, *132*, 154104.

(212) Grimme, S.; Ehrlich, S.; Goerigk, L. *Journal of Computational Chemistry* **2011**, *32*, 1456–1465.

(213) Neese, F. *WIREs Computational Molecular Science* **2022**, *12*, e1606.

(214) Brogaard, R. Y.; Olsbye, U. *ACS Catalysis* **2016**, *6*, 1205–1214.

(215) Kozuch, S.; Shaik, S. *Accounts of Chemical Research* **2010**, *44*, 101–110.

(216) Tuma, C.; Sauer, J. *Phys. Chem. Chem. Phys.* **2006**, *8*, 3955–3965.

(217) Svelle, S.; Tuma, C.; Rozanska, X.; Kerber, T.; Sauer, J. *Journal of the American Chemical Society* **2008**, *131*, 816–825.

(218) Tuma, C.; Kerber, T.; Sauer, J. *Angewandte Chemie International Edition* **2010**, *49*, 4678–4680.

(219) Hansen, N.; Kerber, T.; Sauer, J.; Bell, A. T.; Keil, F. J. *Journal of the American Chemical Society* **2010**, *132*, 11525–11538.

(220) Rybicki, M.; Sauer, J. *Journal of the American Chemical Society* **2018**, *140*, 18151–18161.

(221) Avdoshin, A.; Wenzel, W.; Kozlowska, M. *Chemical Communications* **2025**, *61*, 13711–13714.

(222) Groppo, E.; Martino, G. A.; Piovano, A.; Barzan, C. *ACS Catalysis* **2018**, *8*, 10846–10863.

(223) Harvey, J. N.; Aschi, M. *Physical Chemistry Chemical Physics* **1999**, *1*, 5555–5563.

(224) Amsler, J. Development of Anharmonic Molecular Models and Simulation of Reaction Kinetics in Zeolite Catalysis, en, Open Access, Ph.D. Thesis, Karlsruher Institut für Technologie (KIT), 2023.

(225) Huynh, T.-N.; Mounssef, B.; Sharapa, D. I.; Studt, F.; Bučko, T. *The Journal of Physical Chemistry C* **2025**, *129*, 16742–16755.

(226) Habershon, S.; Manolopoulos, D. E. *Physical Chemistry Chemical Physics* **2011**, *13*, 19714.

(227) Moustafa, S. G.; Schultz, A. J.; Kofke, D. A. *The Journal of Chemical Physics* **2013**, *139*, 084105.

(228) Rossi, M.; Gasparotto, P.; Ceriotti, M. *Physical Review Letters* **2016**, *117*, 115702.

(229) Moustafa, S. G.; Schultz, A. J.; Kofke, D. A. *Journal of Chemical Theory and Computation* **2017**, *13*, 825–834.

(230) Moustafa, S. G.; Schultz, A. J.; Kofke, D. A. *The Journal of Chemical Physics* **2018**, *149*, 124109.

(231) Purohit, A.; Schultz, A. J.; Moustafa, S. G.; Errington, J. R.; Kofke, D. A. *Molecular Physics* **2018**, *116*, 3027–3041.

(232) Moustafa, S. G.; Purohit, A.; Schultz, A. J.; Kofke, D. A. *Computer Physics Communications* **2021**, *258*, 107554.

(233) Pitzer, K. S.; Gwinn, W. D. *The Journal of Chemical Physics* **1942**, *10*, 428–440.

(234) Isaacson, A. D.; Truhlar, D. G. *The Journal of Chemical Physics* **1981**, *75*, 4090–4094.

(235) Jensen, F., *Introduction to computational chemistry*, 2nd ed.; John Wiley & Sons: Nashville, TN, 2007.

(236) Zwanzig, R. W. *The Journal of Chemical Physics* **1954**, *22*, 1420–1426.

(237) Pohorille, A.; Jarzynski, C.; Chipot, C. *The Journal of Physical Chemistry B* **2010**, *114*, 10235–10253.

(238) Bučko, T.; Gešvandtnerová, M.; Rocca, D. *Journal of Chemical Theory and Computation* **2020**, *16*, 6049–6060.

(239) Herzog, B.; Chagas da Silva, M.; Casier, B.; Badawi, M.; Pascale, F.; Bučko, T.; Lebègue, S.; Rocca, D. *Journal of Chemical Theory and Computation* **2022**, *18*, 1382–1394.

(240) Levine, D. S. et al. The Open Molecules 2025 (OMol25) Dataset, Evaluations, and Models, 2025.

(241) Chehaibou, B.; Badawi, M.; Bučko, T.; Bazhirov, T.; Rocca, D. *Journal of Chemical Theory and Computation* **2019**, *15*, 6333–6342.

(242) Rey, J.; Badawi, M.; Rocca, D.; Chizallet, C.; Bučko, T. *Catalysis Science & Technology* **2024**, *14*, 5314–5323.

(243) Gasteiger, J.; Shuaibi, M.; Sriram, A.; Günnemann, S.; Ulissi, Z.; Zitnick, C. L.; Das, A. GemNet-OC: Developing Graph Neural Networks for Large and Diverse Molecular Simulation Datasets, 2022.

(244) Chanussot, L. et al. *ACS Catalysis* **2021**, *11*, 6059–6072.

(245) Bennett, C. H. *Journal of Computational Physics* **1976**, *22*, 245–268.

(246) König, G.; Bruckner, S.; Boresch, S. *Journal of Computational Chemistry* **2009**, *30*, 1712–1718.

(247) Shirts, M. R.; Chodera, J. D. *The Journal of Chemical Physics* **2008**, *129*, 124105.

(248) Collinge, G.; Yuk, S. F.; Nguyen, M.-T.; Lee, M.-S.; Glezakou, V.-A.; Rousseau, R. *ACS Catalysis* **2020**, *10*, 9236–9260.

(249) Tran, R. et al. *ACS Catalysis* **2023**, *13*, 3066–3084.

(250) Campbell, C. T.; Sellers, J. R. V. *Journal of the American Chemical Society* **2012**, *134*, 18109–18115.

(251) Campbell, C. T. *Accounts of Chemical Research* **2019**, *52*, 984–993.

(252) Mavrikakis, M. *Nature Materials* **2006**, *5*, 847–848.

(253) 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. *Nature Chemistry* **2009**, *1*, 552–556.

(254) Nørskov, J. K.; Abild-Pedersen, F.; Studt, F.; Bligaard, T. *Proceedings of the National Academy of Sciences* **2011**, *108*, 937–943.

(255) Studt, F. *Frontiers in Catalysis* **2021**, *1*, 658965.

(256) Liu, W.; Ruiz, V. G.; Zhang, G.-X.; Santra, B.; Ren, X.; Scheffler, M.; Tkatchenko, A. *New Journal of Physics* **2013**, *15*, 053046.

(257) Sheldon, C.; Paier, J.; Usvyat, D.; Sauer, J. *Journal of Chemical Theory and Computation* **2024**, *20*, 2219–2227.

(258) Shi, B. X.; Rosen, A. S.; Schäfer, T.; Grüneis, A.; Kapil, V.; Zen, A.; Michaelides, A. *Nature Chemistry* **2025**, DOI: 10.1038/s41557-025-01884-y.

(259) Sheldon, C.; Paier, J.; Sauer, J. *The Journal of Chemical Physics* **2021**, *155*, 174702.

(260) Nørskov, J. K.; Rossmeisl, J.; Logadottir, A.; Lindqvist, L.; Kitchin, J. R.; Bligaard, T.; Jónsson, H. *The Journal of Physical Chemistry B* **2004**, *108*, 17886–17892.

(261) Panchenko, A.; Koper, M. T. M.; Shubina, T. E.; Mitchell, S. J.; Roduner, E. *Journal of The Electrochemical Society* **2004**, *151*, A2016.

(262) Hyman, M. P.; Medlin, J. W. *The Journal of Physical Chemistry B* **2006**, *110*, 15338–15344.

(263) Qi, L.; Yu, J.; Li, J. *The Journal of Chemical Physics* **2006**, *125*, 054701.

(264) Wasileski, S. A.; Janik, M. J. *Physical Chemistry Chemical Physics* **2008**, *10*, 3613.

(265) Janik, M. J.; Taylor, C. D.; Neurock, M. *Journal of The Electrochemical Society* **2009**, *156*, B126.

(266) Duan, Z.; Wang, G. *Physical Chemistry Chemical Physics* **2011**, *13*, 20178.

(267) Van Spronsen, M. A.; Frenken, J. W. M.; Groot, I. M. N. *Chemical Society Reviews* **2017**, *46*, 4347–4374.

(268) Shubina, T. E.; Hartnig, C.; Koper, M. T. M. *Physical Chemistry Chemical Physics* **2004**, *6*, 4215.

(269) Thermodynamic integration - VASP Wiki - vasp.at, https://www.vasp.at/wiki/index.php/Thermodynamic_integration, [Accessed 24-07-2025].

(270) Wellendorff, J.; Lundgaard, K. T.; Møgelhøj, A.; Petzold, V.; Landis, D. D.; Nørskov, J. K.; Bligaard, T.; Jacobsen, K. W. *Physical Review B* **2012**, *85*, 235149.

(271) Andersen, H. C. *The Journal of Chemical Physics* **1980**, *72*, 2384–2393.

(272) Getman, R. B.; Xu, Y.; Schneider, W. F. *The Journal of Physical Chemistry C* **2008**, *112*, 9559–9572.

(273) De Morais, R. F.; Franco, A. A.; Sautet, P.; Loffreda, D. *Physical Chemistry Chemical Physics* **2015**, *17*, 11392–11400.

(274) Malek, A.; Eikerling, M. H. *Electrocatalysis* **2017**, *9*, 370–379.

(275) Gray, C. M.; Saravanan, K.; Wang, G.; Keith, J. A. *Molecular Simulation* **2017**, *43*, 420–427.

(276) Ford, D. C.; Nilekar, A. U.; Xu, Y.; Mavrikakis, M. *Surface Science* **2010**, *604*, 1565–1575.

(277) Cahyanto, W. T.; Escaño, M. C.; Kasai, H.; Arevalo, R. L. *e-Journal of Surface Science and Nanotechnology* **2011**, *9*, 352–356.

(278) Arumainayagam, C. R.; McMaster, M. C.; Schoofs, G. R.; Madix, R. J. *Surface Science* **1989**, *222*, 213–246.

(279) Oakes, D. J.; McCoustra, M. R. S.; Chesters, M. A. *Faraday Discussions* **1993**, *96*, 325.

(280) Ukraintsev, V. A.; Harrison, I. *The Journal of Chemical Physics* **1994**, *101*, 1564–1581.

(281) Yang, Q. Y.; Maynard, K. J.; Johnson, A. D.; Ceyer, S. T. *The Journal of Chemical Physics* **1995**, *102*, 7734–7749.

(282) Higgins, J.; Conjusteau, A.; Scoles, G.; Bernasek, S. L. *The Journal of Chemical Physics* **2001**, *114*, 5277–5283.

(283) Bisson, R.; Sacchi, M.; Dang, T. T.; Yoder, B.; Maroni, P.; Beck, R. D. *The Journal of Physical Chemistry A* **2007**, *111*, 12679–12683.

(284) Nave, S.; Jackson, B. *The Journal of Chemical Physics* **2009**, *130*, 054701.

(285) Tiwari, A. K.; Nave, S.; Jackson, B. *The Journal of Chemical Physics* **2010**, *132*, 134702.

(286) Karp, E. M.; Silbaugh, T. L.; Campbell, C. T. *Journal of the American Chemical Society* **2013**, *135*, 5208–5211.

(287) Ueta, H.; Chen, L.; Beck, R. D.; Colón-Dìaz, I.; Jackson, B. *Physical Chemistry Chemical Physics* **2013**, *15*, 20526.

(288) Guo, H.; Jackson, B. *The Journal of Chemical Physics* **2016**, *144*, 184709.

(289) Carey, S. J.; Zhao, W.; Frehner, A.; Campbell, C. T.; Jackson, B. *ACS Catalysis* **2017**, *7*, 1286–1294.

(290) Beck, R. D.; Maroni, P.; Papageorgopoulos, D. C.; Dang, T. T.; Schmid, M. P.; Rizzo, T. R. *Science* **2003**, *302*, 98–100.

(291) Kroes, G.-J. *The Journal of Physical Chemistry Letters* **2015**, *6*, 4106–4114.

(292) Nattino, F.; Migliorini, D.; Bonfanti, M.; Kroes, G.-J. *The Journal of Chemical Physics* **2016**, *144*, 044702.

(293) Kroes, G.-J. *Physical Chemistry Chemical Physics* **2021**, *23*, 8962–9048.

(294) Wellendorff, J.; Silbaugh, T. L.; Garcia-Pintos, D.; Nørskov, J. K.; Bligaard, T.; Studt, F.; Campbell, C. T. *Surface Science* **2015**, *640*, 36–44.

(295) Allen, M. P.; Tildesley, D. J., *Computer Simulation of Liquids*; Oxford University PressOxford: 2017.

(296) Scott, D. W., *Multivariate Density Estimation: Theory, Practice, and Visualization*; Wiley: 2015.

(297) Chen, Y.-C. *Biostatistics & Epidemiology* **2017**, *1*, 161–187.

A. Appendix

A.1. Machine Learned Force Field Training and Evaluation

This section outlines the strategy adopted in this work to construct high-quality training datasets and develop robust machine-learned force fields (MLFFs) for use in the MLFF-TI workflow. We employed a multi-stage on-the-fly training protocol involving several density functional theory (DFT)-based molecular dynamics (MD) simulations to train a dedicated MLFF for each system. For the MLFFs used to generate the results in Section 4.2.2.1 (hereafter prefixed as **MLFF400**), the procedure began with a 5-ps temperature-ramping MD simulation, during which the system was gradually heated from 200 K to 500 K. Subsequently, a series of NVT ensemble MD runs at 400 K, approximately 30% higher than the temperature of interest (300 K), were performed consecutively to iteratively retrain the MLFF. This multi-stage training approach serves multiple purposes. The initial temperature ramp and elevated training temperature expand the explored region of phase space, improving the model's robustness against thermal fluctuations. Additionally, the use of successive retraining runs enhances structural diversity by incorporating configurations corresponding to different local minima. As a result, the trained MLFF is capable of making reliable predictions across a wide range of configurations on the potential energy surface (PES)—a particularly valuable feature for catalytic systems, where multiple adsorption minima commonly exist on the surface.

The total training trajectory was ensured to be at least 50 ps for each system (except for the clean Pt slab, where a shorter trajectory sufficed). Each MLFF was trained on independent datasets, with the sole exception of **MLFF400–OH@Pt**, whose training set also incorporated configurations from the Pt and O@Pt systems to improve transferability.

Finally, all MLFFs were refitted with adjusted weights on energies and forces, controlled by the parameters ML_WTOTEN and ML_WTIFOR, respectively. The specific weighting schemes used for each system are listed in Tables A.1–A.4. For this series of force fields, our objective was to obtain highly accurate MLFFs optimized for simulations at 300 K. In particular, for the OH@Pt and OOH@Pt systems, where part of the anharmonicity originates from the adsorbate's rotational degrees of freedom, we systematically refitted the MLFFs using different energy-to-force weight ratios to identify those that best reproduced the rotational potential energy surface (PES). We found that ratios of 50:1 for OH@Pt and 1:1 for OOH@Pt yielded the most reasonable rotational PESs, while a 20:20 ratio provided sufficiently accurate MLFFs for the simpler O@Pt and clean Pt systems.

The final refitting step was performed using regularized singular value decomposition (SVD) in the fast-execution mode, which significantly accelerated production simulations while maintaining excellent predictive accuracy.

We additionally developed three high-temperature MLFFs for the Pt, O@Pt, and OH@Pt systems, which are prefixed as **MLFF1K**, by extending the sampling to include higher-energy regions of phase space. The training procedure closely followed that used for the **MLFF400** models, with two key modifications: (1) the temperature ramp was increased from 200 K to 1200 K, and (2) the subsequent NVT MD simulations were conducted at 1000 K.

These high-temperature MLFFs were employed to examine the temperature effects on adsorption free energy and entropy (presented in Section 4.2.2.4), as well was the sensitivity of the MLFF-TI results to the quality of the underlying force field (discussed in Section 4.1.8.3). For this set of models, a uniform weighting ratio of 20:20 for energies and forces was adopted during refitting to ensure consistency across systems.

The Langevin thermostat was employed for all training simulations[295], as it generates smoother trajectories, which is particularly beneficial during temperature ramping. The cutoff radius for both radial and angular descriptors was set to 5 Å, and these descriptors were represented using Gaussian basis functions with a width of 0.3 Å and 12 radial basis functions. During the constant-temperature retraining MD runs, the force threshold was defined as 90% of the average of all preceding Bayesian errors, thereby triggering additional DFT calculations and expanding the training dataset. All other training parameters were maintained at their default values.

Slightly different strategies were adopted for generating the training structures in each MLFF set. In particular, the force threshold parameter (ML_CTIFOR), which determines whether a DFT calculation is performed and a new configuration is added to the training dataset, was initialized at 0.02 at the start of each retraining run, thereby forcing more DFT evaluations and expanding the training set. In contrast, for the **MLFF1K** set, an adaptive thresholding scheme was applied and the final ML_CTIFOR value from the preceding training trajectory was reused for the subsequent retraining run, which resulted in fewer configurations being collected overall.

However, the total number of training configurations does not solely dictate the accuracy of a given MLFF. Rather, the number of basis functions associated with the constituent elements plays a more decisive role. In this regard, the **MLFF1K** models consistently contain more basis functions than the MLFF400 counterparts, even though their total number of configurations may be smaller (see Tables A.1–A.3).

The accuracy of the trained MLFFs was assessed using multiple quantitative metrics. To this end, a test dataset was constructed by randomly selecting 200 configurations from the 2.5–5 ps MLFF trajectories at 300 K, which were then re-evaluated at the DFT level. The first performance metric was the root-mean-square deviation (RMSD) between MLFF and DFT energies and forces across the test sets. As shown in Tables A.1–A.4, the MLFFs reproduced the DFT reference data with excellent agreement. Interestingly, MLFFs trained at elevated temperatures exhibited higher on-the-fly Bayesian errors during training, yet

their predictions for low-temperature configurations remained comparable in accuracy to those of models trained specifically within the lower temperature range. This trend is expected, as the configurational space at low temperatures is effectively a subset of that explored at higher temperatures, and MLFFs are well known for their strong interpolation capabilities. Similar findings have been reported previously[238].

The second evaluation criterion was the phase-space overlap between MLFF and DFT phase spaces, quantified by the I_w index[164, 238, 239]. As discussed in Section 4.1.8.3, the consistently high I_w values obtained indicate substantial overlap between the two phase spaces, confirming the reliability of the free-energy perturbation (FEP) corrections and, consequently, the robustness of the final free-energy estimates.

A.1.1. Details on the MLFF Used for Adsorption of Oxygen-Containing Intermediates

Table A.1.: Technical details of different MLFFs used for clean Pt(111) system

	MLFF400-Pt	**MLFF1K-Pt**
Number of training configurations	1049	750
Number of Pt basis sets	452	1303
ML_LMAX2	3	3
ML_WTOTEN	20.0	20.0
ML_WTIFOR	20.0	20.0
Energy RMSD (meV/atom)	0.29	0.41
Forces RMSD (eV/Å)	0.019	0.022

Table A.2.: Technical details of different MLFFs used for O@Pt(111) system

	MLFF400-OPt	**MLFF1K-OPt**
Number of training configurations	1807	4075
Number of Pt basis sets	781	2283
Number of O basis sets	364	1321
ML_LMAX2	3	3
ML_WTOTEN	20.0	20.0
ML_WTIFOR	20.0	20.0
Energy RMSD (meV/atom)	0.48	0.42
Forces RMSD (eV/Å)	0.024	0.024

Table A.3.: Technical details of the different MLFFs used for OH@Pt(111) system

	MLFF400-OHPt	**MLFF1K-OHPt**
Number of training configurations	8395	5069
Number of Pt basis sets	2130	3077
Number of O basis sets	2569	3873
Number of H basis sets	1872	3417
ML_LMAX2	3	3
ML_WTOTEN	50.0	20.0
ML_WTIFOR	1.0	20.0
Energy RMSD (meV/atom)	0.49	0.55
Forces RMSD (eV/Å)	0.032	0.031

Table A.4.: Technical details of the MLFF used for OOH@Pt(111) system

	MLFF400-OOHPt
Number of training configurations	4581
Number of Pt basis sets	1171
Number of O basis sets	3833
Number of H basis sets	1862
ML_LMAX2	3
ML_WTOTEN	1.0
ML_WTIFOR	1.0
Energy RMSD (meV/atom)	0.45
Forces RMSD (eV/Å)	0.033

A.1.2. Details on the MLFFs Used for Adsorptions of Methane

To examine the influence of functionals on adsorption free energy and entropy of methane adsorption on Pt(111), two MLFFs were trained for each relevant system at BEEF-vdW and PBE-D3(BJ) levels using similar procedure for training **MLFF1K** MLFFs. Table A.5- A.9 presents technical details on these MLFFs. The values of ML_LMAX2, ML_WTOTEN, and ML_WTIFOR were set to be 3, 20, and 20, respectively, for all MLFFs presented.

Table A.5.: Technical details of the different MLFFs used for clean Pt system

	BEEF-vdW	**PBE-D3**
Number of training configurations	750	853
Number of Pt basis sets	1303	772
Energy RMSD (meV/atom)	0.41	0.45
Forces RMSD (eV/Å)	0.022	0.020

Table A.6.: Technical details of the different MLFFs used for CH_4 gas system

	BEEF-vdW	**PBE-D3**
Number of training configurations	145	121
Number of C basis sets	65	72
Number of H basis sets	242	254
Energy RMSD (meV/atom)	0.11	0.17
Forces RMSD (eV/Å)	0.008	0.011

Table A.7.: Technical details of the different MLFFs used for H@Pt system

	BEEF-vdW	**PBE-D3**
Number of training configurations	4278	3617
Number of Pt basis sets	3640	2914
Number of H basis sets	3625	3130
Energy RMSD (meV/atom)	0.49	0.31
Forces RMSD (eV/Å)	0.024	0.021

Table A.8.: Technical details of the different MLFFs used for CH_3@Pt system

	BEEF-vdW	**PBE-D3**
Number of training configurations	3667	4454
Number of Pt basis sets	3456	6220
Number of C basis sets	1470	2949
Number of H basis sets	5375	6895
Energy RMSD (meV/atom)	0.40	0.29
Forces RMSD (eV/Å)	0.023	0.022

Table A.9.: Technical details of the different MLFFs used for CH_4@Pt system

	BEEF-vdW	**PBE-D3**
Number of training configurations	2014	2184
Number of Pt basis sets	1033	1215
Number of C basis sets	386	492
Number of H basis sets	2852	3597
Energy RMSD (meV/atom)	0.92	0.11
Forces RMSD (eV/Å)	0.025	0.023

A.2. Independence of Constrained Degrees of Freedom

In the Section 4.2.2.1, we have made use of the Eq. (4.15), which allows for the separation of probability distributions for the constraints, provided their mutual independence. Here,

we validate such an assumption of independence as used above for the analysis of probability distributions. To do this, we compare the results obtained from a two-dimensional probability distribution analysis to those from multiple one-dimensional counterparts and show that discrepancies are negligibly small. The validation results shown here are for the OH@Pt(111) system.

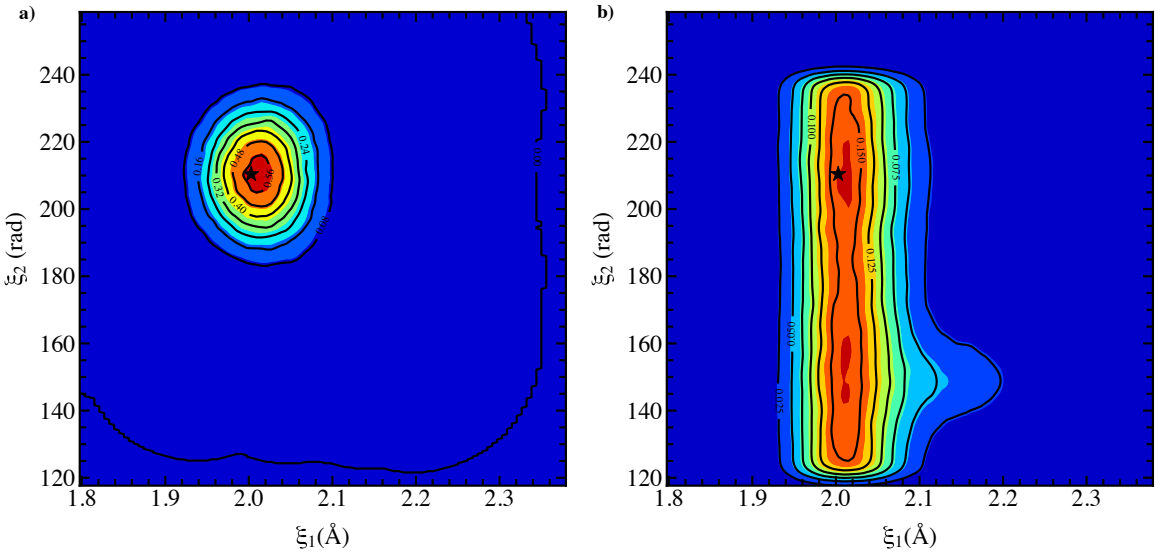


Figure A.1.: Two dimensional probability distributions for a) harmonic reference system ($\lambda = 0$); b) fully interacting system driven by MLFF ($\lambda = 1$). The stars mark the reference values at which the coordinates were fixed in constrained MD runs. In both plots, the isovalues were scaled by a factor of 1000 for clarity.

We constructed the two-dimensional probability densities of the two constrained degrees of freedom from 2.5 ns unconstrained trajectories at $\lambda = 0$ and $\lambda = 1$ (shown in Figure A.1). The probabilities of the system being in the constrained state are 0.000 60 $\mathrm{deg}^{-1}\mathrm{Å}^{-1}$ and 0.000 16 $\mathrm{deg}^{-1}\mathrm{Å}^{-1}$ for the harmonic ($P_0(\xi')$) and anharmonic ($P_1(\xi')$) systems, respectively. Considering the factor of 1/3 for rotational and of 1/9 for translational permutations of the OH adsorbate, this difference yields a contribution of −0.118 eV.

The cases for 1D probability analyses for individual coordinates constrained in TI are discussed in Section 4.2.2.1. It can be shown that the probability of the system being in the state with two specific values of geometric parameters ξ_1 and ξ_2 is indeed the product of the probabilities of the each degree of freedom assuming the corresponding value independently, *i.e.*, $P_\alpha(\xi') = P_\alpha(\xi'_1) \times P_\alpha(\xi'_2)$. Indeed, the probability products are computed to be 0.000 60 and 0.000 15 for $\lambda = 0$ and $\lambda = 1$, respectively. As a result, the sum of contributions from these two probability densities (Table 4.5) is identical to the value obtained via 2D analysis above.

A.3. Comparison to Different Analytical Approximations

In this section, we employ relevant analytical approximations to account for the free energy contribution of the OH's and OOH's translational and rotational modes. Since we could not locate a transition state for the "helicopter-like" rotation of OH adsorbate, we assume the rotation is barrierless and, thus, the one-dimensional (1D) free rotor approximation is employed. The partition function of a 1D free rotor is given as:

$$q^{\text{free rotor}} = \frac{\left(8\pi^3 I_{\text{int}}\, k_{\text{B}} T\right)^{1/2}}{\sigma_{\text{int}} h} \tag{A.1}$$

where k_{B} is the Boltzmann constant, σ_{int} is the reduced symmetry of the rotation, h is the Planck's constant, T is the temperature. For both OH and OOH adsorbates, the moments of inertia are calculated for the rotation around the axis perpendicular to the Pt surface and going through the terminal O atom, as:

$$I_{\text{int}} = \sum_i m_i r_i^2 \tag{A.2}$$

where m_i the mass of the adsorbate's atom i and r_i is the distance between atom i and the terminal O atom. The partition function for this rotation under quasi-classical harmonic approximation is:

$$q^{HA} = \frac{k_{\text{B}} T}{\hbar\omega}. \tag{A.3}$$

For treating the adsorbate's translations, three models are used. First, we adopt the Eqs. 36 and 37 in Ref [147] for constructing the partition function of a 2D ideal gas. Here

$$Q_{\text{xy}}^{\text{free}} = (Q_{\text{x}}^{\text{free}})^2 \tag{A.4}$$

where

$$Q_{\text{x}}^{\text{free}} = (2\pi A m_{\text{ads}} k_{\text{B}} T)^{1/2}/h \tag{A.5}$$

with A being the surface area, m_{ads} is the adsorbate's mass. For the hindered translator treatment, the square lattice approximation is used. That is, we assume the diffusion along the x direction is essentially the same as along the y direction. Therefore, the hindered translator partition function is calculated as:

$$Q_{\text{xy}} = \alpha\pi s_{\text{x}} J f_{\text{anh}} Q_{\text{xx}}^{\text{HO}} \tag{A.6}$$

where $Q_{\text{xx}}^{\text{HO}}$ is the harmonic oscillator approximated partition function.

$$s_{\text{x}} = W_{\text{x}}/k_{\text{B}} T \tag{A.7}$$

$$u_{\text{x}} = \frac{\hbar\omega}{k_{\text{B}} T} \tag{A.8}$$

$$\alpha = A/l^2 \tag{A.9}$$

$$J = e^{-s_x} I_0^2 \left(\frac{s_x}{2}\right) \tag{A.10}$$

$$f_{\text{anh}} = \exp\left(\frac{u_x^2}{u_x + 8s_x}\right) \tag{A.11}$$

Our vibrational analysis shows that the two hindered translational modes correspond to frequencies of 84 and 108 cm^{-1}. We assume now that both modes have the same frequency of 84 cm^{-1} for the simple implementation of hindered translator approximation. (This results in a 8 meV free energy underestimation in this approximation)

Model	**OH**				**OOH**	
	G_{trans}	ΔG_{trans}	G_{rot}	ΔG_{rot}	G_{rot}	ΔG_{rot}
HO	-0.048	0	-0.024	0	-0.018	0
HT/HR	-0.097	-0.049	-	-	-	-
FT/FR	-0.182	-0.134	-0.061	-0.037	-0.096	-0.078

Table A.10.: Translational and rotational free energies in eV estimated by different approximations. HO: harmonic oscillator, HT: hindered translator, HR: hindered rotor, FT: free translator, and FR: free rotor. ΔG values are the deviations with respect to HO values.

The computed translational and rotational free energies for OH and OOH adsorbates on Pt(111) surface are tabulated in Table A.10. Assuming the cancellation of anharmonic effects of surface atoms in clean and adsorbed systems, the anharmonic corrections for the adsorptions of OH and OOH species are −0.090 and −0.072 eV, respectively. The correction for OH is slightly larger (*i.e.,* more negative) than the correction predicted by the hindered-translator (HT) free-rotor (FR) approximation (−0.086 eV) and significantly smaller than the free-translator free-rotor value (−0.171 eV). Additionally, the better agreement between the TI and HTFR values allows us to provide a reasonable decomposition of the TI anharmonic correction shown in Figure 4.6, where we assume −0.037 eV comes from the 1D free rotation, while the rest (−0.053 eV) accounts for translations. Regarding the OOH adsorption, since no diffusion was observed along the simulated trajectory, it is expected that the anharmonic effects come predominantly from rotations. The TI correction agrees well with the 1D free rotor value (−0.078 eV), indicating that this simple model can be used to yield reasonably accurate free energy for this adsorption.

A.4. Diffusion of O on Pt(111)

We present in Figure A.2 the number of adsorption sites visited by atomic oxygen (O) on Pt(111) over a given period of time, as an estimate of the adsorbate's mobility as a function of temperature. This number ranges from 1 to 9 in our simulation model and is approximated by:

$$N_{\text{visited}} = \min\left(\frac{t}{k_{\text{diff}}} + 1, 9\right) \tag{A.12}$$

where t is the elapsed time, and

$$k_{\text{diff}} = \frac{k_{\text{B}}T}{h} \exp\left(\frac{-\Delta G_{\text{diff}}}{k_{\text{B}}T}\right) \tag{A.13}$$

is the diffusion rate of O, with ΔG_{diff} denoting the Gibbs free energy of diffusion at a given temperature T. The predicted values show excellent agreement with those obtained from explicit 2 ns MLFF-MD trajectories across the investigated temperature range.

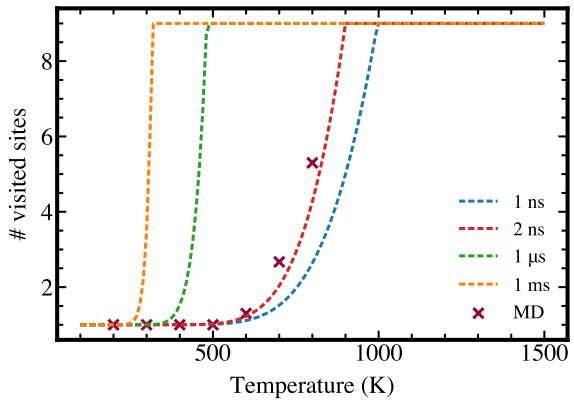


Figure A.2.: Number of adsorption sites visited by the atomic O atom as a function of temperature and time. Crosses are the results from the average of three 2-ns MLFF-MD trajectories at the corresponding temperature.

A.5. Uncertainty Quantification

A.5.1. Thermodynamic Integration Uncertainty

The numerical integration and its associated statistical uncertainties were evaluated using the same method as described in the Supporting Information of Ref.[160, 164].

A.5.2. Free Energy Perturbation Theory Uncertainty

The correction by FEPT is computed as:

$$\Delta A_{MLFF\rightarrow DFT} = A_{1,\text{DFT}} - A_{1,\text{MLFF}} = -k_B T \ln\left\langle \exp\left(-\frac{\mathcal{V}_{\text{DFT}} - \mathcal{V}_{\text{MLFF}}}{k_{\text{B}}T}\right)\right\rangle_{MLFF} \tag{A.14}$$

Using the uncertainty propagration through the logarithm, the uncertainty of the FEPT correction, $\sigma_{\Delta A}$, can be approximated as:

$$\sigma_{\Delta A} \approx \left|\frac{\partial \Delta A}{\partial \langle \exp \rangle} \sigma_{\exp}\right| \tag{A.15}$$

where $\frac{\partial \Delta A}{\partial \langle \exp \rangle}$ is the derivative of the entire expression with respect to the average exponential term and $\sigma_{\exp}$ is the uncertainty of the exponential term inside the logarithm. This derivative can be simplified to:

$$\frac{\partial \Delta A}{\partial \langle \exp \rangle} = -\frac{k_{\mathrm{B}} T}{\langle \exp \rangle} \tag{A.16}$$

Thus, the uncertainty in ΔA can be estimated as:

$$\sigma_{\Delta A} \approx \frac{k_B T}{\left\langle \exp\left(-\frac{\mathcal{V}_{\mathrm{DFT}} - \mathcal{V}_{\mathrm{MLFF}}}{k_{\mathrm{B}} T}\right)\right\rangle} \sigma_{\exp}. \tag{A.17}$$

A.5.3. Kernel Density Estimation Uncertainty

In the constrained TI formalism, the free energy difference between two states α and β can be calculated as:

$$\Delta A_{\alpha \to \beta} = \Delta A_{\alpha \to \beta}(\xi') - k_{\mathrm{B}} T \ln \left[\frac{P_\alpha(\xi')}{P_\beta(\xi')}\right], \tag{A.18}$$

where $\Delta A_{\alpha \to \beta}(\xi')$ is the free energy difference between system α and β with $\xi(\boldsymbol{x})$ fixed at ξ'. The latter term is computed by analyzing the probability distribution of $\xi(\boldsymbol{x})$ in straightforward MD simulations at the two states. Gaussian kernel density estimates (KDEs) are used to estimate the density $\hat{p}(\boldsymbol{x})$ from the geometric data, from which probabilities are calculated. For this, we have used the `gaussian_kde` class from `scipy.stats` in which the bandwidths of gaussians are determined via Scott's Rule[296]. Here, we assume that the trajectories from which geometric data are extracted are sufficiently long (2.5 to 5 ns each) that the uncertainties from sampled probability density are negligible. As such, another source of uncertainties comes from the approximations of the density. For that, we employed the plug-in approach[297] to estimate the 95% confidence interval of $p(\boldsymbol{x})$ as follows:

$$C_{1-\alpha,\mathrm{PI}}(\boldsymbol{x}) = \left[\widehat{p}_n(\boldsymbol{x}) - z_{1-\alpha/2}\sqrt{\frac{\mu_K \cdot \widehat{p}_n(\boldsymbol{x})}{nh^d}}, \quad \widehat{p}_n(x) + z_{1-\alpha/2}\sqrt{\frac{\mu_K \cdot \widehat{p}_n(\boldsymbol{x})}{nh^d}}\right], \tag{A.19}$$

where n is the number of data points, h is the bin's width, d is the dimension of the data (which is 1 for data present in the main text), $\widehat{p}_n(\boldsymbol{x})$ is the estimator at $\boldsymbol{x}$, and $z_{1-\alpha/2}$ is the z-score corresponding to the desired confidence value $1-\alpha$, where $\alpha = 0.05$. From Eq. (A.19), the corresponding standard error of the estimation is obtained as:

$$\sigma_{\widehat{p}(\boldsymbol{x})} = \frac{C^{\mathrm{upper}}_{1-\alpha,\mathrm{PI}}(\boldsymbol{x}) - C^{\mathrm{lower}}_{1-\alpha,\mathrm{PI}}(\boldsymbol{x})}{2z_{1-\alpha/2}} = \sqrt{\frac{\mu_K \cdot \widehat{p}_n(\boldsymbol{x})}{nh^d}}. \tag{A.20}$$

Following the error propagation formula for a function of two variables with uncertainties:

$$\sigma_f^2 = \left(\frac{\partial f}{\partial P_0}\sigma_0\right)^2 + \left(\frac{\partial f}{\partial P_1}\sigma_1\right)^2 \tag{A.21}$$

with $f = -k_\mathrm{B}T \ln \frac{P_0(\xi')}{P_1(\xi')}$, σ_0 and σ_1 being uncertainties of $P_0(\xi')$ and $P_1(\xi')$, respectively, determined via Eq. (A.20). We have:

$$\sigma_f^2 = (k_B T)^2 \left(\frac{\sigma_0^2}{P_0^2} + \frac{\sigma_1^2}{P_1^2} \right). \tag{A.22}$$

Finally, the uncertainty of f is given as:

$$\sigma_f = k_B T \sqrt{\frac{\sigma_0^2}{P_0^2} + \frac{\sigma_1^2}{P_1^2}}. \tag{A.23}$$

We performed the uncertainty quantification for all the kernel density estimations using aforementioned and obtained negligibly small uncertainties (*e.g.,* 0.4 meV for OH@Pt system). This can be explained by the very large numbers of data points used for the estimations (5-10 million points). Therefore, we neglect these uncertainties and assume $\sigma = 0$ for the free energy contributions from the probability analyses, as seen in Table 2 in the main text.

A.6. Influence of Hard-Sphere Settings

Table A.11.: Contributions to adsorption free energies at MLFF level, $\Delta A_{\mathrm{ads,MLFF}}$, in eV, the physisorption of CH_4 on Pt(111) surface calculated at PBE-D3 level at 300 K, using various choices of hard-sphere settings.

Quantity	$r_{\min}$	$2 \times r_{\min}$	10 Å $< z <$ 20 Å	
	MD	MD	MD	analytical
$\Delta A_{0\to1}$	0.0195	0.0382	0.0250	0.0204
$\Delta A_{1\to2}$ (fw, exp)	-0.2010	-0.1679	-0.1946	
$\Delta A_{1\to2}$ (bw, exp)	-0.2066	-0.1684	-0.1940	
$\Delta A_{1\to2}$ (BAR)	-0.2019	-0.1680	-0.1939	
$\Delta A_{2\to3}$	0.0000	-0.0578	-0.0186	
$\Delta A_{\mathrm{ads,MLFF}}$	-0.1824	-0.1876	-0.1875	-0.1921

Table A.11 summarizes the contributions to the adsorption free energy at the MLFF-level, $\Delta A_{\mathrm{ads,MLFF}}$, for methane physisorption on Pt(111) at the PBE-D3 level and at 300 K, obtained using the MLFF-IHSM method with different hard-sphere settings. These settings include the use of Eq. (4.36) to define $\mathcal{V}_{\mathrm{IHSM}}$, where r_{ij}^{hard} corresponds to either the minimum atomic distances between Pt–C and Pt–H atoms obtained from a fully interacting MLFF-MD trajectory at 300 K and at the PBE-D3 level ($r_{\min}$; 2.6 Å and 1.8 Å, respectively), or twice these values ($2 \times r_{\min}$; 5.2 Å and 3.6 Å, respectively). In addition, Eq. (4.38) was also used to define $\mathcal{V}_{\mathrm{IHSM}}$, with parameters $h_{\min} = 10$Å and $h_{\max} = 20$Å.

The resulting free energies are consistent across all settings, with deviations within 0.01 eV. These minor differences are likely attributable to the statistical uncertainty of $\Delta A_{0\to1}$

arising from finite sampling. For example, in the case of 10 Å $< z <$ 20 Å, the MD sampling slightly overestimates the $\Delta A_{0\to1}$ value by approximately 0.005 eV relative to the analytical result.

A.7. Entropy of Individual Systems in the Dissociative Adsorption of CH_4 on Pt(111)

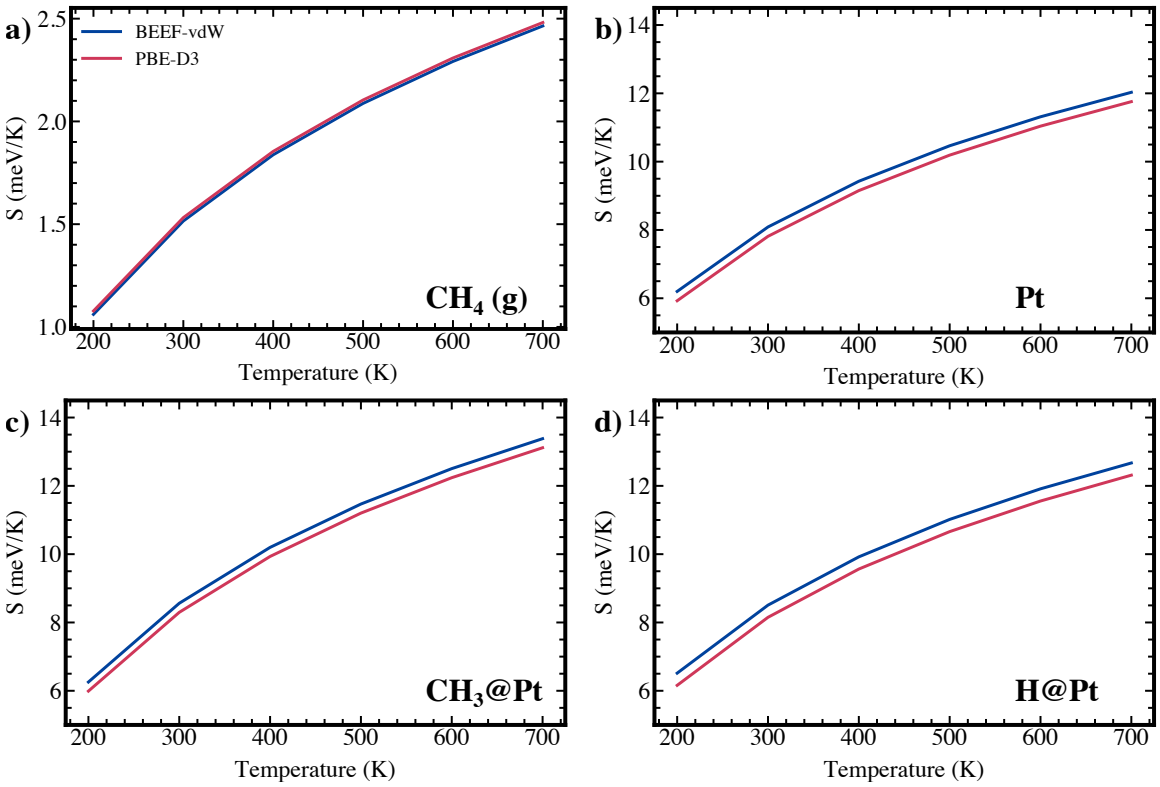


Figure A.3.: The entropy estimated by harmonic approximation for a) gas-phase CH_4 molecule, b) the clean Pt(111) surface, c) the CH_3@Pt(111) system, and d) the H@Pt(111) system, obtained at BEEF-vdW and PBE-D3 levels of theory.

B. Acknowledgments

First and foremost, I would like to express my deepest gratitude to Prof. Felix Studt for giving me the opportunity to join his group, for his continuous guidance, and for the freedom and trust he placed in me to explore my own ideas. His insightful advice, encouragement, and scientific curiosity have been a constant source of inspiration throughout my research journey. Special thanks to my co-supervisor Prof. Karin Fink for her valuable advice and for many fruitful discussions.

I owe special thanks to my mentor (and board game rival), Dr. Dmitry I. Sharapa, for his unwavering support - both scientific and personal. His patience, humor, and wisdom made even the most challenging days enjoyable, and I am truly grateful for all the time and effort he has invested in me.

I have been incredibly fortunate to collaborate with Prof. Dr. Tomáš Bučko, whose enthusiasm, constructive feedback, and deep knowledge have enriched my work in countless ways. I am also grateful to Dr. Jonas Amsler for his patient introduction to the TI method in the very first days. Furthermore, I would like to thank Dr. Bassim Mounssef Jr. for being the right person at the right time to join me on the TI adventure, and Dr. Philipp N. Pleßow for his insightful discussions and valuable perspectives.

My sincere appreciation goes to all collaborators of the P3 project within GRK2450. I would like to thank Nikita Matsokin, Aleksandr Avdoshin, Prof. Dr. Karin Fink, Prof. Dr. Wolfgang Wenzel, and Dr. Mariana Kozlowska for the inspiring teamwork and stimulating scientific exchange. I also gratefully acknowledge the financial support from the DFG through GRK2450, which made this research possible. Additionally, I am very appreciated the helpful support from Sabine Holthoff.

I would also like to acknowledge and thank the rest of the Studt's group, past and present: Dr. Jelena Jelic, Dr. Eduard Araujo-Lopez, Dr. Amir H. Hakimioun, Dr. Zha Shenjun, Dr. Lucas Spiske, Dr. Philipp Huber, Dr. Sarah Bernart, Dr. Chen Jiachen, Dr. Aleksandr Kramarenko, Dr. Héctor Vicente, Claude Coppex, Annika Enß, Aleksandr Maliugin, Enrico Sireci, Liana Savintseva, Aliena Winterbauer, for the great discussions, laughter, and the sense of camaraderie that made the workplace pleasant and productive. Particularly, I thank Annika Enß for proofreading the Zusammenfassung of this dissertation.

I am deeply grateful to Dr. Marjan Krstić and the Habitzreither family for their warmth, hospitality, and for making me feel at home in Germany from the very beginning.

Finally, my heartfelt thanks go to my beloved wife. Thank you for your love, patience, and understanding — for standing by me through every challenge, and for giving me the space

to chase my curiosity wherever it led. I'm also thankful to my parents: even when you didn't fully understand what I was doing, you never stopped me from doing what I love. For that, I am endlessly grateful.

C. List of Publications

List of Publications from the Dissertation

1. **Huynh, T. N.**; Mounssef, B.; Sharapa, D. I.; Studt, F.; Bučko, T. Anharmonic correction to the adsorption free energy of oxygen-containing intermediates on Pt(111) by machine-learned force field-based thermodynamic integrations. *The Journal of Physical Chemistry C* **2025**, *129*, 16742–16755.
2. **Huynh, T. N.***; Matsokin, N.*; Studt, F.; Fink, K.; Sharapa, D. I. Ethylene dimerization on Ni-supported NU-1000: A comprehensive theoretical study. *The Journal of Physical Chemistry C* **2025**, *under Revision.* (* indicates equal contribution)
3. Avdoshin, A.; Matsokin, N. A.; **Huynh, T. N.**; Kozlowska, M.; Sharapa, D. I.; Fink, K.; Studt, F; Wenzel, W. Multiscale Kinetic Modeling of Ethylene Oligomerization in Ni–NU-1000 Single-Atom Catalysts. **2025**, *in Preparation*
4. Mounssef Jr., B.; **Huynh, T. N.**; Sharapa, D. I.; Pleßow, P. N.; Bučko, T.; Studt, F. Temperature trends in anharmonic corrections of methanol adsorption over H-SSZ-13. **2025**, *in Preparation*
5. **Huynh, T. N.**; Mounssef, B.; Sharapa, D. I.; Studt, F.; Bučko, T. The Influence of Density Functionals on Anharmonic Free Energy of Methane Adsorption on Pt(111). **2025**, *in Preparation*
6. **Huynh, T. N.**; Matsokin, N.; Studt, F.; Fink, K.; Sharapa, D. I. Theoretical Elucidation of Ethylene Polymerization on Chromium-based NU-1000 Catalyst. **2025**, *in Preparation*

Publications outside the Scope of this Dissertation

1. **Huynh, T. N.**; Nguyen, T. M. N.; Bui, D. A.; Trung, N. T. Tetrel, nonconventional hydrogen bonds, and noticeable role of dispersion in complexes of fluoroform and carbon dichalcogenides. *Physical Chemistry Chemical Physics* **2025**, *27*, 14966–14975.
2. Jeong, K. J.; Lee, Y.; **Huynh, T. N.**; Nersisyan, H.; Suh, H.; Lee, J. H. Liquid-metal-assisted synthesis of single-crystalline TiC nanocubes with exposed (100) facets for enhanced electrocatalytic activity in the hydrogen evolution reaction. *Small Methods* **2023**, *7*, 2201076.

3. Dang, C. T. P.; Minh Tam, N.; **Huynh, T. N.**; Trung, N. T. Revisiting conventional noncovalent interactions towards a complete understanding: from tetrel to pnicogen, chalcogen, and halogen bond. *RSC Advances* **2023**, *13*, 31507–31517.

4. Lee, J. H.; Nersisyan, H.; **Huynh, T. N.**; Lim, K.-S.; Kim, W.-B.; Choi, W.-S. Combustion–aluminothermic–magnesiothermic reduction of TiO_2 to produce a Ti-rich ingot. *Metallurgical and Materials Transactions B* **2022**, *53*, 3147–3158.

5. Ri, V.; Nersisyan, H.; Lim, K.-S.; Kim, W.-B.; Choi, W.-S.; **Huynh, T. N.**; Lee, J. H. Synthesis and performance of Ti subchlorides ($TiCl_x$, $x = 2, 3$) as a Ti-ion transport agent in NaCl–$CaCl_2$ molten electrolyte. *Materialia* **2022**, *24*, 101498.

6. **Huynh, T. N.***; Babu, M.*; Nersisyan, H.; Hong, S. J.; Jin-Kyu, L.; Ki-Buem, K.; Gian, S.; Jong-Hyeon, L. Enhancement of thermoelectric performance of $Bi_{0.5}Sb_{1.5}Te_3$ alloy by inclusion of $LaVO_3$ Mott insulator. *Chemical Engineering Journal* **2022**, *437*, 135460. (* indicates equal contribution)

7. Nersisyan, H.; **Huynh, T. N.**; Lee, J. H.; Suh, H.; Hong, S. J. Carbon-assisted growth of equiangular 2D hexagons of AlN. *Materials Letters* **2022**, *314*, 131840.

8. **Huynh, T. N.**; Nersisyan, H. H.; Hong, S. J.; Lee, J. H. Co-precipitation synthesis and characterization of strontium lanthanum vanadate nanoparticles. *Korean Journal of Materials Research* **2021**, *31*, 209–218.

9. Nersisyan, H. H.; **Huynh, T. N.**; Park, K. T.; Hong, S. J.; Lee, J. H. AlN nucleation and spontaneous pattern formation via combustion of an Al–C–AlF_3 mixture in nitrogen. *Journal of Crystal Growth* **2021**, *560–561*, 126044.

10. Nersisyan, H.; Woo, H. Y.; Ri, V.; **Huynh, T. N.**; Moon, F.; MacDonald, A.; Earner, N.; Lee, J. H. Hf metal powder synthesis via a chemically activated combustion–reduction process. *Materials Chemistry and Physics* **2021**, *263*, 124417.

11. Nersisyan, H.; **Huynh, T. N.**; Ri, V.; Woo, H. Y.; Hong, S. J.; Lee, J. H. 3D self-assemblies of β-Si_3N_4: Synthesis, characterization and growth mechanism. *Journal of Crystal Growth* **2020**, *549*, 125866.

12. Le, N. N. L.; **Huynh, T. N.**; Chau, H. C.; Trung, N. T.; Ngan, V. T. A computational study on structure and stability of nitrogen-doped titanium clusters Ti_nN ($n = 1$–10). *Vietnam Journal of Chemistry* **2017**, *55*, 744–744.

D. Eidesstattliche Versicherung

Bei dieser Dissertation handelt es sich um meine eigenständig erbrachte Leistung. Ich habe nur die angegebenen Quellen und Hilfsmittel benutzt und mich keiner unzulässigen Hilfe Dritter bedient. Insbesondere habe ich wörtlich oder sinngemäß aus anderen Werken übernommene Inhalte als solche kenntlich gemacht. Ich habe die „Regeln zur Sicherung guter wissenschaftlicher Praxis am Karlsruher Institut für Technologie (KIT)" beachtet. Die Arbeit oder Teile davon habe ich bislang nicht an einer Hochschule des In- oder Auslands als Bestandteil einer Prüfungs- oder Qualifikationsleistung vorgelegt. Die Richtigkeit der vorstehenden Erklärungen bestätige ich. Die Bedeutung der eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unrichtigen oder unvollständigen eidesstattlichen Versicherung sind mir bekannt. Ich versichere an Eides statt, dass ich nach bestem Wissen die reine Wahrheit erklärt und nichts verschwiegen habe.

Karlsruhe, den 03.11.2025

Thanh-Nam Huynh