AdsorbDiff: Adsorbate Placement via Conditional Denoising Diffusion

Anonymous Authors[1]

Abstract

Determining the optimal configuration of adsorbates on a slab (adslab) is pivotal in the exploration of novel catalysts across diverse applications. Traditionally, the quest for the lowest energy adslab configuration involves placing the adsorbate onto the slab followed by an optimization process. Prior methodologies have relied on heuristics, problem-specific intuitions, or brute-force approaches to guide adsorbate placement. In this work, we propose a novel framework for adsorbate placement using denoising diffusion. The model is designed to predict the optimal adsorbate site and orientation corresponding to the lowest energy configuration. Further, we have an end-to-end evaluation framework where diffusion-predicted adslab configuration is optimized with a pretrained machine learning force field and finally evaluated with Density Functional Theory (DFT). Our findings demonstrate an acceleration of up to 5x or 3.5x improvement in accuracy compared to the previous best approach. Given the novelty of this framework and application, we provide insights into the impact of pre-training, model architectures, and conduct extensive experiments to underscore the significance of this approach.

1. Introduction

Heterogenous catalysis plays an important role in developing chemicals in industries, environmental protection through converters, and the synthesis of alternative fuels (Liu & Li, 2017; Zitnick et al., 2020). Modeling these chemical reactions involve an intermediate adsorbate on a catalyst slab which determines the efficacy of the catalyst for that particular reaction. Discovering a novel catalyst computationally involves screening through billions of candidates and finding the lowest energy configuration.

Finding the lowest energy configuration for an adsorbate and slab requires a global optimum (which is non-convex) search across different sites on the slab. Conventional approaches solve this in two steps - (1) heuristically place the adsorbate on certain important sites and (2) perform optimization with quantum mechanical calculators like Density Functional Theory (DFT) on each of these sites. The lowest energy site out of these is considered for calculating adsorption energy, which is a thermodynamic descriptor for how good that catalyst is. With recent advances in machine learning methods for predicting forces, it has become possible to perform optimization with ML force fields (MLFFs) instead of Density Functional Theory (DFT) making this process faster and easier to test many sites and find better minima. These ML force fields are trained on DFT data to predict energies and forces corresponding to different adslab configurations.

The recent release of the OC20-Dense dataset (Lan et al., 2023) signifies a significant advancement in the computation of the lowest energy adslab configuration. This work employs a blend of heuristic and random adsorbate placements across 100 sites, with subsequent optimizations across each site using Density Functional Theory (DFT) to calculate adsorption energy. The study further introduces AdsorbML, a paradigm characterized by a brute-force exploration of initial adsorbate placements. Employing pre-trained machine learning (ML) force fields from OC20, AdsorbML streamlines the optimization process, culminating in the determination of the lowest energy adsorbate-slab (adslab) configuration. The predictive accuracy of these configurations is rigorously validated against DFT single-points or complete DFT optimization. This hybrid approach results in a computational acceleration of 2000-fold in adsorption energy calculations compared to the sole reliance on DFT calculations.

Recent developments in graph neural network (GNN) based ML architectures have increased the accuracies of adsorption energy prediction significantly by encoding geometric information of atoms in more explicit ways. However, there's little to no work done on improving the adsorption site prediction which could help us get away with the currently used brute-force approach.

In this work, we develop a novel conditional denoising diffusion framework for adsorbate placement. We first formulate

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Under review at ICML 2024 AI for Science workshop.

a diffusion framework over the space of the 2D translation and 3D rigid rotation of an adsorbate molecule over the slab considering periodic boundary conditions (PBC) of the slab. Through the learned diffusion process, we sample the most stable site by iteratively updating the center of mass of adsorbate and rigid orientation. Performing a naive unconditional diffusion framework on the most optimal adsorbate site and orientation — corresponding to the lowest energy adslab configuration out of 100 densely sampled calculations in OC20-Dense — leads to throwing away 99% of DFT optimal energy data. Therefore, we modify the diffusion training to be conditional on relative energies (relative across densely sampled sites of an adslab combination). This leads to significant improvements in accuracies and sample efficiency during diffusion training. After sampling for the optimal site and orientation of adsorbate on the slab, we perform ML force field (MLFF) optimization and DFT single-point verification similar to AdsorbML. This comprehensive end-to-end evaluation helps in robust assessment of the practical impact of the learned diffusion model.

There have been significant advances in diffusion generative models in molecular and material discovery, and analogous problems in molecular docking on proteins. However, this is the first work to frame the adsorbate placement problem considering all its symmetries with the slab in a diffusion framework. Intuitively, the reverse diffusion process of AdsorbDiff helps in skipping multiple minima sites due to its energy-based conditional sampling which is followed by a local optimization with a DFT-learned MLFF to find a global optimum. To facilitate further research on this problem, we provide comprehensive results on the importance of GNN architectures for the diffusion task, show the importance of pretraining, and demonstrate the success of our approach to in-distribution (ID) and out-of-distribution (OOD) splits.

The summary of contributions of this work are -

- We propose AdsorbDiff, a novel conditional denoising diffusion framework designed to leverage the translation, rotation, and periodic symmetries inherent in adsorbate and slab interactions. Additionally, this framework is adept at efficiently predicting the lowest energy site by conditional training on relative energies.
- We present our results in a comprehensive end-to-end evaluation framework, integrated with DFT, to accurately gauge the true capability of our approach in predicting optimal adsorption energies.
- We achieve a 31.8% success rate, 3.5x higher than the naive AdsorbML baseline of 9.1% with a single site prediction. Alternatively, we demonstrate that a comparable level of accuracy could be achieved by AdsorbML by employing 5x more placements.
- We demonstrate that pretraining on large-scale local optimization data can significantly improve the results on the search for global optima.
- We show that diffusion results exhibit insignificant dependence on GNN architectures, in contrast to the notable differences observed for the same architectures when trained on DFT forces.
- We highlight the model's generalization capabilities to previously unseen adsorbates and slabs.

2. Background and Related Work

Force-fields: Energy and forces (as a gradient of energy with respect to positions) are calculated using ab initio quantum mechanical methods like Density Functional Theory (DFT). ML models can be trained to predict these energies and forces, and are called ML force-fields (MLFFs). These force fields can be utilized to perform structure optimization to get the lowest energy structures.

Optimization: For adsorption energy prediction, we start with an optimized adsorbate and slab, place the adsorbate on a slab, and perform optimization to get an adslab configuration with the lowest energy. Usually, second-order optimizers like BFGS, L-BFGS, Conjugate gradient descent, etc are used to solve this optimization problem. Since this is non-convex, the initial guess of adsorbate placement or the strategy of optimization is critical to finding an adslab configuration corresponding to the global optimum. AdsorbML (Lan et al., 2023) method starts with combining heuristic and random initial placements which is a brute-force approach to finding better minima. "Easy Potential" from (Schaarschmidt et al., 2022) trains a simple harmonic potential to guess this initial placement. Learn2Hop (Merchant et al., 2021) also learns the optimization landscape to navigate through better and hop through local minima. There are approaches like minima hopping that help in navigating through the entire optimization landscape with a force-field (Jung et al., 2023) and help in finding better minima, but these could be computationally expensive.

GNNs: Message-Passing Neural Networks (MPNN) are a class of graph neural networks (GNN) that are utilized across material property prediction tasks. Different architectures encode the geometric information in different ways. SchNet (Schütt et al., 2018) only encodes the distance information. Including more explicit geometric features have improved the model prediction as DimeNet (Gasteiger et al., 2020b;a) incorporates triplets. SphereNet (Liu et al., 2021), GemNet (Gasteiger et al., 2021; 2022) incorporates complete geometric information explicitly by giving triplets and quadruplets information. PaiNN (Schütt et al., 2021) incorporates directional information and applies only linear operations on those features. Equivariant models like NequIP (Batzner et al., 2022), Allegro (Musaelian et al., 2023), MACE (Batatia et al., 2022), SCN (Zitnick et al., 2022), Equiformer (Liao & Smidt, 2022; Liao et al., 2023)

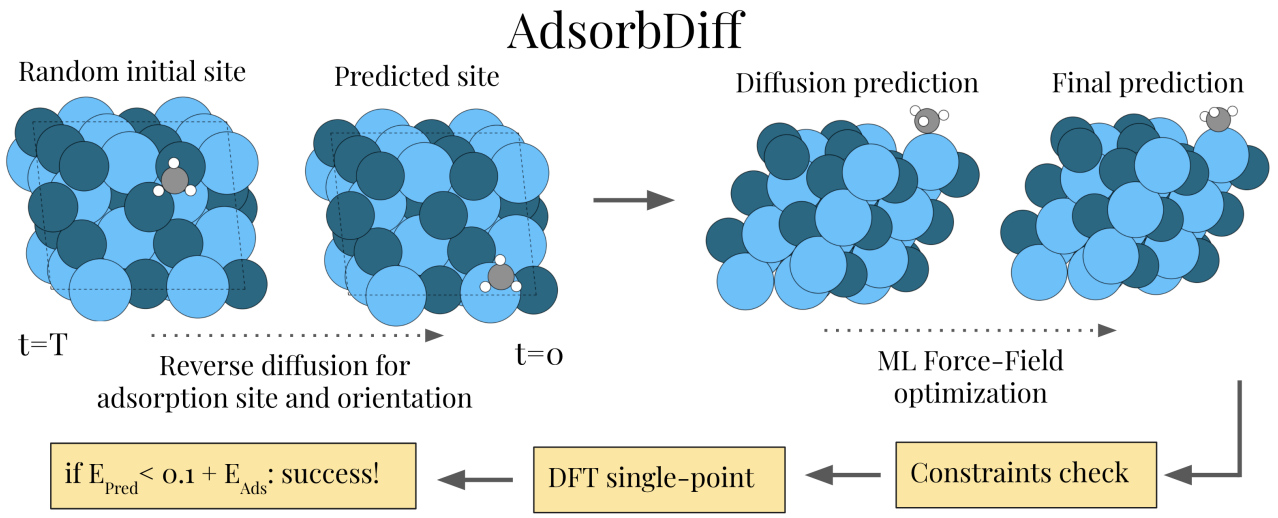

Figure 1. Overview of AdsorbDiff: Random initial site and orientation for the adsorbate are selected, followed by sampling over 2D translation, 3D rigid rotations, and considering periodic boundary conditions (PBC) to predict the optimal site and orientation. MLFF optimization is then conducted from the predicted site with a fixed interstitial gap until convergence. The final prediction undergoes constraint verification, and DFT verification is performed on valid structures to calculate success rates.

utilize spherical harmonics in representing the geometric features.

Diffusion Models: Diffusion models are a class of generative models that have shown impressive results across different domains starting from computer vision (Dhariwal & Nichol, 2021; Croitoru et al., 2023), language models (Gong et al., 2022), temporal data modeling, to applications in molecules (Xu et al., 2022; 2023; Arts et al., 2023; Hoogeboom et al., 2022; Jing et al., 2022), proteins (Wu et al., 2022; Trippe et al., 2022; Watson et al., 2022; 2023) and materials (Xie et al., 2021; Fu et al., 2023; Zeni et al., 2023; Merchant et al., 2023; Yang et al., 2023b).

There are different kinds of formulations proposed for diffusion models like denoising diffusion probabilistic models (DDPMs), score-based generative models (SGMs), and stochastic differential equations (Score SDEs) (Yang et al., 2023a). Many of these formulations have been adapted to problems in molecular and material discovery. For example, CDVAE (Xie et al., 2021) adapts concepts from noise-conditioned score networks (NCSN) for bulk discovery. Conditional diffusion has also been recently utilized across proteins (Krishna et al., 2024), catalyst and materials (Zheng et al., 2023) for generating structures with required properties.

Diffusion models have also been recently utilized for molecular docking on proteins (Corso et al., 2022). Although this problem is somewhat analogous to placing adsorbate on a slab, as far as we know there hasn't been previous work on formulating adsorbate placement in a diffusion framework. AdsorbDiff also differs from molecular docking in several key aspects – 2D translation formulation, periodic boundary conditions, conditional denoising formulation, and the requirement of DFT level accuracy as opposed to simple force-fields for proteins making our end-to-end evaluation with DFT critical.

3. AdsorbDiff

3.1. Overview

The objective of this research is to enhance the efficiency of adsorption energy calculation, representing the lowest energy configuration of an adsorbate on a slab. The methodology of this work involves the initial placement of an adsorbate on a random site within the 2D surface of the slab, followed by reverse diffusion to predict the optimal adsorption site and orientation. Employing machine learning force field optimization, the structure undergoes iterative updates with an optimizer until forces converge close to 0. Subsequently, the final structure is verified for compliance with constraints essential for defining adsorption energy. On the optimized structure, a single Density Functional Theory (DFT) calculation is conducted to obtain the predicted energy (E_{Pred}). A successful outcome is determined by the predicted energy being within 0.1 eV or lower than the DFT baseline of adsorption energy in OC20-Dense data, indicating the model's ability to provide a comparable or superior estimate of adsorption energy (shown in Figure 1).

The code is open-sourced with MIT License[1].

3.2. Adsorbate placement

Various adsorbate placement strategies were explored for the OC20-Dense dataset, incorporating a combination of heuristic and random approaches. Specifically, 100 sites were selected for each adslab configuration, utilizing a blend of heuristic and random placements. The heuristic placement involved strategically situating the adsorbate's binding site on either an on-top site, hollow site, or bridge site, with a specified interstitial gap denoting the distance between the connecting atom of the slab and the corresponding adsorbate atom. Additional random sites are introduced through the random rotation of the adsorbate along the normal of the slab, accompanied by a slight translational wobble along the surface from the heuristic site.

3.3. Diffusion for adsorbate placement

In this work, our objective is to develop a diffusion model aimed at predicting the adsorbate orientation and site corresponding to the lowest energy, as established through benchmarking with the OC20-Dense dataset.

The adsorbate motion is constrained within a manifold (M_c) and utilizes the combined action group (A), as described in DiffDock (Corso et al., 2022). This manifold permits the adsorbate to navigate towards configurations with low-energy adslab states through a combination of translations, rotations, and torsion angle adjustments. Note, for fair comparisons with our baselines, torsion angle alterations are disregarded in our analysis due to the smaller size of the adsorbate employed in this study. This approach aligns with the methodology of AdsorbML, which does not introduce randomness in torsion angles as part of its benchmark.

In our framework, we specifically consider translations in the 2D plane parallel to the slab while accounting for periodic boundary conditions (PBC). The z-coordinate is meticulously aligned to denote the normal direction of the slab and the diffusion process is executed across the xy-coordinates. Therefore, the adsorbate movements are associated with the 2D translation group $T(2)$, and rigid rotations are modeled using the $SO(3)$ group. The translation operation, denoted as $A_{\text{tr}} : T(2) \times \mathbb{R}^{2n} \to \mathbb{R}^{2n}$, is defined as $A_{\text{tr}}(r, x)_i = x_i + r$, employing the isomorphism $T(2) \cong \mathbb{R}^2$, where $x_i \in \mathbb{R}^2$ represents the position of the i-th adsorbate atom. Similarly, the rotation operation, denoted as $A_{\text{rot}} : \text{SO}(3) \times \mathbb{R}^{3n} \to \mathbb{R}^{3n}$, is defined by $A_{\text{rot}}(R, x)_i = R(x_i - \bar{x}) + \bar{x}$, where $\bar{x} = \frac{1}{n}\sum_i x_i$, signifying rotations around the center-of-mass of the adsorbate.

For the initial coordinates of adsorbate, we select a random point on the slab. This point is considered as the center-of-mass of the adsorbate in fractional coordinates. We then convert from fractional coordinates to real coordinates and perform a reverse diffusion process to get to the lowest energy site (as shown in Algorithm 1).

Algorithm 1 Sampling Method

Require: Optimized adsorbate a, interstitial gap z, slab s (normal pointing to z-direction), cell L
Ensure: Sampled adsorbate coordinates x_0
  Sample $\bar{p}^{\text{frac}} \sim \text{Uni}(0, I)$
  $\bar{p} \leftarrow \bar{p}^{\text{frac}} \odot L$
  $a \leftarrow (a - \bar{a}) + \bar{p}$
  Sample $r_N \sim \mathcal{N}(0, \sigma^2_{\text{tr}(T)})$, $R_N \sim \text{Uni}(SO(3))$
  Let $x_N = A((r_N, R_N), a)$
  for $n \leftarrow N$ **to** 1 **do**
    Let $t = \frac{n}{N}$ and $\Delta\sigma^2_{\text{tr}} = \sigma^2_{\text{tr}}(t) - \sigma^2_{\text{tr}}\left(\frac{n-1}{N}\right)$
    $\alpha \in \mathbb{R}^2, \beta \in \mathbb{R}^3 \leftarrow S_\theta(x_n, s, z|E_{\text{rel}} = 0)$
    Sample $z_{\text{tr}} \sim \mathcal{N}(0, \Delta\sigma^2_{\text{tr}})$, $z_{\text{rot}} \sim \mathcal{N}(0, \Delta\sigma^2_{\text{rot}})$
    $\Delta r \leftarrow \Delta\sigma^2_{\text{tr}}\alpha + z_{\text{tr}}$
    $\Delta R \leftarrow R(\Delta\sigma^2_{\text{rot}}\beta + z_{\text{rot}})$
    $\Delta R \leftarrow \text{PBC_corr}(\Delta R, L)$
    $x_{n-1} \leftarrow A((\Delta r, \Delta R), x_n)$
  end for
  return x_0

Algorithm 2 Training algorithm (Single Epoch)

Require: Training pairs $\{(x_0, s)\}$
  for all (x_0, s) **do**
    Sample $t \sim \text{Uni}([0, 1])$
    Sample $\Delta r, \Delta R$ from $p^t_{\text{tr}}(\cdot|0), p^t_{\text{rot}}(\cdot|0)$
    Compute $x_t \leftarrow A((\Delta r, \Delta R), x_0)$
    Predict scores $\alpha \in \mathbb{R}^2$, $\beta \in \mathbb{R}^3$ using $S_\theta(x_t, s, z|E_{\text{rel}})$
    Loss $L = \|\alpha - \nabla \log p^t_{\text{tr}}(\Delta r|0)\|^2_2 + \|\beta - \nabla \log p^t_{\text{rot}}(\Delta R|0)\|^2_2$
  end for

The work conducted by De et al. (De Bortoli et al., 2022) and Corso et al. (Corso et al., 2022) has demonstrated the applicability of the diffusion framework to Riemannian manifolds. In this context, the score model constitutes the tangent space, and a geodesic random walk serves as the reverse stochastic differential equation (SDE) solver. The score model is trained using denoising score matching (Song & Ermon, 2019), wherein a score function $s_\theta(x)$ is learned to approximate the gradient of the probability density $\nabla_x p(x)$ at varying noise levels (as shown in Algorithm 2).

The learned scores for translations and rotations are treated as independent entities, assuming the tangent space is a direct sum of individual tangent spaces, with contributions from torsion being neglected. The forward SDE for both translation and rotation is defined as $dx = \sqrt{\frac{d\sigma^2(t)}{dt}}dw$,

[1] https://anonymous.4open.science/r/AdsorbDiff-D645/

where w represents the corresponding Wiener process. In the translational scenario within $T(2)$, the model learns a score for a standard Gaussian distribution with variance $\sigma^2(t)$. For rotations in $SO(3)$, the diffusion kernel is governed by the $IGSO(3)$ distribution, which can be sampled in the axis-angle parameterization. This involves sampling a unit vector $\omega' \in so(3)$ uniformly and a random angle ω from the interval $[0, \pi]$, as outlined by Equations 1 and 2. The score of diffusion kernel is defined in Equation 3. The computation of $R' = R(\omega\hat{\omega})R$, where R is the result of applying the Euler vector $\omega\hat{\omega}$ to R, has been established in prior work by Yim et al. (Yim et al., 2023). To efficiently carry out the score computation and sampling processes, it is feasible to precompute the truncated infinite series and interpolate the cumulative distribution function (CDF) of $p(\omega)$.

$$p(\omega) = \frac{1 - \cos(\omega)}{\pi} f(\omega) \tag{1}$$

$$f(\omega) = \sum_{l=0}^{\infty} (2l+1) \exp\left(-\frac{l(l+1)\sigma^2}{2}\right) \times \sin\left(\left(l + \frac{1}{2}\right)\omega\right) \sin\left(\frac{\omega}{2}\right) \tag{2}$$

$$\nabla \ln p_t(R'|R) = \left(\frac{d}{d\omega} \log f(\omega)\right) \hat{\omega} \tag{3}$$

3.4. Conditional denoising diffusion for adsorbate placement

While the OC Challenge set provides densely calculated adsorption energies for 244 systems, a total of 244 * 100 DFT optimization benchmarks were conducted. This involved performing 100 different random placements for each configuration. Notably, the naive denoising diffusion setup was exclusively trained on the 244 lowest energy configurations.

To leverage the entirety of the DFT optimization data, a conditional diffusion model is employed. In this model, the optimized position is conditioned on the relative energy, specifically relative to the energy of the lowest energy configuration ($E^c_{\text{rel-i}} = E^c_{\text{min}} - E^c_i$). This approach allows for a more comprehensive utilization of the available DFT optimization data.

3.5. Graph Neural Network (GNN) architecture

The inputs to the ML model are the 3D positions of all input atoms from the adslab configuration and their corresponding atomic numbers. The outputs predict per-atom 3D vectors. These vectors are forces in the case of force fields and the score function in the case of diffusion. To predict multiple score functions (for translation and rotation), multiple output heads are trained each predicting independent score functions.

All architectures used in this work come under the message-passing neural network (MPNN) framework of graph neural networks (GNNs). MPNNs operate by passing messages between nodes in the graph, allowing information to be exchanged and aggregated iteratively. The key components of an MPNN include message passing, updating node states, and global readout. In the message-passing step, nodes exchange information based on their local context, and this information is then used to update the states of the nodes (as shown in Equation 4).

$$h_v^{(t+1)} = \text{Update}\left(h_v^{(t)}, \text{Aggregate}\left(\{m_{u\to v}^{(t)} \mid u \in \mathcal{N}(v)\}\right)\right) \tag{4}$$

Here, $h_v^{(t)}$ represents embeddings of node v at iteration t, $m_{u\to v}^{(t)}$ denotes the message from node u to v at iteration t, $\mathcal{N}(v)$ represents the neighborhood of node v, and Update and Aggregate are differentiable functions for updating node states and aggregating messages, respectively.

In our study, we systematically investigate diverse architectures employed in the training of diffusion models to discern the significance of architectural decisions in this context. Specifically, we have chosen to assess the performance of PaiNN, GemNet-OC, and EquiformerV2, each distinguished by its treatment of explicit geometric information and rotational symmetries (Duval et al., 2023). This selection is grounded in the diverse characteristics they bring to the table. Furthermore, we employ these architectures in benchmarking against OC20 force-field evaluation, thereby facilitating comparative analysis of architectural significance in the realms of force-fields and diffusion.

4. Results

In this section, we present results demonstrating the impact of AdsorbDiff in accelerating the search for adsorption energy or better global optima. Specifically, we demonstrate the impact of conditional denoising training over unconditional training and a randomly placed adsorbate baseline. This random baseline is equivalent to performing AdsorbML on a single site (Nsite=1). Additionally, we demonstrate the impact of pretraining, model architectures, and the generalization of this approach to new adsorbates and slabs.

4.1. Datasets

We utilize two publicly available datasets for this work - OC20-Dense (Lan et al., 2023) and OC20 (Chanussot et al., 2021).

OC20: Open Catalyst 2020 (OC20) is a large-scale dataset that contains converged DFT optimization trajectories of

460k unique adslab configurations, encompassing 55 unique elements and 74 adsorbates. Note that these optimizations are local optimizations performed with a single heuristic placement. ML force field models are trained on the forces derived from these DFT trajectories. Additionally, the optimized structure from OC20 is utilized for pre-training the diffusion model.

OC20-Dense: The OC20-Dense dataset serves as a DFT benchmark for adsorption energies, employing dense placement on 100 random sites per adslab configuration, followed by DFT optimization. This dataset releases both in-distribution (ID) and out-of-distribution (OOD) data, relative to OC20. The ID data incorporates adsorbates and slabs from OC20's training set but presents different combinations and configurations, while OOD introduces new adsorbates and/or slabs not found in the OC20 training set. A subset of OC20-Dense ID and OOD was utilized in the Open Catalyst Challenge 2023, hosted at the AI for Science Workshop during NeurIPS 2023 [2]. We split the ID data into 80/20 ratios for training the diffusion model and validating the sampling process. These smaller subsets make it computationally cheaper to perform end-to-end iterations.

4.2. Metric and constraints

Our success metric is defined by the final energy calculated through DFT. For real-world applications, this energy (D^{DFT}_{Total}) is used in calculating the adsorption energy E^{DFT}_{Ads} as $E^{DFT}_{Adsorption} = E^{DFT}_{Total} - E^{DFT}_{Slab} - E^{DFT}_{Adsorbate}$, where E^{DFT}_{Slab} and $E^{DFT}_{Adsorbate}$ are the independent energies of slab and adsorbate respectively. This adsorption energy acts as a thermodynamic description of how good a catalyst is for downstream application. The DFT Success Rate (SR) is defined as the percentage of *valid* structures within 0.1 eV or lower of the DFT computed adsorption energy benchmark in the OC20-Dense data (as described in AdsorbML). This is computationally expensive to calculate but is accurate. Metrics calculated from ML predictions are inexpensive but are also inaccurate, discussed further in Appendix C.

Since we calculate adsorption energies, the adsorbate and slab must not change during optimization. Therefore, the structures are considered an *anomaly* due to - (1) adsorbate desorption: adsorbate moves far away from the slab, (2) adsorbate dissociation: atoms in adsorbate dissociate into multiple adsorbates, (3) slab mismatch/reconstruction: slab reconstructs into a completely different structure during optimization (4) adsorbate intercalation: when any of the adsorbate atoms detaches and get into the slab.

Experimental setup: All presented results are based on the DFT success rate metric as defined in the preceding section. Throughout the diffusion process, we employ the EquiformerV2 architecture, unless explicitly stated otherwise, owing to its state-of-the-art performance in AdsorbML. Additionally, for MLFF optimization, we utilize GemNet-OC pre-trained on OC20, chosen for its lower inference cost. Further specifics regarding model and training hyperparameters are available in Appendix D. All results are shown on the val ID split apart from the OOD section.

[2] https://opencatalystproject.org/challenge.html

4.3. Conditional vs Unconditional diffusion

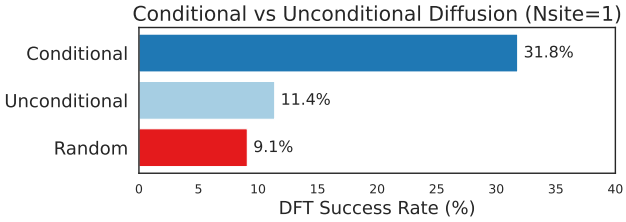

Figure 2. Comparison of conditional and unconditional diffusion with a baseline of random placement. Conditional diffusion training on relative energies of configurations of adslab significantly improves success rates over unconditional training and AdsorbML baseline.

We demonstrate the importance of conditional training on relative energies (as shown in Section 3.4) over unconditional diffusion training in Figure 2. We compare both of these approaches to a naive baseline of AdsorbML with a single site (Nsite=1) where MLFF optimization is performed on a random adsorbate placement. It is noteworthy that the performance of unconditional training is suboptimal, this may be ascribed to the unexploited potential of additional data made available through conditional training.

4.4. AdsorbDiff vs AdsorbML

AdsorbML conducts MLFF optimization and DFT evaluations on adsorption sites randomly placed within the system. A comparative analysis is drawn with AdsorbDiff, where the prediction of adsorption sites is facilitated through the utilization of diffusion models. As depicted in Figure 3, it is evident that AdsorbDiff exhibits notably superior performance, particularly at lower Nsites. However, as the number of adsorption sites (Nsites) increases, AdsorbDiff tends to either converge to or underperform in comparison to the brute force approach employed by AdsorbML. Adsorbate sites sampled from AdsorbDiff have less diversity by design as it's trained to predict the global optima. We calculate the average across the standard deviation of the points sampled at 10 Nsites and get **8.1 Å** for AdsorbML and **2.7 Å** for AdsorbDiff. AdsorbML's brute force placements have more randomness which leads to fewer anomalies post the MLFF

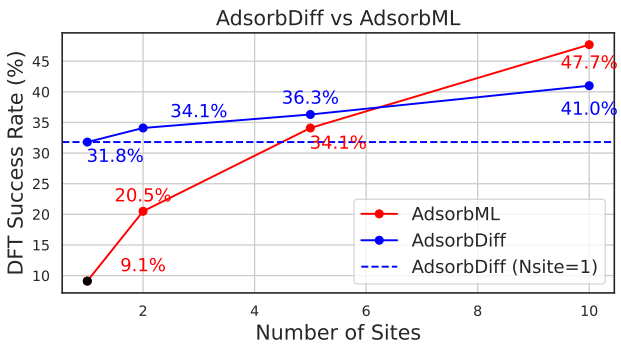

Figure 3. DFT Success Rates (%) for AdsorbDiff and AdsorbML across a varying number of site predictions. AdsorbDiff performs 3.5x better than AdsorbML utilizing a single site prediction. At higher sites, AdsorbML performs better due to the brute-force nature of site prediction that reduces anomalies.

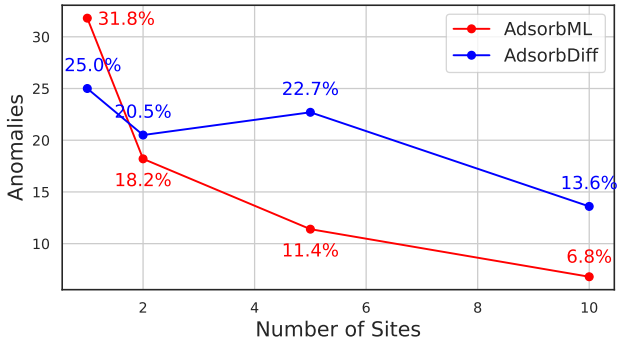

Figure 4. Anomalies in AdsorbDiff and AdsorbML with respect to Nsites. A system is labeled as anomalous if all its predicted sites result in anomalies. AdsorbML has fewer anomalies than AdsorbDiff at higher Nsites due to more randomness in initial sites.

optimization process shown in Figure 4.

4.5. Impact of pretraining

Conditional diffusion benefits from training on a dataset that is 100 times more extensive than the unconditional approach, a consequence of leveraging multiple local optima within a unique adslab configuration. The substantial increase in training data size manifests in a notable enhancement in the success rate for the conditional approach. The OC20 IS2RE dataset, containing optimization data for 460,000 distinct adslab combinations, serves as a valuable resource for pre-training the diffusion model. It is important to acknowledge that this pretraining process results in a model that learns the local optima of an adslab combination, with the caveat that the model may not capture global optima for an adslab combination.

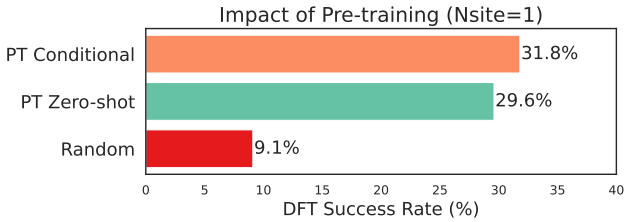

Figure 5. Impact of pretraining on 460k OC20 local optima data on DFT Success Rate. PT Zero-shot measures zero-shot generalization of OC20 pre-trained model to OC20-Dense data. PT Conditional is finetuned on OC20 Dense data conditionally on relative energies of adslab configurations. Random baseline corresponds to randomly placed adsorbate.

IS2RS Pretraining (PT) Zero-shot: Taking advantage of the diffusion model pre-trained on OC20 IS2RE data, we conduct a zero-shot validation on the OC20-Dense ID val split. This experimental setup allows us to assess the model's ability to predict better global optima having trained on a large dataset of local optima. Notably, we observe a substantial increase in DFT success rate in the zero-shot setting (as shown in Figure 5).

IS2RS Pretraining (PT) Conditional: In this approach, we utilize the pre-trained model using the OC20-Dense data as described in Section 3.4. We observe that although this gives a 2% improvement over zero-shot, it converges to the same results as just training conditionally on OC20-Dense (shown in Figure 5).

4.6. Impact of architectures

Architectures characterized by richer geometric information and extensive many-body interaction capabilities, such as eSCN and EquiformerV2, have demonstrated superior performance in force evaluations within the OC20 dataset compared to simpler models like PaiNN, which primarily encode directional information and apply linear transformations. Our benchmarking involves the evaluation of three architectures that exhibit progressively improved performance in OC20 Force MAE, revealing significant differences among them.

This evaluation is specifically conducted in the context of the zero-shot assessment following pretraining (PT zero-shot) on an extensive dataset encompassing 460,000 OC20 instances. This choice is inspired by insights from the GemNet-OC paper (Gasteiger et al., 2022), suggesting that certain architectural choices manifest optimal performance only at higher data scales.

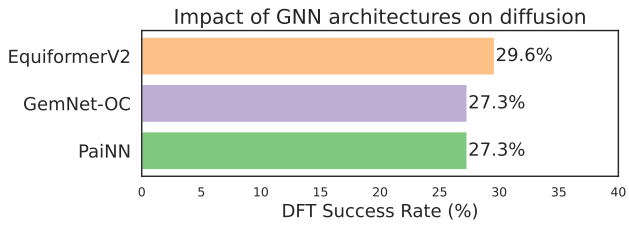

Figure 6. Impact of Graph Neural Network (GNN) architectures on the diffusion process for DFT Success Rate keeping other parts of the framework same. Different architectures perform similarly on the task of diffusion sampling.

Interestingly, in the realm of the diffusion task, we note that the disparity in success rates among these architectures is marginal (as shown in Figure 6) which has been recently demonstrated in applications of molecular generation tasks as well (Wang et al., 2023). The intuition behind this result is that the diffusion model's score function can be thought of as learning a harmonic potential (Xie et al., 2021). Harmonic potentials are simpler force-fields than ab-initio DFT calculations involved in OC20 forces. This could result in simpler architectures being able to capture the underlying complexity of the diffusion task defined in our work.

4.7. OOD generalization

We measure the success of AdsorbDiff in out-of-distribution (OOD) cases where the model hasn't seen the adsorbate or the slab even during the pre-training on OC20. We pick a random 50 samples out of 200 validation OOD split defined in Open Catalyst Challenge 2023. We observe a marginal decrease of only 3.8% in results for the OOD case compared to the ID scenario and consistently observe significant improvement over the AdsorbML (Nsite=1) baseline.

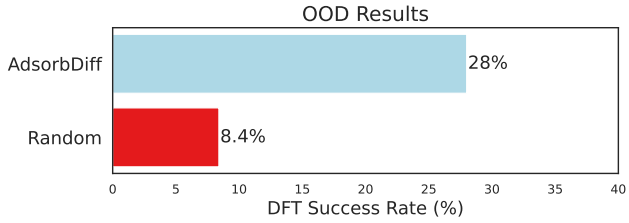

Figure 7. Comparison of DFT Success Rate for In-Distribution (ID) and Out-of-Distribution (OOD) splits using the AdsorbDiff method. Random baseline corresponds to randomly placed adsorbate.

4.8. Inference cost

In the case of conditional diffusion, our approach maintains a maximum step limit of 100, with adsorbate placement converging, on average, within 98 steps. In contrast, for MLFF optimization with a maximum step limit of 300 and Fmax criteria of 0.01 eV/A (consistent with AdsorbML), the convergence occurs in approximately 286 steps. Consequently, for scenarios with a single adsorption site (Nsite 1), AdsorbDiff incurs approximately 34% more inference cost than AdsorbML, given the GNN architecture for diffusion and MLFF optimization is the same. This end-to-end ML framework is $O(10^4)$ times faster than the conventional DFT pipelines (Lan et al., 2023).

In Section 4.6, we illustrate that simpler and faster models such as PaiNN yield comparable performance to more intricate and slower models like EquiformerV2. This enhances the efficiency of our diffusion-based approach, as its computational burden becomes negligible in comparison to MLFF optimization, which would require more computationally intensive ML architectures (details in Appendix B).

5. Conclusion

This work introduces AdsorbDiff, a novel conditional denoising diffusion framework adept at leveraging inherent symmetries in adsorbate and slab interactions, enabling efficient prediction of the lowest energy site. The proposed end-to-end evaluation framework, coupled with Density Functional Theory (DFT), provides a robust assessment of our approach's capability to predict optimal adsorption energies. Notably, AdsorbDiff achieves a remarkable 31.8% success rate with a single site prediction, surpassing the naive AdsorbML baseline (9.1%) by 3.5x. We demonstrate the benefits of pretraining on large-scale local optima of adsorption sites. Interestingly, we find the diffusion method's performance to be not significantly dependent on the GNN architecture choice. Furthermore, our model's demonstrated generalization to previously unseen adsorbates and slabs underscores its adaptability and robustness.

6. Limitations and Future Work

Our findings emphasize that anomalies play a substantial role in diminishing success rates, particularly in the context of multiple site predictions. While certain works have successfully employed constraints, such as Hookean constraints, to mitigate these anomalies, their implementation in a computationally efficient manner for larger adsorbates remains non-trivial. Addressing this challenge stands out as a crucial avenue for future research. Furthermore, the incorporation of torsion angles presents a promising direction for further improvement, especially when dealing with larger adsorbates.

References

Arts, M., Garcia Satorras, V., Huang, C.-W., Zügner, D., Federici, M., Clementi, C., Noé, F., Pinsler, R., and van den Berg, R. Two for one: Diffusion models and force fields for coarse-grained molecular dynamics. *Journal of Chemical Theory and Computation*, 19(18):6151–6159, 2023.

Batatia, I., Kovacs, D. P., Simm, G., Ortner, C., and Csányi, G. Mace: Higher order equivariant message passing neural networks for fast and accurate force fields. *Advances in Neural Information Processing Systems*, 35: 11423–11436, 2022.

Batatia, I., Benner, P., Chiang, Y., Elena, A. M., Kovács, D. P., Riebesell, J., Advincula, X. R., Asta, M., Baldwin, W. J., Bernstein, N., et al. A foundation model for atomistic materials chemistry. *arXiv preprint arXiv:2401.00096*, 2023.

Batzner, S., Musaelian, A., Sun, L., Geiger, M., Mailoa, J. P., Kornbluth, M., Molinari, N., Smidt, T. E., and Kozinsky, B. E (3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature communications*, 13(1):2453, 2022.

Chanussot, L., Das, A., Goyal, S., Lavril, T., Shuaibi, M., Riviere, M., Tran, K., Heras-Domingo, J., Ho, C., Hu, W., et al. Open catalyst 2020 (oc20) dataset and community challenges. *Acs Catalysis*, 11(10):6059–6072, 2021.

Corso, G., Stärk, H., Jing, B., Barzilay, R., and Jaakkola, T. Diffdock: Diffusion steps, twists, and turns for molecular docking. *arXiv preprint arXiv:2210.01776*, 2022.

Croitoru, F.-A., Hondru, V., Ionescu, R. T., and Shah, M. Diffusion models in vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

De Bortoli, V., Mathieu, E., Hutchinson, M., Thornton, J., Teh, Y. W., and Doucet, A. Riemannian score-based generative modelling. *Advances in Neural Information Processing Systems*, 35:2406–2422, 2022.

Dhariwal, P. and Nichol, A. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021.

Duval, A., Mathis, S. V., Joshi, C. K., Schmidt, V., Miret, S., Malliaros, F. D., Cohen, T., Lio, P., Bengio, Y., and Bronstein, M. A hitchhiker's guide to geometric gnns for 3d atomic systems. *arXiv preprint arXiv:2312.07511*, 2023.

Fu, X., Xie, T., Rosen, A. S., Jaakkola, T., and Smith, J. Mofdiff: Coarse-grained diffusion for metal-organic framework design. *arXiv preprint arXiv:2310.10732*, 2023.

Gasteiger, J., Giri, S., Margraf, J. T., and Günnemann, S. Fast and uncertainty-aware directional message passing for non-equilibrium molecules. *arXiv preprint arXiv:2011.14115*, 2020a.

Gasteiger, J., Groß, J., and Günnemann, S. Directional message passing for molecular graphs. *arXiv preprint arXiv:2003.03123*, 2020b.

Gasteiger, J., Becker, F., and Günnemann, S. Gemnet: Universal directional graph neural networks for molecules. *Advances in Neural Information Processing Systems*, 34: 6790–6802, 2021.

Gasteiger, J., Shuaibi, M., Sriram, A., Günnemann, S., Ulissi, Z., Zitnick, C. L., and Das, A. Gemnet-oc: developing graph neural networks for large and diverse molecular simulation datasets. *arXiv preprint arXiv:2204.02782*, 2022.

Gong, S., Li, M., Feng, J., Wu, Z., and Kong, L. Diffuseq: Sequence to sequence text generation with diffusion models. *arXiv preprint arXiv:2210.08933*, 2022.

Hoogeboom, E., Satorras, V. G., Vignac, C., and Welling, M. Equivariant diffusion for molecule generation in 3d. In *International conference on machine learning*, pp. 8867–8887. PMLR, 2022.

Jing, B., Corso, G., Chang, J., Barzilay, R., and Jaakkola, T. Torsional diffusion for molecular conformer generation. *Advances in Neural Information Processing Systems*, 35: 24240–24253, 2022.

Jung, H., Sauerland, L., Stocker, S., Reuter, K., and Margraf, J. T. Machine-learning driven global optimization of surface adsorbate geometries. *npj Computational Materials*, 9(1):114, 2023.

Krishna, R., Wang, J., Ahern, W., Sturmfels, P., Venkatesh, P., Kalvet, I., Lee, G. R., Morey-Burrows, F. S., Anishchenko, I., Humphreys, I. R., et al. Generalized biomolecular modeling and design with rosettafold all-atom. *Science*, pp. eadl2528, 2024.

Lan, J., Palizhati, A., Shuaibi, M., Wood, B. M., Wander, B., Das, A., Uyttendaele, M., Zitnick, C. L., and Ulissi, Z. W. Adsorbml: a leap in efficiency for adsorption energy calculations using generalizable machine learning potentials. *npj Computational Materials*, 9(1):172, 2023.

Liao, Y.-L. and Smidt, T. Equiformer: Equivariant graph attention transformer for 3d atomistic graphs. *arXiv preprint arXiv:2206.11990*, 2022.

Liao, Y.-L., Wood, B., Das, A., and Smidt, T. Equiformerv2: Improved equivariant transformer for scaling to higher-degree representations. *arXiv preprint arXiv:2306.12059*, 2023.

Liu, D. C. and Nocedal, J. On the limited memory bfgs method for large scale optimization. *Mathematical programming*, 45(1-3):503–528, 1989.

Liu, J. and Li, Y. Heterogeneous catalysis: A central science for a sustainable future. *Accounts of Chemical Research*, 50(4):571–572, 2017.

Liu, Y., Wang, L., Liu, M., Zhang, X., Oztekin, B., and Ji, S. Spherical message passing for 3d graph networks. *arXiv preprint arXiv:2102.05013*, 2021.

Ma, Z. and Zaera, F. Heterogeneous catalysis by metals. *Encyclopedia of Inorganic Chemistry*, 2006.

Merchant, A., Metz, L., Schoenholz, S. S., and Cubuk, E. D. Learn2hop: Learned optimization on rough landscapes. In *International Conference on Machine Learning*, pp. 7643–7653. PMLR, 2021.

Merchant, A., Batzner, S., Schoenholz, S. S., Aykol, M., Cheon, G., and Cubuk, E. D. Scaling deep learning for materials discovery. *Nature*, pp. 1–6, 2023.

Musaelian, A., Batzner, S., Johansson, A., Sun, L., Owen, C. J., Kornbluth, M., and Kozinsky, B. Learning local equivariant representations for large-scale atomistic dynamics. *Nature Communications*, 14(1):579, 2023.

Rothenberg, G. *Catalysis: concepts and green applications*. John Wiley & Sons, 2017.

Schaarschmidt, M., Riviere, M., Ganose, A. M., Spencer, J. S., Gaunt, A. L., Kirkpatrick, J., Axelrod, S., Battaglia, P. W., and Godwin, J. Learned force fields are ready for ground state catalyst discovery. *arXiv preprint arXiv:2209.12466*, 2022.

Schütt, K., Unke, O., and Gastegger, M. Equivariant message passing for the prediction of tensorial properties and molecular spectra. In *International Conference on Machine Learning*, pp. 9377–9388. PMLR, 2021.

Schütt, K. T., Sauceda, H. E., Kindermans, P.-J., Tkatchenko, A., and Müller, K.-R. Schnet–a deep learning architecture for molecules and materials. *The Journal of Chemical Physics*, 148(24), 2018.

Shoghi, N., Kolluru, A., Kitchin, J. R., Ulissi, Z. W., Zitnick, C. L., and Wood, B. M. From molecules to materials: Pre-training large generalizable models for atomic property prediction. *arXiv preprint arXiv:2310.16802*, 2023.

Song, Y. and Ermon, S. Generative modeling by estimating gradients of the data distribution. *Advances in neural information processing systems*, 32, 2019.

Trippe, B. L., Yim, J., Tischer, D., Baker, D., Broderick, T., Barzilay, R., and Jaakkola, T. Diffusion probabilistic modeling of protein backbones in 3d for the motif-scaffolding problem. *arXiv preprint arXiv:2206.04119*, 2022.

Wang, Y., Elhag, A. A., Jaitly, N., Susskind, J. M., and Bautista, M. A. Generating molecular conformer fields. *arXiv preprint arXiv:2311.17932*, 2023.

Watson, J. L., Juergens, D., Bennett, N. R., Trippe, B. L., Yim, J., Eisenach, H. E., Ahern, W., Borst, A. J., Ragotte, R. J., Milles, L. F., et al. Broadly applicable and accurate protein design by integrating structure prediction networks and diffusion generative models. *BioRxiv*, pp. 2022–12, 2022.

Watson, J. L., Juergens, D., Bennett, N. R., Trippe, B. L., Yim, J., Eisenach, H. E., Ahern, W., Borst, A. J., Ragotte, R. J., Milles, L. F., et al. De novo design of protein structure and function with rfdiffusion. *Nature*, 620(7976): 1089–1100, 2023.

Wu, K. E., Yang, K. K., Berg, R. v. d., Zou, J. Y., Lu, A. X., and Amini, A. P. Protein structure generation via folding diffusion. *arXiv preprint arXiv:2209.15611*, 2022.

Xie, T., Fu, X., Ganea, O.-E., Barzilay, R., and Jaakkola, T. Crystal diffusion variational autoencoder for periodic material generation. *arXiv preprint arXiv:2110.06197*, 2021.

Xu, M., Yu, L., Song, Y., Shi, C., Ermon, S., and Tang, J. Geodiff: A geometric diffusion model for molecular conformation generation. *arXiv preprint arXiv:2203.02923*, 2022.

Xu, M., Powers, A. S., Dror, R. O., Ermon, S., and Leskovec, J. Geometric latent diffusion models for 3d molecule generation. In *International Conference on Machine Learning*, pp. 38592–38610. PMLR, 2023.

Yang, L., Zhang, Z., Song, Y., Hong, S., Xu, R., Zhao, Y., Zhang, W., Cui, B., and Yang, M.-H. Diffusion models: A comprehensive survey of methods and applications. *ACM Computing Surveys*, 56(4):1–39, 2023a.

Yang, M., Cho, K., Merchant, A., Abbeel, P., Schuurmans, D., Mordatch, I., and Cubuk, E. D. Scalable diffusion for materials generation. *arXiv preprint arXiv:2311.09235*, 2023b.

Yim, J., Trippe, B. L., De Bortoli, V., Mathieu, E., Doucet, A., Barzilay, R., and Jaakkola, T. Se (3) diffusion model with application to protein backbone generation. *arXiv preprint arXiv:2302.02277*, 2023.

Zeni, C., Pinsler, R., Zügner, D., Fowler, A., Horton, M., Fu, X., Shysheya, S., Crabbé, J., Sun, L., Smith, J., et al.

Mattergen: a generative model for inorganic materials design. *arXiv preprint arXiv:2312.03687*, 2023.

Zheng, S., He, J., Liu, C., Shi, Y., Lu, Z., Feng, W., Ju, F., Wang, J., Zhu, J., Min, Y., et al. Towards predicting equilibrium distributions for molecular systems with deep learning. *arXiv preprint arXiv:2306.05445*, 2023.

Zitnick, C. L., Chanussot, L., Das, A., Goyal, S., Heras-Domingo, J., Ho, C., Hu, W., Lavril, T., Palizhati, A., Riviere, M., et al. An introduction to electrocatalyst design using machine learning for renewable energy storage. *arXiv preprint arXiv:2010.09435*, 2020.

Zitnick, L., Das, A., Kolluru, A., Lan, J., Shuaibi, M., Sriram, A., Ulissi, Z., and Wood, B. Spherical channels for modeling atomic interactions. *Advances in Neural Information Processing Systems*, 35:8054–8067, 2022.

Table of Contents

A Background in catalysis

B Compute details

C Metrics other than DFT Success Rate

D Model and training hyperparameters

A. Background in catalysis

Heterogeneous catalysis is a pivotal process in chemistry and industry. Unlike its counterpart, homogeneous catalysis, where reactants, products, and the catalyst coexist in the same phase, heterogeneous catalysis involves catalysts that exist in a different phase from the reactants or products. This phase distinction extends beyond solid, liquid, and gas components—it also encompasses immiscible mixtures (such as oil and water) or any scenario where an interface is present. In most cases, heterogeneous catalysis involves solid-phase catalysts and gas-phase reactants. The heart of this process lies in a cycle of molecular adsorption, reaction, and desorption occurring at the catalyst surface. Thermodynamics, mass transfer, and heat transfer all influence the rate of these reactions. Heterogeneous catalysis plays a crucial role in large-scale production and selective product formation, impacting approximately 35% of the world's GDP (Ma & Zaera, 2006) and aiding in the production of 90% of chemicals by volume (Rothenberg, 2017).

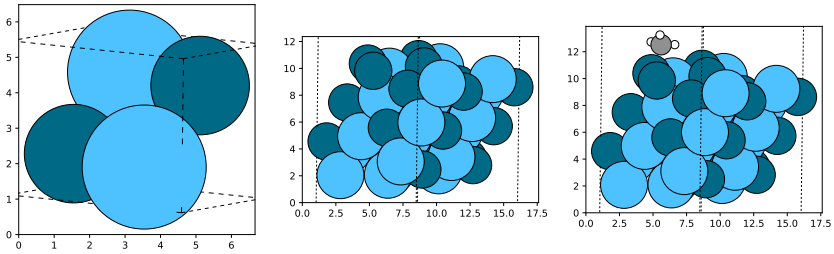

Figure 8. This is a system sampled from the OC20-Dense data. Left: An example bulk structure of Hf2Pd2. Center: Slab created out of the bulk. Right: CH3 adsorbate randomly placed on the slab.

A.1. Bulks

The bulk refers to the solid material that serves as the foundation for catalytic reactions. It constitutes the underlying structure upon which the catalyst operates. Think of it as the bedrock—the stable, three-dimensional lattice of atoms or molecules that provides the framework for further interactions.

A.2. Slabs

A slab is a specific surface layer of the bulk material. Imagine slicing through the bulk crystal lattice to reveal a flat, exposed surface. Slabs are essential because they expose a large number of active sites where reactants can interact with the catalyst. These surfaces play a critical role in determining the overall catalytic behavior. Researchers often explore different slab orientations (such as (100), (111), or (110)) to understand how the arrangement of atoms affects reactivity.

A.3. Adsorbates

An adsorbate is a molecule or atom that adheres to the surface of the catalyst. During heterogeneous catalysis, reactant molecules (the adsorbates) bind to specific sites on the catalyst surface (the adsorbent). This binding occurs via van der Waals forces, including dipole-dipole interactions, induced dipole interactions, and London dispersion forces. Importantly, no chemical bonds form between the adsorbate and adsorbent; their electronic states remain relatively unperturbed. Adsorption is a crucial step in the catalytic process, as it positions reactants near active sites for subsequent reactions.

B. Compute details and discussions

Do AdsorbDiff speedups impact practically?

Supplementary Table VII of AdsorbML (Lan et al., 2023) discusses the computational cost of MLFF optimization (or ML RX as mentioned in that work) and DFT SP / RX costs for different GNN architectures. The computational cost for 300 steps across 100 Nsites of ML RX of PaiNN is 60.4 GPU-hours on 32GB NVIDIA V100 cards. The cost for the best models SCN-MD-Large and GemNet-OC-MD-Large is 1129.2 and 638.3 respectively. The cost of performing diffusion with PaiNN at 1 site (60.4 / 300 GPU-hours) is insignificant to the MLFF optimization cost at 1 Nsite of any of the recent models like SCN-MD-Large (1129.2 / 100 GPU-hours).

Why should we aim to perform only a few MLFF optimizations?

With the MLFF models getting larger, more complex, and more accurate, their inference costs have also increased. The assumption that MLFF optimization inference costs are completely insignificant in comparison to DFT is not valid for larger models. If we assume that only DFT SP is required for the purpose of ranking sites, we can see in the same table of AdsorbML that for SCN-MD-Large the DFT-SP cost is 2645.29 CPU-hours performed across 5 sites. Given that DFT on GPUs is getting faster and it's roughly 5x faster for these systems, DFT SP verifications (2645.39 / 5 GPU-hours) are not truly the bottleneck in comparison to ML RX (1129.2 GPU-hours). To be able to perform billions of these simulations, it's critical to make the ML simulations significantly more efficient.

This hints at a potential progression in future computational workflows, wherein they would commence with a diffusion-based simulation leveraging simple and fast ML architectures. Following this, MLFF optimization might utilize foundation models targeting atomic structures(Shoghi et al., 2023; Batatia et al., 2023), culminating in the validation of results through a select number of DFT calculations.

C. Metrics other than DFT Success Rate

We present all our findings using the DFT Success Rate within our end-to-end framework, recognizing the significance of DFT metrics over purely ML-based metrics, which may suffer from inaccuracies. In this section, we offer the ML Success Rate for key results, comparing it with the DFT Success Rate. Two metrics, solely tracked with our ML models without DFT, are considered. Firstly, the success metric based on the distance between the optimal site and predicted site through diffusion. We define a prediction as successful if the Mean Absolute Error (MAE) distance is less than 1 Å. Secondly, we assess it after MLFF optimization using ML predicted energies. Note that the MLFF optimization is performed using GemNet-OC checkpoint pretraining on OC20 2M splits across all experiments for consistency.

Method	**Diff SR**	**MLFF SR**	**DFT SR**
Unconditional	15.9	22.72	11.36
Conditional	36.4	29.5	31.81
PT zero-shot	38.63	27.27	29.55
PT conditional	29.5	47.72	31.81
PaiNN	29.5	27.3	27.3
GemNet	29.5	27.3	27.3
Equiformer V2	29.5	27.3	29.6

Table 1. Comparison of key results with different metrics for a single-site prediction

We also perform an experiments where we start from the optimal site and perform MLFF optimization and calculate its MLFF Success Rate and find it to be **43%**. With DFT, its 100%. Therefore, we discourage the use of only ML based metrics in evaluating these results. This also demonstrates the importance of our end-to-end framework.

D. Model and training hyperparameters

In the case of diffusion training, an additional head is incorporated for each of these architectures to account for the two vector predictions needed – one for the translation vector and the other for the Euler vector.

D.1. EquiformerV2

Definitions of the attributes mentioned in Table 2 can be found in the open-sourced code of the original work `https://github.com/Open-Catalyst-Project/ocp/tree/main/ocpmodels/models/equiformer_v2`.

Hyperparameter	Value
max_neighbors	20
max_radius	12.0
max_num_elements	90
num_layers	8
sphere_channels	128
attn_hidden_channels	64
num_heads	8
attn_alpha_channels	64
attn_value_channels	16
ffn_hidden_channels	128
norm_type	'layer_norm_sh'
lmax_list	[4]
mmax_list	[2]
grid_resolution	18
num_sphere_samples	128
edge_channels	128
use_atom_edge_embedding	True
distance_function	'gaussian'
num_distance_basis	512
attn_activation	'silu'
use_s2_act_attn	False
ffn_activation	'silu'
use_gate_act	False
use_grid_mlp	True
alpha_drop	0.1
drop_path_rate	0.1
proj_drop	0.0
weight_init	'uniform'

Table 2. EquiformerV2 hyperparameters

D.2. GemNet-OC

Definitions of the attributes mentioned in Table 6 can be found in the open-sourced code of the original work `https://github.com/Open-Catalyst-Project/ocp/tree/main/ocpmodels/models/gemnet_oc`. The same model architecture hyperparameters are utilized for diffusion and MLFF optimization.

D.3. PaiNN

Definitions of the attributes mentioned in Table 3 can be found in the open-sourced code of the work adapted in the OCP repository `https://github.com/Open-Catalyst-Project/ocp/tree/main/ocpmodels/models/painn`.

D.4. MLFF optimization

All MLFF optimization in this study is conducted using the pre-trained GemNet-OC model, trained on the OC20 2M dataset. The preference for GemNet-OC over the more recent EquiformerV2 architecture stems from its up to 8x faster inference speeds, coupled with reasonably good accuracies. This choice shouldn't impact the qualitative results of this work as this method is model agnostic.

Optimizations are performed using L-BFGS algorithm (Liu & Nocedal, 1989). The hyperparameters of L-BFGS are

Hyperparameter	Value
hidden_channels	512
num_layers	6
num_rbf	128
cutoff	12.0
max_neighbors	50

Table 3. PaiNN hyperparameters

mentioned in Table 4. During MLFF optimization, we move both the surface atoms and adsorbate atoms. We keep a maximum step of 300 and an F max of 0.01 eV/Å. All of these settings are consistent across all our runs.

Hyperparameter	Value
maxstep	0.04
memory	50
damping	1.0
alpha	70.0

Table 4. L-BFGS hyperparameters

D.5. Sampling hyperparameters

We sample from the SDE described in Section 3.3. In practice, we find the ODE formulation to be more performant across all experiments and therefore that's our default sampling approach. This could be due to the global optima search that we aim for instead of just sampling from the distribution of minima. We set a max step size for sampling and stop early if the movement of atoms converges. These hyperparameters are shown in Table 5.

We have also tried the Annealed Langevin Sampling formulation proposed in Noise Conditioned Score Network (NCSN) (Song & Ermon, 2019) but find it to be suboptimal taking 10x more steps for convergence and doesn't generalize well outside of the domain of training data.

Hyperparameter	Value
Number of sampling steps	100
Translation low std	0.1
Translation high std	10
Rotation low std	0.01
Rotation high std	1.55

Table 5. Sampling hyperparameters

D.6. Training compute and hyperparameters

We perform all of our training on 2 48GB A6000 GPUs. We use cosine learning rate with linear warmup. The conditional diffusion model has been trained for 47 GPU-hours. Maximum learning rate was kept to be 4.e-4 and 1.e-4 for pretraining and finetuning runs respectively utilizing AdamW optimizer. For each model architecture, maximum batch size was utilized that could fit on the GPU. Overall, we keep the training methods similar to the original works of respective model architectures.

Hyperparameter	Value
num_spherical	7
num_radial	128
num_blocks	4
emb_size_atom	256
emb_size_edge	512
emb_size_trip_in	64
emb_size_trip_out	64
emb_size_quad_in	32
emb_size_quad_out	32
emb_size_aint_in	64
emb_size_aint_out	64
emb_size_rbf	16
emb_size_cbf	16
emb_size_sbf	32
num_before_skip	2
num_after_skip	2
num_concat	1
num_atom	3
num_output_afteratom	3
cutoff	12.0
cutoff_qint	12.0
cutoff_aeaint	12.0
cutoff_aint	12.0
max_neighbors	30
max_neighbors_qint	8
max_neighbors_aeaint	20
max_neighbors_aint	1000
rbf.name	gaussian
envelope.name	polynomial
envelope.exponent	5
cbf.name	spherical_harmonics
sbf.name	legendre_outer
extensive	True
output_init	HeOrthogonal
activation	silu
regress_forces	True
direct_forces	True
forces_coupled	False
quad_interaction	True
atom_edge_interaction	True
edge_atom_interaction	True
atom_interaction	True
num_atom_emb_layers	2
num_global_out_layers	2
qint_tags	[1, 2]

Table 6. GemNet-OC hyperparameters