AdsorbFlow: energy-conditioned flow matching enables fast and realistic adsorbate placement

Jiangjie Qiu [1] **Wentao Li** [1] **Honghao Chen** [1] **Yishen Tew** [1] **Leyi Zhao** [1] **Xiaonan Wang** [1]

[1] *Beijing Key Laboratory of Artificial Intelligence for Advanced Chemical Engineering Materials, Department of Chemical Engineering, Tsinghua University, Beijing, China.*. Correspondence to: Jiangjie Qiu qjj25@mails.tsinghua.edu.cn, Xiaonan Wang wangxiaonan@tsinghua.edu.cn.

1. Introduction

Adsorbate placement—predicting a stable adsorption pose of a molecule on a catalyst surface—is a practical bottleneck in computational heterogeneous catalysis. Classical pipelines (e.g., AdsorbML [1]) enumerate many candidate sites/orientations, relax them with an ML force field (MLFF), and forward only the lowest-energy structure for DFT verification, yet still waste compute on suboptimal candidates. Generative approaches (e.g., diffusion-based AdsorbDiff [2]) propose poses directly, but inference typically requires $O(10^2)$ denoising steps and accuracy still leaves room for improvement. We propose **AdsorbFlow**, a generative model trained by flow matching [3] in pose space, enabling fast sampling with few function evaluations while maintaining physically realistic candidates. We evaluate under the standard **OC20-Dense** [4] protocol: generate N_{site} candidates → MLFF relaxation → DFT verification of the best relaxed candidate; success is defined as adsorption energy within 0.1 eV of the DFT optimum. Under this setting, AdsorbFlow achieves **45.5%** DFT success at $N_{\text{site}} = 1$ (**61.4%** DFT success at $N_{\text{site}} = 10$).

2. AdsorbFlow

2.1 Rigid-body pose representation

We treat the adsorbate as a rigid body with 6 DoF: an anchor translation $\mathbf{z}_{tr} \in \mathbb{R}^3$ and an axis-angle rotation vector $\mathbf{z}_{rot} \in \mathbb{R}^3$. Let $\mathcal{A}$ be adsorbate atoms and define an anchor $\mathbf{c}$ (adsorbate center) from a relaxed reference pose $\{\mathbf{x}_i^0\}$, and a centered rigid template $\mathbf{r}_i = \mathbf{x}_i^0 - \mathbf{c}$ for $i \in \mathcal{A}$. Cartesian positions are constructed by

$$\mathbf{x}_i = \mathbf{R}(\mathbf{z}_{rot})\,\mathbf{r}_i + \mathbf{z}_{tr}, \qquad i \in \mathcal{A}, \tag{1}$$

where $\mathbf{R}(\cdot)$ is the exponential map from axis-angle to SO(3). We only transform adsorbate atoms while keeping the slab fixed.

2.2 Flow matching on a linear pose path

AdsorbFlow learns a time-dependent velocity field in pose space. We use a linear corruption path with $t \sim \mathcal{U}(0,1)$:

$$\mathbf{z}_{tr}(t) = \mathbf{c} + t\,\boldsymbol{\epsilon}_{tr}, \quad \mathbf{z}_{rot}(t) = t\,\boldsymbol{\epsilon}_{rot}, \tag{2}$$

where $\boldsymbol{\epsilon}_{tr}, \boldsymbol{\epsilon}_{rot}$ are sampled translation/rotation noises. At $t = 0$, Eq. (2) recovers the relaxed pose (identity rotation, anchor $\mathbf{c}$); at $t = 1$, it yields a noisy pose. The corresponding target velocities are constant:

$$\mathbf{v}_{tr}^{\star} = \boldsymbol{\epsilon}_{tr}, \qquad \mathbf{v}_{rot}^{\star} = \boldsymbol{\epsilon}_{rot}. \tag{3}$$

Algorithm 1 AdsorbFlow training (flow matching).

Input: relaxed pose $\{\mathbf{x}_i^0\}_{i\in\mathcal{A}}$, slab, energy E
(1) Setup
$\mathbf{c} \leftarrow \frac{1}{|\mathcal{A}|}\sum_{i\in\mathcal{A}} \mathbf{x}_i^0$
$\mathbf{r}_i \leftarrow \mathbf{x}_i^0 - \mathbf{c} \;\; \forall i \in \mathcal{A}$
(2) Corrupt pose on linear path
Sample $t \sim \mathcal{U}(0,1)$ and noises $(\boldsymbol{\epsilon}_{tr}, \boldsymbol{\epsilon}_{rot})$
$\boldsymbol{\epsilon}_{tr} \leftarrow \textsc{PBCWrap}(\boldsymbol{\epsilon}_{tr})$
$\boldsymbol{\epsilon}_{rot} \leftarrow \textsc{SymProj}(\boldsymbol{\epsilon}_{rot})$
$\mathbf{z}_{tr}(t) = \mathbf{c} + t\boldsymbol{\epsilon}_{tr}; \quad \mathbf{z}_{rot}(t) = t\boldsymbol{\epsilon}_{rot}$
Build $\mathbf{x}(t)$ via Eq. (1) (slab fixed)
(3) Predict velocity and minimize loss
$(\hat{\mathbf{v}}_{tr}, \hat{\mathbf{v}}_{rot}) \leftarrow f_\theta(\text{slab}, \mathbf{x}(t), t, E)$
$\mathcal{L} = \|\hat{\mathbf{v}}_{tr} - \boldsymbol{\epsilon}_{tr}\|_2^2 + \|\hat{\mathbf{v}}_{rot} - \boldsymbol{\epsilon}_{rot}\|_2^2$
Minimize $\mathcal{L}$

We train f_θ to predict $(\hat{\mathbf{v}}_{tr}, \hat{\mathbf{v}}_{rot})$ from the corrupted structure, time t, and optional energy conditioning, using an ℓ_2 objective. For periodic slabs, we apply PBC wrapping *only* to the rigid anchor displacement in the surface plane (to avoid splitting a molecule across cell boundaries).

2.3 Equivariant backbone and symmetry-aware rotation learning

Backbone-agnostic design. AdsorbFlow is a framework: the velocity field f_θ can be instantiated with different E(3)-equivariant backbones (e.g., PaiNN [5] or EquiformerV2 [6]). Translation is obtained by mean pooling equivariant per-atom vector outputs over adsorbate atoms.

Rotation head choices. We consider two rotation heads for $\hat{\mathbf{v}}_{rot}$:

(i) Torque–inertia readout Given per-atom vector outputs $\mathbf{f}_{rot}^{(i)}$, we compute an angular-momentum-like aggregate

$$\mathbf{L} = \sum_{i\in\mathcal{A}} \mathbf{r}_i \times \mathbf{f}_{rot}^{(i)}, \qquad \mathbf{I} = \sum_{i\in\mathcal{A}} \left(\|\mathbf{r}_i\|^2 \mathbf{I}_3 - \mathbf{r}_i \mathbf{r}_i^\top\right), \tag{4}$$

and solve $(\mathbf{I} + \lambda \mathbf{I}_3)\hat{\boldsymbol{\omega}} = \mathbf{L}$ with Tikhonov regularization. This head exposes per-atom rotational attributions via $\mathbf{r}_i \times \mathbf{f}_{rot}^{(i)}$.

(ii) Direct axis-angle regression Alternatively, we regress $\hat{\mathbf{v}}_{rot} \in \mathbb{R}^3$ directly via pooled equivariant features and an MLP head.

Unless otherwise stated, our main results use **EquiformerV2 + direct axis-angle head**, which achieves the best OC20-Dense success in our study, while PaiNN+torque provides a lightweight and interpretable variant.

Symmetry handling. To mitigate rotational non-identifiability, we classify adsorbates by PCA (spher-

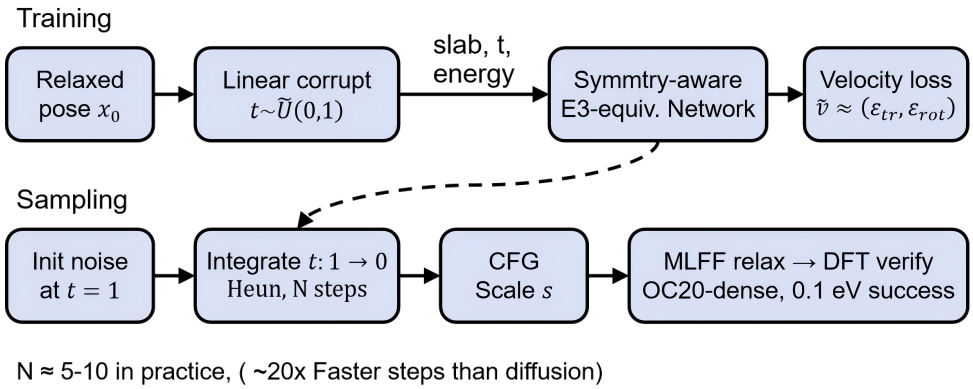

Fig. 1: AdsorbFlow overview. We train a rigid-body velocity field with flow matching on a linear pose interpolant (top), and sample by integrating from noise ($t = 1$) to data ($t = 0$) with a few Heun steps and optional energy-guided CFG (bottom), followed by the standard OC20-Dense evaluation pipeline.

ical/linear/general) and apply a projector to remove symmetry-redundant components; for flip-symmetric linear molecules we align a discrete reflection target for stable supervision.

2.4 *Energy conditioning and classifier-free guidance*

We condition the velocity field on a scalar energy signal via FiLM-style modulation and train with energy dropout [7]. At inference, we use classifier-free guidance (CFG) [8]: $\mathbf{v} = \mathbf{v}_u + s(\mathbf{v}_c - \mathbf{v}_u)$, with guidance scale s, balancing low-energy targeting and diversity.

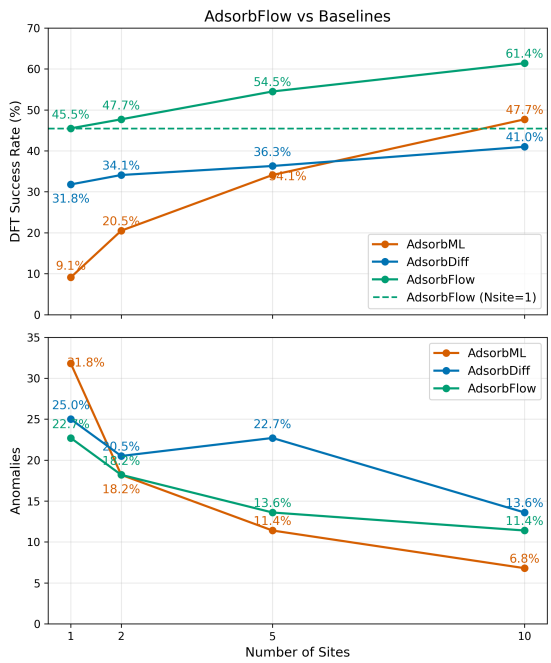

Fig. 2: DFT success rate (top) and anomaly rate after MLFF relaxation (bottom) versus the number of generated candidates N_{site}. Results shown for AdsorbFlow instantiated with **EquiformerV2**. Anomalies include dissociation, desorption, embedding, or surface reconstruction during relaxation.

3. Experiments

Protocol. We follow OC20-Dense and the standard placement evaluation: for each system we sample N_{site} poses, run MLFF relaxation for each, and DFT-verify the single best relaxed candidate (lowest MLFF energy). We report DFT success rate (verified adsorption energy within 0.1 eV of the DFT optimum) and anomaly rate after relaxation. All AdsorbFlow results use Heun sampling with 5 steps and CFG scale $s = 3$.

Results. As shown in Fig. 2, AdsorbFlow consistently outperforms baselines across candidate budgets. With $N_{\text{site}} = 1$, AdsorbFlow achieves **45.5%** DFT success, compared with AdsorbDiff (**31.8%**) and AdsorbML (**9.1%**). Increasing N_{site} further improves success to **61.4%** at $N_{\text{site}} = 10$. Compared with diffusion-based generation, AdsorbFlow yields **lower anomaly rates** after MLFF relaxation across budgets (e.g., 22.7% vs. 25.0% at $N_{\text{site}} = 1$, and 11.4% vs. 13.6% at $N_{\text{site}} = 10$), while requiring only **5** reverse-time integration steps (10 function evaluations with Heun), in contrast to the typical $O(10^2)$ denoising steps of diffusion inference.

4. Conclusion

AdsorbFlow reframes adsorbate placement as rigid-body flow matching with symmetry-aware rotation learning and and energy conditioning with CFG, enabling fast sampling (5-step Heun in our experiments) and strong OC20-Dense success under the MLFF-relax + DFT-verify protocol. Our best results use an EquiformerV2 backbone with a direct rotation head, while a physics-inspired torque–inertia readout provides a lightweight and interpretable alternative.

Acknowledgments

This work was supported by the National Key R&D Program of China (No. 2022ZD0117501), the Scientific Research Innovation Capability Support Project for Young Faculty (ZYGXQNJSKYCXNLZCXM-E7), the Tsinghua University Initiative Scientific Research Program, and the Carbon Neutrality and Energy System Transformation (CNEST) Program.

References

[1] Janice Lan*, Aini Palizhati*, Muhammed Shuaibi*, Brandon M Wood*, Brook Wander, Abhishek Das, Matt Uyttendaele, C Lawrence Zitnick, and Zachary W Ulissi. Adsorbml: Accelerating adsorption energy calculations with machine learning. *arXiv preprint arXiv:2211.16486*, 2022.

[2] Adeesh Kolluru and John R Kitchin. Adsorbdiff: Adsorbate placement via conditional denoising diffusion, 2024.

[3] Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling, 2023.

[4] Lowik Chanussot, Abhishek Das, Siddharth Goyal, Thibaut Lavril, Muhammed Shuaibi, Morgane Riviere, Kevin Tran, Javier Heras-Domingo, Caleb Ho, Weihua Hu, Aini Palizhati, Anuroop Sriram, Brandon Wood, Junwoong Yoon, Devi Parikh, C. Lawrence Zitnick, and Zachary Ulissi. Open catalyst 2020 (oc20) dataset and community challenges. *ACS Catalysis*, 11(10):6059–6072, 2021.

[5] Kristof T. Schütt, Oliver T. Unke, and Michael Gastegger. Equivariant message passing for the prediction of tensorial properties and molecular spectra, 2021.

[6] Yi-Lun Liao, Brandon Wood, Abhishek Das, and Tess Smidt. Equiformerv2: Improved equivariant transformer for scaling to higher-degree representations, 2024.

[7] Ethan Perez, Florian Strub, Harm de Vries, Vincent Dumoulin, and Aaron Courville. Film: Visual reasoning with a general conditioning layer, 2017.

[8] Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance, 2022.