DIFFUSION SAMPLING OF ADSORBATE CONFIGURATIONS ON CATALYST SURFACES

Anonymous authors
Paper under double-blind review

ABSTRACT

Reliable adsorption-energy estimation in heterogeneous catalysis requires sampling diverse low-energy adsorbate configurations on a slab, but existing data-driven methods rely on dense per-system placement labels that are computationally expensive to obtain and typically constrain the search to global rigid-body translations and rotations. We instead cast adsorbate placement as conditional Boltzmann sampling under an energy induced by a pretrained ML interatomic potential, which removes the need for dense supervision and exposes the full adsorbate coordinate space, including internal conformer degrees of freedom. We introduce AdsorbSample, an energy-trained conditional diffusion sampler whose source, controller, and differentiable restraint potentials are tailored to adsorbate–surface geometry, so that physically implausible configurations are suppressed in the proposal. On the OC20-Dense benchmarks, AdsorbSample achieves the strongest MLIP Pass@k across evaluated budgets and substantially improves DFT-verified success at low sampling budgets, while remaining competitive on raw structural validity and diversity.

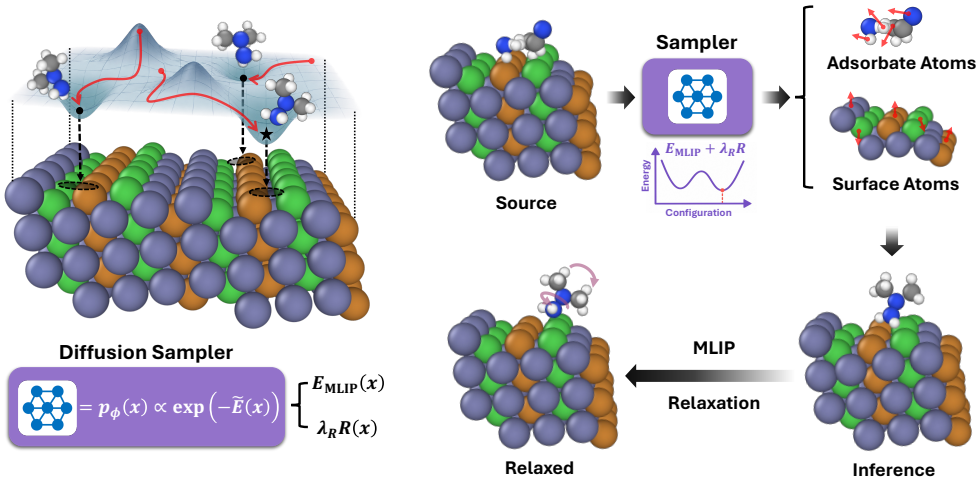


Figure 1: Overview of AdsorbSample. (Left) AdsorbSample trains a diffusion sampler toward the unnormalized density of a target energy, which combines the MLIP energy with geometric restraints. The sampler is trained to generate adsorbate–slab configurations in low-energy regions of the restrained MLIP target. (Right) Starting from a surface-conditioned source, AdsorbSample generates the Cartesian coordinates of the adsorbate atoms (default) and the surface atoms (optional). MLIP relaxation then optimizes the adsorbate and surface atoms jointly and yields the final relaxed configuration.

1 INTRODUCTION

Heterogeneous catalyst discovery hinges on a small number of compact descriptors, most importantly the adsorption energies of key reaction intermediates, which correlate with activation barriers and turnover rates (Nørskov et al., 2009; Medford et al., 2015; Greeley et al., 2006). Computing these

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

descriptors requires identifying the most stable adsorption configuration for each adsorbate–slab pair. This search is challenging: a pair admits many adsorption sites, orientations, and conformers (Figure 1), while a density functional theory (DFT; Kohn & Sham (1965)) relaxation initialized from any one geometry generally reaches only a local minimum and therefore enumeration is computationally slow.

ML interatomic potentials (MLIPs) trained on catalyst configurations (Chanussot et al., 2021) and on broader inorganic and molecular corpora (Chen & Ong, 2022; Deng et al., 2023; Batatia et al., 2025; Wood et al., 2025) now provide computationally inexpensive, differentiable proxies to the DFT potential energy surface across diverse chemistries. Leveraging this, Lan et al. (2023) proposed a search pipeline that enumerates random and heuristic placements of an adsorbate on a slab and relaxes each placement with an MLIP. To evaluate this approach, they also released OC20-Dense, which contains DFT relaxations of 100 random and heuristic placements for each adsorbate–slab system; the lowest adsorption energy among these relaxations serves as the reference for that system. Supervised placement models such as AdsorbDiff (Kolluru & Kitchin, 2024) subsequently learn to propose configurations from OC20-Dense, retaining a dependence on dense placement supervision.

However, constructing such datasets requires many DFT relaxations for every adsorbate–slab system, making extension to new chemical systems expensive. This creates a structural asymmetry in available supervision: pretrained MLIPs make energy and force queries readily available across broad chemical space, while dense samples of low-energy adsorbate configurations remain costly to obtain. This motivates samplers that can exploit the MLIP energy landscape directly, without requiring densely enumerated placements as training targets.

Furthermore, existing approaches (Lan et al., 2023; Kolluru & Kitchin, 2024; Qiu et al., 2026) enumerate or generate only rigid-body translations and rotations of a fixed adsorbate geometry, leaving internal conformational changes to subsequent relaxation. A notable exception is CatFlow (Kim et al., 2026), which co-generates adsorbate and slab atoms in Cartesian space and thus does not impose the rigid-body restriction; however, it is trained as a single-shot generator from supervised placement data and does not address dense *sampling* of diverse low-energy configurations per system.

We address this asymmetry by treating adsorbate placement as conditional sampling from an unnormalized Boltzmann density induced by an MLIP energy, rather than as supervised learning from observed placements. The target requires only energy and gradient queries, which the MLIP supplies, and yields a diverse pool of low-energy candidates that downstream MLIP relaxation and DFT verification can refine. As illustrated in Figure 1, we instantiate this formulation with AdsorbSample, an energy-trained conditional diffusion sampler over the full Cartesian coordinates of the adsorbate.

Our contributions are:

1. We formulate adsorbate placement as conditional sampling from a pretrained MLIP energy landscape, avoiding dense placement supervision.
2. We introduce AdsorbSample, which samples full adsorbate coordinates using a surface-conditioned source, geometric restraints, and a geometry-aware equivariant controller, enabling internal conformational search beyond rigid-body placement.
3. AdsorbSample improves low-energy discovery over search-based and supervised placement baselines on OC20-Dense. We further analyze the roles of pretraining, source placement, controller architecture, restrained relaxation, surface degrees of freedom, and mode coverage.

2 Preliminaries

2.1 Problem formulation

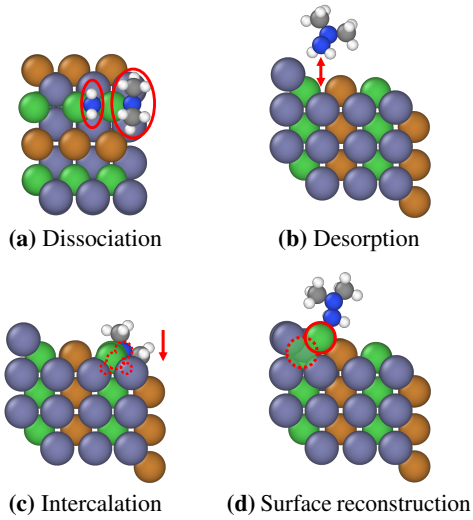
Adsorbate–slab systems. We consider an adsorbate–slab system whose atoms are partitioned into adsorbate atoms $\mathcal{A}$, free surface atoms $\mathcal{S}$, and fixed bulk atoms $\mathcal{C}$. Following common practice in DFT slab calculations (Chanussot et al., 2021), $\mathcal{S}$ relaxes jointly with the adsorbate, whereas $\mathcal{C}$ stays fixed during relaxation. The conditioning information $\mathbf{c}$ comprises the atomic species, the reference slab coordinates, the periodic cell, and the adsorbate bond graph $\mathcal{G}_{\mathcal{A}}$. The graph $\mathcal{G}_{\mathcal{A}}$ specifies the gas-phase covalent connectivity of the adsorbate, which its elemental composition alone cannot determine.

Sampling adsorbate configurations. We sample the adsorbate coordinates $\mathbf{X}_{\mathcal{A}} \in \mathbb{R}^{|\mathcal{A}|\times 3}$, where $\mathbf{X}_{\mathcal{I}}$ denotes the coordinates of the atoms in $\mathcal{I}$. Unlike rigid-body placement, $\mathbf{X}_{\mathcal{A}}$ includes the internal conformational degrees of freedom of the adsorbate. By default, we adopt a *frozen-slab* approximation during sampling, which fixes $\mathbf{X}_{\mathcal{S}}$ and $\mathbf{X}_{\mathcal{C}}$ at the reference slab coordinates and restricts the sampling space to $\mathbf{X}_{\mathcal{A}}$. Given an MLIP energy $E_{\mathrm{MLIP}}(\mathbf{X}_{\mathcal{A}}; \mathbf{c})$ of the adsorbate–slab structure, we define the target distribution π_τ by

$$\pi_\tau(\mathbf{X}_{\mathcal{A}} \mid \mathbf{c}) \propto \exp\big(-E_{\mathrm{MLIP}}(\mathbf{X}_{\mathcal{A}}; \mathbf{c})/\tau\big), \tag{1}$$

where $\tau > 0$ is the temperature in energy units. Unlike energy minimization from a single initialization, sampling from π_τ can visit multiple low-energy basins and thus supplies diverse candidates for relaxation. The normalizing constant of π_τ is intractable, and no samples from π_τ are available, so a sampler can query only E_{MLIP} and its gradient. We learn a single sampler conditioned on $\mathbf{c}$, which amortizes sampling across adsorbate–slab systems and applies to unseen systems without retraining.

Relaxation and invalid adsorption configurations. We relax each generated candidate with the MLIP, updating $\mathbf{X}_{\mathcal{A}}$ and $\mathbf{X}_{\mathcal{S}}$ jointly while keeping $\mathbf{X}_{\mathcal{C}}$ and the periodic cell fixed. The adsorption energy is defined at a local-minimum adsorbate–surface geometry (Chanussot et al., 2021). Relaxation moves a sample from π_τ to such a local energy minimum while allowing the surface to respond to the adsorbate. However, relaxation can also reach a local minimum that no longer contains the intended adsorbate on the intended surface. We classify such a configuration as invalid because its energy does not correspond to the adsorption energy of the intended system. Following Lan et al. (2023) and Kolluru & Kitchin (2024), we consider four invalid cases, namely desorption, dissociation, surface reconstruction, and intercalation (Figure 2), and Appendix B.1 formally defines each case. The goal is to find the lowest-energy valid relaxed configuration for each system, which DFT then verifies.



(a) Dissociation **(b)** Desorption **(c)** Intercalation **(d)** Surface reconstruction

Figure 2: Four post-relaxation invalidity cases.

2.2 Related Work

Catalyst configuration datasets. Large-scale DFT datasets have broadened the chemical space for training and evaluating MLIPs, but dense placement data, which enumerate many initial placements per adsorbate–slab system, remain limited. OC20 (Chanussot et al., 2021) provides DFT relaxations of adsorbates on catalyst surfaces, and its IS2RE task pairs each initial structure with its relaxed structure and energy. OC22 (Tran et al., 2023) extends this effort to oxide electrocatalysts with varying adsorbate coverages, and OC25 (Sahoo et al., 2026) adds solid–liquid interfaces with explicit solvents and ions. In contrast, the only densely sampled dataset, OC20-Dense (Lan et al., 2023), contains only ∼1,000 adsorbate–slab systems per split with 100 random and heuristic placements per system. This scarcity motivates sampling directly from MLIP energies.

Search-based adsorbate placement. Search-based methods enumerate binding sites and orientations, using chemically motivated placement rules or random sampling, and relax each candidate independently. AdsorbML (Lan et al., 2023) accelerates this pipeline by relaxing and ranking a large set of random and heuristic placements with an MLIP before DFT verification. This finite enumeration can miss low-energy basins that none of the chosen placements reach.

Supervised generative placement models. Supervised generative models amortize placement by learning from relaxed configurations. AdsorbDiff (Kolluru & Kitchin, 2024) trains a diffusion model on OC20-Dense after pretraining on OC20 relaxed structures, and AdsorbFlow (Qiu et al., 2026) trains an energy-conditioned flow matching model on OC20-Dense. This dependence on placement

data limits scalability to chemical regimes without densely enumerated configurations, particularly larger and more flexible adsorbates whose conformational spaces are computationally expensive to cover with data. Like AdsorbML, both models generate only the rigid-body translation and rotation of a fixed gas-phase adsorbate geometry, so the internal conformational degrees of freedom of the adsorbate change only incidentally during relaxation. As a result, these methods may miss low-energy basins that require substantially different adsorbate conformations.

Data-free neural sampling. Neural samplers amortize sampling from an unnormalized density using only its log-density and gradient, without samples from the target. Diffusion-based samplers learn stochastic dynamics whose terminal distribution approximates the target (Zhang & Chen, 2022; Vargas et al., 2023; Akhound-Sadegh et al., 2024; Havens et al., 2025; Liu et al., 2025). Unlike supervised diffusion models, they obtain the training signal from the target density rather than from denoising observed samples. We build on Flow Sampling (Havens et al., 2026), which regresses the controller onto self-generated terminal samples and their energy gradients and reuses these pairs through a replay buffer. This design suits our setting, where a pretrained MLIP defines the unnormalized target and supplies its gradient (Section 3.3.2).

3 AdsorbSample

AdsorbSample generates candidate placements by transporting a surface-conditioned source distribution with an equivariant controller, given the conditioning information $\mathbf{c}$ of an adsorbate–slab system. The controller is trained directly against a restrained MLIP energy using Flow Sampling, without samples from the restrained Boltzmann target. At inference, we apply bond-restrained MLIP relaxation, validity filtering, and energy-based ranking to the generated candidates. Throughout this section, we write $\mathbf{X}_t$ for the adsorbate coordinates at diffusion time t. We describe each stage below.

3.1 Surface-conditioned sampler

AdsorbSample generates adsorbate coordinates by integrating the SDE

$$d\mathbf{X}_t = u_t^{\phi}(\mathbf{X}_t, \mathbf{c})\, dt + \sqrt{2\gamma t}\, d\mathbf{W}_t, \qquad \mathbf{X}_0 \sim p_0(\cdot \mid \mathbf{c}), \qquad t \in [0,1], \tag{2}$$

where p_0 is the surface-conditioned source (Section 3.1.1), u_t^{ϕ} is the controller (Section 3.1.2), $\gamma > 0$ sets the noise scale, and $\mathbf{W}_t$ is a standard Wiener process on the adsorbate coordinates. We integrate Equation (2) with the Euler–Maruyama method, and training adjusts u_t^{ϕ} so that $\mathbf{X}_1$ approximately follows the restrained target (Section 3.3.2).

3.1.1 Source distribution

We use a harmonic source on centered local adsorbate coordinates following Jing et al. (2023). Writing $\widetilde{\mathbf{X}}_{\mathcal{A},0} = \{\tilde{\mathbf{x}}_i\}_{i\in\mathcal{A}}$ for the adsorbate coordinates in a local frame, we define

$$p_{\mathrm{harm}}(\widetilde{\mathbf{X}}_{\mathcal{A},0}) \propto \exp\left(-\frac{\alpha}{2}\sum_{i,j\in\mathcal{A}} \|\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_j\|^2\right). \tag{3}$$

We consider two slab-conditioned placement variants built from this same harmonic local distribution: a *centered harmonic* source, which places the adsorbate near the center of the slab, and a *uniform* $\otimes$ *harmonic* source, i.e., the product of a uniform anchor distribution and the harmonic local distribution. The latter samples the anchor uniformly over the slab unit cell. In both cases, the adsorbate is initialized at a sampled height above the top slab layer and shifted upward if needed to satisfy a minimum clearance constraint. Full details are given in Appendix A.1.

3.1.2 Geometry-aware controller

We parameterize the time-dependent control field $u_t^{\phi}(\mathbf{X}_t, \mathbf{c})$ with an equivariant graph neural network specialized to adsorbate–slab systems. The controller receives the current adsorbate coordinates, the reference slab coordinates, the atomic species, the periodic cell, the adsorbate bond graph $\mathcal{G}_{\mathcal{A}}$, and the diffusion time. Under the default frozen-slab approximation, it predicts coordinate updates only

for adsorbate atoms, and the slab atoms provide static geometric context. Section 5.2 examines a variant that also updates $\mathbf{X}_{\mathcal{S}}$.

The controller separates three physically distinct forms of interaction: fixed covalent topology within the adsorbate, intramolecular geometry, and surface interactions. We represent these using the adsorbate bond graph $\mathcal{G}_{\mathcal{A}}$, a complete adsorbate graph, and a local adsorbate–surface graph, respectively. Shortest-path distances in the adsorbate bond graph provide additional topological features, allowing the model to distinguish molecules with identical elemental compositions but different connectivity.

Starting from a PaiNN-style equivariant encoder (Schütt et al., 2021), we introduce several modifications tailored to adsorbate–slab settings. The slab atoms are first contextualized under periodic boundary conditions, after which we replace a single homogeneous interaction pathway with three typed message-passing branches that separately model covalent, nonbonded, and surface interactions. Learned gates combine the three branch outputs as a function of the atom representation and diffusion time, after which a final equivariant readout head produces the drift on adsorbate atoms, $u_t^{\phi}(\mathbf{X}_t, \mathbf{c}) \in \mathbb{R}^{|\mathcal{A}|\times 3}$. Full graph definitions, architectural equations, and feature specifications are given in Appendix A.2.

3.2 MLIP ENERGY AND GEOMETRIC RESTRAINTS

We instantiate E_{MLIP} with the OC20 task of UMA (Wood et al., 2025). Under the frozen-slab approximation, UMA evaluates the full adsorbate–slab structure with the slab at the reference slab coordinates, and its gradient with respect to $\mathbf{X}_{\mathcal{A}}$ supplies the force signal for training the sampler.

Low energy alone does not guarantee a valid adsorption configuration, since the adsorbate may dissociate, desorb, or penetrate beneath the surface. We therefore train the sampler toward the restrained target

$$\widetilde{\pi}_{\tau}(\mathbf{X}_{\mathcal{A}} \mid \mathbf{c}) \propto \exp\big(-E_{\mathrm{MLIP}}(\mathbf{X}_{\mathcal{A}}; \mathbf{c})/\tau - R(\mathbf{X}_{\mathcal{A}}; \mathbf{c})\big), \tag{4}$$

where the dimensionless restraint R does not scale with τ.

The restraint R contains three components. A bond-structure term discourages severe bond deformation and intramolecular clashes, a surface-height term discourages desorption, and a surface-wall term penalizes penetration beneath the exposed surface. The frozen-slab approximation makes a surface-reconstruction penalty unnecessary. These terms provide soft geometric guidance during generation. Appendix A.3 gives the exact restraint definitions and coefficients.

3.3 PRETRAINING AND ENERGY-BASED TRAINING

3.3.1 OC20 IS2RE PRETRAINING

Because samples from the restrained target $\widetilde{\pi}_{\tau}$ are unavailable, energy-based training must learn the transport from the source distribution using only energy and gradient feedback. We reduce the burden of learning this transport from scratch by first initializing the controller to generate plausible adsorbate geometries. Specifically, we pretrain on relaxed structures from the OC20 IS2RE training set (Chanussot et al., 2021) using conditional flow matching (Lipman et al., 2022). Each training structure is obtained by DFT relaxation from a single initial adsorbate placement, so training provides examples of physically plausible adsorbate–slab geometries. We then switch to the Flow Sampling objective, using only MLIP energy and gradient queries. Thus, pretraining initializes the sampler near physically plausible adsorption configurations, reducing the transport that must be learned during energy-based refinement toward low-energy regions.

3.3.2 FLOW SAMPLING

We train AdsorbSample using Flow Sampling (Havens et al., 2026). For each training system $\mathbf{c}$, we generate a terminal configuration $\bar{\mathbf{X}}_1$ from a detached copy of the current sampler and evaluate its target score

$$\mathbf{s}_1 = \nabla_{\mathbf{X}_{\mathcal{A}}} r_{\tau}(\bar{\mathbf{X}}_1; \mathbf{c}), \qquad r_{\tau}(\mathbf{X}; \mathbf{c}) = -E_{\mathrm{MLIP}}(\mathbf{X}; \mathbf{c})/\tau - R(\mathbf{X}; \mathbf{c}). \tag{5}$$

We store $(\mathbf{c}, \bar{\mathbf{X}}_1, \mathbf{s}_1)$ in a replay buffer $\mathcal{B}$.

During optimization, we sample an entry from $\mathcal{B}$, draw a fresh source $\mathbf{X}_0 \sim p_0(\cdot \mid \mathbf{c})$ and $t \sim \mathcal{U}(0,1)$, and construct $\mathbf{X}_t = (1-t)\mathbf{X}_0 + t\bar{\mathbf{X}}_1$. The Flow Sampling target and loss are

$$\mathbf{u}_t^\star = \bar{\mathbf{X}}_1 - \mathbf{X}_0 + \gamma \mathbf{s}_1, \qquad \mathcal{L}_{\mathrm{FS}}(\phi) = \mathbb{E}\left[\left\| u_t^\phi(\mathbf{X}_t, \mathbf{c}) - \mathbf{u}_t^\star \right\|^2\right]. \tag{6}$$

Each MLIP gradient is therefore evaluated only once at a generated endpoint, while the stored endpoint–gradient pair can be reused across source samples and interpolation times. The replay buffer is periodically refreshed with samples from the current sampler.

Damping. Low-temperature energy training can cause successive Flow Sampling updates to rapidly concentrate the sampler on a subset of discovered modes. Following Blessing et al. (2026), we stabilize these updates by regularizing the controller toward a lagged snapshot. At the beginning of each training epoch, we freeze the current controller as $u_t^{\phi_{\mathrm{lag}}}$ and optimize

$$\mathcal{L}_{\mathrm{damp}}(\phi) = \mathbb{E}\left[\left\| u_t^\phi(\mathbf{X}_t, \mathbf{c}) - u_t^{\phi_{\mathrm{lag}}}(\mathbf{X}_t, \mathbf{c}) \right\|^2\right], \qquad \mathcal{L}(\phi) = \mathcal{L}_{\mathrm{FS}}(\phi) + \lambda_{\mathrm{damp}} \mathcal{L}_{\mathrm{damp}}(\phi). \tag{7}$$

The damping term limits abrupt changes to the learned transport between successive updates, helping preserve mode coverage during low-temperature training.

3.4 Inference and bond-restrained relaxation

At inference, we sample adsorbate configurations from the learned dynamics under the frozen-slab approximation, then refine each candidate with UMA relaxation. To reduce relaxation-induced dissociation, we retain the bond-structure restraint used during training and augment the UMA force for $\mathbf{X}_\mathcal{A}$ as

$$\mathbf{F}_{\mathrm{relax}}(\mathbf{X}_\mathcal{A}) = \mathbf{F}_{\mathrm{UMA}}(\mathbf{X}_\mathcal{A}) - \lambda_{\mathrm{bond}} \nabla_{\mathbf{X}_\mathcal{A}} R_{\mathrm{bond}}(\mathbf{X}_\mathcal{A}; \mathcal{G}_\mathcal{A}). \tag{8}$$

During relaxation, we update $\mathbf{X}_\mathcal{A}$ and $\mathbf{X}_\mathcal{S}$ jointly, while $\mathbf{X}_\mathcal{C}$ and the periodic cell remain fixed. We compare bond-restrained and conventional UMA-only relaxation in Section 5.2.

After relaxation, we filter for validity and rank the remaining candidates by their unrestrained energies E_{MLIP} for DFT verification.

4 Experimental Setup

We evaluate AdsorbSample on OC20-Dense (Lan et al., 2023). Following AdsorbDiff (Kolluru & Kitchin, 2024), we use the OC20-Dense ID training systems for energy-based training and evaluate on the ID and OOD validation splits. During energy-based training, each system provides only the conditioning structure $\mathbf{c}$. Supervision comes from UMA energy and force queries, without using OC20-Dense relaxed structures as targets.

We compare against AdsorbML (Lan et al., 2023), which uses heuristic and random placements, and AdsorbDiff (Kolluru & Kitchin, 2024), which learns placements from dense configuration data. Both propose rigid placements of a fixed adsorbate geometry, whereas AdsorbSample generates full adsorbate coordinates. Full hyperparameters are provided in Appendix C.2.

For each evaluation system, we generate and relax $N = 100$ candidates. We filter invalid structures and rank the remaining candidates by their unrestrained UMA energies, with selected candidates verified using DFT single-point calculations. Following AdsorbDiff (Kolluru & Kitchin, 2024) and AdsorbML (Lan et al., 2023), we use a tolerance of $\epsilon_{\mathrm{succ}} = 0.1\,\mathrm{eV}$ for energy-related metrics.

We report four primary metrics. **Validity** is the fraction of generated candidates that pass the post-relaxation anomaly checks. **Diversity** measures the fraction of structurally distinct valid candidates after clustering using slab-aligned adsorbate RMSD (A-RMSD).

MLIP Pass@k measures the probability that a budget of k generated candidates contains a valid structure within ϵ_{succ} of the minimum UMA energy among the OC20-Dense reference structures. For system s, let n_s be the number of generated candidates and c_s the number of valid candidates satisfying the energy threshold. We use the standard unbiased Pass@k estimator (Chen et al., 2021):

$$\mathrm{Pass@}k(s) = 1 - \binom{n_s - c_s}{k} \Big/ \binom{n_s}{k}. \tag{9}$$

DFT SR@k, following AdsorbML (Lan et al., 2023), uses UMA for candidate selection and DFT for verification. Let $\widehat{\mathcal{S}}_{s,1:k}$ denote the first k generated candidates for system s, and let $\widehat{\mathcal{S}}^{\mathrm{val}}_{s,1:k}$ denote the subset that passes the validity checks. We select the valid candidate with the lowest UMA energy,

$$\widehat{\mathbf{X}}^{\star}_{s,k} = \underset{\widehat{\mathbf{X}} \in \widehat{\mathcal{S}}^{\mathrm{val}}_{s,1:k}}{\arg\min} E_{\mathrm{UMA}}(\widehat{\mathbf{X}}), \qquad \mathrm{SR@}k(s) = \mathbf{1}\left[E^{\mathrm{DFT}}_{\mathrm{ads}}(\widehat{\mathbf{X}}^{\star}_{s,k}) - E^{\star,\mathrm{DFT}}_{\mathrm{ads},s} \leq \epsilon_{\mathrm{succ}}\right]. \tag{10}$$

where $E^{\star,\mathrm{DFT}}_{\mathrm{ads},s}$ is the minimum DFT adsorption energy among the OC20-Dense reference structures for system s. If no valid candidate occurs within the first k samples, the system is counted as a failure. Further metric and validity details are given in Appendix B.2.

5 Results and Analysis

Table 1: Result Metrics. All methods use bond-restrained relaxation. For each subset, the best results are shown in **bold**, and the second-best are underlined.

Method	Structural Metrics		MLIP Pass@k [%]				DFT SR@k [%]			
	Val. [%]	Div. [%]	1	2	5	10	1	2	5	10
OC20-Dense ID subset (44 systems)										
AdsorbSample (*this work*)	**81.2**	47.8	**27.2**	**38.7**	**54.7**	**66.0**	**25.0**	**36.4**	**52.3**	56.8
↪ without pretraining	62.3	39.7	15.8	24.6	37.3	46.1	9.1	20.5	34.1	47.7
↪ centered source	56.9	19.1	13.6	19.5	28.4	34.9	9.1	18.2	29.5	34.1
AdsorbDiff (Kolluru & Kitchin, 2024)	64.9	12.5	13.3	18.4	26.0	32.0	18.2	20.5	29.5	34.1
AdsorbML (Lan et al., 2023)	79.6	**50.3**	16.5	27.6	47.3	62.9	13.6	27.3	**52.3**	**63.6**
OC20-Dense OOD subset (50 systems)										
AdsorbSample (*this work*)	80.5	49.5	23.8	**35.9**	**55.8**	**70.9**	**28.0**	**34.0**	40.0	**52.0**
↪ without pretraining	78.1	50.0	21.7	30.6	44.3	55.5	18.0	28.0	36.0	48.0
↪ centered source	62.1	24.3	**24.1**	29.4	36.9	41.5	24.0	28.0	38.0	42.0
AdsorbDiff (Kolluru & Kitchin, 2024)	68.1	17.5	19.3	25.7	33.9	39.5	26.0	26.0	40.0	42.0
AdsorbML (Lan et al., 2023)	**82.6**	**50.6**	17.8	28.5	47.2	63.3	18.0	32.0	**44.0**	**52.0**

5.1 Low-energy placement discovery

Comparison with existing methods. As shown in Table 1, AdsorbSample substantially improves low-energy discovery over existing placement methods. Compared with AdsorbML, the largest gains occur at small sampling budgets: on ID, MLIP Pass@1 increases from 16.5% to 27.2% and DFT SR@1 from 13.6% to 25.0%; on OOD, the corresponding improvements are 17.8% to 23.8% and 18.0% to 28.0%. AdsorbSample also retains an MLIP advantage at larger budgets, reaching Pass@10 of 66.0% vs. 62.9% on ID and 70.9% vs. 63.3% on OOD.

The gains over the supervised AdsorbDiff baseline are larger: AdsorbSample increases MLIP Pass@10 from 32.0% to 66.0% on ID and from 39.5% to 70.9% on OOD, while DFT SR@10 improves from 34.1% to 56.8% and from 42.0% to 52.0%, respectively. These improvements are obtained while maintaining structural validity and diversity comparable to AdsorbML and substantially higher than AdsorbDiff.

The greatest improvements occur at small k, where proposal quality matters most and only a few structures can be forwarded to computationally expensive DFT evaluation. This indicates that AdsorbSample concentrates greater probability mass on useful low-energy configurations, improving the efficiency with which high-quality candidates are discovered.

Discovery beyond OC20-Dense. Because OC20-Dense contains only a finite set of relaxed placements, its reference minimum need not represent the lowest-energy basin accessible to a placement method. Under UMA, AdsorbSample discovers configurations below the OC20-Dense reference minimum on 25/44 ID and 28/50 OOD systems. More substantially, DFT single-point verification confirms configurations at least 0.1 eV below the OC20-Dense DFT minimum on 7/44 ID and 10/50 OOD systems, compared with 6/44 and 3/50 for AdsorbML and 2/44 and 4/50 for AdsorbDiff.

Diversity of low-energy configurations. AdsorbSample discovers substantially greater low-energy configurational diversity than the other placement methods. Among configurations within 0.1 eV of the OC20-Dense UMA reference minimum, it finds 499 distinct basins on ID and 515 on OOD—45% and 26% more than AdsorbML (345/410), and over four times as many as AdsorbDiff (107/106). The difference also persists per system: on OOD, AdsorbSample finds a median of 5 low-energy basins per system, compared with 3 for AdsorbML and 2 for AdsorbDiff. Moreover, 87.2% of its ID basins and 85.8% of its OOD basins do not match any OC20-Dense reference basin. Thus, AdsorbSample not only increases low-energy diversity, but frequently discovers adsorption modes outside those represented by the finite OC20-Dense search. Further details are provided in Appendix C.1.

5.2 Ablations and modeling choices

Pretraining and source distribution. Geometric pretraining substantially improves low-energy discovery (Table 1), with the largest gains on ID: MLIP Pass@1 increases from 15.8% to 27.2% and Pass@10 from 46.1% to 66.0%. The improvement persists on OOD, particularly at larger budgets (55.5% to 70.9% at k=10). Source placement controls a complementary concentration–coverage tradeoff. Relative to the broad source, centering sharply reduces diversity (39.7% to 19.1% on ID; 50.0% to 24.3% on OOD) and slightly improves OOD Pass@1 (21.7% to 24.1%), but performance decreases at larger k (Pass@10: 46.1% to 34.9% on ID and 55.5% to 41.5% on OOD). Thus, pretraining improves low-energy sampling overall, while a broad source is important for covering multiple adsorption basins as the sampling budget grows.

Architecture ablations. We compare our typed-branch controller against PaiNN (Schütt et al., 2021), EquiformerV2 (Liao et al., 2024), and a non-equivariant DiT baseline. Under the same Flow Sampling and evaluation setup, AdsorbSample achieves the strongest OOD Pass@k at every budget. At $k = 10$, it improves over PaiNN, EquiformerV2, and DiT by 22.6%, 88.5%, and 150%, respectively. Full results and architecture details are reported in Appendix D.1.

Surface degrees of freedom. To partially amortize the subsequent surface relaxation, we also train a variant that evolves movable surface atoms during sampling rather than holding the slab fixed. The surface uses a separate equivariant drift head with reduced diffusion noise. Relative to the frozen-surface AdsorbSample under unconstrained relaxation, however, validity decreases by 40% on ID and 37% on OOD, while Pass@10 decreases by 44% and 46%, respectively. This degradation is strongly associated with surface reconstruction, as the fraction of samples flagged for surface change increases nearly $6\times$ on ID (6.4% to 37.1%) and $4\times$ on OOD (10.0% to 37.7%). We therefore retain frozen-surface sampling as the default and leave improved joint adsorbate–surface sampling to future work. Further implementation details and results are provided in Appendix D.2

Bond-restrained relaxation. Bond-restrained relaxation improves post-relaxation validity by suppressing adsorbate dissociation. For AdsorbSample, dissociation decreases from 20.3% to 0% on ID and from 9.1% to 0% on OOD, increasing validity from 65.3% to 81.2% and from 73.1% to 80.5%, respectively. The benefit is most pronounced at larger sampling budgets: Pass@10 increases from 58.2% to 66.0% on ID and from 57.5% to 70.9% on OOD. AdsorbML and AdsorbDiff similarly exhibit reduced dissociation and improved validity and Pass@k, indicating that bond-restrained relaxation provides a general refinement strategy rather than one specific to our sampler. Full results are given in Appendix C.3.

5.3 Mode coverage and sampling robustness

Mode collapse and motif-stratified replay. Strong average discovery rates can conceal failures on individual systems. At low temperatures, energy-based training may repeatedly reinforce common adsorption modes while losing rarer, useful ones. To mitigate this, we use a motif-stratified replay: buffer samples are grouped by the identity of the adsorbate atom contacting the surface and its local surface coordination, and replay samples are drawn approximately uniformly across occupied motif classes rather than uniformly over the buffer. This reduces the tendency of high-frequency motifs to dominate training. Further details are provided in Appendix Section D.3.

We study $C_2H_3O_2$/CaGaN, a system on which both the amortized AdsorbSample model and a model trained only on this system fail to recover the lowest-energy OC20-Dense mode. This reference structure contains a carbon-bound hollow site coordinated by N and Ca. Under standard FIFO replay, this motif is discovered transiently but disappears from the buffer by the end of training (Figure 3b). Motif-stratified replay instead preserves the mode, which comprises 21.9% of the final buffer (Figure 3c). The retained mode also translates to low-energy discovery at inference: the stratified model achieves MLIP Pass@1/5/10 of 14.0%, 53.7%, and 79.5%, whereas the standard AdsorbSample model produces no successful samples. These results indicate that controlling replay composition can prevent the loss of an important mode and improve downstream sampling success.

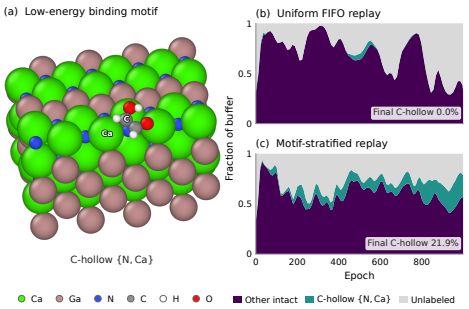


Figure 3: (a) Lowest-energy OC20-Dense C-hollow $\{\mathrm{N}, \mathrm{Ca}\}$ motif for $C_2H_3O_2$/CaGaN, where the carbon binds three or more nearby N/Ca surface atoms. Buffer composition with (b) uniform FIFO and (c) motif-stratified replay.

Scaling to larger flexible adsorbates. We next examine whether full-coordinate sampling can recover low-energy adsorption geometries that require substantial internal conformational changes. We consider guaiacylglycerol-β-guaiacyl ether (GGE, $C_{17}H_{20}O_6$), a 43-atom lignin dimer with seven rotatable bonds. GGE provides a challenging test of conformational search because its low-energy adsorbed geometry on Ni(111) differs substantially from its twisted gas-phase structure. Prior DFT studies find that favorable configurations place both guaiacyl rings approximately parallel to the surface, requiring coordinated changes in the intervening bridge dihedrals (Phongpreecha et al., 2019; Pablo-García et al., 2023). As shown in Figure 4, AdsorbSample recovers the characteristic two-ring adsorbed geometry despite the substantial intramolecular rearrangement required from the gas-phase conformation.

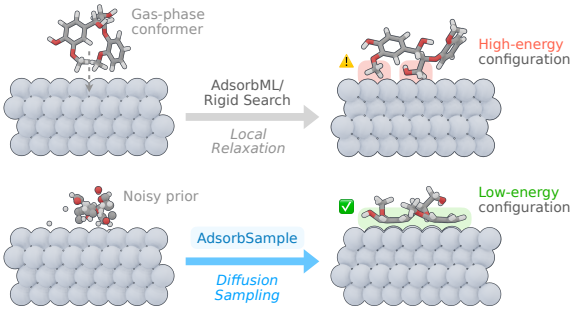


Figure 4: Robustness of AdsorbSample for a flexible adsorbate. The diffusion sampling mechanism facilitates configurational exploration when gas-phase and slab-conditioned energetics and configurations differ significantly.

AdsorbML-style rigid-body placement from standard conformers fails to recover the two-ring adsorbed geometry across 400 trials. Using motif-stratified replay, AdsorbSample reaches the low-energy mode with MLIP Pass@1/2/5/10 of 5.0%, 9.8%, 23.0%, and 41.6%, respectively, relative to the UMA-relaxed literature structure. This demonstrates the benefit of full-coordinate sampling when low-energy adsorption requires substantial intramolecular conformational change. Full experimental details are provided in Appendix Section D.4.

6 Conclusion

We presented AdsorbSample, an energy-trained conditional diffusion sampler for adsorbate placement under a pretrained MLIP energy. By training from energy and forces rather than dense placement labels, AdsorbSample samples full adsorbate coordinates and supports internal conformer motion beyond rigid-body placement. On OC20-Dense benchmarks, AdsorbSample consistently improves MLIP low-energy discovery over existing baselines and substantially improves DFT-verified success at low sampling budgets. Ablations show that pretraining improves the candidate pool, while the source distribution controls the trade-off between concentrated low-budget success and broader multi-basin coverage. Finally, motif-stratified replay can mitigate mode collapse, and experiments on a larger flexible adsorbate demonstrate the value of full-coordinate sampling when low-energy adsorption requires substantial intramolecular rearrangement. Future work includes extending motif-stratified replay to amortized models and developing more robust joint adsorbate–surface sampling.

AI USE STATEMENT

We used generative AI tools to assist with document formatting and presentation. The tools were not used to generate or alter the paper's scientific content. The authors reviewed all AI-assisted changes and take responsibility for the final content of the paper.

REFERENCES

Tara Akhound-Sadegh, Jarrid Rector-Brooks, Avishek Joey Bose, Sarthak Mittal, Pablo Lemos, Cheng-Hao Liu, Marcin Sendera, Siamak Ravanbakhsh, Gauthier Gidel, Yoshua Bengio, et al. Iterated denoising energy matching for sampling from boltzmann densities. In *International Conference on Machine Learning (ICML)*, 2024. (cited on page 4)

Ilyes Batatia, Philipp Benner, Yuan Chiang, Alin M Elena, Dávid P Kovács, Janosh Riebesell, Xavier R Advincula, Mark Asta, Matthew Avaylon, William J Baldwin, et al. A foundation model for atomistic materials chemistry. *The Journal of chemical physics*, 163(18), 2025. (cited on page 2)

Denis Blessing, Lorenz Richter, Julius Berner, Egor Malitskiy, and Gerhard Neumann. Bridge matching sampler: Scalable sampling via generalized fixed-point diffusion matching. In *International Conference on Machine Learning (ICML)*, 2026. (cited on page 6)

Lowik Chanussot, Abhishek Das, Siddharth Goyal, Thibaut Lavril, Muhammed Shuaibi, Morgane Riviere, Kevin Tran, Javier Heras-Domingo, Caleb Ho, Weihua Hu, et al. Open catalyst 2020 (oc20) dataset and community challenges. *Acs Catalysis*, 11(10):6059–6072, 2021. (cited on pages 2, 3, 5, and 19)

Chi Chen and Shyue Ping Ong. A universal graph deep learning interatomic potential for the periodic table. *Nature Computational Science*, 2(11):718–728, 2022. (cited on page 2)

Mark Chen, Jerry Tworek, Heewoo Jun, Qiming Yuan, Henrique Ponde De Oliveira Pinto, Jared Kaplan, Harri Edwards, Yuri Burda, Nicholas Joseph, Greg Brockman, et al. Evaluating large language models trained on code. *arXiv preprint arXiv:2107.03374*, 2021. (cited on pages 6 and 16)

Bowen Deng, Peichen Zhong, KyuJung Jun, Janosh Riebesell, Kevin Han, Christopher J Bartel, and Gerbrand Ceder. Chgnet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nature Machine Intelligence*, 5(9):1031–1041, 2023. (cited on page 2)

Jeff Greeley, Thomas F Jaramillo, Jacob Bonde, IB Chorkendorff, and Jens K Nørskov. Computational high-throughput screening of electrocatalytic materials for hydrogen evolution. *Nature materials*, 5(11):909–913, 2006. (cited on page 1)

Jürgen Hafner. Ab-initio simulations of materials using vasp: Density-functional theory and beyond. *Journal of computational chemistry*, 29(13):2044–2078, 2008. (cited on page 19)

BHLB Hammer, Lars Bruno Hansen, and Jens Kehlet Nørskov. Improved adsorption energetics within density-functional theory using revised perdew-burke-ernzerhof functionals. *Physical review B*, 59(11):7413, 1999. (cited on page 19)

Aaron Havens, Benjamin Kurt Miller, Bing Yan, Carles Domingo-Enrich, Anuroop Sriram, Brandon Wood, Daniel Levine, Bin Hu, Brandon Amos, Brian Karrer, et al. Adjoint sampling: Highly scalable diffusion samplers via adjoint matching. In *International Conference on Machine Learning (ICML)*, 2025. (cited on pages 4 and 14)

Aaron Havens, Brian Karrer, and Neta Shaul. Flow sampling: Learning to sample from unnormalized densities via denoising conditional processes. In *International Conference on Machine Learning (ICML)*, 2026. (cited on pages 4 and 5)

Ask Hjorth Larsen, Jens Jørgen Mortensen, Jakob Blomqvist, Ivano E Castelli, Rune Christensen, Marcin Dułak, Jesper Friis, Michael N Groves, Bjørk Hammer, Cory Hargus, et al. The atomic simulation environment—a python library for working with atoms. *Journal of Physics: Condensed Matter*, 29(27):273002, 2017. (cited on pages 15 and 17)

Bowen Jing, Ezra Erives, Peter Pao-Huang, Gabriele Corso, Bonnie Berger, and Tommi Jaakkola. Eigenfold: Generative protein structure prediction with diffusion models. *arXiv preprint arXiv:2304.02198*, 2023. (cited on page 4)

Minkyu Kim, Kiyoung Seong, Dongyeop Woo, Sungsoo Ahn, and Minsu Kim. On scalable and efficient training of diffusion samplers. *Neural Information Processing Systems (NeurIPS)*, 2025. (cited on page 17)

Minkyu Kim, Nayoung Kim, Honghui Kim, and Sungsoo Ahn. Catflow: Co-generation of slab-adsorbate systems via flow matching. In *International Conference on Machine Learning (ICML)*, 2026. (cited on page 2)

Walter Kohn and Lu Jeu Sham. Self-consistent equations including exchange and correlation effects. *Physical review*, 140(4A):A1133, 1965. (cited on page 2)

Adeesh Kolluru and John R Kitchin. Adsorbdiff: Adsorbate placement via conditional denoising diffusion. In *International Conference on Machine Learning (ICML)*, 2024. (cited on pages 2, 3, 6, 7, 15, and 17)

Georg Kresse and Daniel Joubert. From ultrasoft pseudopotentials to the projector augmented-wave method. *Physical review B*, 59(3):1758, 1999. (cited on page 19)

Janice Lan, Aini Palizhati, Muhammed Shuaibi, Brandon M Wood, Brook Wander, Abhishek Das, Matt Uyttendaele, C Lawrence Zitnick, and Zachary W Ulissi. Adsorbml: a leap in efficiency for adsorption energy calculations using generalizable machine learning potentials. *npj Computational Materials*, 9(1):172, 2023. (cited on pages 2, 3, 6, 7, and 16)

Yi-Lun Liao, Brandon Wood, Abhishek Das, and Tess Smidt. Equiformerv2: Improved equivariant transformer for scaling to higher-degree representations. In *International Conference on Learning Representations (ICLR)*, 2024. (cited on pages 8 and 19)

Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747*, 2022. (cited on page 5)

Dong C Liu and Jorge Nocedal. On the limited memory bfgs method for large scale optimization. *Mathematical programming*, 45(1):503–528, 1989. (cited on page 17)

Guan-Horng Liu, Jaemoo Choi, Yongxin Chen, Benjamin Kurt Miller, and Ricky TQ Chen. Adjoint schrödinger bridge sampler. In *Neural Information Processing Systems (NeurIPS)*, 2025. (cited on page 4)

Andrew J Medford, Aleksandra Vojvodic, Jens S Hummelshøj, Johannes Voss, Frank Abild-Pedersen, Felix Studt, Thomas Bligaard, Anders Nilsson, and Jens K Nørskov. From the sabatier principle to a predictive theory of transition-metal heterogeneous catalysis. *Journal of Catalysis*, 328:36–42, 2015. (cited on page 1)

MPAT Methfessel and Anthony T Paxton. High-precision sampling for brillouin-zone integration in metals. *physical review B*, 40(6):3616, 1989. (cited on page 19)

Juno Nam, Bálint Máté, Artur Toshev, Manasa Kaniselvan, Rafael Gómez-Bombarelli, Ricky TQ Chen, Brandon Wood, Guan-Horng Liu, and Benjamin Miller. Enhancing diffusion-based sampling with molecular collective variables. In *International Conference on Learning Representations (ICLR)*, 2026. (cited on page 14)

Jens Kehlet Nørskov, Thomas Bligaard, Jan Rossmeisl, and Claus Hviid Christensen. Towards the computational design of solid catalysts. *Nature chemistry*, 1(1):37–46, 2009. (cited on page 1)

Sergio Pablo-García, Santiago Morandi, Rodrigo A Vargas-Hernández, Kjell Jorner, Žarko Ivković, Núria López, and Alán Aspuru-Guzik. Fast evaluation of the adsorption energy of organic molecules on metals via graph neural networks. *Nature Computational Science*, 3(5):433–442, 2023. (cited on page 9)

William Peebles and Saining Xie. Scalable diffusion models with transformers. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 4172–4182. IEEE, 2023. (cited on page 19)

Thanaphong Phongpreecha, Jialin Liu, David B Hodge, and Yue Qi. Adsorption of lignin β-o-4 dimers on metal surfaces in vacuum and solvated environments. *ACS Sustainable Chemistry & Engineering*, 7(2):2667–2678, 2019. (cited on page 9)

Jiangjie Qiu, Wentao Li, Honghao Chen, Leyi Zhao, and Xiaonan Wang. Adsorbflow: energy-conditioned flow matching enables fast and realistic adsorbate placement. *arXiv preprint arXiv:2602.19289*, 2026. (cited on pages 2 and 3)

Sushree Jagriti Sahoo, Mikael Maraschin, Joel B Varley, Daniel S. Levine, Zachary Ulissi, C. Lawrence Zitnick, Wayu Takemura, Joseph A. Gauthier, Nitish Govindarajan, and Muhammed Shuaibi. Insights into co dimerization at electrified cu interfaces from large-scale machine learning simulations, 2026. URL https://arxiv.org/abs/2509.17862. (cited on page 3)

Kristof Schütt, Oliver Unke, and Michael Gastegger. Equivariant message passing for the prediction of tensorial properties and molecular spectra. In *International Conference on Machine Learning (ICML)*, 2021. (cited on pages 5, 8, 14, and 19)

Richard Tran, Janice Lan, Muhammed Shuaibi, Brandon M Wood, Siddharth Goyal, Abhishek Das, Javier Heras-Domingo, Adeesh Kolluru, Ammar Rizvi, Nima Shoghi, et al. The open catalyst 2022 (oc22) dataset and challenges for oxide electrocatalysts. *Acs Catalysis*, 13(5):3066–3084, 2023. (cited on page 3)

Francisco Vargas, Will Grathwohl, and Arnaud Doucet. Denoising diffusion samplers. In *International Conference on Learning Representations (ICLR)*, 2023. (cited on page 4)

Brandon M Wood, Misko Dzamba, Xiang Fu, Meng Gao, Muhammed Shuaibi, Luis Barroso-Luque, Kareem Abdelmaqsoud, Vahe Gharakhanyan, John R Kitchin, Daniel S Levine, et al. Uma: A family of universal models for atoms. In *Neural Information Processing Systems (NeurIPS)*, 2025. (cited on pages 2, 5, and 17)

Qinsheng Zhang and Yongxin Chen. Path integral sampler: a stochastic control approach for sampling. In *International Conference on Learning Representations (ICLR)*, 2022. (cited on page 4)

A Additional Definitions

A.1 Source distribution details

Our source distribution $p_0(\mathbf{X}_{\mathcal{A}} \mid \mathbf{c})$ is defined in the frozen-slab setting, with slab coordinates and cell fixed by

$$\mathbf{c} = (\mathbf{Z}_{\mathcal{A}}, \mathcal{G}_{\mathcal{A}}, \mathbf{Z}_{\mathcal{S}}, \mathbf{Z}_{\mathcal{C}}, \mathbf{X}_{\mathcal{S}}, \mathbf{X}_{\mathcal{C}}, \mathbf{L}),$$

and only adsorbate coordinates are sampled.

Harmonic local source. Let $N_{\mathcal{A}} = |\mathcal{A}|$ and let $\widetilde{\mathbf{X}}_{\mathcal{A},0} \in \mathbb{R}^{N_{\mathcal{A}} \times 3}$ denote local adsorbate coordinates before rigid placement. We use a harmonic Gaussian source based on the complete-graph Laplacian

$$L_{\mathrm{cg}} = \left(N_{\mathcal{A}} I_{N_{\mathcal{A}}} - \mathbf{1}\mathbf{1}^{\top}\right) \otimes I_3. \tag{11}$$

If $L_{\mathrm{cg}} = P \,\mathrm{diag}(\lambda)\, P^{\top}$, define

$$T = P \,\mathrm{diag}(\lambda^{-1/2})\, P^{\top}, \tag{12}$$

with $\lambda^{-1/2} = 0$ on the translation nullspace. Then for $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, I_{3N_{\mathcal{A}}})$,

$$\mathrm{vec}(\widetilde{\mathbf{X}}_{\mathcal{A},0}) = \sigma_{\mathrm{ads}}\, T\boldsymbol{\xi}, \tag{13}$$

equivalently

$$\mathrm{vec}(\widetilde{\mathbf{X}}_{\mathcal{A},0}) \sim \mathcal{N}(\mathbf{0}, \sigma_{\mathrm{ads}}^2 L_{\mathrm{cg}}^{\dagger}).$$

Slab-conditioned placement. Let $\mathbf{a}$ and $\mathbf{b}$ be the first two lattice vectors of $\mathbf{L}$, and define

$$\mathbf{n} = \frac{\mathbf{a} \times \mathbf{b}}{\|\mathbf{a} \times \mathbf{b}\|}, \qquad h_{\mathrm{top}} = \max_{j \in \mathcal{S} \cup \mathcal{C}} \mathbf{n}^{\top} \mathbf{x}_j. \tag{14}$$

We then sample a lateral anchor $\mathbf{r}_{\mathrm{xy}}$ and target height h_{tar}.

For the *centered harmonic* source,

$$\mathbf{r}_{\mathrm{xy}} = \mathbf{r}_{\mathrm{ref}} + \sigma_{\mathrm{xy}} \boldsymbol{\varepsilon}_{\mathrm{xy}}, \qquad \boldsymbol{\varepsilon}_{\mathrm{xy}} \sim \mathcal{N}(\mathbf{0}, I_2). \tag{15}$$

For the *uniform* $\otimes$ *harmonic* source,

$$u, v \sim \mathrm{Unif}(0,1), \qquad \mathbf{r}_{\mathrm{xy}} = u\,\mathbf{a}_{1:2} + v\,\mathbf{b}_{1:2}, \tag{16}$$

optionally with small Gaussian jitter. The height is sampled as

$$h_{\mathrm{tar}} = \max(h_{\mathrm{top}} + \mu_z + \sigma_z \varepsilon_z,\ h_{\mathrm{top}} + d_{\mathrm{min}})\,, \qquad \varepsilon_z \sim \mathcal{N}(0,1). \tag{17}$$

The adsorbate is translated rigidly to the chosen anchor and height. If

$$\min_{i \in \mathcal{A}} \mathbf{n}^{\top} \mathbf{x}_i < h_{\mathrm{top}} + d_{\mathrm{min}}, \tag{18}$$

we apply an additional rigid upward shift along $\mathbf{n}$ until the clearance constraint is satisfied.

A.2 Model architecture

We parameterize $u_t^{\phi}(\mathbf{X}_t, \mathbf{c})$ with a typed-branch equivariant graph neural network over the full adsorbate–slab system. The input consists of the current adsorbate coordinates $\mathbf{X}_t = \mathbf{X}_{\mathcal{A},t}$, fixed slab coordinates $(\mathbf{X}_{\mathcal{S}}, \mathbf{X}_{\mathcal{C}})$, atomic numbers, periodic cell $\mathbf{L}$, and adsorbate bond graph $\mathcal{G}_{\mathcal{A}}$. The model predicts nonzero drift only on adsorbate atoms.

Node states and graphs. Each atom i carries scalar features $\mathbf{h}_i \in \mathbb{R}^{d_s}$ and equivariant vector features $\mathbf{v}_i \in \mathbb{R}^{3 \times d_v}$, initialized as

$$\mathbf{h}_i^{(0)} = \mathrm{Emb}(Z_i), \qquad \mathbf{v}_i^{(0)} = \mathbf{0}. \tag{19}$$

Time is encoded by

$$\mathbf{e}_t = \mathrm{MLP}_t(\gamma(t)). \tag{20}$$

At each denoising step, we construct four graphs under the minimum-image convention:

$$\mathcal{E}_{\mathrm{all}} = \{(i,j) : i,j \in \mathcal{A},\ i \neq j\}, \tag{21}$$

$$\mathcal{E}_{\mathrm{bond}} = \mathcal{E}_{\mathcal{A}}, \tag{22}$$

$$\mathcal{E}_{\mathrm{surf}} = \{(i,j) : i \in \mathcal{A},\ j \in \mathcal{S} \cup \mathcal{C},\ r_{ij} < r_{\mathrm{surf}}\}, \tag{23}$$

$$\mathcal{E}_{\mathrm{slab}} = \{(i,j) : i,j \in \mathcal{S} \cup \mathcal{C},\ r_{ij} < r_{\mathrm{slab}}\}. \tag{24}$$

We also compute shortest-path distances on $\mathcal{G}_{\mathcal{A}}$.

Slab encoder and typed branches. Slab atoms are first updated by a PaiNN-style (Schütt et al., 2021) slab encoder on $\mathcal{E}_{\text{slab}}$. The main controller then applies typed interaction layers with three branches: complete adsorbate, bond, and adsorbate–surface. For an edge (i, j), the branch features are

$$\mathbf{e}_{ij}^{(\text{all})} = \left[\text{RBF}(r_{ij}), \text{Emb}_{\text{sp}}(d_{ij}^{\text{sp}}), \text{Emb}_{\text{bond}}(b_{ij})\right], \tag{25}$$

$$\mathbf{e}_{ij}^{(\text{bond})} = \left[\text{RBF}(r_{ij}), \text{Emb}_{\text{bond}}(b_{ij})\right], \tag{26}$$

$$\mathbf{e}_{ij}^{(\text{surf})} = \text{RBF}(r_{ij}), \tag{27}$$

where d_{ij}^{sp} is the graph shortest-path distance and b_{ij} the bond type. Each branch performs PaiNN-style equivariant message passing using sender/receiver hidden states, these edge features, and the normalized relative displacement $\hat{\mathbf{r}}_{ij}$.

Gated fusion and readout. For adsorbate atom i, branch outputs are fused by learned gates conditioned on the hidden state and time embedding. Writing the three branch outputs as $\mathbf{m}_i^{(\text{all})}$, $\mathbf{m}_i^{(\text{bond})}$, and $\mathbf{m}_i^{(\text{surf})}$, the layerwise drift contribution is

$$\mathbf{u}_i^{(\ell)} = \beta_{i,\text{all}}^{(\ell)} U_{\text{all}}^{(\ell)} \mathbf{m}_i^{(\text{all})} + \beta_{i,\text{bond}}^{(\ell)} U_{\text{bond}}^{(\ell)} \mathbf{m}_i^{(\text{bond})} + \beta_{i,\text{surf}}^{(\ell)} U_{\text{surf}}^{(\ell)} \mathbf{m}_i^{(\text{surf})}, \tag{28}$$

while scalar and vector states are updated by residual branch-fusion maps. After L layers, the final drift is

$$u_t^{\phi}(\mathbf{X}_t, \mathbf{c})_i = \sum_{\ell=1}^{L} \mathbf{u}_i^{(\ell)} + \text{Head}\left(\mathbf{h}_i^{(L)}, \mathbf{v}_i^{(L)}, \mathbf{e}_t\right), \qquad i \in \mathcal{A}, \tag{29}$$

with

$$u_t^{\phi}(\mathbf{X}_t, \mathbf{c})_i = \mathbf{0}, \qquad i \in \mathcal{S} \cup \mathcal{C}. \tag{30}$$

Thus the controller separates instantaneous adsorbate geometry, fixed molecular topology, and local surface context while keeping the slab fixed during sampling.

A.3 Restraint potentials

Bond-structure restraint. Let d_{ij} denote the minimum-image distance between adsorbate atoms i and j. We derive a reference pairwise limit ℓ_{ij} from the RDKit molecular graph built from the adsorbate SMILES (with explicit hydrogens added). For bonded pairs $(i, j) \in \mathcal{E}_{\mathcal{A}}$, ℓ_{ij} is constructed from element-specific covalent radii; for non-bonded pairs, ℓ_{ij} is constructed from element-specific van der Waals radii. This restraint is adapted from (Havens et al., 2025; Nam et al., 2026).

Writing $[x]_+ := \max(x, 0)$, the quadratic bond restraint is

$$\begin{aligned} R_{\text{bond}}(\mathbf{X}_{\mathcal{A}}; \mathcal{G}_{\mathcal{A}}) = \alpha \sum_{(i,j)\in\mathcal{E}_{\mathcal{A}}} \frac{1}{2}\left[d_{ij} - c_{\text{long}}\gamma\ell_{ij}\right]_+^2 \\ + \alpha w_{\text{short}} \sum_{(i,j)\in\mathcal{E}_{\mathcal{A}}} \frac{1}{2}\left[\eta\ell_{ij} - d_{ij}\right]_+^2 \\ + \alpha \sum_{(i,j)\notin\mathcal{E}_{\mathcal{A}}} \frac{1}{2}\left[c_{\text{nb}}\gamma^{-1}\ell_{ij} - d_{ij}\right]_+^2, \end{aligned} \tag{31}$$

where η is the minimum bond-length factor, w_{short} weights bonded short-range violations, and the current quadratic implementation uses tighter effective thresholds $c_{\text{long}} = 0.8$ and $c_{\text{nb}} = 1.2$. The first term penalizes stretched bonded pairs, the second penalizes compressed bonded pairs, and the third penalizes near-clashing non-bonded pairs.

Surface-height restraint. Let

$$\mathbf{n} = \frac{\mathbf{a} \times \mathbf{b}}{\|\mathbf{a} \times \mathbf{b}\|} \tag{32}$$

be the surface normal induced by the first two lattice vectors of $\mathbf{L}$, and let

$$h_i = \mathbf{n}^{\top}\mathbf{x}_i, \qquad h_{\text{ref}} = \max_{j\in\mathcal{S}\cup\mathcal{C}} \mathbf{n}^{\top}\mathbf{x}_j. \tag{33}$$

The code defines a target adsorption height

$$h_{\mathrm{eq}} = h_{\mathrm{ref}} + d_{\mathrm{eq}}, \tag{34}$$

and applies a per-atom asymmetric quadratic penalty relative to this target. In the present experiments, the chosen coefficients reduce this to a one-sided attraction term that penalizes atoms lying above the target height, thereby encouraging adsorption-like rather than detached configurations.

Surface wall restraint. The implementation also supports a one-sided harmonic wall of the form

$$R_{\mathrm{wall}} = \frac{k_{\mathrm{wall}}}{2} \sum_{i \in \mathcal{A}} [h_{\mathrm{ref}} - h_i]_+^2, \tag{35}$$

which penalizes adsorbate atoms that move below the top slab plane. When enabled, this term provides an explicit soft barrier against sub-surface penetration.

Overall, the restraint should be interpreted as a geometric prior for frozen-slab sampling: the bond term discourages dissociation and severe self-overlap, the height term discourages desorption, and the wall term can discourage intercalation.

B Evaluation details

B.1 Validity criteria

We adopt the four post-relaxation anomaly criteria of the `DetectTrajAnomaly` utility from Kolluru & Kitchin (2024), as implemented in the `fairchem` package. All four criteria are expressed in terms of an atomic connectivity: the unweighted, undirected graph in which two atoms are linked if and only if their interatomic distance falls within an element-dependent cutoff. We obtain this graph from the neighbor-list utility of the Atomic Simulation Environment (ASE) (Hjorth Larsen et al., 2017), which builds the list under periodic boundary conditions using the minimum-image convention. Two atoms are linked when the distance between them is less than the sum of their per-atom covalent radii (`natural_cutoffs` in ASE). The cutoff radii are often multiplied by $\alpha \geq 1$ to describe loose contacts.

Dissociation. Because AdsorbSample generates adsorbate coordinates directly, a generated configuration may already be dissociated before relaxation. Comparing pre- and post-relaxation connectivity can therefore fail to identify such samples. Instead, we assess the relaxed adsorbate directly against its target SMILES graph. A configuration is flagged as dissociated if the relaxed geometry does not preserve the required adsorbate bonds specified by the SMILES connectivity. This criterion depends only on the final adsorbate geometry and therefore applies consistently to both initially intact and initially dissociated generated configurations.

Desorption. The adsorbate is flagged as desorbed if, in the relaxed structure, no adsorbate atom is linked to any slab atom. The connectivity is computed on the full relaxed adsorbate–slab system with cushion $\alpha = 1.5$; using a cushion here is important because a configuration should be deemed desorbed only when every adsorbate atom is genuinely separated from the surface, not merely sitting just past the strict covalent cutoff (as can occur for physisorbed adsorbates).

Surface reconstruction. The slab is flagged as reconstructed if its connectivity, in the relaxed adsorbate–slab structure, differs from the connectivity of an independently relaxed reference slab beyond a cushion of tolerance. Concretely, the flag is raised in either of two cases. (i) A slab–slab bond is present in the relaxed adsorbate–slab structure (computed at $\alpha = 1$) but is absent from the reference slab even when the reference is computed with the looser cushion $\alpha = 1.5$ (a genuinely new bond). (ii) A slab–slab bond is present in the reference slab (computed at $\alpha = 1$) but is absent from the relaxed adsorbate–slab structure even when the latter is computed with cushion $\alpha = 1.5$ (a genuinely broken bond).

Intercalation. The adsorbate is flagged as intercalated if, in the relaxed structure (computed at $\alpha = 1$), any adsorbate atom is linked to any constrained bulk atom. Intuitively, this catches configurations in which the adsorbate has migrated below the unconstrained surface layers and come into direct covalent-range contact with the frozen subsurface.

B.2 METRIC DEFINITIONS

For each test system $s \in \mathcal{D}$, we generate $N = 100$ candidate adsorbate placements and relax them with the MLIP. Let $\widehat{\mathcal{S}}_s$ denote the resulting MLIP-relaxed candidate set, and $\widehat{\mathcal{S}}_s^{\mathrm{val}} \subseteq \widehat{\mathcal{S}}_s$ the subset that passes the four post-relaxation validity criteria defined in Appendix B.1. For each $\widehat{X} \in \widehat{\mathcal{S}}_s$ we compute the MLIP energy $E_{\mathrm{MLIP}}(\widehat{X}; \mathbf{c})$ and, when scoring final selections, a DFT single-point adsorption energy $E_{\mathrm{ads}}(\widehat{X}; \mathbf{c})$.

Each system is associated with a reference set $\mathcal{R}_s$ of OC20-Dense relaxed structures, and we write $\mathcal{R}_s^{\mathrm{val}} \subseteq \mathcal{R}_s$ for its validity-filtered subset. We define the per-system reference minima under each energy

$$E^{\star}_{\mathrm{MLIP},s} := \min_{Y \in \mathcal{R}_s^{\mathrm{val}}} E_{\mathrm{MLIP}}(Y; \mathbf{c}), \quad E^{\star}_{\mathrm{ads},s} := \min_{Y \in \mathcal{R}_s^{\mathrm{val}}} E_{\mathrm{ads}}(Y; \mathbf{c}), \tag{36}$$

and use a single energy tolerance $\epsilon_{\mathrm{succ}} = 0.1\,\mathrm{eV}$ throughout.

Slab-aligned adsorbate RMSD (A-RMSD). To compare two relaxed adsorbate–slab structures of the same system, we use a slab-aligned adsorbate RMSD that removes two residual ambiguities: in-plane lattice translations between symmetry-equivalent surface sites and permutations of same-element adsorbate atoms. We align on the constrained slab atoms and then solve a species-restricted matching over the adsorbate atoms; we denote the resulting metric by $d(\widehat{X}, Y)$ and provide full details in Appendix B.3.

Validity. Validity is the per-system fraction of generated structures that pass the four post-relaxation criteria, $\mathrm{Val}(s) = |\widehat{\mathcal{S}}_s^{\mathrm{val}}|/|\widehat{\mathcal{S}}_s|$.

Diversity. We greedily cluster $\widehat{\mathcal{S}}_s^{\mathrm{val}}$ under A-RMSD with threshold $d_{\mathrm{cluster}} = 0.05\,\text{Å}$, processing candidates in order of increasing E_{MLIP} and opening a new cluster whenever a candidate lies farther than d_{cluster} from every existing representative; the lowest-MLIP-energy member of each cluster serves as its representative. Letting $\mathcal{C}_s$ denote the resulting set of representatives, diversity is $\mathrm{Div}(s) = |\mathcal{C}_s|/|\widehat{\mathcal{S}}_s|$.

MLIP Pass@k. MLIP Pass@k uses the unbiased combinatorial estimator of Chen et al. (2021) on the MLIP energy. For each system, set $n_s = |\widehat{\mathcal{S}}_s|$ and count valid candidates within ϵ_{succ} of the MLIP reference minimum,

$$c_s = \left| \left\{ \widehat{X} \in \widehat{\mathcal{S}}_s^{\mathrm{val}} : \; E_{\mathrm{MLIP}}(\widehat{X}; \mathbf{c}) - E^{\star}_{\mathrm{MLIP},s} \le \epsilon_{\mathrm{succ}} \right\} \right|, \tag{37}$$

$$\text{MLIP Pass@}k(s) = 1 - \binom{n_s - c_s}{k} \Big/ \binom{n_s}{k}, \tag{38}$$

with MLIP Pass@$k(s) = 0$ when $c_s = 0$ and MLIP Pass@$k(s) = 1$ when $n_s - c_s < k$.

DFT SR@k. Following AdsorbML (Lan et al., 2023), the DFT success rate (SR@k) uses the MLIP energy for selection and DFT for verification. Let $\widehat{\mathcal{S}}_{s,1:k} \subseteq \widehat{\mathcal{S}}_s$ be the first k generated samples in seed/frame order and $\widehat{\mathcal{S}}_{s,1:k}^{\mathrm{val}} = \widehat{\mathcal{S}}_{s,1:k} \cap \widehat{\mathcal{S}}_s^{\mathrm{val}}$ its valid subset. The per-system selection is

$$\widehat{X}^{\star}_{s,k} := \operatorname*{arg\,min}_{\widehat{X} \in \widehat{\mathcal{S}}_{s,1:k}^{\mathrm{val}}} E_{\mathrm{MLIP}}(\widehat{X}; \mathbf{c}), \tag{39}$$

with the system counted as a failure when $\widehat{\mathcal{S}}_{s,1:k}^{\mathrm{val}} = \emptyset$. The per-system success indicator is

$$\mathrm{SR@}k(s) = \mathbf{1}\left[E_{\mathrm{ads}}(\widehat{X}^{\star}_{s,k}; \mathbf{c}) - E^{\star}_{\mathrm{ads},s} \le \epsilon_{\mathrm{succ}} \right]. \tag{40}$$

B.3 SLAB-ALIGNED ADSORBATE RMSD (A-RMSD)

A-RMSD measures geometric similarity between two adsorbate configurations X and Y of the same adsorbate–slab system after factoring out residual rigid-body symmetries of the slab. Two such

configurations may differ by an in-plane lattice translation that maps one site of the periodic surface onto an equivalent site, and by a permutation of same-element adsorbate atoms; A-RMSD eliminates both ambiguities by aligning on the constrained slab atoms $\mathcal{C}$ and then matching adsorbate atoms within each species.

Let $\mathbf{f}_i^X \in \mathbb{R}^3$ denote the fractional coordinates of atom i in configuration X, and let $w(\cdot)$ wrap each in-plane fractional component into $[-\frac{1}{2}, \frac{1}{2})$. Candidate in-plane translations are generated only from $\mathcal{C}$: $\mathcal{T} = \left\{ w\big(\mathbf{f}_{p,xy}^X - \mathbf{f}_{q,xy}^X\big) \,:\, p, q \in \mathcal{C},\ Z_p = Z_q \right\}$. For any $t \in \mathcal{T}$ and any matched pair set M of same-element indices (i, j), define the periodic Cartesian residual and its RMS,

$$\Delta\mathbf{r}_{ij}(t) = \mathbf{L} \begin{bmatrix} w\big(f_{i,x}^X - f_{j,x}^Y - t_x\big) \\ w\big(f_{i,y}^X - f_{j,y}^Y - t_y\big) \\ f_{i,z}^X - f_{j,z}^Y \end{bmatrix}, \tag{41}$$

$$\mathrm{RMS}(M, t) = \left(\frac{1}{|M|} \sum_{(i,j)\in M} \|\Delta\mathbf{r}_{ij}(t)\|^2 \right)^{1/2}. \tag{42}$$

Let $\mathcal{T}^\star \subseteq \mathcal{T}$ be the subset of translations that match the constrained slab, i.e., those under which $\mathcal{C}$ is mapped onto itself by a same-element permutation. For each $t \in \mathcal{T}^\star$, a species-restricted linear-sum assignment over $\mathcal{A}$ minimizing $\sum_{(i,j)\in\pi_\mathcal{A}} \|\Delta\mathbf{r}_{ij}(t)\|^2$ yields an optimal matching $\pi_\mathcal{A}(t)$. The slab-aligned adsorbate RMSD is then

$$\text{A-RMSD}(X, Y) = \min_{t\in\mathcal{T}^\star} \mathrm{RMS}(\pi_\mathcal{A}(t), t). \tag{43}$$

C Experiment Details

C.1 Diversity of low-energy configurations

We analyze the valid relaxed configurations from the main evaluation that lie within 0.1 eV of the OC20-Dense UMA reference minimum. Two configurations of the same system are assigned to the same basin if their slab-aligned adsorbate RMSD (A-RMSD) is at most 0.05 Å and their UMA energies differ by at most 0.02 eV. A-RMSD minimizes over in-plane lattice translations and same-element adsorbate permutations, so distinct adsorption sites and conformers remain separate.

To avoid counting symmetry-equivalent configurations as distinct, we also minimize A-RMSD over point-group operations of the clean slab obtained with `spglib` that preserve the surface normal. Generated basins related by these operations are merged. A generated basin is considered to match OC20-Dense if its symmetry-aware A-RMSD to any reference basin is at most 0.05 Å; otherwise it is counted as a new basin.

C.2 Model and training hyperparameters

All three models use the same typed-branch equivariant controller and the same downstream flow-sampling setup. The shared settings are listed in Table 2. The only differences between models are the source distribution p_0 and whether OC20 IS2RE pretraining is used.

For AdsorbSample, we first pretrain the controller on OC20 IS2RE with conditional flow matching objective $\mathcal{L}_{\mathrm{CFM}}$, using $\gamma_{\mathrm{pre}} = 0$, batch size 64, and 120,000 optimization steps. The downstream amortized run is then initialized from this pretrained controller, and the replay buffer is also initialized by the pretrained controller to start from plausible configurations (Kim et al., 2025). The *centered-source* variant omits this stage and trains only the downstream flow-sampling model from scratch.

C.3 MLIP evaluation and relaxation settings

We use UMA-S-1.1 with the OC20 task head (Wood et al., 2025) as the MLIP for relaxation. Following Kolluru & Kitchin (2024), structures are relaxed with the L-BFGS algorithm (Liu & Nocedal, 1989) as implemented in ASE (Hjorth Larsen et al., 2017), with the constrained slab atoms $\mathcal{C}$ fixed during relaxation. We use a force-convergence threshold $f_{\max} = 0.01$ eV/Å, at most 300 optimization steps, a maximum displacement of 0.04 Å per step, history size 50, damping factor 1.0, and initial inverse-Hessian curvature $\alpha = 70$ eV/Å^2.

Table 2: Shared architecture, energy, restraint, and downstream training hyperparameters used by all three models.

Hyperparameter	Value
Controller widths (d_s, d_v)	$(144, 18)$
Typed interaction layers L	4
Slab encoder layers L_{slab}	2
Cutoffs $(r_{\text{mol}}, r_{\text{surf}}, r_{\text{slab}})$	$(6, 6, 6)$ Å
Downstream objective	$\mathcal{L}_{\text{FS}} + \lambda_{\text{damp}}\mathcal{L}_{\text{damp}}$
Flow-sampling parameters $(\gamma, \tau, \lambda_{\text{damp}})$	$(0.1,\ 5 \times 10^{-3},\ 10)$
Integrator	Euler–Maruyama, 250 steps on $[0, 1]$
Replay buffer size	16,384
Training energy (unitless)	$E_{\text{UMA}}/\tau + R_{\text{bond}} + R_{\text{surf}}$
MLIP instantiation	UMA-S-1.1 with OC20 task head
Bond restraint $(\alpha_{\text{bond}}, \gamma_{\text{bond}}, \eta, w_{\text{short}})$	$(500,\ 1.15,\ 0.55,\ 1.0)$
Surface restraint $(d_{\text{eq}}, k_{\text{attract}}, k_{\text{repel}})$	$(2.0$ Å$,\ 1000,\ 0)$

Table 3: Effect of bond-restrained relaxation. Validity and dissociation are percentages; dissociation denotes the fraction of generated structures flagged as adsorbate-dissociated after relaxation.

Method	Relaxation	Val.	Dissoc.	Pass@1	Pass@2	Pass@5	Pass@10
OC20-Dense ID							
AdsorbSample	unconstrained	65.3	20.3	25.3	35.0	48.1	58.2
	bond-restrained	81.2	0.0	27.2	38.7	54.7	66.0
AdsorbML	unconstrained	77.1	5.1	15.6	25.9	44.1	58.7
	bond-restrained	79.6	0.0	16.5	27.6	47.3	62.9
AdsorbDiff	unconstrained	51.8	31.7	12.7	17.4	24.3	29.2
	bond-restrained	64.9	0.0	13.3	18.4	26.0	32.0
OC20-Dense OOD							
AdsorbSample	unconstrained	73.1	9.1	25.0	34.9	48.1	57.5
	bond-restrained	80.5	0.0	23.8	35.9	55.8	70.9
AdsorbML	unconstrained	82.1	0.5	17.6	28.2	46.4	62.1
	bond-restrained	82.6	0.0	17.8	28.5	47.2	63.3
AdsorbDiff	unconstrained	60.0	18.8	15.7	20.8	27.3	31.3
	bond-restrained	68.1	0.1	19.3	25.7	33.9	39.5

Bond-restrained relaxation. For the bond-restrained variant described in Section 3.4, we augment the UMA force with the adsorbate bond potential used during training. The restraint is constructed from the adsorbate molecular graph with $\lambda_{\text{bond}} = 1$, BondPotential strength $\alpha = 500$, $\gamma = 1.15$, `bond_min_factor`$= 0.55$, and `bond_short_weight`$= 1$, using a quadratic regularizer. No subsequent unconstrained relaxation is applied. The restraint affects only the relaxation trajectory: candidate energies used for ranking and Pass@k are always the unrestrained UMA energies.

Effect of bond-restrained relaxation. Bond-restrained relaxation substantially reduces adsorbate dissociation and improves overall validity across all three methods (Table 3). The effect is largest for AdsorbSample and AdsorbDiff, which have substantial dissociation under unconstrained relaxation. For AdsorbSample, dissociation decreases from 20.3% to 0% on ID and from 9.1% to 0% on OOD, increasing validity from 65.3% to 81.2% and 73.1% to 80.5%, respectively. AdsorbDiff shows a similar reduction (31.7% to 0% on ID and 18.8% to 0.1% on OOD), accompanied by significant validity gains. AdsorbML begins with much lower dissociation (5.1% ID and 0.5% OOD), and accordingly benefits less. These validity gains generally translate to improved low-energy discovery, particularly at larger budgets: AdsorbSample Pass@10 increases from 58.2% to 66.0% on ID and 57.5% to 70.9% on OOD, while AdsorbDiff and AdsorbML also improve across Pass@k. Thus, the bond restraint primarily improves downstream sampling by preventing promising generated configurations from dissociating during local relaxation.

C.4 DFT EVALUATION SETTINGS

Single-point DFT calculations were performed with VASP 6.4.3 (Hafner, 2008) on the generated adsorbate–slab structures. We used the Revised Perdew–Burke–Ernzerhof (RPBE) exchange-correlation functional (Hammer et al., 1999) with projector augmented wave (PAW) pseudopotentials (Kresse & Joubert, 1999) from the VASP PBE 5.4 potential set and a plane-wave kinetic energy cutoff of 350 eV, following the OC20 convention (Chanussot et al., 2021). Brillouin zone sampling used a Monkhorst–Pack mesh determined from the cell dimensions. For each in-plane lattice direction, the number of k-points was set to $\max(1, \mathrm{round}(40/L))$, where L is the corresponding lattice vector length in Å; one k-point was used along the surface normal. Structures were evaluated without further ionic relaxation. Electronic minimization used Methfessel–Paxton smearing (Methfessel & Paxton, 1989) with a smearing width of 0.2 eV and was converged to an electronic tolerance of 10^{-4} eV, with a maximum of 300 electronic iterations. The DFT energy used for evaluation was the final zero-smearing extrapolated total energy reported by VASP. Calculations that failed, did not produce a valid final energy, or reached the electronic iteration limit were treated as unconverged and assigned high energy in the success rate analysis.

D ADDITIONAL RESULTS

D.1 ARCHITECTURE ABLATIONS

The AdsorbSample controller is designed around the adsorbate–slab structure: it separately models fixed covalent topology, instantaneous intramolecular geometry, and surface interactions. We test whether this problem-specific decomposition is necessary, or whether comparable performance can be obtained with more generic neural backbones.

We consider three alternatives spanning different architectural biases. PaiNN (Schütt et al., 2021) provides a standard $E(3)$-equivariant message-passing baseline with coupled scalar and vector features. EquiformerV2 (Liao et al., 2024) tests whether a more expressive higher-order equivariant Transformer can replace the typed interaction structure. Finally, we adapt the non-equivariant Diffusion Transformer (DiT; Peebles & Xie (2023)), providing a substantially higher-capacity baseline without explicit geometric equivariance. The DiT uses an encoder–trunk–decoder architecture with token width 512 and pair width 256, with dense attention over adsorbate and surface atoms.

In each case, the alternative backbone replaces only the drift controller. The source distribution, Flow Sampling objective, UMA energy, geometric restraints, replay buffer, integrator, sampling budget, and evaluation protocol are otherwise shared. Batch size is adjusted for memory. Architecture-specific settings are reported in Table 5.

Table 4: Architecture ablation on OC20-Dense OOD. We use unconstrained UMA relaxation for all methods.

Subset	Model	Validity	Pass@1	Pass@2	Pass@5	Pass@10
OOD	AdsorbSample	**73.1**	**25.0**	**34.9**	**48.1**	**57.5**
	PaiNN	67.8	17.6	25.4	37.2	46.9
	EquiformerV2	38.4	11.4	15.8	23.4	30.5
	DiT	26.2	6.0	10.0	16.4	23.0

Results. The typed-branch controller consistently outperforms all three generic backbones on OOD systems across every sampling budget (Table 4). At Pass@10, AdsorbSample reaches 57.5%, compared with 46.9% for PaiNN, 30.5% for EquiformerV2, and 23.0% for DiT, corresponding to relative improvements of 22.6%, 88.5%, and 150%, respectively. The gap is already present at Pass@1 and widens with sampling budget, indicating that the controller improves not only the probability of finding one successful configuration but also the ability to generate useful low-energy candidates across repeated samples.

Notably, increasing generic backbone complexity does not close the gap: EquiformerV2 underperforms the simpler PaiNN baseline, while the 101.9M-parameter DiT performs worst despite having over 30× more parameters than the equivariant models. These results suggest that adsorbate-

Table 5: Controller-backbone settings for architecture ablations. AdsorbSample settings are reported in Table 2. DiT depth is the number of encoder, trunk, and decoder blocks, and its hidden-channel entry is the token width. Its cutoff is the Gaussian radial-basis support rather than a neighbor-graph radius.

Setting	PaiNN	EquiformerV2	DiT
Parameters	3.191M	3.124M	101.933M
Depth	4	4	2 + 16 + 2
Hidden channels	240	96	512
Cutoff	6.0 Å	6.0 Å	6.0 Å
Radial basis size	18	256	16
$l_{\max}$	–	2	–
Attention heads	–	4	4 / 8 / 4
Batch size	128	8	48

specific geometric inductive biases—in particular, separating covalent, intramolecular, and surface interactions—are more important for this task than generic model capacity alone.

D.2 Joint adsorbate–surface sampling

Motivation. The default AdsorbSample model generates adsorbate coordinates while holding the surface fixed, and surface relaxation is deferred to the subsequent MLIP relaxation. We investigate whether part of this relaxation can instead be amortized by the sampler itself. To do so, we train a variant that jointly evolves the adsorbate and the unconstrained surface atoms during Flow Sampling, while the constrained bulk atoms remain fixed. The goal is for the generated surface geometry to already incorporate part of the adsorbate-induced surface response before local relaxation.

Joint adsorbate–surface dynamics. We maintain separate stochastic dynamics for adsorbate and surface coordinates. The adsorbate follows the standard AdsorbSample dynamics with $\gamma_{\text{ads}} = 0.1$, while the surface uses a smaller diffusion coefficient $\gamma_{\text{surf}} = 0.02$ to limit large slab perturbations. The controller retains the standard adsorbate drift head and adds a separate equivariant output head that predicts velocities for movable surface atoms. Thus, the model predicts

$$u_t^{\phi,\text{ads}} \in \mathbb{R}^{|\mathcal{A}|\times 3}, \qquad u_t^{\phi,\text{surf}} \in \mathbb{R}^{|\mathcal{S}|\times 3}.$$

The adsorbate is initialized from the same source distribution as the standard model. In contrast, surface atoms begin exactly at their clean-slab reference coordinates; we apply no source noise to the surface. Bulk atoms remain fixed throughout.

Surface restraint and target energy. Allowing the slab to move introduces additional degrees of freedom that can produce unrealistically large surface distortions. We therefore tether each movable surface atom to its clean-slab reference position $\mathbf{X}_{\mathcal{S}}^{\text{ref}}$ using

$$R_{\text{tether}} = \frac{k_{\text{tether}}}{2} \sum_{i\in\mathcal{S}} \left\|\mathbf{x}_i - \mathbf{x}_i^{\text{ref}}\right\|^2, \qquad k_{\text{tether}} = 5. \tag{44}$$

The unitless potential used to compute the Flow Sampling score is therefore

$$U_{\text{mov}} = \frac{E_{\text{UMA}}}{\tau} + R_{\text{bond}} + R_{\text{ads-surf}} + R_{\text{tether}}, \qquad \tau = 0.005, \tag{45}$$

where the first three terms are the same UMA energy and adsorbate restraints used by the fixed-surface model. Gradients of U_{mov} are taken with respect to both adsorbate and movable surface coordinates.

Training objective. Flow Sampling targets are computed separately for adsorbate and surface atoms, using their respective diffusion coefficients. We likewise compute separate matching losses, $\mathcal{L}_{\text{FS}}^{\text{ads}}$ and $\mathcal{L}_{\text{FS}}^{\text{surf}}$, and optimize

$$\mathcal{L}_{\text{mov}} = \mathcal{L}_{\text{FS}}^{\text{ads}} + \lambda_{\text{surf}} \mathcal{L}_{\text{FS}}^{\text{surf}}, \qquad \lambda_{\text{surf}} = 100. \tag{46}$$

The raw surface matching error is substantially smaller than the adsorbate error, so λ_{surf} prevents the surface-learning signal from being overwhelmed by the adsorbate term. Except for the additional

surface dynamics, output head, and tethering potential, training follows the same Flow Sampling procedure as the fixed-surface model.

Table 6: Joint adsorbate–surface sampling on OC20-Dense under unconstrained UMA relaxation. Surface change is the fraction of samples flagged by the surface-reconstruction validity criterion.

Subset	Sampling	Validity	Surf. change	Pass@1	Pass@2	Pass@5	Pass@10
ID	Frozen surface	65.3	6.4	25.3	35.0	48.1	58.2
	Movable surface	39.5	37.1	10.3	15.9	24.9	32.6
OOD	Frozen surface	73.1	10.0	25.0	34.9	48.1	57.5
	Movable surface	46.0	37.7	12.1	16.8	24.5	30.9

Low-energy discovery. Joint surface sampling does not improve placement discovery under the current setup (Table 6). On ID, Pass@10 decreases from 58.2% for the fixed-surface model to 32.6%, while on OOD it decreases from 57.5% to 30.9%. Validity similarly decreases from 65.3% to 39.5% on ID and from 73.1% to 46.0% on OOD. The validity loss is strongly associated with slab reconstruction: 37.1% of ID and 37.7% of OOD samples are flagged for surface change, compared with 6.4% and 10.0% for the fixed-surface model. Moreover, 18.4% and 19.8% of movable-surface samples fail solely because of this criterion, with no accompanying dissociation, desorption, or intercalation.

D.3 Motif-stratified buffer

Motivation. Under standard replay, frequently generated adsorption modes can progressively dominate the buffer. This creates a positive-feedback loop: a common motif is sampled more often for training, making the model still more likely to generate it, while rarer modes receive fewer updates and can eventually disappear. Motif-stratified replay instead allocates replay probability at the level of adsorption motifs, reducing this dependence on their raw frequency in the buffer.

Motif assignment. We represent each adsorption configuration by a discrete motif

$$\text{motif} = (\text{contact atom},\ \text{site type},\ \text{neighboring surface species}).$$

The contact atom is the heavy adsorbate atom with the largest number of nearby surface atoms, using a $1.2\times$ covalent-radius cutoff; ties are broken by contact distance. We represent the contact atom by its molecular symmetry class, so chemically equivalent atoms share the same label. The site type is classified as *ontop*, *bridge*, or *hollow* depending on whether the contact atom has one, two, or at least three surface neighbors, and we record the elemental identities of those neighbors. For example, a carbon atom bound in a hollow site coordinated by N and Ca defines a different motif from an oxygen atom bound at an on-top Ca site. Configurations without a well-defined surface contact or with broken adsorbate connectivity are left unlabeled.

Stratified replay. Let n_s denote the number of buffer samples for system s, and let n_m denote the number assigned to motif m. A motif becomes active once at least three examples are present. If M_s motifs are active, a sample i belonging to active motif $m(i)$ receives replay weight

$$w_i = \frac{p_{\text{explore}}}{n_s} + (1 - p_{\text{explore}})\frac{1}{M_s n_{m(i)}}, \tag{47}$$

while inactive or unlabeled samples receive only the uniform term. Thus, p_{explore} preserves uniform exploration, while the remaining mass is divided equally across active motifs regardless of their frequency in the buffer. In the $C_2H_3O_2$/CaGaN experiment, p_{explore} is annealed linearly from 1 to 0.4 over the first 40 flow-training epochs. We additionally use a motif retention floor of two: the two newest samples from each labeled motif are protected from FIFO eviction, while the remaining buffer capacity follows the standard FIFO policy. The buffer contains at most 1024 endpoints.

D.4 LARGER ADSORBATE RESULTS

System. We study guaiacylglycerol-β-guaiacyl ether (GGE, $C_{17}H_{20}O_6$), a 43-atom lignin dimer adsorbed on an $8 \times 8 \times 4$ Ni(111) slab. GGE contains seven rotatable bonds and two aromatic rings. Its low-energy adsorbed geometry places both rings approximately parallel to the surface, requiring substantial internal rearrangement from the twisted gas-phase structure.

Training setup. We train a single-system AdsorbSample model using the same Flow Sampling framework as the amortized model, together with motif-stratified replay (Appendix D.3). We increase the sampling temperature from $\tau = 0.005$ to $\tau = 0.02$ to encourage conformational exploration. The source places the molecular center of mass approximately 2.7 Å above the Ni surface. We retain the standard bond restraint and add the GGE-specific geometric terms described below.

Before insertion into the replay buffer, each newly generated endpoint undergoes up to 10 steps of restrained L-BFGS refinement with $f_{\max} = 0.02$ eV/Å. The refinement minimizes

$$E_{\text{refine}}(\mathbf{X}) = E_{\text{UMA}}(\mathbf{X}) + \tau R_{\text{GGE}}(\mathbf{X}), \tag{48}$$

where R_{GGE} contains the bond and GGE-specific restraints used during training. This provides a short local refinement before the endpoint is stored for replay, while leaving the larger-scale conformational search to the sampler.

GGE-specific restraints. The surface restraints are one-sided quadratic potentials. For a scalar surface distance d, attractive and repulsive terms take the form

$$R_{\text{attr}}(d) = \frac{k}{2}[d - d_0]_+^2, \qquad R_{\text{floor}}(d) = \frac{k}{2}[d_0 - d]_+^2, \tag{49}$$

respectively, where $[x]_+ = \max(x, 0)$. We apply these to the molecular center of mass, heavy atoms, and hydrogens as summarized in Table 7. The COM attraction and heavy-atom attraction prevent desorption and encourage binding, and the atom–surface floor prevents penetration.

For each aromatic ring with atom positions $\{\mathbf{r}_i\}_{i=1}^n$, let $\mathbf{c} = n^{-1}\sum_i \mathbf{r}_i$ and

$$S = \sum_i (\mathbf{r}_i - \mathbf{c})(\mathbf{r}_i - \mathbf{c})^\top.$$

The RMS deviation from the best-fit plane is

$$\rho = \sqrt{\lambda_{\min}(S)/n},$$

and the planarity restraint is

$$R_{\text{ring}} = \frac{1}{2} k n [\rho - d_0]_+^2. \tag{50}$$

Thus, deviations smaller than d_0 incur no penalty. Ring hydrogens are measured relative to the same best-fit plane and receive $\frac{1}{2}k[|d_H| - d_{H,0}]_+^2$ individually.

Rigid-placement baseline. To separate adsorption-site search from intramolecular conformational search, we compare against AdsorbML-style rigid-body placement. Starting from standard RDKit conformers, 400 independent placement-and-relaxation trials fail to recover the characteristic low-energy two-ring adsorbed geometry.

Table 7: Additional geometric restraints for GGE/Ni(111). The standard bond restraint is retained from the main model.

Restraint	Setting	Purpose
COM attraction	$d_0 = 2.5$ Å, $k = 200$	Prevent desorption
Atom–surface floor	$d_0 = 1.5$ Å, $k = 200$	Prevent penetration
Heavy-atom attraction	$d_0 = 2.5$ Å, $k = 4$	Encourage surface contact
Hydrogen floor	$d_0 = 1.9$ Å, $k = 200$	Prevent H–Ni clashes
Ring planarity	$k = 300$, $d_0 = 0.06$ Å, $d_{H,0} = 0.6$ Å	Preserve aromatic geometry