Diffusion Sampling of Adsorbate Configurations on Catalyst Surfaces

Anonymous Authors[1]

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 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 best success rate at low sample budgets while matching or exceeding baselines on validity and diversity.

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 through scaling and volcano-type relations (Nørskov et al., 2009; Medford et al., 2015; Greeley et al., 2006). Computing these descriptors at scale is bottlenecked not by the energy evaluation itself but by the global search over adsorbate configurations: a single adsorbate–slab pair admits many adsorption sites, orientations, and conformers, and a density functional theory (DFT) relaxation initialized from any one geometry returns only a local minimum. The relevant quantity for screening is the lowest-energy basin among these local minima, so accurate adsorption-energy estimation reduces to sampling diverse low-energy configurations from a high-dimensional, multimodal landscape.

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 cheap, differentiable proxies for the underlying potential energy surface across diverse chemistries. This creates a *structural asymmetry* in available supervision: MLIPs span broad chemical space, while datasets that densely enumerate adsorbate placements per system, used by AdsorbML (Lan et al., 2023) and supervised generative models such as AdsorbDiff (Kolluru & Kitchin, 2024), remain expensive and small, capping the scalability of supervised placement learning. Existing approaches (Lan et al., 2023; Kolluru & Kitchin, 2024; Qiu et al., 2026) further restrict the search space by sampling only the global rigid-body translation and rotation of the adsorbate before relaxation, leaving internal conformer degrees of freedom unexplored. 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 (Figure 1). 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. We instantiate this view with an energy-trained conditional diffusion sampler over adsorbate coordinates, drawing on recent diffusion sampler methodologies (Zhang & Chen, 2021; Vargas et al., 2023; Akhound-Sadegh et al., 2024; Havens et al., 2025; Liu et al., 2025). To reduce the cost of post hoc filtering of unphysical configurations, we further guide the sampler at training time with differentiable restraint potentials that penalize desorption, severe atomic overlap, and intercalation, so that anomalies are suppressed in the proposal rather than discarded after relaxation.

---

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

Preliminary work. Submitted to the AI for Science workshop (ICML 2026). Do not distribute.

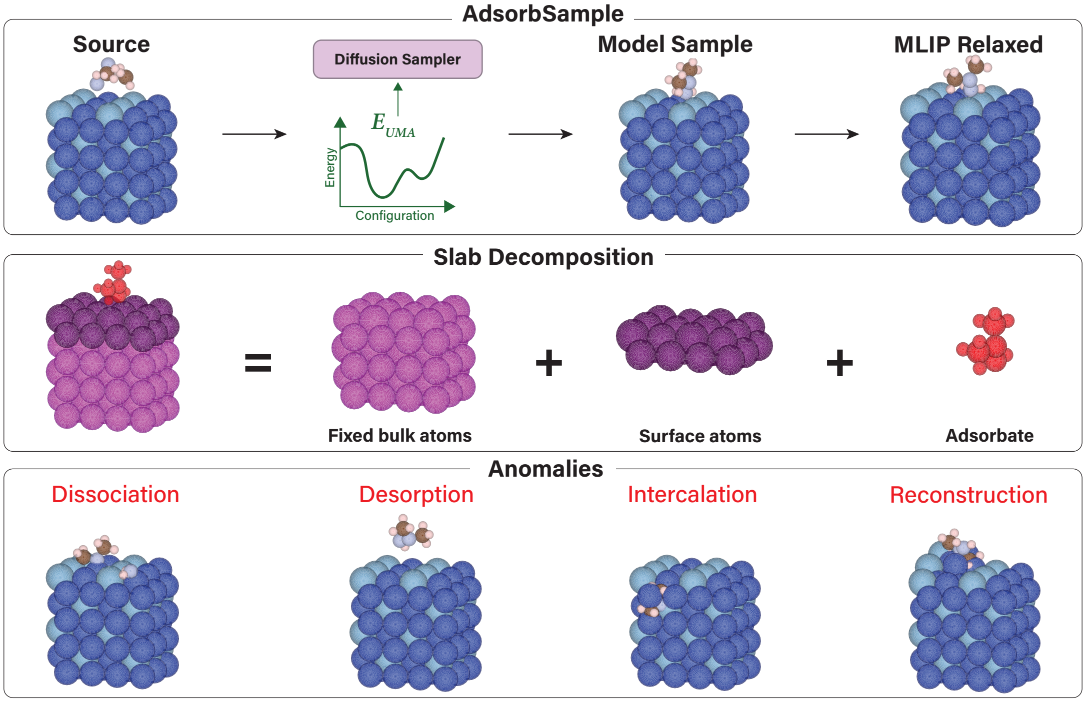


Figure 1. **Overview of AdsorbSample.** (Top) AdsorbSample proposes adsorbate configurations on a catalyst slab under MLIP energy guidance, followed by MLIP relaxation. (Middle) Following OC20 convention (Chanussot et al., 2021), we represent each system as constrained bulk atoms, unconstrained surface atoms, and adsorbate atoms. (Bottom) We validate generated structures using anomaly-based validity criteria.

To our knowledge, this is the first energy-trained conditional diffusion sampler designed specifically for adsorbate–slab configuration generation.

Our contributions are:

1. We formulate adsorbate placement as conditional sampling from an energy landscape induced by a pretrained MLIP, removing the requirement of dense placement labels per adsorbate–surface combination.
2. We introduce AdsorbSample, an energy-trained conditional diffusion sampler over full adsorbate coordinates, allowing internal conformer degrees of freedom beyond rigid-body translation and rotation.
3. We design differentiable restraint potentials that bias proposals away from desorption, severe overlap, and intercalation before relaxation.

2. Background and Preliminaries

2.1. Problem setting

We consider an adsorbate–slab system whose atoms partition into three index sets: the adsorbate $\mathcal{A}$, the unconstrained slab atoms $\mathcal{S}$, and the constrained slab atoms $\mathcal{C}$. In OC20-style configurations, $\mathcal{C}$ corresponds to the bulk layers held fixed during DFT relaxation, whereas $\mathcal{S}$ corresponds to the upper surface layers that are relaxed jointly with the adsorbate. The system is described by atomic numbers $\mathbf{Z}_{\mathcal{A}}, \mathbf{Z}_{\mathcal{S}}, \mathbf{Z}_{\mathcal{C}}$, slab coordinates $\mathbf{X}_{\mathcal{S}} \in \mathbb{R}^{|\mathcal{S}|\times 3}$ and $\mathbf{X}_{\mathcal{C}} \in \mathbb{R}^{|\mathcal{C}|\times 3}$, and periodic cell $\mathbf{L} \in \mathbb{R}^{3\times 3}$. We additionally provide the adsorbate molecular graph $\mathcal{G}_{\mathcal{A}} = (\mathcal{A}, \mathcal{E}_{\mathcal{A}})$, whose edges $\mathcal{E}_{\mathcal{A}} \subseteq \mathcal{A} \times \mathcal{A}$ encode the covalent bond connectivity of the gas-phase adsorbate. This connectivity is fixed across the conditioning and is independent of the current placement of $\mathbf{X}_{\mathcal{A}}$; it supplies internal structure information that is not recoverable from $\mathbf{Z}_{\mathcal{A}}$ alone (e.g., it distinguishes isomers). The constrained atoms $\mathcal{C}$ are held fixed throughout, both during sampling and during any downstream relaxation. The variable to be sampled is the adsorbate coordinate matrix $\mathbf{X}_{\mathcal{A}} \in \mathbb{R}^{|\mathcal{A}|\times 3}$, conditioned on $\mathbf{c} = (\mathbf{Z}_{\mathcal{A}}, \mathcal{G}_{\mathcal{A}}, \mathbf{Z}_{\mathcal{S}}, \mathbf{Z}_{\mathcal{C}}, \mathbf{X}_{\mathcal{S}}, \mathbf{X}_{\mathcal{C}}, \mathbf{L})$.

We adopt a *frozen-slab* approximation during sampling: while the sampler runs, both $\mathbf{X}_{\mathcal{S}}$ and $\mathbf{X}_{\mathcal{C}}$ are held at their reference values and only $\mathbf{X}_{\mathcal{A}}$ varies. This isolates configurational search over adsorbate placement from slab relaxation. Slab relaxation is recovered downstream by an MLIP relaxation that updates $\mathbf{X}_{\mathcal{A}} \cup \mathbf{X}_{\mathcal{S}}$ jointly while keeping $\mathbf{X}_{\mathcal{C}}$ and $\mathbf{L}$ fixed.

Let $E(\cdot)$ denote the DFT total energy of an atomic system. For a relaxed adsorbate–slab structure with adsorbate coordinates $\mathbf{X}^\star_{\mathcal{A}}$, the adsorption energy is

$$\begin{aligned} E_{\text{ads}}(\mathbf{X}^\star_{\mathcal{A}};\mathbf{c}) &= E(\mathbf{X}^\star_{\mathcal{A}},\mathbf{X}^\star_{\mathcal{S}},\mathbf{X}_{\mathcal{C}};\mathbf{c}) \\ &\quad - E^{\text{slab}}(\mathbf{X}_{\mathcal{S}},\mathbf{X}_{\mathcal{C}};\mathbf{Z}_{\mathcal{S}},\mathbf{Z}_{\mathcal{C}},\mathbf{L}) - E^{\text{ads}}(\mathbf{Z}_{\mathcal{A}},\mathcal{G}_{\mathcal{A}}), \end{aligned} \tag{1}$$

where the last two terms are constant for fixed $\mathbf{c}$, so identifying the most stable adsorption configuration reduces to minimizing the relaxed total energy over $\mathbf{X}_{\mathcal{A}}$. Adsorption landscapes are highly multimodal across binding sites, orientations, and conformers, and small perturbations of an initial $\mathbf{X}_{\mathcal{A}}$ can yield distinct relaxed minima; reliable estimation therefore requires sampling a diverse pool of low-energy candidates rather than optimizing a single initialization.

Now, given a differentiable surrogate energy $\widetilde{E}(\mathbf{X}_{\mathcal{A}};\mathbf{c})$ in the *frozen-slab* setting (specified concretely in Section 2.3 and instantiated in Section 3), the induced conditional Boltzmann target is

$$\pi_\tau(\mathbf{X}_{\mathcal{A}} \mid \mathbf{c}) = \frac{1}{Z_\tau(\mathbf{c})}\exp\Big[-\widetilde{E}(\mathbf{X}_{\mathcal{A}};\mathbf{c})/\tau\Big], \tag{2}$$

where the temperature τ controls the trade-off between concentrating mass at the global minimum and covering competing metastable basins, and $Z_\tau(\mathbf{c})$ is the partition function. The sampler must additionally respect slab in-plane periodicity, the physical role of the surface normal, and atom-permutation symmetries within species.

Validity of relaxed configurations. Not every $\mathbf{X}_{\mathcal{A}}$ yields a physically meaningful adsorption configuration after the downstream MLIP relaxation: relaxation can drive an initial placement into a configuration that no longer represents a chemisorbed adsorbate on the intended slab. Following the AdsorbDiff (Kolluru & Kitchin, 2024) anomaly taxonomy, we call a relaxed configuration *valid* only if it preserves adsorbate identity, slab identity, and adsorbate–surface placement, and otherwise flag one of four post-relaxation anomalies: (1) *desorption* – the adsorbate detaches from the slab; (2) *dissociation* – changes in the internal covalent connectivity of the adsorbate; (3) *surface reconstruction* – the slab's connectivity changes beyond a tolerance relative to a reference relaxed slab; (4) *intercalation* – an adsorbate atom comes into contact with a constrained bulk atom. Full description of validity criteria is provided in A.4.

2.2. ML interatomic potentials

An MLIP is a differentiable approximation $E_\theta(\mathbf{X},\mathbf{Z},\mathbf{L})$ to the total DFT energy E, with conservative forces $\mathbf{F}_i = -\nabla_{\mathbf{X}_i} E_\theta$ on the atom i. OC20 enabled large-scale training of such models for catalyst systems (Chanussot et al., 2021), and universal MLIPs extend coverage across the periodic table (Chen & Ong, 2022; Deng et al., 2023; Batatia et al., 2025; Wood et al., 2025). We write

$$E_\theta(\mathbf{X}_{\mathcal{A}};\mathbf{c}) := E_\theta([\mathbf{X}_{\mathcal{A}};\mathbf{X}_{\mathcal{S}};\mathbf{X}_{\mathcal{C}}],\ [\mathbf{Z}_{\mathcal{A}};\mathbf{Z}_{\mathcal{S}};\mathbf{Z}_{\mathcal{C}}],\ \mathbf{L}) \tag{3}$$

for the MLIP energy as a function of the adsorbate coordinates only, with the slab geometry held at its conditioning values.

The salient property for our purposes is that E_θ and $\nabla_{\mathbf{X}_{\mathcal{A}}} E_\theta$ are cheap and broadly available, even where dense placement labels for a given $\mathbf{c}$ are not. Because MLIP accuracy can degrade outside its training distribution, we use E_θ for candidate generation and reserve DFT for final adsorption-energy ranking.

2.3. Diffusion samplers

Diffusion samplers learn an amortized sampler for an unnormalized target density without requiring samples from the target distribution. Instead, they assume access to a differentiable energy and learn stochastic dynamics whose terminal marginal approximates the desired Boltzmann law (Zhang & Chen, 2021; Vargas et al., 2023; Akhound-Sadegh et al., 2024; Havens et al., 2025; Liu et al., 2025).

For a sample variable $\mathbf{X}$ and conditioning $\mathbf{c}$, given a differentiable energy $\widetilde{E}(\mathbf{X};\mathbf{c})$ they target

$$\pi_\tau(\mathbf{X} \mid \mathbf{c}) \propto \exp\big[r_\tau(\mathbf{X};\mathbf{c})\big], \quad r_\tau(\mathbf{X};\mathbf{c}) := -\widetilde{E}(\mathbf{X};\mathbf{c})/\tau, \tag{4}$$

with temperature τ controlling concentration on low-energy regions. The sampler specifies a tractable source $p_0(\mathbf{X} \mid \mathbf{c})$ and an SDE

$$d\mathbf{X}_t = u_t^\phi(\mathbf{X}_t,\mathbf{c})\,dt + g_t\,d\mathbf{B}_t, \qquad \mathbf{X}_0 \sim p_0(\cdot \mid \mathbf{c}), \tag{5}$$

where u_t^ϕ is the learned drift, $g_t \geq 0$ is a prescribed scalar diffusion schedule, and $\mathbf{B}_t$ is a standard Brownian motion of the same dimensionality as $\mathbf{X}$. The terminal marginal p_1^ϕ of this SDE should approximate π_τ.

2.4. Flow sampling

We adopt Flow Sampling, a recent diffusion-sampler objective for learning from unnormalized densities via conditional denoising targets (Havens et al., 2026). At a high level, Flow Sampling replaces target samples with detached terminal samples from the current sampler and trains the drift by regression onto a closed-form denoising target built from the source sample, the terminal sample, and a single score evaluation at the terminal point.

For the linear interpolation process

$$\mathbf{X}_t = (1-t)\mathbf{X}_0 + t\bar{\mathbf{X}}_1, \tag{6}$$

where $\mathbf{X}_0 \sim p_0(\cdot \mid \mathbf{c})$ and $\bar{\mathbf{X}}_1 \sim p_1^{\bar{\phi}}(\cdot \mid \mathbf{c})$ is sampled from a detached copy of the current model, the Flow Sampling objective regresses $u_t^\phi(\mathbf{X}_t,\mathbf{c})$ toward

$$\bar{\mathbf{X}}_1 - \mathbf{X}_0 + \gamma\nabla_{\mathbf{X}} r_\tau(\bar{\mathbf{X}}_1;\mathbf{c}), \tag{7}$$

where γ is the diffusion coefficient parameter. This yields an efficient replay-buffer training scheme, since each terminal sample can be reused together with a single energy-gradient evaluation (Havens et al., 2026).

3. Methods

We first specify the source, control architecture, and restraint terms, and then instantiate the Flow Sampling objective for frozen-slab adsorbate generation.

3.1. Source distribution

We use a harmonic source on centered local adsorbate coordinates (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

$$p_{\mathrm{harm}}(\widetilde{\mathbf{X}}_{\mathcal{A},0}) \propto \exp\left(-\frac{\alpha}{2}\sum_{i,j\in\mathcal{A}}\|\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_j\|^2\right). \quad (8)$$

Equivalently, this is a Gaussian on centered coordinates with covariance proportional to the pseudoinverse of the complete-graph Laplacian. We consider two slab-conditioned placement variants built from this same harmonic local distribution: a *centered harmonic* source, which places the adsorbate near a reference slab anchor, 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.

3.2. Control model parametrization

We parameterize the drift field $u_t^\phi(\mathbf{X}_t, \mathbf{c})$ with a typed-branch equivariant graph neural network designed for frozen-slab adsorbate systems. Here $\mathbf{X}_t \equiv \mathbf{X}_{\mathcal{A},t}$ denotes the current adsorbate configuration along the diffusion trajectory. Thus, the network consumes the current adsorbate coordinates, the fixed slab geometry, atomic numbers, the periodic cell $\mathbf{L}$, and the adsorbate bond graph $\mathcal{G}_\mathcal{A}$. It predicts 3D vectors only for atoms in $\mathcal{A}$; slab atoms are treated as static context and their outputs are masked to zero.

At each denoising step, we construct four interaction graphs under periodic boundary conditions using the minimum-image convention: (i) a complete adsorbate graph over $\mathcal{A}$, (ii) the adsorbate bond graph $\mathcal{G}_\mathcal{A}$, (iii) an adsorbate–surface graph connecting adsorbate atoms to nearby slab atoms, and (iv) a slab graph over $\mathcal{S} \cup \mathcal{C}$. In addition, shortest-path distances on $\mathcal{G}_\mathcal{A}$ are used as topological features, allowing the controller to distinguish adsorbates with identical composition but different covalent connectivity.

Slab atoms are first contextualized by a PaiNN-style (Schütt et al., 2021) slab encoder operating on the slab graph while keeping slab coordinates fixed. The main controller then applies a stack of typed interaction layers with three message-passing branches corresponding to complete adsorbate interactions, covalent-bond interactions, and adsorbate–surface interactions. The branch outputs are fused by learned gates conditioned on the hidden state and diffusion time, separating nonbonded adsorbate geometry, fixed molecular topology, and local surface context within a single controller.

A final equivariant readout head produces the drift on movable adsorbate atoms,

$$u_t^\phi(\mathbf{X}_t, \mathbf{c}) \in \mathbb{R}^{|\mathcal{A}|\times 3}, \quad (9)$$

which is used in the sampler of Section 3.5. Full details of node initialization, graph definitions, branch-specific features, message-passing equations, and gated fusion are deferred to Appendix A.2.

3.3. Restraint

We use a differentiable restraint term $R(\mathbf{X}_\mathcal{A}; \mathbf{c})$ to translate the validity criteria of Section 2.1 into soft geometric penalties during training. The restraint biases the sampler away from placements that are unlikely to relax to valid adsorption configurations.

The restraint has three active components: a bond-structure term, a surface-height term, and a surface wall term. The bond-structure term uses the adsorbate graph $\mathcal{G}_\mathcal{A}$ to discourage dissociation and severe intramolecular clashes, thereby acting as a soft proxy for preserving adsorbate identity during sampling. The surface-height term penalizes adsorbate atoms that drift too far from the slab along the surface normal, and therefore acts as a soft proxy against desorption. The surface wall term penalizes sub-surface penetration and can be viewed as a soft proxy for intercalation.

Because the slab coordinates are held fixed during sampling, we do not include a surface-reconstruction restraint in the frozen-slab objective. Final validity is therefore assessed only after downstream MLIP relaxation, using the anomaly criteria of Section 2.1. Exact definitions of the restraint components are given in Appendix A.3.

3.4. MLIP instantiation

Throughout this work, we instantiate the surrogate energy E_θ with a pretrained Universal Model for Atoms (UMA) potential (Wood et al., 2025). UMA is a universal machine-learning interatomic potential trained across molecular, materials, and catalyst datasets, and maps atomic positions and atomic numbers to a total energy, with forces obtained from energy gradients. Importantly for our setting, UMA is trained with task-specific conditioning that selects the

DFT setting to emulate; here we use the OC20 catalyst-task instantiation throughout. In the frozen-slab setting, we therefore write

$$E_{\mathrm{UMA}}(\mathbf{X}_{\mathcal{A}};\mathbf{c}) = E^{\mathrm{OC20}}_{\mathrm{UMA}}([\mathbf{X}_{\mathcal{A}};\mathbf{X}_{\mathcal{S}};\mathbf{X}_{\mathcal{C}}],[\mathbf{Z}_{\mathcal{A}};\mathbf{Z}_{\mathcal{S}};\mathbf{Z}_{\mathcal{C}}],\mathbf{L})\,. \tag{10}$$

We use this same OC20-instantiated UMA model both to define the training energy and for downstream MLIP relaxation before DFT evaluation.

3.5. Diffusion sampling for adsorbates

We instantiate the Flow Sampling framework of Section 2.4 with sample variable $\mathbf{X}_{\mathcal{A}}$ and conditioning $\mathbf{c}$ from Section 2.1; the drift network $u_t^{\phi}(\mathbf{X}_t,\mathbf{c})$ predicts vectors only on adsorbate atoms and consumes the bond graph $\mathcal{G}_{\mathcal{A}} \subset \mathbf{c}$ as an additional structural conditioning signal (Section 3.2). The source $p_0(\cdot \mid \mathbf{c})$ is as defined in 3.1.

The training energy combines the MLIP surrogate with differentiable restraints,

$$\widetilde{E}(\mathbf{X}_{\mathcal{A}};\mathbf{c}) = E_{\mathrm{UMA}}(\mathbf{X}_{\mathcal{A}};\mathbf{c}) + \lambda\, R(\mathbf{X}_{\mathcal{A}};\mathbf{c}), \tag{11}$$

where R penalizes unphysical placements and conformations.

Given a source $\mathbf{X}_0 \sim p_0(\cdot \mid \mathbf{c})$, a terminal sample $\bar{\mathbf{X}}_1 \sim p_1^{\bar{\phi}}(\cdot \mid \mathbf{c})$ from the detached current sampler, and the linear interpolant

$$\mathbf{X}_t = (1-t)\mathbf{X}_0 + t\bar{\mathbf{X}}_1, \tag{12}$$

we define the Flow Sampling target

$$\mathbf{u}_t^{\star} = \bar{\mathbf{X}}_1 - \mathbf{X}_0 + \gamma\, \nabla_{\mathbf{X}_{\mathcal{A}}} r_{\tau}(\bar{\mathbf{X}}_1;\mathbf{c}). \tag{13}$$

The corresponding training loss is

$$\mathcal{L}_{\mathrm{FS}}(\phi) = \mathbb{E}_{\substack{t\sim\mathcal{U}(0,1),\\ \mathbf{X}_0\sim p_0(\cdot\mid\mathbf{c}),\\ \bar{\mathbf{X}}_1\sim p_1^{\bar{\phi}}(\cdot\mid\mathbf{c})}} \left[\left\| u_t^{\phi}(\mathbf{X}_t,\mathbf{c}) - \mathbf{u}_t^{\star} \right\|^2\right]. \tag{14}$$

Each terminal sample and its energy gradient are reused across many source samples and times, so MLIP gradient evaluations occur only at generated endpoints. This yields the base Flow Sampling objective for adsorbates; we further regularize optimization with a damping term, described next.

3.6. Damping

To stabilize training, we add a damping regularizer that penalizes deviations from a detached lagged copy of the controller. Concretely, at the start of each training epoch we freeze a snapshot of the current controller, denoted $u_t^{\bar{\phi}}$, and keep it fixed while optimizing the updated controller u_t^{ϕ}. Following the damping mechanism used in Blessing et al. (2026), we define

$$\mathcal{L}_{\mathrm{damp}}(\phi) = \mathbb{E}\left[\left\| u_t^{\phi}(\mathbf{X}_t,\mathbf{c}) - u_t^{\bar{\phi}}(\mathbf{X}_t,\mathbf{c}) \right\|^2\right]. \tag{15}$$

The full training objective is then

$$\mathcal{L}(\phi) = \mathcal{L}_{\mathrm{FS}}(\phi) + \lambda_{\mathrm{damp}}\, \mathcal{L}_{\mathrm{damp}}(\phi). \tag{16}$$

Beyond stabilization, this damping term also extends mode coverage: by anchoring updates to the lagged controller $u_t^{\bar{\phi}}$, it slows the contraction onto already-discovered low-energy basins (Blessing et al., 2026).

4. Experiments

4.1. Datasets and splits

We use two datasets in this work: OC20 (Chanussot et al., 2021) and OC20-Dense (Lan et al., 2023).

OC20. Open Catalyst 2020 (OC20) (Chanussot et al., 2021) is a large-scale dataset of DFT optimization trajectories for adsorbate–slab systems. Each OC20 example corresponds to a local optimization from a single heuristic initial placement, so OC20 is well suited for learning general adsorbate–surface representations but does not provide dense supervision over many candidate placements for the same adsorbate–slab combination, and therefore does not help find the global minima.

OC20-Dense. OC20-Dense (Lan et al., 2023) is a dense-placement benchmark for adsorption-energy search. For each adsorbate–slab system, it generates 100 initial placements, relaxes them with DFT, and uses the resulting set of local minima to better approximate the adsorption landscape of that system. The benchmark provides both in-distribution (ID) and out-of-distribution (OOD) subsets relative to the OC20 training distribution: ID uses adsorbates and slabs seen in OC20 training but in new combinations/configurations, while OOD includes new adsorbates and/or slabs not present in OC20 training.

Training and validation protocol. Following AdsorbDiff (Kolluru & Kitchin, 2024), we use the OC20-Dense ID training subset for training and the OC20-Dense ID and OOD validation subsets for model evaluation. Crucially, AdsorbSample does not use the dense OC20-Dense adsorption placements, relaxed structures, or DFT adsorption-energy labels as supervision. For each system, we only use the conditioning information $\mathbf{c}$ and train the sampler from MLIP energy and gradient queries.

4.2. Baselines

We compare against two representative approaches that separate the main distinctions of our method: AdsorbML tests against a non-amortized ML+DFT search pipeline initialized from heuristic and random placements, while AdsorbDiff tests against a supervised generative model trained on dense placement labels. Together, these baselines evaluate whether AdsorbSample can learn an amortized proposal distribution without dense supervision and whether sampling full adsorbate coordinates improves over rigid-body placement generation.

AdsorbML. AdsorbML (Lan et al., 2023) is a hybrid ML+DFT search pipeline. It begins from a set of heuristic and random initial adsorbate placements, relaxes all candidates with an MLIP, ranks the resulting structures by ML energy, and then refines the best k candidates with DFT single-point calculations. The final adsorption-energy prediction is the minimum over these DFT follow-up calculations.

AdsorbDiff. AdsorbDiff (Kolluru & Kitchin, 2024) is a supervised conditional denoising diffusion model for adsorbate placement. It learns to predict an adsorption site and rigid adsorbate orientation on the slab, and is trained on OC20-Dense using relative energies across densely sampled placements for the same adsorbate–slab system. After diffusion sampling, the predicted structure is relaxed with a pretrained ML force field and then verified with DFT. Unlike our approach, AdsorbDiff uses dense placement supervision directly during task-specific training and restricts the generated motion to rigid-body translation and rotation of the adsorbate only.

We note that the recent work AdsorbFlow (Qiu et al., 2026) addresses the same problem setting; however, since its implementation is not publicly available, we do not include it in our benchmark.

4.3. Models

AdsorbSample. Our main model initializes the controller by pretraining on relaxed OC20 IS2RE structures with conditional flow matching, before task-specific flow sampling training on OC20-Dense. Given a source sample $\mathbf{X}_0 \sim p_0(\cdot \mid \mathbf{c})$, an optimized OC20 target structure $\mathbf{X}_1$, and $t \sim \mathcal{U}(0,1)$, we form the linear interpolant

$$\mathbf{X}_t = (1-t)\mathbf{X}_0 + t\mathbf{X}_1, \tag{17}$$

and train the controller with

$$\mathcal{L}_{\text{CFM}}(\phi) = \mathbb{E}\left[\left\| u_t^{\phi}(\mathbf{X}_t, \mathbf{c}) - (\mathbf{X}_1 - \mathbf{X}_0)\right\|^2\right]. \tag{18}$$

After pretraining, we train the same controller on OC20-Dense with Flow Sampling objective described in Section 3.5.

AdsorbSample w/o pretraining. This ablation uses the same controller architecture and downstream Flow Sampling objective, but without the OC20 IS2RE pretraining.

Full model and training details are provided in Appendix B.1.

4.4. Evaluation metrics

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 $\widehat{\mathcal{S}}_s^{\text{val}} \subseteq \widehat{\mathcal{S}}_s$ the subset that passes the four post-relaxation validity criteria of Section 2.1 (full definitions in Appendix A.4). For each $\widehat{X} \in \widehat{\mathcal{S}}_s$ we compute the MLIP energy $E_\theta(\widehat{X}; \mathbf{c})$ and, when scoring final selections, a DFT single-point adsorption energy $E_{\text{ads}}(\widehat{X}; \mathbf{c})$.

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

$$E^\star_{\theta,s} := \min_{Y \in \mathcal{R}_s^{\text{val}}} E_\theta(Y; \mathbf{c}), \quad E^\star_{\text{ads},s} := \min_{Y \in \mathcal{R}_s^{\text{val}}} E_{\text{ads}}(Y; \mathbf{c}), \tag{19}$$

and use a single energy tolerance $\epsilon_{\text{succ}} = 0.1$ 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 A.5.

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

Diversity. We greedily cluster $\widehat{\mathcal{S}}_s^{\text{val}}$ under A-RMSD with threshold $d_{\text{cluster}} = 0.05$ Å, processing candidates in order of increasing E_θ 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 $\text{Div}(s) = |\mathcal{C}_s| / |\widehat{\mathcal{S}}_s|$.

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,

$$c_s = \left|\left\{\widehat{X} \in \widehat{\mathcal{S}}_s^{\mathrm{val}} :\ E_\theta(\widehat{X};\mathbf{c}) - E^\star_{\theta,s} \le \epsilon_{\mathrm{succ}}\right\}\right|, \tag{20}$$

$$\text{MLIP Pass@}k(s) = 1 - \binom{n_s - c_s}{k} \Big/ \binom{n_s}{k}, \tag{21}$$

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 $\widehat{\mathcal{S}}^{\mathrm{val}}_{s,1:k} = \widehat{\mathcal{S}}_{s,1:k} \cap \widehat{\mathcal{S}}_s^{\mathrm{val}}$ its valid subset. The per-system selection is

$$\widehat{X}^\star_{s,k} := \arg\min_{\widehat{X} \in \widehat{\mathcal{S}}^{\mathrm{val}}_{s,1:k}} E_\theta(\widehat{X};\mathbf{c}), \tag{22}$$

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

$$\mathrm{SR@}k(s) = \mathbf{1}\left[E_{\mathrm{ads}}(\widehat{X}^\star_{s,k};\mathbf{c}) - E^\star_{\mathrm{ads},s} \le \epsilon_{\mathrm{succ}}\right]. \tag{23}$$

Coverage (COV). Following conformer-generation evaluation (Xu et al., 2021), we replace molecular RMSD with A-RMSD and report recall- and precision-style coverage against a low-energy MLIP-relaxed reference set $\mathcal{R}_s^{\mathrm{cov}} \subseteq \mathcal{R}_s^{\mathrm{val}}$, defined as the A-RMSD cluster representatives of $\mathcal{R}_s^{\mathrm{val}}$ that lie within $0.2\,\mathrm{eV}$ of $E^\star_{\theta,s}$. As a function of an A-RMSD cutoff d_{tol}, the per-system coverage is

$$\text{COV-R}(s; d_{\mathrm{tol}}) = \frac{1}{|\mathcal{R}_s^{\mathrm{cov}}|} \sum_{Y \in \mathcal{R}_s^{\mathrm{cov}}} \mathbf{1}\left[\min_{\widehat{X} \in \widehat{\mathcal{S}}_s^{\mathrm{val}}} d(\widehat{X}, Y) < d_{\mathrm{tol}}\right], \tag{24}$$

and COV-P$(s; d_{\mathrm{tol}})$ is defined symmetrically by swapping $\widehat{\mathcal{S}}_s^{\mathrm{val}}$ and $\mathcal{R}_s^{\mathrm{cov}}$.

For later analysis, all metrics above are reported as system averages over $s \in \mathcal{D}$, restricted to systems with $\mathcal{R}_s^{\mathrm{val}} \neq \emptyset$ for those that depend on $E^\star_{\theta,s}$ or $E^\star_{\mathrm{ads},s}$.

4.5. Results

We report structural metrics (Validity and Diversity), MLIP Pass@k, and final DFT SR@k in Table 1 for the OC20-Dense ID and OOD subsets, following Kolluru & Kitchin (2024).

Across both ID and OOD splits, AdsorbSample attains the strongest MLIP Pass@k at every budget and the best or tied-best DFT SR@k at all but a few low-k entries, while remaining competitive on raw structural Validity and Diversity. The advantage is therefore not in producing more valid structures, but in placing more of its samples in low-energy basins that survive MLIP- and DFT-based selection.

Compared with AdsorbDiff, the only data-driven baseline trained directly on OC20-Dense, AdsorbSample produces substantially more diverse candidates (45.9%/45.8% vs. 7.4%/10.9% on ID/OOD) and far higher high-budget DFT SR (58.1%/54.0% vs. 29.5%/32.0% at k=10); AdsorbDiff's SR saturates quickly because its learned proposal distribution concentrates on a narrow set of placements and degrades further on OOD (SR@1 drops from 18.2% to 14.0%). AdsorbML reports slightly higher Validity and Diversity, but it shares its placement procedure with the algorithm used to construct the OC20-Dense reference set, so its candidates are biased toward the reference distribution and its raw structural metrics, including coverage, are inflated relative to methods that sample independently. Once samples are scored by MLIP energy this advantage disappears: AdsorbML trails AdsorbSample by 14.3 (ID) and 8.9 (OOD) points at MLIP Pass@1, and by 11.6 (ID) and 8.0 (OOD) points at DFT SR@10.

Effect of pretraining. Removing the pretraining stage causes the largest losses on the ID split, where the pretrained representations align most closely with the target distribution. ID DFT SR drops from 20.9% to 9.1% at k=1 and from 58.1% to 47.7% at k=10, and ID MLIP Pass@k falls by 12.9 points at k=1 and 10.6 points at k=10. The gap widens with k, indicating that pretraining helps populate a broader set of useful basins rather than only sharpening the top-1 prediction. On OOD, where pretrained features are less aligned with the target systems, the gap is smaller (DFT SR@10: 54.0% → 48.0%; the no-pretraining variant even matches AdsorbSample at k=1 with 18.0%). Pretraining therefore primarily boosts sampling success at larger k and is most effective in-distribution.

Effect of source distribution. Replacing the broad surface source with a centered source, which restricts proposals to a narrower region of the surface, trades diversity for concentration. Diversity drops sharply (45.9% → 19.1% on ID; 50.0% → 24.3% on OOD), but the focused source raises low-budget DFT SR on OOD to 24.0% at k=1 and 28.0% at k=2, the highest values among all methods, by spending its samples on a single high-probability region. The advantage disappears at larger k: the centered-source variant plateaus at 34.1% (ID) and 42.0% (OOD) DFT SR@10, well below the broader-source variants (47.7% and 48.0%), because the narrow proposal distribution exhausts its useful basins quickly and cannot reach alternative adsorption sites. A tighter source is thus useful when only a few candidates can be afforded, but a broader source is needed to cover the multi-basin structure of adsorption sites and to scale with the sample budget.

Table 1. **Result Metrics.** 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*)	69.7	45.9	**28.7**	**38.5**	**49.9**	**56.7**	**20.9**	**37.2**	**41.9**	**58.1**
↪ 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)	58.0	7.4	17.5	21.7	25.8	28.6	18.2	18.2	27.3	29.5
AdsorbML (Lan et al., 2023)	**73.1**	**46.6**	14.4	24.2	41.3	55.0	**20.9**	23.3	32.6	46.5
OC20-Dense OOD subset (50 systems)										
AdsorbSample (*this work*)	73.9	45.8	**25.5**	**35.4**	**49.9**	**61.3**	18.0	26.0	**44.0**	**54.0**
↪ 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)	65.0	10.9	16.5	21.4	27.2	30.8	14.0	24.0	30.0	32.0
AdsorbML (Lan et al., 2023)	**83.1**	**54.0**	16.6	26.3	42.3	55.9	14.0	18.0	32.0	46.0

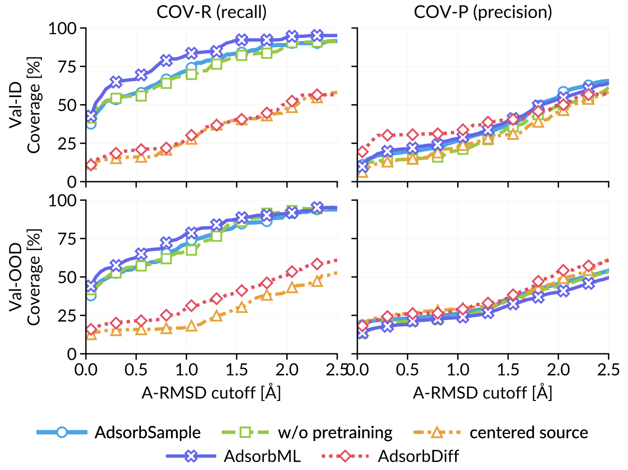


Figure 2. **Structural Coverage.** COV-R and COV-P metrics for each method on MLIP-relaxed structures. Reference structures are filtered to those within 0.2 eV of the system minimum energy.

Coverage analysis. We further analyze sample diversity and coverage against low-energy MLIP-relaxed references in Figure 2. AdsorbDiff's low diversity appears as limited COV-R: it spans fewer distinct low-energy basins despite some high-quality candidates. On the ID subset at small adsorbate RMSD cutoffs, it can achieve slightly higher COV-P than AdsorbSample, indicating concentration near a narrower set of low-energy structures. However, this comes with reduced coverage and depends on DFT-relaxed OC20-Dense labels, which are costly and hard to scale to new adsorbates or catalysts. In contrast, AdsorbSample offers a more scalable way to generate valid, diverse low-energy candidates for unseen systems, while maintaining strong MLIP Pass@k and DFT SR@k after MLIP-based selection. To complement these aggregate coverage metrics, Appendix C.1 provides a system-level mode-collapse diagnostic, tracking how probability mass shifts across adsorbate–surface contact motifs during training.

5. Conclusion

We presented AdsorbSample, an energy-trained conditional diffusion sampler that recasts adsorbate placement as conditional Boltzmann sampling under a pretrained MLIP energy, removing the reliance on dense per-system placement labels and exposing the full adsorbate coordinate space including internal conformer degrees of freedom. The sampler combines a surface-aware source, a conditional control model, and differentiable restraint potentials that suppress desorption, atomic overlap, and intercalation directly in the proposal rather than through post hoc filtering. On the OC20-Dense ID and OOD subsets, AdsorbSample attains the strongest MLIP Pass@k at every sample budget and the best or tied-best DFT SR@k at most operating points, while remaining competitive with the strongest baseline on raw structural validity and diversity. Ablations indicate that pretraining most improves high-k success on the ID split by populating a broader set of useful basins, and that the choice of source trades concentrated low-budget success against the diversity required to scale with the sample budget.

Future works. Energy-trained samplers are susceptible to mode collapse onto a small subset of basins, and a stratified replay buffer that enforces coverage over surface regions or symmetry classes during training could keep the proposal distribution broad without sacrificing low-energy quality. A systematic study of model architecture, in particular equivariance, message-passing depth, and conditioning mechanisms, would also clarify which inductive biases best transfer an MLIP energy landscape into a sampler. Finally, the approach can be extended to larger adsorbates with many torsional degrees of freedom (Pablo-García et al., 2023).

References

Akhound-Sadegh, T., Rector-Brooks, J., Bose, A. J., Mittal, S., Lemos, P., Liu, C.-H., Sendera, M., Ravanbakhsh, S., Gidel, G., Bengio, Y., et al. Iterated denoising energy matching for sampling from boltzmann densities. *arXiv preprint arXiv:2402.06121*, 2024. (cited on pages 1 and 3)

Batatia, I., Benner, P., Chiang, Y., Elena, A. M., Kovács, D. P., Riebesell, J., Advincula, X. R., Asta, M., Avaylon, M., Baldwin, W. J., et al. A foundation model for atomistic materials chemistry. *The Journal of chemical physics*, 163(18), 2025. (cited on pages 1 and 3)

Blessing, D., Richter, L., Berner, J., Malitskiy, E., and Neumann, G. Bridge matching sampler: Scalable sampling via generalized fixed-point diffusion matching. *arXiv preprint arXiv:2603.00530*, 2026. (cited on page 5)

Chanussot, L., Das, A., Goyal, S., Lavril, T., Shuaibi, M., Riviere, M., Tran, K., Heras-Domingo, J., Ho, C., Hu, W., et al. Open catalyst 2020 (oc20) dataset and community challenges. *Acs Catalysis*, 11(10):6059–6072, 2021. (cited on pages 1, 2, 3, 5, and 15)

Chen, C. and Ong, S. P. A universal graph deep learning interatomic potential for the periodic table. *Nature Computational Science*, 2(11):718–728, 2022. (cited on pages 1 and 3)

Chen, M., Tworek, J., Jun, H., Yuan, Q., Pinto, H. P. D. O., Kaplan, J., Edwards, H., Burda, Y., Joseph, N., Brockman, G., et al. Evaluating large language models trained on code. *arXiv preprint arXiv:2107.03374*, 2021. (cited on page 6)

Deng, B., Zhong, P., Jun, K., Riebesell, J., Han, K., Bartel, C. J., and Ceder, G. Chgnet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nature Machine Intelligence*, 5(9):1031–1041, 2023. (cited on pages 1 and 3)

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

Hafner, J. Ab-initio simulations of materials using vasp: Density-functional theory and beyond. *Journal of computational chemistry*, 29(13):2044–2078, 2008. (cited on page 15)

Hammer, B., Hansen, L. B., and Nørskov, J. K. Improved adsorption energetics within density-functional theory using revised perdew-burke-ernzerhof functionals. *Physical review B*, 59(11):7413, 1999. (cited on page 15)

Havens, A., Miller, B. K., Yan, B., Domingo-Enrich, C., Sriram, A., Wood, B., Levine, D., Hu, B., Amos, B., Karrer, B., et al. Adjoint sampling: Highly scalable diffusion samplers via adjoint matching. *arXiv preprint arXiv:2504.11713*, 2025. (cited on pages 1, 3, and 12)

Havens, A., Karrer, B., and Shaul, N. Flow sampling: Learning to sample from unnormalized densities via denoising conditional processes, 2026. URL `https://arxiv.org/abs/2605.03984`. (cited on pages 3 and 4)

Hjorth Larsen, A., Jørgen Mortensen, J., Blomqvist, J., Castelli, I. E., Christensen, R., Dułak, M., Friis, J., Groves, M. N., Hammer, B., Hargus, C., 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 13 and 14)

Jing, B., Erives, E., Pao-Huang, P., Corso, G., Berger, B., and Jaakkola, T. Eigenfold: Generative protein structure prediction with diffusion models. *arXiv preprint arXiv:2304.02198*, 2023. (cited on page 4)

Kim, M., Kim, N., Kim, H., and Ahn, S. Catflow: Co-generation of slab-adsorbate systems via flow matching. *arXiv preprint arXiv:2602.05372*, 2026. (cited on page 1)

Kolluru, A. and Kitchin, J. R. Adsorbdiff: Adsorbate placement via conditional denoising diffusion. *arXiv preprint arXiv:2405.03962*, 2024. (cited on pages 1, 3, 5, 6, 7, 8, 13, and 14)

Kresse, G. and Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Physical review B*, 59(3):1758, 1999. (cited on page 15)

Lan, J., Palizhati, A., Shuaibi, M., Wood, B. M., Wander, B., Das, A., Uyttendaele, M., Zitnick, C. L., and Ulissi, Z. W. Adsorbml: a leap in efficiency for adsorption energy calculations using generalizable machine learning potentials. *npj Computational Materials*, 9(1):172, 2023. (cited on pages 1, 5, 6, 7, and 8)

Liu, D. C. and Nocedal, J. On the limited memory bfgs method for large scale optimization. *Mathematical programming*, 45(1):503–528, 1989. (cited on page 14)

Liu, G.-H., Choi, J., Chen, Y., Miller, B. K., and Chen, R. T. Adjoint schrödinger bridge sampler. *arXiv preprint arXiv:2506.22565*, 2025. (cited on pages 1 and 3)

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

Methfessel, M. and Paxton, A. T. High-precision sampling for brillouin-zone integration in metals. *physical review B*, 40(6):3616, 1989. (cited on page 15)

Nam, J., Máté, B., Toshev, A. P., Kaniselvan, M., Gómez-Bombarelli, R., Chen, R. T., Wood, B., Liu, G.-H., and Miller, B. K. Enhancing diffusion-based sampling with molecular collective variables. *arXiv preprint arXiv:2510.11923*, 2025. (cited on page 12)

Nørskov, J. K., Bligaard, T., Rossmeisl, J., and Christensen, C. H. Towards the computational design of solid catalysts. *Nature chemistry*, 1(1):37–46, 2009. (cited on page 1)

Pablo-García, S., Morandi, S., Vargas-Hernández, R. A., Jorner, K., Ivković, Ž., López, N., and Aspuru-Guzik, A. 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 8)

Qiu, J., Li, W., Chen, H., Zhao, L., and Wang, X. Adsorbflow: energy-conditioned flow matching enables fast and realistic adsorbate placement. *arXiv preprint arXiv:2602.19289*, 2026. (cited on pages 1 and 6)

Schütt, K., Unke, O., and Gastegger, M. Equivariant message passing for the prediction of tensorial properties and molecular spectra. In *International conference on machine learning*, pp. 9377–9388. PMLR, 2021. (cited on pages 4 and 12)

Vargas, F., Grathwohl, W., and Doucet, A. Denoising diffusion samplers. *arXiv preprint arXiv:2302.13834*, 2023. (cited on pages 1 and 3)

Wood, B. M., Dzamba, M., Fu, X., Gao, M., Shuaibi, M., Barroso-Luque, L., Abdelmaqsoud, K., Gharakhanyan, V., Kitchin, J. R., Levine, D. S., et al. Uma: A family of universal models for atoms. *arXiv preprint arXiv:2506.23971*, 2025. (cited on pages 1, 3, 4, and 14)

Xu, M., Luo, S., Bengio, Y., Peng, J., and Tang, J. Learning neural generative dynamics for molecular conformation generation. *arXiv preprint arXiv:2102.10240*, 2021. (cited on page 7)

Zhang, Q. and Chen, Y. Path integral sampler: a stochastic control approach for sampling. *arXiv preprint arXiv:2111.15141*, 2021. (cited on pages 1 and 3)

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 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

$$L_{\mathrm{cg}} = \left(N_\mathcal{A} I_{N_\mathcal{A}} - \mathbf{1}\mathbf{1}^\top\right) \otimes I_3. \tag{25}$$

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

$$T = P \,\mathrm{diag}(\lambda^{-1/2})\, P^\top, \tag{26}$$

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

$$\mathrm{vec}(\widetilde{\mathbf{X}}_{\mathcal{A},0}) = \sigma_{\mathrm{ads}}\, T\boldsymbol{\xi}, \tag{27}$$

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

$$\mathbf{n} = \frac{\mathbf{a} \times \mathbf{b}}{\|\mathbf{a} \times \mathbf{b}\|}, \qquad h_{\mathrm{top}} = \max_{j \in \mathcal{S} \cup \mathcal{C}} \mathbf{n}^\top \mathbf{x}_j. \tag{28}$$

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

For the *centered harmonic* source,

$$\mathbf{r}_{\mathrm{xy}} = \mathbf{r}_{\mathrm{ref}} + \sigma_{\mathrm{xy}} \boldsymbol{\varepsilon}_{\mathrm{xy}}, \qquad \boldsymbol{\varepsilon}_{\mathrm{xy}} \sim \mathcal{N}(\mathbf{0}, I_2). \tag{29}$$

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

$$u, v \sim \mathrm{Unif}(0,1), \qquad \mathbf{r}_{\mathrm{xy}} = u\, \mathbf{a}_{1:2} + v\, \mathbf{b}_{1:2}, \tag{30}$$

optionally with small Gaussian jitter. The height is sampled as

$$h_{\mathrm{tar}} = \max(h_{\mathrm{top}} + \mu_z + \sigma_z \varepsilon_z,\ h_{\mathrm{top}} + d_{\mathrm{min}}), \qquad \varepsilon_z \sim \mathcal{N}(0,1). \tag{31}$$

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

$$\min_{i \in \mathcal{A}} \mathbf{n}^\top \mathbf{x}_i < h_{\mathrm{top}} + d_{\mathrm{min}}, \tag{32}$$

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

$$\mathbf{h}_i^{(0)} = \mathrm{Emb}(Z_i), \qquad \mathbf{v}_i^{(0)} = \mathbf{0}. \tag{33}$$

Time is encoded by

$$\mathbf{e}_t = \mathrm{MLP}_t(\gamma(t)). \tag{34}$$

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

$$\mathcal{E}_{\text{all}} = \{(i,j) : i,j \in \mathcal{A},\ i \neq j\}, \tag{35}$$

$$\mathcal{E}_{\text{bond}} = \mathcal{E}_{\mathcal{A}}, \tag{36}$$

$$\mathcal{E}_{\text{surf}} = \{(i,j) : i \in \mathcal{A},\ j \in \mathcal{S} \cup \mathcal{C},\ r_{ij} < r_{\text{surf}}\}, \tag{37}$$

$$\mathcal{E}_{\text{slab}} = \{(i,j) : i,j \in \mathcal{S} \cup \mathcal{C},\ r_{ij} < r_{\text{slab}}\}. \tag{38}$$

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

$$\mathbf{e}_{ij}^{(\text{all})} = \left[\mathrm{RBF}(r_{ij}), \mathrm{Emb}_{\text{sp}}(d_{ij}^{\text{sp}}), \mathrm{Emb}_{\text{bond}}(b_{ij})\right], \tag{39}$$

$$\mathbf{e}_{ij}^{(\text{bond})} = \left[\mathrm{RBF}(r_{ij}), \mathrm{Emb}_{\text{bond}}(b_{ij})\right], \tag{40}$$

$$\mathbf{e}_{ij}^{(\text{surf})} = \mathrm{RBF}(r_{ij}), \tag{41}$$

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

$$\mathbf{u}_i^{(\ell)} = \beta_{i,\text{all}}^{(\ell)} U_{\text{all}}^{(\ell)} \mathbf{m}_i^{(\text{all})} + \beta_{i,\text{bond}}^{(\ell)} U_{\text{bond}}^{(\ell)} \mathbf{m}_i^{(\text{bond})} + \beta_{i,\text{surf}}^{(\ell)} U_{\text{surf}}^{(\ell)} \mathbf{m}_i^{(\text{surf})}, \tag{42}$$

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

$$u_t^{\phi}(\mathbf{X}_t, \mathbf{c})_i = \sum_{\ell=1}^{L} \mathbf{u}_i^{(\ell)} + \mathrm{Head}\Big(\mathbf{h}_i^{(L)}, \mathbf{v}_i^{(L)}, \mathbf{e}_t\Big), \qquad i \in \mathcal{A}, \tag{43}$$

with

$$u_t^{\phi}(\mathbf{X}_t, \mathbf{c})_i = \mathbf{0}, \qquad i \in \mathcal{S} \cup \mathcal{C}. \tag{44}$$

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., 2025).

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

$$R_{\text{bond}}(\mathbf{X}_{\mathcal{A}}; \mathcal{G}_{\mathcal{A}}) = \alpha \sum_{(i,j)\in\mathcal{E}_{\mathcal{A}}} \frac{1}{2}\big[d_{ij} - c_{\text{long}}\gamma\ell_{ij}\big]_+^2 + \alpha w_{\text{short}} \sum_{(i,j)\in\mathcal{E}_{\mathcal{A}}} \frac{1}{2}\big[\eta\ell_{ij} - d_{ij}\big]_+^2 + \alpha \sum_{(i,j)\notin\mathcal{E}_{\mathcal{A}}} \frac{1}{2}\big[c_{\text{nb}}\gamma^{-1}\ell_{ij} - d_{ij}\big]_+^2, \tag{45}$$

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

$$\mathbf{n} = \frac{\mathbf{a} \times \mathbf{b}}{\|\mathbf{a} \times \mathbf{b}\|} \tag{46}$$

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

$$h_i = \mathbf{n}^\top \mathbf{x}_i, \qquad h_{\mathrm{ref}} = \max_{j \in \mathcal{S} \cup \mathcal{C}} \mathbf{n}^\top \mathbf{x}_j. \tag{47}$$

The code defines a target adsorption height

$$h_{\mathrm{eq}} = h_{\mathrm{ref}} + d_{\mathrm{eq}}, \tag{48}$$

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

$$R_{\mathrm{wall}} = \frac{k_{\mathrm{wall}}}{2} \sum_{i \in \mathcal{A}} [h_{\mathrm{ref}} - h_i]_+^2, \tag{49}$$

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.

A.4. 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. The adsorbate is flagged as dissociated if its internal connectivity changes during relaxation – that is, if any covalent bond between two adsorbate atoms breaks, or any new such bond forms, when comparing the initial and the relaxed structure. The connectivity is computed at $\alpha = 1$ on the adsorbate atoms taken in isolation, so the test is independent of the surrounding slab.

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.

A.5. 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 $\mathcal{C}$: $\mathcal{T} = \left\{ w\left(\mathbf{f}^X_{p,xy} - \mathbf{f}^X_{q,xy}\right) : p, q \in \mathcal{C},\ Z_p = Z_q \right\}$. 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,

$$\Delta \mathbf{r}_{ij}(t) = \mathbf{L} \begin{bmatrix} w\left(f^X_{i,x} - f^Y_{j,x} - t_x\right) \\ w\left(f^X_{i,y} - f^Y_{j,y} - t_y\right) \\ f^X_{i,z} - f^Y_{j,z} \end{bmatrix}, \tag{50}$$

$$\mathrm{RMS}(M, t) = \left(\frac{1}{|M|} \sum_{(i,j) \in M} \|\Delta \mathbf{r}_{ij}(t)\|^2 \right)^{1/2}. \tag{51}$$

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

$$\text{A-RMSD}(X, Y) = \min_{t \in \mathcal{T}^\star} \mathrm{RMS}(\pi_\mathcal{A}(t), t). \tag{52}$$

B. Experiment Details

B.1. 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. The *Uniform-source* and *Static-source* variants omit this stage and train only the downstream flow-sampling model from scratch.

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 $(r_{\mathrm{mol}}, r_{\mathrm{surf}}, r_{\mathrm{slab}})$	$(6, 6, 6)$ Å
Downstream objective	$\mathcal{L}_{\mathrm{FS}} + \lambda_{\mathrm{damp}} \mathcal{L}_{\mathrm{damp}}$
Flow-sampling parameters $(\gamma, \tau, \lambda_{\mathrm{damp}})$	$(0.1,\ 5 \times 10^{-3},\ 10)$
Integrator	Euler–Maruyama, 250 steps on $[0, 1]$
Replay buffer size	16,384
Training energy	$\widetilde{E} = E_{\mathrm{UMA}} + R_{\mathrm{bond}} + R_{\mathrm{surf}}$
MLIP instantiation	UMA-S-1.1 with OC20 task head
Bond restraint $(\alpha_{\mathrm{bond}}, \gamma_{\mathrm{bond}}, \eta, w_{\mathrm{short}})$	$(500,\ 1.15,\ 0.55,\ 1.0)$
Surface restraint $(d_{\mathrm{eq}}, k_{\mathrm{attract}}, k_{\mathrm{repel}})$	(2.0 Å, $1000,\ 0)$

B.2. MLIP evaluation 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 $\mathcal{C}$ constrained during relaxation. We use the following L-BFGS hyperparameters: force-convergence threshold $f_{\max} = 0.01\,\text{eV/Å}$, maximum number of steps 300, per-atom displacement cap $0.04\,\text{Å}$, history size 50, damping factor 1.0, and initial inverse-Hessian curvature $\alpha = 70\,\text{eV/Å}^2$.

B.3. 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 $\max(1, \text{round}(40/L))$, 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.

C. Additional Results

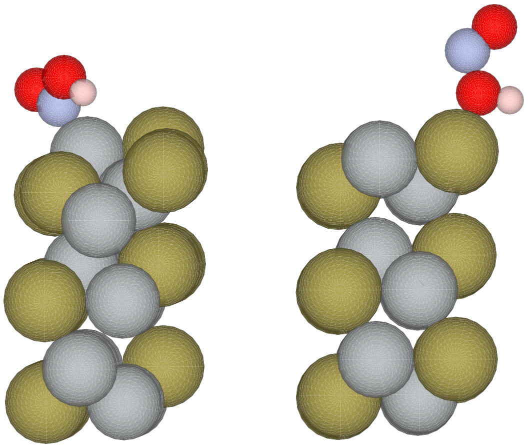
C.1. Mode collapse analysis

To better understand failure modes of energy-trained adsorbate sampling, we analyze an OC20-Dense training system for which AdsorbSample (uniform source) does not generate any candidates within the DFT success threshold of the reference minimum. The reference low-energy structure for this system corresponds to an HNO_2 adsorbate on a Ni–Te slab, with the correct binding motif given by N coordination to a Ni surface atom, as seen in Figure 3a. However, during training the sampler increasingly assigns probability mass to alternative contact motifs that do not recover this low-energy binding mode.

For each saved training epoch, we generated 1000 unrelaxed samples from the current sampler. Each sample was assigned a contact motif by identifying the adsorbate atom closest to the slab and the element of its nearest slab atom. We then computed the fraction of generated samples assigned to each adsorbate–slab contact pair. This diagnostic is applied before MLIP relaxation, so it measures the proposal distribution learned by the sampler rather than corrections introduced by downstream relaxation.

Figure 4 shows that the fraction of samples in the correct N-on-Ni motif decreases over training. Early in training, the sampler still produces a non-negligible fraction of N-on-Ni contacts, but this mode is gradually lost. In particular, late-epoch samples are dominated by O contacts, indicating collapse toward a higher-energy binding family. This provides an example of mode collapse in the learned proposal distribution: although the target system contains a low-energy N-bound basin, the sampler contracts onto a competing motif and therefore fails to produce candidates that can recover the DFT reference minimum.

This analysis suggests that, for some systems, energy-trained sampling can over-concentrate on locally attractive but globally suboptimal contact motifs. Future work could mitigate this behavior by enforcing replay-buffer diversity across surface sites, or using motif-stratified sampling to encourage exploration.



(a) OC20 Dense minimum energy sample binds through N on Ni.

(b) Un-relaxed model sample from final epoch binds through O on Te.

Figure 3. Atom colors: O red, N blue, H pink, Ni gray, Te yellow.

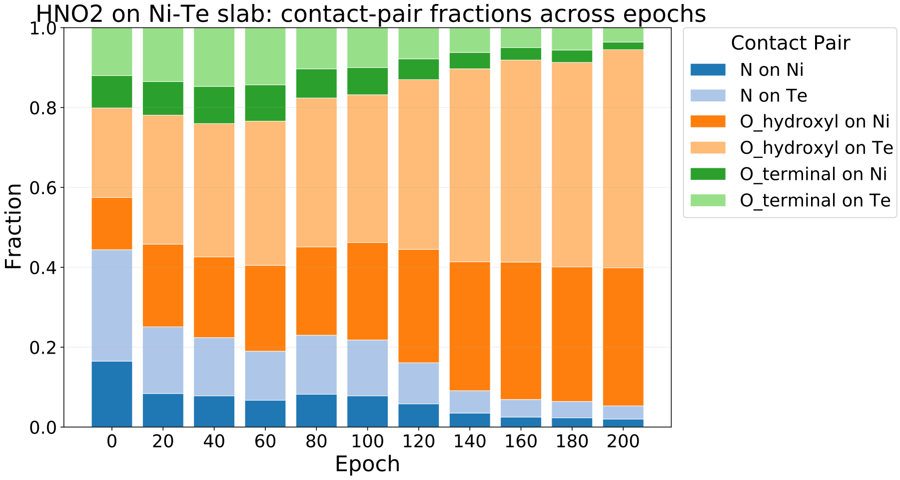


Figure 4. Mode-collapse diagnostic for an HNO_2 adsorbate on a Ni–Te slab. Fraction of correct motif samples (N on Ni, light blue) decreases throughout training.