FLEXMOF: JOINT COMPOSITION AND ALL-ATOM COORDINATE GENERATION FOR METAL–ORGANIC FRAMEWORKS

Anonymous authors
Paper under double-blind review

ABSTRACT

Metal–organic frameworks (MOFs) are porous crystals assembled from metal nodes and organic linkers, offering tunable chemistry and pore structures for gas storage, carbon capture, and catalysis. Their vast design space presents an opportunity for AI-assisted discovery, but useful generation requires more than selecting building blocks: these building blocks must assemble into a periodic crystal with appropriate atomic conformations and pore geometry. Existing approaches often restrict the internal geometry of building blocks or learn composition independently of structure, limiting the connection between composition generation and its three-dimensional realization. We introduce FlexMOF, a unified model for joint composition and all-atom structure generation. Its coordinate diffusion process incorporates molecular topology and periodic interactions while allowing every atomic coordinate and the lattice to evolve. Building on this structural representation, joint training uses a composition encoder shared by the composition and structure pathways, together with latent features, to propagate coordinate supervision into the graph representation used for composition generation. This enables composition learning to benefit directly from modeling three-dimensional structures. FlexMOF achieves significantly better single-sample crystal structure recovery than the reported state-of-the-art all-atom baseline. Joint composition and structure generation also yields chemically and structurally valid, novel, and unique MOF candidates. By connecting flexible all-atom modeling with composition learning, FlexMOF represents an important step toward AI-assisted MOF discovery through the unified design of composition and all-atom periodic structure.

1 INTRODUCTION

Metal–organic frameworks (MOFs) are porous crystals assembled from building blocks including metal nodes and organic linkers. Their tunable chemistry and pore structures support applications in gas storage, carbon capture, and catalysis (Fu et al., 2024; Sriram et al., 2025). Exploring their vast design space requires more than selecting building blocks. These building blocks must form a periodic crystal with appropriate atomic conformations and pore geometry. Crystal structure prediction of MOFs determines atomic positions and lattice parameters from given building-block graphs, whereas de novo generation must also propose the building blocks themselves. Connecting composition to its three-dimensional realization is therefore central to both tasks.

Modeling this geometry is challenging for MOF unit cells that can contain hundreds of atoms. Coarse-grained methods simplify the problem by treating building blocks as rigid or partially constrained units, limiting changes in their internal geometry (Fu et al., 2024; Kim et al., 2025b; Jiao et al., 2025; Kim et al., 2025a). Recent all-atom models enable flexible structure prediction directly from molecular graphs (Kim et al., 2026), but increase the dimensionality of the generative state. Periodicity further increases the complexity by coupling atomic geometry to the lattice parameters: atoms across cell boundaries can be nearest neighbors, and their relationships change as atomic positions and the cell evolve during structure generation. Representing atom interactions across periodic images explicitly with separate attention operations introduces significant additional computational and memory costs.

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.

Generating MOF compositions raises a complementary learning problem: their composition representation should reflect the periodic structures they can support. Two-stage approaches connect composition generation to a separate structure model (Kim et al., 2025a). Passing generated building blocks between stages, however, does not by itself couple the learning of composition and geometry. Joint latent models already learn chemical and geometric information together through structural reconstruction (Joshi et al., 2025; Šimkus et al., 2025). In these models, the denoising process for generation operates on a compressed latent representation, with atomic positions and lattice parameters recovered by a decoder. Learning accurate composition representations requires accounting for how chemical choices constrain internal conformations and interactions across cell boundaries at the all-atom level.

We introduce FlexMOF, a unified all-atom periodic model of MOF composition and structure. Its central design is to use an explicit structural model both to predict crystals from building-block graphs and to supervise the representation from which those graphs are generated. The structural pathway denoises all atomic coordinates and the lattice explicitly in Cartesian space. Building-block graphs supply chemical context without fixing bond lengths, angles, or conformations. Corner vertices of the unit cell are modeled as cell-corner virtual tokens to represent lattice degrees of freedom, and periodic neighborhoods are rebuilt as the geometry evolves during diffusion. Periodic copies share a source atom's scalar features but have different spatial displacements. A source-atom aggregation method combines the attention to the same atom across different periodic images while retaining image-specific interactions. This avoids expanding attention over separate image tokens and achieves efficient periodicity modeling.

Joint training connects composition graph reconstruction and coordinate denoising through a shared composition encoder, while a composition latent-flow model learns to sample the resulting graph representations. With a separately trained metadata prior, FlexMOF can propose building blocks and construct their periodic structures without a reference MOF composition. On the Boyd–Woo benchmark, FlexMOF achieves a single-sample crystal matching rate of 46.31%, compared with the reported 30.0% of AtomMOF-L—an approximately 54% relative improvement. It also predicts pore geometry more accurately than the evaluated baselines. Among all generated candidates, 55.86% are simultaneously valid, novel, and unique under the strict evaluation protocol. Ablations show that periodic interactions benefit large and boundary-rich frameworks, while molecular-topology biases improve prediction beyond raw elemental and bond features. After adaptation, the structural pathway improves system-macro reference coverage (COV-R) of complete MOF–adsorbate configurations over a reproduced AtomMOF-L baseline using the same number of samples.

Figure 1 compares structure recovery, MOF generation validity, and adsorption configuration and energy results. These complementary evaluations support coupling composition learning to explicit periodic all-atom modeling.

Our contributions are:

- **Flexible all-atom periodic modeling.** We propose a diffusion process in Cartesian space for explicit all-atom coordinate and lattice generation. Our building-block-graph-conditioned model jointly updates atomic coordinates and the lattice, using source-atom aggregation to incorporate periodic interactions without expanding scalar attention over individual images.
- **Joint composition–structure learning.** We train the composition representation used for composition generation with both graph reconstruction and explicit periodic coordinate denoising. Accurate all-atom structural knowledge is incorporated into the latent representation for composition generation.
- **State-of-the-art prediction and effective generation.** FlexMOF predicts crystal structures and pore geometries more accurately by a large margin than the evaluated baselines and significantly increases the quality of de novo generation in terms of validity, novelty, and uniqueness.

2 Related Work

Modeling periodic materials. Learning atomic interactions and generating crystal structures both require representations that account for the periodic repetition of the unit cell. MatterSim learns

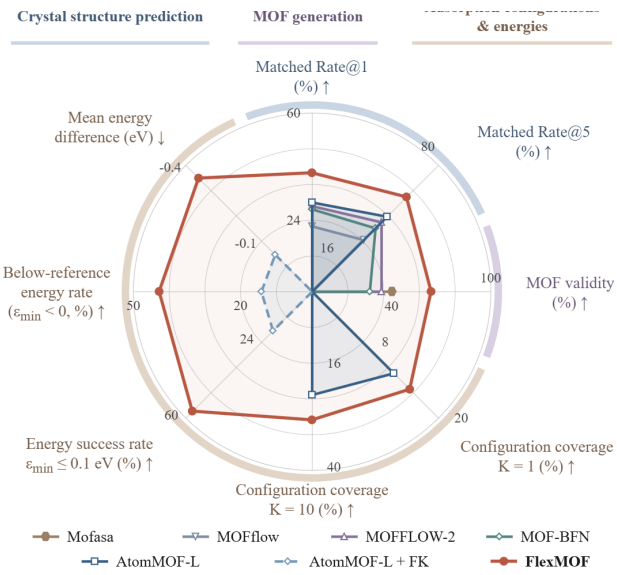


Figure 1: Cross-task comparison of structure recovery, MOF generation validity, and adsorption configurations and energies. Axes have independent scales; unavailable values appear at the origin. Adsorption protocols differ across sources (Table 2).

interatomic interactions from first-principles energies, forces, and stresses using graph-based representations of periodic atomic environments, enabling structural relaxation and molecular dynamics simulations (Yang et al., 2024). Generative models learn distributions over crystal structures: DiffCSP jointly diffuses fractional atomic coordinates and the lattice using a periodic equivariant denoising network (Jiao et al., 2023), while MatterGen jointly generates atom types, coordinates, and the lattice and supports generation under chemical, symmetry, and property constraints (Zeni et al., 2025). The All-atom Diffusion Transformer (ADiT) uses an autoencoder and latent diffusion to model both periodic materials and non-periodic molecules within a shared representation (Joshi et al., 2025). Direct application to MOF generation remains challenging: interatomic models require a supplied structure, while the crystal benchmarks used by these generative models primarily involve small unit cells (Jiao et al., 2023; Zeni et al., 2025; Joshi et al., 2025). MOFs often contain hundreds of atoms per cell, increasing the computational cost of modeling periodic interactions and the difficulty of preserving linker connectivity and metal–linker coordination during generation (Kim et al., 2025b; 2026).

MOF structure prediction. MOF structure prediction seeks a periodic arrangement of specified building blocks. MOFFlow uses flow matching to predict building-block translations, rotations, and lattice parameters, treating the supplied building blocks as rigid bodies (Kim et al., 2025b). MOF-BFN retains fixed local building-block geometries and explicitly incorporates periodicity through fractional coordinates and a periodic graph network while predicting building-block positions, orientations, and the lattice (Jiao et al., 2025). MOFFlow-2 introduces torsional flexibility into organic linkers while retaining bond lengths and angles from initialized conformers (Kim et al., 2025a). AtomMOF predicts all-atom structures with freely varying coordinates and lattice parameters, conditioning on chemical graphs and local geometric features obtained from RDKit conformers and metal-node templates (Kim et al., 2026). FlexMOF combines molecular connectivity with periodic neighborhoods recomputed during all-atom denoising, enabling flexible structure prediction from building-block graphs without precomputed local conformers. Aggregating features from periodic images of the same atom reduces the memory cost of modeling interactions across cell boundaries, while all atomic coordinates and the lattice remain free to evolve.

De novo MOF generation. De novo MOF generation requires selecting building blocks together with their periodic arrangement. MOFDiff generates coarse-grained structures, retrieves building blocks from a fixed library, and reconstructs all-atom frameworks through assembly (Fu et al., 2024). MOFGPT generates textual MOF descriptions in the MOFid representation for property-

guided design (Badrinarayanan et al., 2025). MOFFlow-2 trains its autoregressive building-block generator and structure predictor separately, linking them through conformer initialization (Kim et al., 2025a). BBA MOF Diffusion jointly models inorganic nodes, organic edges, and topological nets, generating building blocks in three dimensions before assembling complete frameworks (Duan et al., 2025). LEGO-MOF learns equivariant linker representations for generation, editing, and property optimization, with generated linkers incorporated into frameworks through rigid-body assembly (Zhang et al., 2026). Mofasa jointly generates atom types, fractional coordinates, and the lattice through latent diffusion, using an autoencoder of periodic structures to learn representations of chemical identity and geometry (Šimkus et al., 2025). FlexMOF couples building-block graph generation to explicit periodic all-atom denoising through shared representations and joint training. Coordinate supervision shapes the graph latent space, allowing composition generation to benefit from learned molecular geometry and periodic interactions. The generated graphs then condition structure generation, enabling a unified model to generate building blocks and predict their complete framework geometry with full atomic and lattice flexibility.

3 Joint Composition and All-Atom Structure Generation

Complete MOF generation requires chemical connectivity and compatible periodic geometry. FlexMOF couples their learning (Figure 2): composition $C = (\mathcal{H}, \mathcal{G})$ contains building-block graphs $\mathcal{G} = \{G_b\}_{b=1}^{B}$ and metadata $\mathcal{H}$ (defined below); structure $S = (\mathbf{X}, \mathbf{L})$ contains Cartesian coordinates $\mathbf{X} \in \mathbb{R}^{N\times 3}$ for N physical atoms and lattice vectors as rows of $\mathbf{L} \in \mathbb{R}^{3\times 3}$. Chemical connectivity guides geometry, while structural supervision trains building-block representations.

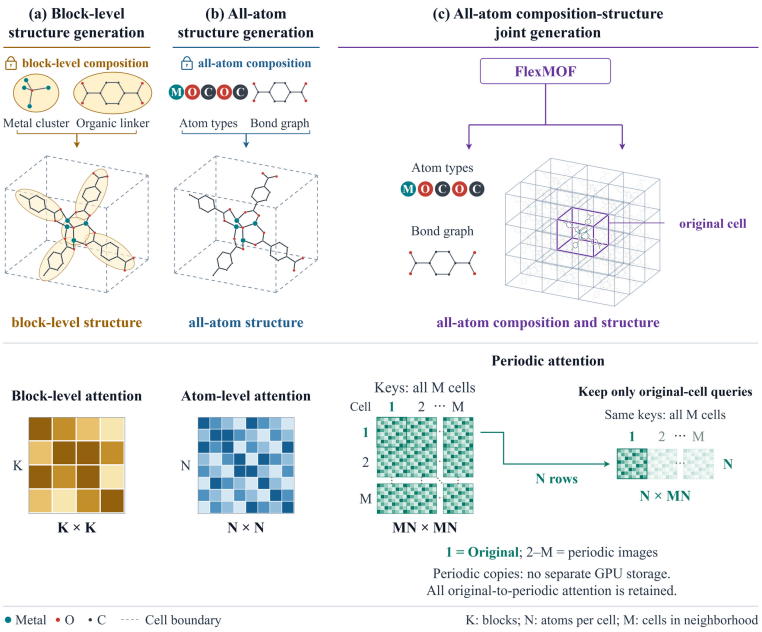


Figure 2: Joint composition and all-atom generation. The lower panel restricts queries to the original cell; Equation 3 aggregates scalar attention by source atom. M counts cells retained under the global 40 Å cutoff and image budget (Section 3.2); K counts building blocks, not experimental candidates. Images share scalar keys/values but still require storage for displacements and geometric weights.

3.1 All-Atom Structure Diffusion

Atomic conformations and the periodic cell must evolve together. We propose an all-atom diffusion process in Cartesian position space using a denoising diffusion probabilistic model (DDPM) (Ho et al., 2020). Building-block graphs specify chemistry while allowing bond lengths, angles, and conformations to vary. Virtual nodes at unit-cell vertices expose cell geometry; lattice-translated interactions capture cross-boundary neighbors during denoising (Jiao et al., 2023).

Four independent corners define the cell; all eight vertices are virtual nodes without element identities, allowing lattice vectors to adapt alongside atoms. State $\mathbf{Y}_0 \in \mathbb{R}^{(N+4)\times 3}$ contains atoms and independent corners; dependent corners are reconstructed for attention. To simplify learning across framework sizes, we adopt a count-adaptive centered cubic reference $\mathbf{Y}_{\mathrm{ref}}(N)$. Its edge length and atomic noise standard deviation scale as $N^{1/3}$, matching the initial spatial scale to atom count (Appendix D.1):

$$\mathbf{Y}_\tau = \mathbf{Y}_{\mathrm{ref}} + \sqrt{\bar{\alpha}_\tau}(\mathbf{Y}_0 - \mathbf{Y}_{\mathrm{ref}}) + \sqrt{1-\bar{\alpha}_\tau}\,\boldsymbol{\eta}_N, \qquad \bar{\alpha}_\tau = \prod_{\ell=1}^{\tau}(1-\beta_\ell). \tag{1}$$

Here τ is the discrete diffusion step, β_ℓ the noise schedule, and $\boldsymbol{\eta}_N$ Gaussian noise with count-scaled atomic variance and correlated corner displacements (Appendix D.1). The denoiser minimizes root-mean-square noise error:

$$\mathcal{L}_{\mathrm{coord}} = \mathbb{E}\left[\sqrt{\frac{\|\boldsymbol{\eta}_N - \widehat{\boldsymbol{\eta}}_\theta(\mathbf{Y}_\tau, \tau, \mathcal{G}, \mathcal{H})\|_F^2}{3(N+4)} + 10^{-5}}\right]. \tag{2}$$

The expectation covers training structures, diffusion steps, and noise; $\|\cdot\|_F$ denotes the Frobenius norm. Dependent corners and padding are excluded.

To guide denoising with connectivity, Graphormer-style shortest-path, multi-hop bond, and disconnected-pair biases encode building-block topology (Ying et al., 2021). A spatial diffusion Transformer (Peebles & Xie, 2023) combines them with radial distances; a scalar/vector decoder (Schütt et al., 2021) predicts atomic and corner noise using displacement directions. This supplies chemical context while preserving atomic flexibility (Appendix B).

3.2 Periodic attention with source-atom aggregation

To reduce the cost of explicit periodicity modeling, we propose source-atom aggregation. Atom j at integer translation $\mathbf{n}$ has displacement $\mathbf{r}_{ij\mathbf{n},\tau} = \mathbf{x}_{i,\tau} - (\mathbf{x}_{j,\tau} + \mathbf{n}\mathbf{L}_\tau)$. Translations in $[-5,5]^3$ are restricted by current cell dimensions. Images within 40 Å of eligible central tokens undergo nearest-distance selection, capped at $\max(896-N,0)$ extra tokens; the eight central corners are additional. Thus M in Figure 2 varies with geometry and budget (Appendix D.2).

Periodic copies share their source atom's scalar key/value but have distinct displacements, allowing geometric weights to be summed before scalar attention. For the retained image set $\mathcal{I}_j$ (including the central copy), bias $b_{ij\mathbf{n}}$, query/key vectors $\mathbf{q}_i, \mathbf{k}_j$, value $\mathbf{v}_j$, and head width d_h,

$$w_{ij} = \frac{\sum_{\mathbf{n}\in\mathcal{I}_j}\exp(b_{ij\mathbf{n}})}{\sum_k \sum_{\mathbf{m}\in\mathcal{I}_k}\exp(b_{ik\mathbf{m}})}, \qquad \mathbf{o}_i = \sum_j \mathrm{softmax}_j\left(\frac{\mathbf{q}_i^\top \mathbf{k}_j}{\sqrt{d_h}} + \log w_{ij}\right)\mathbf{v}_j. \tag{3}$$

Masked pairs have zero weight. For fixed images and shared keys/values, aggregation preserves expanded-image scalar attention before dropout, reducing its repeated weight array from TT_{ext} to T^2 for $T = N+8$ central and T_{ext} central-plus-image tokens. Geometry is reused across spatial layers within each denoiser call; vector updates retain image-specific directions. This avoids repeated scalar attention while retaining periodic geometry (Appendix D.2).

3.3 Composition generation and the joint distribution

In composition diffusion, we explicitly model the multiplicity of unique building blocks to reduce the complexity of composition learning. Each distinct block is generated once and replicated by its multiplicity. Metadata $\mathcal{H}$ specifies an atom-count estimate, unique-block count B, retained-node counts n_b, multiplicities m_b, and linker/singleton roles r_b. Graph $G_b = (\mathbf{a}_b, \mathbf{A}_b, \mathbf{h}_b)$ stores elements, binary intramolecular adjacency, and attached-hydrogen counts. Hydrogen expansion gives $N = \sum_{b=1}^{B} m_b(n_b + \sum_{i=1}^{n_b} h_{bi})$. We therefore decompose the joint distribution of composition and structure into metadata generation, conditional graph generation, and conditional structure generation:

$$p_{\eta,\phi,\theta}(C,S) = p_\eta(\mathcal{H})\, p_\phi(\mathcal{G} \mid \mathcal{H})\, p_\theta(\mathbf{X}, \mathbf{L} \mid \mathcal{G}, \mathcal{H}). \tag{4}$$

Here η, ϕ, θ index metadata, graph, and coordinate distributions; the latter two share parameters. The learned metadata prior supplies sizes and roles without a reference MOF.

To generate discrete chemistry through a continuous flow, we adopt LG-Flow's graph-latent formulation (Siraudin & Morris, 2026). A graph encoder maps chemical labels and Laplacian features to per-node Gaussian latents (Kingma & Welling, 2014; Dwivedi et al., 2023). Reconstruction and Kullback–Leibler regularization define $\mathcal{L}_{\mathrm{AE}}$ (Appendix C); standardized posterior means $\mathbf{Z}_{b,0} \in \mathbb{R}^{n_b \times d}$ provide flow-matching targets (Lipman et al., 2023):

$$\mathbf{Z}_{b,t} = (1-t)\mathbf{Z}_{b,0} + t\boldsymbol{\epsilon}_b, \quad \boldsymbol{\epsilon}_b \sim \mathcal{N}(0, I), \quad \mathbf{v}_b = \boldsymbol{\epsilon}_b - \mathbf{Z}_{b,0}. \tag{5}$$

Latent time $t \in [0, 1]$ runs from data to noise independently of DDPM step τ. Conditioned on $\mathcal{H}$, a self-conditioned Transformer exchanges block summaries alongside graph-local attention to supply context from other constituents and predicts $\widehat{\mathbf{Z}}_{b,0}$ with objective:

$$\mathcal{L}_{\mathrm{latent}} = \mathbb{E}\left[\frac{1}{B}\sum_{b=1}^{B}\frac{1}{dn_b}\left\|\frac{\widehat{\mathbf{Z}}_{b,0} - \mathbf{Z}_{b,0}}{\max(t, t_{\min})}\right\|_F^2\right]. \tag{6}$$

Here d is the latent width; $t_{\min} > 0$ caps near-clean weighting. For $t \geq t_{\min}$, this equals velocity-prediction error in Equation 5. The expectation covers graph targets, latent times, and noise.

3.4 Joint training and sequential sampling

To let structural supervision shape composition representations, we couple learning through a shared composition encoder and atom-level latent injection into the coordinate denoiser. We jointly optimize reconstruction, composition generation, and periodic denoising:

$$\mathcal{L}_{\mathrm{joint}} = \mathcal{L}_{\mathrm{AE}} + \mathcal{L}_{\mathrm{latent}} + \mathcal{L}_{\mathrm{coord}}. \tag{7}$$

Coordinate gradients reach the graph posterior, training composition representations to support periodic denoising and graph reconstruction. Latent-flow targets are detached for stability; the metadata prior is fitted separately (Appendix E).

Sampling follows Equation 4: draw metadata, generate graphs with 100 Heun intervals from $t = 1$ to $t = 0$, then denoise atoms and the cell with 50 second-order DPM-Solver++ network evaluations (Lu et al., 2022). This supplies chemistry before geometry generation without a reference MOF (Appendix E.2; Figure 6).

4 Experiments

We evaluate structure recovery and pore properties, ablate periodic and molecular-graph information, then test MOF–adsorbate transfer and complete generation.

4.1 Structural matching and geometric fidelity

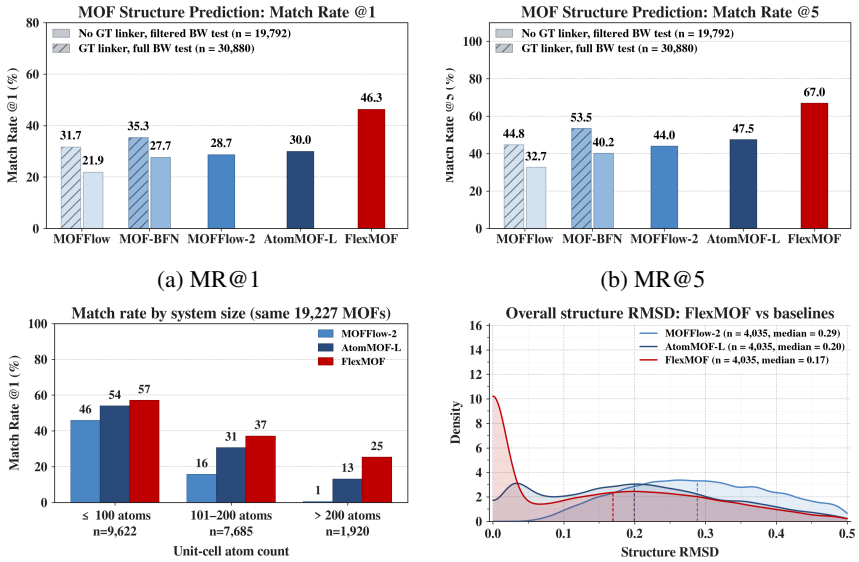
The Boyd–Woo database (BW-DB) contains hypothetical MOFs assembled from building blocks and crystal topologies (Boyd et al., 2019; Fu et al., 2024). Match rate $\mathrm{MR@}K$ is the percentage of references recovered by any of K candidates under periodic structure matching (Ong et al., 2013). The main test set comprises 19,792 references passing MOFChecker's chemical and geometric checks, aligned with the filtered MOFFlow-2 test set. Missing or failed predictions are failures.

FlexMOF recovers 9,165 references at MR@1 (46.31%) and 13,256 at MR@5 (66.98%; Figure 3). Reported AtomMOF-L scores are 30.0% and 47.5%: gains of 16.31 and 19.48 percentage points, or 54.4% relative at MR@1. The results support higher overall structure prediction quality with FlexMOF.

Among 19,227 shared MOFs, the ≤ 100, 101–200, and > 200-atom groups contain 9,622, 7,685, and 1,920 structures. FlexMOF achieves 57%, 37%, and 25% MR@1, versus AtomMOF-L's 54%, 31%, and 13% and MOFFlow-2's 46%, 16%, and 1%. Recovery declines with size for all methods; FlexMOF's lead over AtomMOF-L grows from 3 to 12 points. The advantage of FlexMOF becomes more evident as the size of MOFs grows.

Matched-geometry error uses aligned root-mean-square deviation (RMSD), which is shown in Figure 3(d), normalized by atomic spacing $(V/N)^{1/3}$ for cell volume V and atom count N. Medians are

0.17, 0.20, and 0.29 for FlexMOF, AtomMOF-L, and MOFFlow-2. Each contributes 4,035 method-specific matches, so the comparison is unpaired. These results support FlexMOF for more accurate recovering of periodic MOF geometry.



(a) MR@1 (b) MR@5

(c) MR@1 by atom count (d) Overall RMSD density

Figure 3: Structure recovery: (a,b) MR@1 and MR@5 on filtered BW ($n = 19,792$); gray bars retain published unfiltered-population results. (c) MR@1 by atom count on 19,227 shared references. (d) Normalized RMSD density on 4,035 matches per method from method-specific sets; dashed lines mark medians. Higher MR and lower RMSD are better.

4.2 Pore Property Evaluation

We evaluate volumetric and gravimetric surface area (VSA, GSA), accessible void fraction (VF), pore-limiting diameter (PLD), largest-cavity diameter (LCD), and density using mean absolute error (MAE). Zeo++ computes pore geometry (Willems et al., 2012). At $K = 1$, shared calculable pairs within 19,227 conditions include unmatched structures. FlexMOF lowers all six MAEs against MOFFlow-2 and five against AtomMOF-L (Table 1). Against AtomMOF-L, GSA and VF errors fall by 21.4% and 28.4%; density remains less accurate. These errors support all-atom structure generation for reproducing accessible pore geometry.

Table 1: Pore-property MAE on the shared calculable intersection within 19,227 conditions. ↑/↓ indicate higher/lower is better. n is the number of shared calculable pairs for each property.

Property	n	MOFFlow-2 ↓	AtomMOF-L ↓	FlexMOF ↓
VSA (m^2/cm^3)	19,018	197.059	108.849	**85.816**
GSA (m^2/g)	19,018	293.709	140.014	**110.064**
Accessible VF	19,044	0.032301	0.012075	**0.008651**
PLD (Å)	19,029	0.9112	0.5527	**0.4778**
LCD (Å)	19,029	0.9697	0.5187	**0.4385**
Density (g/cm^3)	19,044	0.021535	**0.006143**	0.008041

4.3 Ablation Studies

Periodic boundary conditions. We study the effect of explicit modeling of periodicity under matched data and sampling on 30,880 references at MR@5. Cross-boundary fraction is the share of reference adjacency relations crossing cell boundaries. Periodicity is more important for large or

boundary-rich structures: +4.78 percentage points for $N \geq 400$ and +2.63 for the highest boundary-fraction quartile, supporting explicit cross-cell contacts. The largest bin gain is 21.2 points (201–300 atoms, fraction ≥ 0.25, $n = 52$; Figure 4).

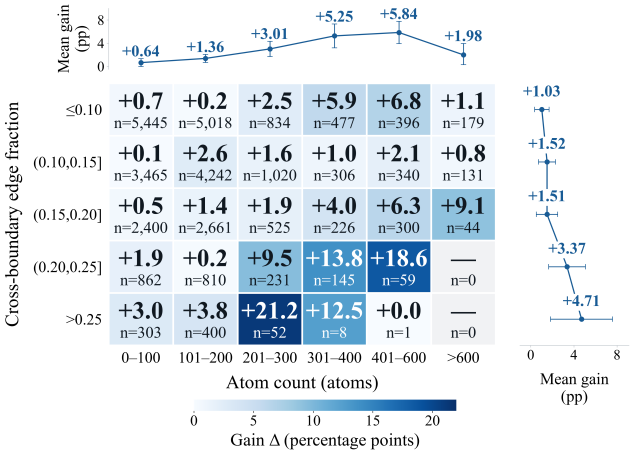


Figure 4: PBC-on minus PBC-off MR@5 gain on BW. Cells give percentage-point gains and counts by atom count and reference cross-boundary fraction; empty cells have $n = 0$. Top/right panels show sample-weighted marginal means with 95% paired confidence intervals.

Molecular-graph topology ablation. Removing shortest-path, multi-hop bond-sequence, and disconnected-pair attention biases, while retaining element/bond features and the periodic coordinate pathway, lowers MR@1 to 41.02% (8,119/19,792) and MR@5 to 62.25% (12,321/19,792). Against the rounded same-cohort joint reference (46.3%, 67.0%), losses are 5.28 and 4.75 percentage points. Molecular topology therefore adds chemical context beyond local features (Appendix F.3). Both ablations support periodic interactions and topology-aware attention as effective denoising guidance.

4.4 Transfer to MOF–adsorbate configurations

ODAC25 provides MOF–adsorbate structures for direct-air-capture modeling (Sriram et al., 2025). We finetune the model on ODAC25 to adapt the coordinate pathway to generate framework atoms, guests, and the cell jointly from specified chemistry and guest counts. The test set comprises 3,582 references in 683 systems. System-macro reference coverage (COV-R) averages complete MOF–adsorbate reference-match fractions over equally weighted systems, pooling candidates within each system (Appendix G.2).

At $K = 10$, COV-R reaches 28.75%, exceeding reproduced AtomMOF-L's 22.52% by 6.23 points (27.7% relative). Published AtomMOF-L coverage (23.11%) uses a different time budget (Table 2). The coverage gain supports transferring all-atom denoising to joint framework–guest generation.

Table 2: MOF–adsorbate coverage and energy results. FlexMOF and reproduced AtomMOF-L use system-macro COV-R with K candidates per reference; published coverage uses its time budget. SR: energy success rate; Mean: mean $\varepsilon_{\min}$ over converged systems. Published FK-steered AtomMOF-L uses a separate relaxation protocol (Kim et al., 2026); – denotes unreported values.

Method	Budget	COV-R (%) ↑	SR (%) ↑	$\varepsilon_{\min} < 0$ (%) ↑	Mean (eV) ↓
FlexMOF	$K = 10$	**28.75**	**56.81**	**42.75**	**-0.3486**
AtomMOF-L (reproduced)	$K = 1$	12.92	–	–	–
AtomMOF-L (reproduced)	$K = 10$	22.52	–	–	–
AtomMOF-L (reported)	< 1 s	23.11	–	–	–
AtomMOF-L+FK (reported)	$K = 1$	–	12.30	6.30	0.0841
AtomMOF-L+FK (reported)	$K = 10$	–	18.59	14.20	−0.0462

Adsorption energies. After fixed-cell relaxation with the eSEN-ODAC25-Full interatomic potential, $\varepsilon_{\min}$ subtracts each system's lowest reference adsorption energy from its lowest converged candidate energy. Success requires convergence and $\varepsilon_{\min} \leq 0.1\,\mathrm{eV}$. At $K = 10$, 56.81% of all 683 systems succeed and 42.75% fall below the reference minimum. Mean $\varepsilon_{\min}$ over 459 systems with convergence is $-0.3486\,\mathrm{eV}$. Published FK-steered AtomMOF-L uses Feynman–Kac (FK) steering with energy or overlap penalties and a different relaxation protocol (Kim et al., 2026).

4.5 Complete MOF generation

We generate composition and structure without a reference MOF or post-generation relaxation. Validity (V) requires passing all 15 MOFChecker checks (Jablonka, 2023); novelty (N), absence from indexed BW training structures; and uniqueness (U), a MOFid identifier occurring exactly once (Bucior et al., 2019). Strict VNU intersects all three. Of all 10,000 draws, 6,669 are valid (66.69%), 8,393 novel (83.93%), 9,621 unique (96.21%), and 5,586 satisfy VNU (55.86%).

Joint training improves validity and strict VNU over separate training by 8.54 and 7.73 percentage points, respectively. VNU improves by 16.1% relative; novelty and singleton uniqueness rise by 1.39 and 2.33 points. These gains support coupling composition learning with structure supervision to improve complete MOF generation.

Compared with published Mofasa (Šimkus et al., 2025), BW-only validity without relaxation improves by 21.89 points (Table 3), showing higher de novo generation quality even without higher-quality experiment MOF structures for training and force field relaxation. Figure 8 shows representative outputs and declining validity with size.

Table 3: Complete-generation results (%). V: MOFChecker validity; N: novelty; U: singleton uniqueness; VNU: their intersection. Mofasa's EU/EVNU additionally require successful MOFid extraction, which holds for all FlexMOF outputs. – denotes unreported results.

Method	Training data	Relaxation	V ↑	N ↑	U (EU) ↑	VNU (EVNU) ↑
Mofasa	BW	None	44.8	–	–	–
Mofasa	QMOF	None	52.9	–	73.5	42.4
Mofasa-opt	QMOF	Orb + D3	59.9	–	72.7	48.2
Mofasa	Exp.	None	52.2	–	71.8	36.8
Mofasa-opt	Exp.	Orb + D3	62.8	–	72.6	44.8
FlexMOF (Separate training)	BW	None	58.15	82.54	93.88	48.13
FlexMOF (Joint training)	BW	None	**66.69**	83.93	**96.21**	**55.86**

FlexMOF (Separate training) trains composition and structure models separately; FlexMOF (Joint training) trains them jointly. QMOF contains quantum-chemical MOF data; Exp. combines Mofasa's experimental collections. Orb + D3 denotes interatomic-potential relaxation with dispersion correction. All FlexMOF rates use 10,000 draws; novelty compares against 213,583 of 247,062 BW training records. Baseline sample counts and identity protocols differ (Appendix F.5).

5 Discussion

FlexMOF couples composition generation and flexible periodic all-atom denoising through a shared composition encoder and coordinate supervision, learning building-block graphs under geometric supervision and transferring to MOF–adsorbate configurations. Evaluation on topology-assembled frameworks and computational MOF–adsorbate data leaves synthetic accessibility and operating stability unresolved. Large-framework generation remains difficult, motivating generalization tests beyond the evaluated chemical and structural distributions. Our next step is property-guided MOF design: conditioning composition and structure generation on target pore or adsorption properties while preserving chemical feasibility and diversity. Selected candidates will undergo independent property calculations, relaxation, stability tests, and experimental validation.

AI USE STATEMENT

Generative AI assisted manuscript preparation and editing, literature searches, and experimental planning and execution. The authors reviewed the outputs and take responsibility for the manuscript.

ETHICS STATEMENT

This study uses computational materials datasets and involves no human participants.

REPRODUCIBILITY STATEMENT

Section 3 defines the model and objectives; Appendix B details the architecture. Appendix A describes the data; Appendices C–E give representation, computation, and sampling details. Appendices F and G specify evaluation populations, matching conventions, uncertainty estimates, and energy evaluation.

REFERENCES

Srivathsan Badrinarayanan, Rishikesh Magar, Akshay Antony, Radheesh Sharma Meda, and Amir Barati Farimani. MOFGPT: Generative design of metal–organic frameworks using language models. *Journal of Chemical Information and Modeling*, 65(17):9049–9060, 2025. doi: 10.1021/acs.jcim.5c01625.

Erik Bitzek, Pekka Koskinen, Franz Gähler, Michael Moseler, and Peter Gumbsch. Structural relaxation made simple. *Physical Review Letters*, 97(17):170201, 2006. doi: 10.1103/PhysRevLett.97.170201.

Peter G. Boyd, Arunraj Chidambaram, Enrique García-Díez, Christopher P. Ireland, Thomas D. Daff, Richard Bounds, Andrzej Gładysiak, Pascal Schouwink, Seyed Mohamad Moosavi, M. Mercedes Maroto-Valer, Jeffrey A. Reimer, Jorge A. R. Navarro, Tom K. Woo, Susana Garcia, Kyriakos C. Stylianou, and Berend Smit. Data-driven design of metal–organic frameworks for wet flue gas CO2 capture. *Nature*, 576:253–256, 2019. doi: 10.1038/s41586-019-1798-7.

Benjamin J. Bucior, Andrew S. Rosen, Maciej Haranczyk, Zhenpeng Yao, Michael E. Ziebel, Omar K. Farha, Joseph T. Hupp, J. Ilja Siepmann, Alán Aspuru-Guzik, and Randall Q. Snurr. Identification schemes for metal–organic frameworks to enable rapid search and cheminformatics analysis. *Crystal Growth & Design*, 19(11):6682–6697, 2019. doi: 10.1021/acs.cgd.9b01050.

Chenru Duan, Aditya Nandy, Sizhan Liu, Yuanqi Du, Liu He, Yi Qu, Haojun Jia, and Jin-Hu Dou. Building-block aware generative modeling for 3D crystals of metal organic frameworks. *arXiv preprint arXiv:2505.08531*, 2025. URL https://arxiv.org/abs/2505.08531.

Vijay Prakash Dwivedi, Chaitanya K. Joshi, Anh Tuan Luu, Thomas Laurent, Yoshua Bengio, and Xavier Bresson. Benchmarking graph neural networks. *Journal of Machine Learning Research*, 24(43):1–48, 2023. URL https://www.jmlr.org/papers/v24/22-0567.html.

Xiang Fu, Tian Xie, Andrew S. Rosen, Tommi S. Jaakkola, and Jake Allen Smith. MOFDiff: Coarse-grained diffusion for metal–organic framework design. In *International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=0VBsoluxR2.

Xiang Fu, Brandon M. Wood, Luis Barroso-Luque, Daniel S. Levine, Meng Gao, Misko Dzamba, and C. Lawrence Zitnick. Learning smooth and expressive interatomic potentials for physical property prediction. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 17875–17893. PMLR, 2025. URL https://proceedings.mlr.press/v267/fu25h.html.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851, 2020. URL https://papers.neurips.cc/paper_files/paper/2020/hash/4c5bcfec8584af0d967f1ab10179ca4b-Abstract.html.

Kevin Maik Jablonka. mofchecker. Open-source software, 2023. URL https://github.com/lamalab-org/mofchecker.

Rui Jiao, Wenbing Huang, Peijia Lin, Jiaqi Han, Pin Chen, Yutong Lu, and Yang Liu. Crystal structure prediction by joint equivariant diffusion. In *Advances in Neural Information Processing Systems*, volume 36, pp. 17464–17497, 2023. doi: 10.52202/075280-0767. URL https://papers.neurips.cc/paper_files/paper/2023/hash/38b787fc530d0b31825827e2cc306656-Abstract-Conference.html.

Rui Jiao, Hanlin Wu, Wenbing Huang, Yuxuan Song, Yawen Ouyang, Yu Rong, Tingyang Xu, Pengju Wang, Hao Zhou, Wei-Ying Ma, Jingjing Liu, and Yang Liu. MOF-BFN: Metal–organic frameworks structure prediction via Bayesian flow networks. In *Advances in Neural Information Processing Systems*, volume 38, 2025. doi: 10.52202/085713-5410. URL https://openreview.net/forum?id=pNwiFucAtA.

Chaitanya K. Joshi, Xiang Fu, Yi-Lun Liao, Vahe Gharakhanyan, Benjamin Kurt Miller, Anuroop Sriram, and Zachary Ward Ulissi. All-atom diffusion transformers: Unified generative modelling of molecules and materials. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 28393–28417, 2025. URL https://proceedings.mlr.press/v267/joshi25a.html.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. In *Advances in Neural Information Processing Systems*, volume 35, pp. 26565–26577, 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/a98846e9d9cc01cfb87eb694d946ce6b-Abstract.html.

Nayoung Kim, Seongsu Kim, and Sungsoo Ahn. Flexible MOF generation with torsion-aware flow matching. In *Advances in Neural Information Processing Systems*, volume 38, 2025a. doi: 10.52202/085713-0464.

Nayoung Kim, Seongsu Kim, Minsu Kim, Jinkyoo Park, and Sungsoo Ahn. MOFFlow: Flow matching for structure prediction of metal–organic frameworks. In *International Conference on Learning Representations*, 2025b. URL https://openreview.net/forum?id=dNT3abOsLo.

Nayoung Kim, Honghui Kim, Sihyun Yu, Minkyu Kim, Seongsu Kim, and Sungsoo Ahn. AtomMOF: All-atom flow matching for MOF–adsorbate structure prediction. *arXiv preprint arXiv:2602.07351*, 2026. URL https://arxiv.org/abs/2602.07351.

Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. In *International Conference on Learning Representations*, 2014. URL https://arxiv.org/abs/1312.6114.

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. doi: 10.1088/1361-648X/aa680e.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=PqvMRDCJT9t.

Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. DPM-Solver++: Fast solver for guided sampling of diffusion probabilistic models. *arXiv preprint arXiv:2211.01095*, 2022. URL https://arxiv.org/abs/2211.01095.

Shyue Ping Ong, William Davidson Richards, Anubhav Jain, Geoffroy Hautier, Michael Kocher, Shreyas Cholia, Dan Gunter, Vincent L. Chevrier, Kristin A. Persson, and Gerbrand Ceder. Python materials genomics (pymatgen): A robust, open-source Python library for materials analysis. *Computational Materials Science*, 68:314–319, 2013. doi: 10.1016/j.commatsci.2012.10.028. URL https://pymatgen.org/.

William Peebles and Saining Xie. Scalable diffusion models with transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4195–4205, 2023. URL https://openaccess.thecvf.com/content/ICCV2023/html/Peebles_Scalable_Diffusion_Models_with_Transformers_ICCV_2023_paper.html.

Andrew S. Rosen, Shaelyn M. Iyer, Debmalya Ray, Zhenpeng Yao, Alán Aspuru-Guzik, Laura Gagliardi, Justin M. Notestein, and Randall Q. Snurr. Machine learning the quantum-chemical properties of metal–organic frameworks for accelerated materials discovery. *Matter*, 4(5):1578–1597, 2021. doi: 10.1016/j.matt.2021.02.015.

Andrew S. Rosen, Victor Fung, Patrick Huck, C. T. O'Donnell, Matthew K. Horton, Donald G. Truhlar, Kristin A. Persson, Justin M. Notestein, and Randall Q. Snurr. High-throughput predictions of metal–organic framework electronic properties: Theoretical challenges, graph neural networks, and data exploration. *npj Computational Materials*, 8:112, 2022. doi: 10.1038/s41524-022-00796-6.

Kristof T. Schütt, Oliver T. Unke, and Michael Gastegger. Equivariant message passing for the prediction of tensorial properties and molecular spectra. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139, pp. 9377–9388. PMLR, 2021. URL https://proceedings.mlr.press/v139/schutt21a.html.

Vaidotas Šimkus, Anders Christensen, Steven Bennett, Ian Johnson, Mark Neumann, James Gin, Jonathan Godwin, and Benjamin Rhodes. Mofasa: A step change in metal–organic framework generation. *arXiv preprint arXiv:2512.01756*, 2025. URL https://arxiv.org/abs/2512.01756.

Antoine Siraudin and Christopher Morris. Principled latent diffusion for graphs via laplacian autoencoders. *arXiv preprint arXiv:2601.13780*, 2026. URL https://arxiv.org/abs/2601.13780.

Anuroop Sriram, Logan M. Brabson, Xiaohan Yu, Sihoon Choi, Kareem Abdelmaqsoud, Elias Moubarak, Pim de Haan, Sindy Löwe, Johann Brehmer, John R. Kitchin, Max Welling, C. Lawrence Zitnick, Zachary Ulissi, Andrew J. Medford, and David S. Sholl. The Open DAC 2025 dataset for sorbent discovery in direct air capture. arXiv preprint arXiv:2508.03162, 2025. URL https://arxiv.org/abs/2508.03162.

Thomas F. Willems, Chris H. Rycroft, Michaeel Kazi, Juan C. Meza, and Maciej Haranczyk. Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials. *Microporous and Mesoporous Materials*, 149(1):134–141, 2012. doi: 10.1016/j.micromeso.2011.08.020.

Han Yang, Chenxi Hu, Yichi Zhou, Xixian Liu, Yu Shi, Jielan Li, Guanzhi Li, Zekun Chen, Shuizhou Chen, Claudio Zeni, Matthew Horton, Robert Pinsler, Andrew Fowler, Daniel Zügner, Tian Xie, Jake Smith, Lixin Sun, Qian Wang, Lingyu Kong, Chang Liu, Hongxia Hao, and Ziheng Lu. MatterSim: A deep learning atomistic model across elements, temperatures and pressures. *arXiv preprint arXiv:2405.04967*, 2024. URL https://arxiv.org/abs/2405.04967.

Chengxuan Ying, Tianle Cai, Shengjie Luo, Shuxin Zheng, Guolin Ke, Di He, Yanming Shen, and Tie-Yan Liu. Do transformers really perform badly for graph representation? In *Advances in Neural Information Processing Systems*, volume 34, pp. 28877–28888, 2021. URL https://papers.neurips.cc/paper_files/paper/2021/hash/f1c1592588411002af340cbaedd6fc33-Abstract.html.

Claudio Zeni, Robert Pinsler, Daniel Zügner, Andrew Fowler, Matthew Horton, Xiang Fu, Zilong Wang, Aliaksandra Shysheya, Jonathan Crabbé, Shoko Ueda, Roberto Sordillo, Lixin Sun, Jake Smith, Bichlien Nguyen, Hannes Schulz, Sarah Lewis, Chin-Wei Huang, Ziheng Lu, Yichi Zhou, Han Yang, Hongxia Hao, Jielan Li, Chunlei Yang, Wenjie Li, Ryota Tomioka, and Tian Xie. A generative model for inorganic materials design. *Nature*, 639(8055):624–632, 2025. doi: 10.1038/s41586-025-08628-5. URL https://www.nature.com/articles/s41586-025-08628-5.

Chaoran Zhang, Guangyao Li, and Dongxu Ji. LEGO-MOF: Equivariant latent manipulation for editable, generative, and optimizable MOF design. *arXiv preprint arXiv:2604.13520*, 2026. URL https://arxiv.org/abs/2604.13520.

A Data and Composition Representation

The Boyd–Woo database (BW-DB) is an existing topology-assembled corpus of hypothetical periodic MOFs (Boyd et al., 2019; Fu et al., 2024). The split used here contains 247,062 training, 30,881 validation, and 30,880 test MOFs. Building-block graphs preserve element identities and binary intramolecular connectivity. Terminal hydrogens attached to non-hydrogen atoms are encoded as counts in $\{0, 1, 2, 3, 4\}$; other hydrogens remain explicit. Expanding these counts and replicating building blocks exactly recovers the original chemical specification for all 238,946 building-block graphs. Non-linker singletons use element predictions with zero attached-hydrogen count. Coordination between building blocks is determined by geometry, rather than building-block-graph bond outputs.

Graph identities are assigned to codec training if they occur in an original training MOF; remaining identities are assigned to validation and then test. Linker identity counts are 200,927 / 19,328 / 18,683. Auxiliary singleton repetition adds 10,032 training records. The element vocabulary is $\{1, 6, 7, 8, 9, 15, 16, 17, 23, 24, 28, 29, 30, 35, 53, 56\}$. The full corpus reaches 2,368 explicit atoms, 9 unique building blocks, 517 retained nodes before replication, and a multiplicity of 96; coordinate training uses an input limit of 768 atoms.

The composition latent-flow model conditions on retained-node counts, multiplicities, and roles. The separately fitted metadata prior samples these quantities with an atom-count estimate, using a feasibility grammar for requested counts. The retained-node total is $N_{\mathrm{ret}} = \sum_{b=1}^{B} m_b n_b$, while N counts expanded physical atoms after hydrogen decoding. The sampled estimate does not constrain decoded hydrogens exactly. Padding and cross-MOF attention are masked; unsuccessful attempts remain in evaluation denominators.

B Model Architecture

FlexMOF maps building-block graphs and an evolving periodic state to atomic and cell-corner noise, and learns to generate those graphs through a graph autoencoder and a composition latent-flow model. The composition encoder is the attention backbone shared by graph encoding and coordinate denoising. The graph encoder comprises graph-specific input embeddings, this backbone, and posterior heads; the graph autoencoder additionally includes the graph decoder. The composition latent-flow model has its own Transformer parameters. We describe these interfaces below; the posterior and reconstruction objective are defined in Appendix C, and the coordinate state and periodic aggregation are derived in Appendix D.

B.1 Shared composition encoder and graph autoencoder

For a building-block graph with n_b retained nodes, element and attached-hydrogen-count embeddings are added to a degree projection and Laplacian spectral features. Degree includes the attached hydrogens. A two-layer network embeds each retained eigenvector/eigenvalue pair; the embeddings are summed over available frequencies. The graph encoding call supplies learned binary-bond attention biases to the composition encoder and masks padding. It uses graph inputs without coordinate or periodic features.

The composition encoder alternates pre-normalized multi-head self-attention and a two-layer ReLU feed-forward network, each with a residual connection. After graph-path layer normalization, a linear projection produces the mean and log variance of the node posterior in Equation 11. The graph decoder maps each node latent to element and attached-hydrogen-count logits. Its binary-bond head forms bilinear pair scores, symmetrizes them, and combines each pair score with contexts obtained by masked row-wise averaging at its two endpoints before classifying each unordered node pair. Self edges are excluded. Element, bond, and hydrogen-count argmax decisions recover a building-block graph; expanding attached hydrogens and replicating blocks yields the full atom-level composition.

The coordinate path calls the same composition encoder on physical atoms and eight cell-corner virtual tokens. Its input starts with element/type embeddings plus a common learned token embedding. Generated building-block graphs are re-encoded: their normalized posterior means are projected to the token width and added to the corresponding physical-atom embeddings. Replicated atoms share the latent of their retained source node, attached hydrogens inherit their parent node's

latent, and cell-corner virtual tokens receive zero latent injection. Thus the coordinate input reflects the decoded graph. Initial pair features are elementwise products of projected token features; each encoder layer adds analogous products from its hidden states. Pair features are normalized after the encoder stack. The encoder returns scalar node features and pair features for the spatial network.

B.2 Composition latent-flow model and metadata interface

The composition latent-flow model receives noisy normalized node latents, latent time, and building-block metadata. A linear map embeds the concatenation of the noisy latent and the self-conditioning prediction. A sinusoidal time embedding, log-transformed node counts and multiplicities, and a role embedding condition local diffusion-Transformer layers through shifts, scales, and residual gates. Attention is local to each building-block graph; block-summary exchange supplies information about the other blocks in the same MOF. Layer normalization and a linear readout predict the clean normalized latent $\widehat{\mathbf{Z}}_{b,0}$. This clean-latent prediction parameterizes the flow field used by the sampler in Appendix E.2; it is the prediction in Equation 6.

Exchange between building blocks. The composition latent-flow model concatenates noisy latents with a detached previous clean prediction, or zeros without self-conditioning. Every third local layer pools node features, exchanges building-block summaries within each MOF, and broadcasts a projected update:

$$\bar{\mathbf{f}}_b = \frac{1}{n_b}\sum_i \mathbf{f}_{bi}, \qquad \mathbf{f}_{bi} \leftarrow \mathbf{f}_{bi} + \mathbf{W}\, F_{\text{exchange}}(\{\bar{\mathbf{f}}_{b'}\}_{b'=1}^{B}; t, \mathcal{H})_b. \tag{8}$$

Here $\mathbf{f}_{bi}$ are node features, $\bar{\mathbf{f}}_b$ is their block mean, F_{exchange} is a Transformer over block summaries, and $\mathbf{W}$ is a learned projection.

The exchange projection starts at zero. The global condition is $\log(1+B), \log(1+N_{\text{ret}}), \log(1+\sum_b m_b)$; each local condition includes $\log(1+n_b), \log(1+m_b)$ and a role embedding. Shared attention, masked mean pooling, and broadcasting preserve equivariance under atom permutations and consistent building-block reindexing for fixed latent inputs.

Before graph generation, a separately trained autoregressive metadata prior supplies the requests. It embeds the preceding categorical value, the current metadata field, and a normalized feasibility state, then uses a two-layer gated recurrent unit (GRU) to predict the next value. Feasibility masks restrict the categorical distribution. The sequence specifies an atom-count estimate, the number of blocks, and each block's role, multiplicity, explicit size, and retained-node count. The latent-flow model receives retained-node counts, multiplicities, and roles. Hydrogen counts are decoded from the generated latents, so the coordinate-stage atom count N is recomputed after graph expansion rather than fixed exactly by the metadata estimate.

B.3 Periodic coordinate denoiser

Spatial features and vector readout. The spatial network receives the composition encoder's node and pair features together with the current and reference geometries. Element and degree features, graph shortest-path and multi-hop bond biases, disconnected-pair biases, and radial distances provide its inputs. Projected pair features contribute a headwise attention bias. At each spatial layer, the diffusion-time embedding is added to the spatial stream, while the composition encoder's node features produce adaptive layer-normalization shifts, scales, and residual gates. Periodic displacements are

$$\mathbf{r}_{ij\mathbf{n},\tau} = \mathbf{x}_{i,\tau} - (\mathbf{x}_{j,\tau} + \mathbf{n}\mathbf{L}_\tau), \qquad \mathbf{n} \in \mathbb{Z}^3. \tag{9}$$

Periodic images reuse their central atom's chemical identity while carrying lattice-translated coordinates. The neighborhoods are rebuilt as the current cell and positions change. For the evaluated structure configuration, image selection uses a system-wide token budget and the scalar attention is aggregated as in Section 3.2.

The scalar/vector decoder initializes channels from radial features and displacement directions over central and retained image neighbors. Its layers exchange scalar and vector information through within-channel and cross-channel attention, followed by scalar/vector feed-forward updates. Two

gated equivariant readout blocks reduce the channel width and produce one Cartesian three-vector per atom or cell-corner virtual token. Its scalar time gate blends that output with the displacement from the reference geometry:

$$\widehat{\boldsymbol{\eta}}_\theta = g_\tau(\mathbf{Y}_\tau - \mathbf{Y}_{\mathrm{ref}}) + (1 - g_\tau)\mathbf{v}_\theta, \quad g_\tau = \mathrm{sigmoid}\left(h_\theta(\tau) + \mathrm{logit}\,\sqrt{1 - \bar{\alpha}_\tau}\right). \tag{10}$$

Equation 10 is written for the physical atoms and four independent cell-corner virtual tokens; $\mathbf{v}_\theta$ denotes the corresponding readout rows, and h_θ is a learned scalar function of diffusion time. The scalar/vector decoder is rotationally equivariant with respect to its geometric inputs. The full denoiser additionally conditions on the fixed reference cell.

B.4 Network configuration

Table 4 collects network dimensions together with the existing coordinate-noise and sampling settings. The composition latent-flow Transformer has 12 local layers of width 768, eight attention heads, and a feed-forward width of 3,072; the shared composition encoder and spatial Transformer use 32 heads and feed-forward width 768. The graph posterior has 32 channels per retained node, and the graph decoder's bond classifier has hidden width 24. The metadata prior has two GRU layers of hidden width 128, with previous-value and field embeddings of widths 64 and 16. Optimization and latent normalization are specified in Appendix E.

Table 4: Joint model periodic-coordinate architecture, training noise schedule, and sampling settings.

Setting	Value
Composition encoder, spatial, decoder layers	14 / 12 / 12
Hidden width / heads / pair width	768 / 32 / 64
Dropout / token-attention dropout	0.1 / 0.0
Periodic distance cutoff	40 Å
Image-budget parameter / cell-range parameter	896 / 5
Radial kernels / graph path length	32 / 5
Noise schedule steps	5,000
Sigmoid beta endpoints	10^{-7} / 0.002
Reference edge coefficient s	2.859496852322873
Atomic noise coefficient a	1.0531306506190654
Independent-corner noise scale σ_c	0.5
Sampling stride / recycling	-100 / 0
Coordinate-network evaluations	50

C Composition Representation and Generation

For each graph, $\mathbf{D}_b$ is the diagonal degree matrix for adjacency matrix $\mathbf{A}_b$. We form the combinatorial Laplacian $\mathbf{D}_b - \mathbf{A}_b$ and retain up to 32 of the lowest eigenpairs, including zero modes. Missing frequencies are masked. Element, degree, hydrogen-count, and spectral embeddings enter the graph encoder. Training randomizes eigenvector signs (Dwivedi et al., 2023) and uses amplitude modulation of scale 0.005. The per-node posterior is

$$q_\psi(\mathbf{U}_b \mid G_b) = \prod_{i=1}^{n_b} \mathcal{N}(\mathbf{u}_{bi}; \boldsymbol{\mu}_{bi}, \mathrm{diag}(\boldsymbol{\sigma}_{bi}^2)). \tag{11}$$

The graph decoder is described in Appendix B.1. Its loss is

$$\mathcal{L}_{\mathrm{AE}} = \mathcal{L}_{\text{atom-CE}} + \mathcal{L}_{\text{bond-CE}} + \mathcal{L}_{\text{H-CE}} + \lambda_{\mathrm{KL}} D_{\mathrm{KL}}\big(q_\psi(\mathbf{U} \mid G) \| \mathcal{N}(0, I)\big). \tag{12}$$

Node losses average over valid nodes, bond loss averages over unordered valid pairs, and the KL term sums over latent channels before averaging over nodes. Singleton bond loss is zero. The KL weight is 10^{-6}. Posterior means are normalized channel-wise for latent flow matching and restored before decoding:

$$\mathbf{Z}_{b,0} = (\boldsymbol{\mu}_b - \mathbf{c}) \oslash \mathbf{s}, \qquad \widehat{\mathbf{U}}_b = \mathbf{c} + \mathbf{s} \odot \widehat{\mathbf{Z}}_{b,0}. \tag{13}$$

Here $\mathbf{c}, \mathbf{s} \in \mathbb{R}^d$ are running channel-wise means and positive scales, broadcast over nodes; $\oslash$ and $\odot$ denote elementwise division and multiplication.

C.1 Reconstruction diagnostics

Graph reconstruction is evaluated on held-out identities. On 18,683 test linker identities, posterior means exactly reconstruct 18,554 graphs, with all 639,115 element labels and hydrogen counts correct. The remaining errors concern bonds. Full-validation exact reconstruction is 19,205/19,328 (99.364%).

Table 5: Hydrogen-count codec reconstruction on 18,683 test graphs. Accuracies and F1 are percentages.

Latent	Exact count ↑	Exact (%) ↑	Atom (%) ↑	Bond F1 (%) ↑	H count (%) ↑
Posterior mean	18,554	99.310	100.000	99.986	100.000
Posterior sample	18,503	99.037	100.000	99.982	100.000

D Periodic Geometry and the Structure Denoiser

The structure denoiser combines a cell-corner representation with chemical and periodic geometric features. We give the corner construction, noise distribution, and periodic-aggregation derivation below; network details are in Appendix B.

D.1 Cell-corner representation and correlated noise

Four points $\mathbf{u}_0, \mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ determine the cell:

$$\mathbf{L} = \begin{bmatrix} \mathbf{u}_1 - \mathbf{u}_0 \\ \mathbf{u}_2 - \mathbf{u}_0 \\ \mathbf{u}_3 - \mathbf{u}_0 \end{bmatrix}, \qquad \mathbf{u}_{\boldsymbol{\beta}} = \mathbf{u}_0 + \sum_{k=1}^{3} \beta_k (\mathbf{u}_k - \mathbf{u}_0), \quad \boldsymbol{\beta} \in \{0,1\}^3. \tag{14}$$

The eight tokens follow the order $(000, 001, 010, 011, 100, 101, 110, 111)$, with independent indices $(0, 4, 2, 1)$. The remaining corners are reconstructed after each update. Centering uses the midpoint of opposite corners and translates atoms consistently. For a reference-cell edge of $sN^{1/3}$, the atoms start at the origin and the corners form a centered cube.

Atomic noise has a standard deviation of $aN^{1/3}$. To construct cell noise, draw independent $\boldsymbol{\delta}_k \sim \mathcal{N}(0, \sigma_c^2 I)$ for $k = 0, 1, 2, 3$, form the affine corners, and subtract their opposite-corner midpoint:

$$\widetilde{\boldsymbol{\eta}}_{\boldsymbol{\beta}} = \boldsymbol{\delta}_0 + \sum_{k=1}^{3} \beta_k (\boldsymbol{\delta}_k - \boldsymbol{\delta}_0), \qquad \boldsymbol{\eta}_{\boldsymbol{\beta}} = \widetilde{\boldsymbol{\eta}}_{\boldsymbol{\beta}} - \frac{\widetilde{\boldsymbol{\eta}}_{000} + \widetilde{\boldsymbol{\eta}}_{111}}{2}. \tag{15}$$

This construction gives correlated displacements that preserve the parallelepiped geometry. Equation 2 evaluates the atoms and independent corners, with denominator $3(N + 4)$.

D.2 Periodic attention aggregation and storage

This section derives Equation 3 and describes the computational path used by the structure model. Let $s_{ij} = \mathbf{q}_i^\top \mathbf{k}_j / \sqrt{d_h}$. For images that reuse source keys and values, expanded scalar attention is

$$\mathbf{o}_i^{\text{expanded}} = \frac{\sum_j \sum_{\mathbf{n} \in \mathcal{I}_j} \exp(s_{ij} + b_{ij\mathbf{n}}) \mathbf{v}_j}{\sum_k \sum_{\mathbf{m} \in \mathcal{I}_k} \exp(s_{ik} + b_{ik\mathbf{m}})} = \frac{\sum_j \exp(s_{ij}) w_{ij} \mathbf{v}_j}{\sum_k \exp(s_{ik}) w_{ik}}. \tag{16}$$

The normalization shared by the geometric weights cancels. Thus the aggregation preserves the pre-dropout scalar attention for a fixed retained image set, up to floating-point rounding and the numerical weight floor. The implementation uses a floor of 10^{-10} before taking logarithms, with zero weights masked. Atom pairs retain the multiplicity and geometric weights of their images. Image-specific directions are evaluated separately when forming vector features. The composition encoder uses the same source-atom aggregation with uniform valid-image weights; the spatial encoder includes the learned chemical and geometric biases. Here source-index denotes the index mapping used to implement source-atom aggregation.

At each denoiser call, candidate translations are drawn from the configured integer range and restricted using the current cell dimensions. The global image selector retains candidates within the cutoff of a valid central token of a different type. If too many candidates remain, it applies a threshold on their minimum distance to the central tokens. The retained-image allowance is $\max(896 - N, 0)$, where N counts physical atoms. The eight central-cell corner tokens are additional to this count. The cutoff is 40 Å and the integer range is $[-5, 5]^3$. These settings govern the evaluated structure path.

Let $\mathcal{T}_\tau \subseteq [-5, 5]^3 \cap \mathbb{Z}^3$ contain candidate translations after screening by the current cell dimensions. Write $\mathbf{p}_{i,\tau}$ for the position of a valid central token (atom or corner), and a_i for its token type. The cutoff-eligible image tokens are

$$\begin{aligned} \mathcal{E}_\tau = \big\{ (j, \mathbf{n}) : \mathbf{n} \in \mathcal{T}_\tau \setminus \{\mathbf{0}\}, \\ \min_{i:\, a_i \neq a_j} \|\mathbf{p}_{i,\tau} - (\mathbf{p}_{j,\tau} + \mathbf{n}\mathbf{L}_\tau)\|_2 < r_c \big\}, \qquad r_c = 40\,\text{Å}. \end{aligned} \tag{17}$$

The budget threshold uses each candidate's minimum distance to any valid central token, retaining at most $\max(896 - N, 0)$ image tokens in $\widehat{\mathcal{E}}_\tau$. The represented cells and their count are

$$\mathcal{M}_\tau = \{\mathbf{0}\} \cup \{\mathbf{n} : \exists j,\ (j, \mathbf{n}) \in \widehat{\mathcal{E}}_\tau\}, \qquad M_\tau = |\mathcal{M}_\tau|. \tag{18}$$

Thus M in Figure 2 depends on the evolving geometry; a retained cell need not contribute every token. The cutoff is measured between token positions, not cell centers.

Let R be the number of candidate translations and Q the distance chunk size. A dense candidate-distance array has $O(RT^2)$ entries. Chunking bounds that temporary array by $O(TQ)$, while candidate positions, membership flags, and minimum distances require $O(RT)$ storage. For T_{ext} retained central-plus-image tokens, image-specific geometry costs $O(TT_{\text{ext}})$ entries. Aggregation produces $O(T^2)$ scalar weights, which are reused across spatial layers within a denoiser call. Pair representations and the central-cell attention remain quadratic in T. The distance chunk size is 4,096. These storage estimates describe the intermediate arrays used to construct periodic neighborhoods.

E Parameterization, Training, and Sampling

FlexMOF trains the graph encoder, composition latent-flow model, and periodic coordinate denoiser in one joint network. The graph encoder and coordinate pathway share the composition encoder. Graph reconstruction updates this trunk and the graph posterior; coordinate loss reaches both through direct trunk use and undetached latent injection into atom features. The latent-flow target is detached from the graph posterior. An autoregressive prior for atom-count and building-block metadata is trained on metadata likelihood and supplies requests to the joint network at sampling time. Sequential sampling of graphs and then coordinates implements the joint generative distribution. The reported joint-model checkpoint is E381 after 10,000 fine-tuning updates; its architecture is detailed in Appendix B.

E.1 JOINT NETWORK AND OPTIMIZATION

Architecture dimensions are collected in Appendix B; the graph reconstruction KL weight is 10^{-6}. Online latent normalization tracks posterior means during joint training, with momentum 0.99 and a 20,000-node warmup before applying the latent-flow loss.

The composition latent-flow model's time floor is 0.1, self-conditioning probability is 0.5, and the time logit follows $\mathcal{N}(1, 1)$. The training objective adds coordinate denoising, graph reconstruction, and latent flow with unit weights for the two composition losses. Generated graphs are decoded before coordinate denoising; the coordinate pathway never receives a noisy graph.

The joint network configuration uses AdamW with $\beta_1 = 0.9$ and $\beta_2 = 0.999$, learning rate 10^{-4}, 262,500 warmup updates, zero weight decay, clipping at 1.0, FP32, and seed 12347. Its configured microbatch size is 2 per device across eight devices, with eight accumulation steps. Antithetic coordinate times have uniform marginals. Table 4 gives the periodic coordinate architecture and noise settings.

E.2 SAMPLING DETAILS

Composition generation uses 100 second-order Heun intervals (Karras et al., 2022), integrating from $t = 1$ toward $t = 0$. The final step returns the clean prediction directly, avoiding division by zero. For $t > 0$, the clean-prediction velocity is $\widehat{\mathbf{v}}_{b,t} = (\mathbf{Z}_{b,t} - \widehat{\mathbf{Z}}_{b,0})/t$. Initial and refresh noise use independent streams. For $0.1 \le t \le 0.8$, set $t' = \min(t(1 + 8/100), 0.99)$ and refresh by

$$\mathbf{Z}_{t'} = r\mathbf{Z}_t + \sqrt{t'^2 - r^2 t^2}\,\boldsymbol{\epsilon}', \qquad r = \frac{1 - t'}{1 - t}, \quad \boldsymbol{\epsilon}' \sim \mathcal{N}(0, I). \tag{19}$$

The update preserves the conditional Gaussian marginal when clean targets and noise are independent. Element, symmetrized bond, and hydrogen-count argmax decisions define the expanded chemical graph. The coordinate sampler initializes atoms and corners from the count-dependent reference plus noise, uses 50 second-order DPM-Solver++ evaluations, rebuilds periodic neighborhoods on each call, and reconstructs and centers the cell after every update. Coordinates remain unwrapped during denoising and are wrapped for export. All attempts are recorded without replacement of failed samples.

F EVALUATION PROTOCOLS AND ADDITIONAL RESULTS

F.1 STRUCTURE MATCHING AND COMPARISON POPULATIONS

The official MOFChecker-filtered BW test set aligned with the MOFFlow-2 filtered test set contains 19,792 references; FlexMOF matches 9,165 at MR@1 (46.31%) and 13,256 at MR@5 (66.98%). Published baselines retain their reported evaluation populations and supplied-geometry settings. Matching tolerances are 0.3 for relative lattice length, 0.5 for normalized site distance, and 10° for angles. The comparison disables lattice rescaling and enables supercell attempts.

The size-stratified panel (c) of Figure 3 uses a shared cohort of 19,227 structures; panels (a,b) report the official filtered BW-DB cohort of 19,792 references alongside separately identified published baselines. Pore evaluation uses the same 19,227 shared conditions, and the PBC ablation uses 30,880 test structures. The updated overall-RMSD panel uses 4,035 successful matches per method on method-specific matched sets; linker RMSD is measured after local alignment in Å, while overall RMSD is dimensionless after normalization by $(V/N)^{1/3}$.

F.2 PAIRED PBC ABLATION

The periodic-interaction ablation compares PBC-on and PBC-off coordinate pathways under matched data, optimization, and sampling settings. Neighborhoods use a 40 Å cutoff and image allowance $\max(896 - N, 0)$, excluding the eight central-cell corners from N. For MR@5, five independently seeded candidates (12347–12351) are generated per reference, and a reference succeeds if any candidate matches. Atom count and cross-cell-boundary adjacency fraction are measured on

the reference. The high-boundary group starts at the reference upper-quartile threshold 0.1582; its complement forms the lower-boundary group. The $N \geq 400$ and high-boundary strata overlap. The MR@1 stratum summary is reported in Table 7.

Table 6: PBC-on minus PBC-off MR@5 gains in broader, overlapping strata of the paired ablation. Entries give the paired bootstrap 95% interval in percentage points.

Structure stratum	n	Gain [95% interval] (pp) ↑
$N \geq 400$ atoms	1,463	+4.78 [3.28, 6.29]
Highest boundary-fraction quartile	7,724	+2.63 [1.73, 3.52]
Lower-boundary complement	23,156	+1.16 [0.66, 1.65]

Table 7: Separately reported PBC-on minus PBC-off MR@1 gains by structure stratum; these values are distinct from the paired MR@5 run in the main-text PBC heatmap.

Structure stratum	ΔMR@1 (pp) ↑
$N \geq 400$ atoms	+2.72
High cross-cell-boundary adjacency	+2.00
Other structures	+1.02

Table 8: Matched outcomes for the five-candidate MR@5 PBC ablation. Figure 4 shows the fine-grained gain bins.

Structure stratum	n	PBC on ↑	PBC off ↑
All structures	30,880	16,453	15,982
$N \geq 400$ atoms	1,463	221	151
High boundary fraction	7,724	3,807	3,604
Lower boundary fraction	23,156	12,646	12,378

F.3 Molecular-graph topology ablation

The topology-bias ablation uses the MOFFlow-2 MetalOxo final-test cohort of 19,792 structures without an atom-count cutoff. The reported checkpoint is at global step 65,382, with minimum validation loss 1.0115296841. It produces 8,119/19,792 matches at MR@1 (41.0216%) and 12,321/19,792 at MR@5 (62.2524%); relative to the supplied rounded same-cohort joint reference (46.3% and 67.0%), the differences are -5.2784 and -4.7476 percentage points. These differences use the rounded reports rather than subtraction of the main result counts.

F.4 Pore calculation, coverage, and uncertainty

We compute volumetric and gravimetric surface area (VSA, GSA), accessible void fraction (VF), pore-limiting diameter (PLD), largest-cavity diameter (LCD), and density. Zeo++ 0.4.7 (Willems et al., 2012) uses a 1.2755 Å probe/channel radius, 10,000 surface samples per atom, and 100,000 volume samples. The fixed atomic radius table includes a default 2 Å radius for some metals. For this MOF benchmark, hydrogens and guests are retained as stored. Density is calculated from explicit atomic masses and cell volume.

For each property, we use all candidate–reference pairs that are calculable for all three methods, including matched and unmatched structures. Zero values are retained; calculation failures are recorded as missing and included in the coverage report. Pointwise confidence intervals use 2,000 paired condition-bootstrap resamples without adjustment for multiple comparisons.

For property p, define the shared calculable index set I_p and the paired absolute-error difference against a baseline as

$$\Delta_p = \frac{1}{|I_p|} \sum_{i \in I_p} \left(|\hat{p}_i^{\text{ours}} - p_i| - |\hat{p}_i^{\text{base}} - p_i| \right). \tag{20}$$

Negative values favor FlexMOF. Resampling condition indices jointly keeps candidates, references, and baselines paired. Table 9 reports pointwise 95% intervals for all properties and both baselines. The intervals favor FlexMOF for the five pore descriptors and AtomMOF-L for density. Table 10 reports the calculable pairs.

Table 9: Paired MAE difference (FlexMOF minus baseline) and pointwise 95% bootstrap interval. Units are property-specific; negative differences favor FlexMOF.

Property	Versus MOFFlow-2 ↓	Versus AtomMOF-L ↓
VSA (m^2/cm^3)	-111.24 [$-114.57, -108.14$]	-23.03 [$-25.37, -20.81$]
GSA (m^2/g)	-183.64 [$-188.28, -179.06$]	-29.95 [$-32.47, -27.59$]
VF	-0.023650 [$-0.024109, -0.023164$]	-0.003424 [$-0.003669, -0.003187$]
PLD (Å)	-0.4334 [$-0.4477, -0.4193$]	-0.0749 [$-0.0845, -0.0645$]
LCD (Å)	-0.5312 [$-0.5480, -0.5145$]	-0.0801 [$-0.0903, -0.0696$]
Density (g/cm^3)	-0.013494 [$-0.013816, -0.013157$]	0.001898 [$0.001660, 0.002117$]

Table 10: Calculable candidate–reference pairs out of 19,227 conditions. The final column is the three-method intersection used for paired accuracy.

Property	MOFFlow-2 ↑	AtomMOF-L ↑	FlexMOF ↑	Shared
VSA (m^2/cm^3)	19,211	19,041	19,202	19,018
GSA (m^2/g)	19,211	19,041	19,202	19,018
VF	19,225	19,052	19,217	19,044
PLD (Å)	19,218	19,045	19,208	19,029
LCD (Å)	19,218	19,045	19,208	19,029
Density (g/cm^3)	19,225	19,052	19,217	19,044

Using the shared calculable pairs holds the comparison set fixed for each property. The coverage table reports how many conditions contribute to each comparison.

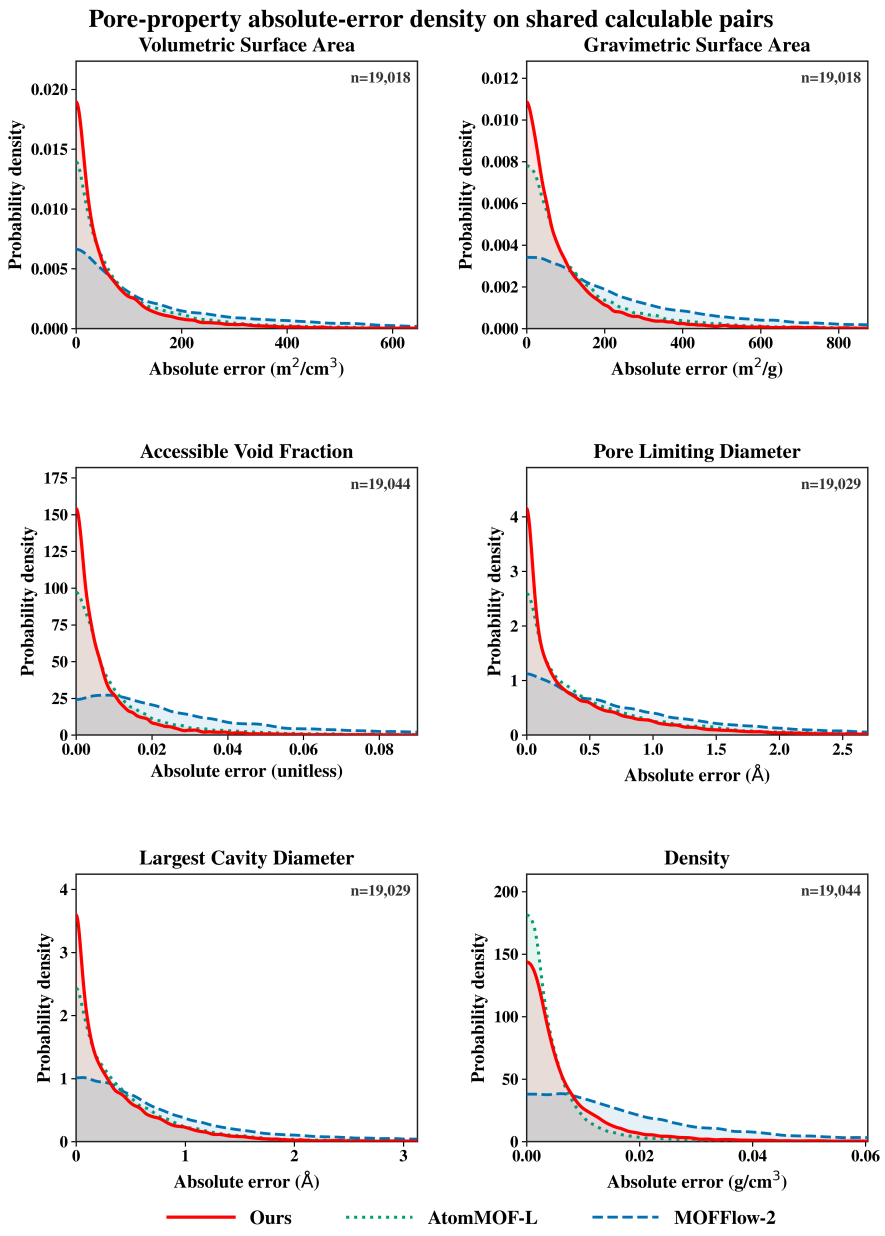


Figure 5: Distributions of per-structure absolute errors in six pore properties. Curves show probability-density estimates on property-specific shared calculable pairs within the 19,227-condition pool, using one candidate per condition ($K = 1$) without structure-match filtering. Panel annotations give the shared sample counts. Solid, dotted, and dashed curves denote FlexMOF ("Ours"), AtomMOF-L, and MOFFlow-2, respectively. Horizontal axes show absolute errors; vertical axes show probability density. Table 1 reports the means. Panels have different scales and finite displayed ranges.

F.5 Complete-generation criteria and literature protocols

All FlexMOF metrics use the same 10,000-draw denominator, including failed attempts. Validity requires all 15 MOFChecker 2.0 checks. Uniqueness uses singleton MOFids: a framework identifier must occur exactly once among the 10,000 outputs. A MOFid was obtained for every FlexMOF output, so U equals EU and V equals EV, where E denotes successful identifier extraction. Novelty compares normalized structural keys against 213,583 of 247,062 BW training records; 33,479 records lack the paired RDKit data needed by this index. Strict VNU is the intersection of validity, this novelty criterion, and singleton uniqueness.

Mofasa values in Table 3 are literature results (Šimkus et al., 2025). Its validity values come from Table 1: 10,000 samples for BW, approximately 202,000 for each QMOF variant, 250,000 for Exp., and 225,000 for Exp.-opt. Its U and VNU values are the unrescaled EU and EVNU entries at 10,000 samples in Table 6. Samples without a MOFid fail EU and EVNU. Standalone N is unreported; ENU combines novelty and uniqueness and is not substituted for N. Mofasa tests novelty by training-set MOFid matching, whereas FlexMOF uses the indexed structural keys described above.

QMOF provides MOF structures with quantum-chemical calculations (Rosen et al., 2021; 2022). The Exp. collection combines QMOF, computation-ready CoRE-MOF-2024, and ARC-MOF DB12/DB14, as defined by Mofasa. The suffix opt denotes post-generation relaxation with the Orb interatomic potential and D3 dispersion correction. FlexMOF uses BW training data without post-generation force-field or interatomic-potential relaxation. Reported comparisons therefore retain differences in training collections, sample populations, and novelty criteria.

G Adsorption Configuration and Energy Protocols

G.1 Data, adaptation, and candidate budgets

Processed ODAC25 contains 153,249 training, 3,130 validation, and 3,582 test records. Adaptation of the periodic coordinate pathway uses all training records, AdamW, learning rate 3×10^{-5}, 500 warmup updates, zero weight decay, gradient clipping at 1, and FP32. Framework and guest chemistry and loading are specified by this transfer task; coverage sampling uses 50 network evaluations per candidate.

At $K = 10$, the candidate pool has 35,820 slots. Within system s, J_s reference records contribute $10J_s$ candidates. Complete-complex StructureMatcher uses relative lattice tolerance 0.3, normalized site tolerance 0.5, angle tolerance 10°, and lattice rescaling.

G.2 System-macro reference coverage

$$\text{COV-R} = \frac{100}{n_{\text{sys}}} \sum_{s=1}^{n_{\text{sys}}} \frac{1}{J_s} \sum_{j=1}^{J_s} \mathbb{I}\left[\exists i \le KJ_s : \text{match}(R_{sj}, P_{si})\right], \qquad n_{\text{sys}} = 683. \tag{21}$$

All 683 systems receive equal weight, including systems with no matches; missing or unusable candidate slots do not cover a reference. At $K = 10$, 1,030 of 3,582 references are covered (28.75% micro average), and system-macro reference coverage (COV-R) is 28.75% (Table 2). Reproduced AtomMOF-L system-macro reference coverage (COV-R) uses the same grouping and matching rule at $K = 1$ and $K = 10$.

G.3 Adsorption-energy evaluation

Energy evaluation uses $K = 10$ and relaxes candidates with the cell fixed using the Fast Inertial Relaxation Engine (FIRE) (Bitzek et al., 2006) in the Atomic Simulation Environment (ASE) (Larsen et al., 2017), using the eSEN-ODAC25-Full interatomic potential (Fu et al., 2025; Sriram et al., 2025). Adsorption energies are evaluated after relaxation against stored reference geometries.

For a converged candidate, adsorption energy is the complete-complex energy minus the energies of the separated host and guest subsets. The subsets retain their coordinates, cell, and original atom assignments; they are not separately relaxed. Reference adsorption energies are recomputed with

the same eSEN model at the stored reference geometries. Let $\mathcal{C}_s$ denote the converged candidates for system s:

$$\Delta E_{\text{ads}} = E_{\text{complex}} - E_{\text{host}} - E_{\text{guest}}, \qquad \varepsilon_{\min,s} = \min_{i \in \mathcal{C}_s} \Delta E_{\text{ads}}(P_{si}) - \min_{j \le J_s} \Delta E_{\text{ads}}(R_{sj}), \quad (22)$$

where P_{si} and R_{sj} are candidate and reference configurations. We set $\varepsilon_{\min,s} = +\infty$ for $\mathcal{C}_s = \varnothing$ when computing success rates; these systems are excluded from the finite-energy mean. Energy success requires a converged candidate and $\varepsilon_{\min,s} \le 0.1\,\text{eV}$; the below-reference rate uses $\varepsilon_{\min,s} < 0$. Both rates use all 683 systems, including systems with no converged candidate. The mean $\varepsilon_{\min}$ is computed over the 459 systems with at least one converged candidate. Table 2 reports these results alongside the published FK-steered AtomMOF-L evaluation at $K = 10$ (Kim et al., 2026).

H Supplementary Visualizations

The workflow overview, conditional structure examples, and complete-generation examples supplement the main quantitative comparisons.

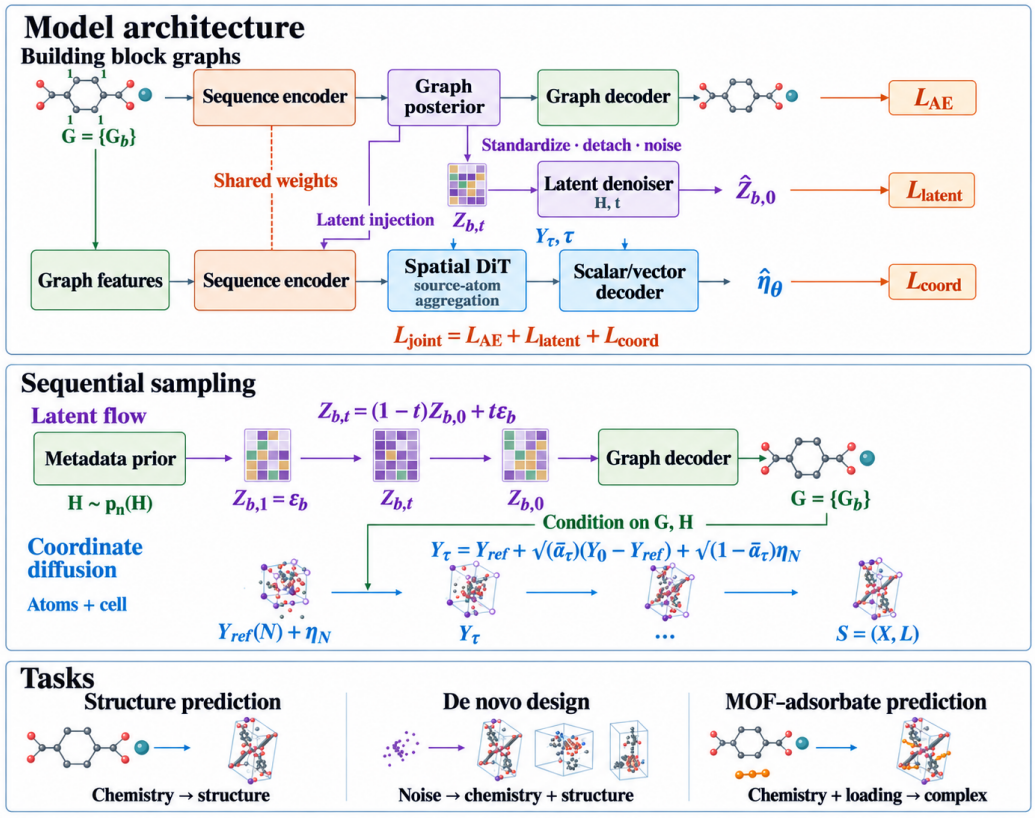


Figure 6: Overview of the FlexMOF architecture, sequential sampling procedure, and supported tasks.

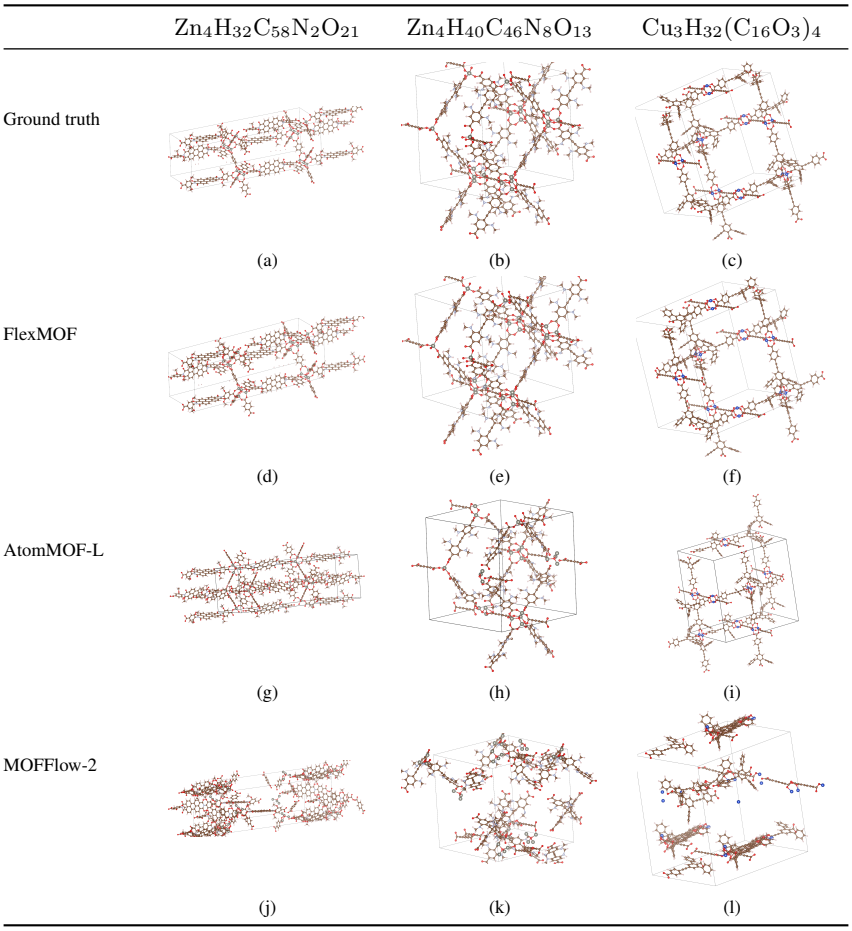
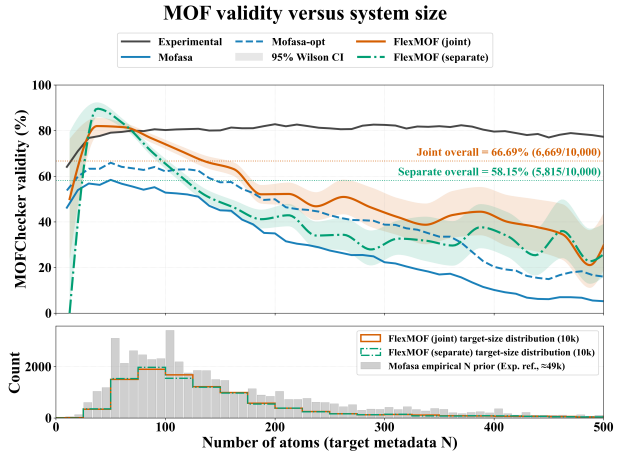
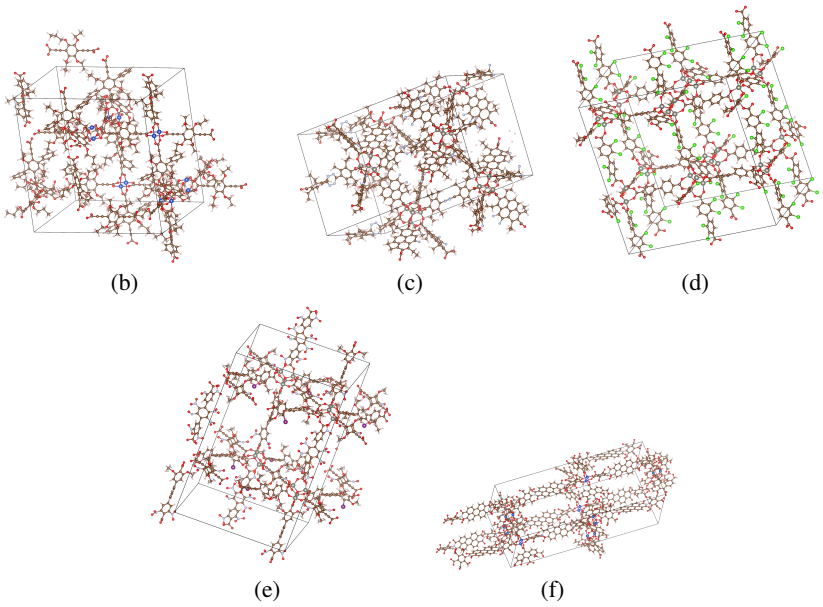


Figure 7: Crystal structure prediction examples. Columns share a composition; rows show references, FlexMOF, AtomMOF-L, and MOFFlow-2.



(a) MOFChecker validity by target metadata atom count



(b) (c) (d)

(e) (f)

Figure 8: Complete generation without post-generation relaxation. (a) Validity by requested atom count, with all draws in bin denominators and 95% Wilson intervals. Background curves and the size histogram use different data and sampling conditions. (b)–(f) Five valid generated structures illustrate variation in cell shape, size, and composition.