MADFIELD: MULTI-FIDELITY AMORTIZED DENSITY FIELD FOR ADSORPTION IN NANOPOROUS MATERIALS

Anonymous authors
Paper under double-blind review

ABSTRACT

High-throughput computational screening of nanoporous materials for gas storage and separation requires fast and accurate characterization of adsorption equilibrium. Particle-based grand canonical Monte Carlo (GCMC) and density-based classical density functional theory (cDFT) provide simulation-based estimates of gas uptake and adsorbate density fields, but their speed–accuracy tradeoff remains insufficient for large-scale screening. In this work, we address this gap with Multi-fidelity Amortized Density Field (MADField), which reframes adsorption prediction as equilibrium density-field estimation. MADField learns from two complementary fidelities, combining broad and scalable cDFT density supervision with GCMC density labels, and recovers gas uptake by integrating the predicted density field. MADField improves uptake accuracy over the strongest baselines by $5.7\times$ for cDFT and $3.35\times$ for GCMC, and its predicted fields accelerate cDFT solvers with a $5.7\times$ reduction in walltime while recovering 37.78% of cases that fail under standard settings. We also evaluate MADField under realistic screening scenarios for CH_4 working-capacity and low-pressure CO_2 uptake screening, where it reduces the number of required GCMC candidate evaluations to obtain the top 0.1% pool by 8.0 and 11.5 times, respectively.

1 INTRODUCTION

High-throughput computational screening is central to discovering nanoporous materials for gas storage and separation (Wilmer et al., 2012; Lin et al., 2012; Mason et al., 2014; Charalambous et al., 2024). Among these materials, metal–organic frameworks (MOFs) have attracted particular attention because their tunable pore geometries and chemical environments offer a vast design space for controlling adsorption. Screening this design space requires predicting gas uptake, the amount adsorbed at a given pressure and temperature. This, in turn, requires characterizing adsorption equilibrium: how gas molecules distribute within a material's pores while interacting with the pore surfaces and with one another.

However, the cost of evaluating adsorption equilibrium constrains the scope of materials screening. Grand canonical Monte Carlo (GCMC) provides a particle-based description of this equilibrium, sampling gas configurations through insertion, deletion, and roto-translational moves (Figure 1(a)) (Adams, 1975). Although GCMC has enabled large screens, the possible combinations of framework chemistry and pore structure extend far beyond what can be evaluated exhaustively (Liu et al., 2024; Majumdar et al., 2021). Bucior et al. (2019) used a surrogate to rank 54,776 candidates for hydrogen storage and reserved GCMC only on top-ranked structures, because simulating the full library was too expensive (Bucior et al., 2019). The possible MOF chemistries and pore structures extend well beyond existing libraries (Liu et al., 2024; Majumdar et al., 2021). Faster, reliable adsorption prediction could broaden screening within MOFs and across other nanoporous materials, including amorphous carbons and porous polymers (Thyagarajan & Sholl, 2020).

Classical density functional theory (cDFT) offers a lower-cost, density-based description. It replaces particle sampling with optimization of a grand-potential functional over the one-body adsorbate density field, from which uptake follows by integration (Figure 1(b)) (Evans, 1979; Roth, 2010). Recent work shows that cDFT can reproduce GCMC adsorption trends in MOFs at lower cost (Dufour-

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.

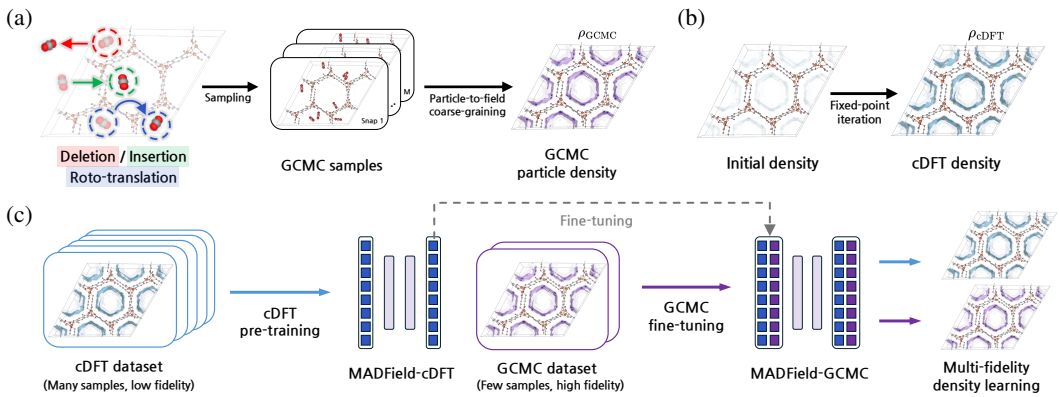


Figure 1: **Multi-fidelity adsorbate density learning from cDFT and GCMC. (a)** GCMC samples a particle ensemble through deletion, insertion, and roto-translation moves, and particle-to-field coarse-graining converts the ensemble into ρ_{GCMC}. **(b)** cDFT solves for ρ_{cDFT} by fixed-point iteration from an initial density. **(c)** MADField is pre-trained on broad cDFT labels and fine-tuned on sparse GCMC labels, combining cDFT coverage with GCMC fidelity.

Décieux et al., 2025; Thiele et al., 2026). Yet practical cDFT relies on approximate free-energy functionals, so its predicted densities and uptake can deviate from GCMC.

Contribution. In this work, we introduce Multi-fidelity Amortized Density Field (MADField) for adsorption simulation in nanoporous materials. At a high level, MADField predicts the equilibrium adsorbate density field and obtains scalar uptake by integration. Our key idea is to pre-train on abundant low-fidelity cDFT density labels and fine-tune on a smaller set of high-fidelity GCMC-derived labels. We use the predicted fields to initialize cDFT calculations and to rank materials for follow-up GCMC. We also examine whether a model trained on MOFs can extend adsorption prediction to other nanoporous materials, where class-specific labels may be scarce.

We develop MADField to predict unnormalized density in physical units, preserving spatial adsorption structure while recovering uptake by integration. Density provides a common prediction target for cDFT and GCMC: we pre-train on cDFT fields across nine adsorbates, then fine-tune on GCMC-derived fields for CH_4, Xe, and CO_2 (Figure 1(c)). We further evaluate the MOF-trained model on disordered nanoporous materials without retraining on those material classes.

We use predicted densities to initialize the original cDFT solver, including for calculations that failed with its standard initialization. Only solver-converged outputs become additional training labels. We also use the predicted uptakes to prioritize candidate materials for follow-up GCMC evaluation. To support further work on density-based adsorption models, we will release a dataset of 566,193 converged cDFT calculations.

Experiments show that density supervision improves uptake prediction over direct scalar regression in matched-backbone ablations, even when both models receive cDFT pre-training. On MOFs, MADField reduces uptake MAE by $5.7\times$ on cDFT labels and $3.35\times$ on GCMC labels relative to the strongest baselines. It also transfers zero-shot to disordered materials, reducing uptake MAE by $4.9\times$ on a cDFT amorphous-carbon benchmark and $3.1\times$ on a GCMC benchmark spanning porous polymers, kerogens, and amorphous carbons. Learned initialization recovers 37.8% of otherwise failed cDFT states and reduces median solver walltime by $5.7\times$ on jointly converged states. We evaluate MADField for CH_4 working-capacity and low-pressure CO_2 uptake screening on 254,198 ARC-MOF and 15,572 MOSAEC-DB frameworks, respectively. Compared with the best scalar baseline for each task, MADField reduces the number of GCMC candidate evaluations needed to recover all materials in the true top 0.1% by factors of approximately 8.0 and 11.5, respectively.

2 Background and Related Work

This section introduces adsorption equilibrium and its particle- and density-based formulations (Section 2.1), followed by related work on machine learning for adsorption prediction (Section 2.2).

2.1 ADSORPTION EQUILIBRIUM AS PARTICLE AND DENSITY FIELDS

Adsorption is the process by which guest gas molecules accumulate on the surfaces and within the pores of a host porous material. For a rigid host framework $\mathcal{M}$, guest species s, temperature T, and pressure P, the central field-level object is the equilibrium one-body number density $\rho_{\mathrm{eq}}(\mathbf{r})$, defined as the expected number of adsorbate molecule centers per unit volume at position $\mathbf{r}$ in the periodic unit cell $\mathcal{V}$. The expected molecule count N and corresponding *adsorption isotherm* $\mathcal{I}_T$ are obtained from this field:

$$N(P,T;\mathcal{M},s) = \int_{\mathcal{V}} \rho_{\mathrm{eq}}(\mathbf{r})\, d\mathbf{r}, \qquad \mathcal{I}_T(\mathcal{M},s) = \{N(P,T;\mathcal{M},s) : P \in \mathcal{P}\}. \tag{1}$$

Uptake follows by unit conversion (Section H) and is the primary quantity used to rank porous materials in high-throughput screening, since practical applications often require maximizing the amount of gas stored or captured at a target pressure and temperature. The adsorption isotherm further characterizes how uptake changes with pressure, enabling evaluation of process-relevant quantities such as working capacity between adsorption and desorption pressures, pressure-dependent selectivity, and operating-window robustness.

Particle view: grand canonical Monte Carlo. GCMC simulates adsorption by sampling many-body particle configurations in the grand canonical (μVT) ensemble (Adams, 1975). For adsorbate species s, a configuration of n rigid molecules with centers $\mathbf{r}_{1:n}$ and orientations $\boldsymbol{\omega}_{1:n}$ has energy

$$U_n(\mathbf{r}_{1:n},\boldsymbol{\omega}_{1:n};\mathcal{M},s) = \sum_{i=1}^{n} V_{\mathrm{ext}}^{\mathrm{GCMC}}(\mathbf{r}_i,\omega_i;\mathcal{M},s) + \sum_{i<j} u_{ss}(\mathbf{r}_i,\omega_i,\mathbf{r}_j,\omega_j), \tag{2}$$

where $V_{\mathrm{ext}}^{\mathrm{GCMC}}$ is the framework–fluid potential and u_{ss} the fluid–fluid pair interaction under the GCMC force field. Orientation matters for rigid multi-site CO_2 and drops out for spherical single-site adsorbates. At chemical potential $\mu = \mu(P,T;s)$, a Markov chain proposes insertion, deletion, translation, and rotation moves, accepted by a Metropolis criterion so that the chain samples the μVT distribution over variable-size configurations. Adsorption observables are then computed as ensemble averages, *i.e.*, uptake is the expected particle count, and the one-body density field is obtained by coarse-graining the particles:

$$N(P,T;\mathcal{M},s) = \langle n \rangle_{\mu,V,T}, \qquad \rho_{\mathrm{eq}}(\mathbf{r}) = \left\langle \sum_{i=1}^{n} \delta(\mathbf{r}-\mathbf{r}_i) \right\rangle_{\mu,V,T}. \tag{3}$$

The ensemble average marginalizes molecular orientation, yielding the molecule-center density. Integrating this density recovers the uptake consistent with Equation (1). In practice, GCMC with classical force fields is a standard simulation tool for large-scale MOF screening, with studies evaluating over 10^5 of structures for gas storage and separation (Wilmer et al., 2012; Colón & Snurr, 2014; Simon et al., 2015).

Density view: classical density functional theory. cDFT reformulates adsorption equilibrium as a variational problem over the one-body density field (Evans, 1979; Roth, 2010). For a guest species s at fixed (T,μ), the equilibrium density ρ_{eq} is defined as the minimizer of the grand-potential functional

$$\Omega[\rho;\mathcal{M},s] = F_{\mathrm{id}}[\rho] + F_{\mathrm{exc}}[\rho;s] + \int_{\mathcal{V}} \rho(\mathbf{r})\big[V_{\mathrm{ext}}(\mathbf{r};\mathcal{M},s) - \mu\big]\, d\mathbf{r}, \tag{4}$$

where F_{id} is the ideal-gas free energy, F_{exc} captures non-ideal fluid correlations, and V_{ext} is the framework–fluid external potential. Setting $\delta\Omega/\delta\rho = 0$ yields the Euler–Lagrange condition

$$\rho_{\mathrm{eq}}(\mathbf{r}) = \rho_{\mathrm{bulk}}\, \exp\Big[-\beta\, V_{\mathrm{ext}}(\mathbf{r}) - \beta\, \frac{\delta F_{\mathrm{exc}}}{\delta\rho}\Big|_{\rho_{\mathrm{eq}}} + \beta\, \mu_{\mathrm{exc}}\Big], \tag{5}$$

where $\beta = 1/k_B T$, ρ_{bulk} is the bulk fluid density at the given (T,μ), and μ_{exc} is the excess chemical potential of the bulk fluid. Because F_{exc} depends nonlinearly on ρ, this self-consistent equation is solved by fixed-point iteration on a volumetric grid until convergence. Uptake is then computed by integrating the converged density field. Recent work has shown that cDFT can reproduce GCMC-level gas uptake in MOFs at substantially lower computational cost (Dufour-Décieux et al., 2025;

Thiele et al., 2026). The two simulation references supervise the same physical observable under different reference models. We use the PC-SAFT functional (Sauer & Gross, 2017), which parameterizes each species by segment number m, segment diameter σ, and dispersion energy ε. Supporting derivations and numerical details for the Euler–Lagrange equation, Boltzmann reference, Picard iteration, and PC-SAFT functional are given in Section A.

2.2 Machine learning for adsorption observables

Machine learning approaches to adsorption have mostly targeted scalar uptake at a given thermodynamic state. These models differ in representation and backbone. Existing models have progressed from building-block and topological embeddings (Lee et al., 2021) to energy-based representations such as 2D energy histograms (Shi et al., 2023) and voxelized potential-energy fields processed by 3D CNNs (Sarikas et al., 2024). More recent work adopts graph and transformer architectures that combine local atomic environments, global textural features, and transferable pre-training across materials and gas species (Chen et al., 2022; Lin et al., 2025; Cui et al., 2023; Kang et al., 2023; Wang et al., 2024).

In their original formulations, existing density predictors condition reconstruction on scalar uptake or predict unit-normalized fields, rather than independently determining uptake. Sun and Siepmann (Sun & Siepmann, 2024) predict voxel-wise differential contributions to global quadratic isotherm parameters, enabling thermodynamic-state-dependent density reconstruction but requiring global uptake information. Burner et al. (Burner et al., 2026) use an equivariant graph network based on DeepDFT (Jørgensen & Bhowmik, 2020; 2022) to predict normalized adsorbate density fields in MOFs. While suitable for binding sites, this approach cannot recover global uptake N, since ρ is normalized to integrate to unity.

3 Method

MADField is a multi-fidelity neural density operator for amortized adsorption simulation. Given a porous framework $\mathcal{M}$, an adsorbate species s, and a thermodynamic condition (P, T), MADField predicts the equilibrium adsorbate density field in physical units. Gas uptake is obtained by integrating this field, while the same prediction preserves spatial adsorption structure and can serve as an initial density for iterative cDFT solvers. The training pipeline first learns MADField-cDFT from solver-converged cDFT density fields, then uses its predictions to reinitialize failed cDFT calculations and expand the cDFT label set, and finally adapts the resulting density operator to sparse GCMC-derived density fields to obtain MADField-GCMC.

3.1 Multi-fidelity density training

Shared density target. MADField is trained to predict the equilibrium one-body density field $\rho_{\mathrm{eq}}(\mathbf{r}; \mathcal{M}, s, P, T)$. For MADField-cDFT, this density is supervised by solver-converged PC-SAFT cDFT labels; for MADField-GCMC, it is supervised by particle-to-grid averaged GCMC labels. Thus, both stages learn the same field-level object, while the source of supervision changes. The scalar uptake is recovered by integrating the predicted density.

cDFT pre-training. The first stage trains MADField-cDFT on PC-SAFT cDFT density fields. These labels are available at larger scale and across more adsorbates than GCMC-derived density fields, but inherit the approximation and convergence properties of the cDFT solver. This stage learns a broad adsorption-density prior that captures how framework geometry, adsorbate identity, and thermodynamic condition shape the equilibrium density field.

cDFT refinement from predicted densities. Because MADField-cDFT predicts an unnormalized equilibrium adsorbate density field, its output can be used not only as a one-shot prediction but also as an initial density for the original cDFT solver. For PC-SAFT cDFT, the equilibrium density is a fixed point of the Picard map, $\rho_{\mathrm{eq}} = \mathcal{T}_{\boldsymbol{\xi}}[\rho_{\mathrm{eq}}]$, with iterations $\rho^{(n+1)} = \mathcal{T}_{\boldsymbol{\xi}}[\rho^{(n)}]$, where $\boldsymbol{\xi}$ denotes the PC-SAFT adsorbate parameters. In a standard cDFT run, the initial density is the Boltzmann density ρ_{B}. In our refinement procedure, we keep the same solver but replace this initial density with the predicted density, $\rho^{(0)} = \hat{\rho}_\theta$. The subsequent iterations are still performed by the original

cDFT solver, and only converged runs are retained. Thus, the refined outputs are solver-converged cDFT densities rather than neural pseudo-labels.

Label bootstrapping. We use this refinement procedure for label bootstrapping, as illustrated in Figure 2. After training MADField-cDFT on converged cDFT calculations, we use its predictions as initial densities for cases that previously failed under the same standard initialization and rerun the original solver. When the reinitialized solve converges, the resulting fixed-point density satisfies the same cDFT convergence criteria and is added to the supervised cDFT dataset. This expands cDFT supervision without treating neural predictions as ground truth.

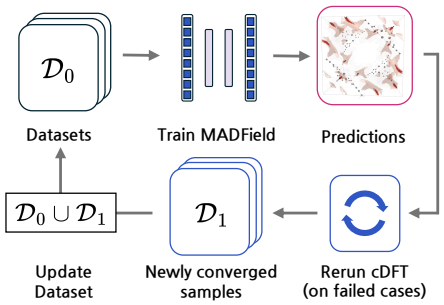


Figure 2: **Bootstrapping cDFT labels.** MADField reinitializes failed cDFT cases with predicted density fields and reruns the original solver. Only solver-converged outputs become additional training labels.

GCMC adaptation. After cDFT pre-training and bootstrapping, we adapt the resulting MADField-cDFT checkpoint to GCMC-derived density fields. GCMC provides particle-simulation references under the chosen force field, but field-level GCMC labels are more expensive and therefore sparser. MADField-GCMC is obtained by fine-tuning MADField-cDFT on paired GCMC density grids. This stage uses the cDFT-trained model as an adsorption-density prior and updates it toward particle-simulation fidelity, rather than learning the GCMC density operator from scratch. Thus, the abundant cDFT labels provide broad geometric and thermodynamic coverage, while the sparse GCMC labels correct the remaining fidelity gap.

Training objective. Training combines density supervision and integrated-uptake loss. cDFT and initial CH_4/Xe adaptation use

$$\mathcal{L} = \mathcal{L}_{\text{field}} + \lambda_N \mathcal{L}_N, \tag{6}$$

where $\mathcal{L}_{\text{field}}$ combines log-residual, relative pore-density, high-density-site, and coarse-branch supervision; $\mathcal{L}_N$ compares integrated prediction and target uptake. For framework–CO_2 electrostatics, the extension uses charge-derived features and charge-conditioned density refinement. Section B.3 gives its stage-specific objectives and final reconstruction; Section B defines the base loss.

3.2 Periodic volumetric density operator

Periodic volumetric representation. The continuous density fields are represented on a periodic discrete voxel grid for learning and evaluation. The operator is designed around three requirements: periodic cell geometry, three-dimensional pore connectivity, and continuous dependence on thermodynamic and molecular conditions. The volumetric input contains the reduced external potential and the log-Boltzmann density. The convolutional stem operations are periodic: the convolutional stem uses circular padding, shifted-window attention uses circular indexing, and positional information is encoded with periodic Fourier features. We provide implementation details in Section D.

Boltzmann-residual parameterization. To encode the dominant one-body framework response, we use the ideal-gas Boltzmann density $\rho_{\text{Boltz}}(\mathbf{r}) = \rho_{\text{bulk}} \exp[-\beta V_{\text{ext}}(\mathbf{r})]$, where V_{ext} is the framework–fluid external potential, ρ_{bulk} is the bulk number density at (P, T), and $\beta = (k_B T)^{-1}$. Its derivation and relation to the cDFT Euler–Lagrange equation are given in Section A. The base density operator predicts a gated correction in log-density space:

$$\log \hat{\rho}_\theta(\mathbf{r}) = \log \rho_{\text{Boltz}}(\mathbf{r}) + \sigma(h_\theta(\mathbf{r}))\, \delta_\theta(\mathbf{r}), \tag{7}$$

where $\delta_\theta(\mathbf{r})$ is a residual field and $h_\theta(\mathbf{r})$ is a gate field. This ensures positive density. The CO_2 extension adds charge-conditioned refinement and a positive amplitude correction (Section B.3).

Backbone and conditioning. The backbone is a U-shaped 3D Swin Transformer (Liu et al., 2021). A two-stage convolutional stem downsamples the discrete voxel grid to a token resolution where windowed attention is tractable. Shifted windows, skip connections, and the decoder

allow adsorption features to propagate across connected pores while keeping memory cost manageable. Non-volumetric information is injected through a conditioning vector containing $\log \rho_{\text{bulk}}$, $\log P$, lattice parameters, and a three-scalar molecular descriptor. This vector modulates every Swin block through adaptive normalization. For MADField-cDFT, the descriptor is the PC-SAFT triplet (m, σ, ε) (segment number, segment diameter, and dispersion energy); for MADField-GCMC, the same descriptor slot is filled with Lennard-Jones parameters from the GCMC references. For CO_2, additional charge-derived features are incorporated as described in Section B.3.

4 Experiments

4.1 Setup

Datasets. We evaluate two simulation fidelities with framework-disjoint train/validation/test splits. cDFT uses QMOF (Rosen et al., 2021) across nine adsorbates; GCMC uses ARC-MOF CH_4/Xe references (Burner et al., 2022; 2026) and our CO_2 density dataset generated for 2,000 MOFs from MOSAEC-DB (Gibaldi et al., 2025). Zero-shot transfer uses amorphous carbons (Gardner et al., 2023) and PRAM disordered materials (Thyagarajan & Sholl, 2020). Screening evaluates 254,198 ARC-MOF frameworks for CH_4 and 15,572 MOSAEC-DB frameworks for CO_2, excluding the respective GCMC training structures. Data generation, splits, and evaluation populations are detailed in Section F.

Table 1: **Model capabilities.** Original-model capabilities: independent uptake (N), density (ρ), adsorbate (Ads.), pressure (P), material class (Mat.); ✔/✘/▲: full/no/partial support.

Model	N	ρ	Ads.	P	Mat.
RetNet	✔	✘	✘	✘	✔
MOFTransformer	✔	✘	✘	✘	✘
DeepSorption	✔	✘	✘	▲	✔
IsothermNet	✔	✘	✔	✔	✔
Uni-MOF	✔	✘	✔	✔	✘
SorbIIT	✘	▲	✔	✔	✔
DeepAPD	✘	▲	✘	✘	✔
MADField (Ours)	✔	✔	✔	✔	✔

Baselines. We compare scalar uptake against RetNet, MOFTransformer, DeepSorption, Uni-MOF, and IsothermNet, and density fields against DeepAPD and SorbIIT. Table 1 summarizes their capabilities. MADField-cDFT is pre-trained on cDFT fields and adapted to GCMC labels as MADField-GCMC. Density-baseline reconstruction, uptake dependencies, and training details are given in Section G.

Metrics. We evaluate uptake using mean absolute error N_{err} in $\text{cm}^3(\text{STP})/\text{g}$ and spatial density agreement using the Tanimoto coefficient T_ρ between unit-integral density fields. Scores are averaged over evaluated states within each species; **Mean** then weights species equally, covering nine species for cDFT and CH_4/Xe/CO_2 for GCMC. For PRAM, we average the four material-family means equally. Full metric definitions are provided in Section H.

Screening protocol. For CH_4, candidates are ranked by working capacity between $P_{\text{ads}} = 65$ bar and $P_{\text{des}} = 5.8$ bar at $T = 298$ K:

$$\text{WC}(\mathcal{M}, \text{CH}_4) = q_{\text{vol}}(P_{\text{ads}}, T; \mathcal{M}, \text{CH}_4) - q_{\text{vol}}(P_{\text{des}}, T; \mathcal{M}, \text{CH}_4) \tag{8}$$

Here, q_{vol} denotes volumetric uptake in $\text{cm}^3(\text{STP})/\text{cm}^3$, obtained from the predicted molecule count by the conversion in Appendix H. For CO_2, candidates are ranked by uptake at 0.15 bar. Targets are the reference-defined top 0.1%, 1%, or 10% of each candidate pool. We measure their recall as an increasing fraction of candidates is evaluated in predicted rank order. Additional fixed-threshold retrieval results are reported in Section I.4.

4.2 Density-field prediction on MOFs

Table 2 reports density-field prediction on the cDFT and three-species GCMC benchmarks. MADField achieves mean $T_\rho = 0.994$ on cDFT and 0.858 on GCMC, compared with 0.909 and 0.705 for DeepAPD, the strongest density baseline. For CO_2, MADField achieves $T_\rho = 0.634$, compared with 0.329 for DeepAPD and 0.241 for SorbIIT. Additional density results on amorphous carbons are provided in Section I.3. Representative fields are shown in Figure 6, and physical-density errors of the final three-species model are reported in Table 19.

Table 2: **Density field prediction on MOFs.** Tanimoto similarity between unit-integral density fields (↑); **Mean** weights species equally. cDFT uses 3,366 common test states (Section F.1). Ads. and P: adsorbate and pressure conditioning.

Cond.		Model	cDFT density										GCMC density			
Ads.	P		H_2	Ar	Kr	Xe	N_2	CH_4	CO_2	C_2H_6	C_3H_8	**Mean**	CH_4	Xe	CO_2	**Mean**
✔	✔	SorbIIT	0.842	0.675	0.605	0.726	0.733	0.632	0.858	0.777	0.819	0.741	0.730	0.690	0.241	0.554
✖	✖	DeepAPD	0.964	0.951	0.918	0.869	0.957	0.903	0.926	0.881	0.810	0.909	0.913	0.873	0.329	0.705
✔	✔	**MADField**	**1.000**	**1.000**	**0.998**	**0.990**	**1.000**	**0.995**	**0.993**	**0.992**	**0.974**	**0.994**	**0.979**	**0.963**	**0.634**	**0.858**

Table 3: **Gas uptake prediction on MOFs.** MAE in $\mathrm{cm}^3(\mathrm{STP})/\mathrm{g}$ (↓); **Mean** weights species equally. Ads. and P denote adsorbate and pressure conditioning.

Cond.		Model	cDFT uptake										GCMC uptake			
Ads.	P		H_2	Ar	Kr	Xe	N_2	CH_4	CO_2	C_2H_6	C_3H_8	**Mean**	CH_4	Xe	CO_2	**Mean**
✖	✖	RetNet	4.83	3.93	9.46	22.05	3.31	17.44	17.91	27.70	46.62	17.03	10.72	10.65	27.72	16.36
		MOFTransformer	0.12	1.87	6.21	15.09	1.32	5.32	12.12	15.66	23.40	9.01	7.22	12.32	22.08	13.87
✖	▲	DeepSorption	4.07	5.00	10.65	21.24	3.66	12.71	21.19	19.19	45.37	15.90	12.73	20.19	28.55	20.49
		IsothermNet	5.21	7.53	18.71	40.25	5.53	21.71	44.11	52.15	81.59	30.75	16.26	21.78	32.39	23.48
✔	✔	Uni-MOF	0.46	1.81	5.11	12.21	1.22	4.13	10.39	11.81	22.12	7.70	8.11	12.29	22.24	14.21
		MADField	**0.08**	**0.11**	**0.49**	**2.86**	**0.13**	**0.79**	**1.17**	**2.00**	**4.49**	**1.35**	**1.41**	**2.25**	**8.76**	**4.14**

4.3 Gas uptake prediction

Table 3 reports uptake MAE (N_{err}) on MOFs. MADField achieves mean $N_{\mathrm{err}} = 1.35\,\mathrm{cm}^3(\mathrm{STP})/\mathrm{g}$ on the complete cDFT test set and $4.14\,\mathrm{cm}^3(\mathrm{STP})/\mathrm{g}$ on the CH_4/Xe/CO_2 GCMC test set. Compared with the strongest baselines (Uni-MOF at 7.70 and MOFTransformer at 13.87), these correspond to approximately $5.7\times$ and $3.35\times$ lower MAE, respectively. On the common CO_2 test set, MADField achieves $8.76\,\mathrm{cm}^3(\mathrm{STP})/\mathrm{g}$, compared with 22.08 for MOFTransformer, the lowest-error evaluated scalar baseline (about $2.5\times$ lower MAE).

To examine the prediction target, we compare MADField with a scalar-output counterpart using the same backbone. On cDFT labels, density-field prediction reduces uptake MAE from 7.47 to 1.31 under the common evaluation protocol in Section I.1. For CH_4/Xe GCMC, we compare scalar- and density-supervised variants with and without cDFT pre-training (Table 15). The pre-trained scalar variant learns cDFT uptake before fine-tuning on GCMC uptake. Density supervision reduces uptake MAE from 7.950 to 2.152 without pre-training and from 7.531 to 1.996 with pre-training.

To evaluate multi-fidelity transfer, we compare the two density-supervised variants. cDFT pre-training reduces uptake MAE from 2.152 to 1.996; the relative reduction increases to 41.0% when only 10% of GCMC labels are available (Section I.2). Section 4.4 evaluates how the complete pipeline translates prediction accuracy into reduced screening budgets.

Transferability to other materials. Table 4 reports zero-shot transfer to disordered porous materials unseen during training. On the cDFT amorphous-carbon benchmark, MADField achieves $2.71\,\mathrm{cm}^3(\mathrm{STP})/\mathrm{g}$, improving $4.9\times$ over the strongest baseline. On the GCMC benchmark, MADField achieves mean MAE $1.62\,\mathrm{cm}^3(\mathrm{STP})/\mathrm{g}$, improving $3.1\times$ over the strongest baseline. The family-wise MAEs are 1.72, 1.21, 2.11, and 1.45 for PIM, HCP, Kerogen, and AC, respectively (Section F.1). Figure 3 shows representative frameworks, predicted densities, and isotherms for QMOF (ID) and JLA amorphous carbon (OOD).

Table 4: **Transfer to disordered materials.** Uptake MAE in $\mathrm{cm}^3(\mathrm{STP})/\mathrm{g}$ (↓). cDFT: CH_4 on AC at 1 bar; GCMC: PRAM.

	cDFT	GCMC				
Model	AC	PIM	HCP	Ker.	AC	**Mean**
RetNet	21.30	2.28	1.94	6.36	9.51	5.02
MOFTF	56.40	6.90	13.10	13.60	30.60	16.05
DeepSorp	52.50	6.14	6.72	26.80	23.10	15.69
IsoNet	13.33	9.63	11.41	44.23	46.03	27.83
Uni-MOF	31.30	8.20	6.10	27.80	26.90	17.25
MADField	**2.71**	**1.72**	**1.21**	**2.11**	**1.45**	**1.62**

Ker.: Kerogen; MOFTF: MOFTransformer; DeepSorp: DeepSorption; IsoNet: IsothermNet.

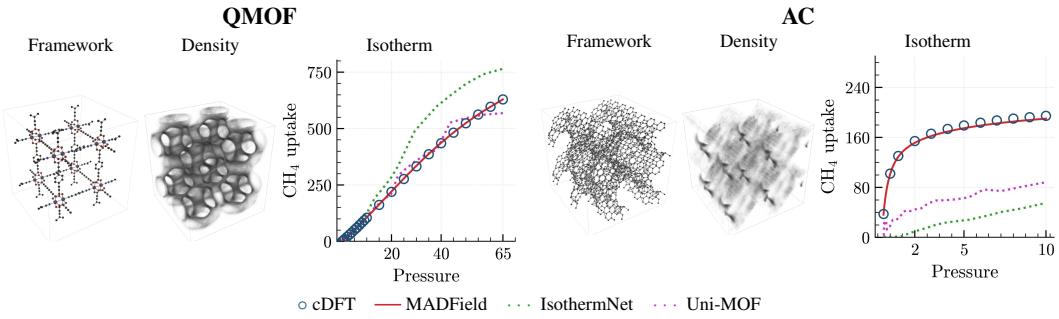


Figure 3: **Frameworks, predicted densities, and adsorption isotherms.** CH_4 predictions for QMOF qmof-0e8c4b6 (left) and JLA amorphous carbon jla_d1p1_T3300_t210 (right). Grayscale isosurfaces show the predicted cDFT-level density at 65 bar (QMOF) and 10 bar (AC); isotherm uptake is in cm^3/g and pressure in bar.

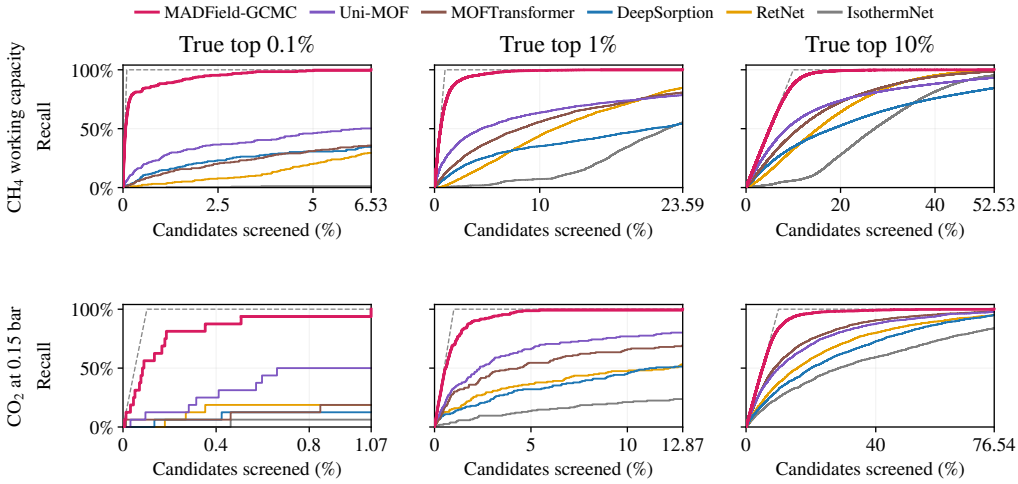


Figure 4: **Cumulative recovery in CH_4 and CO_2 screening.** Rows show CH_4 working capacity (254,198 ARC-MOF candidates) and CO_2 uptake at 0.15 bar (15,572 MOSAEC-DB candidates). Columns target the true top 0.1%, 1%, and 10%. Each panel ends at MADField-GCMC's first 100% recall, so horizontal scales differ; gray dashed lines show ideal rankings. Full 0–100% curves appear in Figure 10.

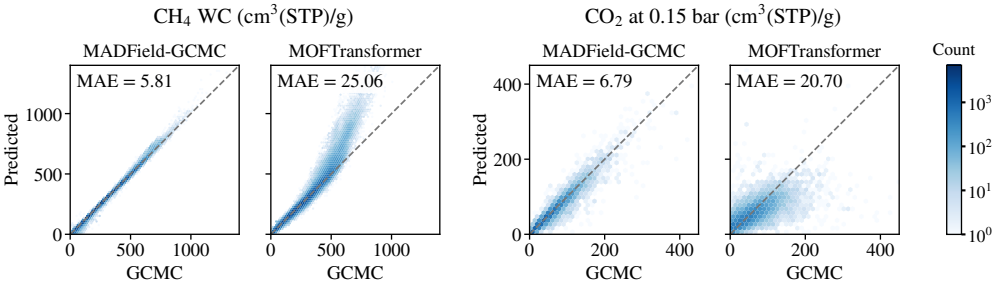


Figure 5: **Screening prediction parity.** MADField-GCMC and the lowest-MAE scalar baseline, MOFTransformer, predict CH_4 working capacity (left) and CO_2 uptake at 0.15 bar (right). MAEs $(\mathrm{cm}^3(\mathrm{STP})/\mathrm{g})$ use the GCMC-training-excluded ARC-MOF/MOSAEC-DB screening cohorts (254,198/15,572 frameworks), distinct from the uptake-test cohorts in Table 3. Hexbins show log-counts; full baseline comparisons appear in Section I.4.

4.4 High-throughput screening for methane storage and carbon capture

Beyond per-condition accuracy, practical screening must recover high-performing frameworks while minimizing follow-up GCMC evaluations. We evaluate two tasks: CH_4 working-capacity screening

on ARC-MOF using publicly available GCMC uptake data (Burner et al., 2022), and CO_2 uptake screening on a subset of MOSAEC-DB (Gibaldi et al., 2025) using GCMC uptake references that we computed. The CH_4 pool contains 254,198 candidates after excluding 16,385 frameworks in the GCMC CH_4/Xe training union. The CO_2 comparison uses 15,572 common-support frameworks held out from CO_2 supervision at 0.15 bar (Section F.3). We use the CH_4/Xe-adapted checkpoint for CH_4 and the CO_2-extended checkpoint for CO_2.

Figure 4 compares cumulative recovery of the true top 0.1%, 1%, and 10% of each pool. All models rank the same candidates within each task, and targets are fixed by the GCMC references before varying the screening budget. These sets contain 255, 2,542, and 25,420 CH_4 targets, and 16, 156, and 1,558 CO_2 targets, respectively. This tests retrieval across different target prevalences rather than a single performance threshold. Each panel displays the budget range up to MADField-GCMC's first complete recovery, with its own horizontal scale. The complete-recovery budgets reported below are computed from the full rankings.

MADField-GCMC concentrates high-performing candidates near the top of its rankings in both tasks. For the rarest targets (top 0.1%), it recovers all targets after screening 6.53% of the CH_4 pool and 1.07% of the CO_2 pool; the strongest scalar baseline at complete recall requires 52.16% (RetNet) and 12.27% (Uni-MOF), respectively. The advantage persists for the top 1%: complete recovery requires 23.59% and 12.87% of the two pools, compared with 57.56% (RetNet) and 86.37% (DeepSorption) for the respective best scalar baselines. These rankings therefore reduce the simulation budget needed to recover high-performing candidates across both working-capacity and uptake screening.

Figure 5 complements the ranking results with screening parity: MADField-GCMC reduces MAE relative to MOFTransformer from 25.06 to 5.81 $\mathrm{cm}^3(\mathrm{STP})/\mathrm{g}$ for CH_4 WC and uptake MAE from 20.70 to 6.79 $\mathrm{cm}^3(\mathrm{STP})/\mathrm{g}$ for CO_2 at 0.15 bar. This reduction in follow-up evaluations is complemented by fast neural inference. In the CH_4 timing benchmark, GPU inference takes 0.043 s compared with approximately 3 h for the single-core CPU RASPA reference per pressure state, corresponding to a 2.5×10^5 speedup (Section I.6).

4.5 Accelerating cDFT with MADField

We use predicted equilibrium densities to initialize the original cDFT solver (Table 5). On a controlled panel of 180 states spanning nine adsorbates and the same 20 frameworks at 10 bar, both initializations use identical damping schedules, convergence tolerances, and iteration caps. Standard and MADField initialization converge on 132 and 137 states, respectively. On states converging under both, median iterations decrease from 379.5 to 62.0, a 6.1× reduction. This corresponds to 5.7× faster solver execution, with median walltime decreasing from 27.02 to 4.73 s.

Table 5: **MADField-initialized cDFT.** Dataset recovery and the separate 180-state solver panel.

Metric	Standard	**MADField**
Dataset convergence (%)	90.06	**93.82**
Panel median solver steps	379.5	**62.0**
Panel solver speedup (×)	1.0	**5.7**

Two-stage bootstrapping recovers 23,602 of the 62,475 failed states (37.78%, approximately 38%), increasing dataset convergence from 90.06% to 93.82% (+3.75 percentage points). Only solver-converged outputs augment the label set. The recovery protocol and bootstrap model progression are detailed in Section C and Section I.5.

5 Conclusion

We introduced MADField, which predicts equilibrium adsorbate density fields and obtains uptake by integration. The resulting fields support uptake prediction, spatial adsorption analysis, and cDFT solver initialization. Controlled ablations establish the uptake-accuracy benefit of physical-density supervision beyond scalar regression and cDFT pre-training alone. Database-scale CH4 and CO2 screening demonstrates the practical payoff: substantially fewer follow-up GCMC evaluations at complete target recall, with the largest gains for rare high-performing candidates. Performance depends on the simulation references and varies across unseen material families.

Reproducibility

Sections F and H specify the evaluation sets and metrics. Architecture, training, and simulation settings are provided in the appendices. We provide the code as supplementary material and will publicly release the generated datasets to support reproduction and further research.

AI use statement

We used generative AI tools to assist with minor manuscript revisions and proofreading, to support the initial search for related work, and to generate portions of the research code. All AI-assisted outputs were reviewed and verified by the authors. The authors take responsibility for the final content of this work.

References

DJ Adams. Grand canonical ensemble monte carlo for a lennard-jones fluid. *Molecular Physics*, 29 (1):307–311, 1975. 1, 3

Benjamin J Bucior, N Scott Bobbitt, Timur Islamoglu, Subhadip Goswami, Arun Gopalan, Taner Yildirim, Omar K Farha, Neda Bagheri, and Randall Q Snurr. Energy-based descriptors to rapidly predict hydrogen storage in metal–organic frameworks. *Molecular Systems Design & Engineering*, 4(1):162–174, 2019. doi: 10.1039/C8ME00050F. 1

Jake Burner. Metal-organic framework ch4@1bar adsorbate probability distributions, August 2025a. URL https://doi.org/10.5281/zenodo.16800893. 20

Jake Burner. Metal-organic framework ch4@65bar adsorbate probability distributions, August 2025b. URL https://doi.org/10.5281/zenodo.16801034. 20

Jake Burner. Metal-organic framework xe@1bar adsorbate probability distributions, August 2025c. URL https://doi.org/10.5281/zenodo.16801181. 20

Jake Burner, Jun Luo, Andrew White, Adam Mirmiran, Ohmin Kwon, Peter G. Boyd, Stephen Maley, Marco Gibaldi, Scott Simrod, and Tom K. Woo. ab initio repeat charge mof database (arc-mof), July 2022. URL https://doi.org/10.5281/zenodo.6908728. 6, 9, 20

Jake Burner, Olivier Marchand, Rosa Cicciarella, Marco Gibaldi, and Tom K Woo. Rapid prediction of single-site adsorbate probability distributions in metal–organic frameworks using graph neural networks. *Digital Discovery*, 2026. 4, 6

Carlos Campañá, Bastien Mussard, and Tom K. Woo. Electrostatic potential derived atomic charges for periodic systems using a modified error functional. *Journal of Chemical Theory and Computation*, 5(10):2866–2878, 2009. doi: 10.1021/ct9003405. 16, 22

Charithea Charalambous, Elias Moubarak, Johannes Schilling, Eva Sanchez Fernandez, Jin-Yu Wang, Laura Herraiz, Fergus Mcilwaine, Shing Bo Peh, Matthew Garvin, Kevin Maik Jablonka, et al. A holistic platform for accelerating sorbent-based carbon capture. *Nature*, 632(8023):89–94, 2024. 1

Pin Chen, Rui Jiao, Jinyu Liu, Yang Liu, and Yutong Lu. Interpretable graph transformer network for predicting adsorption isotherms of metal–organic frameworks. *Journal of Chemical Information and Modeling*, 62(22):5446–5456, 2022. 4

Yamil J Colón and Randall Q Snurr. High-throughput computational screening of metal–organic frameworks. *Chemical Society Reviews*, 43(16):5735–5749, 2014. 3

Jiyu Cui, Fang Wu, Wen Zhang, Lifeng Yang, Jianbo Hu, Yin Fang, Peng Ye, Qiang Zhang, Xian Suo, Yiming Mo, et al. Direct prediction of gas adsorption via spatial atom interaction learning. *Nature Communications*, 14(1):7043, 2023. 4

Vincent Dufour-Décieux, Philipp Rehner, Johannes Schilling, Elias Moubarak, Joachim Gross, and André Bardow. Classical density functional theory as a fast and accurate method for adsorption property prediction of porous materials. *AIChE Journal*, 71(6):e18779, 2025. 1, 3, 15

Timm Esper, Gernot Bauer, Philipp Rehner, and Joachim Gross. Pcp-saft parameters of pure substances using large experimental databases. *Industrial & Engineering Chemistry Research*, 62 (37):15300–15310, 2023. 20

Robert Evans. The nature of the liquid-vapour interface and other topics in the statistical mechanics of non-uniform, classical fluids. *Advances in physics*, 28(2):143–200, 1979. 1, 3

Almudena García-Sánchez, Conchi O. Ania, José B. Parra, David Dubbeldam, Thijs J. H. Vlugt, Rajamani Krishna, and Sofía Calero. Transferable force field for carbon dioxide adsorption in zeolites. *The Journal of Physical Chemistry C*, 113(20):8814–8820, 2009. doi: 10.1021/jp810871f. 22

John LA Gardner, Zoé Faure Beaulieu, and Volker L Deringer. Synthetic data enable experiments in atomistic machine learning. *Digital Discovery*, 2(3):651–662, 2023. 6, 20

Marco Gibaldi, Anna Kapeliukha, Andrew White, Jun Luo, Robert Alex Mayo, Jake Burner, and Tom K Woo. MOSAEC-DB: a comprehensive database of experimental metal–organic frameworks with verified chemical accuracy suitable for molecular simulations. *Chemical Science*, 16 (9):4085–4100, 2025. doi: 10.1039/D4SC07438F. 6, 9, 22

Joachim Gross and Gabriele Sadowski. Perturbed-chain saft: An equation of state based on a perturbation theory for chain molecules. *Industrial & engineering chemistry research*, 40(4):1244–1260, 2001. 15

Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*, 2022. 15

Peter Bjørn Jørgensen and Arghya Bhowmik. Deepdft: Neural message passing network for accurate charge density prediction. In *Machine Learning for Molecules Workshop@ NeurIPS 2020*, 2020. 4

Peter Bjørn Jørgensen and Arghya Bhowmik. Equivariant graph neural networks for fast electron density estimation of molecules, liquids, and solids. *npj Computational Materials*, 8(1):183, 2022. 4

Yeonghun Kang, Hyunsoo Park, Berend Smit, and Jihan Kim. A multi-modal pre-training transformer for universal transfer learning in metal–organic frameworks. *Nature Machine Intelligence*, 5(3):309–318, 2023. 4

Ohmin Kwon, Marco Gibaldi, Kasturi Nagesh Pai, Arvind Rajendran, and Tom K Woo. Identification of metal–organic frameworks for near practical energy limit CO2 capture from wet flue gases: An integrated atomistic and process simulation screening of experimental MOFs. *ACS Central Science*, 11(8):1438–1451, 2025. doi: 10.1021/acscentsci.5c00777. 22

Sangwon Lee, Baekjun Kim, Hyun Cho, Hooseung Lee, Sarah Yunmi Lee, Eun Seon Cho, and Jihan Kim. Computational screening of trillions of metal–organic frameworks for high-performance methane storage. *ACS Applied Materials & Interfaces*, 13(20):23647–23654, 2021. 4

Emily Lin, Yang Zhong, Gang Chen, and Sili Deng. Unified physio-thermodynamic descriptors via learned co2 adsorption properties in metal-organic frameworks. *npj Computational Materials*, 11 (1):225, 2025. 4

Li-Chiang Lin, Adam H Berger, Richard L Martin, Jihan Kim, Joseph A Swisher, Kuldeep Jariwala, Chris H Rycroft, Abhoyjit S Bhown, Michael W Deem, Maciej Haranczyk, et al. In silico screening of carbon-capture materials. *Nature materials*, 11(7):633–641, 2012. 1

Tsung-Wei Liu, Quan Nguyen, Adji Bousso Dieng, and Diego A. Gómez-Gualdrón. Diversity-driven, efficient exploration of a MOF design space to optimize MOF properties. *Chemical Science*, 15(45):18903–18919, 2024. doi: 10.1039/D4SC03609C. 1

Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 10012–10022, 2021. 5, 19

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017. 15

Sauradeep Majumdar, Seyed Mohamad Moosavi, Kevin Maik Jablonka, Daniele Ongari, and Berend Smit. Diversifying databases of metal organic frameworks for high-throughput computational screening. *ACS Applied Materials & Interfaces*, 13(51):61004–61014, 2021. doi: 10.1021/acsami.1c16220. 1

Jarad A Mason, Mike Veenstra, and Jeffrey R Long. Evaluating metal–organic frameworks for natural gas storage. *Chemical Science*, 5(1):32–51, 2014. 1

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. 19

YA Ran, S Sharma, SRG Balestra, Z Li, S Calero, TJH Vlugt, RQ Snurr, and D Dubbeldam. Raspa3: A monte carlo code for computing adsorption and diffusion in nanoporous materials and thermodynamics properties of fluids. *The Journal of Chemical Physics*, 161(11), 2024. 22

Anthony K Rappé, Carla J Casewit, Kent S Colwell, William A Goddard III, and W Mason Skiff. Uff, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *Journal of the American chemical society*, 114(25):10024–10035, 1992. 15, 20, 22

Andrew S Rosen, Shaelyn M Iyer, Debmalya Ray, Zhenpeng Yao, Alan 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. 6, 20

Roland Roth. Fundamental measure theory for hard-sphere mixtures: a review. *Journal of Physics: Condensed Matter*, 22(6):063102, 2010. 1, 3, 15

Antonios P Sarikas, Konstantinos Gkagkas, and George E Froudakis. Gas adsorption meets deep learning: voxelizing the potential energy surface of metal-organic frameworks. *Scientific Reports*, 14(1):2242, 2024. 4

Elmar Sauer and Joachim Gross. Classical density functional theory for liquid–fluid interfaces and confined systems: A functional for the perturbed-chain polar statistical associating fluid theory equation of state. *Industrial & Engineering Chemistry Research*, 56(14):4119–4135, 2017. 4, 15

Noam Shazeer. Glu variants improve transformer. *arXiv preprint arXiv:2002.05202*, 2020. 19

Kaihang Shi, Zhao Li, Dylan M Anstine, Dai Tang, Coray M Colina, David S Sholl, J Ilja Siepmann, and Randall Q Snurr. Two-dimensional energy histograms as features for machine learning to predict adsorption in diverse nanoporous materials. *Journal of chemical theory and computation*, 19(14):4568–4583, 2023. 4

Cory M Simon, Jihan Kim, Diego A Gomez-Gualdron, Jeffrey S Camp, Yongchul G Chung, Richard L Martin, Rocio Mercado, Michael W Deem, Dan Gunter, Maciej Haranczyk, et al. The materials genome in action: identifying the performance limits for methane storage. *Energy & Environmental Science*, 8(4):1190–1199, 2015. 3

Rolf Stierle, Gernot Bauer, Nadine Thiele, Benjamin Bursik, Philipp Rehner, and Joachim Gross. Classical density functional theory in three dimensions with gpu-accelerated automatic differentiation: Computational performance analysis using the example of adsorption in covalent-organic frameworks. *Chemical Engineering Science*, 298:120380, 2024. 15

Yangzesheng Sun and J Ilja Siepmann. Understanding and predicting the spatially resolved adsorption properties of nanoporous materials. *Journal of Chemical Theory and Computation*, 20(12):5259–5275, 2024. 4

Nadine Thiele, Tiong Wei Teh, Benjamin Bursik, Marcel Granderath, Gernot Bauer, Vincent Dufour-Décieux, Philipp Rehner, Rolf Stierle, André Bardow, Niels Hansen, et al. Efficient prediction of multicomponent adsorption isotherms and enthalpies of adsorption in mofs using classical density functional theory. *The Journal of Physical Chemistry B*, 2026. 2, 4

Raghuram Thyagarajan and David S Sholl. A database of porous rigid amorphous materials. *Chemistry of Materials*, 32(18):8020–8033, 2020. doi: 10.1021/acs.chemmater.0c03057. 1, 6, 20

Jingqi Wang, Jiapeng Liu, Hongshuai Wang, Musen Zhou, Guolin Ke, Linfeng Zhang, Jianzhong Wu, Zhifeng Gao, and Diannan Lu. A comprehensive transformer-based approach for high-accuracy gas adsorption predictions in metal-organic frameworks. *Nature Communications*, 15 (1):1904, 2024. 4

Christopher E Wilmer, Michael Leaf, Chang Yeon Lee, Omar K Farha, Brad G Hauser, Joseph T Hupp, and Randall Q Snurr. Large-scale screening of hypothetical metal–organic frameworks. *Nature chemistry*, 4(2):83–89, 2012. 1, 3

APPENDIX

A Additional Background on Adsorption Simulation **14**
A.1 Equilibrium Density and Fixed-Point Solution 14
A.2 PC-SAFT Functional and Framework Potential 15

B Training **15**
B.1 Density and Integrated-Uptake Losses 15
B.2 Pre-training and Initial GCMC Adaptation 15
B.3 Three-species Adaptation and CO_2 Refinement 16

C cDFT Warm-Start Protocol **18**

D Architecture Details **19**
D.1 Inputs and Periodic Backbone 19
D.2 Density Reconstruction 19

E PC-SAFT Parameters **20**

F Data Generation and Evaluation Sets **20**
F.1 Sources and Framework Partitions 20
F.2 cDFT and Published GCMC References 21
F.3 CO_2 Density Labels and Screening Pools 22

G Baseline Implementations **22**

H Evaluation Metrics **24**

I Additional Experiments **25**
I.1 cDFT Model Variants 25
I.2 Multi-fidelity Ablations 26
I.3 Density Accuracy and Spatial Analysis 27
I.4 Additional Screening Results 30
I.5 Bootstrap Model Progression 34
I.6 Computational Cost 34
I.7 Reference Fidelity between cDFT and GCMC 34

J Broader Impacts **35**

A ADDITIONAL BACKGROUND ON ADSORPTION SIMULATION

A.1 EQUILIBRIUM DENSITY AND FIXED-POINT SOLUTION

The number-density integral in Equation (1) gives the expected molecule count. GCMC estimates this one-body field by averaging molecule-center counts over configurations; for rigid multi-site molecules, the field marginalizes molecular orientation. The particle energy and grand potential are defined in Equations (2) and (4). Here we give the cDFT equation used to generate density labels and initialize the solver.

For $F_{\mathrm{id}} = k_B T \int \rho[\log(\rho\Lambda^3) - 1]d\mathbf{r}$ and $\mu_{\mathrm{exc}} = \delta F_{\mathrm{exc}}/\delta\rho$, stationarity of the grand potential gives

$$k_B T \log(\rho_{\mathrm{eq}}\Lambda^3) + \mu_{\mathrm{exc}}[\rho_{\mathrm{eq}}] + V_{\mathrm{ext}} - \mu = 0, \tag{9}$$

$$\rho_{\mathrm{eq}}(\mathbf{r}) = \rho_{\mathrm{bulk}} \exp\left[-\beta V_{\mathrm{ext}}(\mathbf{r}) - \beta\mu_{\mathrm{exc}}(\mathbf{r}; [\rho_{\mathrm{eq}}]) + \beta\mu_{\mathrm{exc}}^{\mathrm{bulk}}\right]. \tag{10}$$

The bulk equation of state fixes μ and ρ_{bulk} at (P, T). Dropping excess interactions yields the Boltzmann reference $\rho_B = \rho_{\text{bulk}} e^{-\beta V_{\text{ext}}}$. Denoting the right-hand side of Equation (10) by $\mathcal{T}[\rho]$, damped iteration uses

$$\rho^{(n+1)} = (1-\alpha)\rho^{(n)} + \alpha \mathcal{T}[\rho^{(n)}], \qquad 0 < \alpha \leq 1. \tag{11}$$

Standard initialization uses ρ_B; learned initialization uses the predicted physical density. The solver subsequently seeks the same fixed point (Section C).

A.2 PC-SAFT Functional and Framework Potential

We use the effective non-polar, non-associating PC-SAFT functional (Gross & Sadowski, 2001; Sauer & Gross, 2017):

$$F_{\text{exc}} = F_{\text{hs}} + F_{\text{chain}} + F_{\text{disp}}. \tag{12}$$

The hard-sphere contribution uses fundamental measure theory (Roth, 2010); the other terms account for segment connectivity and dispersion. Periodic convolutions are evaluated on the density grid (Stierle et al., 2024; Dufour-Décieux et al., 2025). The adsorbate triplet (m, σ, ε) specifies segment number, diameter, and dispersion energy (Table 9).

The cDFT input potential is a rigid-framework Lennard–Jones field,

$$V_{\text{ext}}^{\text{cDFT}}(\mathbf{r}) = m \sum_{j \in \mathcal{N}(\mathbf{r})} 4\varepsilon_{fj} \left[\left(\frac{\sigma_{fj}}{r_j(\mathbf{r})} \right)^{12} - \left(\frac{\sigma_{fj}}{r_j(\mathbf{r})} \right)^{6} \right]. \tag{13}$$

Framework atoms use UFF parameters (Rappé et al., 1992), with Lorentz–Berthelot mixing $\sigma_{fj} = (\sigma_f + \sigma_j)/2$ and $\varepsilon_{fj} = \sqrt{\varepsilon_f \varepsilon_j}$. Periodic distances and cell replication respect the 12.8 Å cutoff, including non-orthorhombic cells. GCMC reference interactions and the charge-aware CO_2 labels are specified in Sections F.2 and F.3.

B Training

B.1 Density and Integrated-Uptake Losses

For cDFT pre-training and initial CH_4/Xe adaptation, let $\rho^\star$ be the reference density, $\mathcal{P} = \{v : \beta V_v^{\text{input}} < 50\}$ the pore region, and $\Delta = \log \rho - \log \rho_B$ the log-residual. Positive numerical floors stabilize logarithms and relative errors at zero-count voxels. Define relative density error over a voxel set S by $E_S = \sum_{v \in S} |\hat{\rho}_v - \rho_v^\star| / (\sum_{v \in S} \rho_v^\star + \epsilon)$. The objective is

$$\mathcal{L} = \eta_{\text{res}} \mathcal{L}_{\text{res}} + \eta_\rho \mathcal{L}_\rho + \eta_{\text{top}} \mathcal{L}_{\text{top}} + \eta_N \mathcal{L}_N + \eta_{\text{hier}} \mathcal{L}_{\text{hier}} + 0.01 \mathcal{L}_{\text{wall}}. \tag{14}$$

$$\mathcal{L}_{\text{res}} = |\mathcal{P}|^{-1} \sum_{v \in \mathcal{P}} (\hat{\Delta}_v - \Delta_v^\star)^2, \qquad \mathcal{L}_\rho = E_{\mathcal{P}}, \quad \mathcal{L}_{\text{top}} = E_{\mathcal{P}_{\text{top}}},$$

$$\mathcal{L}_N = \frac{|N[\hat{\rho}] - N[\rho^\star]|}{N[\rho^\star] + \epsilon}, \qquad \mathcal{L}_{\text{hier}} = |\mathcal{P}|^{-1} \sum_{v \in \mathcal{P}} (c_v - \text{LP}(\Delta^\star)_v)^2.$$

Here $\mathcal{P}_{\text{top}}$ contains the highest-density 10% of pore voxels, c is the coarse residual branch, and LP averages over 4^3 voxels before upsampling. The weight vector is $(0.25, 1.0, 0.5, 0.25, 0.05)$ for cDFT phase 1 and initial CH_4/Xe adaptation, and $(0.5, 2.0, 1.5, 0.25, 0.05)$ for cDFT phase 2. The wall term $\mathcal{L}_{\text{wall}} = |\overline{\mathcal{P}}|^{-1} \sum_{v \notin \mathcal{P}} \hat{\rho}_v^2$ penalizes predicted density outside the pore region. Log-domain stabilization affects the residual targets; the density and uptake terms use the physical reference field.

B.2 Pre-training and Initial GCMC Adaptation

The cDFT operator is learned from solver-converged density labels. Initial CH_4/Xe adaptation freezes the backbone and updates rank-4 LoRA adapters in attention/feed-forward layers together with conditioning and output modules (Hu et al., 2022). Both stages use AdamW (Loshchilov & Hutter, 2017), warmup followed by cosine decay, and BF16 training. The optimizer budgets are summarized in Table 6; checkpoints are selected using validation performance. The

Table 6: Base and initial-adaptation training. Budgets are maximum optimizer updates; effective batch sizes count density states.

Stage	Initialization	Updated modules	Peak LR / budget	Batch
cDFT phase 1	Scratch	Full operator	6×10^{-4} / 50k	64
cDFT phase 2	Phase 1	Full operator	2×10^{-4} / 30k	64
CH_4/Xe	cDFT	LoRA, conditioning, heads	5×10^{-4} / 50k	64

molecular-conditioning triplet changes from the PC-SAFT registry to the GCMC reference registry: $(1, 3.73, 148.0)$ for CH_4 and $(1, 4.10, 221.0)$ for Xe, with lengths in Å and energies in K. Bulk density is obtained from the PC-SAFT equation of state. All states of each framework remain within one data partition (Section F).

B.3 Three-species Adaptation and CO_2 Refinement

Shared adaptation. The three-species operator continues from the frozen cDFT backbone with rank-32 LoRA and gas-balanced sampling of CH_4/Xe/CO_2 fields. It uses density, high-density, normalized coarse-shape, uptake, and log-uptake terms with weights $1/0.5/1/0.5/0.5$. Training uses effective batch 64 and BF16. The final shared continuation has a 10k-update budget, base learning rate 2×10^{-4}, and 2% warmup. Learning-rate multipliers are 0.25 for shared LoRA, 0.5 for conditioning/output delta branches, and 1 for the molecular polar branch. The budget counts optimizer updates; validation selects a checkpoint within the run. Stage-specific schedules and trainable-module configurations accompany the released implementation.

Charge-derived features. The CO_2 extension combines the non-polar PC-SAFT/UFF input potential with REPEAT framework charges (Campañá et al., 2009) and the rigid-molecule descriptor $(q_C, q_O, d_{CO}) = (0.6512e, -0.3256e, 1.149\,\text{Å})$. Signed, positive, and negative charge-deposition channels and a periodic unit-probe potential are constructed on a 64^3 grid and aligned with the 32^3 token grid. Global charge statistics, a spatial charge encoder, block modulation, attention adapters, and electrogeometry features provide charge-conditioned density refinement. The last stage learns an additional six-block feature-residual branch with expanded width 3,072 while freezing the inherited modules.

From framework charges to token features. The four spatial inputs are signed, positive, and negative REPEAT charge deposits and the dimensionless potential $\beta\phi$ of a unit-charge probe. Deposits are constructed in fractional-cell coordinates and coarse-grained from 64^3 to 32^3, so their grid indices align with the backbone tokens. The charge channels are compressed by an inverse-hyperbolic-sine transform, while the probe-potential channel is scaled and clipped. A periodic convolutional encoder maps the four channels to 32 features per token using three residual blocks. A linear projection adds these encoded features to the stem tokens before the first Swin block.

Two additional local descriptors summarize the magnitude of the discrete probe-potential gradient and the traceless part of its second derivative, using periodic differences and cell geometry. A two-layer projection adds their embedding to the spatial charge encoding. These descriptors supply local electrostatic geometry alongside the LJ reference; the model still predicts an orientation-marginalized molecule-center density. Separately, six global framework statistics—atom count, total absolute charge, minimum and maximum charge, charge range, and mean absolute charge—condition a bounded, spatially uniform log-density correction.

Where charge information enters the backbone. At each of the six Swin blocks, a charge-conditioned residual branch combines normalized backbone tokens with the aligned charge encoding. In a compact notation, its update has the form

$$h_j \leftarrow h_j + \mathbf{1}_{s=\mathrm{CO_2}} W_j^{\mathrm{up}} \,\mathrm{SiLU}\big[(1 + \tanh \gamma_j(e)) \odot W_j^{\mathrm{down}} \,\mathrm{LN}(h_j) + b_j(e)\big] , \tag{15}$$

where e is the encoded charge grid and the affine modulation (γ_j, b_j) is predicted from its layer-normalized features. The inherited post-block branches have width 32. Separate width-16 adapters generate additive query/key and value corrections inside attention, so charge information affects

both attention weights and the features being aggregated. The final refinement adds another residual branch of width 3,072 after each block, with the inherited backbone and earlier branches frozen. The species indicator bypasses these charge updates for CH_4/Xe.

Spatial correction and amplitude. The density decoder receives the charge-conditioned tokens. In addition to its fine/coarse reconstruction, a spatial head concatenates each final token with its charge encoding and predicts a 4^3-voxel log-density correction. The correction is bounded by $3\tanh(u/3)$. After exponentiation and clipping, the spatially corrected field is multiplied by the ratio of its inherited predicted pore integral to its corrected predicted pore integral. This assigns the spatial head a shape-refinement role while preserving the inherited predicted uptake at that interface. Both integrals use predictions and the input pore mask. The resulting field, including the inherited global charge correction, is ρ_{pre} in Equation (16). Table 7 summarizes the inputs and insertion points of these modules.

Table 7: Data flow through the CO_2 extension. Spatial features align with the 32^3 token grid; the final output remains a 128^3 physical-density field.

Module	Input	Output / insertion point
Spatial charge encoder	Four charge/probe-potential grids	32-channel encoding; projected into stem tokens
Electrostatic geometry	Periodic potential derivatives, cell	Local features added to the charge encoding
Block and attention adapters	Tokens and aligned charge features	Post-block residuals and query/key/value corrections
Spatial density head	Final tokens and charge encoding	Bounded voxel correction; predicted pore integral preserved
Global charge branch	Six framework charge statistics	Uniform log-density correction before spatial decoding
Final amplitude head	Predicted-field/token summaries and physical inputs	One bounded scalar multiplying ρ_{pre}

Final physical density. Let ρ_{pre} be the positive field reconstructed from the charge-conditioned backbone and density decoder. A learned amplitude refinement gives

$$\hat{\rho}(\mathbf{r}) = \rho_{\mathrm{pre}}(\mathbf{r}) \exp(\mathbf{1}_{s=\mathrm{CO_2}} a_\phi(\mathbf{x})), \qquad a_\phi(\mathbf{x}) = 2\tanh(f_\phi(\mathbf{x})/2). \tag{16}$$

The standardized features $\mathbf{x}$ summarize predicted tokens/density and framework/condition-derived inputs; they contain no reference uptake. The positive multiplier is applied only to CO_2. For every GCMC species, uptake and normalized Tanimoto are computed from this same final prediction. Normalization is used for the shape metric; the uptake integral retains the physical density scale.

The feature vector $\mathbf{x}$ includes five summaries of final tokens: their uniform mean, pore-weighted mean, predicted-density-weighted mean, standard deviation, and Boltzmann-weighted mean. Scalar features include predicted density concentration, pore fraction, pressure, bulk density, cell lengths/angles, global charge statistics, potential-threshold occupancies, and weighted moments of local charge and electrostatic-geometry fields. Feature means and scales are fixed from training data. A two-hidden-layer, width-64 MLP predicts the single scalar a_ϕ for each framework–condition state. Applying this scalar uniformly preserves the normalized spatial shape of ρ_{pre}, while changing its integral. During final feature refinement, gradients pass through this frozen amplitude head to the newly trained token branches. Thus the final stage can improve both shape through token/decoder features and amplitude through the features presented to the fixed head.

Refinement objective. With $M_v = \mathbf{1}[\beta V_v^{\mathrm{input}} < 50]$, define pore-normalized shapes $p_v = M_v\hat{\rho}_v / \sum_u M_u\hat{\rho}_u$ and $t_v = M_v\rho_v^\star / \sum_u M_u\rho_u^\star$. The final-stage objective uses integrated uptake U in $\mathrm{cm}^3(\mathrm{STP})/\mathrm{g}$ and within-framework pressure differences:

$$\begin{aligned}\mathcal{L}_{\mathrm{ref}} &= H_{0.05}(\hat{U}/20, U/20) + 0.2\mathcal{L}_{\mathrm{coarse}} \\ &\quad + 0.1H_{0.05}(\Delta\hat{U}/20, \Delta U/20) + 0.3\|p - t\|_1, \\ \mathcal{L}_{\mathrm{coarse}} &= \|\bar{p} - \bar{t}\|_1 + 0.2[1 - \cos(\bar{p}, \bar{t})].\end{aligned} \tag{17}$$

Here $H_{0.05}$ is smooth-L_1 with beta 0.05; bars denote pooling by four followed by normalization. Batches contain four frameworks with their available pressure pair, augmented by synchronized periodic translations. This stage uses AdamW at 10^{-4}, cosine decay, and a 12,800-update budget. The reported three-species MOF results use a single validation-selected model fixed before its reported evaluation; charge branches are bypassed for CH_4/Xe. The CH_4 screening and original spatial-site experiments use the earlier CH_4/Xe-adapted model, as specified with their results.

Training responsibilities. The shared three-species continuation updates rank-32 adapters and conditioning/output adjustments while retaining the cDFT backbone weights. Gas-balanced sampling keeps the minority CO_2 task represented. Charge-conditioned branches are then trained on the CO_2 labels: global statistics establish an amplitude correction; spatial and block/attention branches introduce local charge dependence; the pooled-feature head learns its bounded multiplier. The final stage optimizes only the newly added six token-residual branches through the inherited decoder and amplitude head using Equation (17). The inherited modules participate in the forward and backward computation but their parameters stay fixed. Validation uptake error selects the final model; test and screening predictions use that fixed selection. The supplied module configuration specifies the active branches and frozen parameters, without requiring internal run names to define the method.

C cDFT Warm-Start Protocol

Two-stage label recovery. Bootstrapping targets the failed states of the cDFT generation campaign. Predicted physical densities initialize the original solver at the same framework, species, pressure, and temperature. A converged field is accepted under the stage's convergence criterion; the second-stage residual threshold is 10^{-5}. The two-stage recovery result is the deduplicated union of successful state identities. The aggregate recovery is reported in Section 4.5; Table 23 evaluates models trained with the expanded labels.

Connection to the training labels. Starting from the available solver-converged density set, we train a density operator and predict initial fields for failed states. The original cDFT solver then iterates from these fields; only accepted solver-converged outputs become new supervision. Retraining on the enlarged set supplies the operator for the next recovery round. Framework partitions remain fixed, and newly accepted labels enter the corresponding partition rather than creating a new random split. The progression experiment in Table 23 evaluates the operators before recovery, after one recovery round, and after two rounds on the same held-out states.

The main cDFT benchmark uses the operator trained with the one-bootstrap label set. Its two optimizer phases in Table 6 are successive learning-rate phases on that training corpus; they do not denote two recovery rounds. The two-bootstrap operator is reported separately in the progression experiment. Likewise, the cumulative dataset recovery statistic and the independently initialized solver panel below answer different questions: additional label availability and solver cost, respectively.

For the fixed failure cohort, the first stage contributes 12,601 recovered states and the second adds 11,001 distinct states, for 23,602 in the union. The second-stage acceptance checks the convergence flag, finite output/residual, residual at most 1.01×10^{-5}, and a maximum density at least $0.1\rho_{\mathrm{bulk}}$. The learned field supplies the initial condition without reference-uptake rescaling. Stage-specific solver controls and acceptance functions are retained in the companion methods/configuration files.

Controlled initialization panel. The acceleration experiment uses nine adsorbates on the same 20 frameworks at 10 bar, giving 180 independently initialized states. Boltzmann and MADField initializations are compared with identical damping schedule, convergence tolerance, iteration cap, and external potential. Pressure continuation is disabled. Iteration and solver-time summaries use states converging under both initializations, and speedup is the ratio of their medians. Solver-only timing excludes neural initialization and preprocessing. The panel results are reported in Section 4.5.

D Architecture Details

D.1 Inputs and Periodic Backbone

The base operator configuration is summarized in Table 8. The operator represents each periodic unit cell on a 128^3 voxel grid. Two input channels are

$$\tilde{V} = \frac{\mathrm{clip}(\beta V_{\mathrm{ext}}, -15, 60)}{15}, \qquad \tilde{\rho}_B = \frac{\mathrm{clip}(\log \rho_{\mathrm{bulk}} - \beta V_{\mathrm{ext}}, -30, 5)}{15}. \tag{18}$$

The first encodes framework geometry and interaction wells; the second supplies a condition-dependent density scale. The conditioning vector contains $\log \rho_{\mathrm{bulk}}$, $\log P$, three log cell lengths, three cell-angle deviations, and the molecular triplet. Fixed normalization is applied before AdaLN conditioning (Peebles & Xie, 2023). Physical cell geometry therefore remains available despite the common voxel resolution.

A two-stage convolutional stem with circular padding produces 32^3 tokens. Six Swin blocks (Liu et al., 2021) form a shallow encoder/decoder with three skip projections. Shifted-window indexing wraps periodically, and integer-harmonic sine/cosine features encode token positions on each axis. The blocks use eight-head attention, SwiGLU feed-forward layers (Shazeer, 2020), and condition-dependent adaptive normalization.

Position convention. For an axis with n tokens, the integer index $i \in \{0, \ldots, n-1\}$ corresponds to fractional coordinate i/n. Positional channels use

$$\mathrm{PE}_k^{\sin}(i) = a_k \sin(2\pi k i/n), \qquad \mathrm{PE}_k^{\cos}(i) = a_k \cos(2\pi k i/n), \tag{19}$$

with integer harmonics through the Nyquist limit; repeated harmonics have decreasing amplitudes a_k. Channels from the three axes are concatenated. Consequently, translating an index by n leaves its encoding unchanged. The physical metric is supplied separately through the cell-conditioning inputs. Circular stem padding and periodic window shifts let opposite faces communicate without treating the cell boundary as an empty wall.

The encoder and decoder operate at the same token resolution. Skip projections concatenate corresponding encoder and decoder features and project them back to width 288. This preserves local information for voxel reconstruction while the attention blocks propagate information across windows.

D.2 Density Reconstruction

Fine residual, gate, and coarse heads reconstruct the density from tokens:

$$\log \hat{\rho}_\theta = \log \rho_B + \sigma(g_\theta)\left[r_\theta + \lambda \mathrm{Up}(r_\theta^{\mathrm{coarse}})\right]. \tag{20}$$

The fine and gate heads predict 4^3 voxel patches; the coarse branch is interpolated from the token grid. The base log-density is clamped to $[\log \rho_{\mathrm{bulk}} - 20, \log \rho_{\mathrm{bulk}} + 10]$. The CO_2 refinement in Equation (16) is subsequently applied to the positive reconstructed field. The base operator has 6.35M parameters; the three-species extension has 21.36M, including 11.84M updated in final refinement.

Here σ is the sigmoid, λ is an unconstrained learned scalar initialized to 0.3, and Up denotes trilinear interpolation from the token grid to the density grid. Each fine-head token predicts a distinct 4^3 voxel patch; the coarse head predicts one value per token, providing smoothly varying cell-scale structure. Their sum is modulated by a voxelwise gate before it corrects the Boltzmann log-density.

The coarse-head weights and bias start at zero. The gate-head weights also start at zero, with bias -2, so its initial sigmoid value is approximately 0.119. The fine residual uses small random weights and zero bias. Together these choices keep the initial density near the Boltzmann reference while permitting gradients to train the fine and coarse corrections. They specify the base density operator; the absolute-log-density ablation uses a different output parameterization (Section I.1).

Table 8: Base density-operator hyperparameters.

Component	Configuration
Input / token grid	128^3 / 32^3
Stem	Two stride-2 convolutions; channels $2 \to 48 \to 288$
Backbone	Width 288; six blocks; eight attention heads
Window / shift	8^3 tokens / four tokens per axis
Encoder / decoder	Three blocks each; three skip projections
Feed-forward	SwiGLU; inner width 384
Conditioning	11 inputs; AdaLN MLP $11 \to 128 \to 128 \to 1728$
Fine / gate / coarse head	$288 \to 64$ / $288 \to 64$ / $288 \to 1$
Drop path	Linear schedule to 0.1

E PC-SAFT Parameters

The non-polar adsorbate registry is given in Table 9. The UFF framework parameter file (Rappé et al., 1992) and complete registry are supplied with the potential-generation code; the mixing rule is given in Section A.2.

Table 9: PC-SAFT non-polar parameters for the nine adsorbates. The archived registry combines the Esper records (Esper et al., 2023) with its separately attributed H_2 entry and the recorded non-polar CO_2 extension.

Adsorbate	m	σ (Å)	ε/k_B (K)	M_w (g/mol)
H_2	1.000	2.960	34.20	2.016
Ar	1.000	3.378	117.81	39.95
Kr	1.000	3.608	164.02	83.80
Xe	1.000	3.927	227.70	131.29
CH_4	1.000	3.701	150.07	16.04
N_2	1.238	3.300	89.41	28.01
C_2H_6	1.607	3.517	191.45	30.07
C_3H_8	1.986	3.624	209.09	44.10
CO_2	2.531	2.579	153.32	44.01

F Data Generation and Evaluation Sets

F.1 Sources and Framework Partitions

QMOF supplies the crystalline frameworks for cDFT supervision (Rosen et al., 2021); ARC-MOF supplies the published CH_4/Xe GCMC references (Burner et al., 2022; Burner, 2025a;b;c). JLA-Gardner amorphous carbons (Gardner et al., 2023) and PRAM disordered materials (Thyagarajan & Sholl, 2020) are used for zero-shot evaluation. Simulations treat rigid frameworks and single-component adsorption. Density labels use a periodic 128^3 cell grid; all gas/pressure states of a framework stay in one training, validation, or test partition.

Common cDFT density test support. The cDFT portion of Table 2 evaluates all three models on the same 3,366 test states across 289 frameworks. Checkpoints remain fixed by their validation selection rules, including maximum validation-median similarity for SorbIIT; test scores are not used for checkpoint selection. The comparison uses DeepAPD's native pressure heads: 1/100 bar for H_2, 1/65 bar for CH_4, and 1/10 bar for the other seven species. We intersect this support with converged references and SorbIIT's native isotherm support, which requires at least three converged pressure states per framework–species pair for quadratic fitting. This support rule excludes 126 framework–species pairs (169 states) from the full 28,692-state test pool before the pressure-head intersection. Similarities use the same 128^3 reference grid and are averaged arithmetically within each species, then equally across species. In the order H_2, Ar, Kr, Xe, N_2, CH_4, CO_2, C_2H_6, and

Table 10: Framework partitions and evaluation support. Reference test states count gas–pressure fields unless noted. The CO_2 scalar benchmark uses the 90-MOF common subset.

Source / reference	Train	Validation	Test	Evaluated states
QMOF / cDFT, full test	4,904	577	289	28,692
QMOF / cDFT, common density	—	—	289	3,366
ARC-MOF / GCMC, CH_4	16,426	1,932	967	1,934
ARC-MOF / GCMC, Xe	17,769	1,932	967	967
CO_2 density campaign	1,697	199	101	201 density fields
CO_2 scalar, common support	—	—	90	180 uptake states
JLA-Gardner / cDFT transfer	—	—	Zero-shot	52 shared density fields
PRAM / GCMC, CH_4	—	—	Zero-shot	138 uptake states

C_3H_8, the common test-state counts are 574, 498, 436, 236, 418, 346, 418, 272, and 168. The 28,692-state cDFT uptake benchmark remains separate.

PRAM model and physical inputs. The GCMC transfer row in Table 4 uses the final three-species composite model without PRAM fine-tuning or checkpoint selection on PRAM. Evaluation uses FP32 at 298.15 K and 1 bar, the CH_4 molecular-conditioning triplet $(1, 3.73, 148.0)$, and the PC-SAFT bulk number density. The CO_2-specific charge branch is disabled for methane. Inaccessible pores are encoded in the external-potential grid; the dimensionless potential is clipped to $[-2100/298.15, 50]$. Uptake is obtained by summing the final predicted absolute field over the full grid and converting with the framework-cell volume and mass. The four material-family errors are averaged equally over the fixed 138-state PRAM cohort.

The PRAM evaluation comprises 75 PIM, 44 HCP, 16 kerogen, and three Bhatia-carbon states. GCMC density-baseline comparisons use the same field IDs per task. Species and material-family aggregation are defined in Section H. Exact framework partitions and the reference-generation configurations accompany the data.

Campaign and supervised states. The cDFT campaign records 628,668 attempted framework–gas–pressure states, of which 566,193 converged under the reference generation settings and 62,475 form the fixed failure cohort. The planned benchmark release comprises these 566,193 converged states. Attempted states include failed solves and therefore are not all density labels. Training consumes solver-converged fields in its assigned framework split; the model-progression experiment distinguishes the label sets before and after recovery as described in Section C. These simulation counts describe the generation campaign, whereas Table 10 identifies the held-out support used to report model errors.

Partitions are defined at framework-ID level before expanding gas and pressure conditions. Stored symmetry variants retain their parent framework's assignment. Structural matching was used to check the QMOF partitions after construction; the training split is not defined by a newly imposed clustering threshold. Exact IDs in the companion manifest specify the evaluated split. The CO_2 screening exclusion remains specific to CO_2 supervision.

F.2 cDFT and Published GCMC References

The cDFT campaign uses the non-polar PC-SAFT functional in Section A.2 at 298.15 K. One external-potential grid is reused across pressures for each framework/adsorbate pair. Isotherms are solved from low to high pressure, starting from the nearest converged lower-pressure field when available and otherwise from the Boltzmann density. Failed states are excluded from initial supervised training and retained for recovery. The pressure grids are given in Table 11.

The published ARC-MOF CH_4/Xe densities and uptake references define the GCMC adaptation targets. Voxel density is the ensemble-averaged molecule-center count divided by voxel volume, $\rho_v = \langle n_v \rangle / \Delta V$. PRAM methane references are used only for transfer evaluation. Reference interactions and the molecular-conditioning inputs are distinguished in Sections B.2 and I.7.

Table 11: **cDFT pressure grids at $T = 298.15$ K.**

Fluid	# pressures	Pressure range
Ar, C_2H_6, C_3H_8, CO_2, Kr, N_2, Xe	12	0.1, 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 bar
CH_4	24	0.1, 0.5, 1, 2, 3, 4, 5, 5.8, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65 bar
H_2	30	0.1, 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100 bar

F.3 CO_2 Density Labels and Screening Pools

The framework CIFs for our CO_2 dataset are drawn from MOSAEC-DB (Gibaldi et al., 2025). We generated GCMC densities for a 2,000-MOF subset of a 19,152-MOF scalar-reference library at 298 K and 0.15/0.35 bar. We select these two pure-component pressures to represent CO_2 partial pressures for 15% and 35% flue-gas compositions at a total pressure of 1 bar, motivated by the coal-combustion and steel-production cases studied by Kwon et al. (2025). Spatial acceptance retains 3,969 fields from 1,997 frameworks. The framework partitions are in Table 10; their density-field counts are 3,373/395/201 for train/validation/test. Scalar references at both pressures give 3,394/398/202 states on those partitions.

Reference simulations use RASPA3 (Ran et al., 2024), UFF framework Lennard–Jones parameters (Rappé et al., 1992), REPEAT framework partial charges (Campañá et al., 2009), and rigid Garcia-Sanchez CO_2 (García-Sánchez et al., 2009). The Lennard–Jones cutoff is 12.8 Å with analytic tail corrections; electrostatics use Ewald summation. Scalar runs use 10,000 initialization and 10,000 production cycles. Density production uses 2×10^7 equilibration moves and up to 5×10^8 production moves, checking spatial agreement every 5×10^6 moves.

Density acceptance requires mean pairwise Tanimoto at least 0.75 between replicate-cell adsorption probability distributions (APDs). Accepted APDs are folded to the crystallographic cell. With APD weights a_v and simulated molecule count N_{cell}, absolute labels are constructed as

$$\rho_v^\star = \frac{a_v}{\sum_u a_u} \frac{N_{\text{cell}}}{\Delta V}, \qquad \sum_v \rho_v^\star \Delta V = N_{\text{cell}}. \tag{21}$$

This label construction combines spatial sampling with the scalar reference amplitude. Densities are stored in $\mathrm{mol/m^3}$; multiplying by $6.02214076 \times 10^{-7}$ gives Å^{-3}. MADField receives framework and thermodynamic inputs, not reference uptake, during prediction.

All 2,000 density-selected MOFs are excluded from the CO_2 screening pool, leaving 17,152 frameworks held out from CO_2 supervision. The six-model screening comparison uses a fixed 15,572-framework common support.

On that common support, the two fixed-threshold retrieval tasks contain 153 positives at 0.15 bar (uptake at least 9 mmol/g) and 118 at 0.35 bar (at least 12 mmol/g). These counts use the same candidate intersection as the AP results in Table 22. Cross-species framework membership and exact partitions are provided with the data. The CH_4 GCMC-training-excluded screening cohort is defined in Section I.4.

G Baseline Implementations

We compare scalar uptake models and spatial-density models using their task-appropriate inputs and conditioning. The compact specifications in Table 12 distinguish pressure-specific models from conditional predictors. Scalar baselines are trained on uptake labels; density baselines are compared using normalized-density similarity. Their detailed optimizers, validation-selection settings, graph construction, and environment configurations accompany the implementation.

Table 12: Baseline inputs and task adaptation. Explicit framework charge channels are provided to MADField's CO_2 extension, not these baselines.

Model	Main input / output	Conditioning and initialization
RetNet	External-potential voxels / uptake	Per-condition CNN; scratch for original tasks, transferred for CO_2.
MOFTransformer	Atomic structure and energy grid / uptake	Pretrained encoder; per-condition fine-tuning, with CH_4 initialization for CO_2.
DeepSorption	Atomic structure and gas parameters / uptake	Per-gas structure model; pressure outputs or condition-specific tasks.
Uni-MOF	Structure, gas identity and pressure / uptake	Released pretrained encoder; pressure-conditioned adaptation.
IsothermNet	Textural/energy descriptors, gas and pressure / uptake	Scratch conditional isotherm model.
DeepAPD	Atom/probe graphs / normalized density	cDFT task training; released CH_4/Xe models; new CO_2 training.
SorbIIT	External-potential grid / density representation	Isotherm-based reconstruction; task-specific density adaptation.

Scalar representations and targets. Table 13 summarizes the CO_2 training configurations for RetNet, DeepSorption, MOFTransformer, and Uni-MOF. RetNet uses a three-dimensional convolutional network on external-potential voxels, with one model per gas–pressure condition. Voxel transforms and standardization are fitted on training inputs; its regression target is uptake in physical units and its objective is MSE. The original task training starts from scratch. For CO_2, the two models instead transfer the CH_4 5.8/65-bar backbones and reinitialize the final regression layer. This changes initialization, not the fact that the downstream labels are CO_2 uptake.

DeepSorption encodes atom types, coordinates, and cell information with its sequence-transformer implementation. A gas-specific model outputs a vector over the pressure grid; missing reference states are masked out of the loss. Targets are standardized separately by pressure using valid training labels, and predictions are transformed back before computing MAE. CO_2 uses two outputs, a transferred CH_4 backbone, and a reinitialized output head. Validation selection uses the native robust relative-error criterion on reference loadings above 0.05 mol/kg.

MOFTransformer combines an atomic graph, the UFF interaction-energy grid, and cell geometry. Because the regression head does not take pressure or gas identity, each evaluated condition has its own checkpoint. Training uses standardized uptake MSE and restores physical units with that training split's mean and standard deviation. Original tasks start from the released pretrained encoder; the CO_2 pressure models use the corresponding adapted CH_4 weights as initialization. Selection uses validation MAE.

Uni-MOF is conditional on gas attributes, pressure, and temperature, so one CO_2 model predicts both pressures. It uses the pretrained atom-pair transformer and standardized $\log(1+q)$ targets, with the corresponding inverse transform at inference. The reported CO_2 run computes the target normalizer from its 1,697 training frameworks and selects the checkpoint by validation R^2. IsothermNet combines structural/bond, interaction, and textural descriptors with gas and thermodynamic conditions. Its scalar target is $\log(q + 10^{-3})$, optimized by MSE and converted back to physical uptake for evaluation. Neither model receives reference uptake as a prediction input.

The scalar models retain their native representations and target transforms. Their optimization criteria therefore need not equal the final physical-unit MAE; Section H defines the common reporting metric and Table 10 its evaluation support.

Density reconstruction and uptake dependence. Original SorbIIT reconstructs density from predicted isotherm coefficients and supplied global adsorption information. Our CO_2 labels cover only one temperature and two pressures, rather than the original multi-temperature isotherm setting; we therefore adapt its CellUNet to predict density directly, training a separate model at each pressure without reference uptake as input. The reported SorbIIT GCMC results use direct-density adaptations (including the earlier CH_4/Xe runs), not a reproduction of the original coefficient-

Table 13: Compact CO_2 scalar-training recipes for RetNet, DeepSorption, MOFTransformer, and Uni-MOF. Native selection rules are preserved; the epoch budget is a maximum except for RetNet's fixed final-epoch output.

Model	Optimizer / peak LR	Budget / batch	Checkpoint criterion
RetNet	Adam / 10^{-3}	60 epochs / 256	Final epoch; StepLR every 10 epochs
DeepSorption	Adam / 10^{-4}	200 epochs / 64	Validation robust relative error; early stopping
MOFTransformer	AdamW / 10^{-4}	40 epochs / 64	Validation MAE; 5% warmup
Uni-MOF	Adam / 3×10^{-4}	50 epochs / 64	Validation R^2; 6% warmup

reconstruction method. We evaluate these adaptations and DeepAPD on normalized-density similarity.

CO_2 adaptation. All scalar models are trained on the new CO_2 uptake labels using their native representations; two-pressure outputs or pressure conditioning follow each architecture. Uni-MOF uses training-only target-normalization statistics. DeepAPD is newly trained on the CO_2 density labels and sums the losses of available pressure heads. SorbIIT is trained separately at the two pressures using direct density supervision. Density scores are computed on the 201 accepted test fields, averaging the two seed scores per field. Checkpoints and inputs for each reported task are mapped in the accompanying configurations.

Density-baseline task recipes. DeepAPD's cDFT comparison evaluates task-trained density predictions for the nine species using normalized similarity, whereas the CH_4/Xe GCMC comparison uses the released multitask model and its three gas–pressure heads. The new CO_2 experiment starts from random weights, using the native atom/probe PaiNN model and two of its existing task heads. It minimizes negative normalized Tanimoto, summing the available pressure-head losses for a framework. Adam uses learning rate 5×10^{-5} with a maximum of 30 epochs; the native last-task validation rule selects on the 0.35-bar validation fields. Test predictions cover every voxel on the 128^3 grid, and target densities enter only the metric after inference.

SorbIIT's cDFT adaptation uses one model per fluid. It predicts voxelwise coefficients of a fixed-temperature quadratic isotherm and reconstructs density at the requested pressure. Training combines coefficient error, density error, integrated-loading error, non-negativity, and amplitude regularization with weights $0.1/1/0.1/0.05/0.05$. Its external-potential transform follows the original fixed-temperature convention, and grid resampling preserves the cell integral. The direct-density GCMC adapters instead predict a field for each pressure condition. For CO_2, the input is the potential in K clipped to $[-10^4, 10^4]$ and divided by 10^3, together with cell geometry; the target is absolute density in $\mathrm{mol/m^3}$ divided by 50,000. Full-grid L_1 training uses Adam at 10^{-3} for 20 epochs, batch one with accumulation four, and minimum validation L_1 selection. Undoing this fixed target scale differs from normalizing a field to unit mass. The reported shape metric applies unit-mass normalization after prediction.

We report SorbIIT adaptations as density baselines only; this evaluation scope does not imply that their fixed-scaled outputs lack amplitude information. Explicit REPEAT atomic charges and derived charge grids are supplied only to MADField's extension, while the baselines retain their structural or interaction inputs. All models' CO_2 labels nevertheless come from the same charge-aware GCMC reference physics.

H Evaluation Metrics

Uptake and integration support. For voxel volume ΔV, the molecular count estimator is $\hat{N} = \sum_v M_v \hat{\rho}_v \Delta V$. The main cDFT and CH_4/Xe estimates integrate the full cell ($M_v = 1$); CO_2 uses the accessible-pore mask $M_v = \mathbf{1}[\beta V_{\mathrm{ext}}^{\mathrm{input}} < 50]$. Gravimetric uptake is $q = \hat{N} V_m^{\mathrm{STP}} / M_{\mathrm{cell}}$, and volumetric uptake is $q_{\mathrm{vol}} = \hat{N} V_m^{\mathrm{STP}} / (N_A V_{\mathrm{cell}})$, where M_{cell} is the molar mass of the framework unit cell, with consistent length and mass units. We use $V_m^{\mathrm{STP}} = 22,414\,\mathrm{cm^3/mol}$ and report

gravimetric MAE in $\mathrm{cm}^3(\mathrm{STP})/\mathrm{g}$. Benchmark scalar references retain their recorded units after conversion; separately masked diagnostics are identified with their results.

Density similarity and physical error. Normalize each full-grid field to unit mass, $p_v = \rho_v / \sum_u \rho_u$. The Tanimoto coefficient is

$$T(\hat{p}, p^\star) = \frac{\sum_v \hat{p}_v p_v^\star}{\sum_v \hat{p}_v^2 + \sum_v (p_v^\star)^2 - \sum_v \hat{p}_v p_v^\star}. \tag{22}$$

Physical-density errors retain the original amplitude and use the stated pore mask on both prediction and reference: MAE is $|M|^{-1} \sum_{v \in M} |\hat{\rho}_v - \rho_v^\star|$ and relative L_1 is $\sum_{v \in M} |\hat{\rho}_v - \rho_v^\star| / \sum_{v \in M} \rho_v^\star$. Density MAE is reported in $10^{-3}\ \text{Å}^{-3}$ and relative L_1 in percent.

Aggregation. We first average state-level errors or similarities within each species. **Mean** then averages species equally: nine species for cDFT, three for the main GCMC benchmark, and two for controlled CH_4/Xe studies. PRAM weights its four material-family means equally. For density baselines with two seeds, per-state metric scores are averaged before species aggregation.

Screening. Models rank candidates by predicted working capacity or uptake for the specified task. We distinguish top-fraction retrieval from fixed-threshold retrieval. In Figure 4, a top-fraction target set contains the $K = \lceil f N_{\mathrm{cand}} \rceil$ highest-reference candidates for $f \in \{0.001, 0.01, 0.1\}$; reference-value ties are resolved by framework ID. Cumulative recall is the fraction of this fixed target set recovered as the prediction-ranked screening budget grows. In Tables 21 and 22, targets instead satisfy the stated physical threshold. For these fixed-threshold tasks, AP groups tied prediction scores and sums precision weighted by recall increments. Retrieval budgets use descending prediction order, with framework ID as the tie-breaker. The full-recall budget is the smallest ranking prefix containing every reference-defined target in the evaluated library. A CH_4 candidate evaluation comprises the two pressures needed to obtain its working capacity. Candidate-count ratios compare follow-up evaluation budgets; neural and reference-simulation timing is reported separately.

I Additional Experiments

I.1 cDFT Model Variants

We compare scalar uptake regression, removal of the Boltzmann residual, and an FNO backbone on the same 28,692 cDFT test states. Table 14 uses the saved BF16 inference records for every variant and one common reference: uptake integrated within $\beta V_{\mathrm{ext}} < 50$. All three density models, including the full density operator, integrate predictions over that same mask; scalar predictions are compared with the same reference uptake. We average absolute errors within species and then average species equally. This controlled evaluation convention is distinct from the main FP32/full-cell cDFT benchmark. Under this common evaluator, density-field prediction reduces mean uptake MAE from 7.47 to $1.31\ \mathrm{cm}^3(\mathrm{STP})/\mathrm{g}$, an approximately $5.7\times$ reduction. Evaluator configurations accompany the results.

Table 14: **Ablation study on cDFT uptake prediction.** Errors are MAE of gas uptake in cm^3(STP)/g on the held-out cDFT test set. **Density Head** is the full MADField-cDFT model that predicts $\rho(\mathbf{r})$ and integrates it to obtain uptake. **Scalar Head** uses the same backbone but directly predicts uptake. **w/o Boltzmann residual** removes the Boltzmann-residual output parameterization. **FNO baseline** replaces the Swin backbone with a Fourier neural operator baseline.

Configuration	H_2	Ar	Kr	Xe	N_2	CH_4	CO_2	C_2H_6	C_3H_8	**Mean**
Scalar Head	2.34	2.20	4.71	10.90	1.81	8.77	8.18	10.90	17.40	7.47
w/o Boltzmann residual	0.15	0.21	0.72	3.28	0.17	1.26	1.46	2.15	4.55	1.55
FNO baseline	0.23	0.65	2.98	9.76	0.44	5.31	6.43	8.01	16.08	5.54
Density Head	**0.07**	**0.11**	**0.47**	**2.62**	**0.13**	**0.79**	**1.19**	**1.92**	**4.47**	**1.31**

Variant definitions. The scalar head mean-pools the final tokens and applies a $288 \to 144 \to 1$ MLP with GELU to predict log uptake; exponentiation yields positive scalar uptake. It replaces the voxel decoder in the prediction path. The no-Boltzmann variant keeps the Swin backbone but directly predicts absolute log-density: its fine, intermediate, and upsampled coarse log-fields are averaged before exponentiation and the density clamp. Its training target is log density rather than the Boltzmann log-residual. The FNO variant retains density supervision and the Boltzmann-reference reconstruction but replaces window attention by spectral mixing and local channel mixing. These variants compare changes in target/head, reference parameterization, and backbone, respectively. The scalar output and density-integrated outputs are converted to the common reporting units after their respective prediction paths.

I.2 Multi-fidelity Ablations

The controlled CH_4/Xe experiment holds backbone capacity and framework partitions fixed while varying scalar/density supervision and cDFT pre-training (Table 15). The variants retain the width-288, six-block, eight-head backbone and fixed evaluation support, while their heads, objectives, and adaptation parameter sets follow the target formulation. The training definitions below specify these differences.

Table 15: Matched-backbone target/transfer comparison on the fixed CH_4/Xe test split. Uptake MAE is the equal-species mean in $\mathrm{cm}^3(\mathrm{STP})/\mathrm{g}$, after validation-only checkpoint selection by the same statistic. MF denotes cDFT pretraining followed by GCMC adaptation.

Model	cDFT target	GCMC target	Uptake MAE
Scalar scratch	None	Uptake	7.950
Scalar-MF	Uptake	Uptake	7.531
Density scratch	None	Density	2.152
Density-MF	Density	Density	1.996

Scalar and density supervision. Scalar-MF is first trained from random weights on *scalar cDFT uptake*, using the scalar head defined above; its GCMC phase starts from those scalar-pretrained weights. It does not initialize from a density-trained checkpoint. For log uptake $z = \log q$ in mol/kg, scalar training uses

$$\mathcal{L}_{\mathrm{scalar}} = \mathrm{mean}(\hat{z} - z)^2 + 0.05\,\mathrm{mean}\,\frac{|\hat{q} - q|}{\max(q, 10^{-6})}. \tag{23}$$

Small positive target floors stabilize the logarithm and relative term. cDFT scalar pretraining uses 50k and 30k optimizer phases at 6×10^{-4} and 2×10^{-4}, with effective batch 64. The GCMC adaptation uses rank-4 attention/feed-forward LoRA with trainable conditioning and uptake head, a 50k-update maximum at 5×10^{-4}, and 5% warmup. During training it tracks validation mean absolute log error, averaged equally over gases. Scalar scratch trains the same scalar architecture directly on GCMC labels from random weights, with 50k and 30k phases at 6×10^{-4} and 2×10^{-4}; the second phase starts from the best first-phase scalar checkpoint.

Density scratch and Density-MF use the voxel loss components in Section B.1, with weights $0.25/1/0.5/0.25/0.05$ for residual, density, high-density, uptake, and hierarchy, and an additional wall penalty of weight 0.01. Both have a 50k-update maximum and effective batch 64, using BF16, AdamW, 5% warmup, and cosine decay. Scratch updates the full model at 10^{-4} with weight decay 0.01; Density-MF starts from cDFT density pretraining and uses rank-4 adaptation at 5×10^{-4} with zero weight decay.

All variants are selected by equal-species absolute uptake MAE on the full 5,796-state validation split, without using test errors. The reported comparison uses matched backbone capacity and a fixed 2,901-state test split; target-specific heads and update configurations are summarized in Table 16. Scalar inference uses FP32 and density inference uses BF16.

Within each target formulation, pretraining changes the source of the initialization. Across scalar and density formulations, both the supervised observable and the output head change. The reported comparison therefore tests the complete supervised formulations at matched backbone capacity.

Table 16: Controlled target/transfer definitions for Table 15. Shared backbone capacity and test support do not imply identical trainable modules or optimizer budgets.

Variant	Initial training target	GCMC updates	Selection statistic
Scalar scratch	None; random weights	Full scalar model	Gas-mean absolute uptake MAE
Scalar-MF	Scalar cDFT uptake	Rank-4 adapters, conditioning, scalar head	Gas-mean absolute uptake MAE
Density scratch	None; random weights	Full density model	Gas-mean absolute uptake MAE
Density-MF	cDFT density fields	Rank-4 adapters, conditioning, density heads	Gas-mean absolute uptake MAE

Density-label efficiency uses nested subsets sampled by framework; all available gases and pressures of a selected framework stay together, and validation/test partitions remain fixed (Table 17). The pre-training gain grows as density supervision becomes sparse.

Table 17: Mean uptake MAE ($\mathrm{cm}^3(\mathrm{STP})/\mathrm{g}$) versus GCMC label fraction in the controlled training series.

Labels (%)	Density scratch	Density-MF	Error reduction (%)
10	4.715	2.782	41.0
20	3.297	2.292	30.5
50	2.423	2.068	14.7
100	2.152	1.996	7.2

Adaptation and reliability. Full fine-tuning achieves uptake MAE 1.930 with 6.354M trainable parameters, compared with 1.996 and 1.573M for rank-4 LoRA. For the controlled Density-MF model, 10,000 framework bootstrap resamples give a 95% interval of $[1.864, 2.130]$ for mean uptake MAE. Across three fine-tuning seeds with the pre-trained checkpoint fixed, mean uptake MAE is 2.018 ± 0.021.

Condition-level uncertainty. Table 18 resolves the overall interval by gas and pressure. Each of the 10,000 bootstrap replicates samples frameworks with replacement and keeps all conditions of a sampled framework together. We average state errors within each species, then average species equally in each replicate; the 2.5th and 97.5th percentiles give the interval. Seed variation instead changes fine-tuning randomness with pretraining and partitions fixed. These summarize different sources of variation in the controlled CH_4/Xe series.

Table 18: Framework-bootstrap intervals for the controlled Density-MF model. Uptake MAE is in $\mathrm{cm}^3(\mathrm{STP})/\mathrm{g}$.

Group	MAE	95% CI
Overall	1.996	[1.864, 2.130]
CH_4, 1 bar	0.290	[0.260, 0.322]
CH_4, 65 bar	1.725	[1.624, 1.829]
Xe, 1 bar	2.984	[2.744, 3.230]

I.3 Density Accuracy and Spatial Analysis

On the same 52 zero-shot JLA-Gardner CH_4 density fields, MADField-cDFT reaches Tanimoto 0.885, compared with 0.492 for SorbIIT and 0.422 for DeepAPD. Figure 6 shows representative crystalline-MOF and amorphous-carbon fields.

Table 19(a) reports unnormalized physical-density accuracy of the final three-species model on all 3,102 GCMC test states, using the pore mask and no reference-uptake rescaling. Panel (b) retains a

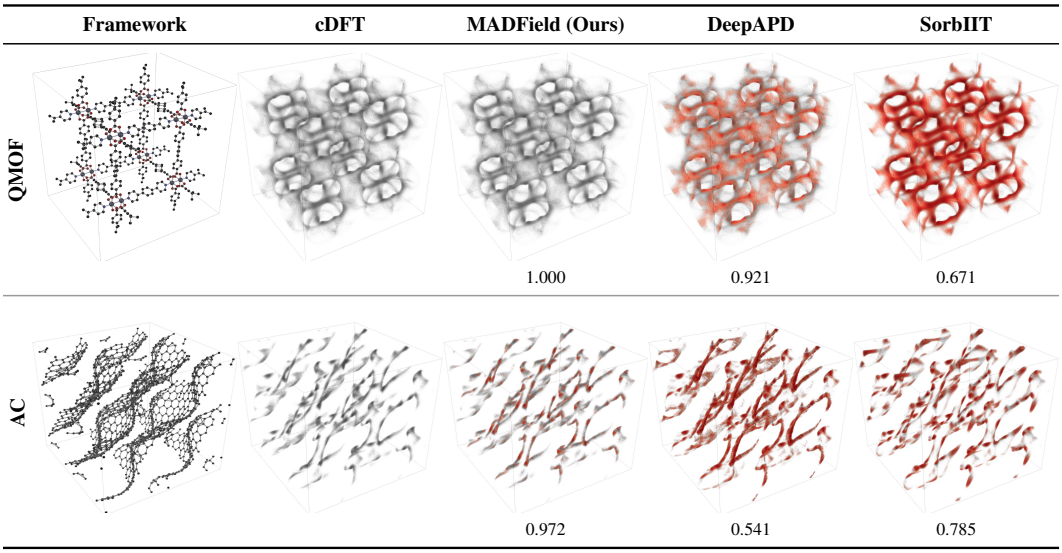


Figure 6: **Visualization of CH_4 density field and error at cDFT-level.** ID QMOF (top) and OOD AC (bottom): framework, **cDFT** reference density (gray), and model errors. Red marks larger normalized absolute error on a shared per-row scale; Tanimoto scores appear below the predictions.

separate controlled CH_4/Xe comparison: supplying DeepAPD with reference uptake fixes its amplitude but yields about $2.2\times$ higher physical-density MAE than Density-MF. This is an oracle shape diagnostic.

Table 19: Physical-density errors within $\beta V_{\text{ext}}^{\text{input}} < 50$. MAE is in 10^{-3} Å^{-3}; relative L_1 is in percent. (a) Final three-species model. (b) Controlled CH_4/Xe model and DeepAPD scaled by ground-truth uptake.

(a) Final three-species MADField-GCMC			
Gas	States	Density MAE	Relative L_1 (%)
CH_4	1,934	1.808	12.89
Xe	967	2.672	18.14
CO_2	201	7.292	61.78
Mean	3,102	3.924	30.94
(b) Separate ablation and oracle diagnostics			
Model	Reference	Density MAE	Relative L_1 (%)
Density-MF	GCMC (CH_4/Xe)	2.243	14.93
DeepAPD + GT uptake	GCMC (CH_4/Xe)	4.904	33.53

Peak detection and localization. We identify adsorption sites as local maxima after periodic Gaussian smoothing. The detector uses voxel standard deviation $\sqrt{0.4\,\text{Å}/\Delta x}$, where $\Delta x = \|\mathbf{a}\|/n_x$, an iterated 18-connected maximum-filter neighbourhood with radius parameter 0.45 Å, and a relative peak-height cutoff of 0.1. Predicted and reference maxima are converted to cell coordinates; periodic pair distances minimize over neighbouring images. Hungarian assignment minimizes the total pair distance, then pairs farther than 1 Å are discarded. F1 is averaged per framework, while localization distance first averages matched pairs within a framework and then frameworks with at least one match. Table 20 reports binding-site identification and localization accuracy.

Table 20: Binding-site F1 and matched localization distance on the same 967 ARC-MOF frameworks using the CH_4/Xe-adapted checkpoint. Distance is in Å; aggregation is defined above.

	F1		Distance (Å)	
Condition	DeepAPD	MADField	DeepAPD	MADField
CH_4, 1 bar	0.914	0.949	0.167	0.084
CH_4, 65 bar	0.799	0.936	0.261	0.112
Xe, 1 bar	0.847	0.921	0.215	0.120

I.4 Additional Screening Results

Candidate pools and screening setup. The screening checkpoint is the CH_4/Xe-adapted model, trained on GCMC CH_4 density labels at 1 and 65 bar. It is evaluated at 5.8 and 65 bar without additional 5.8-bar GCMC supervision. Scalar baselines use uptake supervision at both screening pressures. Candidate ranking uses the corresponding working-capacity predictions; model-specific pressure handling is documented with the screening configurations.

The main CH_4 screening cohort is constructed from the 270,583-framework ARC-MOF library. The GCMC training partitions contain 16,426 unique CH_4 and 17,769 unique Xe framework IDs; the former is a subset of the latter. Of their 17,769-ID union, 16,385 occur in this library. Excluding these IDs leaves $270,583 - 16,385 = 254,198$ candidates and 166 targets with reference WC at least $200\,\mathrm{cm}^3(\mathrm{STP})/\mathrm{cm}^3$. The exclusion is defined by membership in the GCMC training partitions. All models in Table 21 rank the same 254,198 candidates containing these 166 targets. The common-support screening set contains 15,572 frameworks held out from CO_2 supervision (Section F.3). Targets have uptake at least 9 mmol/g at 0.15 bar (153 targets) or 12 mmol/g at 0.35 bar (118 targets); each pressure defines a separate task.

Target definitions. The top-fraction benchmark in Figure 4 uses the true top 0.1%, 1%, and 10% within each task: 255, 2,542, and 25,420 targets among 254,198 CH_4 candidates, and 16, 156, and 1,558 targets among 15,572 CO_2 candidates at 0.15 bar. The fixed-threshold results below use $\mathrm{WC} \geq 200\,\mathrm{cm}^3(\mathrm{STP})/\mathrm{cm}^3$ for CH_4 (166 targets), and uptake thresholds of 9 and 12 mmol/g for CO_2 at 0.15 and 0.35 bar (153 and 118 targets). Thus the 255-member CH_4 top-0.1% set and the 166-member threshold set define different retrieval tasks. Their MADField-GCMC complete-recovery ranks both happen to be 16,600; the target sets and recall denominators remain distinct.

Screening inference. For IsothermNet, elements outside the checkpoint vocabulary use an all-zero element-category block while retaining structural and physical descriptors. U/Np Lennard–Jones descriptors use UFF parameters, with σ converted to the model's $r_{\min}$ convention; checkpoint weights and input dimensions remain unchanged.

For each CO_2 pressure, AP, top-fraction recall curves, and parity errors use the same predictions on the same candidate IDs. At 0.15 bar, recovering all 16 true top-0.1% targets requires 166 of 15,572 candidates for MADField-GCMC and 1,910 for Uni-MOF, the best scalar baseline at complete recall: 1.0660% versus 12.2656%, a $1910/166 = 11.506$-fold reduction.

Screening parity and full-range recovery. Figures 7 and 8 compare CO_2 uptake predictions on the 15,572-framework screening cohort. MADField-GCMC achieves MAE 6.79 and $7.79\,\mathrm{cm}^3(\mathrm{STP})/\mathrm{g}$ at 0.15 and 0.35 bar, respectively. These screening-cohort errors are distinct from the held-out test-set uptake MAEs in Table 3.

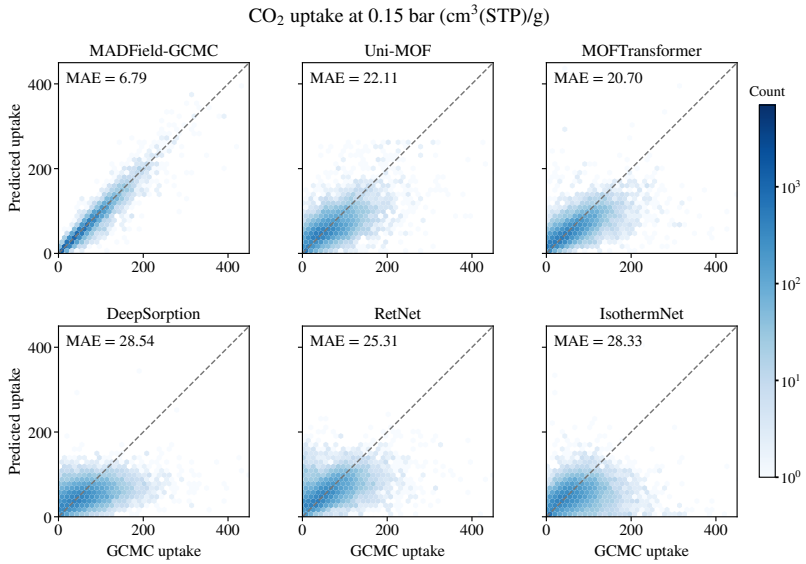


Figure 7: **CO_2 screening uptake parity at 0.15 bar.** Hexagonal bins show MOF counts on a logarithmic scale shared across models and both figures; the dashed line is $y = x$. Axes show 0–450 $\mathrm{cm}^3(\mathrm{STP})/\mathrm{g}$; MAEs use all 15,572 screening frameworks.

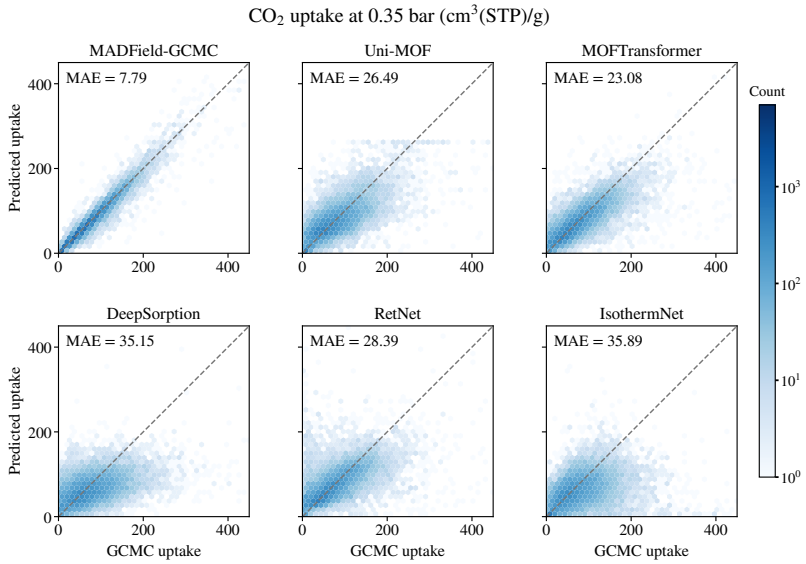


Figure 8: **CO_2 screening uptake parity at 0.35 bar.** The models, 15,572-framework cohort, bin-count scale, and metric conventions match Figure 7. Each panel reports full-cohort MAE in $\mathrm{cm}^3(\mathrm{STP})/\mathrm{g}$.

CH_4 working-capacity parity. Figure 9 compares predictions on the same 254,198 GCMC-training-excluded ARC-MOF candidates used in the main screening benchmark. MADField-GCMC achieves gravimetric WC MAE 5.81 $cm^3(STP)/g$ over the complete candidate set. This differs from the test-set uptake metric in Table 3; retrieval rankings remain based on volumetric WC.

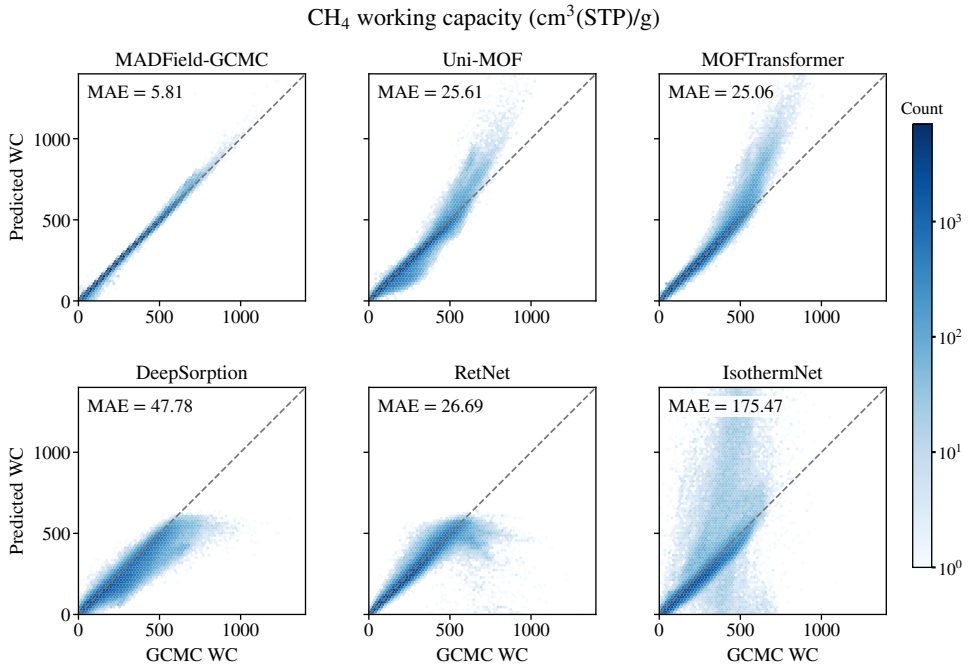


Figure 9: **CH_4 working-capacity parity on the screening cohort.** All six models are evaluated on the same 254,198 candidates. Hexagonal bins share a logarithmic count scale; dashed lines mark $y = x$. The displayed range is 0–1400 $cm^3(STP)/g$ on both axes; each MAE includes all candidates, including predictions outside that range.

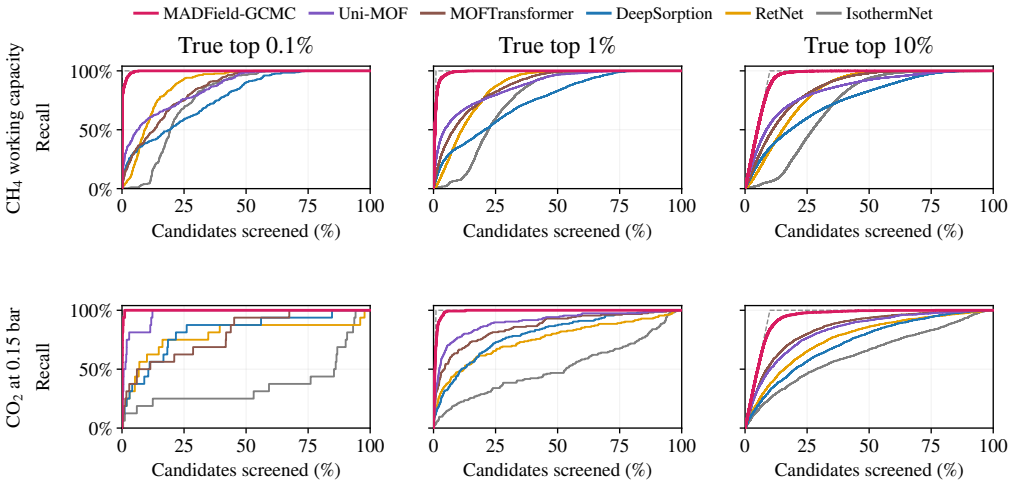


Figure 10: **Full-range cumulative recovery for Figure 4.** The same predictions, candidate sets, top-fraction targets, and ranking rules are shown over 0–100% of candidates screened, with a common horizontal scale across all panels. Rows show CH_4 working capacity and CO_2 uptake at 0.15 bar; columns target the true top 0.1%, 1%, and 10%. Gray dashed lines indicate ideal rankings.

CH_4 fixed-threshold retrieval. MADField-GCMC reaches at least 95% recall after 4,688 evaluations and complete recall after 16,600, corresponding to approximately 1.84% and 6.53% of the cohort. The complete-recovery evaluation budget is approximately $8.1\times$ smaller than MOFTransformer's and $8.0\times$ smaller than RetNet's, the smallest budget among the scalar baselines; compared with exhaustive evaluation of the 254,198-candidate cohort, it is $15.3\times$.

Table 21: Fixed-threshold CH_4 screening after excluding GCMC training frameworks. All models rank the same 254,198 candidates containing 166 targets with WC $\geq 200\,\mathrm{cm}^3(\mathrm{STP})/\mathrm{cm}^3$. The final column is the fraction needed for complete target recovery.

Model	Candidates	AP	100% recall (%)
MADField-GCMC	254,198	0.557	6.53
Uni-MOF	254,198	0.010	67.94
MOFTransformer	254,198	0.004	52.81
DeepSorption	254,198	0.004	73.51
RetNet	254,198	0.003	52.16
IsothermNet	254,198	0.001	60.19

CO_2 fixed-threshold retrieval. Table 22 reports AP for the three-species model and baselines.

Table 22: Fixed-threshold CO_2 screening AP on the 15,572-framework common support: uptake ≥ 9 mmol/g at 0.15 bar (153 targets) and ≥ 12 mmol/g at 0.35 bar (118 targets).

Model	AP, 0.15 bar	AP, 0.35 bar
MADField-GCMC	**0.7815**	**0.8122**
MOFTransformer	0.1539	0.2070
Uni-MOF	0.2430	0.2296
RetNet	0.0871	0.1160
IsothermNet	0.0180	0.0127
DeepSorption	0.0771	0.0876

Additional GCMC checkpoint comparison with fixed-threshold targets. For a separate evaluation on the full 270,583-framework library with 167 targets satisfying WC $\geq 200\,\mathrm{cm}^3(\mathrm{STP})/\mathrm{cm}^3$, using the same AP evaluator, the CO_2-extended checkpoint gives AP 0.4623 and a 4.733% full-recall fraction. On the same training-excluded pool, it gives AP 0.4635 and a 5.003% full-recall fraction.

The 254,198-candidate cohort contains 108 additional frameworks used for CO_2 training by the extended checkpoint. Excluding training IDs from all three GCMC species instead gives 254,090 candidates and 165 targets; the extended checkpoint has AP 0.4604 and a 5.003% full-recall fraction on this stricter pool. This three-species exclusion defines the evaluation cohort for the extended checkpoint; the main CH_4/Xe-adapted screening model uses the 254,198-candidate cohort.

The extended checkpoint has lower AP but requires a smaller screened fraction for complete target recovery, reflecting the different ranking properties measured by the two metrics.

I.5 Bootstrap Model Progression

The recovery and solver-acceleration results are reported in Section 4.5; the two-stage protocol is specified in Section C. Table 23 evaluates the separate benefit of training on the expanded cDFT label sets. All three models use the same test states and the physical-density evaluator, with pore-masked prediction/reference integration.

Table 23: Bootstrap model progression on the same 28,692-state cDFT test cohort. Density MAE is in 10^{-3} Å^{-3}; relative density L_1 is in percent; uptake MAE is in $\mathrm{cm}^3(\mathrm{STP})/\mathrm{g}$. Each metric first averages states within species and then weights the nine species equally.

Training labels	Density MAE	Relative L_1	Tanimoto	Uptake MAE
Before bootstrap	0.891	7.696	0.9727	2.180
After one bootstrap	0.520	4.416	0.9891	1.307
After two bootstraps	0.493	4.188	0.9896	1.276

I.6 Computational Cost

Screening timing. The screening timing comparison uses GPU inference and our CPU RASPA reference, separately from the GPU gRASPA label-generation measurements below. The reported screening timing estimates are 0.043 s for neural inference and approximately 3 h for the CPU reference, respectively. CPU RASPA simulations used Intel Xeon Gold 6126 processors at 2.60 GHz (12 physical cores per CPU), with one core per independent simulation. The CPU time is the arithmetic mean over 200 single-pressure CH_4 simulations: 100 at 5.8 bar and 100 at 65 bar. It is a per-pressure-state time, not the combined time for the two-pressure working-capacity calculation.

Reference-label generation. Table 24 compares scalar-uptake and converged-density production. gRASPA scalar settings follow the published uptake protocol; density settings follow the DeepAPD reference-generation protocol. cDFT pressure continuation is disabled. These are mean simulation times for the evaluated generation protocols, separate from the median solver-initialization ratio in Section 4.5.

Table 24: Mean simulation walltime under the evaluated protocols. Relative times are calculated from the displayed rounded seconds. cDFT pressure continuation is disabled.

Method	Output	Mean time (s)	Relative to cDFT
gRASPA	3D density	18,468	681.5
gRASPA	Scalar uptake	118.3	4.37
cDFT	3D density	27.1	1.00

I.7 Reference Fidelity between cDFT and GCMC

Table 25 compares non-polar PC-SAFT/UFF cDFT with the corresponding GCMC references on 967 ARC-MOF frameworks and 4,835 gas–pressure states. Its CO_2 reference conditions are 0.1/1 bar; the new charge-aware CO_2 benchmark uses 0.15/0.35 bar. The comparison motivates adaptation from the broad cDFT reference to GCMC fidelity.

Table 25: cDFT–GCMC reference uptake agreement. MAE is in mmol/g, rather than the $\mathrm{cm}^3(\mathrm{STP})/\mathrm{g}$ used for model benchmarks.

Gas	Pressure (bar)	Uptake MAE	Spearman
CH_4	1	0.108	0.999
CH_4	65	0.331	0.999
Xe	1	0.229	0.995
CO_2	0.1	0.394	0.845
CO_2	1	1.689	0.687

J BROADER IMPACTS

Learned adsorption-density surrogates can reduce screening costs and provide spatial information for gas storage, separation, and carbon-capture research. Their use should retain the reference model's physical assumptions and validate shortlisted materials with appropriate simulations or experiments.