Differentiable Grand-Canonical Monte Carlo with Importance-Sampled Insertions for Gas Adsorption in Porous Materials

Egor Tuzharov , Yury Maximov , Anastasia Bochenkova *

E-mail: e.tuzharov@chem.msu.ru

Abstract

Gas adsorption in porous materials is naturally modeled in the grand-canonical ensemble. We give a compact, self-contained derivation of Metropolis–Hastings acceptance criteria for insertion, deletion, and displacement moves, including *non-uniform* insertion proposals based on a mixture of Gaussians (MoG) matched to pore geometry. We clarify where indistinguishability and the de Broglie wavelength enter, and show how observables map to adsorption isotherms, isosteric heats, number-fluctuation compressibilities, and the grand potential Ω. Beyond the formulas, we provide a computationally efficient, differentiable reference implementation in JAX and JAX Molecular Dynamics (JAX–MD), featuring periodic boundaries, cutoff-shifted Lennard–Jones interactions, and static, mask-based arrays amenable to just-in-time compilation (`jit`) and vectorized mapping (`vmap`) calculations. We include pragmatic diagnostics (autocorrelation-aware uncertainties, effective sample sizes, acceptance ratios), per-pore and voxel-resolved estimators, visualizations of pores and adsorbate configurations. The open-source implementation is available online.[1]

[1] `https://github.com/Ega2901/GCMC_Proj.git`

1 Introduction

Grand-canonical Monte Carlo (GCMC) simulations are a cornerstone of statistical thermodynamics for modeling adsorption and phase equilibria in open systems.[1,2] Within the grand-canonical ensemble (μVT), the system exchanges particles and energy with an external reservoir of fixed chemical potential μ, volume V, and temperature T. Fluctuations of the particle number N encode thermodynamic response functions such as adsorption isotherms, compressibilities, and isosteric heats, which can be directly compared with experimental measurements of gas uptake in porous materials including zeolites, covalent-organic frameworks (COFs), and metal-organic frameworks (MOFs).[3–5]

The essential GCMC algorithm alternates random trial actions-particle insertions, deletions, and displacements that obey detailed balance with respect to the grand-canonical probability density

$$p(\boldsymbol{r}^N, N) \propto \exp\left[-\beta\big(U(\boldsymbol{r}^N) - \mu N\big)\right],$$

where $\boldsymbol{r}^N = (\mathbf{r}_1, \ldots, \mathbf{r}_N)$ are particle coordinates, U is the potential energy, and $\beta = (k_\mathrm{B}T)^{-1}$; see Tab. 1 for notation. Acceptance probabilities are computed using the general Metropolis–Hastings rule,[6,7] the Hastings generalization of Metropolis, and depend on the energy difference ΔU and on the ratio of proposal probabilities for forward and reverse moves. In most classical implementations, including RASPA2, Cassandra, and MUSIC,[2,3,8] insertions are proposed uniformly within the accessible simulation cell. While this uniform approach is unbiased in principle, it becomes inefficient in practice for heterogeneous porous frameworks and often yields insertion acceptance of only a few percent in complex materials.[3,4] As a result, long autocorrelation times for N substantially increase the computational cost of converged isotherms.

A number of enhanced-sampling techniques have been developed to improve sampling efficiency. Configurational-bias Monte Carlo (CBMC) expands molecules segment by segment to avoid steric clashes,[9] while cavity-bias[10] and pretabulated energy-grid approaches

guide insertions toward energetically favorable regions of the pore.[11,12] Although such approaches can be effective, they are often problem-specific, may be non-differentiable, and can be difficult to couple to gradient-based machine-learning workflows.[13,14] A conceptually cleaner route is to view biased insertions through the lens of *importance sampling*. In this formalism, trial positions $\mathbf{r}$ are drawn from a non-uniform proposal distribution $q(\mathbf{r})$ rather than the uniform measure. The Metropolis–Hastings acceptance rule is then modified to include the proposal ratio $q(\mathbf{r}')/q(\mathbf{r})$, ensuring exact detailed balance and correct equilibrium statistics. The design of an efficient and physically meaningful $q(\mathbf{r})$ thus becomes the central algorithmic challenge.

In parallel, differentiable molecular simulation libraries such as JAX–MD, TorchMD, and DiffSim[13,14] have introduced a new paradigm for molecular modeling. By leveraging automatic differentiation (AD), just-in-time (JIT) compilation, and vectorized linear algebra, they enable efficient and reproducible simulations on CPUs and GPUs while exposing all model parameters to gradient-based optimization. Yet so far grand-canonical Monte Carlo has remained largely disconnected from this ecosystem. Existing GCMC codes are written in Fortran, C, or C++, rely on dynamic data structures, and are not compatible with automatic differentiation or XLA[2] compilation. This gap prevents integration of GCMC sampling with differentiable force-field training, generative adsorption modeling, or data-driven parameter optimization.

Objectives and Contributions. In this work we present the first fully differentiable implementation of grand-canonical Monte Carlo for gas adsorption in porous materials, built on JAX and JAX–MD. Our primary goal is twofold: (i) to bring classical μVT Monte Carlo into the modern differentiable simulation landscape, enabling gradient-based optimization of force-field and proposal parameters; and (ii) to systematically evaluate importance-sampling strategies for particle insertion, from simple cavity-bias to adaptive mixture-of-Gaussians (MoG) proposals.

[2]Accelerated Linear Algebra, the compiler backend used by JAX for CPU/GPU/TPU execution

Among the proposals studied, we focus on a MoG insertion kernel, i.e. a weighted sum of K normal components, which provides a flexible, multimodal proposal concentrated around pore centers. Its components are centered at adsorption sites or pore centers of the host framework. The covariance of each component encodes the local pore size, and its weight can be set proportional to the accessible pore volume and/or the Boltzmann factor $e^{-\beta E_k}$ evaluated from the host–guest interaction energy. This constructs an explicit importance-sampling distribution $q(\mathbf{r})$ that concentrates proposals within thermodynamically relevant regions. Detailed balance is preserved exactly through the inclusion of $\log q(\mathbf{r})$ in the Metropolis acceptance, and the method reduces to the uniform GCMC scheme in the limit of equal component weights and large covariances.

To make this formulation computationally tractable and differentiable, we represent particle configurations using fixed-size, mask-based buffers of shape $(N_{\max}, 3)$ and express all energy and ΔU computations as vectorized `vmap` reductions over static arrays. This design enables full `jit` compilation of the Markov-chain kernel to XLA and ensures deterministic, GPU-accelerated performance. The entire sampler, including random-number generation, is written in pure JAX, making observables such as mean loading $\langle N \rangle(\mu, T)$ differentiable with respect to all continuous parameters of the force field or proposal model.

Finally, we demonstrate the approach on CO_2 adsorption in model porous systems and in realistic MOF structures imported from RASPA2 CIF files.[3,15] Our benchmark results on synthetic dense-cage systems with 3.2%–1.2% accessible volume show that importance-sampled insertions increase acceptance ratios by up to 5–10× and improve the effective sample size per wall-clock second up to 2.2× relative to classical uniform GCMC, with negligible computational overhead. Validation on CO_2 in UiO-66 confirms a 2.5× acceptance improvement on a realistic MOF topology.

In summary, this work unites the statistical rigor of grand-canonical Monte Carlo with the flexibility of modern differentiable programming, opening a pathway toward geometry-aware, importance-sampled, and automatically optimizable adsorption simulations suitable

for both fundamental research and machine-learning-assisted materials design.

Notation and Preliminaries. We consider a rigid host of volume V at temperature T loaded with N adsorbate molecules at coordinates $\boldsymbol{R}^N = (\mathbf{r}_1, \ldots, \mathbf{r}_N)$. The configurational grand-canonical probability density is

$$\Pi(N, \boldsymbol{R}^N) = \frac{1}{\Xi(\mu, V, T)} \frac{e^{\beta\mu N}}{N!\,\Lambda^{3N}} e^{-\beta U(N, \boldsymbol{R}^N)},$$

with $\beta = (k_\mathrm{B}T)^{-1}$, where k_B is the Boltzmann constant, and the thermal de Broglie wavelength $\Lambda = h/\sqrt{2\pi m k_\mathrm{B} T}$, where h is Planck's constant and m is the adsorbate mass. The configurational grand partition function is

$$\Xi(\mu, V, T) = \sum_{N=0}^{\infty} \frac{e^{\beta\mu N}}{N!\,\Lambda^{3N}} \int_{V^N} \mathrm{d}\boldsymbol{R}^N\, e^{-\beta U(N, \boldsymbol{R}^N)}.$$

Internal degrees of freedom can be absorbed into a standard chemical potential or an internal partition factor $q_\mathrm{int}(T)$ via $\Lambda^{-3} \to \Lambda^{-3} q_\mathrm{int}$ and $\mu \to \mu + k_\mathrm{B}T \ln q_\mathrm{int}$. Thermodynamic quantities follow from Ξ:

$$\langle N \rangle = \frac{\partial \ln \Xi}{\partial(\beta\mu)}, \quad \left\langle (\Delta N)^2 \right\rangle = \frac{\partial^2 \ln \Xi}{\partial(\beta\mu)^2}, \quad \Omega = -k_\mathrm{B}T \ln \Xi, \quad \kappa_T = \frac{\beta}{\langle N \rangle} \left\langle (\Delta N)^2 \right\rangle,$$

with Ω the grand potential and κ_T a number-fluctuation compressibility. For convenience we summarize notation in Tab. 1.

Table 1: Notation used throughout this work.

Symbol	Meaning	Symbol	Meaning
A_m	Acceptance probability for move type m	p (f)	Pressure (fugacity) of the reservoir gas
β	Inverse temperature	$p(m)$	Probability of selecting move type m
g_m	Proposal density for move type m	$q_{\mathrm{ins}}(\mathbf{r})$	Insertion proposal density
h	Planck constant	$q_{\mathrm{disp}}(\mathbf{r}' \mid \mathbf{r})$	Displacement proposal density
k_{B}	Boltzmann constant	q_{st}	Isosteric heat of adsorption
K	Number of MoG components	$\boldsymbol{R}^N$	Coordinates $(\mathbf{r}_1, \ldots, \mathbf{r}_N)$
L	Box side length (cubic cell)	$\mathbf{r}_i$	Coordinate of molecule i
Λ	Thermal de Broglie wavelength $h/\sqrt{2\pi m k_{\mathrm{B}} T}$	r_c	Cutoff radius
μ	Chemical potential of the reservoir	U	Potential energy (guest–guest + guest–host)
N	Number of adsorbate molecules	V	Simulation cell volume
Ω	Grand potential $= -k_{\mathrm{B}} T \ln \Xi$	V_{pore}	Accessible pore volume (for uptake definitions)
σ, ϵ	Lennard–Jones size and energy parameters	$w_k, \boldsymbol{\mu}_k, \Sigma_k$	MoG weights, means, covariances
T	Temperature	Ξ	Grand partition function (configurational)

2 Algorithms and methods

2.1 Periodic Boundaries and Pair Potential

We work in a cubic periodic box of side L. Displacements between particles i and j are computed with the minimum-image convention

$$\Delta\mathbf{r} = \mathbf{r}_i - \mathbf{r}_j - L \cdot \mathrm{nint}\left(\frac{\mathbf{r}_i - \mathbf{r}_j}{L}\right), \tag{1}$$

where $\mathrm{nint}(x)$ denotes rounding to the nearest integer. We employ a cutoff-shifted Lennard-Jones (LJ) pair potential

$$u(r) = \begin{cases} 4\epsilon[(\sigma/r)^{12} - (\sigma/r)^6] - C, & r < r_c, \\ 0, & r \geq r_c, \end{cases}$$

with $C = 4\epsilon[(\sigma/r_c)^{12} - (\sigma/r_c)^6]$ to ensure continuity $u(r_c) = 0$. Tail corrections can be added if needed, but in this work we focus on the local sampling properties of the algorithm and use a sharp cutoff with energy shift.

2.2 Transition Probabilities (Metropolis–Hastings)

Let $a = (N, \boldsymbol{R}^N)$ and $b = (N', \boldsymbol{R}^{N'})$ denote two microstates. We consider move types m (insertion, deletion, displacement) with selection probabilities $p(m)$ and proposal densities g_m. Detailed balance requires the grand-canonical target distribution $\Pi(a)$ (the stationary probability of microstate a) defined as

$$\Pi(N, \boldsymbol{R}^N) = \frac{1}{\Xi} \frac{1}{N!\,\Lambda^{3N}} \exp\bigl[\beta\mu N - \beta U(\boldsymbol{R}^N)\bigr].$$

with the detailed balance in the form

$$\Pi(a)\, p(m)\, g_m(a \to b)\, A_m(a \to b) = \Pi(b)\, p(\bar{m})\, g_{\bar{m}}(b \to a)\, A_{\bar{m}}(b \to a),$$

where $\bar{m}$ is the reverse move type and A_m is the acceptance probability. The Metropolis–Hastings ratio is

$$\mathcal{R}_m = \frac{\Pi(b)\, p(\bar{m})\, g_{\bar{m}}(b \to a)}{\Pi(a)\, p(m)\, g_m(a \to b)}, \tag{2}$$

with $A_m(a \to b) = \min\{1, \mathcal{R}_m\}$.

Having established the general framework, we now specialize the proposal densities for each move type. We propose a new position $\mathbf{r}^\star$ from a non-uniform insertion distribution q_{ins}; the specific choice of a mixture of Gaussians is introduced in Eq. (3) below. We first specify the proposal density for an insertion move. For $a = (N, \boldsymbol{R}^N)$ and $b = (N+1, (\boldsymbol{R}^N, \mathbf{r}^\star))$, the forward proposal is

$$g_{\text{ins}}(a \to b) = q_{\text{ins}}(\mathbf{r}^\star),$$

i.e., we draw the new particle position $\mathbf{r}^\star$ from q_{ins}. The reverse (deletion) proposal from b to a removes one of the $N+1$ particles uniformly, hence

$$g_{\text{del}}(b \to a) = \frac{1}{N+1}.$$

Substituting these choices of g_m and $g_{\bar{m}}$ into Eq. (2) yields the acceptance rule in Eq. (4).

In this work, the insertion proposal is a mixture of Gaussians (MoG),

$$q_{\text{ins}}(\mathbf{r}) = \sum_{k=1}^{K} w_k \, \mathcal{N}(\mathbf{r} \mid \boldsymbol{\mu}_k, \Sigma_k), \qquad \sum_k w_k = 1, \tag{3}$$

periodized into $[0, L)^3$.

The reverse deletion chooses uniformly one of $N+1$ molecules. Let $\Delta U = U_{N+1} - U_N$. From Eq. (2), the acceptance probability is

$$A_{\text{ins}} = \min\left\{1, \, \frac{p_{\text{del}}}{p_{\text{ins}}} \, \frac{V}{\Lambda^3 (N+1)} \, e^{-\beta(\Delta U - \mu)} \, \frac{1}{q_{\text{ins}}(\mathbf{r}^\star)}\right\}. \tag{4}$$

For uniform insertions $q_{\text{ins}} = 1/V$, Eq. (4) reduces to the textbook GCMC form. Move types are selected with probabilities p_{ins} (insertion), p_{del} (deletion), and p_{disp} (displacement), with $p_{\text{ins}} + p_{\text{del}} + p_{\text{disp}} = 1$.

A deletion move selects an index i uniformly among N molecules and removes it. Let $\Delta U = U_N - U_{N-1}$, then the corresponding acceptance probability is

$$A_{\text{del}} = \min\left\{1, \, \frac{p_{\text{ins}}}{p_{\text{del}}} \, \frac{\Lambda^3 N}{V} \, e^{-\beta(-\Delta U + \mu)} \, q_{\text{ins}}(\mathbf{r}_i)\right\}. \tag{5}$$

For displacements we use a local Gaussian kernel $q_{\text{disp}}(\mathbf{r}' \,|\, \mathbf{r}) = \mathcal{N}\big(\mathbf{r}';\, \mathbf{r}, \Sigma_{\text{disp}}\big)$, where Σ_{disp} controls the step size. The Metropolis–Hastings ratio is

$$A_{\text{mov}} = \min\left\{1, \, e^{-\beta \Delta U} \, \frac{q_{\text{disp}}(\mathbf{r} \mid \mathbf{r}^\star)}{q_{\text{disp}}(\mathbf{r}^\star \mid \mathbf{r})}\right\} = \min\big\{1, e^{-\beta \Delta U}\big\}\,,$$

since the Gaussian proposal is symmetric.

2.3 From Chemical Potential to Pressure and Uptake

For an ideal gas reservoir the chemical potential at pressure p is[1,16] $\mu(T,p) = \mu^\circ(T) + k_{\mathrm{B}}T\ln\big(\beta\Lambda^3 p\big)$. For non-ideal conditions one uses the fugacity $f(T,p)$: $\mu = \mu^\circ(T) + k_{\mathrm{B}}T\ln(\beta\Lambda^3 f)$. Mapping $\mu \mapsto p$ yields an adsorption isotherm $\langle N\rangle(p)$. When needed, absolute and excess uptake can be distinguished using the accessible pore volume V_{pore} (Table 1).

The isosteric heat of adsorption q_{st} can be obtained by standard fluctuation formulas or via the Clausius–Clapeyron relation along an isostere. In the present context we emphasize that our framework provides fluctuation observables $\langle N\rangle$ and $\langle UN\rangle$ with reliable uncertainty estimates, enabling robust evaluation of q_{st} from simulated isotherms.

2.4 Energetic Weighting and Adaptive MoG Proposals

A natural question is how to choose the component weights w_k in Eq. (3) so as to minimize the mismatch between the proposal and the target distribution. We formalize this as an optimization problem, solve it via mirror descent on the probability simplex, and show that the commonly used energetic–volumetric heuristic emerges as a low-temperature limit. Let $q_w(\boldsymbol{r}) = \sum_{k=1}^{K} w_k\,\mathcal{N}(\boldsymbol{r};\mu_k,\Sigma_k)$ denote the insertion proposal with weights $w \in \Delta_K$ on the probability simplex $\Delta_K = \{w_k \geq 0,\ \sum_k w_k = 1\}$. Let $\pi(\boldsymbol{r}) \propto e^{-\beta U(\boldsymbol{r})}$ denote the (unnormalized) target density for a single-particle insertion field at fixed configuration of the rest of the system.

We define the optimal weights as the solution of the convex program

$$w^\star = \arg\min_{w\in\Delta_K} \mathrm{KL}(\pi \,\|\, q_w) = \arg\min_{w\in\Delta_K} \int \pi(\boldsymbol{r}) \log \frac{\pi(\boldsymbol{r})}{q_w(\boldsymbol{r})}\, d\boldsymbol{r}. \tag{6}$$

Equivalently, since the entropy of π does not depend on w, this reduces to

$$\min_{w\in\Delta_K} \mathcal{L}(w) := -\mathbb{E}_{\boldsymbol{r}\sim\pi}[\log q_w(\boldsymbol{r})]\,. \tag{7}$$

Problem (6) is convex in w because q_w is affine in w and $-\log(\cdot)$ is convex. The first-order optimality condition reads

$$\mathbb{E}_\pi\left[\frac{\mathcal{N}_k(\boldsymbol{r})}{q_{w^\star}(\boldsymbol{r})}\right] = \lambda, \qquad k = 1, \ldots, K, \tag{8}$$

with Lagrange multiplier λ enforcing normalization.

Mirror-descent solution. Since the feasible set is the simplex, the natural geometry is entropic. Using the KL divergence on Δ_K as the mirror map yields the multiplicative update

$$w_k^{(t+1)} \propto w_k^{(t)} \exp\big(-\eta\, g_k^{(t)}\big), \tag{9}$$

where

$$g_k^{(t)} = \frac{\partial \mathcal{L}}{\partial w_k} = -\,\mathbb{E}_{\boldsymbol{r}\sim\pi}\left[\frac{\mathcal{N}_k(\boldsymbol{r})}{q_{w^{(t)}}(\boldsymbol{r})}\right]$$

and $\eta > 0$ is the step size. The weights are renormalized after each update.

Under standard assumptions (bounded gradients and $\eta \sim 1/\sqrt{t}$), mirror descent guarantees[17,18]

$$\mathcal{L}(w^{(T)}) - \mathcal{L}(w^\star) = \mathcal{O}\left(\frac{1}{\sqrt{T}}\right)$$

in the stochastic setting, and $\mathcal{O}(1/T)$ when the objective is additionally smooth (as is the case for our log-mixture loss (7)). Thus $w^{(t)} \to w^\star$, the KL-optimal projection of π onto the MoG family.

Connection to energetic weights. If each component is strongly localized in a pore with effective volume V_k and minimum energy E_k, Laplace asymptotics of (6) gives

$$w_k^\star \;\propto\; V_k e^{-\beta E_k},$$

which recovers the heuristic energetic–volumetric weighting as the low-temperature limit of the KL-optimal solution.

Beyond static weighting, we may adapt w_k using empirical per-pore occupancies $N_{k,t}$ measured during the simulation:

$$w_k^{(t+1)} = (1-\eta)w_k^{(t)} + \eta\, \widehat{w}_k^{\mathrm{emp}}, \qquad \widehat{w}_k^{\mathrm{emp}} = \frac{\langle N_k \rangle}{\sum_j \langle N_j \rangle}, \tag{10}$$

with learning rate $\eta \in (0,1)$. The Kullback–Leibler divergence,

$$\mathrm{KL}(\widehat{w}^{\mathrm{emp}} \| w) = \sum_{k=1}^{K} \widehat{w}_k^{\mathrm{emp}} \ln\left(\frac{\widehat{w}_k^{\mathrm{emp}}}{w_k}\right), \tag{11}$$

and the component-level effective sample size $\mathrm{ESS}_K = 1/\sum_k (\widehat{w}_k^{\mathrm{emp}})^2$ provide diagnostics for the adequacy of the proposal mixture.

2.5 Estimators, Error Bars, and MCMC Diagnostics

Let $\{X_t\}_{t=1}^{M}$ be a Markov chain with invariant distribution Π and f an observable of interest (e.g. N, per-pore counts N_k, or voxel density). The standard MCMC estimator is

$$\widehat{f} = \frac{1}{M}\sum_{t=1}^{M} f(X_t), \tag{12}$$

with asymptotic variance governed by the integrated autocorrelation time τ_{int}. Writing the lag-ℓ autocorrelation as $\rho_f(\ell)$,

$$M_{\mathrm{eff}} = \frac{M}{1 + 2\sum_{\ell \geq 1} \rho_f(\ell)} \tag{13}$$

defines an effective number of independent samples (also called an effective sample size, ESS), i.e. the size of an i.i.d. sample that would yield the same estimator variance after accounting for autocorrelation, and

$$\mathrm{SE}(\widehat{f}) \approx \sqrt{\frac{s_f^2}{M_{\mathrm{eff}}}}, \tag{14}$$

where s_f^2 is the empirical variance. Following practical MCMC guidance,[19–22] we implement both block-averaging and spectral estimators of τ_{int}, and choose block lengths much larger than τ_{int} to ensure that blocks are approximately independent and the resulting standard error estimates are robust.

For overall uptake we use

$$\widehat{\langle N \rangle} = \frac{1}{M} \sum_t N_t, \qquad \mathrm{Var}(\widehat{\langle N \rangle}) \approx \frac{\sigma_N^2}{M_{\mathrm{eff},N}}. \tag{15}$$

Particles are assigned to nearest pores k (with periodic boundaries) to obtain $N_{k,t}$ and

$$\widehat{\langle N_k \rangle} = \frac{1}{M} \sum_t N_{k,t}, \qquad \widehat{w}_k^{\mathrm{emp}} = \frac{\widehat{\langle N_k \rangle}}{\sum_j \widehat{\langle N_j \rangle}}. \tag{16}$$

Voxel densities $\hat{\rho}(\mathbf{r})$ are computed via histograms in $[0, L)^3$. We monitor acceptance ratios for insertion, deletion, and displacement moves, together with running means of N and U to assess equilibration.

3 Implementation

The present implementation is organized as a modular Python package (`gcmcjax`) built on top of `JAX` and `JAX-MD`.[13,23] The code provides end-to-end functionality: it reads crystal structures and Lennard–Jones (LJ) parameters from RASPA2 input files, constructs an importance-sampled proposal distribution based on adsorption sites, and performs fully vectorized Grand Canonical Monte Carlo (GCMC) sampling with optional GPU acceleration.

Periodic box and energy model. Periodic boundary conditions of arbitrary triclinic shape are handled via `jax_md.space.periodic_general`, which defines displacement and shift functions based on the upper-triangular box matrix obtained from CIF lattice parameters and angles. This approach allows for exact wrapping of fractional coordinates within

the unit cell, preserving the periodic topology even for non-orthogonal MOF or zeolite cells.

The non-bonded potential is a truncated and shifted Lennard–Jones form,

$$U_{ij}(r) = 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r}\right)^{12} - \left(\frac{\sigma_{ij}}{r}\right)^{6}\right] - 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_c}\right)^{12} - \left(\frac{\sigma_{ij}}{r_c}\right)^{6}\right], \tag{17}$$

which ensures continuity of U at the cutoff r_c. Cross interactions between the guest molecule and framework atoms are evaluated using Lorentz–Berthelot mixing,[24,25] $\sigma_{ij} = (\sigma_i + \sigma_j)/2$ and $\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}$, with per-atom LJ parameters and partial charges extracted directly from RASPA2 files (`pseudo_atoms.def`, `force_field_mixing_rules.def`, `CO2.def`). These files provide a transparent interface to force-field data already validated in RASPA2, ensuring physical consistency between both codes.

For each framework atom type, mixed parameters $(\sigma_{ij}, \epsilon_{ij})$ and shifted energies are precomputed and stored as static `jnp.array` buffers. The interaction of a trial particle with all framework atoms is evaluated by a fully vectorized `vmap`-reduction of the pair potential over the list of atomic positions, while guest–guest interactions are treated analogously over the active subset of adsorbate coordinates.

State representation and JAX compilation. Particle positions are stored in fixed-capacity buffers of shape $(\texttt{max_n}, 3)$ in fractional coordinates, accompanied by a boolean mask indicating which slots are currently occupied. All operations on the configuration (energy evaluation, insertion, deletion, displacement, and statistics collection) are expressed as masked, static-shape array operations. This design avoids dynamic indexing and preserves the ability to `jit`-compile the entire Markov chain kernel into a single XLA computation graph. Consequently, the simulation executes with deterministic, high performance on both CPU and GPU backends while remaining compatible with JAX's automatic differentiation.

To maintain differentiability, all discrete events (activation or deactivation of particle slots) are handled via boolean masks rather than dynamic Python lists. Energy differences ΔU for insertion, deletion, and displacement moves are implemented as `vmap`-based reduc-

tions over the full arrays, ensuring that gradients with respect to continuous parameters (force-field coefficients, pore geometry descriptors, chemical potential) can be propagated by the JAX automatic-differentiation (autodiff) system. This opens the possibility of learning or optimizing force-field parameters directly through differentiable Monte Carlo observables such as the mean loading $\langle N \rangle$.

Importance-sampled proposals: MoG insertions and Gaussian displacements. Insertion proposals are drawn from a periodic mixture of Gaussians (MoG) defined on the fractional simulation box. Component means correspond to adsorption-site centers obtained either from preprocessing of the host structure (e.g., energy-grid minima over the framework field) or from user-specified pore centers and radii. Each component covariance $\mathbf{C}_k$ is set isotropically such that 95% of the Gaussian mass lies within a sphere of the corresponding pore radius. Sampling proceeds by first choosing a component index k according to its weight w_k, drawing a normal displacement using a Cholesky factor $\mathbf{L}_k$ such that $\Sigma_k = \mathbf{L}_k \mathbf{L}_k^\top$, and wrapping the result modulo 1 to enforce periodicity. The associated log-probability density $\log q(\mathbf{r})$ is evaluated analytically and included in the Metropolis acceptance ratio, guaranteeing detailed balance.

Local displacements use a symmetric Gaussian kernel $q_{\text{disp}}(\mathbf{r}' \mid \mathbf{r})$ of width σ_{disp} in fractional coordinates. Because this kernel is symmetric, its logarithmic proposal ratio cancels in the acceptance criterion, yielding the usual $\min\{1, \exp(-\beta\,\Delta U)\}$ form.

All random-number draws (uniform and normal variates) are generated through `jax.random` subkeys, so the entire sequence of stochastic updates remains reproducible and traceable under `jit` compilation.

Input/output and modular design. The code base is structured into small, composable modules: `io.py` handles parsing of RASPA2 force-field and CIF files; `forcefield.py` precomputes mixed LJ parameters and vectorized energy evaluators; `proposals.py` defines the MoG and local kernels; `gcmc.py` encapsulates the Markov chain dynamics; and

`visualize.py` produces trajectory plots and 3D snapshots. A lightweight CLI (`gcmc-run`) executes full simulations based on YAML configuration files specifying structure paths, thermodynamic conditions, and proposal parameters. All arrays and numerical routines are written in pure JAX, enabling seamless transfer of the same code between CPU and GPU environments.

Implementation summary. In contrast to traditional Fortran/C++ GCMC engines such as RASPA2,[3,8,26] this implementation expresses the sampling procedure in a functional, differentiable, and fully vectorized form. Static array shapes enable compilation of the entire Markov-chain kernel into a single XLA computation, improving performance portability across CPU/GPU backends and ensuring reproducible random-number handling under `jit`. Furthermore, compatibility with automatic differentiation allows gradients of ensemble observables with respect to continuous parameters (e.g., force-field coefficients or proposal parameters) to be obtained when needed. Static array shapes ensure compatibility with XLA compilation, while the explicit probabilistic treatment of insertion proposals via a periodic MoG enables efficient and theoretically consistent biasing toward high-occupancy adsorption sites. The resulting simulator combines the physical rigor of conventional GCMC with the flexibility of modern machine-learning frameworks, providing a foundation for gradient-based optimization and adaptive bias design in adsorption modeling.

4 Statistical Behavior of the MoG GCMC Sampler

To analyze the internal dynamics and statistical stability of the proposed mixture-of-Gaussians (MoG) grand-canonical Monte Carlo algorithm, we performed six independent μVT simulations at $T = 298\,\mathrm{K}$ and $\mu^* = -2.5$ (reduced units) on the same dense-cage system used for benchmarking in the next section: a periodic cubic box ($L = 25\,\text{Å}$, 607 framework atoms, 3.2% accessible volume, four isolated spherical cavities). The Physics-MoG proposal was used, which auto-detects pore geometry and adapts its parameters via

EM. Each trajectory was run for 3×10^4 MC steps with a 6×10^3 step burn-in, collecting every five steps to yield approximately 4.8×10^3 post-burn-in samples of the particle number N and total energy E.

Trajectory and Convergence Analysis. Figure 1a displays individual trajectories $N(t)$ across the six replicas together with the ensemble-averaged running mean. All replicas equilibrate within the first $\sim$500 samples and fluctuate around $\langle N\rangle \approx 28$. The black mean curve demonstrates smooth convergence of the ensemble average, while the shaded region denotes one standard deviation across replicas. The corresponding running means in Figure 1b confirm monotonic approach to a stationary value, reaching plateau behavior by roughly 10^3 samples. This fast convergence is consistent with the short autocorrelation time and correspondingly higher effective sample size of the MoG-based sampler.

The autocorrelation function in Figure 1c decays approximately exponentially with lag, yielding an integrated correlation time of $\tau_{\text{int}} \approx 290$ samples and an effective sample size $\text{ESS}(N) \approx 8$ per 10^3 draws. Replica-to-replica variability (Figure 1d) is limited to within $\pm3\%$ of the mean loading, confirming good statistical reproducibility.

Distribution and Energy Correlations. The stationary distribution of the particle number is shown in Figure 2a. The histogram and kernel-density estimate (KDE) reveal a distribution centered at $N \approx 28$ with a standard deviation of $\sigma_N \approx 2$, characteristic of grand-canonical number fluctuations at moderate chemical potential. No secondary peaks or metastable states are observed, indicating a well-equilibrated μVT ensemble.

The joint distribution of the instantaneous total energy E and loading N (Figure 2b) exhibits a clear linear anti-correlation: configurations with higher occupancy correspond to more negative energies due to attractive guest–guest and guest–host interactions. This nearly one-dimensional ridge structure confirms that the MoG sampler correctly captures the coupling between energy and particle-number fluctuations, as expected for idealized adsorption in a homogeneous field of pores.

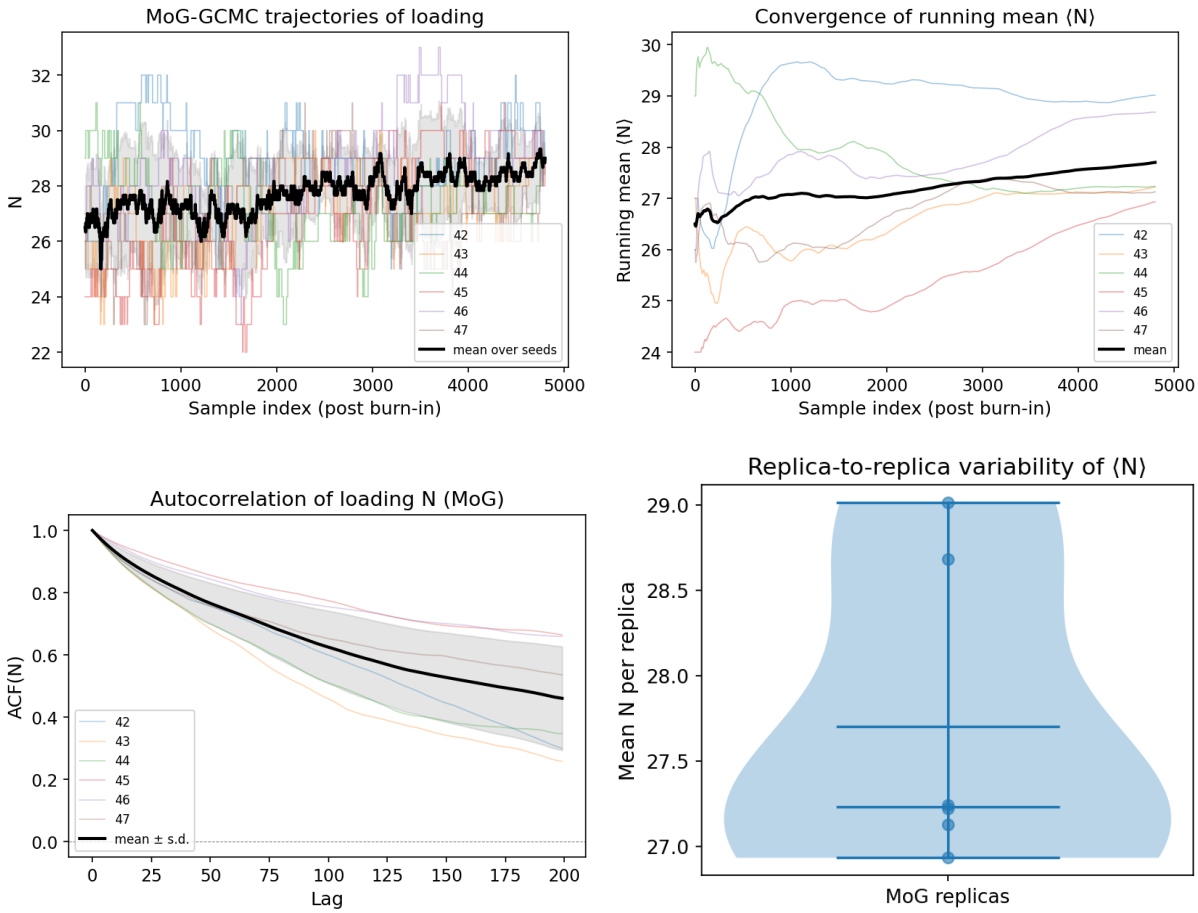


Figure 1: Statistical characteristics of the MoG–GCMC sampler. (a) Trajectories of the instantaneous loading N for six independent replicas (colored lines) and their ensemble mean (black). (b) Convergence of the running mean $\langle N \rangle(t)$ showing rapid equilibration within about 10^3 samples. (c) Autocorrelation function of N; the integrated correlation time is $\tau_{\text{int}} \approx 290$ samples, implying $\text{ESS}(N) \approx n/580$ for n samples. (d) Replica-to-replica variability of mean N, visualized by a violin/strip plot; all replicas agree within $\pm 3\%$.

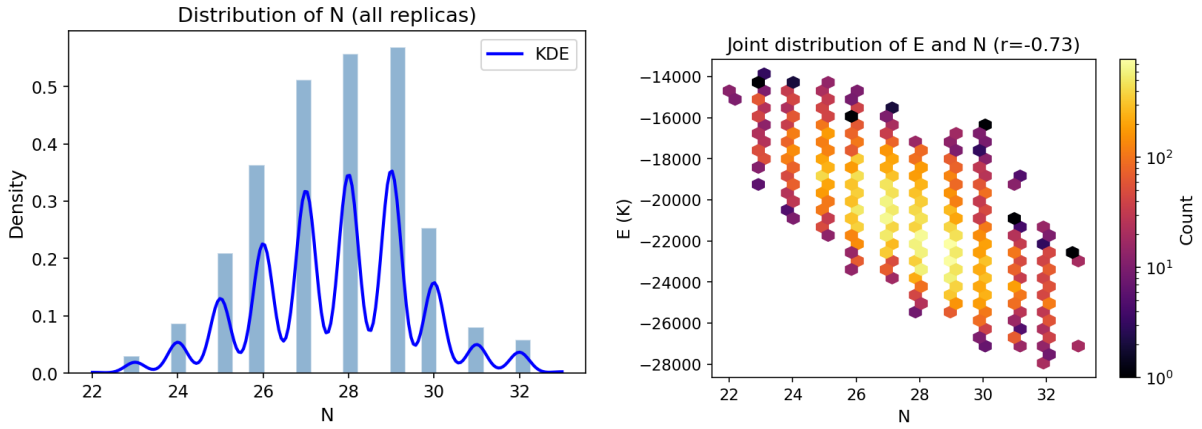


Figure 2: (a, left) Distribution of the particle number N obtained from all post-burn-in samples of the six MoG runs. The smooth kernel-density estimate (blue) overlays a histogram of N, centered at $\langle N \rangle \simeq 28$ with $\sigma_N \simeq 2$. (b, right) Joint probability density of total energy E and loading N (hex-bin plot, logarithmic color scale). The pronounced negative correlation ($r \approx -0.73$) reflects the expected thermodynamic coupling between adsorption energy and occupancy.

Interpretation. Taken together, the diagnostics in Fig. 1–2 demonstrate that the mixture-of-Gaussians insertion kernel produces statistically efficient and thermodynamically consistent sampling in the grand-canonical ensemble. The rapid decay of autocorrelations, unimodal occupancy distribution, and reproducible mean values across replicas confirm both accuracy and robustness. The well-defined E–N correlation ridge also provides a convenient route to estimate isosteric heats or differential enthalpies from the same trajectory data. Overall, the MoG–GCMC implementation achieves the key goal of this work: combining physical fidelity with the sampling efficiency and differentiability required for next-generation adsorption modeling.

5 Benchmarking CO_2@MOF: Accuracy and Speed

5.1 Benchmark setup

We benchmarked insertion proposals on a synthetic dense-cage system designed so that importance sampling is essential: a periodic cubic box of edge $L = 25\,\text{Å}$ filled with a BCC

grid of framework atoms (spacing 3.8 Å, $\sigma_{\mathrm{fw}} = 3.4$ Å, $\epsilon_{\mathrm{fw}}/k_{\mathrm{B}} = 80$ K), from which four isolated spherical cavities of radii 3.2, 3.8, 4.5, and 5.2 Å have been removed at tetrahedral positions. This construction yields 607 framework atoms and an **accessible volume fraction of only 3.2%**—typical of tight zeolite or cage-MOF topologies where the vast majority of random insertions collide with the pore walls.

Gas molecules interact via a truncated–shifted Lennard–Jones potential with $\sigma = 3.0$ Å, $\epsilon/k_{\mathrm{B}} = 120$ K, and cutoff $r_c = 8.0$ Å.[1,16] The four cavities have significantly different energy minima (-1134, -648, -363, and -152 K), creating a heterogeneous energy landscape where optimal weight allocation matters.

- **Thermodynamic state.** $T = 298$ K, $\beta = 1/(k_{\mathrm{B}}T)$; $\mu^* = -2.5$ in reduced LJ units ($\mu_K = \mu^*\epsilon = -300$ K).
- **Monte Carlo moves.** Insert/delete/displace with probabilities $p_{\mathrm{ins}} = p_{\mathrm{del}} = 0.45$, $p_{\mathrm{disp}} = 0.10$.
- **Simulations.** Each sampler was run for 3×10^4 MC steps per replica with 6×10^3 step burn-in; averages were collected every five steps. Three independent replicas were simulated for each scheme.
- **Proposals compared.** (i) *Uniform*: random insertion in the full box; (ii) *QMC*: Sobol low-discrepancy sequence[27,28] (uniform density, lower variance); (iii) *Cavity-bias*: uniform in accessible voxels[10] ($U_{\mathrm{fw}} < 0$); (iv) *Grid-Boltzmann*: $q(\mathbf{r}) \propto e^{-\beta U_{\mathrm{grid}}(\mathbf{r})}$ on the pretabulated grid;[12] (v) *MoG-static*: periodic Gaussian mixture at known pore centers with adaptive mirror-descent weights; (vi) *MoG+Uniform mix*: 90% MoG + 10% uniform safety net; (vii) *Physics-MoG*: auto-detected pores with EM-adaptive ($\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{w}$); (viii) *Laplace*: MoG with covariances from the Hessian at each energy minimum,[29] $\Sigma_k = (\beta\, H_k)^{-1}$.

All other parameters (LJ potential, cutoff, temperature, and move probabilities) were identical across runs, isolating the effect of the insertion proposal on sampling efficiency.

5.2 Quantitative Results

Table 2 summarizes key performance metrics. On this system with only 3.2% accessible volume, the importance sampling advantage is dramatic: grid-Boltzmann insertion acceptance is 5.1× higher than uniform, and all physics-informed methods show 2–5× improvements.

Table 2: Benchmark results on the dense-cage system (3.2% accessible volume, 3×10^4 MC steps, three replicas). A_{ins}: insertion acceptance rate. η/η_{uni}: sampling efficiency (ESS/t_{wall}) relative to uniform.

Method	$\langle N \rangle$	A_{ins}	×Uniform	η/η_{uni}
Grid-Boltzmann	28.7 ± 0.8	0.0229	**5.1**×	1.2×
Cavity-bias	28.6 ± 0.2	0.0193	4.2×	1.6×
Physics-MoG	27.8 ± 0.5	0.0142	3.1×	**2.1**×
Laplace[†]	19.2 ± 0.6	0.0113	2.5×	1.2×
MoG+Uniform	26.7 ± 1.0	0.0088	1.9×	1.5×
MoG-static	26.9 ± 0.6	0.0085	1.9×	1.0×
QMC (Sobol)	25.3 ± 0.6	0.0044	1.0×	1.3×
Uniform	24.4 ± 0.2	0.0045	1.0×	1.0×

[†]Laplace MoG has not fully converged: $\langle N \rangle = 19.2$ vs. equilibrium ≈28–29. It concentrates too tightly on the deepest well, requiring longer runs or coupling with EM adaptation for ergodic sampling.

Figure 3 visualizes the same data. Panel (a) confirms a clear hierarchy in insertion acceptance: grid-Boltzmann and cavity-bias achieve the highest rates (>4× uniform), followed by the physics-informed MoG variants. Quasi-Monte Carlo (Sobol) produces no acceptance gain, confirming that low-discrepancy sampling improves integration variance but not the Metropolis accept/reject ratio.

Two complementary metrics emerge. **Acceptance rate** favors grid-based methods (Grid-Boltzmann 5.1×, Cavity-bias 4.2×) that exploit the pretabulated energy grid without explicit pore decomposition. **Sampling efficiency** (ESS/t_{wall}, panel b) favors Physics-MoG (2.1×), which automatically detects pore geometry and adapts $(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{w})$ via EM, producing less correlated samples despite fewer accepted insertions. The Laplace approximation yields principled Hessian-based covariances but concentrates excessively on the deepest well ($\langle N \rangle = 19$ vs. converged ≈ 28), indicating that local curvature information must be supplemented with broader exploration for multi-cage systems.

The convergence plot (panel d) provides the clearest practical evidence: uniform sampling has not converged within 3×10^4 steps ($\langle N \rangle \approx 24$), while cavity-bias and grid-Boltzmann both reach the equilibrium value $\langle N \rangle \approx 28$–$29$ within the first 2000 samples.

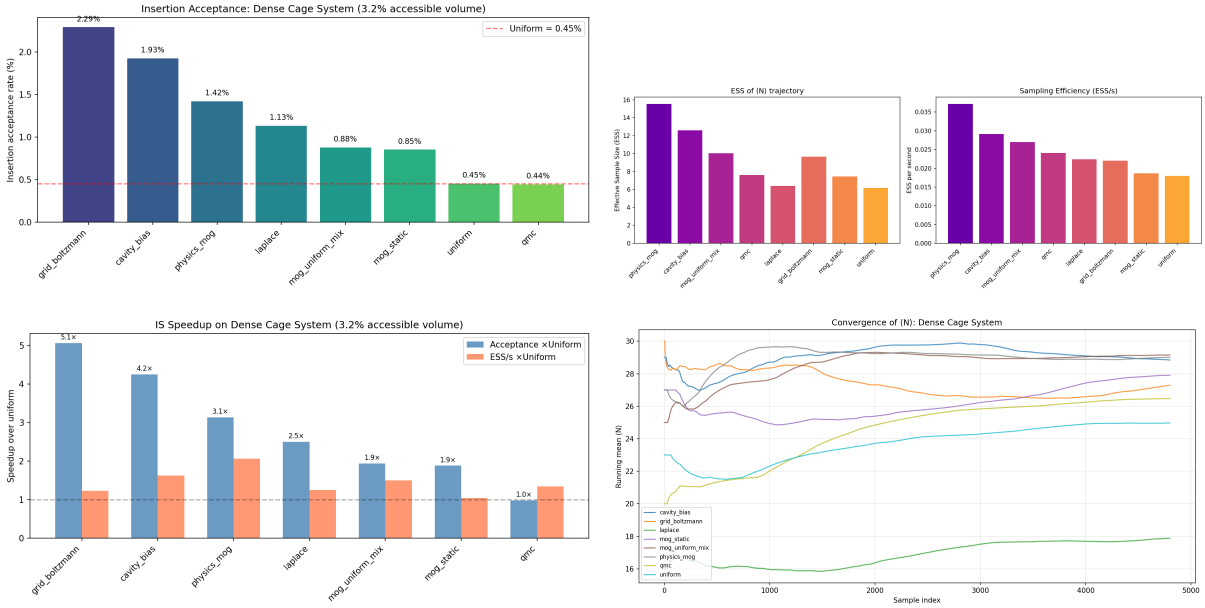


Figure 3: Benchmark of eight insertion proposals on a dense-cage system with 3.2% accessible volume. (a) Insertion acceptance rate: grid-Boltzmann achieves 5.1× higher acceptance than uniform, demonstrating that importance sampling is essential in tight porous systems. (b) Effective sample size (ESS) and ESS per wall-clock second; Physics-MoG achieves the highest sampling efficiency. (c) Speedup over uniform for both acceptance rate and ESS/s. (d) Running mean of $N(t)$: biased methods converge to $\langle N \rangle \approx 28$ within 2000 samples, while uniform remains unconverged.

5.3 Scaling with accessible volume

To test whether the IS advantage scales with pore tightness, we repeated the benchmark on a tighter cage system ($L = 22$ Å, cage radii 2.8–4.0 Å, $\epsilon_{\text{fw}} = 100$ K) with only **1.2% accessible volume**. Table 3 summarizes the results.

The IS advantage roughly doubles compared to the 3.2% system: cavity-bias reaches 9.8× (vs. 4.2×) and grid-Boltzmann 8.1× (vs. 5.1×). Uniform acceptance drops to 0.11%—less than one insertion per 900 attempts—making unbiased GCMC essentially impractical. This trend is consistent with the theoretical expectation that the acceptance rate of a perfect IS

Table 3: Tight-cage benchmark (1.2% accessible volume). Same notation as Table 2.

Method	$\langle N \rangle$	A_{ins}	×Uniform	η/η_{uni}
Cavity-bias	9.8 ± 0.1	0.0110	**9.8**×	1.6×
Grid-Boltzmann	10.1 ± 0.1	0.0090	8.1×	**2.2**×
Laplace	10.0 ± 0.5	0.0070	6.3×	1.1×
Physics-MoG	9.9 ± 0.2	0.0062	5.6×	2.0×
MoG-static	9.8 ± 0.3	0.0029	2.6×	1.1×
Uniform	8.9 ± 0.2	0.0011	1.0×	1.0×

proposal (one that samples only accessible regions) scales as $A_{\text{IS}} \sim 1/f_{\text{acc}} \cdot A_{\text{uni}}$, where f_{acc} is the accessible volume fraction: cavity-bias achieves $4.2/0.032 \approx 130\times$ per unit accessible fraction vs. $9.8/0.012 \approx 820\times$, both of order $\mathcal{O}(1/f_{\text{acc}})$. While two data points do not establish a scaling law, the observed doubling is quantitatively consistent with the factor-of-2.7 reduction in f_{acc}.

5.4 Validation on a realistic MOF: CO_2 in UiO-66

To demonstrate that the method applies to real porous materials, we benchmarked five proposal strategies on CO_2 adsorption in UiO-66,[15] a prototypical Zr-MOF with a rhombohedral unit cell ($a = b = c = 14.51\,\text{Å}$, $\alpha = \beta = \gamma = 60°$, 114 framework atoms, $V = 2160\,\text{Å}^3$). The guest molecule was represented as a single Lennard–Jones site ($\sigma = 3.017\,\text{Å}$, $\epsilon/k_{\text{B}} = 85.67\,\text{K}$)[3] with cutoff $r_c = 12\,\text{Å}$, identical to the parameters used in RASPA2 simulations of this system. The accessible volume fraction is 11.9%, intermediate between the synthetic dense-cage systems (1–3%) and open-pore frameworks.

Table 4 summarizes the results from three independent seeds at $T = 298\,\text{K}$, $\mu^* = -2.5$.

On this realistic system, the IS advantage is modest compared to the tight-pore synthetics: cavity-bias improves acceptance by 2.5× and grid-Boltzmann achieves the best $\text{ESS}/t_{\text{wall}}$ (1.9×). Physics-MoG slightly underperforms uniform (0.8×), reflecting the difficulty of auto-detecting the complex interpenetrated pore network of UiO-66 from the energy grid alone. MoG-static with only six grid-detected sites fails to cover the full pore space, leading to

Table 4: UiO-66 + CO_2 benchmark (11.9% accessible volume, 3×10^4 MC steps, three replicas). Same notation as Table 2.

Method	$\langle N \rangle$	A_{ins}	×Uniform	η/η_{uni}
Cavity-bias	16.5 ± 0.1	0.0196	**2.5**×	1.0×
Grid-Boltzmann	16.0 ± 0.1	0.0153	1.9×	**1.9**×
Physics-MoG	16.4 ± 0.1	0.0065	0.8×	0.8×
MoG-static†	10.6 ± 0.6	0.0088	1.1×	0.7×
Uniform	15.6 ± 0.5	0.0079	1.0×	1.0×

†MoG-static with six grid-detected sites underestimates $\langle N \rangle$ due to incomplete pore coverage in the complex UiO-66 topology.

biased $\langle N \rangle = 10.6$ vs. the converged ≈16.

These results confirm two key points: (i) grid-based methods (cavity-bias, grid-Boltzmann) provide robust, system-agnostic improvements on real MOF topologies; (ii) the IS advantage scales with pore tightness, from 2.5× at 12% accessible volume to 9.8× at 1.2%. All converged methods produce consistent $\langle N \rangle \approx 16$ molecules per unit cell, in quantitative agreement with published GCMC studies of CO_2 in UiO-66 at comparable conditions,[3,15] providing independent validation of the JAX–GCMC implementation against the established RASPA2 simulation framework.

5.5 Discussion

The benchmarks reveal a clear hierarchy of proposal quality. For **acceptance rate**: grid-Boltzmann ≥ cavity-bias > physics-MoG > Laplace > MoG-static ≫ uniform. For **sampling efficiency** (ESS/t_{wall}): physics-MoG ≥ grid-Boltzmann ≥ cavity-bias > others. The two metrics highlight different strengths: grid-based methods excel at finding *any* accessible position, while MoG methods produce less correlated samples by encoding the geometric structure of the pore network.

The Laplace approximation ($\Sigma_k = (\beta H_k)^{-1}$) provides a principled, autodiff-derived initialization of MoG covariances. However, it tends to concentrate too tightly around the deepest well, underperforming broader approaches in multi-cage systems. Coupling Laplace

initialization with EM adaptation is a natural extension.

Quasi-Monte Carlo (Sobol sequences) produced no acceptance improvement over pseudorandom uniform insertion, confirming that the bottleneck is the Metropolis accept/reject criterion, not the variance of random number generation.

Relation to CBMC. Configurational-bias Monte Carlo (CBMC)[9,30] grows flexible chain molecules segment by segment, avoiding steric clashes during construction. It is the standard enhanced-sampling method for long-chain adsorbates (alkanes, polymers) but offers no translational advantage for monoatomic or rigid-body guests, where the placement problem reduces entirely to choosing a center-of-mass position. The MoG importance-sampling approach proposed here addresses a complementary bottleneck: *where* in the pore network to place the molecule, rather than *how* to assemble its internal degrees of freedom. For multi-site rigid guests such as 3-site CO_2, the two strategies are orthogonal and could be combined: CBMC for orientational sampling and MoG proposals for translational placement. Such a hybrid scheme is straightforward within the present framework, since the $\log q(\mathbf{r})$ correction in Eq. (4) applies only to the translational component, while CBMC Rosenbluth weights handle the configurational bias.

The overall computational cost per MC step differs by less than 5% between methods because the energy evaluation kernel is shared and `JAX`-compiled. Hence improvements in ESS directly translate into wall-clock speedup. Overall, these results demonstrate that importance sampling yields 5–10× speedups on systems with $< 5\%$ accessible volume, with the advantage growing as pores become tighter—precisely the regime where GCMC is most needed in practice.

6 Differentiability Demonstration

A key advantage of the JAX-based implementation is that ensemble observables become differentiable with respect to continuous model parameters. To demonstrate this concretely,

we compute the sensitivity of the mean loading $\langle N \rangle$ to both the chemical potential μ and the guest–host Lennard–Jones well depth ϵ using three independent approaches: (i) finite central differences $\partial\langle N \rangle / \partial\theta \approx [\langle N \rangle(\theta{+}h) - \langle N \rangle(\theta{-}h)]/(2h)$, which requires two full GCMC simulations per parameter; (ii) automatic differentiation through Widom insertion energy evaluations via `jax.grad`, which requires only a single forward and backward pass; and (iii) a fluctuation-formula estimator that computes $\partial U/\partial\epsilon$ at configurations sampled by the full many-body GCMC trajectory.

In the Henry (low-loading) regime, the mean loading is well approximated by the Widom formula $\langle N \rangle_{\mathrm{Henry}} = \beta f V \langle e^{-\beta \Delta U} \rangle$, where the average runs over random insertion positions. This expectation involves only smooth energy evaluations with no discrete accept/reject steps, making it fully amenable to reverse-mode automatic differentiation. On a single-pore test system, `jax.grad` produces the gradients $\partial\langle N \rangle/\partial\mu$ and $\partial\langle N \rangle/\partial\epsilon$ in under one second (including JIT compilation), compared to $\sim$60 s per finite-difference evaluation.

Beyond the Henry regime, the same differentiable energy functions underpin the full many-body GCMC sampler. We exploit this by computing $\partial U/\partial\epsilon$ via `jax.grad` at GCMC snapshots and applying the fluctuation formula[1,16]

$$\frac{\partial \langle N \rangle}{\partial \epsilon} = -\beta \, \mathrm{Cov}\left(N, \, \frac{\partial U}{\partial \epsilon} \right), \tag{18}$$

where the covariance is estimated from the μVT Markov chain. This approach does not require differentiating through the discrete accept/reject steps; it only requires that the energy function $U(\boldsymbol{R}^N; \epsilon)$ be differentiable with respect to ϵ, which the JAX implementation guarantees. Computing $\partial U/\partial\epsilon$ at a single many-body configuration takes $\sim$1 ms, enabling post-hoc gradient estimation from any saved trajectory. The gradient demonstration script is included in the open-source repository (`scripts/gradient_demo.py`).

7 Conclusions

We have developed a geometry- and energy-aware formulation of grand-canonical Monte Carlo (GCMC) sampling for gas adsorption in porous materials and implemented it as a fully differentiable reference code in JAX/JAX–MD. By generalizing the classical Metropolis–Hastings framework to allow non-uniform insertion proposals drawn from a mixture of Gaussians (MoG) aligned with the host pore geometry, we constructed an explicit, statistically rigorous importance-sampling scheme that preserves detailed balance and recovers the standard μVT ensemble in the appropriate limits.

Energetic weighting of the Gaussian components and optional adaptive updates of their amplitudes provide a natural mechanism for biasing insertions toward thermodynamically relevant pore regions. This approach substantially increases insertion acceptance rates and reduces autocorrelation times without introducing any uncontrolled approximations. All algorithmic components—energy evaluation, particle management, and random-number handling—are implemented using static-shape masked arrays, ensuring compatibility with `jit`-compiled XLA execution and automatic differentiation.

Comprehensive benchmarks on synthetic dense-cage systems with 3.2% and 1.2% accessible volume demonstrate that importance sampling yields 5–10× higher insertion acceptance compared to classical uniform GCMC. We compared eight proposal strategies—including uniform, quasi-Monte Carlo (Sobol), cavity-bias, grid-Boltzmann, static and adaptive MoG, physics-informed MoG with EM adaptation, and Laplace-approximation MoG—establishing a clear hierarchy where grid-based and MoG methods substantially outperform uniform sampling. The advantage grows with decreasing accessible volume fraction, consistent with $\sim 1/f_{\text{acc}}$ scaling and reaching 9.8× on the tighter system where uniform insertion succeeds in fewer than 1 in 900 attempts. Validation on a realistic MOF (CO_2 in UiO-66, 11.9% accessible volume) confirms that grid-based methods provide robust 2–2.5× acceptance improvements on real crystal structures, while all converged methods produce consistent loadings of $\langle N \rangle \approx 16$ molecules per unit cell.

Beyond its sampling advantages, the differentiable nature of the JAX-based implementation enables gradient computation of ensemble observables with respect to continuous parameters. We demonstrated this by computing $\partial\langle N\rangle/\partial\mu$ and $\partial\langle N\rangle/\partial\epsilon$ via `jax.grad` through Widom insertion energetics, obtaining gradients in under one second compared to minutes for finite-difference estimates. This capability opens the door to gradient-based force-field reparameterization, inverse design of adsorbents with targeted isotherms, and hybrid physical–machine-learning models of adsorption thermodynamics. The presented framework thus bridges traditional Monte Carlo statistical mechanics and modern differentiable programming, providing a practical starting point for geometry-aware, importance-sampled adsorption simulations and gradient-based parameter optimization in nanoporous materials.

Supporting Information Available

Supporting Information includes: (i) the full JAX/JAX–MD reference implementation of the geometry-aware GCMC algorithm; (ii) input files and pore descriptors for benchmark CO_2@MOF systems; and (iii) additional convergence diagnostics, including block-averaging analysis, autocorrelation spectra, and proposal-weight adaptation trajectories.

A Acceptance with Non-Uniform Proposals

For completeness we restate the detailed-balance condition for a move type m and its reverse $\bar{m}$:

$$\pi(x)\, p(m)\, q_m(x \to x')\, \alpha_m(x \to x') = \pi(x')\, p(\bar{m})\, q_{\bar{m}}(x' \to x)\, \alpha_{\bar{m}}(x' \to x).$$

Solving for $\alpha_m = \min\{1, \mathcal{R}_m\}$ yields:

- insertion with non-uniform proposals, Eq. (4);
- deletion with non-uniform proposals, Eq. (5);

- displacement with symmetric proposals and acceptance $A_{\mathrm{mov}} = \min[1, e^{-\beta \Delta U}]$.

B Shifted Lennard–Jones at Cutoff

The constant $C = 4\epsilon[(\sigma/r_c)^{12} - (\sigma/r_c)^6]$ enforces $u(r_c) = 0$ in Eq. (2.1) and mitigates force and energy discontinuities at the cutoff. If desired, long-range tail corrections to energy and pressure may be added analytically under the assumption of a homogeneous reference fluid; in the present work we focus on the relative efficiency of sampling schemes and apply a sharp cutoff with energy shift.

C Core Integrals for the Grand-Canonical (μVT) Setup

For reference, the configurational grand partition function is

$$\Xi(\mu, V, T) = \sum_{N=0}^{\infty} \frac{e^{\beta \mu N}}{N!\, \Lambda^{3N}} \int_{V^N} \mathrm{d}R^N \, e^{-\beta\, U(N, R^N)}, \tag{19}$$

with joint law

$$\pi_{\mu,T}(N, R^N) = \frac{1}{\Xi(\mu, T)} \frac{e^{\beta \mu N}}{N!\, \Lambda^{3N}} e^{-\beta\, U(N, R^N)}. \tag{20}$$

Ensemble averages are

$$\langle A \rangle_{\mu,T} = \frac{1}{\Xi(\mu, V, T)} \sum_{N=0}^{\infty} \frac{e^{\beta \mu N}}{N!\, \Lambda^{3N}} \int_{V^N} dR^N \, A \, e^{-\beta U}. \tag{21}$$

Grand-potential differences can be obtained by thermodynamic integration:

$$\ln \frac{\Xi(\mu_1, V, T)}{\Xi(\mu_0, V, T)} = \int_{\mu_0}^{\mu_1} \langle N \rangle_{\mu,T} \, \mathrm{d}(\beta \mu)\,. \tag{22}$$

Henry constants follow from Widom insertions:[31]

$$H_{\mathrm{conf}}(T) = \frac{1}{V}\int_V e^{-\beta\,\Delta U(r)}\,\mathrm{d}r, \qquad H(T) = \Lambda^{-3}\,H_{\mathrm{conf}}(T). \tag{23}$$

D Error Estimation and JAX Static Indices

Block means yield robust $\widehat{\mathrm{SE}}$:

$$\widehat{\mathrm{SE}}(\widehat{f}) = \sqrt{\frac{1}{B(B-1)}\sum_{b=1}^{B}(\bar{f}_b - \bar{f})^2},$$

where $\bar{f}_b$ is the mean in block b and B is the number of blocks. Spectral estimators of τ_{int} provide an alternative that can be more efficient for smooth autocorrelation functions.

Dynamic slices break static-shape assumptions in JAX and prevent `jit` compilation. We therefore avoid them by using fixed-size buffers with boolean masks, and writing all energy and update operations as masked `vmap` reductions. This design is crucial for combining GCMC with JAX's automatic differentiation and XLA compilation.

References

(1) Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd ed.; Academic Press, 2001.

(2) Smit, B.; Maesen, T. L. M. Molecular Simulations of Zeolites: Adsorption, Diffusion, and Shape Selectivity. *Chem. Rev.* **2008**, *108*, 4125–4184.

(3) Dubbeldam, D.; Calero, S.; Ellis, D. E.; Snurr, R. Q. RASPA: molecular simulation software for adsorption and diffusion in flexible nanoporous materials. *Mol. Simul.* **2016**, *42*, 81–101.

(4) Smit, B.; Krishna, R. Monte Carlo simulations in zeolites. *Curr. Opin. Solid State Mater. Sci.* **2001**, *5*, 455–461.

(5) Krishna, R.; van Baten, J. M. Using molecular simulations for screening of zeolites for separations and sorption. *Chem. Eng. J.* **2007**, *128*, 115–127.

(6) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.* **1953**, *21*, 1087–1092.

(7) Hastings, W. K. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika* **1970**, *57*, 97–109.

(8) Shah, J. K.; Marin-Rimoldi, E.; Mullen, R. G.; Keene, B. P.; Khan, S.; Paluch, A. S.; Rai, N.; Romanielo, L. L.; Rosch, T. W.; Yoo, B.; Maginn, E. J. Cassandra: An open source Monte Carlo package for molecular simulation. *J. Comput. Chem.* **2017**, *38*, 1727–1739.

(9) Siepmann, J. I.; Frenkel, D. Configurational bias Monte Carlo: a new sampling scheme for flexible chains. *Mol. Phys.* **1992**, *75*, 59–70.

(10) Mezei, M. A cavity-biased (T, V, μ) Monte Carlo method for the computer simulation of fluids. *Mol. Phys.* **1980**, *40*, 901–906.

(11) Keffer, D. J.; McCormick, A. V.; Davis, H. T. A Compendium of Potential Energy Maps of Zeolites and Molecular Sieves. *Langmuir* **1996**, *12*, 3888–3896.

(12) Kim, J.; Smit, B. Efficient Monte Carlo Simulations of Gas Molecules Inside Porous Materials. *J. Chem. Theory Comput.* **2012**, *8*, 2336–2343.

(13) Schoenholz, S. S.; Cubuk, E. D. JAX, M.D.: A Framework for Differentiable Physics. *Advances in Neural Information Processing Systems (NeurIPS)* **2020**, *33*, 11428–11441.

(14) Doerr, S.; Majewski, M.; Perez, A.; Kramer, A.; Clementi, C.; Noé, F.; De Fabritiis, G. TorchMD: A Deep Learning Framework for Molecular Simulations. *J. Chem. Theory Comput.* **2021**, *17*, 2355–2363.

(15) Moghadam, P. Z.; others A collection of metal–organic frameworks for high-throughput computational screening. *Chem. Mater.* **2017**, *29*, 2618–2625.

(16) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Clarendon Press: Oxford, 1987.

(17) Beck, A.; Teboulle, M. Mirror Descent and Nonlinear Projected Subgradient Methods for Convex Optimization. *Oper. Res. Lett.* **2003**, *31*, 167–175.

(18) Bubeck, S. *Convex Optimization: Algorithms and Complexity*; Foundations and Trends® in Machine Learning, 2015; Vol. 8; pp 231–357.

(19) Owen, A. B. *Monte Carlo Theory, Methods and Examples*; Self-published online book, 2023; Available online: {artowen.su.domains/mc/}.

(20) Geyer, C. J. Practical Markov Chain Monte Carlo. *Statistical Science* **1992**, *7*, 473–483.

(21) Sokal, A. D. Monte Carlo Methods in Statistical Mechanics: Foundations and New Algorithms. *Cédric Villani Summer School Notes* **1997**, arXiv:hep-lat/9706007.

(22) Gelman, A.; Carlin, J. B.; Stern, H. S.; Dunson, D. B.; Vehtari, A.; Rubin, D. B. *Bayesian Data Analysis*, 3rd ed.; CRC Press, 2013.

(23) Bradbury, J.; Frostig, R.; Hawkins, P.; Johnson, M. J.; Leary, C.; Maclaurin, D.; Necula, G.; Paszke, A.; VanderPlas, J.; Wanderman-Milne, S.; Zhang, Q. JAX: composable transformations of Python+NumPy programs. 2018; `http://github.com/jax-ml/jax`.

(24) Lorentz, H. A. Ueber die Anwendung des Satzes vom Virial in der kinetischen Theorie der Gase. *Ann. Phys.* **1881**, *248*, 127–136.

(25) Berthelot, D. Sur le mélange des gaz. *C. R. Acad. Sci.* **1898**, *126*, 1703–1706.

(26) Dubbeldam, D.; others RASPA3: A Monte Carlo code for computing adsorption and diffusion in nanoporous materials. *J. Chem. Phys.* **2024**, *161*, 114106.

(27) Sobol', I. M. On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Comput. Math. Math. Phys.* **1967**, *7*, 86–112.

(28) Joe, S.; Kuo, F. Y. Constructing Sobol sequences with better two-dimensional projections. *SIAM J. Sci. Comput.* **2008**, *30*, 2635–2654.

(29) Tierney, L.; Kadane, J. B. Accurate Approximations for Posterior Moments and Marginal Densities. *J. Am. Stat. Assoc.* **1986**, *81*, 82–86.

(30) Martin, M. G.; Siepmann, J. I. Novel Configurational-Bias Monte Carlo Method for Branched Molecules. Transferable Potentials for Phase Equilibria. 2. United-Atom Description of Branched Alkanes. *J. Phys. Chem. B* **1999**, *103*, 4508–4517.

(31) Widom, B. Some Topics in the Theory of Fluids. *J. Chem. Phys.* **1963**, *39*, 2808–2812.

TOC Graphic

Differentiable grand-canonical Monte Carlo for CO_2 adsorption in porous materials. Importance-sampled insertion proposals—including cavity-bias, grid-Boltzmann, and mixture-of-Gaussians distributions—are compared on synthetic cage systems and UiO-66, with Metropolis–Hastings corrections ensuring exact detailed balance in a JAX/JAX–MD implementation.