DFT-quality adsorption simulations in metal-organic frameworks enabled by machine learning potentials

Ruben Goeminne, Louis Vanduyfhuys, Veronique Van Speybroeck, and Toon Verstraelen*

Center for Molecular Modeling (CMM), Ghent Univeristy, Technologiepark 46, 9052 Zwijnaarde, Belgium

E-mail: Toon.Verstraelen@UGent.be

Abstract

Nanoporous materials such as metal-organic frameworks (MOFs) have been extensively studied for their potential for adsorption and separation applications. In this respect, grand canonical Monte Carlo (GCMC) simulations have become a well-established tool for computational screenings of the adsorption properties of large sets of MOFs. However, their reliance on empirical force field potentials have limited the accuracy with which this tool can be applied to MOFs with challenging chemical environments such as open-metal sites. On the other hand, density-functional theory (DFT) is too computationally demanding to be routinely employed in GCMC simulations due to the excessive number of required function evaluations. Therefore, we propose in this paper a protocol for training machine leaning potentials (MLPs) on a limited set of DFT intermolecular interaction energies (and forces) of CO_2 in ZIF-8 and the open-metal site containing Mg-MOF-74, and use the MLPs to derive adsorption isotherms from first principles. We make use of equivariant NequIP model which

has demonstrated excellent data efficiency, and as such an error on the interaction energies below 0.2 kJ·mol^{-1} per adsorbate in ZIF-8 was attained. Its use in GCMC simulations results in highly accurate adsorption isotherms and heats of adsorption. For Mg-MOF-74, a large dependence of the obtained results on the used dispersion correction was observed, where PBE-MBD performs the best. Lastly, to test the transferability of the MLP trained on ZIF-8, it was applied to ZIF-3, ZIF-4 and ZIF-6, which resulted in large deviations on the predicted adsorption isotherms and heats of adsorption. Only when explicitly training on data for all ZIFs, accurate adsorption properties were obtained. As the proposed methodology is widely applicable to guest adsorption in nanoporous materials, it opens up the possibility for training general-purpose MLPs to perform highly accurate investigations of guest adsorption.

Introduction

Porous materials such as metal-organic frameworks (MOFs) have drawn widespread interest in the last few decades for their potential for gas adsorption and separation applications.[1–3] Consisting of metal-oxide building blocks and organic ligands, their permanent porosity under guest removal was first established in the late 1990s.[4–8] Through their high degree of versatility, rational design of these materials for targeted adsorption properties became possible through the concept of isoreticular design. In seminal work by the group of Yaghi, the pore size of MOF-5 could be incrementally varied from 3.8 to 28.8 Å by substitution of the constituting organic linker, giving rise to the IRMOF isoreticular series of MOFs. Of these, IRMOF-6 demonstrated state-of-the-art methane storage capacity at the time.[9] Furthermore, linkers can be functionalized to tune the host-guest interactions.[10–12] Clearly, the chemical space of MOFs arising from combining linkers, metal-oxide bricks, topologies and linker functionalizations is vast, precluding an exhaustive search of this space through experimental characterization. Herein lies the prime value of computational screening studies.[13–23] Through grand canonical Monte Carlo simulations, the adsorption properties of large sets of MOF structures can be quickly extracted. Early studies following this approach focused on the screening of MOFs for their uptake and heat of adsorption of hydrogen, carbon dioxide and methane,[13–15] with later studies investigating separation properties.[16–19] Prime examples of the success of such computational screenings include the identification of NU-1103 out of a set of 13 512 MOFs with outstanding hydrogen storage capacity at cryogenic temperatures and the identification of 300 out of a set of 137 953 MOFs with exceptional methane storage capacity.[20,21] Lin *et al.* leveraged GPU hardware to significantly speed up adsorption simulations, enabling them to screen hundreds of thousands of zeolites and ZIFs (zeolitic imidazolate framework; a class of MOFs with isomorphic topology to zeolites) for carbon capture applications.[22,23] Furthermore, recent developments in machine learning (ML) models have been leveraged to predict the adsorption characteristics in MOFs.[24–30] By training such models on reference calculations or experimental uptakes, the adsorption properties of

novel MOFs can be predicted with high accuracy (compared to the reference calculations or experiments).

While all of the aforementioned screening studies have clearly demonstrated their merit in identifying promising nanoporous for applications, they critically rely on an accurate description of the underlying potential energy surface (PES). Preferentially, DFT or wavefunction-based methods would be used to provide a highly accurate description of the PES. However, such methods are excessively computationally demanding for routine usage in GCMC simulations, due to the required number of evaluations needed to converge a typical simulation. Therefore, computationally cheap force fields are usually employed. To describe the host-guest and guest-guest interactions, non-covalent force field potentials are employed. Usually, a Lennard-Jones (LJ) or Buckingham potential is used, supplemented with an electrostatic interaction arising from atomic charges assigned to the framework and adsorbates. Examples of force fields used for this task are the Universal Force Field (UFF), DREIDING, MM3 and the general Amber force field (GAFF).[31–34] In later years, these force fields were adapted for specific use with MOFs, including parameterizations of transition metals commonly encountered in these materials.[35,36] However, many examples exist in the literature where force fields required re-parameterization due to poor agreement with experiments. Boulanger *et al.* rescaled the GAFF van der Waals dispersion parameters by a factor of 1.115 to obtain accurate hydration free energies, Wu *et al.* rescaled the UFF LJ parameter ϵ by a factor of 0.635 to reproduce methane adsorption isotherms in ZIF-8 and similarly, Pérez-Pellitero *et al.* found that a rescaling of the LJ parameters was required for agreement with experimental adsorption isotherms of N_2 and CO_2 in ZIFs.[37–39] Moreover, force fields such as UFF and DREIDING have been shown to significantly underestimate the interaction strength of adsorbates with open-metal sites present in many MOFs.[40] This deficiency has spurred the development of more advanced polarizable force fields in which polarization of the adsorbate on open-metal sites is modeled with the induced dipole method.[41? –45] However, even when employing these more advanced force fields, significant deviations between the experimental

and predicted CO_2 adsorption isotherms and heats of adsorption remain in the open-metal site containing M-MOF-74 (M=Co,Fe,Mg,Mn,Ni,Zn) series of MOFs.[45] Such limits in the attainable accuracy could be related to the functional form of the employed force fields, which might not fully capture the interaction of adsorbates with the open-metal site.

The preceding discussion demonstrates the need for methods that can bridge the gap between a force field and DFT description of the intermolecular interactions between adsorbates and a MOF adsorbent. On one hand, DFT calculations are prohibitively expensive for use in GCMC simulations due to the required number of iterations needed for convergence.[46] On the other hand, force field based GCMC simulations are computationally (relatively) cheap, but are limited in accuracy when applied to challenging intermolecular interactions such as those present in open-metal site containing MOFs. To bridge this gap, different approaches have been investigated. Lee *et al.* evaluated the Henry regime (characterizing the low-pressure adsorption behavior) in M-MOF-74(M=Zn,Mg) for several adsorbents by performing biased single particle insertions, evaluated at the DFT level of theory.[47] In a similar work, Vandenbrande *et al.* proposed an importance sampling scheme based on preceding force field calculations to evaluate the Henry coefficients and heats of adsorption at the DFT level of theory at infinite dilution for methane in UiO-66 and CO_2 in Mg-MOF-74 by means of single particle insertions.[48] However, the efficiency of this method (the number of required DFT evaluations) still depends critically on the quality of the used force field. Moreover, only the low-pressure regime can be characterized from this method. Alternatively, Kundu *et al.* predicted the adsorption isotherms of CO_2 in Mg-MOF-74 from calculations of the interaction energy of an adsorbate on the open-metal site and linker with a correlated wavefunction-corrected DFT method.[49–51] With these adsorption sites, a lattice of sites was defined, on which grand canonical Monte Carlo simulations were performed. Although successful in predicting the adsorption isotherms at different temperatures, this method relies on an initial accurate determination of well-defined adsorption sites, as guest molecules are not considered to adsorb on other sites of the framework, limiting the general

applicability of this method.

Hence, this work aims at a generally applicable methodology to perform GCMC calculations in MOFs at DFT accuracy. For this, we leverage the recent developments of highly accurate and data-efficient equivariant machine learning potentials (MLPs) to describe the interaction of guest molecules in MOFs. In the message passing NequIP potential used in this work, each atom is characterized by a feature vector which is refined during training through a series of interactions layers.[52] Tested on datasets of energies and forces of small molecules and periodic materials, it demonstrated state-of-the-art accuracy as well as exceptional data efficiency. In recent years, MLPs have been employed to study the mechanical, diffusion and phase transition properties of MOFs by means of MD simulations.[53–56] In these applications, the training set consists of DFT energies and atomic forces. However, for our application of employing an MLP in GCMC simulations, the requirements on the training set are fundamentally different. Through the integration of the equations of motion, an equilibrium canonical MD simulation samples regions of the PES proportional to a Boltzmann factor of the potential energy. While the exact same distribution is sampled from (canonical) Monte Carlo simulations, this is achieved by performing trial moves of the system to regions of the PES which might be very high in energy. Even though these trial moves will not be accepted and are therefore not relevant, they still need to be reliably rejected. This implies that, contrary to the case of MD simulations, the MLP needs to be explicitly trained on structures high on the PES. Secondly, only the intermolecular or non-covalent energy is of importance in GCMC simulations, as both the framework and adsorbates are generally considered rigid.

To address these considerations, a methodology is proposed to efficiently generate datasets of frameworks loaded with guests which will serve as the input data to train an MLP. Subsequently, tests are performed to determine both the hyperparameters of the network, as well as the required data set size to derive an accurate model for CO_2 in ZIF-8. For adsorption in this MOF, high quality experimental data is available, allowing for a unambiguous

comparison with the results obtained with our proposed methodology.[57] With the trained MLP, GCMC calculations are performed to extract the isotherms and heats of adsorption, demonstrating excellent agreement with experiments. The proposed methodology is also applied to model CO_2 adsorption in the open-metal site containing Mg-MOF-74 with reference data generated at the PBE-D3(BJ), PBE-TS and PBE-MBD level of theory, highlighting the important differences between the used dispersion corrections.[58–62] Lastly, the ability of an MLP trained on ZIF-8 to model adsorption in ZIF-3, ZIF-4 and ZIF-6 was investigated, as this type of transferability of interactions to similar chemical environments is usually implicitly assumed in classical force fields. The MLP is shown to be unable to capture the stronger interactions present in ZIF-3 and ZIF-4. Even in ZIF-6, where the error on the test set is acceptable, significant errors in the predicted heats of adsorption are obtained, highlighting the risk of extrapolation for MLPs. Only when explicitly training on all ZIFs, accurate results are obtained.

Methodology

Non-covalent dataset generation

As mentioned before, the requirements on a dataset to train an MLP employed in GCMC simulations are different than the ones needed for performing MD simulations, due to the need to reliably reject trial moves that are high in potential energy. On the other hand, MLPs can generally be trained to higher accuracy when the training set is not excessively unbalanced, as would be the case when combining repulsive interactions order of magnitudes stronger than the attractive interactions. Therefore, an inclusion in the dataset of extremely repulsive interactions as encountered when atoms approach each other closely, is not desirable. As a compromise between these considerations, the method illustrated in Figure 1 for generating a dataset of non-covalent interactions is proposed.

First, the van der Waals radii derived by Alvarez are assigned to each element present in

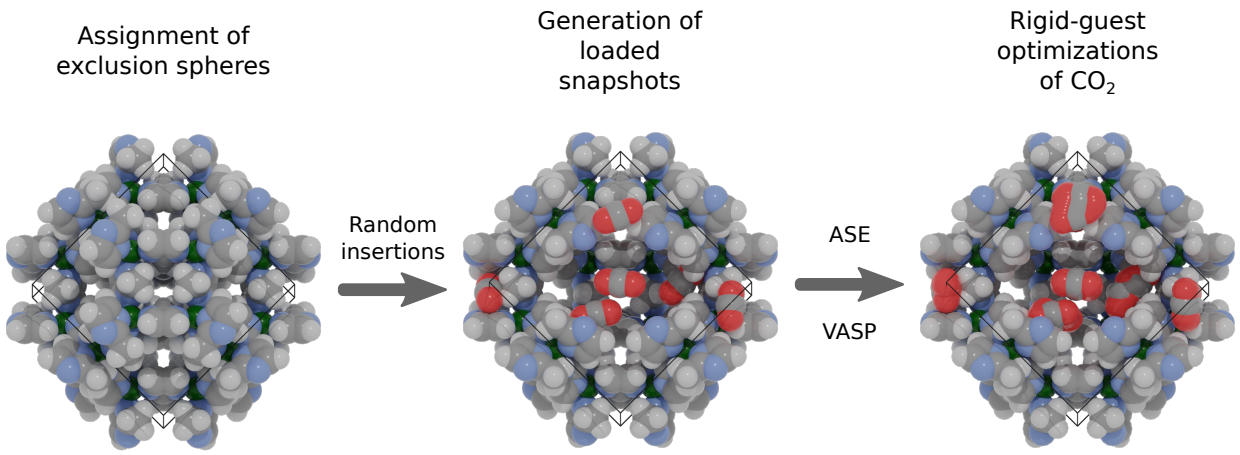

Figure 1: Proposed methodology for generating the dataset to train an MLP on intermolecular energies (and forces). First, snapshots of the framework loaded with adsorbates respecting minimum intermolecular distances defined by the exclusion spheres are generated. Subsequently, rigid-guest optimizations of the adsorbates are performed (with a fixed framework) to guide the adsorbates to favorable interaction sites in the framework.

the framework and adsorbates.[63] Their analysis of more than 5 million non-covalent distances shows that no non-bonded interactions occur at inter-atomic distances below the sum of two atomic radii minus 0.7 Å. Therefore, the van der Waals radius for each atom minus 0.35 Å can be seen as an 'exclusion sphere' within which no adsorbate-adsorbate or adsorbate-framework interactions occur. The exclusion spheres will be used further on (with proper validation) in GCMC simulations to reject any trial step in which an overlap occurs. Because of this, there is no need to explicitly train the MLP to these very short-ranged (repulsive) interactions.

Subsequently, a set of N guest-loaded snapshots constrained by the exclusion spheres are generated by means of random insertions, implemented in a custom python script. The number of adsorbates in each snapshot N_{ads} is distributed uniformly between 0 and N_{max}, with N_{max} the maximum number of guests which can reliably be inserted without violating the constraints. Although the snapshots generated from this procedure respect the minimum intermolecular distances defined by the exclusion spheres, they still mostly represent high energy structures. Therefore, in a last step, each snapshot is optimized using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) optimizer from the Atomic Simulation Environ-

ment (ASE) at the relevant level of theory for N_{steps} iterations, keeping both the framework coordinates fixed and the adsorbates rigid ($N_{\text{steps}} = 30$ in this work, as will be shown further).[64] While extensions to flexible frameworks and adsorbates are possible, these will not be considered in this work. Every snapshot from the optimizations is retained in the dataset, for a total of $N \times N_{\text{steps}}$ snapshots. By retaining all snapshots, both the high energy (at the beginning of the optimization), as well as the low energy structures (at the end of the optimization) are sampled. By subtracting the energies and forces of the constituting isolated framework and adsorbates from the total energies and forces, the intermolecular energies E_{inter} and forces $\mathbf{F}_{\text{inter}}$ are obtained. As the framework remains fixed, the energy and forces of the framework are the same for each snapshot, requiring only 1 single point calculation. In this work, the experimentally resolved structures for all MOFs are used. Similarly, because of the use of rigid adsorbates, only a single calculations of the energy and forces of an isolated adsorbate are required. Note that this calculation of intermolecular energies and forces is not subject to a basis set superposition error, as plane waves were used as basis set in this work. Further details on the DFT settings are given in the Computational Details.

MLP training

The Neural Equivariant Interatomic Potential (NequIP) is used in this work as MLP to represent the intermolecular interaction energy and forces of guest-loaded frameworks, chosen for its excellent accuracy and data efficiency.[52] However, future more advanced models could also be employed as simple drop-in replacements. To train the network, an appropriate cost function is required. In this work, a slight adaptation of the existing cost functions in NequIP is made. As N_{ads} varies across snapshots in the training set and the intermolecular energy is an extensive property, the energy cost is normalized with N_{ads} (see Equation 1). Neglecting to do so would artificially bias training towards snapshots with higher N_{ads}. Additionally, the intermolecular forces are also included in the cost function. While not strictly needed, including forces during training is known to improve data efficiency. An L1 cost function is

used for increased robustness with respect to outliers:

$$\mathcal{C}(\boldsymbol{\theta}) = \frac{\lambda_E}{N}\sum_{i=1}^{N}\left\|\frac{E^{(i)}(\boldsymbol{\theta}) - \hat{E}^{(i)}_{\text{inter}}}{N_{\text{ads,i}}}\right\| + \frac{\lambda_F}{3NN_a}\sum_{i=1}^{N}\sum_{a=1}^{N_a^{(i)}}\left\|\boldsymbol{F}_a^{(i)}(\boldsymbol{\theta}) - \hat{\boldsymbol{F}}^{(i)}_{\text{inter,a}}\right\| \tag{1}$$

with the sum over i running over the snapshots, the sum over a running over atoms, $N_a^{(i)}$ the number of atoms in each snapshot i and $N_a = \sum_i N_a^{(i)}$. In this cost function, the parameters of the network $\boldsymbol{\theta}$ are to be optimized to minimize the difference between the predicted energies and forces ($E^{(i)}(\boldsymbol{\theta})$ and $\boldsymbol{F}_a^{(i)}(\boldsymbol{\theta})$) and the reference energies and forces ($\hat{E}^{(i)}_{\text{inter}}$ and $\hat{\boldsymbol{F}}^{(i)}_{\text{inter,a}}$). The hyperparameters λ_E and λ_F determine the relative weight of the energies and forces. The effect of different hyperparameters of the network such as the cutoff radius, the number of interaction layers, the number of features and the maximum rotation order are investigated in the Supporting Information.

GCMC calculations

With the trained MLPs, GCMC calculations are performed using the same exclusion spheres as for the dataset generation to reject unphysical trial moves. Care is taken that the exclusion spheres were not too large such that energetically favorable moves would be rejected (see further). In other words, the exclusion spheres should strictly encompass regions of the pores associated with unfavorable insertion energies. The Peng-Robinson equation of state is used to convert the gas pressure to a fugacity used in the simulations. Apart from the gas uptake, the isosteric heat of adsorption Q_{st} as a function of the gas pressure can be derived from GCMC calculations from the following Equation:[65]

$$Q_{\text{st}} = -\frac{\langle E \cdot N\rangle - \langle E\rangle\langle N\rangle}{\langle N^2\rangle - \langle N\rangle^2} + RT \tag{2}$$

with N and E the instantaneous number of adsorbates and intermolecular energy, the brackets $\langle\cdots\rangle$ denoting an ensemble average and R and T are the molar gas constant and tem-

perature. In this Equation, the first term represents the difference in interaction energy between a MOF loaded with $N+1$ molecules, compared to N molecules, and the second term represents the work required to compress an ideal gas at a fixed pressure and temperature. Alternatively, the isosteric heat of adsorption at infinite dilution $Q_{st,\text{dilution}}$ can also be computed using the Widom insertion method, performing a set of random insertions of the adsorbate in the framework and computing the interaction energy ΔE:[66]

$$Q_{st,\text{dilution}} = -\frac{\langle \Delta E\, e^{-\beta \Delta E} \rangle}{\langle e^{-\beta \Delta E} \rangle} + RT \tag{3}$$

Results

Carbon dioxide adsorption in ZIF-8

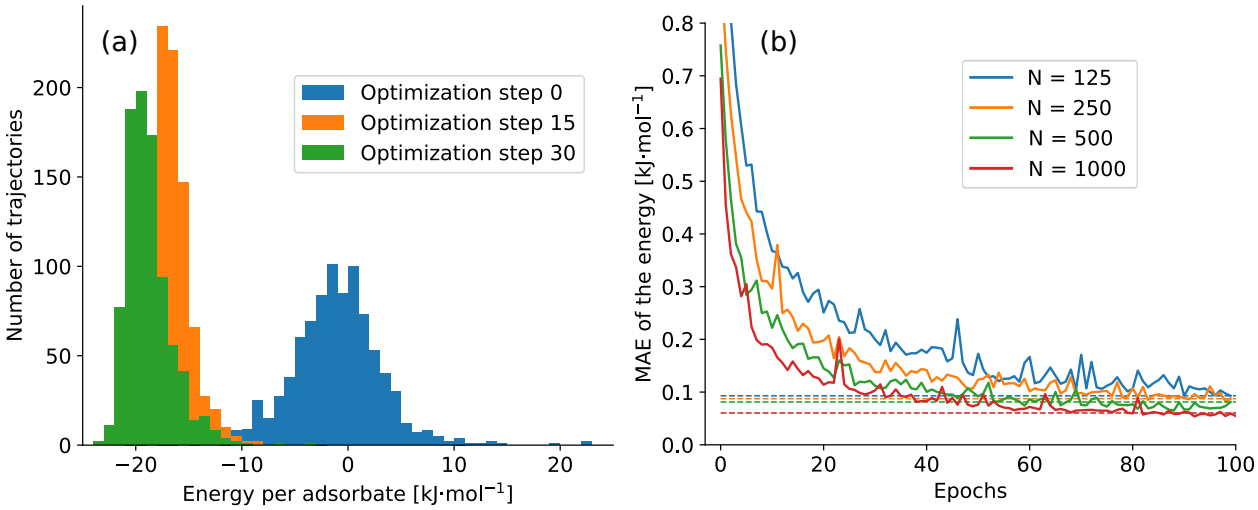

Figure 2: (a) Histogram of the interaction energy per adsorbate along three steps in optimization trajectories (step 0, 15 and 30) out of a total of 1000 optimizations, performed at the PBE-D3(BJ) level of theory. (b) The mean absolute error (MAE) on the energy of the validation set during training for different sizes of the dataset (between 125 and 1000 optimizations).

With the procedure outlined before, 1000 snapshots containing between 0 and 32 CO_2 adsorbates were generated after which each was optimized at the PBE-D3(BJ) level of theory for a total of 30 optimization steps. To determine whether the geometric optimizations have

steered the adsorbates to the favorable interaction sites, the histograms of the interaction energy per adsorbate are shown at the start of the optimization trajectories (step 0), in the middle (step 15) and at the end (step 30) in Figure 2(a). As can be seen, the initial snapshots represent mostly configurations between -10 and 10 kJ·mol^{-1} per adsorbate. Optimizing the snapshots, most have reached interaction sites between -20 and -15 kJ·mol^{-1} per adsorbate at the end. Note that it is not vitally important for all optimization trajectories to fully converge. As long as sufficient favorable interactions are included in the dataset, the MLP is able to accurately represent the minima (see further). To determine the required number of trajectories to train an accurate MLP, the total dataset of 1000 optimization trajectories was subdivided into sets containing N = 125, 250, 500 and 1000 trajectories. For each set, a separate MLP was trained (denoted further as MLP$_{125}$, MLP$_{250}$, MLP$_{500}$ and MLP$_{1000}$), in which 80% of the data is used for training and 20% for validation. The validation mean absolute error (MAE) is shown in Figure 2(b). As can be seen, the MAE on the interaction energies is low even for the dataset containing 125 trajectories. Increasing N from 125 to 1000 decreases the MAE from 0.093 to 0.060 kJ·mol^{-1} per adsorbate. However, evaluating each model on a test set containing 100 optimizations to which no training was performed, elevated errors of 0.45 to 0.15 kJ·mol^{-1} per adsorbate are obtained for MLP$_{125}$ and MLP$_{1000}$.

As a first validation of the derived MLPs, their accuracy in reproducing interaction energies of a single adsorbate in the framework is investigated. A set of 50 rigid-body optimizations of an adsorbate in ZIF-8 performed with MLP$_{1000}$ revealed that the strongest interaction site of CO_2 is a configuration in which the adsorbate interacts with a pore window, perpendicular to the plane of the window (see Figure 3(b)). To explore the interaction around this minimum, the adsorbate was subsequently translated along the normal of the pore window plane. Relative to the configuration centered in the pore window, the adsorbate is translated with distances between 0 and 4 Å with steps of 0.1 Å. For each structure, the interaction energy is recomputed with all MLPs and the reference PBE-D3(BJ) level of theory. As shown in Figure 3(a), significant deviations are observed for MLP$_{125}$. For the

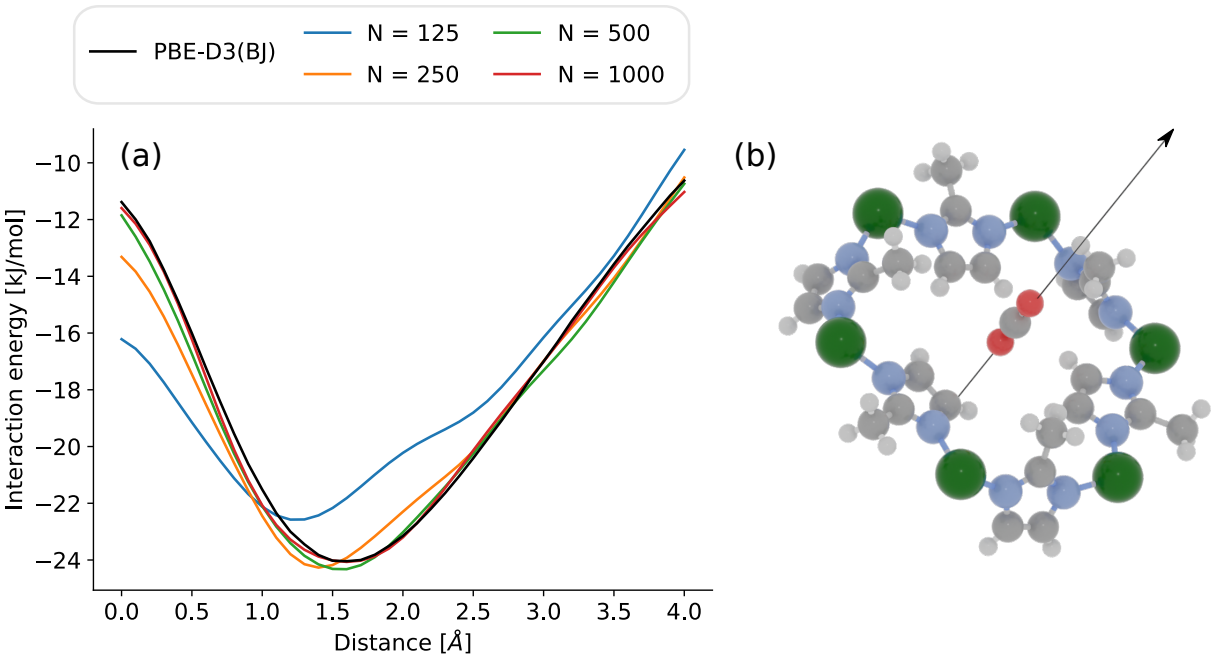

Figure 3: (a) Interaction energies of an adsorbate with the pore window in ZIF-8, perpendicular to the plane of the window, computed with all four MLPs and the reference PBE-D3(BJ) level of theory. The adsorbate is translated along the normal (see (b)) with distances between 0 and 4 Å. (b) Visualization of the adsorbate interacting with the pore window of ZIF-8, and the translation vector along the normal on the pore window. For clarity, only the Zn atoms and imidazolate linkers of the six-membered pore ring are shown.

other MLPs, good agreement with the reference energies are obtained. Furthermore, as the optimizations were performed with an MLP, this demonstrates that the forces learned during training enable accurate optimizations towards the minimum.

While this first validation demonstrates that an MLP trained on a sufficient number of optimizations can accurately reproduce interaction energies around minima on the potential energy surface, an additional validation of the used exclusion spheres is vital to ensure that their exact values do not influence the adsorption properties obtained further. To this end, the heats of adsorption Q_{st} of CO_2 in ZIF-8 were computed at 273 K by means of Widom insertions with rescaled values of the radii of the exclusion spheres between a factor of 0.80 and 1.20. In case the exclusion spheres would erroneously exclude relevant regions of the potential energy surface, Q_{st} would be sensitive to rescaling factors smaller than 1. As shown in Figure 4, Q_{st} is indeed insensitive to the exact choice of the radii for the exclusion spheres

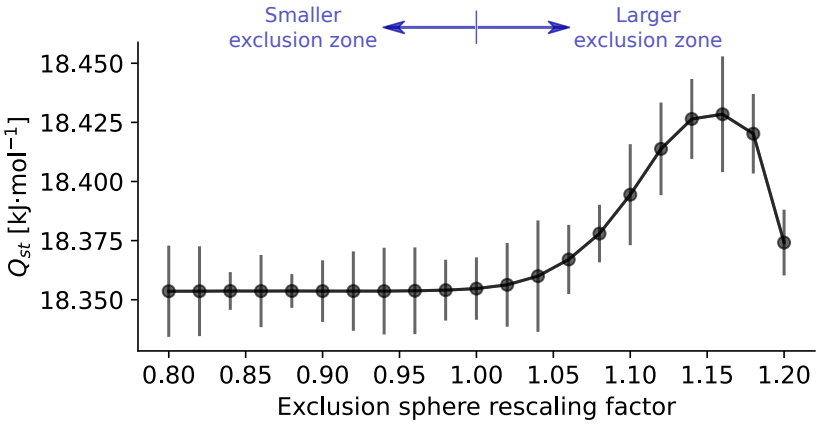

Figure 4: Isosteric heat of adsorption Q_{st} of CO_2 in ZIF-8 obtained from $5 \cdot 10^6$ Widom insertions using MLP_{1000} with rescaled exclusion sphere radii. Errors are obtained from dividing the set of $5 \cdot 10^6$ Widom insertions in 5 subsets, and computing the standard deviation between Q_{st} obtained for each subset.

(with only small deviations occurring from rescaling factors larger than 1), demonstrating that the region excluded by the exclusion spheres is indeed an irrelevant part of the phase space.

With each of the MLPs trained on varying dataset sizes, GCMC calculations were subsequently performed at a temperature of 273 K and a range of gas pressures between 0.1 bar and 50 bar. Each simulation consists of $5 \cdot 10^6$ MC steps, with equal probability for translations, rotations, insertions and deletions. The resulting uptake as a function of the gas pressure and heat of adsorption as a function of the guest loading are shown in Figure 5(a) and 5(b). The adsorption isotherms obtained with the MLPs trained on between 125 and 1000 optimizations are nearly identical. Comparing these isotherms with experimental results obtained by Simmons *et al.*, Abraha *et al.* and Gracés *et al.*, good agreement between both is observed, as seen in Figure 5(a).[57,67,68] While a slight over-prediction of the uptake at low gas pressure and an under-prediction of the uptake at high gas pressures is observed, this could be due to the rigid-framework approximation employed in GCMC simulations, as the flexible behavior of ZIF-8 and its effects on adsorption have been well-established.[69] As shown in Figure 5(b), MLP_{1000} predicts an increase in the heat of adsorption of CO_2 in ZIF-8 from 18.0 kJ·mol^{-1} at the smallest loadings to approximately 25.8 kJ·mol^{-1} at the highest

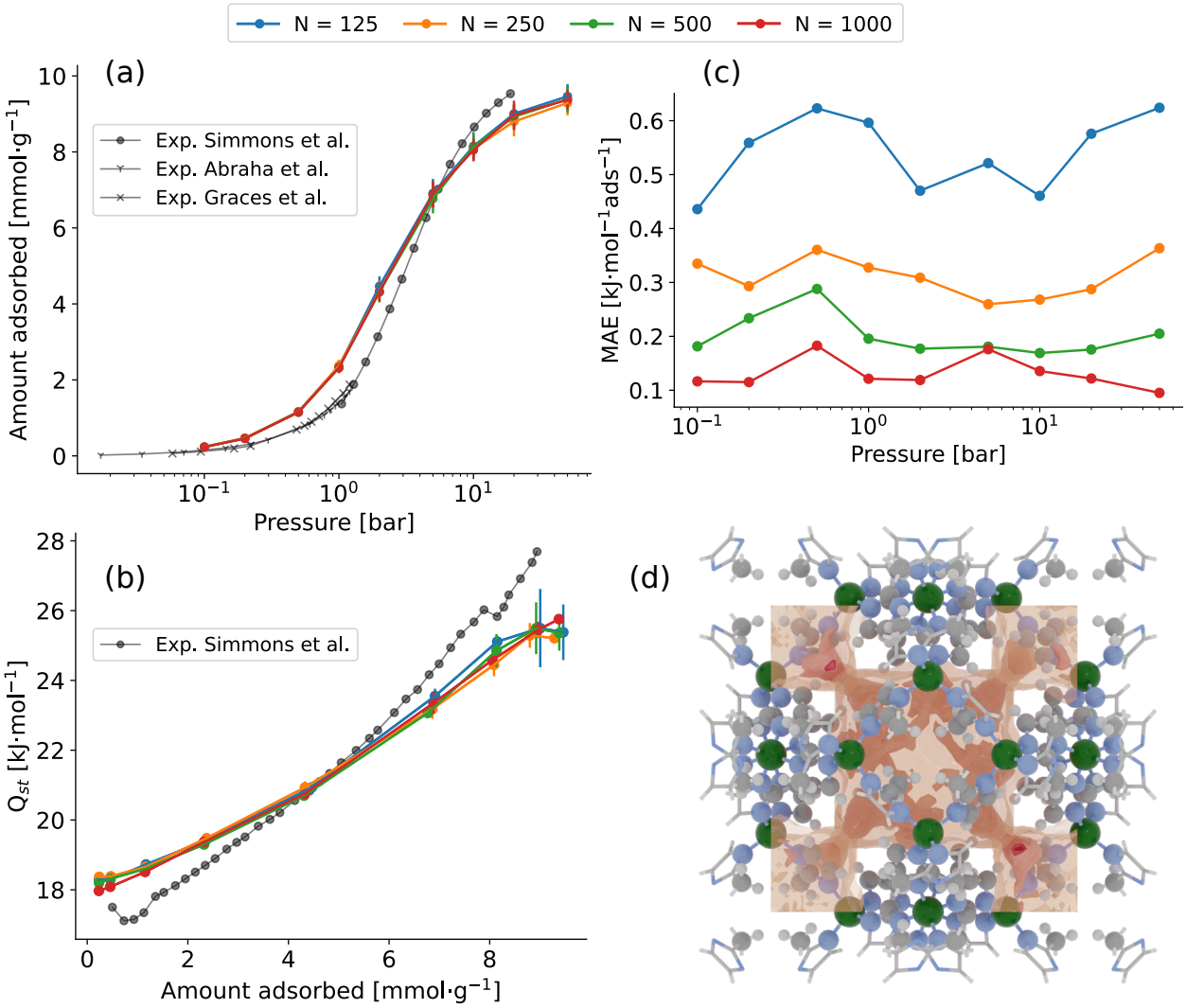

Figure 5: (a) The gas uptake and (b) isosteric heats of adsorption of CO_2 in ZIF-8 obtained with MLPs trained on between 125 and 1000 optimizations. Error bars are obtained from block averaging. (c) The test MAE on the energy of snapshots extracted from GCMC simulations as a function of the gas pressure and (d) a density plot of adsorbate atoms extracted from the MLP trained on 1000 optimizations at 10 bar.

loadings, in good agreement with experimentally derived heats of adsorption.[57] This increase in heat of adsorption as a function of the CO_2 loading is indicative for the relatively weak host-guest interactions in ZIF-8 compared to the guest-guest interactions. Even though the most favorable interaction sites for an adsorbate in ZIF-8 are already occupied at higher guest loadings, the heat of adsorption still increases due to the guest-guest stabilization.

To determine the accuracy of the performed GCMC simulations compared to the reference level of theory, 50 snapshots were taken from the GCMC simulations performed with each of the MLPs at each gas pressure. The interaction energy of these snapshots was recomputed with the reference PBE-D3(BJ) level of theory. As shown in Figure 5(c), the resulting average MAE on the energy decreases with increasing N from 0.54 to 0.13 kJ·mol^{-1} per adsorbate. Already for $N = 250$ (representing 7500 total DFT evaluations), a MAE smaller than 0.40 kJ·mol^{-1} per adsorbate is attained. As illustration, the density of adsorbate oxygen atoms extracted from the GCMC simulations using $N = 1000$ at a pressure of 10 bar is shown in Figure 5(d).

Carbon dioxide adsorption in Mg-MOF-74

Having validated our proposed methodology, it is now applied to the challenging case of CO_2 adsorption on the open-metal sites present in Mg-MOF-74. Due to the undercoordination of the metal, CO_2 molecules can approach the Mg site in this MOF to within 2.2 Å, which is closer than would be allowed by the tabulated van der Waals radius for Mg.[45,63] Therefore, the Mg van der Waals radius was set to 1.0 Å, allowing insertions of CO_2 in the framework to within 1.8 Å of the Mg metal site. This choice was validated *a posteriori*, as was done for ZIF-8 in Figure 4, and is shown in Section 3 of the Supporting Information.

From the literature, it is known that the interaction energy of a CO_2 molecule with the Mg open-metal site in Mg-MOF-74 depends significantly on the choice of dispersion correction.[70] Therefore, 1000 optimizations were performed in the framework at the PBE-D3(BJ), PBE-TS and PBE-MBD levels of theory. This selection of dispersion corrections

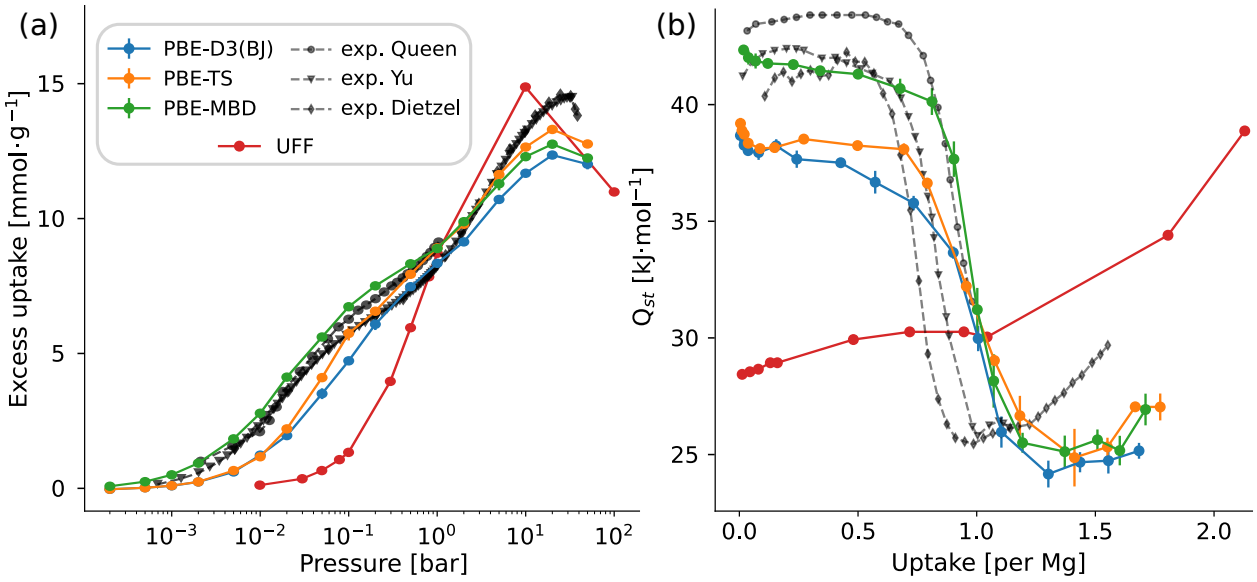

Figure 6: (a) The excess CO_2 uptake and (b) isosteric heats of adsorption in Mg-MOF-74 computed with MLPs trained on PBE-D3(BJ), PBE-TS and PBE-MBD reference optimizations, compared to experiments by Dietzel *et al.*,[71] Yu *et al.*[72] and Queen *et al.*[73] and UFF simulations.[45] All MLP GCMC simulations were performed at 298 K and pressures between $2{\cdot}10^{-4}$ and 50 bar.

was inspired by previous work by Rehak *et al.* in which the adsorption energy on the Mg site obtained with different levels of theory was benchmarked.[70] Therein, the PBE-TS and PBE-MBD functionals most closely agreed with the experimentally derived adsorption energy. Therefore, these two functionals, as well as the commonly used PBE-D3(BJ) functional are considered further.

Training an MLP on the energies and forces results in a validation MAE on the energies of 0.16, 0.20 and 0.16 kJ·mol^{-1} per adsorbate for the PBE-D3(BJ), PBE-TS and PBE-MBD functionals, respectively. Subsequently, GCMC simulations were performed for $2.5{\cdot}10^6$ iterations at 298 K and pressures between $2{\cdot}10^{-4}$ and 50 bar. The results are shown in Figure 6 and compared to experiments performed by Dietzel, Yu and Queen *et al.*, as well as simulations performed previously with the Universal Force Field (UFF).[45,71–74] As seen in Figure 6(a), UFF underestimates the uptake at low pressures and overestimates the uptake at the highest pressures. This is due to its limited accuracy in describing the interaction strength with the Mg site (underestimating the heats of adsorption at low uptakes, see Figure 6(b)), as well as

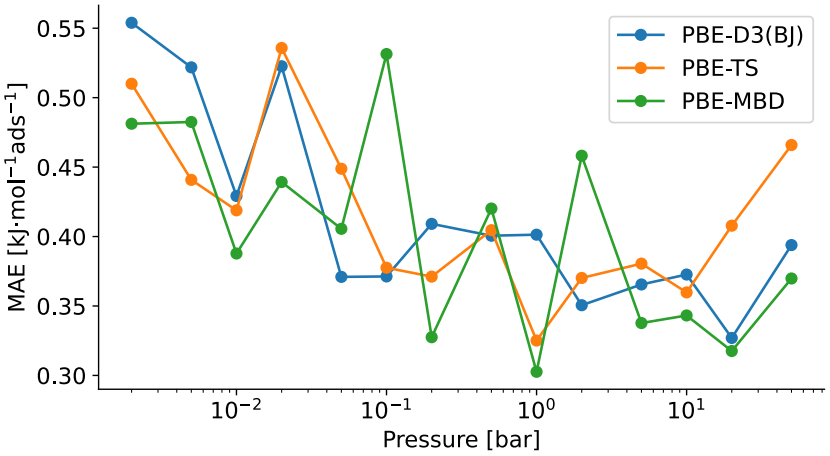

Figure 7: The test MAE on the energy of snapshots extracted from GCMC simulations as a function of the gas pressure for the MLPs trained on PBE-D3(BJ), PBE-TS and PBE-MBD reference optimizations, respectively.

the mutual interaction between CO_2 adsorbates (as seen from the heats of adsorption at the highest uptakes). GCMC simulations performed with the MLPs at each of the three levels of theory significantly improves the description of guest adsorption in the framework. Between these, PBE-MBD performs best, accurately capturing the shape of the adsorption isotherm at pressures below 1 bar. Moreover, as seen in Figure 6(b), the MLP trained on PBE-MBD reference data results in highly accurate heats of adsorption. Compared to the dependence of the heats of adsorption on the uptake seen for ZIF-8 (see Figure 5(b)), the opposite behavior is observed. At loadings below 1 adsorbate per Mg site, a heat of adsorption of approximately 42 kJ·mol^{-1} is obtained, in good agreement with experiments.[71–73] At higher loadings, the heat of adsorption suddenly drops to around 26 kJ·mol^{-1} due to the full occupation of Mg interaction sites, as also observed experimentally. This investigation demonstrates how, compared to UFF, a model which captures the interaction energy with the framework and other adsorbates with high accuracy can result in quantitative predictions of the adsorption properties in an open-metal site containing MOF.

As before, 50 snapshots were extracted from simulations at each gas pressure for the three levels of theory and recomputed. As shown in Figure 7, the MAEs are equal to 0.41, 0.42 and

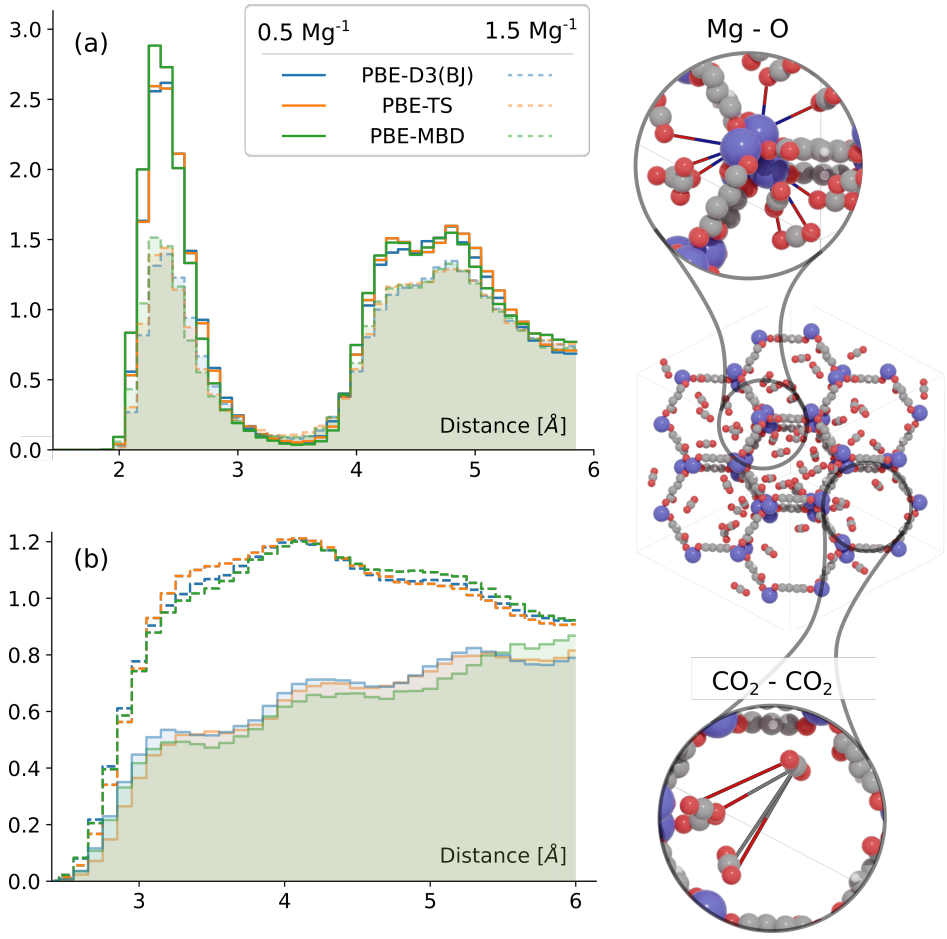

Figure 8: (a) RDF of magnesium–adsorbate oxygen atoms and (b) RDF of the adsorbate–adsorbate atoms computed from GCMC calculations in Mg-MOF-74 at each of the three levels of theory. The full lines correspond to a loading of half an adsorbate per magnesium site, and the dotted lines correspond to a loading of one and a half adsorbates per magnesium site.

0.40 kJ·mol^{-1} per adsorbate for PBE-D3(BJ), PBE-TS and PBE-MBD, respectively. A slight increase in error is seen at the lowest pressures, as for these snapshots all CO_2 molecules are adsorbed on Mg sites, on which the interaction is most difficult to capture. The errors are significantly higher than for CO_2 adsorption in ZIF-8, and can be rationalized by the more diverse adsorption environment in Mg-MOF-74, with significantly stronger interactions on the Mg site compared to the linkers.[51] However, these errors are still deemed acceptable, as differences between the levels of theory can be significantly larger (as seen from the heats of adsorption shown in Figure 6(b)).

To investigate the difference in structure of adsorbed CO_2 in the region of high heats of adsorption at low loadings and lower heats of adsorption at higher loadings (see Figure 6(b)) as well as differences between functionals, radial distribution functions (RDFs) of the magnesium–adsorbate oxyen atoms and adsorbate–adsorbate atoms are computed and shown in Figures 8(a) and 8(b), respectively. Comparing the Mg–O RDFs at low and high adsorbate loadings (0.5 adsorbates versus 1.5 adsorbates per Mg site), shown in Figure 8(a), it is clear that at low loadings guests preferentially adsorb at the Mg site, with a peak in the RDF around 2.4 Å and a secondary broad peak between 4 and 5 Å, representing the interaction with a neighboring Mg site. Compared to this, the high framework loading exhibits a significantly lower first peak as more adsorbates are located in the middle of the pores, further away from the Mg sites. The difference between simulations performed with MLPs trained on each of the three levels of theory is minor. Only in the first peak, a slight shift of approximately 0.05 Å is seen for PBE-MBD compared to the other functionals. In Figure 8(b), the RDF of adsorbate atoms is shown at low and high loadings. As seen from the low occupation of distances between 3 and 4 Å at low loadings, adsorbates are mostly separated from each other, located at each Mg site. Contrary to this, due to the filling of the center of the pore at higher loadings, interatomic distances between 3 and 4 Å are much more represented, with a peak around 4 Å. Again, only minor differences are observed between different functionals.

An investigation into transferability

Given an MLP trained on interaction energies of an adsorbate in a MOF consisting of certain building blocks, one could be interested in modeling adsorption in MOFs comprised of the same building blocks but with different topologies. This approach is exemplified in most common force fields making use of atom types: parameters derived for atoms in certain chemical environments are assumed to transfer well between different systems containing those environments. To investigate whether this concept of transferability can be applied to non-covalent interactions in MOFs modeled by MLPs, GCMC simulations were performed on ZIF-3, ZIF-4 and ZIF-6 (see Figure 9(a)) by use of the MLP trained on 1000 optimizations of CO_2–loaded ZIF-8 (denoted MLP_{ZIF8}). This approach is shown schematically in Figure 9(b), with the resulting isotherms and heats of adsorption shown as dotted curves in Figure 9(c). While the results seem acceptable at first sight, a problem arises when recomputing extracted snapshots at the PBE-D3(BJ) level of theory. MAEs of 2.38, 8.27 and 0.97 $kJ \cdot mol^{-1}$ are obtained for ZIF-3, ZIF-4 and ZIF-6.

As comparison to this, a single MLP was trained instead on 250 optimization in each of the four ZIFs (denoted $MLP_{ZIF\{3,4,6,8\}}$). Note that the same number of total optimizations (1000) are performed for this MLP as for MLP_{ZIF8}. This approach is shown in Figure 9(d), and the resulting adsorption isotherms and heats of adsorption are shown in Figure 9(c) with solid lines. A significantly different adsorption behavior is observed, with consistently higher heats of adsorption and uptakes. For $MLP_{ZIF\{3,4,6,8\}}$, recomputing the extracted snapshots results in MAEs of 0.40, 0.39, 0.32 and 0.32 $kJ \cdot mol^{-1}$ for ZIF-3, ZIF-4, ZIF-6 and ZIF-8. From the difference in heats of adsorption obtained from both methods, it is clear that MLP_{ZIF8} is not capable of reproducing the adsorption behavior in the other ZIFs. This can be rationalized by their substantially different density, probe-occupiable pore volume and surface area (computed using a nitrogen probe with PoreBlazer,[75] see Table 1). For example, ZIF-4 is the densest of the four MOFs, resulting in a significantly stronger interaction of an adsorbate with the adsorbent (as seen from the heats of adsorption

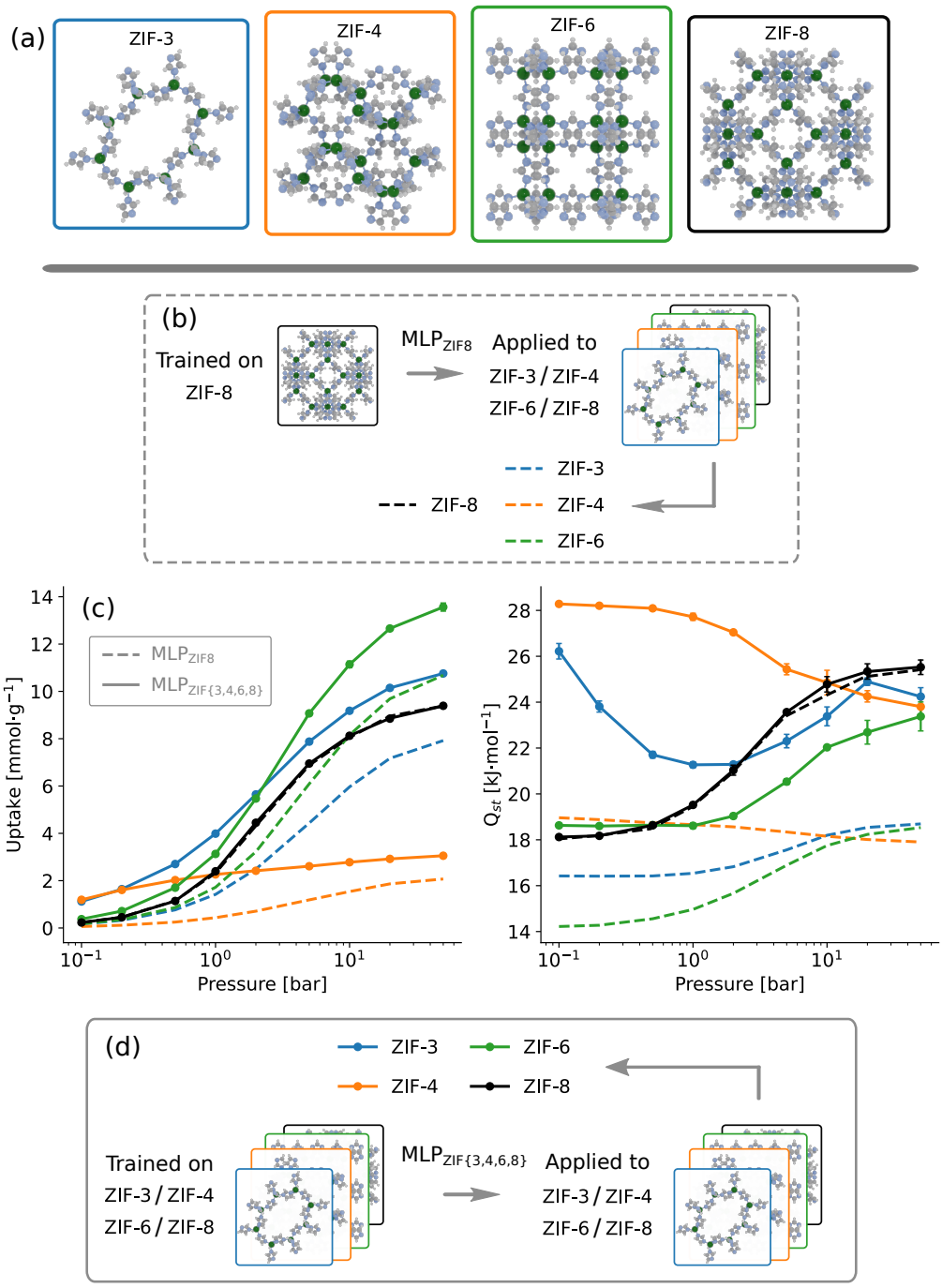

Figure 9: (a) Illustrations of ZIF-3, ZIF-4, ZIF-6 and ZIF-8. (b) An MLP trained on optimizations in ZIF-8 (MLP_{ZIF8}) is applied to perform GCMC simulations in all of the four ZIFs. (c) The uptake and heats of adsorption as a function of the gas pressure obtained from GCMC simulations obtained with both MLP_{ZIF8} in dotted lines and $MLP_{ZIF\{3,4,6,8\}}$ in full lines. (d) An MLP is trained on optimizations in all four ZIFs ($MLP_{ZIF\{3,4,6,8\}}$) and subsequently applied to each.

Table 1: Topology, density, probe-occupiable pore volume (V_{PO}) and surface area of all four ZIFs, computed with PoreBlazer.[75]

	Topology	Density [g/cm^3]	V_{PO} [cm^3/g]	Surface area [m^2/g]
ZIF-3	dft	0.88	0.61	1476
ZIF-4	cag	1.22	0.22	166
ZIF-6	gls	0.76	0.80	2612
ZIF-8	sod	0.92	0.52	1173

in the low pressure regime from Figure 9(c)). Even for ZIF-6, on which MLP_{ZIF8} exhibited a MAE of 0.97 kJ·mol^{-1}, heats of adsorption consistently off by approximately 4 kJ·mol^{-1} are obtained compared to $MLP_{ZIF\{3,4,6,8\}}$. The reason for this discrepancy is the content of the test set. Due to the lacking ability of MLP_{ZIF8} to recognize the favorable adsorption locations in the ZIF-6 framework, the test set does not include these locations. Therefore, the test error is deceptively low, while the actual error at these sites is significantly higher. Only when explicitly training an MLP to reference data which includes these important sites, this failure is alleviated.

To validate whether this statement holds true, it is possible to compute the heats of adsorption at infinite dilution at the reference PBE-D3(BJ) level of theory by means of an importance sampling approach, proposed by Vandenbrande *et al.*[48] First, a set of 10^6 random insertions in each of the ZIFs are constructed for which the $MLP_{ZIF\{3,4,6,8\}}$ interaction energies are computed. Subsequently, a set of $N_{\text{importance}}$ insertions are extracted from this wider set with a probability proportional to a Boltzmann factor. Recomputing those $N_{\text{importance}}$ snapshots, the heats of adsorption at the reference level of theory can be obtained from the following Equation:

$$Q_{st} = \frac{\langle \Delta U \exp\left(-\beta(\Delta U - \Delta\tilde{U})\right)\rangle}{\langle \exp\left(-\beta(\Delta U - \Delta\tilde{U})\right)\rangle} \tag{4}$$

with $\Delta\tilde{U}$ and ΔU the interaction energies computed with the MLP and PBE-D3(BJ), respectively, $\beta = 1/k_B T$ and the ensemble average a canonical average over the MLP energies. In case the MLP accurately reproduces the reference PES, only a limited number of insertions $N_{\text{importance}}$ are required, keeping the computational cost for this procedure manageable. In

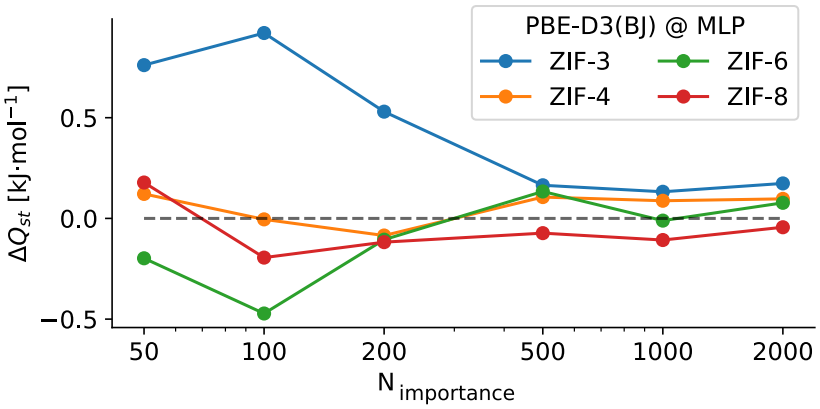

Figure 10: Difference between the MLP and reference PBE-D3(BJ) heats of adsorption for all four ZIFs. MLP results are obtained from 10^6 Widom insertions, and the reference heats of adsorption are obtained from recomputing importance sampled configurations from the MLP insertions. Convergence with respect to the number of recomputed snapshots $N_{\text{importance}}$ is shown for $N_{\text{importance}}$ between 50 and 2000.

Figure 10, the difference between the MLP and PBE-D3(BJ) importance sampled heats of adsorption obtained in this manner are shown. With $N_{\text{importance}} = 1000$, convergence within 0.10 kJ·mol^{-1} is obtained. As can be seen, the heats of adsorption obtained with the MLP compare well to those computed from importance sampled PBE-D3(BJ) insertions. For all four ZIFs, the error on the predicted heats of adsorption are lower than 0.20 kJ·mol^{-1}, further validating the derived MLP.

This investigation highlights how even seemingly well-performing MLPs can fail in unforeseen ways when sampling parts of the PES on which it was not adequately trained. While such failures can be detected by validation on optimizations performed at the reference DFT level of theory, this implies additional computationally demanding calculations. This clear failure in transferability can be attributed to the short-range nature of the employed MLPs. As the MLPs only describe the short-range interactions within a certain cutoff radius, longer ranged interactions are not explicitly modeled, and are only incorporated implicitly through error cancellation with the short-range interactions. Therefore, the merging of an explicit model for the long-range interactions by classical force fields with the short-range description by neural networks appears the most promising method to improve transferability. For example, MLPs could be trained to predict short-range atomic energies as well trainable atomic

force field parameters such as charges or Lennard-Jones parameters.[76] In this way, both the short-range and long-range interactions are modeled explicitly. However, such approaches were not investigated in this work, as these would require extensive benchmarking on which functional forms and trainable force field parameters increase the transferability across sets of MOFs with different topologies.

However, from the results obtained in this work, it is clear that accurate MLPs can be derived from a dataset containing between 250 and 1000 optimizations performed at the DFT level of theory (between 7500 and 30000 single point calculations, respectively). To illustrate the computational speedup enabled by the MLP compared to reference DFT calculations, a single energy evaluation of a guest-loaded ZIF-8 framework was performed with both methods on a single core of an Intel Xeon Gold 6140 processor. On this CPU, a DFT calculations takes approximately 6300 seconds, while the same energy evaluation takes only 490 milliseconds with the MLP. This represents a speedup of more than four orders of magnitude. To speed up both the training and use of the MLPs for GCMC simulations further, NVIDIA V100 GPUs were employed. On one such GPU, training an MLP takes on average 27 hours, while a typical GCMC calculation consisting of $5 \cdot 10^6$ iterations can be performed in approximately 50 hours. It is therefore clear that properly trained MLPs can vastly extend the applicability of DFT calculations, and can be expected to be of significant help in modeling adsorption in challenging systems such as open-metal site containing MOFs.

Conclusions

In this work, we proposed a generally applicable methodology for performing GCMC simulations by use of MLPs trained on DFT intermolecular reference energies and forces. First, an appropriate dataset was constructed to train the MLPs. For this, snapshots of guest-loaded frameworks are generated based on van der Waals 'exclusion spheres' and subsequently (rigidly) optimized using the relevant level of theory. With 7500 single point calculations

(250 trajectories, with 30 optimizations steps each) a test MAE on the energy of less than 0.4 kJ·mol^{-1} per adsorbate can be achieved, as calculated from snapshots extracted from GCMC simulations performed with the MLP. Having initially benchmarked the methodology on CO_2 adsorption in ZIF-8, it was subsequently applied to model CO_2 adsorption in Mg-MOF-74, an open-metal site containing MOF which has attracted attention for its interesting adsorption properties, as well as the challenge in modeling these properties computationally using force fields. Comparing three different dispersion corrections combined with the PBE functional for modeling adsorption in this MOF, PBE-MBD showed excellent agreement between the computed and experimentally obtained adsorption isotherms and heats of adsorption at 298 K. The characteristic drop in the heat of adsorption of CO_2 in this MOF between low and high uptakes from approximately 42 kJ·mol^{-1} to 26 kJ·mol^{-1} is quantitatively predicted from the performed GCMC simulations. Lastly, also the transferability of an MLP trained on ZIF-8 towards other ZIFs was tested. In this case, significant under-predictions of the uptakes and heats of adsorption were obtained for a set of other ZIFs composed of the same building blocks. This investigation highlighted the difficulty in training transferable MLPs, warranting caution when MLPs are employed to model adsorption in MOFs with structural differences to those/that in the training set. Only when explicitly training to data for all ZIFs, accurate results are obtained.

The general methodology proposed here has potential to be widely applicable not only to model adsorption in MOFs, but also in other nanoporous materials such as covalent organic frameworks (COFs), zeolites or graphene derivatives. By training an MLP on limited set of DFT reference data, a 'best of both worlds' compromise is made between the accuracy of DFT and the computational efficiency of force fields, unlocking the potential for routine GCMC simulations at DFT accuracy. Moreover, it could be employed as a tool to benchmark computationally cheaper classical force fields where experimental reference data is not available. Therefore, this work holds potential to aid accurate computational screening of adsorption properties in nanoporous materials, supporting application-focused future

research.

Care should still be taken in choosing an accurate reference level of theory. As shown in this work, the choice of dispersion correction alone can affect the obtained heat of adsorption of CO_2 in Mg-MOF-74 by as much as 5 kJ·mol^{-1}. For MOFs containing transition metals, the presence of multiple magnetic configurations could further complicate the generation of accurate reference data.[48,72] Secondly, this work only considered rigid frameworks. However, it could easily be extended to flexible adsorbents by including different framework geometries and training on the total energy and atomic forces (not just the intermolecular energy and forces). With this, hybrid GCMC/MD schemes could be employed to properly sample the adsorbate-induced flexibility of frameworks.[77]

Computational details

DFT optimizations

All single point calculations were performed with VASP, using the projected augmented wavefunction (PAW) PBE potentials.[78–81] The electronic convergence threshold for the energy was set to 10^{-5} eV. For ZIF-8, a single k-point was used, while for Mg-MOF-74 a $2\times2\times2$ k-point mesh was used for all PBE-D3(BJ) and PBE-TS calculations and a $3\times3\times3$ k-point mesh for all PBE-MBD calculations (see Supporting Information). An appropriate value for the cutoff energy for CO_2 adsorption in ZIF-8 was determined by performing 20 optimizations of guest-loaded frameworks for 30 steps using a cutoff of 700 eV. The final 20 structures were recomputed with cutoff energies between 350 eV and 650 eV. As shown in the Supporting Information, the interaction energy is converged to within 0.05 kJ·mol^{-1} using a cutoff energy of 500 eV. For Mg-MOF-74, a cutoff energy of 600 eV was used in accordance with previous work.[70] Furthermore, as the difference between calculations performed with the precision set to `normal` compared to `accurate` is also within 0.05 kJ·mol^{-1}, the precision was set to `normal`. For the optimizations, the BFGS optimizer from ASE was used with a

maximum step size of 0.5 Å.[64]

MLP hyperparameters

The effect of different hyperparameters of the network on both the accuracy and computational efficiency are tested and are included in the Supporting Information. The used hyperparameters are as follows: a cutoff radius of 5 Å is used, combined with 4 interaction layers, 32 features and a maximum rotation order of 1. The weights $\lambda_E = 50$ and $\lambda_F = 1$ are used to train the network. Although forces are not strictly necessary (only the energy is required in GCMC simulations), including forces is known to improve data efficiency during training.

Acknowledgement

The authors acknowledge the financial support from the Research Board of Ghent University (BOF), the Flemish Government and Flanders Innovation & Entrepreneurship (VLAIO) through the Moonshot project MOONRISE, the computational resources (Stevin Supercomputer Infrastructure) and services provided by the VSC (Flemish Supercomputer Center), funded by Ghent University, FWO, and the Flemish Government, department EWI. The authors would like to thank dr. Jelle Vekeman for his helpful comments to improve the quality of the manuscript.

Supporting Information Available

Benchmarking of the used DFT settings, determination of the MLP hyperparameters and validation of the used van der Waals radius for the Mg site in Mg-MOF-74. A zip file containing the simulation data and software implementations is archived on Zenodo (10.5281/zenodo.7904959).

References

(1) Li, J. R.; Kuppler, R. J.; Zhou, H. C. Selective gas adsorption and separation in metal-organic frameworks. *Chemical Society Reviews* **2009**, *38*, 1477–1504.

(2) Li, J. R.; Ma, Y.; McCarthy, M. C.; Sculley, J.; Yu, J.; Jeong, H. K.; Balbuena, P. B.; Zhou, H. C. Carbon dioxide capture-related gas adsorption and separation in metal-organic frameworks. 2011.

(3) Sumida, K.; Rogow, D. L.; Mason, J. A.; McDonald, T. M.; Bloch, E. D.; Herm, Z. R.; Bae, T. H.; Long, J. R. Carbon dioxide capture in metal-organic frameworks. *Chemical Reviews* **2012**, *112*, 724–781.

(4) Li, H.; Eddaoudi, M.; Groy, T. L.; Yaghi, O. M. Establishing microporosity in open metal-organic frameworks: Gas sorption isotherms for Zn(BDC) (BDC = 1,4-benzenedicarboxylate) [28]. *Journal of the American Chemical Society* **1998**, *120*, 8571–8572.

(5) Yaghi, O. M.; Li, H.; Davis, C.; Richardson, D.; Groy, T. L. *Synthetic Strategies, Structure Patterns, and Emerging Properties in the Chemistry of Modular Porous Solids* ; 1998.

(6) Kepert, C. J.; Rosseinsky, M. J. Zeolite-like crystal structure of an empty microporous molecular framework. *Chemical Communications* **1999**, 375–376.

(7) Li, H.; Eddaoudi, M.; O'Keeffe, M.; Yaghi, O. M. Design and synthesis of an exceptionally stable and highly porous metal- organic framework. *Nature* **1999**, *402*, 276–279.

(8) Chui, S. S.; Lo, S. M.; Charmant, J. P.; Orpen, A. G.; Williams, I. D. A chemically functionalizable nanoporous material [Cu3(TMA)2 (H2O)3](n). *Science* **1999**, *283*, 1148–1150.

(9) Eddaoudi, M.; Kim, J.; Rosi, N.; Vodak, D.; Wachter, J.; O'Keeffe, M.; Yaghi, O. M. Systematic design of pore size and functionality in isoreticular MOFs and their application in methane storage. *Science* **2002**, *295*, 469–472.

(10) Klontzas, E.; Mavrandonakis, A.; Tylianakis, E.; Froudakis, G. E. Improving hydrogen storage capacity of MOF by functionalization of the organic linker with lithium atoms. *Nano Letters* **2008**, *8*, 1572–1576.

(11) Lyu, J.; Zhang, X.; Otake, K. I.; Wang, X.; Li, P.; Li, Z.; Chen, Z.; Zhang, Y.; Wasson, M. C.; Yang, Y.; Bai, P.; Guo, X.; Islamoglu, T.; Farha, O. K. Topology and porosity control of metal-organic frameworks through linker functionalization. *Chemical Science* **2019**, *10*, 1186–1192.

(12) Ali Akbar Razavi, S.; Morsali, A. Linker functionalized metal-organic frameworks. *Coordination Chemistry Reviews* **2019**, *399*, 213023.

(13) Düren, T.; Bae, Y. S.; Snurr, R. Q. Using molecular simulation to characterise metalorganic frameworks for adsorption applications. *Chemical Society Reviews* **2009**, *38*, 1237–1247.

(14) Han, S. S.; Mendoza-Cortés, J. L.; Goddard, W. A. Recent advances on simulation and theory of hydrogen storage in metalorganic frameworks and covalent organic frameworks. *Chemical Society Reviews* **2009**, *38*, 1460–1476.

(15) Getman, R. B.; Bae, Y. S.; Wilmer, C. E.; Snurr, R. Q. Review and analysis of molecular simulations of methane, hydrogen, and acetylene storage in metal-organic frameworks. *Chemical Reviews* **2012**, *112*, 703–723.

(16) Yang, Q.; Liu, D.; Zhong, C.; Li, J. R. Development of computational methodologies for metal-organic frameworks and their application in gas separations. *Chemical Reviews* **2013**, *113*, 8261–8323.

(17) Wu, D.; Wang, C.; Liu, B.; Liu, D.; Yang, Q.; Zhong, C. Large-scale computational screening of metal-organic frameworks for CH4/H2 separation. *AIChE Journal* **2012**, *58*, 2078–2084.

(18) Colón, Y. J.; Snurr, R. Q. High-throughput computational screening of metal-organic frameworks. *Chemical Society Reviews* **2014**, *43*, 5735–5749.

(19) Yu, J.; Xie, L. H.; Li, J. R.; Ma, Y.; Seminario, J. M.; Balbuena, P. B. CO2 Capture and Separations Using MOFs: Computational and Experimental Studies. *Chemical Reviews* **2017**, *117*, 9674–9754.

(20) Wilmer, C. E.; Leaf, M.; Lee, C. Y.; Farha, O. K.; Hauser, B. G.; Hupp, J. T.; Snurr, R. Q. Large-scale screening of hypothetical metal-organic frameworks. *Nature Chemistry* **2012**, *4*, 83–89.

(21) Gómez-Gualdrón, D. A.; Colón, Y. J.; Zhang, X.; Wang, T. C.; Chen, Y. S.; Hupp, J. T.; Yildirim, T.; Farha, O. K.; Zhang, J.; Snurr, R. Q. Evaluating topologically diverse metal-organic frameworks for cryo-adsorbed hydrogen storage. *Energy and Environmental Science* **2016**, *9*, 3279–3289.

(22) Lin, L. C.; Berger, A. H.; Martin, R. L.; Kim, J.; Swisher, J. A.; Jariwala, K.; Rycroft, C. H.; Bhown, A. S.; Deem, M. W.; Haranczyk, M.; Smit, B. In silico screening of carbon-capture materials. *Nature Materials* **2012**, *11*, 633–641.

(23) Kim, J.; Martin, R. L.; Rübel, O.; Haranczyk, M.; Smit, B. High-Throughput Characterization of Porous Materials Using Graphics Processing Units. *Journal of Chemical Theory and Computation* **2012**, *8*, 1684–1693.

(24) Ahmed, A.; Siegel, D. J. Predicting hydrogen storage in MOFs via machine learning. *Patterns* **2021**, *2*, 100291.

(25) Altintas, C.; Altundal, O. F.; Keskin, S.; Yildirim, R. Machine Learning Meets with Metal Organic Frameworks for Gas Storage and Separation. 2021; `https://doi.org/10.1021/acs.jcim.1c00191`.

(26) Guo, W.; Liu, J.; Dong, F.; Chen, R.; Das, J.; Ge, W.; Xu, X.; Hong, H. Deep Learning Models for Predicting Gas Adsorption Capacity of Nanomaterials. *Nanomaterials* **2022**, *12*, 3376.

(27) Borboudakis, G.; Stergiannakos, T.; Frysali, M.; Klontzas, E.; Tsamardinos, I.; Froudakis, G. E. Chemically intuited, large-scale screening of MOFs by machine learning techniques. *npj Computational Materials* **2017**, *3*, 40.

(28) Shi, K.; Li, Z.; Anstine, D. M.; Tang, D.; Colina, C. M.; Sholl, D. S.; Siepmann, J. I.; Snurr, R. Q. Two-Dimensional Energy Histograms as Features for Machine Learning to Predict Adsorption in Diverse Nanoporous Materials. *Journal of Chemical Theory and Computation* **2023**, *19*, 4568–4583.

(29) Gheytanzadeh, M.; Baghban, A.; Habibzadeh, S.; Esmaeili, A.; Abida, O.; Mohaddespour, A.; Muhammad, .; Munir, T. Towards estimation of CO2 adsorption on highly porous MOF-based adsorbents using gaussian process regression approach. *Scientific Reports* — **2021**, *11*, 15710.

(30) Hung, T. H.; Xu, Z. X.; Kang, D. Y.; Lin, L. C. Chemistry-Encoded Convolutional Neural Networks for Predicting Gaseous Adsorption in Porous Materials. *Journal of Physical Chemistry C* **2022**, *126*, 2813–2822.

(31) Rappé, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations. *Journal of the American Chemical Society* **1992**, *114*, 10024–10035.

(32) Mayo, S. L.; Olafson, B. D.; Goddard, W. A. DREIDING: A generic force field for molecular simulations. *Journal of Physical Chemistry* **1990**, *94*, 8897–8909.

(33) Allinger, N. L.; Yuh, Y. H.; Lii, J. H. Molecular Mechanics. The MM3 Force Field for Hydrocarbons. 1. *Journal of the American Chemical Society* **1989**, *111*, 8551–8566.

(34) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general Amber force field. *Journal of Computational Chemistry* **2004**, *25*, 1157–1174.

(35) Addicoat, M. A.; Vankova, N.; Akter, I. F.; Heine, T. Extension of the universal force field to metal-organic frameworks. *Journal of Chemical Theory and Computation* **2014**, *10*, 880–891.

(36) Bureekaew, S.; Amirjalayer, S.; Tafipolsky, M.; Spickermann, C.; Roy, T. K.; Schmid, R. MOF-FF - A flexible first-principles derived force field for metal-organic frameworks. *Physica Status Solidi (B) Basic Research* **2013**, *250*, 1128–1141.

(37) Boulanger, E.; Huang, L.; Rupakheti, C.; Mackerell, A. D.; Roux, B. Optimized Lennard-Jones Parameters for Druglike Small Molecules. *Journal of Chemical Theory and Computation* **2018**, *14*, 3121–3131.

(38) Wu, X.; Huang, J.; Cai, W.; Jaroniec, M. Force field for ZIF-8 flexible frameworks: Atomistic simulation of adsorption, diffusion of pure gases as CH4, H2, CO 2 and N2. *RSC Advances* **2014**, *4*, 16503–16511.

(39) Pérez-Pellitero, J.; Amrouche, H.; Siperstein, F. R.; Pirngruber, G.; Nieto-Draghi, C.; Chaplais, G.; Simon-Masseron, A.; Bazer-Bachi, D.; Peralta, D.; Bats, N. Adsorption of CO2, CH4, and N2 on zeolitic imidazolate frameworks: Experiments and simulations. *Chemistry - A European Journal* **2010**, *16*, 1560–1571.

(40) Fischer, M.; Gomes, J. R.; Jorge, M. Computational approaches to study adsorption in MOFs with unsaturated metal sites. *Molecular Simulation* **2014**, *40*, 537–556.

(41) Belof, J. L.; Stern, A. C.; Eddaoudi, M.; Space, B. On the mechanism of hydrogen storage in a metal-organic framework material. *Journal of the American Chemical Society* **2007**, *129*, 15202–15210.

(42) Franz, D. M.; Dyott, Z. E.; Forrest, K. A.; Hogan, A.; Pham, T.; Space, B. Simulations of hydrogen, carbon dioxide, and small hydrocarbon sorption in a nitrogen-rich rht-metal-organic framework. *Physical Chemistry Chemical Physics* **2018**, *20*, 1761–1777.

(43) Becker, T. M.; Heinen, J.; Dubbeldam, D.; Lin, L. C.; Vlugt, T. J. Polarizable Force Fields for CO2 and CH4 Adsorption in M-MOF-74. *Journal of Physical Chemistry C* **2017**, *121*, 4659–4673.

(44) Becker, T. M.; Luna-Triguero, A.; Vicent-Luna, J. M.; Lin, L. C.; Dubbeldam, D.; Calero, S.; Vlugt, T. J. Potential of polarizable force fields for predicting the separation performance of small hydrocarbons in M-MOF-74. *Physical Chemistry Chemical Physics* **2018**, *20*, 28848–28859.

(45) Becker, T. M.; Lin, L. C.; Dubbeldam, D.; Vlugt, T. J. Polarizable Force Field for CO2 in M-MOF-74 Derived from Quantum Mechanics. *Journal of Physical Chemistry C* **2018**, *122*, 24488–24498.

(46) Chang, C.; Deringer, V. L.; Katti, K. S.; Van Speybroeck, V.; Wolverton, C. M. Simulations in the era of exascale computing. *Nature Reviews Materials* **2023**, 1–5.

(47) Lee, Y.; Poloni, R.; Kim, J. Probing gas adsorption in MOFs using an efficient ab initio widom insertion Monte Carlo method. *Journal of Computational Chemistry* **2016**, *37*, 2808–2815.

(48) Vandenbrande, S.; Waroquier, M.; Van Speybroeck, V.; Verstraelen, T. Ab Initio Evaluation of Henry Coefficients Using Importance Sampling. *Journal of Chemical Theory and Computation* **2018**, *14*, 6359–6369.

(49) Kundu, A.; Piccini, G.; Sillar, K.; Sauer, J. Ab Initio Prediction of Adsorption Isotherms for Small Molecules in Metal-Organic Frameworks. *Journal of the American Chemical Society* **2016**, *138*, 14047–14056.

(50) Sillar, K.; Kundu, A.; Sauer, J. Ab Initio Adsorption Isotherms for Molecules with Lateral Interactions: CO2 in Metal-Organic Frameworks. *Journal of Physical Chemistry C* **2017**, *121*, 12789–12799.

(51) Kundu, A.; Sillar, K.; Sauer, J. Ab Initio Prediction of Adsorption Isotherms for Gas Mixtures by Grand Canonical Monte Carlo Simulations on a Lattice of Sites. *Journal of Physical Chemistry Letters* **2017**, *8*, 2713–2718.

(52) Batzner, S.; Musaelian, A.; Sun, L.; Geiger, M.; Mailoa, J. P.; Kornbluth, M.; Molinari, N.; Smidt, T. E.; Kozinsky, B. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature Communications* **2022**, *13*, 1–11.

(53) Eckhoff, M.; Behler, J. From Molecular Fragments to the Bulk: Development of a Neural Network Potential for MOF-5. *Journal of Chemical Theory and Computation* **2019**, *15*, 3793–3809.

(54) Achar, S. K.; Wardzala, J. J.; Bernasconi, L.; Zhang, L.; Johnson, J. K. Combined Deep Learning and Classical Potential Approach for Modeling Diffusion in UiO-66. *Journal of Chemical Theory and Computation* **2022**, *18*, 3593–3606.

(55) Yu, Y.; Zhang, W.; Mei, D. Artificial Neural Network Potential for Encapsulated Platinum Clusters in MOF-808. *Journal of Physical Chemistry C* **2022**, *126*, 1204–1214.

(56) Vandenhaute, S.; Cools-Ceuppens, M.; DeKeyser, S.; Verstraelen, T.; Van Speybroeck, V. Machine learning potentials for metal-organic frameworks using an incremental learning approach. *npj Computational Materials* **2023**, *9*, 1–8.

(57) Simmons, J. M.; Wu, H.; Zhou, W.; Yildirim, T. Carbon capture in metal-organic frameworks - A comparative study. *Energy and Environmental Science* **2011**, *4*, 2177–2185.

(58) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Physical Review Letters* **1996**, *77*, 3865–3868.

(59) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *Journal of Chemical Physics* **2010**, *132*, 154104.

(60) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *Journal of Computational Chemistry* **2011**, *32*, 1456–1465.

(61) Tkatchenko, A.; Scheffler, M. Accurate molecular van der Waals interactions from ground-state electron density and free-atom reference data. *Physical Review Letters* **2009**, *102*, 073005.

(62) Ambrosetti, A.; Reilly, A. M.; Distasio, R. A.; Tkatchenko, A. Long-range correlation energy calculated from coupled atomic response functions. *Journal of Chemical Physics* **2014**, *140*, 18A508.

(63) Alvarez, S. A cartography of the van der Waals territories. *Dalton Transactions* **2013**, *42*, 8617–8636.

(64) Hjorth Larsen, A.; JØrgen Mortensen, J.; Blomqvist, J.; Castelli, I. E.; Christensen, R.; Dułak, M.; Friis, J.; Groves, M. N.; Hammer, B.; Hargus, C.; Hermes, E. D.; Jennings, P. C.; Bjerre Jensen, P.; Kermode, J.; Kitchin, J. R.; Leonhard Kolsbjerg, E.; Kubal, J.; Kaasbjerg, K.; Lysgaard, S.; Bergmann Maronsson, J.; Maxson, T.; Olsen, T.; Pastewka, L.; Peterson, A.; Rostgaard, C.; SchiØtz, J.; Schütt, O.;

Strange, M.; Thygesen, K. S.; Vegge, T.; Vilhelmsen, L.; Walter, M.; Zeng, Z.; Jacobsen, K. W. The atomic simulation environment - A Python library for working with atoms. *Journal of Physics Condensed Matter* **2017**, *29*, 30.

(65) Nicholson, D.; Parsonage, N. G. *Computer simulation and the statistical mechanics of adsorption*; Academic Press: London, 1982.

(66) Widom, B. Some topics in the theory of fluids. *The Journal of Chemical Physics* **1963**, *39*, 2808–2812.

(67) Abraha, Y. W.; Tsai, C. W.; Niemantsverdriet, J. W.; Langner, E. H. Optimized CO2 Capture of the Zeolitic Imidazolate Framework ZIF-8 Modified by Solvent-Assisted Ligand Exchange. *ACS Omega* **2021**, *6*, 21850–21860.

(68) Garcés, S. I.; Villarroel-Rocha, J.; Sapag, K.; Korili, S. A.; Gil, A. Comparative study of the adsorption equilibrium of CO2 on microporous commercial materials at low pressures. *Industrial and Engineering Chemistry Research* **2013**, *52*, 6785–6793.

(69) Fairen-Jimenez, D.; Moggach, S. A.; Wharmby, M. T.; Wright, P. A.; Parsons, S.; D Uren, T. Opening the Gate: Framework Flexibility in ZIF-8 Explored by Experiments and Simulations. *J. Am. Chem. Soc* **2011**, *133*, 8900–8902.

(70) Rehak, F. R.; Piccini, G.; Alessio, M.; Sauer, J. Including dispersion in density functional theory for adsorption on flat oxide surfaces, in metal-organic frameworks and in acidic zeolites. *Physical Chemistry Chemical Physics* **2020**, *22*, 7577–7585.

(71) Dietzel, P. D.; Besikiotis, V.; Blom, R. Application of metal-organic frameworks with coordinatively unsaturated metal sites in storage and separation of methane and carbon dioxide. *Journal of Materials Chemistry* **2009**, *19*, 7362–7370.

(72) Yu, D.; Yazaydin, A. O.; Lane, J. R.; Dietzel, P. D.; Snurr, R. Q. A combined experi-

mental and quantum chemical study of CO2 adsorption in the metal-organic framework CPO-27 with different metals. *Chemical Science* **2013**, *4*, 3544–3556.

(73) Queen, W. L.; Hudson, M. R.; Bloch, E. D.; Mason, J. A.; Gonzalez, M. I.; Lee, J. S.; Gygi, D.; Howe, J. D.; Lee, K.; Darwish, T. A.; James, M.; Peterson, V. K.; Teat, S. J.; Smit, B.; Neaton, J. B.; Long, J. R.; Brown, C. M. Comprehensive study of carbon dioxide adsorption in the metal-organic frameworks M2(dobdc) (M = Mg, Mn, Fe, Co, Ni, Cu, Zn). *Chemical Science* **2014**, *5*, 4569–4581.

(74) Rappé, A. K.; Goddard, W. A. Charge equilibration for molecular dynamics simulations. *Journal of Physical Chemistry* **1991**, *95*, 3358–3363.

(75) Sarkisov, L.; Bueno-Perez, R.; Sutharson, M.; Fairen-Jimenez, D. Materials Informatics with PoreBlazer v4.0 and the CSD MOF Database. *Chemistry of Materials* **2020**, *32*, 9849–9867.

(76) Behler, J.; Csányi, G. Machine learning potentials for extended systems: a perspective. *European Physical Journal B* **2021**, *94*, 142.

(77) Rogge, S. M. J.; Goeminne, R.; Demuynck, R.; GutiérrezSevillano, J. J.; Vandenbrande, S.; Vanduyfhuys, L.; Waroquier, M.; Verstraelen, T.; Van Speybroeck, V. Modeling Gas Adsorption in Flexible MetalOrganic Frameworks via Hybrid Monte Carlo/Molecular Dynamics Schemes. *Advanced Theory and Simulations* **2019**, *2*, 1800177.

(78) Kresse, G.; Hafner, J. Ab initio molecular dynamics for liquid metals. *Physical Review B* **1993**, *47*, 558–561.

(79) Kresse, G.; Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Physical Review B - Condensed Matter and Materials Physics* **1996**, *54*, 11169–11186.

(80) Kresse, G.; Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Computational Materials Science* **1996**, *6*, 15–50.

(81) Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Physical Review B - Condensed Matter and Materials Physics* **1999**, *59*, 1758–1775.

Graphical TOC Entry