

pubs.acs.org/jcim

Interpretable Graph Transformer Network for Predicting Adsorption Isotherms of Metal–Organic Frameworks

Pin Chen,^{||} Rui Jiao,^{||} Jinyu Liu, Yang Liu,* and Yutong Lu*



ABSTRACT: Predicting interactions between metal—organic frameworks (MOFs) and their adsorbates based on structures is critical to design high-performance porous materials. Many gas uptake prediction models have been proposed, but adsorption isotherm prediction is still challenging for most existing models. Here, we report a deep learning approach (MOFNet) that can predict adsorption isotherms for MOFs based on hierarchical representation and pressure adaptive mechanism. We elaborately design a hierarchical representation to encode the MOF structures. We adopt a graph transformer network to capture atomic-level information, which can help learn chemical features required under low-pressure conditions. A pressure adaptive mechanism is employed to interpolate and extrapolate the given limited data points by transfer learning, which can predict adsorption isotherms on a wider pressure range by only one model. We demonstrate that our predictor outperformed other traditional machine learning as well as graph neural network models on the challenging benchmarks and also achieves high performance on the real-world experimental observed adsorption isotherms. Finally, we interpret the models to discover and present potential structure—property relationships using the self-attention mechanism in the network. The proof-of-concept applications, such as disordered MOF predictions and missing data imputation of gas adsorption isotherms, showcase the generality and usability of our model to improve MOF material design.

INTRODUCTION

The rapid development of experimental and theoretical calculated databases has made machine learning (ML) methods play an important role in the field of material design and discovery. The combination of big data and ML has also been referred to as the "fourth paradigm of science".¹ In the past years, the usefulness of ML has been proved in the field of material property predicting,² surface adsorption,³ forcefield parameter fitting,⁴ protein structure predicting,^{5,6} and so forth.

Metal–organic frameworks (MOFs) have attracted increasing attention because of their capabilities in many applications, including gas separation⁷ and storage,^{8,9} sensing,¹⁰ and catalysis^{11,12} to name a few. The unique structural diversities of uniform pore structures, tunable porosity, flexibility in geometry, and chemical functionality make MOFs highly distinct from traditional materials. Experiments to collect gas uptake of MOFs are expensive and time-consuming. Therefore, research on gas uptake of MOFs commonly uses molecular dynamics (MD) or grand canonical Monte Carlo (GCMC) methods to obtain gas adsorption isotherms.^{13–15} However, these methods still have computing costs, which limit their application in large-scale material screening. Recently, ML methods have begun to be applied to gas uptake as well as gas adsorption isotherms. Traditional ML has been widely explored for predicting gas uptake of MOFs, for example, decision tree (DT),¹⁶ random forest (RF),¹⁷ Gaussian process (GP) regression,¹⁸ support vector machine (SVM),¹⁹ and so forth. In addition, deep learning models have also emerged for

Received: July 12, 2022 Published: November 1, 2022



gas uptake prediction, for example, multilayer perceptrons (MLPs),^{20–22} graph convolution neural network (GCN),²³ transfer learning,²⁴ and so forth.

Despite the remarkable success, current ML models in adsorption properties prediction still encounter several limitations. First, when the atoms and bonds of the MOF structure are represented as nodes and edges in graph neural networks, the calculation is very time-consuming due to the large number of atoms (ranging from hundreds to thousands). Too much redundant information can easily lead to over-fitting of the deep learning model. Second, the bottleneck of the adsorption isotherm prediction lies in low-pressure gas uptake. The reason behind this phenomenon is that the chemical contributions of the MOF pores are likely to play an important role in determining adsorption capacity. However, commonly used chemical representing methods such as chemical motif density, atomic property-weighted radial distribution functions (AP-RDF),²⁵ and Bag-of-Atom²¹ cannot fully capture the chemical information. Third, previous works mainly focused on predicting gas uptake at a specific gas pressure, and few studies have been done to predict the adsorption isotherm. This curve-like property prediction is more practical and challenging. One of the challenges is that the prediction model needs to accurately predict not only the adsorption capacities at each pressure but also the adsorption isotherm types.²⁶ Another challenge is that the efficient adsorption isotherm prediction model should have generalization capabilities to enable the interpolation and extrapolation from the labeled pressure conditions.

In this study, we propose a deep learning model, named MOFNet, to address the above limitations. In our model, we design a hierarchical representation to describe the structure graph of MOFs. We leverage both local and global environment features of MOFs in our model. The local environment features only use the minimum motif of MOFs based on the symmetric information to reduce redundant graph structures, which can capture fine-grained, atomic-scale information required under the low-pressure conditions. Meanwhile, the global features utilize the geometric structure information of the MOFs, which are critical contributions for adsorption under the high-pressure conditions. Specially, we employ a graph transformer to capture the key contributions of the atoms or functional groups relative to the adsorption properties. We design a pressure adaptive mechanism to interpolate and extrapolate the given limited data points by transfer learning, which can predict adsorption isotherm on a wider pressure range by only one model. We also construct a large-scale calculated adsorption capacity database with N₂, CO2, and CH4 based on a Cambridge Structural Database MOF subset (CSD-MOFDB) for training our model and manually collect data set containing experimental observed gas isotherms from the NIST/ARPA-E Database of Novel and Emerging Adsorbent Materials (NIST-ISODB)²⁷ to verify our model. Comparing with other traditional ML as well as graph neural network models, MOFNet achieves excellent performance for predicting the adsorption isotherm on the CSD-MOFDB data set. Furthermore, MOFNet also achieves high accuracy for predicting adsorption isotherms as well as adsorption isotherm types on NIST-ISODB data set. The proof-of-concept applications, such as disordered MOF predictions and missing data imputation of gas adsorption isotherms, showcase the generality and usability of our model to improve MOF material design.

METHODS

Data Set. CSD-MOFDB Data Set. The MOF database²⁸ contained 72,618 non-disordered structures integrated within the Cambridge Structural Database (CSD, version 5.4) were used. The solvent molecules were cleaned from the structures by using an in-house script based on the graph-labeling algorithm.²⁹ The important structural properties, including largest cavity diameter (LCD), pore-limiting diameter (PLD), helium void fraction (VF), and so forth, were calculated using open-source software Zeo++.30 We refined this data set by removing structures based on the PLD less than the diameter of three adsorbed molecules so that all three adsorbents can access to the MOFs' pores, and the diameters of N₂, CO₂, and CH₄ are 3.64, 3.3, and 3.8 Å, respectively. Some structures failed in GCMC calculation are also removed. As shown in Table S1, we finally eliminated those structures and ended up with 7304, 6997, and 8539 MOFs for N₂, CO₂, and CH₄, respectively. We depicted the structure distribution of MOFs for three adsorbents of N2, CO2, and CH4 in Figure S1 and selected the CH₄ data set to show the frequency of appearance of various elemental types in Figure S2. The results indicate that the atom type diversity of our data set contains 81 types in total, which covers the most used types in the element periodic table. We also evaluated the structure's diversity by comparing with other data sets in Figures S3 and S4. The detailed description of the calculated data set can be found in the Supporting Information.

GCMC simulations were carried out to calculate the adsorption data of MOFs for CO2, N2, and CH4 using RASPA software.³¹ A Lennard-Jones (LJ) plus Coulomb potential described the interactions between MOF frameworks and the gas molecules. The parameters for the MOF framework were obtained from the universal force field (UFF),³² while the TraPPE forcefield³³ was used for the guest model of CO2, N2, and CH4. The cutoff distance for truncation of the intermolecular interactions was set to 12.8 Å. The equilibration method $(QEq)^{34}$ as implemented in RASPA was assigned to frameworks. The supercells were used in simulations by replicating the minimum images to meet the requirement that each cell length was not less than 26 Å. Number of cycles were set to 10,000 with the first 5000 cycles for initialization and the last 1000 cycles for taking the ensemble averages. The frameworks were fixed during simulations to save calculation time. The external temperatures for CO₂, N₂, and CH₄ were set to 273, 77, and 298 K, respectively. We set eight pressure points from the range of 0.2-100, 50-50,000, and 50-10,000 kPa for N2, CO2, and CH₄, respectively. We evaluated the accuracy of our calculated data set by comparing with another data set in Figure S5. The detailed description of the calculated data set can be found in the Supporting Information.

NIST-ISODB Data Set. The NIST/ARPA-E database^{*a*} contains experimental gas adsorption data of MOFs, while few of them have atomic coordinate files. We obtained all gas adsorption capacities (CO₂ at 273 K, N₂ at 77 K, and CH₄ at 298 K) through the application programming interface (API) provided by the official website. We further manually collected MOF structures with crystallographic information file (CIF) that are not used in the CSD-MOFDB data set. Finally, we obtained 54 MOFs with 1876 pressure data points covering N₂, CO₂, and CH₄ adsorbate molecules.

MOFNet Model. *Hierarchical Representation*. The prediction of MOFNet originates from both local representation and global representation. Consider an MOF unit with N atoms and M global features. To aggregate the local feature, the input node feature of the MOF minimum unit is first transformed by an atom embedding network to generate the hidden node embedding H. Then, the embedding is processed by L transformer-style message passing layers $f_{local}(\mathcal{E}, \mathbf{H}, \mathbf{X})$ composed of self-attention and feed forward layers, where $\mathcal{E} \in \{0,1\}^{N \times N}$ and $\mathbf{X} \in \mathbb{R}^{N \times 3}$ denote the adjacency matrix and the atom coordinates of the MOF unit, respectively. In the other path, the global geometric features $T \in \mathbb{R}^{M}$ are fed into an MLP $f_{global}(T)$ to obtain the global representation of the MOF unit. Finally, the local and global representations are summed up and transformed by an MLP, where the final predicted gas uptake Y can be acquired. That is

$$Y = MLP(f_{local}(\mathcal{E}, H, X) + f_{global}(T))$$
⁽¹⁾

Local Node and Edge Features. To minimize the size of the graph, we used the minimum repeating unit to characterize local features. We used 81 element types, and classify the outside types as "other". Apart from the atom type, the other attributes of a node include its node degree, whether it is a hydrogen bond acceptor, whether it has chirality, whether it belongs to the ring system, whether it is a metal, and whether it is a spiro atom. The method of constructing the edge uses the interface provided by the CSD Python API to calculate the adjacency matrix.

Global Features. We used eight crystal geometric attributes, for example, gravimetric surface area (GSA), LCD, PLD, volumetric surface area (VSA), channel number (CN), crystal density, pore volume (PV), VF, and open metal site per cell (OMS).

We give the detailed information for local and global features in the Supporting Information (Table S2).

Graph Transformer. Message Passing in the MOF Unit. An MOF unit is a molecule graph that comprises atoms with interactions. To capture the inter-atomic information, we apply the message passing operation on the MOF unit. Particularly, we iteratively apply the following aggregation scheme on a fully connected graph formed by an MOF unit

$$m_i^{(l)} = \text{AGGREGATE}_{j \neq i} \psi(h_i^{(l)}, h_j^{(l)})$$
(2)

$$h_i^{(l+1)} = \text{UPDATE}(h_i^{(l)}, m_i^{(l)})$$
 (3)

where we employ the average aggregation operator as the aggregation function in eq 2 and a feed forward network after residual connection as the feature update in eq 3. ψ is a neural network parameterized layer that conveys the message between atoms *i* and *j*.

Graph Transformer Network. Considering the highly versatile structures of MOFs, we further employ a 3D-aware self-attention mechanism on the message passing to expand the representation capacity of MOFNet. Inspired by the molecule attention transformer module,³⁵ we adopt a threefold attention aggregation that simultaneously considers interactions from atom representation, bond connection, and 3D geometric information. For node *i*, *j* with the hidden state h_i , $h_j \in \mathbb{R}^h$, coordinates x_i , $x_i \in \mathbb{R}^3$, and the bond connection \mathcal{E}_{ii} denoting

pubs.acs.org/jcim

the i - j entry of the graph adjacency matrix, the transformerstyle message passing layer ψ is defined as

$$\psi(i, j) = \lambda_{a} \frac{\exp(g_{Q}(h_{i})^{T}g_{K}(h_{j}))}{\sum_{k} \exp(g_{Q}(h_{i})^{T}g_{K}(h_{k}))}g_{V}(h_{i})}$$

attention on node representation
$$+ \lambda_{b} \underbrace{\mathbb{I}[\mathcal{E}_{ij}]/d_{i}}_{\text{bond messages}} + \lambda_{d} \underbrace{\phi(||x_{i} - x_{j}||_{2})}_{\text{geometrical information}}$$
(4)

where λ_a , λ_b , and λ_g are the balanced coefficients for the hybrid message function. $g_{Q_j} g_{K_j} g_{V}$ are the query, key, and value functions parameterized by MLPs. d_i is the node degree of node *i*. ϕ is a function over the inter-atomic distances $||x_i - x_j||_n$, which satisfies the E(n) invariance such that the predicted gas absorption remains unchanged under all rotation, reflection, and translation transformations on the input MOF unit. We apply a set of radial Bessel basis functions to construct the invariant function ϕ as

$$\phi(d) = \sigma\left(\sum_{i=1}^{N} \alpha_i \frac{\sin\left(\frac{\pi d}{c}\right)}{d}\right) \phi(d)$$
(5)

where σ is the activation function with the formula of SiLU(x) = $\frac{x}{1 + e^{-x}}$, N is the number of radial Bessel basis functions, α_i is the learnable parameter, c = 5.0 is the cutoff distance, and $\varphi(d)$ is the cosine cutoff function as

$$\varphi(d) = \begin{cases} \frac{1}{2} \left(\cos\left(\frac{\pi d}{c}\right) + 1 \right) & d \le c \\ 0 & d > c \end{cases}$$
(6)

Pressure Adaptive Mechanism. In order to further predict the isotherm of MOFs under varying pressures, we propose a novel transfer learning approach to acquire the interpolated and extrapolated isotherm given a limited data regime.

For each MOF, we have calculated the gas uptake $\{Y^{(i)}\}_{i=1}^{K}$ at *K* different pressure levels $\{P_i\}_{i=0}^{K}$. To take pressure into consideration, we extend our model in eq 1 with pressure *P*

$$\tilde{Y} = \text{MLP}(f_{\text{local}}(\mathcal{E}, \mathbf{H}, \mathbf{X}) + f_{\text{global}}(\mathbf{T}) + (\log P - \log P_0)\beta_p)$$
(7)

where $\beta_{\rm p}$ is a trainable adapted tensor and P_0 is a basic pressure level.

To obtain the adapted model, we first pretrain the model with data under pressure P_0 by eq 1, and β_p is initialized randomly. As mentioned, the model pretrained at low pressure focuses more on the local representations, which leads to a better initialization of the graph transformer module f_{local} . Therefore, we empirically set P_0 as the lowest pressure in our simulated training data $P_0 = \min\{P_i\}_{i=0}^K$. As $P = P_0$, the adaptation term $(\log P - \log P_0)\beta_p$ is constant to 0 during the pretraining procedure.

After pretraining, we transfer the model across all available pressures $\{P_i\}_{i=1}^{K}$. Particularly, for each MOF unit in the training set, we first obtain the hierarchical representation f_r by summing up the local and global representations, that is, $f_r = f_{\text{local}}(\mathcal{E}, \mathbf{H}, \mathbf{X}) + f_{\text{global}}(\mathbf{T})$. We repeat $f_r K$ times, and each of them is assigned with an available pressure P_i .

pubs.acs.org/jcim





According to eq 7, the prediction at P_i is derived as $\tilde{Y}^{(i)} = \text{MLP}(f_r + (\log P - \log P_0)\beta_p)$. The predictions of multiple pressures $\{\tilde{Y}^{(i)}\}_{i=1}^K$ are then compared with the corresponding labels $\{Y^{(i)}\}_{i=1}^K$ and the optimization objective is the summation of losses from various pressures. That is

$$\mathcal{L} = \sum_{i=1}^{K} \left\| \tilde{Y}^{(i)} - Y^{(i)} \right\|^{2}$$
(8)

Experimental Details. Multiple hyperparameters were used in our MOFNet. After testing hyperparameters on the validation data set, the graph transformer network layers was set to 2 and the hidden state of each layer was set to 1024. We applied a three-layer MLP as the global feature encoder, and the dimensions of these three layers were set to 128, 512, and 1024. The model was trained for 300 epoch using a batch size of 32 MOF structures. We use mean square error (MSE) as the training loss function. We used Adam optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.98$, and $\epsilon = 10^{-9}$ with the warmup strategy proposed by

	N_2			CO ₂			CH_4		
	low	medium	high	low	medium	high	low	medium	high
model	0.2 kPa	20 kPa	100 kPa	10 kPa	500 kPa	$5 \times 10^4 \text{ kPa}$	50 kPa	200 kPa	10 ⁴ kPa
SVR	0.289	0.244	0.240	0.691	0.359	0.296	0.462	0.367	0.186
DT	0.355	0.304	0.305	0.732	0.424	0.350	0.548	0.444	0.249
GBRT	0.296	0.251	0.246	0.475	0.372	0.304	0.496	0.387	0.198
RF	0.273	0.231	0.227	0.671	0.342	0.278	0.443	0.348	0.182
SchNet	0.293	0.247	0.244	0.695	0.359	0.296	0.444	0.334	0.190
PaiNN	0.281	0.250	0.240	0.632	0.348	0.282	0.402	0.309	0.189
EGNN	0.295	0.249	0.245	0.697	0.356	0.297	0.445	0.334	0.188
DimeNet++	0.276	0.228	0.227	0.595	0.331	0.272	0.398	0.296	0.187
MOFNet-L ^b	0.377	0.379	0.379	0.770	0.405	0.382	0.515	0.405	0.406
MOFNet-G ^c	0.289	0.243	0.240	0.720	0.362	0.293	0.471	0.368	0.190
MOFNet	0.251	0.216	0.215	0.569	0.311	0.258	0.359	0.281	0.171

Table 1. MOFNet Predicting Performance (sMAPE) Comparing with Other Traditional Machine Learning Methods and GNN-Based Models^a

^{*a*}All models are trained on the same data set and display the predicting performance on the independent test set. Here, we only select three pressure points per each adsorption gas type in the picture. More pressure data points and metrics (MAE and PCC) can be found in Tables S7–S15. ^{*b*}MOFNet only trained on local features. ^{*c*}MOFNet only trained on global features.

ref 36. We follow the same warmup strategy to schedule the learning rate during the training steps. Specifically, the learning rate is a function of training steps as learning rate = $(d_{\text{model}}, \text{warmup steps})^{-0.5}$ min(steps/warmup steps, (warmup steps/ steps)^{0.5}), where d_{model} is the hidden dimension of the model. In our experiments, we set $d_{\text{model}} = 1024$ and warmup steps = 2000. Hence, the maximum learning rate is set to $(d_{\text{model}}, \text{warmup steps})^{-0.5} = 7 \times 10^{-4}$. More tested hyperparameters can be found in Table S3.

We used a 10-fold cross-validation method and randomly split our calculated data set 8:1:1 for training, validation, and testing. We give all the data sets used and the exact number of training data sizes, validation data sizes, and test data sizes in Table S4. Finally, the ensemble model with 10 trained models is used for testing the experimental data. We evaluated our MOFNet model by Pearson correlation coefficient (PCC), mean absolute error (MAE), and symmetric mean absolute percentage error (sMAPE). Specially, we adopted sMAPE measured based on relative errors which is suitable for evaluating the performance of the gas uptake at various pressures. We compared our model with both traditional ML algorithms (e.g., SVR, DT, GBRT, and RF) and graph-based neural networks (e.g., SchNet,³⁷ DimeNet++,³⁸ EGNN,³⁹ and PaiNN⁴⁰). All models trained on the same data as MOFNet. The detailed hyperparameters of traditional ML methods and graph-based neural networks can be found in Tables S5 and S6, respectively.

We focus on two tasks for predicting properties of MOFs. (1) Gas uptake prediction evaluates the ability of the model to predict gas adsorption values under different pressures for each gas on CSD-MOFDB and NIST-MOFDB databases. (2) Gas isotherm prediction evaluates the ability of the interpolation and extrapolation performance of our model on NIST-ISODB. Obtaining accurate gas adsorption isotherms is based on accurately predicting the gas uptake under different pressures. For the gas isotherm prediction, we require not only the predicted adsorption values to be close to the target but also the shape of the predicted adsorption isotherm to be consistent with the target. Therefore, we used the consistency of adsorption isotherm types as an additional evaluation metric.

DATA AND MODEL AVAILABILITY

To ensure reproducibility of the results, the CSD-MOFDB and NIST-ISODB data set used in this work has been made available as a compressed file at https://matgen.nscc-gz.cn/dataset.html. The MOFNet model and related data processing scripts have been released as an open-source code in a GitHub repository at https://github.com/Matgen-project/MOFNet.

RESULTS

MOFNet Architecture. As shown in Figure 1, MOFNet consists of two essential sub-models. The first model is the hierarchical representation (Figure 1a). The second model is the pressure adaptive mechanism (Figure 1b). We represent the structures of MOFs into two separate parts: the global part contains crystal geometric attributes, and the local part learns the fine-grained atomic and functional group information from a symmetric unit. We will demonstrate in the next section that the hierarchical representation approach focuses on different features under different pressure conditions. Specially, an end-to-end graph transformer network is applied to learn central features contributed to gas uptake. Finally, we use a pressure adaptor by a novel transfer learning approach to obtain the interpolated and extrapolated isometric curve given the limited data points.

Results of Predicting Performance on CSD-MOFDB Data Set. We compared our MOFNet with two lines of baseline models. The first line is traditional ML methods such as SVR,¹⁹ DT,¹⁶ GBRT,⁴¹ and RF,¹⁷ which have shown superior performance in gas uptake prediction. The second line is GNN-based deep learning methods. We involve SchNet,³⁷ DimeNet++,³⁸ EGNN,³⁹ and PaiNN⁴⁰ as state-of-the-art models which have been validated on other material benchmarks to verify that the graph transformer is a more suited local feature encoder.

We displayed our results of predicting performance on the calculated data set in Table 1. Our MOFNet outperforms the best traditional RF model. This phenomenon is in line with our expectations. Our predictor contains local representation, which is helpful to learn a more informative representation of MOFs. MOFNet also outperforms the best GNN-based DimeNet++ model, and the results are consistent in MAE



Figure 2. (a) Scatter plot of MOFNet on CH₄ at a high pressure of 10,000 kPa. (b) Case example of MOF isotherms (CSD code: BAZFUF).

and PCC performance, as shown in Tables S7–S15. The difference between MOFNet and other GNN models lies in the processing of local features, which shows that the self-attention mechanism of MOFNet can better capture local features related to adsorption properties.

In particular, we found that predicting the gas uptake at low pressure is a more challenging task for N_2 , CO_2 , and CH_4 . The sMAPE metric of N_2 , CO_2 , and CH_4 increases as the pressure decreases. Previous studies²¹ have explained that at high pressure, the geometric information dominates adsorption properties, so the ML models containing geometric descriptors achieved high performance. At low pressure, the chemical information related to the interactions between the gas guest and the MOF framework may play a more important role in determining the absorption capacities. We plotted the scatter points of predictions against simulation data for CH_4 at the pressure of 10^4 kPa, and MOFNet achieves a high accuracy on all metrics (Figure 2a). We give a detailed example of MOFs (CSD code: BAZFUF) in Figure 2b, and the adsorption isotherm of CH_4 is close to the simulation result.

We also carried out ablation studies on MOFNet to investigate factors that influence the performance of the proposed MOFNet framework. The results are shown in Tables S7–S15. MOFNet achieves the lowest sMAPE, MAE, RMSE, and PCC metrics against MOFNet-L (only using local features) and MOFNet-G (only using global features) for N₂, CO_{2} and CH_4 , respectively.

Results of Predicting Performance on NIST-ISODB Data Set. We further explored the interpolation and extrapolation performance of MOFNet on the real-world adsorption isotherms. We collected data from the NIST/ ARPA-E Database of Novel and Emerging Adsorbent Materials (NIST-ISODB),²⁷ which is a collection of compiled gas adsorption measurements in MOFs from both experimental and simulation sources for various gases under a wide range of conditions. The main challenges include the following: first, the pressure interval of the real adsorption isotherm is likely to be much wide, which requires the extrapolation of our predictor trained on limited pressure intervals. Second, some adsorption isotherms are very complex within specific pressure intervals, which requires the interpolation capability of MOFNet trained on only eight pressure points. Third, there are six types of representative adsorption isotherms proposed by IUPAC conventions²⁶ that reflect the relationship between porous structures and their adsorption types. Therefore, the isotherm predictor needs to not only predict the gas uptake at different pressures accurately but also generate the entire isotherm consistent with the adsorption type.

To verify our MOFNet, we obtained all gas adsorption data from NIST-ISODB and selected the data containing coordinate files. Finally, we collected 54 MOFs with 1876 data points covering N₂, CO₂, and CH₄. As shown in Figure 3a, our MOFNet achieves high performance in interpolated and extrapolated capacities. We checked all the isotherm types, and depicted the prediction accuracy in Figure 3b. We observed that MOFNet achieves the best MAE and PCC metric for CH₄ gas, while it obtains moderate accuracy in isotherm types. Specifically, we presented two case examples in Figure 3c,d. In Figure 3c, we showed a successful case by the MOF-74-Ni material. MOFNet accurately predicts the gas uptake at various pressures with a MAE of 0.155 and a PCC of 0.997 and obtains the consistent adsorption type (type I) with experimental observations. In another case example (Figure 3d), although MOFNet achieves high performance in MAE (0.919) and PCC (0.816) metrics, it predicted the wrong isotherm type. Our predictor gets type I isotherm, but the truth is type II. We found that the PLD of ZIF-8 is 3.27 Å, while the kinetic diameter of N_2 is 3.64 Å, which indicates that the ZIF-8 may have a larger PLD in the experimental environment. This may explain the biases when predicting the ZIF-8 adsorption isotherm at high pressures.

Interpretable Prediction of Gas Adsorption Isotherms in MOFs. We attempted to interpret our MOFNet predictor by determining which feature(s) a given ML model weighs most heavily when making the prediction. We first employed Shapley Additive Explanations (SHAP),⁴² a gametheoretic approach to model interpretation, to explore this issue. Figure 4a–c ranks the local and global features to MOFNet under low, medium, and high pressure conditions for N₂, CO₂, and CH₄, respectively. Some observations stand out







Figure 3. Predicting performance on real-world adsorption isotherms (NIST-ISODB data set): (a) the half violin plots of MAE and PCC for N_2 , CO_2 , and CH_4 on independent adsorption data points, respectively. SD means standard deviation. (b) The accuracy of isotherm types prediction for N_2 , CO_2 , and CH_4 , respectively. (c) Case example of the MOF-74-Ni isotherm. (d) Case example of the ZIF-8 isotherm.

when comparing these rankings. First, in all these three adsorption gases, the relative importance distributions of local features appear to narrow with the increase in pressure. This phenomenon further validates our hypothesis that as the pressure increases, the geometric features of MOFs gradually become the dominant factors. Second, the relative importance of local features is not the same in different gases. We can observe that the order of the distribution interval width in different gases is $CH_4 > CO_2 > N_2$, possibly because the temperature of N_2 is the lowest (77 K) compared to the others. In addition, N₂ is relatively inert and has a weaker interaction with MOFs. Third, PV, density, and VF are the most important global features to MOFNet, which confirms the findings of multiple high-throughput studies of adsorption of MOFs and other materials.^{22,43} Additionally, we can see that the relative importance of the PV descriptor varies greatly in CH₄ and CO₂ at different pressures, especially at high pressures.

Let us illustrate the findings of gas uptake sensitivity to pressure using a specific MOF example. As shown in Figure 4e, we depict the density of CH_4 sorbate component on a

framework calculated by GCMC methods at different pressures. At 50 kPa, the uptake loading is 0.601 mol/kg and few CH₄ adsorbed on the framework. Under this condition, the interactions between the CH4 molecules and the framework are rare, and the CH4 molecules are more likely to appear on the open metal sites or the specific functional group, which is also called the monolayer adsorption stage.²⁶ Thus, the chemical information of the framework correlates well with adsorption properties. As the pressure increases, we can see from Figure 4e that more CH₄ distributed in the pore of the framework with an uptake loading of 1.277 mol/kg at 200 kPa pressure. As a result, the geometric descriptors, for example, PLD, LCD and GSA, and so forth, start to work. When the high-pressure environment of 10,000 kPa is carried out, the maximum uptake loading reaches to 3.984 mol/kg. The cavity of the framework structure will be exploited by CH₄ adsorbed gas (multilayer adsorption stage), and thus, the geometric features will dominate the adsorption properties.

We further explored the local feature learned by MOFNet by determining which nodes (actually the chemical atoms and

pubs.acs.org/jcim



Figure 4. Feature importance, as ranked by SHAP for (a) N_{22} (b) CO_{22} , and (c) CH_4 . The feature's name is shown on the *y*-axis, and its corresponding importance is shown on the *x*-axis. (d) Molecule minimum unit of MOFs (CSD code: LAGCIH). (e) CH_4 uptake of LAGCIH material and its uptake behavior on low, medium (Med.), and high pressures. The atom surface is colored by blue (inside) and gray (outside). Heatmaps of self-attention weights from the first layer of MOFNet on a molecule minimum unit of LAGCIH for (f) N_{22} , (g) CO_{22} and (h) CH_4 .

functional groups) in the graph transformer network gain the most attention weight. We displayed the averaged attention matrix of the first layer learned by our MOFNet in Figure 4f—h for N₂, CO₂, and CH₄, respectively. It is observed that the attention matrixes focus more on the metals, such as the sliver and zinc atoms in the minimum MOF unit. This phenomenon consists with the prior knowledge that the gases are mostly absorbed around the open metal atoms, ^{44,45} so that the absorption is mainly determined by the metals. Overall, the self-attention mechanism considers the inter-atomic interactions in MOF units and captures the relationship between the predicted property and different types of atoms.

DISCUSSION

In this work, we establish an efficient method for describing the relationship between MOFs and its adsorbates to predict the curve-like properties. First, it enables us to train the big structures of MOFs collected in the CSD database by hierarchical representation, which is general and applicable to other macro-molecular systems with certain symmetry. Second, we have proved that the graph transformer can effectively capture the relationship between the MOF structure and gas uptake by focusing on open metal sites and functional groups. Unlike other models developed for single system property prediction, such as DimNet++, PaiNN, EGNN, and SchNet, the transformer model with the self-attention mechanism may be more suitable for two interacting systems. The reason behind this phenomenon is that some atoms of the interacting system (the parts that are in contact or close to each other) play a central role for the concerned properties, and the attention mechanism of the transformer is likely to help to capture the important features. Moreover, due to the good parallax of the transformer models, the self-supervised learning strategy has been widely used to pre-train these models, such as the work⁴⁶ in the biochemistry and the work⁴⁷ in biology. Third, in our proposed pressure adaptive mechanism, we introduce an adapted tensor to balance the contribution weights of global and local features at different pressures. This mechanism can interpolate and extrapolate the isotherm using limited data and thus allow us to predict the isotherm of MOFs under varying pressures. We also compare with the method by inputting pressure as the feature into MLPs,^{22,48} of which



Figure 5. (a) Ordered and (b) disordered structure of MOF-205. White, gray, red, and dark blue spheres stand for H, C, O, and Zn atoms, respectively. The atom surface is colored by blue (inside) and gray (outside). (c) MOFNet predictions of ordered and disordered structures of MOF-205. (d) Distribution of difference between predictions of the ordered and disordered structures.

prediction performance is worse than our method as shown in Table S16. The reason behind this could be the model pretrained at low pressure used in pressure adaptive mechanism focuses more on the local representations, which leads to a better initialization of the graph transformer module.

Meanwhile, we noticed that the node contributed weights of some non-metals in the structure are very weak, which may allow MOFNet to predict the disordered MOFs caused by these atoms with reasonable results. We depicted the ordered and disordered structure in Figure 5a,b. The disordered atoms are carbon. We found that the predictions of MOF-205 material in ordered and disordered structures are close (Figure 5c), and their MAE and PCC against experimental observations are 5.937 (PCC = 0.987) and 4.832 (PCC = 0.987), respectively. We further showed the prediction difference between these two structures against various pressures in Figure 5d, and those disordered atoms have weak effect at low pressures and little effect at high pressures. Additionally, due to the capacities of interpolation and extrapolation, MOFNet can potentially be applied to missing data imputation of gas adsorption isotherms in databases. We found reports of related attempts,⁴⁹ and our model may help to complete missing adsorption properties of nanoporous materials. In terms of the training data set, previous studies are mainly based on the hMOF data set,⁵⁰ which has limited diversity in types of metal atoms and metallic corners. Our MOFNet is trained on CSD-MOFDB, which has the most complete collection of MOF structures at present, with wide distribution of element types and more diverse structures.

CONCLUSIONS

In summary, we develop a deep learning model based on hierarchical representation, graph transformer, and pressure adaptive mechanism for predicting the gas adsorption isotherm of MOFs. To train our model, we construct the CSD-MOFDB data set with N_2 , CO_2 , and CH_4 calculated adsorption capacities for training and manually collect the NIST-ISODB data set containing experimental observed gas isotherms to verify our model. We first evaluate our model on the CSD-MOFDB data set, and MOFNet achieves excellent performance on predicting the gas absorption on various pressure levels compared with prior state-of-art methods. We further extend our model in a transfer learning fashion to learn an interpolation across arbitrary pressures and fit the isotherms. Finally, we assess MOFNet on real-world NIST-ISODB and discover that MOFNet, despite trained with simulation data, is capable of generalizing to real-world data. We showcase the generality and usability of our model in the proof-of-concept applications, which might be valuable to the material science community.

ASSOCIATED CONTENT

1 Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.2c00876.

Data set processing workflow and relevant data set size; detailed information for local and global environment features; evaluation for the CSD-MOFDB data set; hyperparameters for all ML models; and model performance on the CSD-MOFDB data set (PDF)

AUTHOR INFORMATION

Corresponding Authors

- Yang Liu Department of Computer Science and Technology, Tsinghua University, Beijing 100084, P. R. China; Institute for AI Industry Research, Tsinghua University, Beijing 100084, P. R. China; Email: liuyang2011@tsinghua.edu.cn
- Yutong Lu National Supercomputer Center in Guangzhou, School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou 510006, P. R. China; Email: luyutong@mail.sysu.edu.cn

Authors

- Pin Chen National Supercomputer Center in Guangzhou, School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou 510006, P. R. China; orcid.org/ 0000-0001-8746-9917
- Rui Jiao Department of Computer Science and Technology, Tsinghua University, Beijing 100084, P. R. China; Institute for AI Industry Research, Tsinghua University, Beijing 100084, P. R. China
- Jinyu Liu National Supercomputer Center in Guangzhou, School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou 510006, P. R. China

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.2c00876

Author Contributions

^{II}P.C. and R.J. contributed equally to this work.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the Program for Guangdong Introducing Innovative and Entrepreneurial Teams (no. 2016ZT06D211), Guangdong Province Key Area R&D Program (no. 2019B010940001) and Nature Science Foundation of China (no. U1811461).

ADDITIONAL NOTE

^{*a*}https://adsorbents.nist.gov.

REFERENCES

(1) Agrawal, A.; Choudhary, A. Perspective: Materials informatics and big data: Realization of the "fourth paradigm" of science in materials science. *APL Mater.* **2016**, *4*, 053208.

(2) Xie, T.; Grossman, J. C. Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties. *Phys. Rev. Lett.* **2018**, *120*, 145301.

(3) Back, S.; Yoon, J.; Tian, N.; Zhong, W.; Tran, K.; Ulissi, Z. W. Convolutional neural network of atomic surface structures to predict binding energies for high-throughput screening of catalysts. *J. Phys. Chem. Lett.* **2019**, *10*, 4401–4408.

(4) Wang, H.; Zhang, L.; Han, J.; E, W. DeePMD-kit: A deep learning package for many-body potential energy representation and molecular dynamics. *Comput. Phys. Commun.* **2018**, 228, 178–184.

(5) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S.; Ballard, A.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589.

(6) Baek, M.; DiMaio, F.; Anishchenko, I.; Dauparas, J.; Ovchinnikov, S.; Lee, G. R.; Wang, J.; Cong, Q.; Kinch, L. N.; Schaeffer, R. D.; Millán, C.; Park, H.; Adams, C.; Glassman, C. R.; DeGiovanni, A.; Pereira, J. H.; Rodrigues, A. V.; van Dijk, A. A.; Ebrecht, A. C.; Opperman, D. J.; Sagmeister, T.; Buhlheller, C.; Pavkov-Keller, T.; Rathinaswamy, M. K.; Dalwadi, U.; Yip, C. K.; Burke, J. E.; Garcia, K. C.; Grishin, N. V.; Adams, P. D.; Read, R. J.; Baker, D. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **2021**, *373*, 871–876.

(7) Zhou, D.-D.; Chen, P.; Wang, C.; Wang, S.-S.; Du, Y.; Yan, H.; Ye, Z.-M.; He, C.-T.; Huang, R.-K.; Mo, Z.-W.; Huang, N.-Y.; Zhang, J.-P. Intermediate-sized molecular sieving of styrene from larger and smaller analogues. *Nat. Mater.* **2019**, *18*, 994–998.

(8) Mason, J. A.; Veenstra, M.; Long, J. R. Evaluating metal-organic frameworks for natural gas storage. *Chem. Sci.* **2014**, *5*, 32–51.

(9) Li, B.; Wen, H. M.; Zhou, Z.; Chen, B. Porous Metal-Organic Frameworks for Gas Storage and Separation: What, How, and Why? J. Phys. Chem. Lett. 2014, 5, 3468–3479.

(10) Lustig, W.; Mukherjee, S.; Rudd, N.; Desai, A.; Li, J.; Ghosh, S. Metal-organic frameworks: functional luminescent and photonic materials for sensing applications. *Chem. Soc. Rev.* **201**7, *46*, 3242.

(11) Dhakshinamoorthy, A.; Li, Z.; Garcia, H. Catalysis and photocatalysis by metal organic frameworks. *Chem. Soc. Rev.* 2018, 47, 8134–8172.

(12) Zhu, L.; Liu, X.-Q.; Jiang, H.-L.; Sun, L.-B. Metal-Organic Frameworks for Heterogeneous Basic Catalysis. *Chem. Rev.* 2017, *117*, 8129–8176.

(13) Chen, C.; Zhang, M.; Zhang, W.; Bai, J. Stable Amide-Functionalized Metal-Organic Framework with Highly Selective CO2 Adsorption. *Inorg. Chem.* **2019**, *58*, 2729–2735.

(14) Zhang, H.; Snurr, R. Q. Computational Study of Water Adsorption in the Hydrophobic Metal-Organic Framework ZIF-8: Adsorption Mechanism and Acceleration of the Simulations. *J. Phys. Chem. C* 2017, *121*, 24000–24010.

(15) Park, J.; Suh, B. L.; Kim, J. Computational Design of a Photoresponsive Metal-Organic Framework for Post Combustion Carbon Capture. *J. Phys. Chem. C* **2020**, *124*, 13162–13167.

(16) Pardakhti, M.; Moharreri, E.; Wanik, D.; Suib, S. L.; Srivastava, R. Machine learning using combined structural and chemical descriptors for prediction of methane adsorption performance of metal organic frameworks (MOFs). ACS Comb. Sci. 2017, 19, 640–645.

(17) Fanourgakis, G. S.; Gkagkas, K.; Tylianakis, E.; Klontzas, E.; Froudakis, G. A robust machine learning algorithm for the prediction of methane adsorption in nanoporous materials. *J. Phys. Chem. A* **2019**, *123*, 6080–6087.

(18) Ohno, H.; Mukae, Y. Machine Learning Approach for Prediction and Search: Application to Methane Storage in a Metal-Organic Framework. *J. Phys. Chem. C* **2016**, *120*, 23963–23968.

(19) Anderson, R.; Rodgers, J.; Argueta, E.; Biong, A.; Gómez-Gualdrón, D. A. Role of pore chemistry and topology in the CO2 capture capabilities of MOFs: from molecular simulation to machine learning. *Chem. Mater.* **2018**, *30*, 6325–6337.

(20) Anderson, G.; Schweitzer, B.; Anderson, R.; Gómez-Gualdrón, D. A. Attainable volumetric targets for adsorption-based hydrogen storage in porous crystals: molecular simulation and machine learning. *J. Phys. Chem. C* **2018**, *123*, 120–130.

(21) Burner, J.; Schwiedrzik, L.; Krykunov, M.; Luo, J.; Boyd, P. G.; Woo, T. K. High-Performing Deep Learning Regression Models for Predicting Low-Pressure CO2 Adsorption Properties of Metal-Organic Frameworks. J. Phys. Chem. C 2020, 124, 27996–28005.

(22) Gurnani, R.; Yu, Z.; Kim, C.; Sholl, D. S.; Ramprasad, R. Interpretable Machine Learning-Based Predictions of Methane Uptake Isotherms in Metal-Organic Frameworks. *Chem. Mater.* **2021**, *33*, 3543–3552.

(23) Wang, R.; Zhong, Y.; Bi, L.; Yang, M.; Xu, D. Accelerating Discovery of Metal-Organic Frameworks for Methane Adsorption with Hierarchical Screening and Deep Learning. *ACS Appl. Mater. Interfaces* **2020**, *12*, 52797–52807.

(24) Ma, R.; Colón, Y. J.; Luo, T. Transfer Learning Study of Gas Adsorption in Metal-Organic Frameworks. *ACS Appl. Mater. Interfaces* **2020**, *12*, 34041–34048.

(25) Fernandez, M.; Boyd, P. G.; Daff, T. D.; Aghaji, M. Z.; Woo, T. K. Rapid and accurate machine learning recognition of high performing metal organic frameworks for CO2 capture. *J. Phys. Chem. Lett.* **2014**, *5*, 3056–3060.

(26) Thommes, M.; Kaneko, K.; Neimark, A. V.; Olivier, J. P.; Rodriguez-Reinoso, F.; Rouquerol, J.; Sing, K. S. Physisorption of gases, with special reference to the evaluation of surface area and pore size distribution (IUPAC Technical Report). *Pure Appl. Chem.* **2015**, *87*, 1051–1069.

(27) Siderius, D.; Shen, V.; Johnson, R., III; van Zee, R.Nist/arpa-e Database of Novel and Emerging Adsorbent Materials; National Institute of Standards and Technology: Gaithersburg, MD, 2014; Vol. 10, p T43882.

(28) Moghadam, P. Z.; Li, A.; Wiggin, S. B.; Tao, A.; Maloney, A. G.; Wood, P. A.; Ward, S. C.; Fairen-Jimenez, D. Development of a Cambridge Structural Database Subset: A Collection of Metal-Organic Frameworks for Past, Present, and Future. *Chem. Mater.* **2017**, *29*, 2618–2625.

(29) Goldsmith, J.; Wong-Foy, A. G.; Cafarella, M. J.; Siegel, D. J. Theoretical Limits of Hydrogen Storage in Metal-Organic Frameworks: Opportunities and Trade-Offs. *Chem. Mater.* **2013**, *25*, 3373–3382.

(30) Willems, T. F.; Rycroft, C. H.; Kazi, M.; Meza, J. C.; Haranczyk, M. Algorithms and tools for high-throughput geometrybased analysis of crystalline porous materials. *Microporous Mesoporous Mater.* **2012**, *149*, 134–141.

(31) 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.

(32) Rappe, A. K.; Casewit, C. J.; Colwell, K.; Goddard, W. A., III; Skiff, W. M. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.* **1992**, *114*, 10024–10035.

(33) Potoff, J. J.; Siepmann, J. I. Vapor-liquid equilibria of mixtures containing alkanes, carbon dioxide, and nitrogen. *AIChE J.* **2001**, *47*, 1676–1682.

(34) Wilmer, C. E.; Snurr, R. Q. Towards rapid computational screening of metal-organic frameworks for carbon dioxide capture:

Calculation of framework charges via charge equilibration. *Chem. Eng.* J. 2011, 171, 775–781.

pubs.acs.org/jcim

(35) Maziarka, Ł.; Danel, T.; Mucha, S.; Rataj, K.; Tabor, J.; Jastrzebski, S.Molecule attention transformer. **2020**, arXiv preprint arXiv:2002.08264,

(36) Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; Polosukhin, I.Attention is all you need. **2017**, pp 5998–6008, arXiv:1706.0376.

(37) Schütt, K. T.; Sauceda, H. E.; Kindermans, P.-J.; Tkatchenko, A.; Müller, K.-R. SchNet - A deep learning architecture for molecules and materials. *J. Chem. Phys.* **2018**, *148*, 241722.

(38) Klicpera, J.; Giri, S.; Margraf, J. T.; Günnemann, S.Fast and uncertainty-aware directional message passing for non-equilibrium molecules. **2020**. arXiv preprint arXiv:2011.14115,

(39) Satorras, V. G.; Hoogeboom, E.; Welling, M.E(n) equivariant graph neural networks. In *International Conference on Machine Learning*, 2021; pp 9323–9332.

(40) Schütt, K. T.; Unke, O. T.; Gastegger, M.Equivariant message passing for the prediction of tensorial properties and molecular spectra. **2021**, arXiv preprint arXiv:2102.03150,

(41) Friedman, J. H. Greedy function approximation: a gradient boosting machine. *Ann. Stat.* 2001, 29, 1189–1232.

(42) Lundberg, S. M.; Lee, S.-I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 4678.

(43) 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 metalorganic frameworks. *Nat. Chem.* **2012**, *4*, 83–89.

(44) Maia, R. A.; Louis, B.; Gao, W.; Wang, Q. CO2 adsorption mechanisms on MOFs: a case study of open metal sites, ultramicroporosity and flexible framework. *React. Chem. Eng.* **2021**, *6*, 1118–1133.

(45) Karra, J. R.; Walton, K. S. Effect of Open Metal Sites on Adsorption of Polar and Nonpolar Molecules in Metal–Organic Framework Cu-BTC. *Langmuir* **2008**, *24*, 8620–8626.

(46) Chen, D.; Zheng, J.; Wei, G.-W.; Pan, F. Extracting Predictive Representations from Hundreds of Millions of Molecules. *J. Phys. Chem. Lett.* **2021**, *12*, 10793–10801.

(47) Rives, A.; Meier, J.; Sercu, T.; Goyal, S.; Lin, Z.; Liu, J.; Guo, D.; Ott, M.; Zitnick, C. L.; Ma, J.; Fergus, R. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proc. Natl. Acad. Sci. U.S.A.* **2021**, *118*, No. e2016239118.

(48) Anderson, R.; Biong, A.; Gómez-Gualdrón, D. A. Adsorption isotherm predictions for multiple molecules in MOFs using the same deep learning model. *J. Chem. Theory Comput.* **2020**, *16*, 1271–1283.

(49) Sturluson, A.; Raza, A.; McConachie, G. D.; Siderius, D. W.; Fern, X. Z.; Simon, C. M. Recommendation System to Predict Missing Adsorption Properties of Nanoporous Materials. *Chem. Mater.* **2021**, *33*, 7203–7216.

(50) Boyd, P. G.; Chidambaram, A.; García-Díez, E.; Ireland, C. P.; Daff, T. D.; Bounds, R.; Gładysiak, A.; Schouwink, P.; Moosavi, S. M.; Maroto-Valer, M. M.; Reimer, J. A.; Navarro, J. A. R.; Woo, T. K.; Garcia, S.; Stylianou, K. C.; Smit, B. Data-driven design of metalorganic frameworks for wet flue gas CO2 capture. *Nature* **2019**, *576*, 253–256.