HyMARC Seedling: Optimized Hydrogen Adsorbents via Machine Learning and Crystal Engineering

Donald J. Siegel (Primary Contact, UM), Alauddin Ahmed (UM), Suresh Kuthuru (UM), Darpandeep Aulakh (UM), Adam Matzger (UM), Justin Purewal (Ford), and Mike Veenstra (Ford)

University of Michigan (UM)
Ann Arbor, MI 48109
Phone: (734) 764-4808
Email: djsiege@umich.edu

DOE Manager: Zeric Hulvey
Phone: 202-586-1570
Email: zeric.hulvey@ee.doe.gov

Contract Number: DE-EE0008093
Subcontractors: Ford Motor Company, Dearborn, MI
Project Start Date: September 1, 2017
Project End Date: September 30, 2021

Acknowledgment: This material is based upon work supported by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under the Fuel Cell Technologies Office (FCTO) under Award Number DE-EE0008093.

Project Objectives

Objective 1: Apply machine learning (ML) techniques to design and experimentally demonstrate new MOFs having usable volumetric capacities exceeding 50 H_2 g/L (single-crystal/pressure swing), without compromising gravimetric capacity, kinetic performance, or reversibility.

Objective 2: Control MOF crystal morphology and crystallite size distribution to increase packing density of target high capacity MOF by at least 30% (compared to its powder tap density) with less than 15% decrease in gravimetric performance.

Technical Scope Summary

This project aims to overcome volumetric limitations associated with physisorptive hydrogen storage at both the materials and system level. This goal will be achieved using a combination of computational techniques and experimental synthesis and testing. Our efforts will target storage media based on metal-organic frameworks (MOFs), a class of hydrogen adsorbents with highly tunable properties.

At the materials level, machine learning methods will be applied to our database of 476,007 real and hypothetical MOFs. This analysis will guide the discovery of new compounds that can break through the so-called Volumetric Ceiling. This performance ceiling was identified in our prior screening studies; it reveals that no known MOFs can surpass a usable volumetric capacity of 40 g H2/L (assuming an isothermal pressure swing between 100 and 5 bar at 77 K). In contrast to the conventional approach to MOF discovery, where capacity is predicted from a known crystal structure, this project aims to invert this process and 'reverse engineer' optimal MOFs with the aid of machine learning. Starting instead from a target hydrogen capacity, MOF crystallographic properties such as surface area, pore volume, etc. will be identified. Next, these properties will be translated into specific combinations of linkers, metal clusters, and MOF topologies consistent with those properties. The end result will be a catalog of MOFs that are purpose built for high volumetric capacity. The most promising compounds will be synthesized and assessed experimentally with respect to their usable hydrogen capacities.

At the system level, we will develop crystal growth and processing techniques that result in MOF- based adsorbent beds with low void fractions. Analysis by the Hydrogen Storage Engineering Center of Excellence has shown that inefficient materials packing can result in density reductions of more than 60% compared to the single-crystal level. Therefore, packing inefficiencies have the potential to negate improvements in volumetric performance achieved at the materials level. Strategies to engineer particle properties and pack adsorbents with low void fraction – beyond routine "shake and compress" techniques – have not been widely explored. This project aims to close this performance gap by developing synthetic procedures that optimize particle morphology and size distribution.

Organization of this Final Report

This final project report is presented in two parts. **Part 1** describes efforts focused on materials development. This includes the identification, synthesis, and characterization of MOFs that exhibit high volumetric and gravimetric capacities simultaneously. **Part 2** describes the outcome of crystal engineering by controlling MOF crystal morphology and crystallite size distribution.

PART 1: Materials Development

Introduction

Hydrogen (H_2) is considered to be a future automotive fuel.[1–6] This potential reflects its high specific energy compared to competing fuels such as natural gas and gasoline, and the ability of H_2 to be produced renewably and consumed without CO_2 emissions.[2,7] Nevertheless, the adoption of hydrogen in mobile applications such as fuel cell (FC) vehicles has been limited by its low volumetric energy density.[2,6,7] Consequently, the design of low-cost H_2 storage systems that overcome these volumetric limitations has been the focus of recent research.[4,8–12] At present, FC vehicles employ storage systems based on gaseous H_2 compressed to pressures up to 700 bar.[13] This approach is costly and can incur limitations in driving range.[7,11,13,14]

Storage based on adsorption in porous hosts is an alternative to high-pressure compression.[15] Due to their high gravimetric densities, fast kinetics, and reversibility, metal-organic frameworks (MOFs) have emerged as one of the most promising classes of hydrogen sorbents.[2,7] MOFs are crystalline materials formed by the self-assembly of inorganic metal clusters and organic linkers.[16–22] By virtue of their building-block structure and the large number of potential components, the number of MOFs is potentially limitless.[21–25] Further modifications to MOF chemistry can be achieved by introducing functional groups, substituting different metals, and by mixing metals and/or linkers.[26–28]

Despite these many possibilities, a relatively small fraction of MOFs have been synthesized.[29,30] While the crystal structures of these 'real' MOFs are available in the Cambridge Structural Database (CSD),[29,30] many exhibit disorder, missing atoms, or have negligible porosity; consequently, these materials are not immediately amenable to assessment via computational modeling.[29,31–35]

One way to bypass these complications is through computational design. To date, nearly a million 'hypothetical' MOFs have been reported, [1,36,45,46,37–44] and it is reasonable to expect that many more materials will be proposed.[47–51] High-throughput screening using Grand Canonical Monte Carlo (GCMC)[52–56] has been successful in identifying promising candidates with superior gas storage capacities on sub-sets of these catalogs.[36,38,39,46,50,57–60] Nevertheless, given the large number of possibilities, a systematic search across all of these materials is challenging even

Table 1: Summary of recent studies that use machine learning (ML) to predict H_2 adsorption in MOFs. ρ_{crys}, vf, vsa, mpd, lcd represent single crystal density, void fraction, volumetric surface area, maximum pore diameter, and largest cavity diameter, respectively. R^2, AUE, and RMSE represent the coefficient of determination, average unsigned error, and root-mean-square error, respectively.

Study	ML Features	ML Method	Properties Predicted	Accuracy
Anderson et al. (2019)[45]	Epsilon, temperature, pressure, ρ_{crys}, vf, vsa, mpd, lcd, alchemical catecholate site density, unit cell volume	Neural network[77]	Total volumetric H_2 for pressures 0.1, 1, 5, 35, 65, and 100 bar at 77, 160, and 295 K	AUE = 0.75 - 2.93 g-H_2 L^{-1}
Bucior et al. (2019)[82]	Energetics of MOF-guest interactions	Multilinear regression with LASSO[77]	Deliverable H_2 storage capacity between 2-100 bar at 77 K	R^2 = 0.96; MAE = 1.4 - 3.4 g-H_2 L^{-1}; RMSE = 3.1 - 4.4 g-H_2 L^{-1}
Borboudakis et al. (2017)[63]	92 binary features based on linker, metal cluster, and 12 functional groups	Ridge linear regression and support vector machine with polynomial/Gaussian kernel[77,79,80]	Total H_2 storage capacity at 1 bar and 77 K	AUE = 0.47 (ridge regression), 0.50 (SVM) g-H_2 g^{-1}-MOF
Thornton et al. (2017)[78]	Adsorption energy, ρ_{crys}, vf, gsa, vsa, lcd	Neural network[72]	Net H_2 capacity for pressure swing between 1 and 100 bar at 77 and 298 K	R^2 = 0.88; RMSE = 3.6 g-H_2 L^{-1}

with high-throughput techniques.[1,61] Furthermore, differences in the implementation (i.e., use of different temperature/pressure conditions or interatomic potentials) can complicate comparisons between screening studies. Thus, more efficient and consistent screening approaches are desirable for predicting the gas storage properties of MOFs in existing and future databases.

Machine learning (ML) could provide a path forward.[62–65] For ML to be helpful, access to high-quality training data is essential. Unfortunately, training on experimental H_2 storage data in MOFs is non-trivial:[1,2,6,66–68] experimental uptake data is generally restricted to a relatively small number of MOFs, and can depend sensitively upon the experimental conditions and the purity of the sample.[2,67,69] Employing a dataset based on a consistent set of computational predictions may be a better choice.[62,63]

Earlier work has demonstrated that accurate isotherms for H_2 uptake in MOFs can be predicted using the pseudo-Feynman-Hibbs potential (to describe H_2) combined with general interatomic potentials to describe the MOF.[1,2,6,70] This approach was used to screen a database of 5,309 real MOFs, from which IRMOF-20 was identified and experimentally demonstrated to have a favorable balance of high gravimetric and volumetric H_2 density.[2] In a follow-on study a larger database of 495,305 MOFs was compiled from several publicly-available databases (see Table S1 for details).[1,29,31,33,36–40,45] Following a pre-screen based on crystallographic properties and empirical correlations, the H_2 capacities of a subset of 43,777 MOFs were evaluated using GCMC. Three additional MOFs – SNU-70, UMCM-9, and PCN-610/NU-100 – were identified and shown experimentally to out-perform the leading MOF candidate, IRMOF-20.[1]

Table 2. MOF datasets employed in this study.

Source	Database Identity	Number of MOFs
Goldsmith et al.,[31] Chung et al.,[33] Moghadam et al.,[29] Groom et al.[30]	Real MOFs: UM[31]+CoRE[33]+CSD[29,30]	15,235
Chung et al.[34]	CoRE 2019[34]	14,142
Moghadam et al.,[29] Groom et al.[30]	*CSD 2017 additional[29,30]	48,696
Martin et al.[38]	Mail-order[38]	112
Bao et al.[46]	In-silico deliverable[46]	2,816
Bao et al.[39]	In-silico surface[39]	8, 885
Witman et al.[40]	MOF-74 analogs[40]	61
Colón et al.[59]	ToBaCCo[59]	13,512
Gomez-Gualdron et al.[45]	Zr-MOFs[45]	204
Wilmer et al.[36]	Northwestern[36]	137,000
Aghaji et al.,[37] Boyd et al.[91,92]	**Univ. of Ottawa[37,91,92]	317,462
Lan et al.[83]	BJT MOFs[83]	303,793
Chung et al.[41,93]	***R-WLLFHS[41,93]	51,163
Li et al.[84]	MTV[84]	11,555
Anderson et al.[42]	CSM-2018-I[42]	117
Anderson et al.[43]	CSM-2018-II[43]	32
Anderson et al.[44]	CSM-2019-I[44]	99
Ahmed et al.[1]	In-house[1]	18
	Total	**918,734**

*A subset of the CSD 2017 MOF dataset[29,30] whose crystallographic properties were found to exhibit extremely low values (e.g GSA ~ 0) in a prior study.
**A recent version of this database is available publicly;[91,92] however, the present study employs an earlier version[37] that was shared privately.
***A curated subset of the Northwestern[36] database.

The database of MOF properties[71] generated in these prior studies presents an opportunity to develop ML models that can predict H_2 uptake across even larger MOF datasets.[1,71] Table 1 summarizes prior ML studies of H_2 storage in MOFs. (Reports employing ML for other adsorbates such as CH_4,[72,73] CO_2[74,75], N_2[74,75] are summarized in Table S2.) To the best of our knowledge, ML was first used to predict H_2 uptake in compounds from the Nanoporous Materials Genome.[76] A neural network (NN)[77] was used to predict usable capacities on a test set of ~1000 compounds, including MOFs.[78] In the same year, Borboudakis et al.[63] predicated H_2 capacities in 100 MOFs using 92 binary features related to a MOF's linker, metal cluster, and functional group(s). Ridge linear regression (RR)[77,79,80] and support vector machine (SVM)[77,81] algorithms were used to predict gravimetric capacity. Later, Bucior et al[82] predicted the H_2 capacities of 50,000 MOFs extracted from the CSD using multi-linear regression (MLR).[77] The models were trained using the energetics of H_2-MOF interactions and the usable volumetric capacities predicted by GCMC. More recently, ML was used to predict H_2 storage capacities in 105 hypothetical MOFs constructed from 17 different topologies, 4 distinct metal clusters, and 5 unique organic linkers.[43] NN[77] models employing 11 features were trained to predict total volumetric uptake at various temperatures and pressures.[43]

Table 3. Machine learning regression algorithms employed in this work.

Machine Learning Algorithm	Abbreviation
Extremely Randomized Trees[76,83,103,104]	ERT
Boosted Decision Trees[76,92,102, 103, 104]	BDT
Bagging with Decision Trees[76,9093,103,104]	B/DT
Random Forest[76,90,94,103,104]	RF
Bagging with Random Forest[76,93,94,103,104]	B/RF
Gradient Boosting[76,92,95,102, 103, 104]	GB
Decision Trees[76,90,103104]	DT
Nu-Support Vector Machine with Radial Basis Function (RBF) Kernel[76,79,90,96,98,103104]	Nu-SVM/RBF-K
Support Vector Machine with RBF Kernel[76,79,90,97,98,103104]	SVM/RBF-K
Support Vector Machine with Linear Kernel[76,79,96,99,103104]	SVM/L-K
Linear Regression[76, 77, 78,99,100,103,104]	LR
Ridge Regression[76, 77, 78,99,100,103,104]	RR
K-Nearest Neighbors[76,90,101,103104]	K-NN
AdaBoost[76,92,102, 103, 104]	AB

Expanding upon these prior reports, the present study applies ML to explore a large database of 918,734 known and proposed MOFs. The database was assembled from a diverse collection of publicly-available MOF repositories,[1,29,41–45,83,84,31,33,34,36–40] and allows for a wide-ranging and consistent assessment of H_2 uptake in MOFs.

Here the extremely randomized trees (ERT)[77,85] algorithm was identified as the most accurate ML model for predicting H_2 uptake. A training set comprising 24,674 MOFs was sufficient to enable accurate predictions of usable capacities across 820,039 unseen compounds.[71] These predictions were made using a small set of seven crystallographic features as input: single-crystal density, pore volume, gravimetric & volumetric surface area, void fraction, largest cavity diameter, and pore limiting diameter. Importantly, ML identified 8,282 MOFs – 8,187 appropriate for pressure swing operation and 95 for temperature/pressure swing use – with the potential to exceed both the gravimetric and volumetric capacities of state-of-the-art materials. These compounds are comprised predominantly of hypothetical MOFs, and exhibit low densities (<0.31 g cm^{-3}) in combination with high surface areas (> 5,300 m^2 g^{-1}), void fractions (~0.90), and pore volumes (>3.3 cm^3 g^{-1}). In addition to identifying high-capacity MOFs, the relative importance of the input features is quantified; dependencies on the ML algorithm and training set size and are also assessed. The single most important features for predicting H_2 uptake are pore volume (for gravimetric capacity) and void fraction (for volumetric capacity). A simplified model using only two input features is demonstrated to predict capacities with high accuracy – within 0.2 wt.% and 1.4 g g-H_2 L^{-1} of more expensive Monte Carlo calculations. The ML models are available for use via the web,[86] allowing for rapid and accurate predictions of hydrogen capacities with only a small amount of structural data required as input.

Methodology

MOF database

A database of crystal structures for 918,734 MOFs was created by combining 19 existing databases.[1,29,41–45,83,84,31,33,34,36–40] Table 2 summarizes the source databases and the number of MOFs contained in each. Out of these 19 databases, only the UM,[31] CSD,[29,30] and CoRE[33,34] databases contain data on MOFs that have been previously

synthesized. (MOFs listed in these datasets are referred to as 'real' MOFs.) The remaining databases contain data for proposed, or 'hypothetical,' MOFs. The seven crystallographic properties for all MOFs in the database were calculated using the zeo++ code[25,87–90] with a probe radius of 1.86 Å. This data is available at the HyMARC data hub.[71] Additional details can be found in our prior work.[1] These properties include: single crystal density (d), pore volume (pv), gravimetric surface area (gsa), volumetric surface area (vsa), void fraction (vf), largest cavity diameter (lcd), and pore limiting diameter (pld).

A prior study examined a subset of the present database, wherein the hydrogen uptake in 495,305 MOFs was estimated using the Chahine rule.[71,1,2] Subsequently, usable uptake in a portion of this subset comprising 43,777 MOFs predicted to be promising based on the Chahine rule was evaluated using Grand Canonical Monte Carlo (GCMC). This GCMC-evaluated dataset contained a mix of real and hypothetical MOFs: 15,235 real MOFs were sourced from the UM,[31] CoRE,[33] and Cambridge Structural Databases,[29,30] and 28,542 hypothetical MOFs were extracted from the Mail-order,[38] *in silico* deliverable,[46] *in silico* surface,[39] MOF-74 analogs,[40] ToBaCCo,[59] Zr-MOFs,[45] Northwestern,[36] University of Ottawa,[37,87,88] and in-house[1] hypothetical MOF databases (see Ref. 1 or Table S1 for details).[1,29,31,33,36–40] Hydrogen uptake isotherms for two operating conditions were predicted: for an isothermal pressure swing (PS) at T = 77 K between 5 and 100 bar, and for a combined temperature-pressure swing (TPS) between 77 K/100 bar (filled state) and 160 K/5 bar (empty state). Usable gravimetric (UG) and volumetric (UV) capacities were then calculated based on the isotherm data.

In addition to the 43,777 MOFs examined in Ref. 1, in the present study GCMC isotherms were evaluated for an additional 54,918 MOFs (see Ref. 1 and Section 3 of the SI for further details). These additional MOFs were selected at random from the 495,305 entry HyMARC database and therefore represent a more diverse sampling of the MOF property-space. To this dataset 423,429 additional compounds were added from 7 additional datasets: BJT MOFs,[83] R-WLLFHS,[41,89] MTV,[84] CSM-2018-I,[42] CSM-2018-II,[43] and CSM-2019-I,[44] and selected MOFs from the CSD 2017 dataset.[29,30] Subsequently, the capacities of the MOFs from these additional datasets were predicted by the ML models without retraining (i.e., no MOFs from these datasets were used for training or testing, and none of their isotherms were evaluated in advance with GCMC). In total, the dataset employed in the present study contains H_2 uptake data for 98,695 MOFs[71] and crystallographic property data for 918,734 MOFs.

We note that the present MOF dataset includes approximately 74,000 compounds having open metal sites (OMS), comprising roughly 8% of the total dataset. As the interatomic potential used in our GCMC calculations is not tuned to capture the unique aspects of the H_2-OMS interaction, it is possible that the calculated capacities for this class of MOFs will be less accurate. Fig. S1 and Table S3 compares experiments and the present GCMC calculations of H_2 capacities across a benchmark set of OMS MOFs discussed by García-Holley et al.[90] and in our prior work.[1] These data show that GCMC calculations using the pseudo-Feynman-Hibbs potential are in good agreement with experimental data for these OMS MOFs. The good agreement between theory and experiments is a consequence of the low temperature operating conditions used in our study, combined with the relatively low density of OMS in these MOFs.

Machine learning model

The No Free Lunch Theorem[91] implies that the optimal choice of ML algorithm is problem-specific. The differing performance of the algorithms summarized in Tables 1 and S2 is consistent with this notion. Identifying the best algorithm for a given dataset requires comparing multiple ML methods, each with optimized hyperparameters. Unfortunately, few comparisons of ML methods for gas adsorption exist; although dozens of ML algorithms are

available,[77,79,97–106,80,81,85,92–96] only RR,[77,79,80] MLR,[77] SVM,[77,81] and NN[77] have been examined for predicting H_2 storage.[43,63,78,82,105] The present study casts a wider net by comparatively assessing 14 ML algorithms (Table 3).[77,79,97–106,80,81,85,92–96]

The crystallographic properties of MOFs are known to correlate with H_2 capacities.[2,31,90,107–110] The ML models developed here exploit these correlations by adopting only crystallographic properties as input features. Moreover, the number of features was restricted to a small set comprising 7 properties: d, pv, gsa, vsa, vf, lcd, and pld. These are the same properties employed in our prior work.[1,2,111,112] Figure S2 shows the distribution of crystallographic properties for the training, test, and unseen datasets. Also, Table S4 summarizes five descriptive (minimum, maximum, mean, median, and % of 0's) and two distribution statistics (skew and kurtosis) of all crystallographic features for the training, test, and unseen datasets. (The details regarding these statistics and the definitions of skew and kurtosis can be found in Section S5 of the SI). The maxima's and minima's of the features in the training set establish the validity ranges of the ML models developed here.

The goal of the ML models is to predict 4 output properties: UG and UV for each of PS and TPS operating conditions. This was accomplished by developing separate ML models for each of the four targeted capacities. Figure S3 illustrates the overall work-flow.

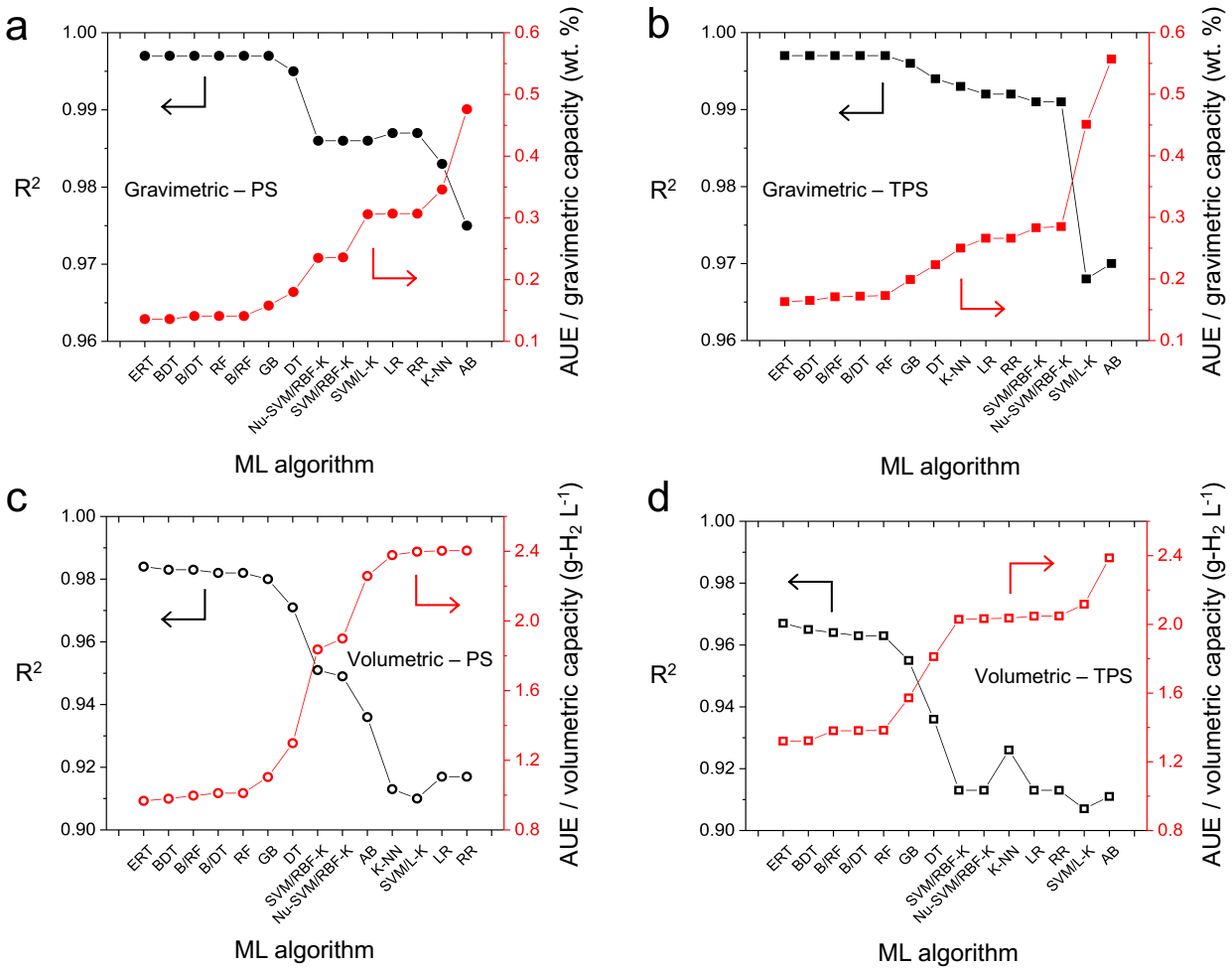


Figure 1. Comparison of ML algorithms for predicting hydrogen uptake in MOFs. Left (a,c) and right (b,d) panels report performance for PS and TPS conditions, respectively. Top (a,b) and bottom (c,d) panels report performance for usable gravimetric and volumetric capacities, respectively. The abbreviations for the ML methods are defined in Table 3.

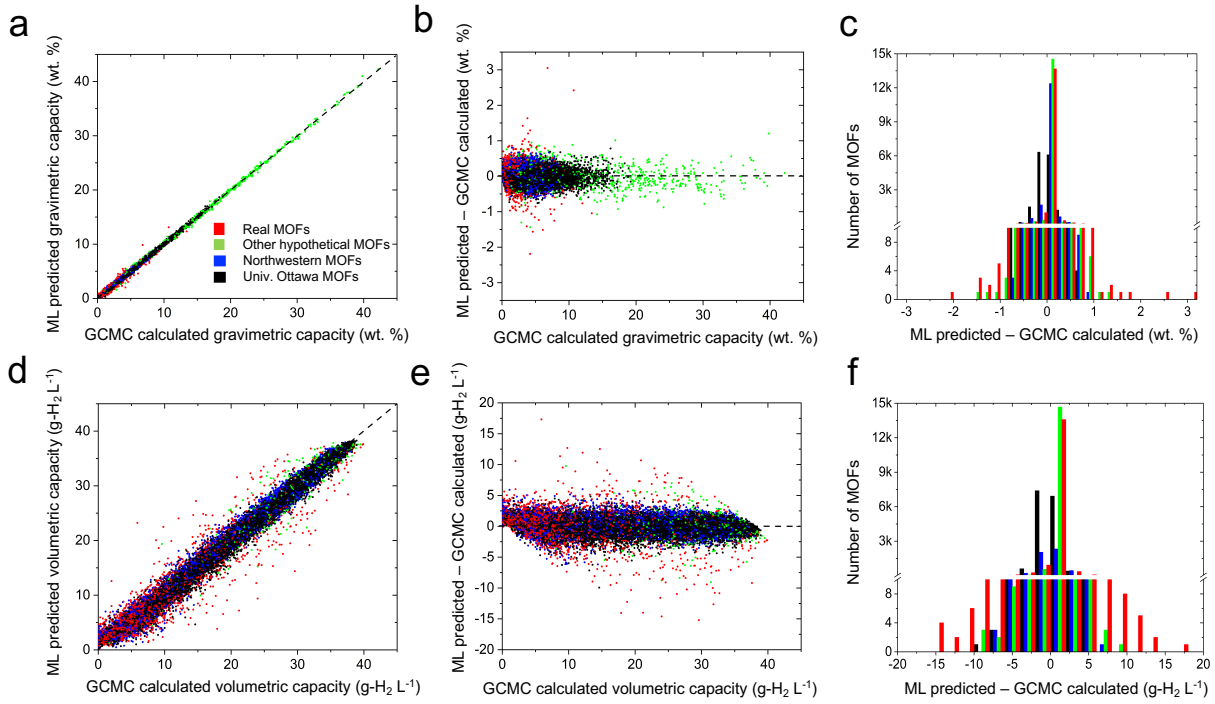


Figure 2. Performance of the Extremely Randomized Trees ML algorithm with respect to GCMC calculations for predicting usable H_2 capacities in MOFs. Data is collected at 77 K for a pressure swing between 100 and 5 bar on a test set of 24,674 MOFs. Different colors represent different categories of MOFs. Top (**a**-**c**) and bottom (**d**-**f**) panels illustrate performance for usable gravimetric and volumetric capacities, respectively. (**a**, **d**): Agreement between ML and GCMC predictions. (**b**, **e**): Difference between ML and GCMC as a function of GCMC capacity. (**c**, **f**) Distribution of differences in predictions between ML and GCMC.

The existing dataset of 98,695 MOFs (for which both crystallographic and capacity data are available)[71] was initially split into training and test sets of 74,201 and 24,674 MOFs, respectively, after shuffling the entire dataset.[106] ML algorithms[92,95,96,106] (Table 3) were implemented using the scikit-learn library.[106] Both scaled and unscaled features were used in training ML models. 10-fold cross-validation was used to optimize the hyperparameters of each model. The performance of the ML algorithms was assessed by comparing the predicted H_2 capacities to the capacity predicted by GCMC for the MOFs in the test set. The metrics used for the performance assessment of ML models were the coefficient of determination (R^2), average unsigned error (AUE), root-mean-squared error (RMSE), median absolute error (MAE), mean absolute percentage error (MAPE), and the Kendall rank correlation coefficient (τ). Additional details regarding these calculations can be found in Section S7 of the SI.

Dataset size

An obstacle to wider adoption of ML in materials science is the availability of sufficient quantities of high-quality training data.[113,114] Unfortunately, it is not yet clear how much data is needed to construct a useful ML model for a given system. Fernandez et al.[73] found that a reasonable balance between accuracy (R^2 ~0.85 to 0.93) and computational expense for predicting methane storage in MOFs was achieved for a training set containing data on 10,000 MOFs with 3 features. In contrast, Fanourgakis et al. [115] showed that a much smaller training set of ~1000 MOFs was sufficient to predict methane uptake when using six crystallographic features and four fictitious features. The different training set sizes required in these prior studies arise from the differing numbers and types of features used.

The present study explores this issue further by systematically examining the effect of training set size, and the training set to test set ratio, on ML accuracy. For each of the four targeted capacity outputs, 100 independent ML models were developed by varying the size of the training set between 100 and 74,000 MOFs (see Table S5 for a

list of the training set sizes). The four best-performing ERT ML algorithms identified earlier were used with 10-fold cross-validation. The resulting models were assessed using a common test set of 24,674 MOFs.

Feature importance/selection

The well-known Chahine rule proposes a linear correlation between gravimetric surface area and excess gravimetric H_2 capacity in adsorbents.[116,117] Nevertheless, the Chahine rule overpredicts H_2 capacities for MOFs with high surface areas,[117] and has not been extended to predict usable capacities.[1,2,6] Hence, a model for predicting H_2 uptake that is more general than the Chahine rule, yet requires limited input data, would be very helpful. In principle, ML could be used to generate such a predictive model if the features that are the most important for predicting H_2 uptake could be identified. Along these lines, Pardakhti et al. reported improved accuracy in predicting CH_4 adsorption when using a combination of (7) crystallographic and (19) chemical features.[72] Recently, Moosavi et al. explored feature importance in predicting the synthesis of MOFs.[118]

The present study determines the minimum number and optimal combination of crystallographic features necessary to achieve a specified accuracy in predicting H_2 uptake. The relative importance of the input features was assessed for all possible univariate and multivariate feature combinations using ERT ML models. The number of multivariate feature combinations, M, is given by: $M(n_{tot}, n_{sub}) = \frac{n_{tot}!}{n_{sub}!(n_{tot}-n_{sub})!}$, where $n_{tot} = 7$ is the total number of available features, and $1 <= n_{sub} <= 7$ is the number of features used as input to a given ML model. A total of 127 feature combinations are possible. ML models were developed for each of these feature combinations for each of the 4 output capacities, resulting in a total of 508 distinct ML models. All models were trained using a dataset of 74,021 MOFs and tested on a common set of 24,674 MOFs. 10-fold cross-validation was used for tuning and validating the models using only the training set. Univariate feature importance was further assessed using (*i.*) Pearson's correlation coefficient (r),[119–121] (*ii*), Breiman and Friedman's tree-based algorithm as implemented in Scikit-learn,[92,106] and (*iii*) the permutation importance method as implemented in rfpimp package.[122] Additional details regarding these methods can be found in Section 13 of the SI.

Results and Discussion

Evaluating ML algorithms

Tables S6-S9 illustrate the effect of several feature scaling methods on the performance of the ML algorithms examined here. Only the SVM family of models (SVM/L-K, SVM/RBF-K, and Nu-SVM/RBF-K)[77,92,98,100,101,106] were impacted by the choice of scaling method.

Figure 1 compares the accuracy of the ML algorithms for predicting hydrogen uptake in MOFs. R^2 and AUE were used as performance metrics. SVM variants were trained using min-max feature scaling; un-scaled features were

Table 4. Performance of the Extremely Randomized Trees ML algorithm in predicting UG and UV H_2 capacities of MOFs under PS and TPS conditions. R^2, AUE, RSME, and MAE represent the coefficient of determination, average unsigned error, root-mean-squared error, and median absolute error, respectively.

H_2 capacity type	R^2	AUE (capacity units)	RMSE (capacity units)	Kendall τ	MAE (capacity units)
UG at PS (wt. %)	0.997	0.14	0.18	0.961	0.10
UV at PS (g-H_2 L^{-1})	0.984	0.97	1.40	0.922	0.69
UG at TPS (wt. %)	0.997	0.16	0.23	0.966	0.10
UV at TPS (g-H_2 L^{-1})	0.967	1.32	1.92	0.819	0.91

used in training the remaining models. The performance of the algorithms as measured by 4 additional metrics –

root mean square error (RMSE), explained variance (EV), MAE (median absolute error), and Kendall τ – is reported in Tables S6-9.

Overall, these data indicate that the tree-based ensemble methods are superior to the other methods examined. In particular, the Extremely Randomized Trees (ERT)[77,85,106] algorithm exhibited the best performance overall. Boosted Decision Trees (BDT),[77,92–94,104,106] Random Forest (RF),[77,96,106] and Bagging algorithm variants[77,95,106,123,124] (with tree-based base estimators) are nearly as accurate. The R values for ERT predictions exceed 0.997 for gravimetric capacities, which are equivalent to errors of ~0.14 wt.%. Volumetrically, the accuracy of the ERT algorithm is slightly worse than its gravimetric performance: R^2 = 0.967-0.984, equivalent to errors of ~1.1 g-H_2 L^{-1} on average. In general, the worst-performing algorithms were linear regression, ridge regression, and support vector machine with linear kernel. For these algorithms R^2 varies between 0.913 and 0.992 depending on the conditions (i.e., gravimetric/volumetric and PS/TPS). As expected, the linear nature of these algorithms fails to fully capture the nonlinear dependence of output capacities on the multiple input features.

Figure 1 also shows that all the algorithms tested yield more accurate predictions of UG capacities compared to those for UV. Likewise, all algorithms more accurately predict usable capacities under PS conditions than under TPS conditions. This reflects the fact that the functional relationships between output capacities (UG/UV) and input features under PS and TPS conditions are likely different, as was observed in previously reported structure(feature)-property(capacity) relationships.[1,6,125] Table 4 summarizes the performance of the ERT algorithm in further detail. A comparison of Tables 1 and 4 indicates that the accuracy of the present ML models surpass previously-reported models for H_2 uptake. Furthermore, the present models also appear to be an improvement over earlier models that aim to predict the adsorption capacities of MOFs for any gas species, Table S2. This improved performance can be attributed to the exploration and optimization of multiple ML algorithms, use of an appropriate feature set, and the relatively large size of the present training set.

Figure 2 illustrates the degree of agreement between ERT ML predictions and GCMC calculations of usable H_2 capacities under PS conditions as a function of MOF source database. (Fig. S4 shows similar data for TPS conditions; see also Table 4.) As mentioned above, the present ML models more accurately predict UG capacities than

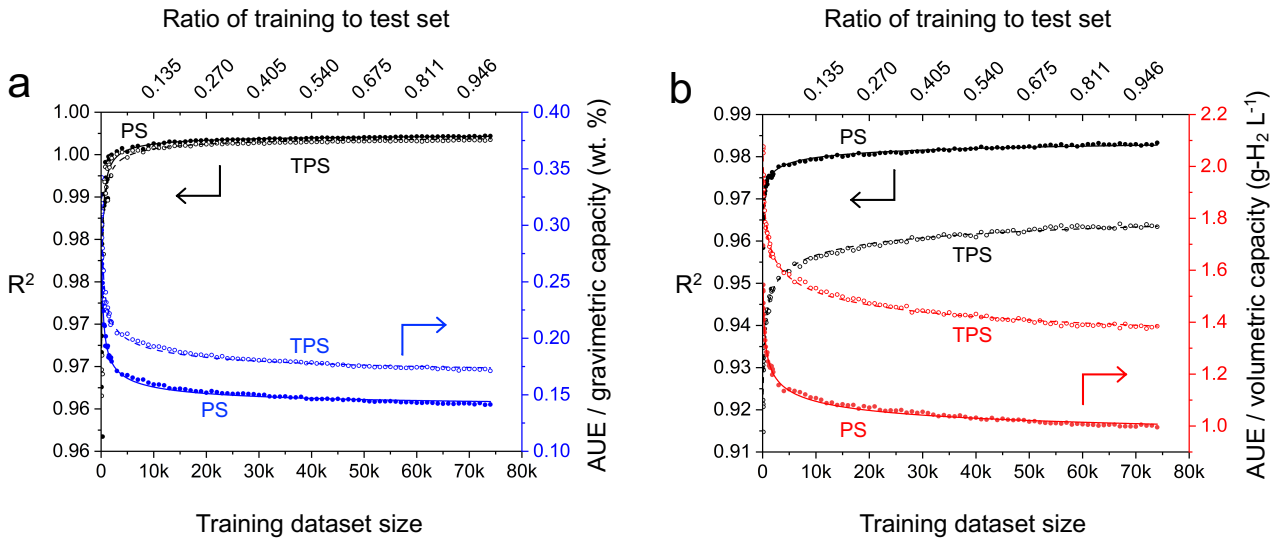


Figure 3. Performance of Extremely Randomized Trees ML models for predicting usable (a) gravimetric and (b) volumetric H_2 capacity as a function of training set size and the ratio of training to test set size. 100 different training sets ranging in size between 100 and 74,021 MOFs were examined. A common set of 24,674 MOFs was used for testing. Performance is quantified using R^2 (left axis, black) and the average unsigned error, AUE (right axis, blue and red for UG and UV, respectively). Lines represent a power-law fit to the data.

UV capacities. The largest differences between ML and GCMC capacities (Fig. 2c,f & Fig. S4c,f) primarily occur for the real MOF dataset. In principle, these differences may arise either from ML overfitting or from inaccurate GCMC predictions caused by non-ideal/incomplete MOF crystal structure data (i.e., missing atoms, disorder, etc.) as mentioned in prior studies.[1,35,126–129] ERT algorithms are fairly robust against overfitting.[85] To examine the possibility for overfitting, test set errors were compared with training set errors as shown in Fig. S5 and Table 4. These data suggest that the outliers are not a consequence of over fitting; hence, inaccuracies in the crystal structure data is proposed as the most likely source of this disagreement. [1,35,126–129]

Effect of training set size

Figure 3 illustrates the impact of training set size on the accuracy of the ERT ML models as quantified using R^2 and AUE. (Table S5 summarizes the dataset sizes used in these plots.) For training sets containing more than 5000 MOFs R^2 and AUE vary slowly and in a monotonic fashion, with AUE decreasing and R^2 increasing. The accuracy of the models is more sensitive to the size of the training set for smaller training sets containing roughly 5,000 or fewer MOFs. Figure S6 highlights the variation in performance for these smaller training sets.

The trends AUE as a function of training set size can be fit to a power-law expression of the form $AUE(m) = \alpha m^{\beta} + \gamma$ where m represents the size of the training set and β is the power law exponent. Fitting this model to the data shown in Fig. 3 reveals that the AUE for UG converges faster with training set size (β = -0.37 and -0.43) than it does for UV (β = -0.16 and -0.23). A full tabulation of the power-law parameters is given in Table S10. Based on these power-law expressions, one can determine the necessary size of the training set to achieve a desired level of accuracy. For example, assuming pressure swing operation, to achieve an AUE of approximately 0.25 wt.% and 1.5 g-H_2 L^{-1} requires training set sizes (for UG and UV) of less than 300 MOFs randomly selected from the diverse datasets used here.

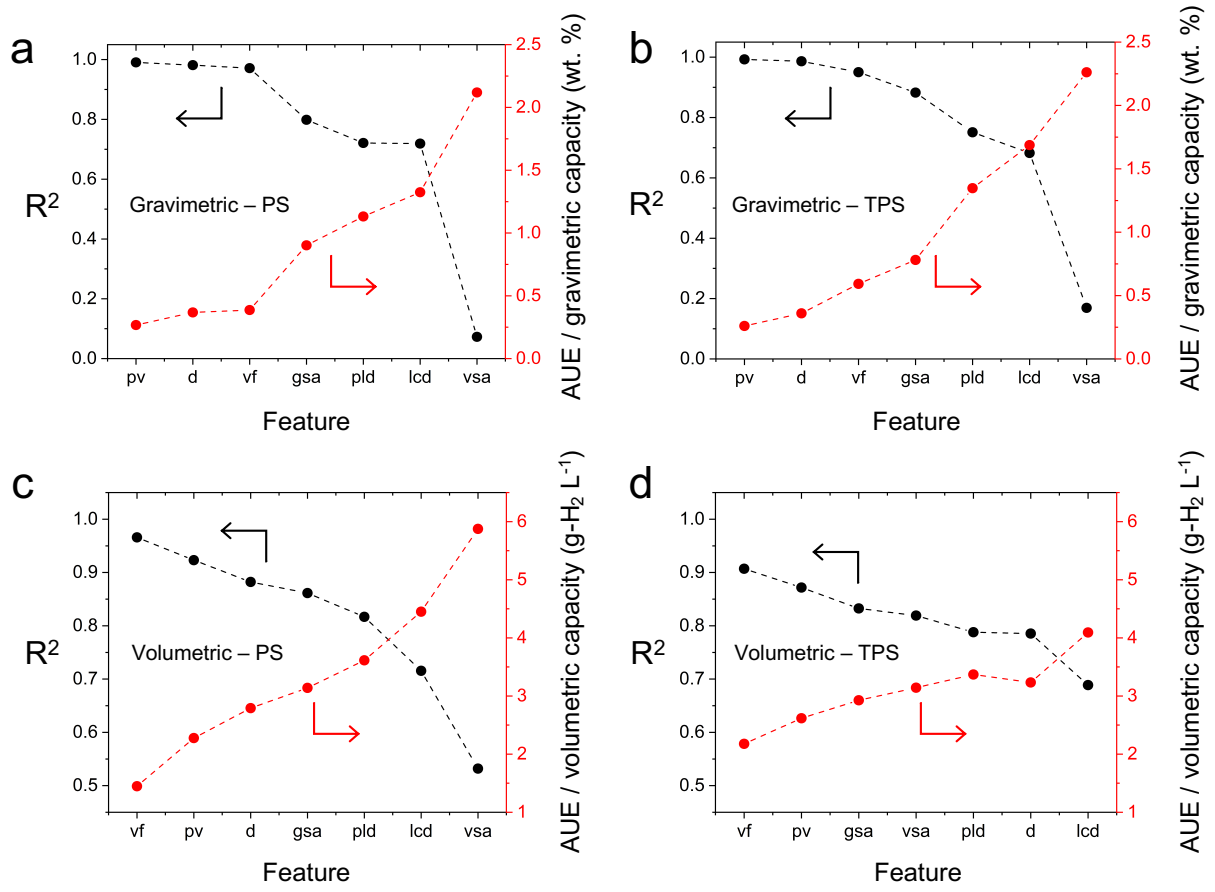


Figure 4. Importance of seven features in predicting usable H_2 storage capacities of MOFs. Feature importance was determined by developing distinct ERT models for each individual feature. The accuracy of the resulting models was assessed using R^2 (left axis; black data set) and AUE (right axis; red data set). Models were trained on a dataset of 74,201 MOFs and tested on a set of 24,674 MOFs. pv = pore volume; d = density; vf = void fraction; gsa = gravimetric surface area; pld = pore limiting diameter; lcd = limiting cavity diameter; vsa = volumetric surface area.

Univariate feature importance

Figure 4 illustrates the relative importance of the seven crystallographic features in predicting usable hydrogen uptake in MOFs. Feature importance was determined by developing ERT models for each single feature individually. Additional details for these models are provided in the Supporting Information. Based on these models, it is evident that pore volume (pv) and void fraction (vf) are the dominant features in predicting H_2 capacity; these two properties appear as the first- or second-most important single features regardless of operating condition or capacity type. The importance of these features can be rationalized by two factors. First, based on the empirical Chahine rule, the pore volume of a MOF correlates with its excess uptake.[116] Second, pore volume and void fraction are related (since pv = vf d^{-1}) – MOFs with larger pv have larger vf, and *vice versa*.[1]

Conversely, the largest cavity diameter (lcd) and volumetric surface area (vsa) are the single features whose ML models yield the lowest accuracy. The relative importance of the individual features for predicting UG capacities is: pv > d > vf > gsa > pld > lcd > vsa. This ordering is the same for PS and TPS conditions. In contrast, the importance ordering for UV capacities differs based on the operating condition. Nevertheless, vf and pv remain the two most important single features for both UV conditions, in that order (Fig. 4).

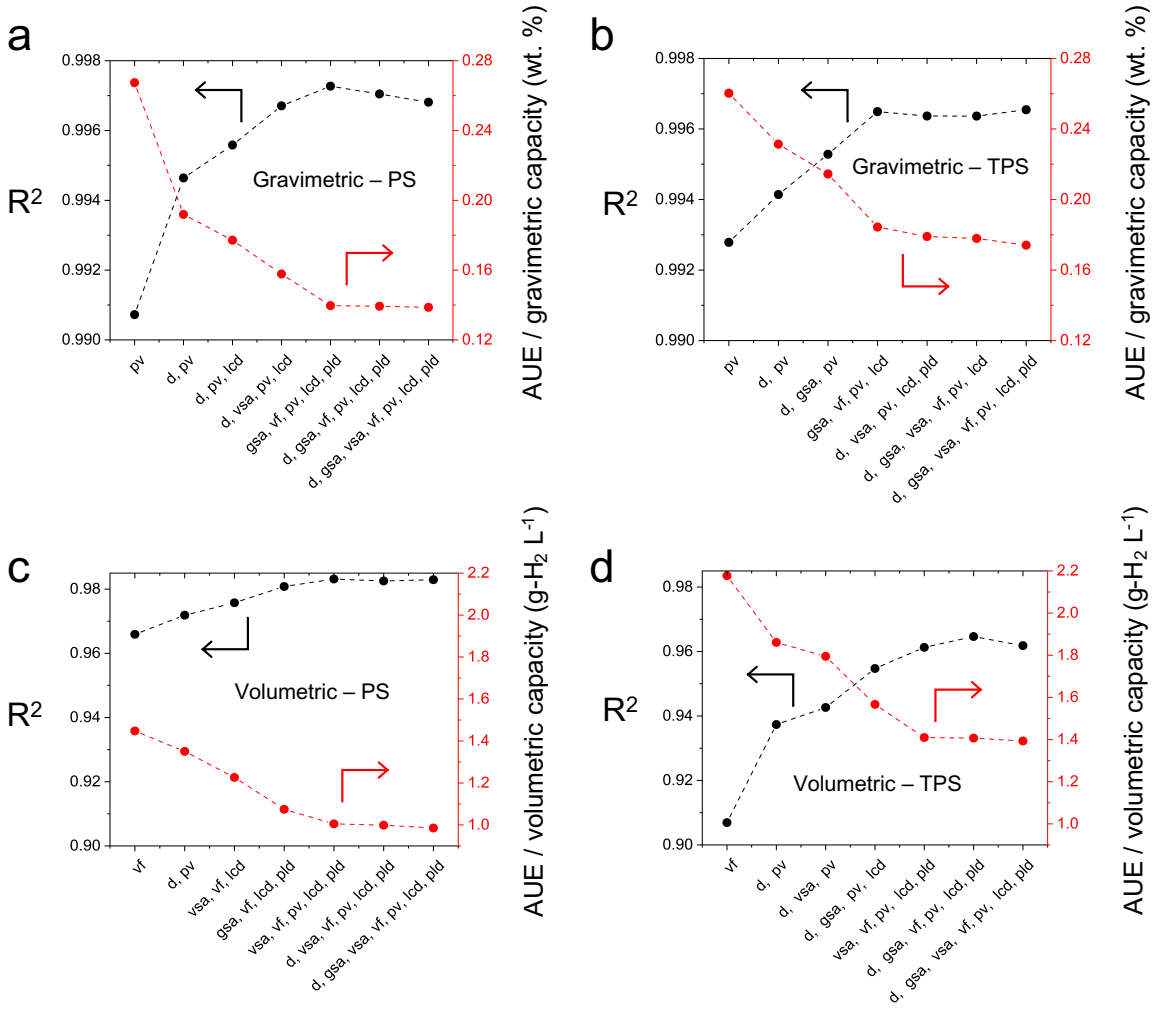


Figure 5. Accuracy of Extremely Randomized Trees (ERT) ML models as determined by R^2 and AUE as a function of the number and combination of input features. Each data point represents the most accurate feature combination for a given number of features. ERT models were trained on a dataset of 74,201 MOFs. R^2 and AUE were calculated using a test of 24,674 MOFs. Feature abbreviations are defined in Fig. 4.

Despite their limited input, the single-feature ML models illustrated in Fig. 4 achieve high accuracy. For example, any of the three independent models for UG-PS based only on pv, d, or vf can predict capacities with $R^2 > 0.95$ and with AUE of less than 0.5 wt.%. The accuracy and simplicity of the univariate ML models suggest that they can be used to quickly screen new MOFs for their utility in hydrogen storage. To that end, optimized single-feature ML models for the four categories of usable capacities considered here have been made available for use on the web with an interactive web form or with a python API.[86] Furthermore, the ML models can be downloaded via figshare.[130] These models take as input either pv (for UG predictions) or vf (for UV) of a given MOF. These input data can be quickly calculated from a MOF's crystal structure using modern structure analysis codes.[25,111,131–

Table 5. The best combinations of features for predicting UG and UV H_2 storage capacities at PS and TPS conditions.

Condition	Feature Combination	No. features	R^2	AUE	RMSE	Kendall τ
UG at PS	gsa, vf, pv, lcd, pld	5	0.997	0.14 wt. %	0.19 wt. %	0.959
UG at TPS	d, vsa, pv, lcd, pld	5	0.996	0.18 wt. %	0.25 wt. %	0.959
UV at PS	vsa, vf, pv, lcd, pld	5	0.983	1.01 g-H_2 L^{-1}	1.45 g-H_2 L^{-1}	0.920
UV at TPS	vsa, vf, pv, lcd, pld	5	0.961	1.41 g-H_2 L^{-1}	2.10 g-H_2 L^{-1}	0.814

[134] As shown in Fig. 4, these models can predict UG with an average error of less than 0.4 wt.%, and UV with errors less than 2.2 g-H_2 L^{-1}.

Figure S7 compares the single feature importance assessments based on ERT ML models (as reported in Fig. 4) with three popular methods for determining feature importance: Pearson's correlation coefficient (r),[119–121] Breiman and Friedman's tree-based algorithm as implemented in Scikit-learn,[92,106] and the permutation importance method as implemented in rfpimp package.[122] It is clear that the feature importance methods do not reproduce *in detail* the rank-ordering of feature importance that is suggested by our ERT ML models. Nevertheless, good agreement is evident more broadly. For example, in the case of UG (Fig. S8a,c), the three feature importance methods suggest that in aggregate pv is the most important feature, while vsa is the least, in agreement with the ERT models (Fig. 4a,b). Similarly, for UV, the importance methods suggest that vf and lcd are among the most and least important features, respectively. This is the same trend found in the univariate ERT models (Fig. 4c,d).

Multivariate feature importance

Figure 5 illustrates how the accuracy of the ML models varies with the number and combination of features. Assuming 7 features, 2^7-1 = 127 possible combinations exist. For a given number of features, Fig. 5 plots the combination of features resulting in the highest accuracy model. The last section of Appendix A summarizes the performance for all 508 possible feature combinations and capacity/operating condition types.) As expected, Fig. 5 shows that ML accuracy generally increases as the number of input features increases. As previously discussed, when limited to a single feature, vf yields the best accuracy for predicting UV, while pv is the best choice for UG. When the feature set is extended to 2 features, the combination of d and pv is the optimal choice among the $\binom{7}{2} = 21$ possible pairs regardless of the capacity (UG vs UV) or operating condition (PS vs TPS). For larger numbers of features, the optimal feature combination depends upon the operating condition and the capacity type. Based on the AUE, whose value tends to plateau as more features are added, highly accurate ML models can be generated using only 5 input features (Table 5). These data lend further support to the notion that the accuracy of a given ML model depends on both the number and identity of the input features. As a slightly more accurate alternative to the univariate web models described above, a subset of the present multivariate ML models that use 4, 5, and 7 input features are also available on the web using an interactive web form and via a python API.[86] The ML models can also be downloaded via figshare.[130]

H_2 uptake in unseen MOFs

Figure 6 illustrates the H_2 storage capacities of 820,039 MOFs as predicted by the 7-feature ERT ML models developed here. (This dataset is publicly accessible via HyMARC data hub.[71]) These MOFs are referred to as 'unseen,' in that they have not been included in the training or test sets used to develop the models. Figures 6a,b show UV capacities as functions of UG capacities under PS and TPS conditions, respectively. Both plots exhibit a rapid increase in UV at low values of UG, and reach a maximum in UV at UG values of approximately 9 wt.%. Beyond the maximum, UV decreases relatively slowly with increasing UG. These trends are consistent with our earlier findings derived from GCMC calculations on smaller datasets.[1,2,6]

In the case of PS operation, the maximum UV across the MOFs in the dataset is 37.4 g-H_2 L^{-1}; for TPS operation the maximum UV is 48.5 g-H_2 L^{-1}. In the case of UG, the maximum value predicted is 39 wt.% for PS operation and 42 wt.% for TPS. These values can be placed in context by comparing against the DOE hydrogen storage targets, which stipulate system-level hydrogen densities of 5.5 wt% and 40 g-H_2 L^{-1} by 2030 and 6.5 wt.%/50 g-H_2 L^{-1} longer-term ('Ultimate target').[6] Given that the tank and balance-of-plant for the storage system have non-zero mass and volume, the MOFs examined here cannot meet the Ultimate target for UV, regardless of operating condition.[12] More optimism exists, however, for meeting the gravimetric targets given the high UG exhibited by these

systems on a MOF-only basis. Of course, an additional challenge is to identify MOFs that excel both gravimetrically and volumetrically.[1,2,6,31,135]

It is also helpful to compare the performance predictions in Figs. 6a,b with that of state-of-the-art materials. In the case of PS operation, our prior study demonstrated that PCN-610 (NU-100) exhibits a hydrogen capacity of 10.1 wt.% & 35.5 g-H_2 L^{-1},[1] which to our knowledge is the best combination of gravimetric and volumetric capacities reported for any MOF under these conditions. The data in Figure 6a reveals that 16,345 MOFs can, in principle, exceed this capacity on both a UG and UV basis. In the case of TPS operation (Figure 6b), MOF-5 remains the benchmark, which a measured capacity of 7.8 wt.% & 51.9 g-H_2 L^{-1}.[2] Figure 8c shows that only 21 MOFs outperform MOF-5 under these conditions.

Regarding the accuracy of the present ML predictions, Table 4 shows that the AUE of these models are on the order of 0.15 wt.% and 1.3 g-H_2 L^{-1}. Although these errors are small, a more rigorous validation of the ML can be achieved with GCMC calculations. Thus, GCMC calculations were performed on a subset of MOFs that ML predicted to exhibit high UV and UG capacities. These MOFs fall within the rectangular regions shown in Figs. 6a,b, and exhibit capacities that meet or exceed 36 g-H_2 L^{-1} & 7.5 wt.% for PS conditions and 48 g-H_2 L^{-1} & 7.5 wt.% under TPS conditions. In total, 21,700 compounds were re-examined with GCMC based on their ML-predicted PS capacities, and another 7,901 were re-examined for TPS.

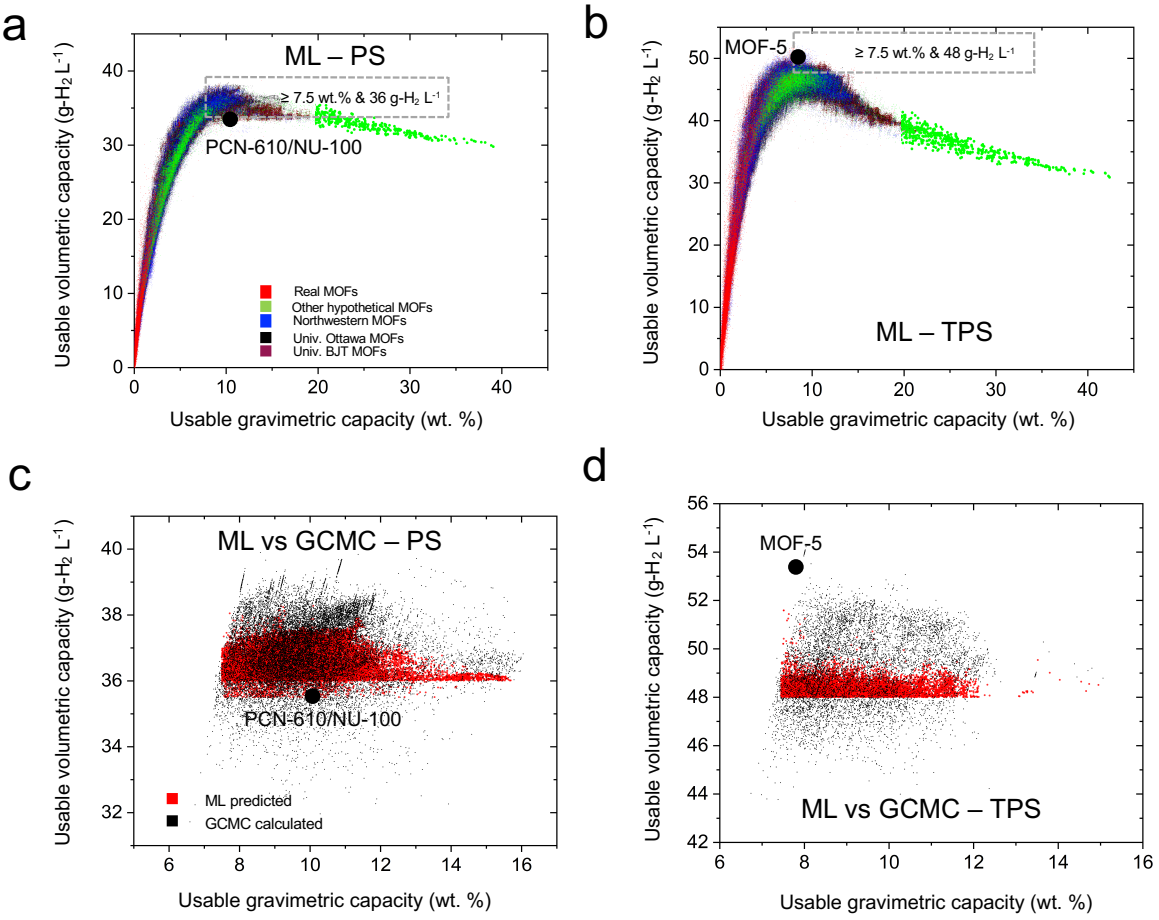


Figure 6. (a,b) Machine learning predictions of usable hydrogen capacities of 820,093 MOFs. Colors indicate the originating database for a given MOF. (c,d) Validation of ML-predicted capacities for the highest capacity MOFs identified by ML (shown in the rectangular regions in c,d) using GCMC simulations. For comparison, the capacities of PCN-610/NU-100 (PS: 10.1 wt.%, 35.5 g-H_2 L^{-1}) and MOF-5 (TPS: 7.8 wt.%, 51.9 g -H_2 L^{-1}) are shown.

Figure 6c compares ML and GCMC predictions for usable capacities for 21,700 high-capacity MOFs under PS conditions. The strong overlap in the two datasets further highlights the accuracy of the ML models. A total of 8,187 MOFs were predicted by GCMC to outperform PCN-610/NU-100 under these conditions. A summary of the 10 highest-capacity MOFs, sorted based on their GCMC capacities, is provided in Table 6. (A more extensive listing is provided in Table S11.) The highest capacity MOFs are all hypothetical compounds: 5 originate from the ToBaCCo database,[59] two are from the Univ. of Ottawa database,[37] and the remainder are from the Northwestern[36] database. These MOFs all exhibit high surface areas (average =5746, range = 4346 – 7835 $m^2 g^{-1}$) and large void fractions of 0.89, on average. The range of these property values are consistent with those reported in an earlier study,[1,136,137] and suggest that maximizing the surface area is an important design guideline for PS operation. The highest capacity MOF, mof_7642,[59] is predicted to exhibit capacities of 11.1 wt. % and 40.5 g-H_2 L^{-1}, surpassing that of PCN-610/NU-100, the record-holder under PS conditions. The crystal structure of mof_7642 is shown in Fig. 7a.

A search in the CCDC[138] was performed to identify MOFs that have been synthesized that are similar to the high-capacity compounds identified here. The existence of similar MOFs may suggest synthetic procedures that could be adapted to the present systems. The top 5 MOFs under PS conditions contain relatively long tritopic linkers. In the case of mof_7642, this search identified the interpenetrated MOF RANCEQ[139] as having a similarly index of 0.82. Interpenetration is fairly common in MOFs (such as mof_7642) with longer linkers, and is generally undesirable for achieving high uptake. Nevertheless, several examples of successful synthesis of MOFs with long, multi-topic linkers that do not undergo interpenetration have been reported. These include MOF-180 and MOF-200,[140] the PCN-6X series,[141] and NOTT-112.[142] The next 4 PS candidates in Table 6 exhibit pillared Zn paddlewheel clusters with long ditopic linkers. Karagiaridi et al.[143] demonstrated the feasibility of synthesizing pillared paddlewheel MOFs with long linkers; the SALEM-X series are examples.[143] Finally, str_m3_o5_o20_f0_nbo.sym.1.out is based on a Zn paddlewheel cluster and a ditopic linker. HOFSUS is an example of such a MOF.[144]

Figure 6d provides a similar comparison between ML predictions and GCMC calculations for MOFs expected to exhibit high capacities under TPS conditions. Under these conditions only 95 MOFs were predicted by GCMC

Table 6. Highest capacity MOFs, as identified by ML and verified with GCMC, under pressure swing and temperature + pressure swing conditions. Here NW and UO refer to the Northwestern[36] and University of Ottawa databases.[39]

Name	Source	Density ($g\ cm^{-3}$)	Grav. surface area ($m^2 g^{-1}$)	Vol. surface area ($m^2 cm^{-3}$)	Void fraction	Pore volume ($cm^3 g^{-1}$)	Largest cavity diameter (Å)	Pore limiting diameter (Å)	Usable grav. capacity (wt. %)		Usable vol. capacity ($g\text{-}H_2\ L^{-1}$)	
									GCMC	ML	GCMC	ML
Pressure swing												
mof_7642	ToBaCCo	0.30	5561	1695	0.89	2.93	12.8	11.8	11.1	10.3	40.5	37.4
mof_7690	ToBaCCo	0.30	5715	1706	0.89	2.98	12.8	12.0	11.3	10.4	40.3	37.3
mof_7594	ToBaCCo	0.40	5070	2031	0.86	2.15	11.2	9.7	8.6	7.9	39.9	37.0
mof_7210	ToBaCCo	0.29	5936	1730	0.89	3.04	13.4	11.7	11.4	10.5	39.8	37.1
mof_7738	ToBaCCo	0.25	6054	1502	0.90	3.64	14.5	13.5	13.0	12.0	39.7	37.0
hypotheticalMOF_5045702_i_1_j_24_k_20_m_2	NW	0.31	5926	1820	0.88	2.87	16.0	11.0	10.9	10.1	39.7	37.2
str_m3_o19_o19_f0_nbo.sym.1.out	UO	0.31	5073	1583	0.90	2.88	17.7	12.9	10.8	10.1	39.7	37.1
hypotheticalMOF_5037315_i_1_j_20_k_12_m_1	NW	0.31	5818	1787	0.88	2.86	16.0	11.0	10.9	10.0	39.7	37.0
hypotheticalMOF_5037467_i_1_j_20_k_12_m_8	NW	0.31	5860	1800	0.88	2.85	16.0	11.0	10.9	10.0	39.7	37.0
str_m3_o5_o20_f0_nbo.sym.1.out	UO	0.39	4772	1882	0.87	2.22	14.1	9.6	8.7	8.1	39.7	37.2
Temperature + pressure swing												
str_m1_o1_o11_f0_pcu.sym.102.out	UO	0.45	4352	1974	0.84	1.84	12.9	10.1	10.4	9.7	53.1	48.1
str_m1_o1_o11_f0_pcu.sym.117.out	UO	0.47	4162	1977	0.83	1.74	12.8	9.9	9.9	9.0	52.8	48.0
str_m1_o1_o11_f0_pcu.sym.121.out	UO	0.47	4263	2006	0.83	1.76	12.1	10.2	10.0	9.4	52.7	48.1
str_m1_o1_o11_f0_pcu.sym.13.out	UO	0.46	4326	2005	0.83	1.79	12.7	9.9	10.1	9.3	52.6	48.0
str_m1_o1_o11_f0_pcu.sym.159.out	UO	0.58	3703	2138	0.80	1.38	10.4	8.6	8.3	7.6	52.6	48.5
str_m1_o1_o11_f0_pcu.sym.200.out	UO	0.45	4359	1978	0.84	1.84	12.9	10.1	10.3	9.6	52.6	48.1
str_m1_o1_o11_f0_pcu.sym.212.out	UO	0.60	3417	2035	0.83	1.39	12.0	10.1	8.1	7.5	52.5	48.1
str_m1_o1_o11_f0_pcu.sym.51.out	UO	0.46	4330	2007	0.83	1.79	11.9	9.9	10.1	9.3	52.5	48.1
str_m1_o1_o11_f0_pcu.sym.71.out	UO	0.45	4436	1980	0.84	1.87	13.0	10.9	10.4	9.7	52.5	48.1
str_m1_o1_o11_f0_pcu.sym.89.out	UO	0.58	3507	2043	0.83	1.42	12.4	9.8	8.2	7.7	52.5	48.1

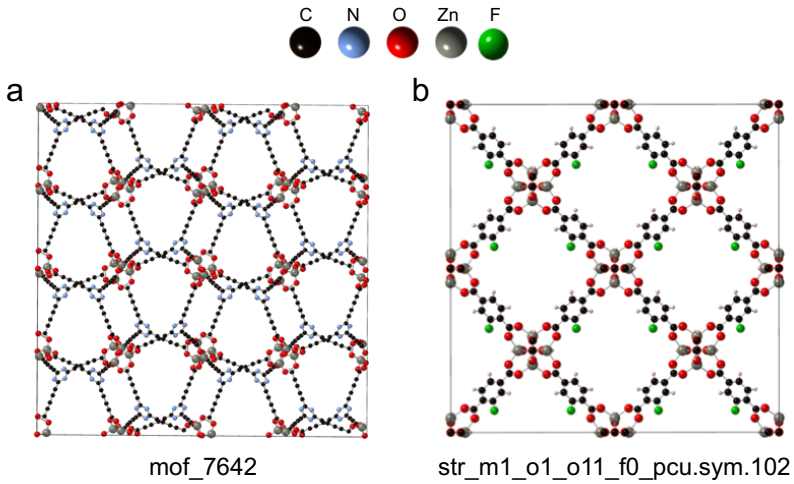


Figure 7. Crystallographic images of the highest-capacity MOFs under (a) pressure swing and (b) temperature swing conditions. These MOFs originate from the ToBaCCo[59] and University of Ottawa[39] databases, respectively.

to outperform MOF-5. A summary of the 10 highest-capacity MOFs, sorted by their GCMC capacities, is provided in Table 6. (See Table S12 for a more extensive tabulation.) As found for PS operation, all of the top performing candidates are hypothetical compounds. One difference with the PS case is that all of these MOFs originate from the Univ. of Ottawa database.[37] Furthermore, none of the highest capacity MOFs identified for PS operation appear as top candidates for TPS. Comparing the highest-capacity MOFs for both operating conditions, it can be seen that the high-capacity TPS MOFs systematically exhibit lower surface areas (avg. = 4073 m^2 g^{-1}), smaller void fractions (avg. = 0.83), and higher densities. Hence, the categories of MOFs that maximize uptake under PS and TPS conditions exhibit distinct properties. These differences suggest that maximizing the surface area – which, as discussed above, is desirable for maximizing PS capacity – is not advantageous for TPS operation. This behavior can be explained by trends in *total* capacities,[6] which the TPS capacities reported here approximate. More specifically, it is known that total volumetric capacities are maximized for intermediate values of the surface area; for larger surface areas the volumetric capacity decreases.

Returning to the list of promising MOFs for TPS operation, Table 6 reports that the highest-capacity MOF, str_m1_o1_o11_f0_pcu.sym.102.out, has a GCMC-predicted capacity of 10.4 wt.% and 53.1 g-H_2 L^{-1}. This capacity surpasses that of MOF-5, which to our knowledge holds the capacity record under these conditions. The crystal structure of this MOF is shown in Fig. 7b.

The top 10 MOFs under TPS conditions contain the same Zn metal cluster and terephthalic acid linkers, where the linkers have been modified with varying functional groups. The slight differences in the capacities of these MOFs can be traced to differences in the functional groups. A similarity search based on str_m1_o1_o11_f0_pcu.sym.117.out identified 40 similar MOFs. Approximately 30 of these (for example, HIFTOG, MIBQAR, UNIGEE, VUSJUP, and ZELROZ) contain Zn metal clusters and linkers based on variants of terephthalic acid.

Figures S8-S9 and Table S13 quantify the differences between ML and GCMC predictions on the subset of high-capacity MOFs shown in Figs. 6c,d. For PS operation, the AUE of ML relative to GCMC is 0.24 wt.% and 0.66 g-H_2 L^{-1}, while for TPS the AUE is 0.24 wt.% and 1.28 g-H_2 L^{-1}. Both sets of errors are comparable to the errors reported in Table 4 for the original test set of MOFs. Figures S8(c, f) and S9(c, f) plot the frequency distribution of the differences between GCMC and ML. These distribution plots suggest that the largest differences occur for predictions involving real MOFs and for hypothetical MOFs extracted from databases other than those from

Northwestern,[36] University of Ottawa,[37] and BJT.[83] (These MOFs are referred to as "other hypothetical MOFs" in Figure 6). These MOFs, along with the real compounds, exhibit higher structural diversity than those contained in the other databases. For example, the diversity of the topologies used in the ToBaCCo[59] and Zr-MOFs[45] databases and in the linkers used in MTV-MOF[84] database are larger than what is found in the databases from Northwestern,[36] University of Ottawa,[37] and BJT.[83]

Limitations of this Study

As previously described, some of the high-capacity MOFs identified here may prove difficult to synthesize. Although this limitation applies primarily to the hypothetical MOFs, in some cases real MOFs are also known to undergo framework collapse during activation, which would reduce capacity.[1,2] Nevertheless, future improvements to synthesis techniques may overcome these limitations – what is difficult to make today may be possible in the future. Secondly, our models do not distinguish between realistic MOFs having non-defective crystal structures and those for which the structures are defective/unrealistic. Unrealistic structures can result from incomplete or imperfect virtual solvent removal and the presence of partial occupancies or symmetry disorder in the crystal structure.[31] Consequently, a defective/unrealistic MOF could be erroneously predicted to be a promising candidate. Follow-up calculations using GCMC and visual inspection of the crystal structure are recommended for all promising candidates identified by ML. Finally, the ML models developed here are non-interpretable, 'black-box' models. Although these models are demonstrated to be highly accurate, additional effort is required to assess the relative importance of their input data. (The approach demonstrated here for evaluating feature importance involved the development of multiple models with varying numbers and combinations of features.) Alternatively, interpretable white-box ML models could be developed to provide more insight into feature importance. However, our experience suggests that white-box models generate less accurate predictions.

Public accessibility of ML models and datasets

General public can use all our ML models free of cost for the prediction of H_2 storage capacities of an arbitrary MOF via the HyMARC website.[1] An user can provide input features required by ML models either via an web form (Figure 8a)[1] or a Python API (Figure 8b). The detailed user instructions are also provided in the HyMARC website.[1] In addition, 16 downloadable ML models with user instructions were made publicly available via figshare,[2] an online open access repository. All ML predicted usable gravimetric and volumetric H_2 storage capacities of 820K MOFs under PS (Fig. 8c) and TPS (Fig. 8d) conditions were made public via HyMARC datahub,[3] including their 7 crystallographic features.

Summary – Materials Discovery

Figure 8. Open-access ML models and datasets for H_2 storage in MOFs publicly accessible via the HyMARC website. (a) Interactive web form. (b) Python API. Datasets containing usable gravimetric and volumetric capacities of 820K MOFs under (c) PS and (d) TPS conditions, including 7 crystallographic properties.

The H_2 storage capacities of nearly a million MOFs have been predicted via machine learning. The predictions span a diverse collection of MOFs sourced from 19 databases and reveal performance under two operating conditions: pressure swing and temperature + pressure swing. More than a dozen ML algorithms were benchmarked, with the extremely randomized trees method found to be the most accurate. The resulting ML models are accessible on the web at the HyMARC Data Hub.[86] These models allow for accurate, rapid screening of the hydrogen storage properties of new MOFs using minimal structural data as input; only a single feature is needed for the simplest models.

The accuracy of the ML models was characterized as a function of training set size and the number/combination of input features. Regarding the dependence on the training set, the accuracy of the models can be well-described using a simple power-law function of the training set size. The dependence on the number and combination of input features was determined by evaluating 508 independent ML models generated from all possible combinations of the seven features. The single most important features for predicting H_2 uptake are pore volume (for gravimetric capacity) and void fraction (for volumetric capacity).

Using these models, 8,282 MOFs are identified that have the potential to exceed the capacities of state-of-the-art materials under usable conditions. The identified MOFs are predominantly hypothetical compounds, which (for pressure-swing operation) exhibit low densities (<0.31 g cm^{-3}) in combination with high surface areas (> 5,300 m^2 g^{-1}), void fractions (~0.90), and pore volumes (>3.3 cm^3 g^{-1}). These MOFs are suggested as targets for experimental synthesis.

PART 2: CRYSTAL ENGINEERING

Introduction

Storage of sufficient quantities of fuel on automobiles presents one of the greatest challenges to realizing a hydrogen economy. A number of technologies have been pursued for improving energy density of hydrogen; these are divided into physical-based (cold or cryo- compressed hydrogen storage) and materials-based (chemical hydrogen storage materials and metal hydrides, sorbents) approaches.[7,145–148] Storage in metal-organic framework (MOF) sorbents via adsorption presents one of the most promising approaches due to fast charge/discharge kinetics, facile reversibility, and high gravimetric capacities.[149] However, high volumetric densities are uncommon in MOFs, and these densities can be impacted by multiple factors such as MOF structure, the nature of the interaction of H_2 with the MOF, and the packing of the MOF material.[150] The issue of materials packing, although acknowledged as an important factor, has not been widely examined.

Calculations of volumetric performance often assume (unrealistic) single crystal packing densities, and it must be recognized that this represents an upper limit to performance. Analysis by the Hydrogen Storage Engineering Center of Excellence (HSECoE) demonstrated that inefficient material packing can result in a >60 % volumetric density reduction compared to the single crystal.[149] In other words, volumetric performance of the MOF in a real system will be profoundly impacted by the discrepancy between MOF material packing density and crystallographic density. Consequently, improvements to the intrinsic capacity of the adsorbent, which have been the focus of materials research for more than a decade, can be 'undone' by poor packing of the media in the storage system. Therefore, MOF packing efficiency is an important parameter and plays a critical role in improving volumetric hydrogen storage density in real systems.

The packing of uniform spheres has an upper limit of 74% packing efficiency.[151] Achieving this threshold requires an ordered arrangement in space and it has recently been shown that random packing of spheres cannot exceed 63% packing efficiency.[151] For other particle shapes the results differ: randomly jammed tetrahedral dice[152] exhibit packing efficiencies of 76% compared to 100% packing efficiency for the regularly arranged platonic solids. However, there exists no theoretical framework for predicting the minimum void space that is practically achievable for polydisperse shapes characteristic of MOFs. Thus, packing density studies on real MOF samples whose crystals exhibit different shapes and sizes is currently best pathway to understand packing efficiency. Unfortunately, strategies to engineer crystal shape/size properties and pack MOFs with low void fraction have not been widely explored. Here, MOFs with targeted crystal shapes (cubic, cuboctahedral, octahedral, and spherical) and sizes are synthesized. Their packing density and hydrogen uptake are characterized and compared with BASF-produced (hereafter "commercial") material. The optimization of crystal size and engineering of crystal morphology for MOF-5 is demonstrated to dramatically improve volumetric hydrogen storage performance, both in terms of packing density (up to 100% improvement) and compacted density (up to 33%) with respect to what can be achieved with commercial MOF-5 powders.

Experimental section

Materials

Terephthalic acid (H_2BDC, 98.0%), 1,3,5-tris(4-carboxyphenyl)benzene (H_3BTB, ≥98%), oxalic acid (≥99.0%), malonic acid (99%), succinic acid (≥99.0%), glutaric acid (99%), adipic acid (99.5%), and suberic acid (98%) were purchased from Sigma Aldrich. 1,3,5-Benzenetricarboxylic acid (≥98%) N,N-dimethylformamide (DMF, ACS grade), methylene chloride (DCM, HPLC grade, 99.9%), and zinc nitrate hexahydrate ($Zn(NO_3)2{\cdot}6H_2O$, ACS grade) were purchased from Fisher Scientific. N,N-diethylformamide (DEF, 99.0%) was purchased from Acros Organics. [1,1'-biphenyl]-3,4',5-tricarboxylic acid,[153] 1 2,4,6-tris(4-carboxyphenyl)aniline (NH_2-H_3BTB, H_3L),[153] 5'-((3,5-dicarboxyphenyl)ethynyl)-[1,1':3',1''-terphenyl]-4,4''-dicarboxylic acid (H_4L),[154] and 5'-(4-carboxyphenyl)-[1,1':3',1''-terphenyl]-3,4'',5-tricarboxylic acid (H_4L)[155] were synthesized according to the literature procedure.

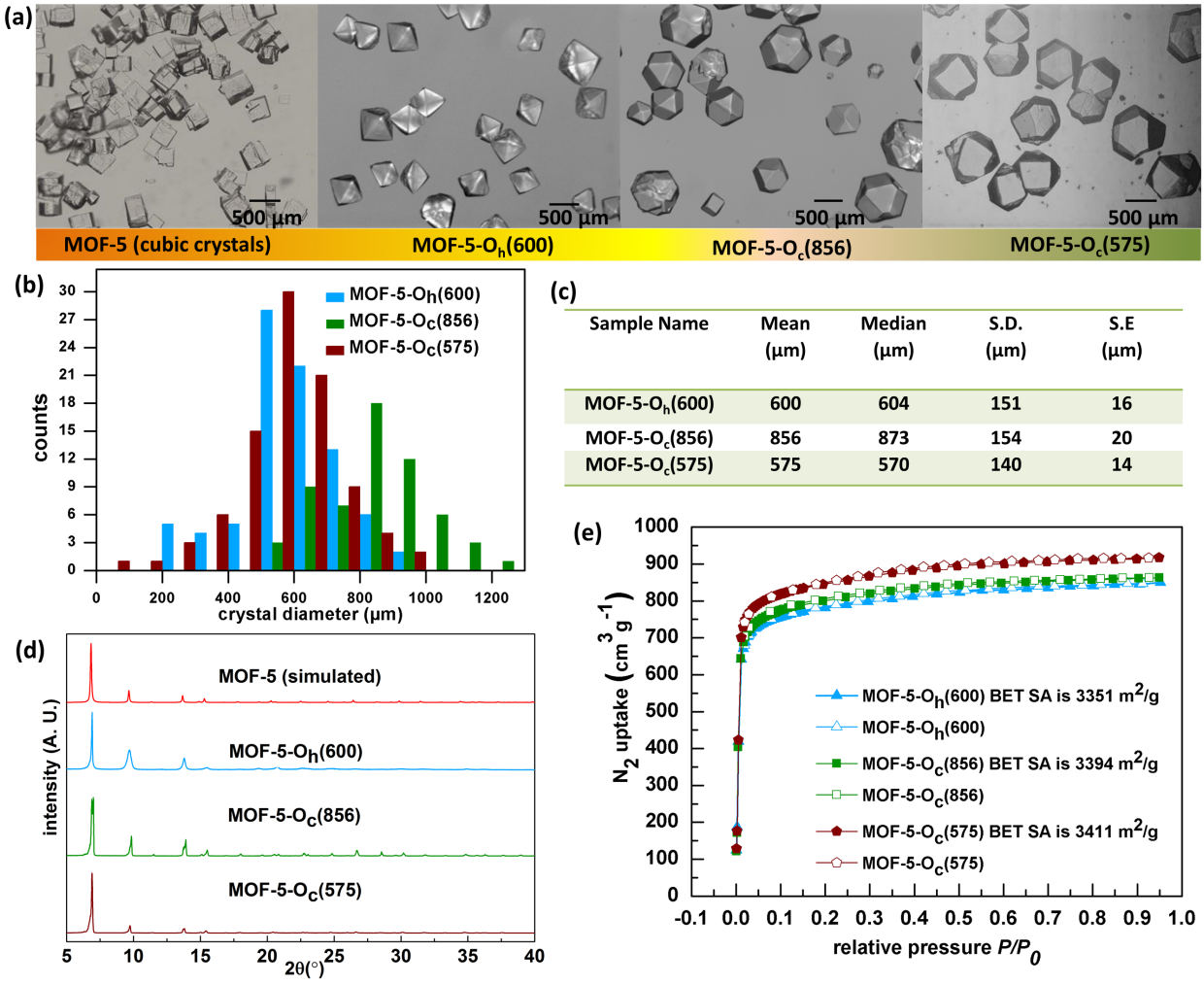
Results and discussion

To assess the influence of crystal size and shape on packing efficiency, the benchmark compound MOF-5 was selected as a model.[156] MOF-5 shows one of the highest deliverable hydrogen volumetric capacities among all MOFs based on its single crystal density[1] and, based on techno-economic analysis, can meet adsorbent cost targets.[157] Nevertheless, the poor packing density of MOF-5 limits its volumetric hydrogen capacity in practice.[1,158] In the present study, cubic MOF-5 crystals with four different sizes, varying from hundreds of micrometers to millimeter scale, are obtained by modulating the metal:ligand (M:L) molar ratio as well as the reactant concentration. In addition, new methods for synthesizing MOF-5 crystals with different shapes by the action of additives are described. Controlling these aspects of MOF crystals is demonstrated to dramatically improve volumetric hydrogen storage performance. It is anticipated that these lessons are directly applicable to the large family of cubic MOFs existing in the literature.

Control of crystal size

The effect of synthetic process parameters such as concentration of reagents, temperature, and time were investigated for their effect on crystal size and size distribution. Reactions for shorter duration (12 to 18 h) at high temperature (110 to 150 °C) afforded cubic crystals with a broad size distribution (on the order of 200-1300 μm). Longer durations (24 to 72 h) at lower temperatures (60 to 90 °C) yielded crystals with narrow size distributions, but led to a greater than 15% reduction in BET surface area. The optimal conditions for narrowing the size distribution without compromising surface area were 18-24 h at 100 °C. Variation of the metal:ligand (M:L, $Zn(NO_3)_2.6H_2O$:H_2BDC) molar ratio was studied over the range of 1:1 to 5:1. For samples with M:L =1 or M:L ≥ 4, the final product had an additional solid phase other than MOF-5. Accordingly, efforts were focused on 1 < M:L < 4. Additionally, the reactant concentration was varied to tune crystal size distributions.

Eventually, the optimized M:L molar ratio mixtures (3.8:1, 2.3:1, 1.7:1, and 1.7: 1(twofold dilution) were heated at 100 °C for 24 h. These four recipes resulted in four different crystal size distributions: MOF-5(2349), MOF-5(1500), MOF-5(808), and MOF-5(279), respectively where the number in parentheses indicates the mean crystal size in microns (Figure 1a, Supporting Information). The crystal sizes of all samples and their distributions and



Sample Name	Mean (μm)	Median (μm)	S.D. (μm)	S.E (μm)
MOF-5-O_h(600)	600	604	151	16
MOF-5-O_c(856)	856	873	154	20
MOF-5-O_c(575)	575	570	140	14

Figure 9. (a) Optical images of MOF-5 crystals with controlled morphologies. (b) and (c). Histogram plot and tabulated crystal size distributions range and median values for different morphologies of MOF-5. (d) PXRD patterns of new MOF-5 morphologies compared with the simulated pattern of MOF-5 computed from the crystal structure (refcode SAHYOQ). (e) N2 sorption isotherms for different morphologies of MOF-5 (adsorption data are shown with filled symbols while desorption data are shown with empty symbols).

statistics are represented in Figure 9b and 9c. The crystal sizes of these samples were compared to commercial MOF-5 (Figure 9a) which has submicron average crystal size. Phase purity of all samples was confirmed through powder X-ray diffraction, and it was observed that powder patterns of all samples agree with the calculated pattern of MOF-5 simulated from the crystal structure (Figure 9d). All samples exhibit very good surface areas (3505-3464 m^2/g) matching the expected theoretical surface area (3563 m^2/g) of MOF-5 (Figure 9e).

Morphology engineering

Inspired by observations of different morphologies of MOF-5 by the action of a tricarboxylic acid linker,[158] we conjectured that carboxylic acids can serve as additives for synthesizing new morphologies. An array of di-, tri-, and tetra-topic carboxylic acids were examined as morphology modifiers (Figure S1, Supporting Information); the role of additive concentration in influencing morphology by varying the additive molar ratio while keeping the ratio of BDC and $Zn(NO_3)_2.6H_2O$ (1:2.8 molar ratio) unchanged was also examined (Supporting Information).

Ditopic carboxylic acid (H_2L) linkers

Several alkane dicarboxylic acids including oxalic acid, malonic acid, succinic acid, glutaric acid, adipic acid, and suberic acid were tested for their ability to modify morphology when employed as additives at the 5-10 mol% level. It was observed that introducing these carboxylic acids to the MOF-5 reaction mixture afforded no change in MOF-5 cubic crystal morphology. It is known that some linear (aromatic) linkers can give rise to new phases incorporating zinc and two linkers,[159] but in all cases examined here the predominant phase was MOF-5.

Tritopic carboxylic acid (H_3L) linkers

Four tritopic carboxylic acids (trimesic acid, [1,1'-biphenyl]-3,4',5-tricarboxylic acid, 1,3,5-tris(4-carboxyphenyl)benzene (H_3BTB) and 2,4,6-tris(4-carboxyphenyl)aniline (NH_2-H_3BTB)) were screened. Among these, the addition of H_3BTB and NH_2-H_3BTB to the initial MOF-5 reagent mixture were observed to generate different shaped crystals. When ~4 mol% H_3BTB was added to the MOF-5 reaction mixture and heated at 100 °C for 24 h, uniform octahedral (O_h) morphology crystals (MOF-5-O_h(600)) are synthesized, while the addition of lower amounts of H_3BTB (~2 mol%) to the MOF-5 reaction mixture at 100 °C for 48 h yielded uniform cuboctahedral (O_c) shaped crystals (MOF-5-O_c(856)). The addition of H_3BTB at concentrations greater than ~4 mol% resulted in both needle- and octahedral-shaped crystals which is consistent with a mixed linker MOF where the needles correspond to UMCM-1.[160,161] Similarly, the addition of NH_2-H_3BTB (~3 mol%) to the reaction mixture at 100 °C for 24 h resulted in cuboctahedral shaped crystals (MOF-5-O_c(575)). Addition of this additive in higher mol% resulted in the final product having an additional crystalline solid phase other than MOF-5 as well as a greater than 15% reduction in BET surface area.

The obtained new morphologies (MOF-5-O_h(600), MOF-5-O_c(856), and MOF-5-O_c(575) where the number indicates the mean crystal size in microns) and their crystal size distributions and statistics are represented in Figures 9a, 9b, and 9c). PXRD patterns of these non-cubic morphologies are consistent with the pattern for MOF-5 simulated from the single crystal X-ray structure (Figure 9d). This reveals that the overall framework structure remains unchanged signaling the ability of additives to engineer morphology without substantially changing crystal structure. All samples (MOF-5-O_h(600), MOF-5-O_c(856), and MOF-5-O_c(575)) exhibit a small reduction (at most ~150 m^2/g) in BET surface area compared to optimal cubic MOF-5 (Figure 9e).

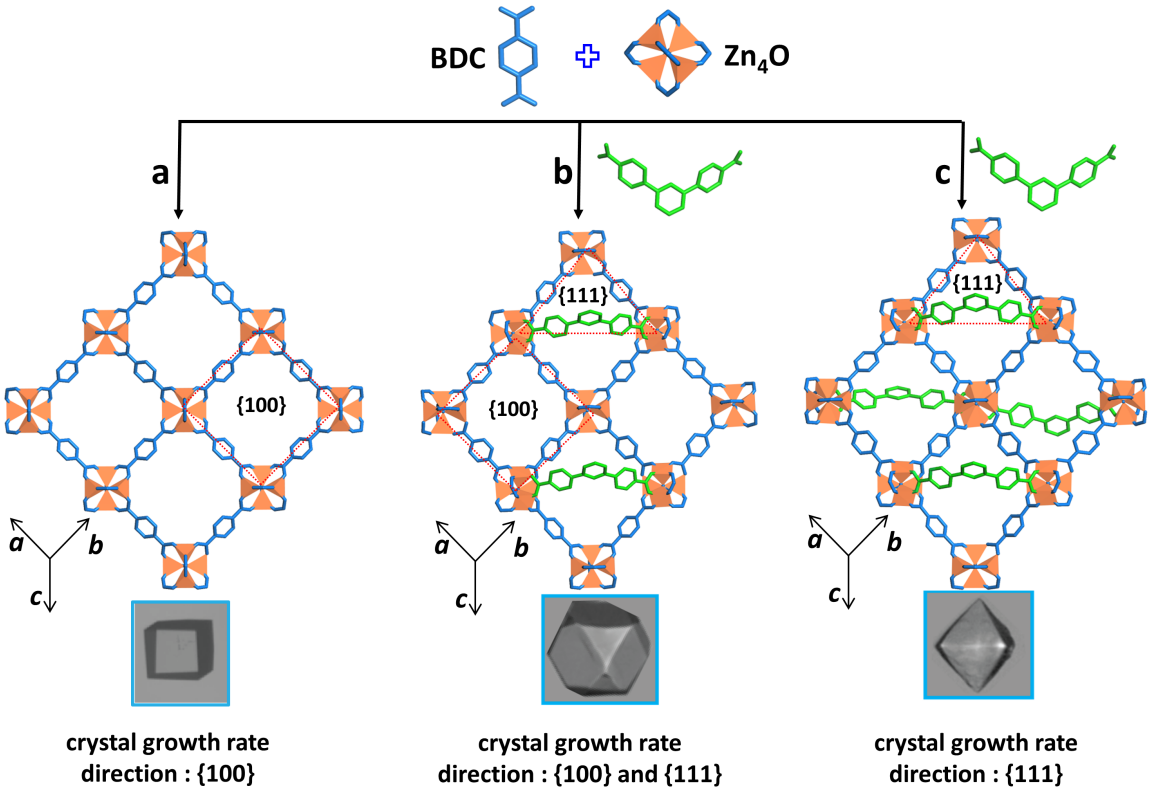
Tetratopic carboxylic acid (H_4L) linkers

Tetratopic (5'-((3,5-dicarboxyphenyl)ethynyl)-[1,1':3',1''-terphenyl]-4,4''-dicarboxylic acid and 5'-(4-carboxyphenyl)-[1,1':3',1''-terphenyl]-3,4'',5-tricarboxylic acid) carboxylic acids added to the initial MOF-5 reagent mixture generated different cuboctahedral, octahedral, and spherical shaped crystals in 24 h at 100 °C respectively

(Figure S3a, Supporting Information). The PXRD patterns of these samples demonstrate that they are MOF-5 (Figure S3b, Supporting Information). However, with these new morphologies reproducibility is a major issue and a greater than 15% reduction in BET surface area was observed making them unsuitable for hydrogen storage.

Mechanism of crystal morphology engineering

To understand the evolution of MOF-5 morphologies in the presence of additives, the relative growth rates of the different surface facets of the crystal must be accounted for. Changing reagent concentration, the presence of additives, and/or modulation of synthesis conditions can suppress the growth rate of certain crystallographic facets. Regardless of changes made to the relative concentrations of reagents (M:L molar ratio), temperature, and reaction time for MOF-5 synthesis, cubic morphology crystals with {100} crystallographic facets were consistently obtained. This indicates that slow crystal growth occurs along the {100} facet (Scheme 1a).



Scheme 1. A schematic representation of crystal growth mechanism for all three morphologies of the MOF-5. The cubic crystal morphology (a) is controlled by the slower crystal growth rate along {100} facet direction. The additive (*m*-terephenyl-4,4′-dicarboxylate shown in green) blocks MOF-5 growth along the {111} direction partially or totally at the expense of all {100} facets during crystal growth which results in the formation of cuboctahedral (b) and octahedral (c) crystal morphologies, respectively.

By contrast, H_3L or H_4L additives significantly alter the relative growth rates of crystal facets, resulting in different crystal morphologies. Among (H_3L or H_4L) additives, only those containing a *m*-terephenyl-4,4′-dicarboxylic acid moiety resulted in non-cubic morphologies. We hypothesize that the *m*-terephenyl-4,4′-dicarboxylic acid moiety is responsible for the formation of the modified crystal morphology because it can bridge two Zn_4O clusters situated diagonally across a pore window in the (100) plane of MOF-5[7] (oxo-oxo length of 18.303 Å). The notion that a *m*-terephenyl-4,4′-dicarboxylate can bridge this distance is supported by inspection of the MOF-177[162] structure; the closest distance between Zn_4O units in this MOFs is 18.432 Å and these are bridged by *m*-terephenyl-4,4′-dicarboxylate units of 1,3,5-tris(4-benzoate)benzene. Thus, during the initial stage of crystal growth, the additive containing the *m*-terephenyl-4,4′-dicarboxylic acid moiety (Scheme 1) blocks MOF-5 growth along the {111} facet direction which results in a slow growth rate of this facet and expression of this crystal surface. This additive blocks the {111} facet direction accompanied by the comparable growths rate of both {100} and {111} facets on the crystal surface when the modulator is at low concentrations. Therefore, cuboctahedral shaped crystals covered by six {100} and eight {111} facets are observed (Scheme 1b). At higher additive concentrations, MOF-5 growth along the {111} facet direction is blocked at the expense of all regularly observed {100} facets resulting in the formation of octahedral crystal morphology entirely covered by eight {111} facets (Scheme 1c). In contrast, additives lacking the *m*-terephenyl-4,4″-dicarboxylic acid motif are larger or smaller than the cluster spacing on the MOF-5 {111} facet and accordingly these additives do not yield new morphologies. These morphological transitions clearly demonstrate that employing additives that selectively interact with certain crystallographic facets and varying their concentration can change the relative growth rate of different crystallographic facets to control morphology and provides a designed way to control morphology that complements approaches based on surfactants.[163,164]

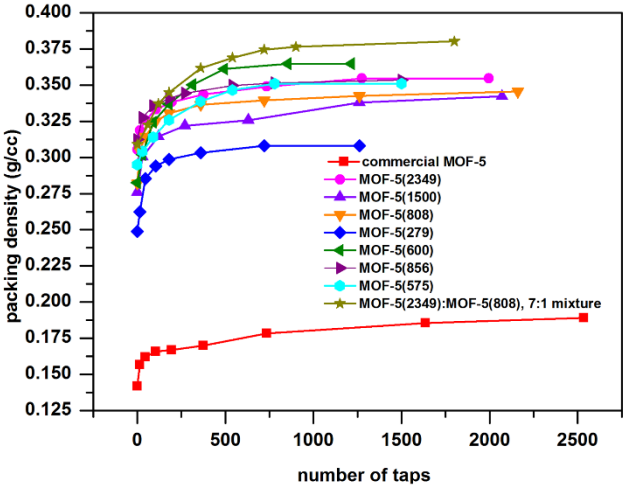


Figure 10. Representative packing density measurements of MOF-5 cubic size-controlled, and different non-cubic crystal morphology samples.

Table 7. Representative packing density improvement for MOF-5 cubic size-controlled, and different non-cubic crystal morphology samples compared to commercial MOF-5.

Sample name	Packing density (g/cc)	Improvement over commercial MOF-5 (in %)
commercial MOF-5	0.189	-
MOF-5(2349)	0.355	87
MOF-5(1500)	0.342	81
MOF-5(808)	0.346	83
MOF-5(279)	0.308	63
MOF-5-O_h(600)	0.365	93
MOF-5-O_c(856)	0.353	87
MOF-5-O_c(575)	0.352	86
MOF-5(2349):MOF-5(808), 7:1 mixture	0.380	100

Packing density

In hydrogen storage applications it is desirable to fill the storage vessel with sorbent in a manner that minimizes the presence of voids and thereby maximizes the volumetric density. The packing density for all samples was optimized and measured using a jolting volumeter (see experimental details in Supporting Information). The number of taps vs. packing density for commercial MOF-5, size-controlled MOF-5 (MOF-5(2349), MOF-5(1500), MOF-5(808), and MOF-5(279)), and controlled crystal morphology (MOF-5-O_h(600), MOF-5-O_c(856), and MOF-5-O_c(575)) samples was monitored. Packing density increases with increasing number of taps until it converges to a constant value (Figure 10). It was observed that all samples exhibit a notable improvement in packing

density of (63% to 93%) compared to commercial MOF-5 (Table 7). Additionally, the packing density of a mixture of size-controlled of MOF-5(2349) and MOF-5(808) in a 7:1 mass ratio[165] displays a remarkable (100%) improvement over commercial MOF-5. All newly synthesized size-controlled cubic and non-cubic morphology crystals possess low external surface area that allows the larger size crystals to flow more easily. Consequently, these samples becomes more compact upon tapping when compared to small crystallite sample of commercial MOF-5, which exhibits high external surface area with more cohesive behavior impeding free flow of crystals.[166,167] Empirically, non-cubic crystal morphology samples (MOF-5-O_h(600), MOF-5-O_c(856), and MOF-5-O_c(575)) have less electrostatic charge than cubic MOF-5 samples and are relatively free-flowing leading to the greater individual packing densities. However, in a practical storage system some degree of mechanical compaction may be used to further increase capacity beyond powder density. It is not certain that these improved powder densities can translate to improvements after densification.

Compaction study

Hydrogen gas storage measurements were performed for commercial MOF-5, size-controlled MOF-5, a bimodal mixture of MOF-5(2349) and MOF-5(808), and octahedral crystal morphology samples to quantify the influence of mechanical compaction on the hydrogen adsorption capacity at 77K. Compaction represents an approach to further densify materials beyond the observed tap densities albeit with the risk of structural damage due to high contact pressures. Selected measurements are provided in Figure S6, Supporting Information, which demonstrate the excess hydrogen adsorption isotherms after compacting samples to successively higher densities. The commercial MOF-5 and size-controlled MOF-5 (MOF-5(2349), MOF-5(1500), MOF-5(808), and MOF-5(279)) samples retain greater than 95% of their excess gravimetric hydrogen capacity up to a compaction density fraction (packing density/crystal density) of between 57-70%; thereafter their hydrogen capacity decreases with increasing compaction density of the sample (Figure 11)[168] consistent with structural damage.[167] By contrast, the size-controlled MOF-5 bimodal mixture of MOF-5(2349) and MOF-5(808) in a 7:1 mass ratio and MOF-5-O_h(600) samples retain high gravimetric capacity up to compaction density of approximately 75% vs. excess hydrogen gas adsorption. Importantly, the compaction density for MOF-5(2349), MOF-5(1500), MOF-5(808) and MOF-5(279), 7:1 mixture, and MOF-5-O_h(600) samples exhibit improvements over the compacted density of commercial MOF-5 with negligible performance loss respectively. The 7:1 mixture resulted in the most notable density improvement of 33% when comparing the inflection point from the commercial MOF-5 to maintain powder adsorption performance (green dashed vs black dashed trend lines). This result indicates that an improvement in powder density results in less damage upon compaction, presumably due to the presence of an optimized arrangement of crystallites and minimization of void space prior to compression. Conse-

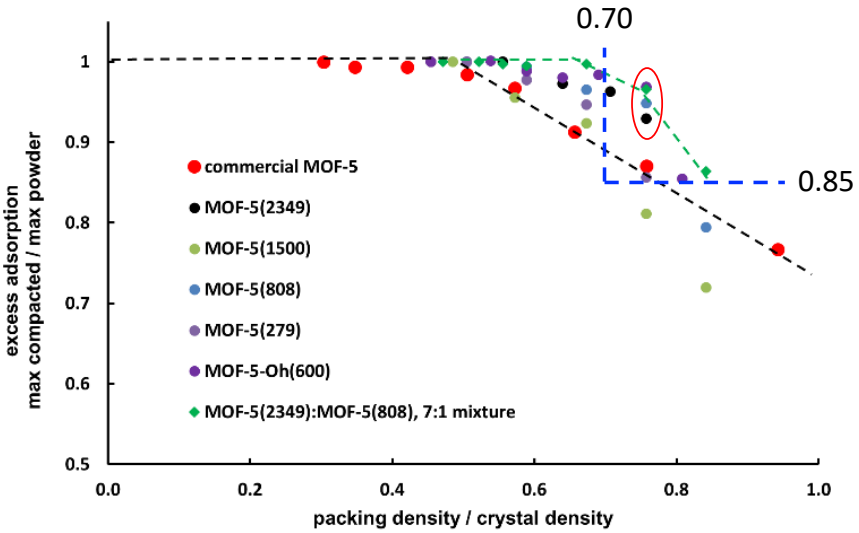


Figure 11. Excess adsorption of hydrogen gas vs. compaction density. The x-axis corresponds to the density of the compacted MOF-5 sample divided by its crystal density (0.594 g/cc).[168] The y-axis corresponds to the ratio between the maximum excess hydrogen adsorption at 77 K for a MOF-5 sample compacted to a specific density by the value for the initial value measured for the MOF-5 sample.

quently, the present samples achieve an unprecedented density of 0.45 g/cm^3 while maintaining gravimetric performance (See Table S5, Supporting Information). Any percentage gain in volumetric capacity from the density improvement is a direct gain in driving range contributing to a projected system can exceed the performance of the current state-of-the art 700 bar compressed gas system in tandem with system engineering improvements as shown in Figure 12.

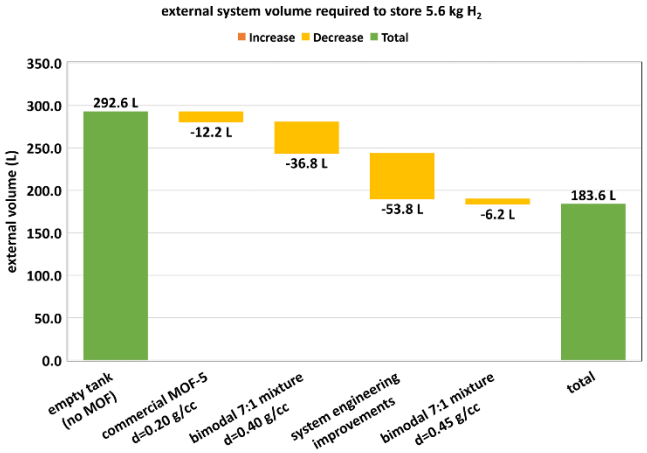


Figure 12. Waterfall chart depicting the total external system volume of bimodal MOF-5(2349):MOF-5(808), 7:1 mixture sample required to store 5.6 kg of usable hydrogen gas. Starting from an empty tank storing hydrogen gas at 77 K and 100 bar, the reduction in external volume is shown for MOF-5(2349):MOF-5(808), 7:1 mixture sample (at 0.40 g/cc and 0.45 g/cc without performance loss) to the system.

Storage System Projections

The correlation between material level and system-level volumetric hydrogen capacities represented in Figure 12. The HSECoE adsorbent system model used for these system predictions takes as input MOF material properties and as output predicts the system characteristics of a full scale 5.6 kg storage c system. The system assumptions include a baseline multi-layer vacuum insulation (MLVI) thickness of 23 mm and a 3/8-inch liquid N_2 channels to reduce the type-1 tank temperature during fueling. The initial system model capacity prediction is based on a full state and an empty state of 100 bar/77 K and 5 bar/160 K, respectively. The complete set of assumptions and schematics of the HSECoE system adsorbent system model have been published by Purewal et al.[169] The effects of the improvements on the system-level volumetric capacity from the 7:1 mixture is depicted in the Figure 12 (waterfall chart). The system engineering improvements include optimization of the vacuum insulation (10 mm), reduction of the N_2 channel diameter (1/4”), and SS type 1 tank. Controlling crystal size and use of bimodal distribution of MOF-5 in tandem with system optimization leads to a large decrease in the external volume of about 109L and exhibits a 30.5 g/L volumetric capacity, sufficient to surpass the 25 g/L volumetric capacity of a typical 700 bar compressed storage system and exceed the DOE 2020 target for volumetric capacity (30 g/L). Similarly impressive results are obtained with MOF-5-O_h(600) of a single size suggesting a pathway for even more dramatic improved by controlling size distribution in tandem with crystal shape.

Summary – Crystal Engineering

The enhancement volumetric storage density of hydrogen storage systems to levels required for automotive applications is a longstanding problem that particularly plagues physical adsorbents. Although compaction of MOFs has been extensively investigated, the challenge of performance erosion as single crystal densities are approached is universally observed. Here, the ability to achieve high volumetric storage densities (75% of single crystal density) without substantial degradation is achieved. A key to success is the use of crystals with a bimodal size distribution or octahedral morphology sample that can fill space efficiently within a bed and can be compacted with minimal structural damage. A complementary approach based on morphology engineering is also shown to be promising and relies on face-selective inhibition of crystal growth to induce non-cubic morphologies. The principles developed here are fully applicable to the large class of cubic MOF sorbents currently being considered in areas as diverse batteries, carbon capture, separations, and fuel gas storage.

End of project goal : *Demonstrate either: (A) 1-2 MOFs having usable volumetric capacities exceeding that of the state of the art material (50 g/L, single-crystal/pressure swing), with less than 10% compromise to gravimetric capacity, or (B) at least one MOF with a packing density equivalent to 70% of the single crystal density with <15% decrease in gravimetric performance.*

As shown in Fig. 11, we successfully achieved the end of project goal "(B)" by demonstrating "at least one MOF with a packing density equivalent to 70% of the single crystal density with <15% decrease in gravimetric performance." Four MOFs samples/morphologies satisfy this milestone: MOF-5-Oh(600), MOF-5(808), MOF-5(2349), and a 7:1 mixture of MOF-5(2349):MOF-5(808). Importantly, the packing density of a mixture of size-controlled of MOF-5(2349) and MOF-5(808) in a 7:1 mass ratio[165] displays a remarkable (100%) improvement over commercial MOF-5.

Acknowledgement

Financial support for this study was provided by the US Department of Energy, Office of Energy Efficiency and Renewable Energy, Grant no. DE-EE0008093. Partial computing resources were provided by the NSF via grant 1531752 MRI: Acquisition of Conflux, A Novel Platform for Data-Driven Computational Physics (Tech. Monitor: Ed Walker). The authors acknowledge Jesse Adams, Dr. Zeric Hulvey, Ms. Courtney Pailing, Mr. Nick Wunder, Ms. Nalinrat Guba and Dr. Caleb Phillips for facilitating web hosting of the ML models and the development of an application programmers interface. A.A. acknowledges Profs. Randall Snurr and Tom Woo for providing access to their MOF databases; Dr. Maciej Haranczyk for use of the Zeo++ code and the mail-order MOF database; and Prof. Adam J. Matzger, Dr. Antek G. Wong-Foy, Dr. Saona Seth, and Dr. Yiyang Liu for support for MOF design.

References

(1) Ahmed, A.; Seth, S.; Purewal, J.; Wong-Foy, A. G.; Veenstra, M.; Matzger, A. J.; Siegel, D. J. Exceptional Hydrogen Storage Achieved by Screening Nearly Half a Million Metal-Organic Frameworks. *Nat. Commun.* **2019**, *10* (1), 1568. https://doi.org/10.1038/s41467-019-09365-w.

(2) Ahmed, A.; Liu, Y.; Purewal, J.; Tran, L. D.; Veenstra, M.; Wong-Foy, A.; Matzger, A.; Siegel, D. Balancing Gravimetric and Volumetric Hydrogen Density in MOFs. *Energy Environ. Sci.* **2017**, *10*, 2459–2471. https://doi.org/10.1039/C7EE02477K.

(3) Wong-Foy, A. G.; Matzger, A. J.; Yaghi, O. M. Exceptional H2 Saturation Uptake in Microporous Metal-Organic Frameworks. *J. Am. Chem. Soc.* **2006**, *128*, 3494–3495. https://doi.org/10.1021/ja058213h.

(4) Satyapal, S.; Petrovic, J.; Read, C.; Thomas, G.; Ordaz, G. The U.S. Department of Energy's National Hydrogen Storage Project: Progress towards Meeting Hydrogen-Powered Vehicle Requirements. *Catal. Today* **2007**, *120* (3), 246–256. https://doi.org/10.1016/j.cattod.2006.09.022.

(5) Greene, D. L.; Joseck, F.; Duleep, G.; -Oak, G. *XI-13 FY 2013 Annual Progress Report DOE Hydrogen and Fuel Cells Program DOE Manager Subcontractor: Overall Objectives Contribution to Achievement of DOE Systems Analysis Milestones XI.1 Worldwide Status of Hydrogen Fuel Cell Vehicle Technology and Prospe*.

(6) Allendorf, M. D.; Hulvey, Z.; Gennett, T.; Ahmed, A.; Autrey, T.; Camp, J.; Seon Cho, E.; Furukawa, H.; Haranczyk, M.; Head-Gordon, M.; et al. An Assessment of Strategies for the Development of Solid-State Adsorbents for Vehicular Hydrogen Storage. *Energy Environ. Sci.* **2018**, *11* (10), 2784–2812. https://doi.org/10.1039/C8EE01085D.

(7) Yang, J.; Sudik, A.; Wolverton, C.; Siegel, D. J. High Capacity Hydrogen Storage Materials: Attributes for Automotive Applications and Techniques for Materials Discovery. *Chem. Soc. Rev.* **2010**, *39*, 656–675.

https://doi.org/10.1039/b802882f.
(8) Long, J. R. *201 "Hydrogen Storage in Metal-Organic Frameworks," J. R. Long; U.S. Department of Energy, Hydrogen and Fuel Cells Program 2015 Annual Merit Review Proceedings: Project ST103. Https://Www.Hydrogen.Energy.Gov/Pdfs/Review15/St103_long_2015_o.Pdf.*; 2015.
(9) DOE Technical Targets for Onboard Hydrogen Storage for Light-Duty Vehicles, Https://Energy.Gov/Eere/Fuelcells/Doe-Technical-Targets-Onboard-Hydrogen-Storage-Light-Duty-Vehicles.
(10) Astiaso Garcia, D.; Barbanera, F.; Cumo, F.; Di Matteo, U.; Nastasi, B. Expert Opinion Analysis on Renewable Hydrogen Storage Systems Potential in Europe. *Energies* **2016**, *9* (11), 963. https://doi.org/10.3390/en9110963.
(11) Riis, T.; Sandrock, G.; Ulleberg, Ø.; Vie, P. J. S. Hydrogen Storage R&D: Priorities and Gaps. In *Hydrogen Production and Storage: R&D Priorities and Gaps*; International Energy Agency: Paris, France, 2006; pp 19–33.
(12) Purewal, J.; Veenstra, M.; Tamburello, D.; Ahmed, A.; Matzger, A. J.; Wong-Foy, A. G.; Seth, S.; Liu, Y.; Siegel, D. J. Estimation of System-Level Hydrogen Storage for Metal-Organic Frameworks with High Volumetric Storage Density. *Int. J. Hydrogen Energy* **2019**, *44* (29). https://doi.org/10.1016/j.ijhydene.2019.04.082.
(13) Manoharan, Y.; Hosseini, S. E.; Butler, B.; Alzhahrani, H.; Senior, B. T. F.; Ashuri, T.; Krohn, J.; Manoharan, Y.; Hosseini, S. E.; Butler, B.; et al. Hydrogen Fuel Cell Vehicles; Current Status and Future Prospect. *Appl. Sci.* **2019**, *9* (11), 2296. https://doi.org/10.3390/app9112296.
(14) Makridis, S. S. Hydrogen Storage and Compression. In *Methane and Hydrogen for Energy Storage*; Carriveau, R., Ting, D. S.-K., Eds.; The Institution of Engineering and Technology, 2016; pp 1–28.
(15) "Ford/BASF-SE/UM Activities in Support of the Hydrogen Storage Engineering Center of Excellence," Mike Veenstra, Jun Yang, Chunchuan Xu, Manuela Gaab, Lena Arnold, Ulrich Muller, Donald J. Siegel, and Yang Ming.; U.S. Department of Energy, Hydrogen and Fuel Cells Program 2014 Annual Merit Review Proceedings: Project ST010.https://www.hydrogen.energy.gov/pdfs/review14/st010_veenstra_2014_o.pdf.
(16) Öhrström, L. Let's Talk about MOFs—Topology and Terminology of Metal-Organic Frameworks and Why We Need Them. *Crystals* **2015**, *5* (1), 154–162. https://doi.org/10.3390/cryst5010154.
(17) Fischer, R. A.; Schwedler, I. Terminologie von Metall-Organischen Gerüstverbindungen Und Koordinationspolymeren (IUPAC-Empfehlungen 2013). *Angew. Chemie* **2014**, *126* (27), 7209–7214. https://doi.org/10.1002/ange.201400619.
(18) Batten, S. R.; Champness, N. R.; Chen, X.-M.; Garcia-Martinez, J.; Kitagawa, S.; Öhrström, L.; O'Keeffe, M.; Paik Suh, M.; Reedijk, J. Terminology of Metal–Organic Frameworks and Coordination Polymers (IUPAC Recommendations 2013). *Pure Appl. Chem.* **2013**, *85* (8), 1715–1724. https://doi.org/10.1351/PAC-REC-12-11-20.
(19) Thommes, M.; Kaneko, K.; Neimark, A. V.; Olivier, J. P.; Rodriguez-Reinoso, F.; Rouquerol, J.; Sing, K. S. W. Physisorption of Gases, with Special Reference to the Evaluation of Surface Area and Pore Size Distribution (IUPAC Technical Report). *Pure Appl. Chem.* **2015**, *87* (9–10), 1051–1069. https://doi.org/10.1515/pac-2014-1117.
(20) Batten, S. R.; Champness, N. R.; Chen, X.-M.; Garcia-Martinez, J.; Kitagawa, S.; Öhrström, L.; O'Keeffe, M.; Suh, M. P.; Reedijk, J. Coordination Polymers, Metal–Organic Frameworks and the Need for Terminology Guidelines. *CrystEngComm* **2012**, *14* (9), 3001. https://doi.org/10.1039/c2ce06488j.
(21) O'Keeffe, M. Nets, Tiles, and Metal-Organic Frameworks. *APL Mater.* **2014**, *2* (12), 124106. https://doi.org/10.1063/1.4901292.
(22) Tranchemontagne, D. J.; Mendoza-Cortés, J. L.; O'Keeffe, M.; Yaghi, O. M. Secondary Building Units, Nets and Bonding in the Chemistry of Metal–Organic Frameworks. *Chem. Soc. Rev.* **2009**, *38* (5), 1257. https://doi.org/10.1039/b817735j.

(23) Reymond, J.-L. The Chemical Space Project. *Acc. Chem. Res.* **2015**, *48* (3), 722–730. https://doi.org/10.1021/ar500432k.
(24) Kontijevskis, A. Mapping of Drug-like Chemical Universe with Reduced Complexity Molecular Frameworks. *J. Chem. Inf. Model.* **2017**, *57* (4), 680–699. https://doi.org/10.1021/acs.jcim.7b00006.
(25) Martin, R. L.; Smit, B.; Haranczyk, M. Addressing Challenges of Identifying Geometrically Diverse Sets of Crystalline Porous Materials. *J. Chem. Inf. Model.* **2012**, *52* (2), 308–318. https://doi.org/10.1021/ci200386x.
(26) Sun, D.; Sun, F.; Deng, X.; Li, Z. Mixed-Metal Strategy on Metal–Organic Frameworks (MOFs) for Functionalities Expansion: Co Substitution Induces Aerobic Oxidation of Cyclohexene over Inactive Ni-MOF-74. *Inorg. Chem.* **2015**, *54* (17), 8639–8643. https://doi.org/10.1021/acs.inorgchem.5b01278.
(27) Deng, H.; Doonan, C. J.; Furukawa, H.; Ferreira, R. B.; Towne, J.; Knobler, C. B.; Wang, B.; Yaghi, O. M. Multiple Functional Groups of Varying Ratios in Metal-Organic Frameworks. *Science* **2010**, *327* (5967), 846–850. https://doi.org/10.1126/science.1181761.
(28) Park, J.; Kim, H.; Han, S. S.; Jung, Y. Tuning Metal–Organic Frameworks with Open-Metal Sites and Its Origin for Enhancing CO_2 Affinity by Metal Substitution. *J. Phys. Chem. Lett.* **2012**, *3* (7), 826–829. https://doi.org/10.1021/jz300047n.
(29) Moghadam, P. Z.; Li, A.; Wiggin, S. B.; Tao, A.; Maloney, A. G. P.; 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* (7), 2618–2625. https://doi.org/10.1021/acs.chemmater.7b00441.
(30) Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C. The Cambridge Structural Database. *Acta Crystallogr. Sect. B Struct. Sci. Cryst. Eng. Mater.* **2016**, *72* (2), 171–179. https://doi.org/10.1107/S2052520616003954.
(31) 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* (16), 3373–3382. https://doi.org/10.1021/cm401978e.
(32) Altintas, C.; Avci, G.; Daglar, H.; Nemati Vesali Azar, A.; Erucar, I.; Velioglu, S.; Keskin, S. An Extensive Comparative Analysis of Two MOF Databases: High-Throughput Screening of Computation-Ready MOFs for CH4 and H2 Adsorption. *J. Mater. Chem. A* **2019**, *7* (16), 9593–9608. https://doi.org/10.1039/c9ta01378d.
(33) Chung, Y. G.; Camp, J.; Haranczyk, M.; Sikora, B. J.; Bury, W.; Krungleviciute, V.; Yildirim, T.; Farha, O. K.; Sholl, D. S.; Snurr, R. Q. Computation-Ready, Experimental Metal–Organic Frameworks: A Tool To Enable High-Throughput Screening of Nanoporous Crystals. *Chem. Mater.* **2014**, *26*, 6185–6192.
(34) Chung, Y. G.; Haldoupis, E.; Bucior, B. J.; Haranczyk, M.; Lee, S.; Zhang, H.; Vogiatzis, K. D.; Milisavljevic, M.; Ling, S.; Camp, J. S.; et al. Advances, Updates, and Analytics for the Computation-Ready, Experimental Metal–Organic Framework Database: CoRE MOF 2019. *J. Chem. Eng. Data* **2019**, *64* (12), 5985–5998. https://doi.org/10.1021/acs.jced.9b00835.
(35) Chen, T.; Manz, T. A. Identifying Misbonded Atoms in the 2019 CoRE Metal–Organic Framework Database. *RSC Adv.* **2020**, *10* (45), 26944–26951. https://doi.org/10.1039/D0RA02498H.
(36) 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. *Nat. Chem.* **2011**, *4* (2), 83–89. https://doi.org/10.1038/nchem.1192.
(37) Aghaji, M. Z.; Fernandez, M.; Boyd, P. G.; Daff, T. D.; Woo, T. K. Quantitative Structure – Property Relationship Models for Recognizing Metal Organic Frameworks (MOFs) with High CO 2 Working Capacity and CO2/CH4 Selectivity for Methane Purification. **2016**, 4505–4511. https://doi.org/10.1002/ejic.201600365.
(38) Martin, R. L.; Lin, L. C.; Jariwala, K.; Smit, B.; Haranczyk, M. Mail-Order Metal-Organic Frameworks (MOFs): Designing Isoreticular MOF-5 Analogues Comprising Commercially Available Organic

Molecules. *J. Phys. Chem. C* **2013**, *117* (23), 12159–12167. https://doi.org/10.1021/jp401920y.
(39) Bao, Y.; Martin, R. L.; Haranczyk, M.; Deem, M. W. In Silico Prediction of MOFs with High Deliverable Capacity or Internal Surface Area. *Phys. Chem. Chem. Phys.* **2015**, *17* (18), 11962–11973. https://doi.org/10.1039/C5CP00002E.
(40) Witman, M.; Ling, S.; Anderson, S.; Tong, L.; Stylianou, K. C.; Slater, B.; Smit, B.; Haranczyk, M. In Silico Design and Screening of Hypothetical MOF-74 Analogs and Their Experimental Synthesis. *Chem. Sci.* **2016**, *7* (9), 6263–6272. https://doi.org/10.1039/C6SC01477A.
(41) Chung, Y. G.; Gómez-gualdrón, D. A.; Li, P.; Leperi, K. T.; Deria, P.; Zhang, H.; Vermeulen, N. A.; Stoddart, J. F.; You, F.; Hupp, J. T.; et al. In Silico Discovery of Metal-Organic Frameworks for Precombustion CO 2 Capture Using a Genetic Algorithm. **2016**, No. October.
(42) Anderson, R.; Rodgers, J.; Argueta, E.; Biong, A.; Go, D. A. Role of Pore Chemistry and Topology in the CO 2 Capture Capabilities of MOFs: From Molecular Simulation to Machine Learning. *Chem. Mater* **2018**, *30*, 11. https://doi.org/10.1021/acs.chemmater.8b02257.
(43) 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* **2019**, *123* (1), 120–130. https://doi.org/10.1021/acs.jpcc.8b09420.
(44) Anderson, R.; Gómez-Gualdrón, D. A. Increasing Topological Diversity during Computational "Synthesis" of Porous Crystals: How and Why. *CrystEngComm* **2019**, *21* (10), 1653–1665. https://doi.org/10.1039/c8ce01637b.
(45) Gomez-Gualdron, D. A.; Gutov, O. V.; Krungleviciute, V.; Borah, B.; Mondloch, J. E.; Hupp, J. T.; Yildirim, T.; Farha, O. K.; Snurr, R. Q. Computational Design of Metal–Organic Frameworks Based on Stable Zirconium Building Units for Storage and Delivery of Methane. *Chem. Mater.* **2014**, *26* (19), 5632–5639. https://doi.org/10.1021/cm502304e.
(46) Bao, Y.; Martin, R. L.; Simon, C. M.; Haranczyk, M.; Smit, B.; Deem, M. W. In Silico Discovery of High Deliverable Capacity Metal–Organic Frameworks. *J. Phys. Chem. C* **2015**, *119* (1), 186–195. https://doi.org/10.1021/jp5123486.
(47) Willems, T. F.; Rycroft, C. H.; Kazi, M.; Meza, J. C.; Haranczyk, M. Algorithms and Tools for High-Throughput Geometry-Based Analysis of Crystalline Porous Materials. *Microporous Mesoporous Mater.* **2012**, *149* (1), 134–141. https://doi.org/10.1016/j.micromeso.2011.08.020.
(48) Addicoat, M. A.; Coupry, D. E.; Heine, T. AuToGraFS: Automatic Topological Generator for Framework Structures. *J. Phys. Chem. A* **2014**, *118* (40), 9607–9614. https://doi.org/10.1021/jp507643v.
(49) Boyd, P. G.; Woo, T. K. A Generalized Method for Constructing Hypothetical Nanoporous Materials of Any Net Topology from Graph Theory. *CrystEngComm* **2016**, *18* (21), 3777–3792. https://doi.org/10.1039/C6CE00407E.
(50) 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 Environ. Sci.* **2016**, *9* (10), 3279–3289. https://doi.org/10.1039/C6EE02104B.
(51) Yao, Z.; Sanchez-Lengeling, B.; Bobbitt, N. S.; Bucior, B. J.; Kumar, S. G. H.; Collins, S. P.; Burns, T.; Woo, T. K.; Farha, O.; Snurr, R. Q.; et al. Inverse Design of Nanoporous Crystalline Reticular Materials with Deep Generative Models. Https://Doi.Org/10.26434/Chemrxiv.12186681.V2. **2020**. https://doi.org/10.26434/CHEMRXIV.12186681.V2.
(52) Sadus, R. J. *Molecular Simulation of Fluids: Theory, Algorithms, and Object-Orientation.*; Elsevier: Amsterdam, 1999.
(53) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press: New York, NY, 1989.
(54) Frenkel, D.; Smit, B. *Understanding Molecular Simulation : From Algorithms to Applications*, 2nd ed.;

Academic Press, Inc.: Orlando, FL, 2001.
(55) Hill, T. L. *An Introduction to Statistical Thermodynamics*; Dover Publications, 1986.
(56) Dubbeldam, D.; Torres-Knoop, A.; Walton, K. S. Molecular Simulation On the Inner Workings of Monte Carlo Codes On the Inner Workings of Monte Carlo Codes. *Mol. Simul.* **2013**, *39*, 14–15. https://doi.org/10.1080/08927022.2013.819102.
(57) 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 CO 2 Capture. *J. Phys. Chem. Lett.* **2014**, *5* (17), 3056–3060. https://doi.org/10.1021/jz501331m.
(58) Martin, R. L.; Simon, C. M.; Smit, B.; Haranczyk, M. *In Silico* Design of Porous Polymer Networks: High-Throughput Screening for Methane Storage Materials. *J. Am. Chem. Soc.* **2014**, *136* (13), 5006–5022. https://doi.org/10.1021/ja4123939.
(59) Colón, Y. J.; Gómez-Gualdrón, D. A.; Snurr, R. Q. Topologically Guided, Automated Construction of Metal–Organic Frameworks and Their Evaluation for Energy-Related Applications. *Cryst. Growth Des.* **2017**, *17* (11), 5801–5810. https://doi.org/10.1021/acs.cgd.7b00848.
(60) Boyd, P. G.; Moosavi, S. M.; Witman, M.; Smit, B. Force-Field Prediction of Materials Properties in Metal-Organic Frameworks. *J. Phys. Chem. Lett.* **2017**, *8* (2), 357–363. https://doi.org/10.1021/acs.jpclett.6b02532.
(61) Thornton, A. W.; Simon, C. M.; Kim, J.; Kwon, O.; Deeg, K. S.; Konstas, K.; Pas, S. J.; Hill, M. R.; Winkler, D. A.; Haranczyk, M.; et al. Materials Genome in Action: Identifying the Performance Limits of Physical Hydrogen Storage. *Chem. Mater.* **2017**, *29* (7), 2844–2854. https://doi.org/10.1021/acs.chemmater.6b04933.
(62) Bobbitt, N. S.; Snurr, R. Q. Molecular Simulation Molecular Modelling and Machine Learning for High-Throughput Screening of Metal-Organic Frameworks for Hydrogen Storage Molecular Modelling and Machine Learning for High-Throughput Screening of Metal-Organic Frameworks for Hydrogen Storage. **2019**. https://doi.org/10.1080/08927022.2019.1597271.
(63) 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 Comput. Mater.* **2017**, *3*. https://doi.org/10.1038/s41524-017-0045-8.
(64) Broom, D. P.; Webb, C. J.; Hurst, K. E.; Parilla, P. A.; Gennett, T.; Brown, C. M.; Zacharia, R.; Tylianakis, E.; Klontzas, E.; Froudakis, G. E.; et al. Outlook and Challenges for Hydrogen Storage in Nanoporous Materials. *Appl. Phys. A* **2016**, *122* (3), 151. https://doi.org/10.1007/s00339-016-9651-4.
(65) Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine Learning for Molecular and Materials Science. *Nature* **2018**, *559* (7715), 547–555. https://doi.org/10.1038/s41586-018-0337-2.
(66) Wahiduzzaman, M.; Walther, C. F. J.; Heine, T. Hydrogen Adsorption in Metal-Organic Frameworks: The Role of Nuclear Quantum Effects. *J. Chem. Phys.* **2014**, *141* (6), 064708. https://doi.org/10.1063/1.4892670.
(67) Durette, D.; Bénard, P.; Zacharia, R.; Chahine, R. Investigation of the Hydrogen Adsorbed Density inside the Pores of MOF-5 from Path Integral Grand Canonical Monte Carlo at Supercritical and Subcritical Temperature. *Sci. Bull.* **2016**, *61* (8), 594–600. https://doi.org/10.1007/s11434-016-1027-9.
(68) Fischer, M.; Hoffmann, F.; Fröba, M. Preferred Hydrogen Adsorption Sites in Various MOFs-A Comparative Computational Study. *ChemPhysChem* **2009**, *10* (15), 2647–2657. https://doi.org/10.1002/cphc.200900459.
(69) Furukawa, H.; Miller, M. A.; Yaghi, O. M. Independent Verification of the Saturation Hydrogen Uptake in MOF-177 and Establishment of a Benchmark for Hydrogen Adsorption in Metal–Organic Frameworks. *J. Mater. Chem.* **2007**, *17* (30), 3197. https://doi.org/10.1039/b703608f.
(70) Fischer, M.; Hoffmann, F.; Fröba, M. Preferred Hydrogen Adsorption Sites in Various MOFs-A Comparative Computational Study. *ChemPhysChem* **2009**, *10* (15), 2647–2657. https://doi.org/10.1002/cphc.200900459.

(71) Ahmed, A.; Siegel, D. J. HyMARC datahub https://datahub.hymarc.org/dataset/computational-prediction-of-hydrogen-storage-capacities-in-mofs.
(72) 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* (10), 640–645. https://doi.org/10.1021/acscombsci.7b00056.
(73) Fernandez, M.; Woo, T. K.; Wilmer, C. E.; Snurr, R. Q. Large-Scale Quantitative Structure–Property Relationship (QSPR) Analysis of Methane Storage in Metal–Organic Frameworks. *J. Phys. Chem. C* **2013**, *117*, 7681–7689. https://doi.org/10.1021/jp4006422.
(74) Fernandez, M.; Trefiak, N. R.; Woo, T. K. Atomic Property Weighted Radial Distribution Functions Descriptors of Metal–Organic Frameworks for the Prediction of Gas Uptake Capacity. *J. Phys. Chem. C* **2013**, *117* (27), 14095–14105. https://doi.org/10.1021/jp404287t.
(75) Fernandez, M.; Barnard, A. S. Geometrical Properties Can Predict CO_2 and N_2 Adsorption Performance of Metal–Organic Frameworks (MOFs) at Low Pressure. *ACS Comb. Sci.* **2016**, *18* (5), 243–252. https://doi.org/10.1021/acscombsci.5b00188.
(76) Nanoporous Materials Genome Center. http://www.chem.umn.edu/nmgc/.
(77) Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning*; Springer Series in Statistics; Springer New York: New York, NY, 2009. https://doi.org/10.1007/978-0-387-84858-7.
(78) Thornton, A. W.; Simon, C. M.; Kim, J.; Kwon, O.; Deeg, K. S.; Konstas, K.; Pas, S. J.; Hill, M. R.; Winkler, D. A.; Haranczyk, M.; et al. Materials Genome in Action: Identifying the Performance Limits of Physical Hydrogen Storage. *Chem. Mater.* **2017**, *29* (7), 2844–2854. https://doi.org/10.1021/acs.chemmater.6b04933.
(79) Dorugade, A. V; Kashid, D. N. *Alternative Method for Choosing Ridge Parameter for Regression*; 2010; Vol. 4.
(80) Van Wieringen, W. N. *Lecture Notes on Ridge Regression*; 2020.
(81) Smola, A. J.; Smola, A. J.; Schölkopf, B. A Tutorial on Support Vector Regression. **2004**.
(82) Bucior, B. J.; Bobbitt, N. S.; Islamoglu, T.; Goswami, S.; Gopalan, A.; Yildirim, T.; Farha, O. K.; Bagheri, N.; Snurr, R. Q. Energy-Based Descriptors to Rapidly Predict Hydrogen Storage in Metal-Organic Frameworks. *Mol. Syst. Des. Eng. 2018. DOI 10.1039/c8me00050f.* https://doi.org/10.1039/c8me00050f.
(83) Lan, Y.; Yan, T.; Tong, M.; Zhong, C. Large-Scale Computational Assembly of Ionic Liquid/MOF Composites: Synergistic Effect in the Wire-Tube Conformation for Efficient CO2/CH4 Separation. *J. Mater. Chem. A* **2019**, *7* (20), 12556–12564. https://doi.org/10.1039/c9ta01752f.
(84) Li, S.; Chung, Y. G.; Simon, C. M.; Snurr, R. Q. High-Throughput Computational Screening of Multivariate Metal– Organic Frameworks (MTV-MOFs) for CO 2 Capture. *J. Phys. Chem. Lett* **2017**, *8*, 19. https://doi.org/10.1021/acs.jpclett.7b02700.
(85) Geurts, P.; Ernst, D.; Wehenkel, L. Extremely Randomized Trees. *Mach. Learn.* **2006**, *63* (1), 3–42. https://doi.org/10.1007/s10994-006-6226-1.
(86) Ahmed, A.; Siegel, D. J. HyMARC Sorbent Machine Learning Model: Predicting the hydrogen storage capacity of metal-organic frameworks via machine learning. https://sorbent-ml.hymarc.org/.
(87) 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.; et al. Data-Driven Design of Metal–Organic Frameworks for Wet Flue Gas CO2 Capture. *Nature* **2019**, *576* (7786), 253–256. https://doi.org/10.1038/s41586-019-1798-7.
(88) 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.; et al. Data-Driven Design of Metal-Organic Frameworks for Wet Flue Gas CO2 Capture, Materials Cloud Archive 2018.0016/v3 (2019), Doi: 10.24435/Materialscloud:2018.0016/V3. *Nature* **2019**, *576* (7786), 253–256.

https://doi.org/10.1038/s41586-019-1798-7.
(89) R-WLLFHS: Https://Github.Com/Snurr-Group/Reduced-HMOF-Database.
(90) García-Holley, P.; Schweitzer, B.; Islamoglu, T.; Liu, Y.; Lin, L.; Rodriguez, S.; Weston, M. H.; Hupp, J. T.; Gómez-Gualdrón, D. A.; Yildirim, T.; et al. Benchmark Study of Hydrogen Storage in Metal–Organic Frameworks under Temperature and Pressure Swing Conditions. *ACS Energy Lett.* **2018**, 748–754. https://doi.org/10.1021/acsenergylett.8b00154.
(91) Wolpert, D. H.; Macready, W. G. No Free Lunch Theorems for Optimization. *IEEE Trans. Evol. Comput.* **1997**, *1* (1).
(92) Breiman, L.; Friedman, J. H.; Olshen, R. A.; Stone, C. J. *Classification And Regression Trees*; Routledge, 2017. https://doi.org/10.1201/9781315139470.
(93) Freund, Y.; Schapire, R. E. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *J. Comput. Syst. Sci.* **1997**, *55* (1), 119–139. https://doi.org/10.1006/jcss.1997.1504.
(94) Drucker, H. Improving Regressors using Boosting Techniques https://dl.acm.org/doi/10.5555/645526.657132 (accessed Nov 18, 2020).
(95) Breiman, L. Bagging Predictors. *Mach. Learn.* **1996**, *24* (2), 123–140. https://doi.org/10.1023/A:1018054314350.
(96) Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45* (1), 5–32. https://doi.org/10.1023/A:1010933404324.
(97) Friedman, J. Greedy Function Approximation: A Gradient Boosting Machine. *Ann. Stat.* **2001**, *29* (5), 1189–1232.
(98) Chang, C.-C.; Lin, C.-J. *LIBSVM: A Library for Support Vector Machines*; 2001.
(99) Platt, J. C.; Platt, J. C. Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. *Adv. LARGE MARGIN Classif.* **1999**, 61--74.
(100) Buhmann, M. D. *Radial Basis Functions: Theory and Implementations*; Cambridge University Press: Cambridge, United Kingdom, 2002.
(101) Fan, R.-E.; Chang, K.-W.; Hsieh, C.-J.; Wang, X.-R.; Lin, C.-J. *LIBLINEAR: A Library for Large Linear Classification*; 2008; Vol. 9.
(102) Rifkin, R. M. .; Lippert, R. A. *Notes on Regularized Least Squares*; 2007.
(103) Altman, N. S. An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression. *Am. Stat.* **1992**, *46* (3), 175–185. https://doi.org/10.1080/00031305.1992.10475879.
(104) Freund, Y.; Schapire, R. E. A Short Introduction to Boosting. *J. Japanese Soc. Artif. Intell.* **1999**, *14* (5), 771–780.
(105) Fernández-Delgado, M.; Sirsat, M. S.; Cernadas, E.; Alawadi, S.; Barro, S.; Febrero-Bande, M. An Extensive Experimental Survey of Regression Methods. *Neural Networks* **2019**, *111*, 11–34. https://doi.org/10.1016/J.NEUNET.2018.12.010.
(106) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
(107) Richard, M.-A.; Bénard, P.; Chahine, R. Gas Adsorption Process in Activated Carbon over a Wide Temperature Range above the Critical Point. Part 1: Modified Dubinin-Astakhov Model. *Adsorption* **2009**, *15* (1), 43–51. https://doi.org/10.1007/s10450-009-9149-x.
(108) Gomez-Gualdron, D. A.; Wang, T. C.; García-Holley, P.; Sawelewa, R. M.; Argueta, E.; Snurr, R. Q.; Hupp, J. T.; Yildirim, T.; Farha, O. K. Understanding Volumetric and Gravimetric Hydrogen Adsorption Trade-off in Metal–Organic Frameworks. *ACS Appl. Mater. Interfaces* **2017**, *9* (39), 33419–33428. https://doi.org/10.1021/acsami.7b01190.
(109) Düren, T.; Bae, Y.-S.; Snurr, R. Q. Using Molecular Simulation to Characterise Metal–Organic Frameworks for Adsorption Applications. *Chem. Soc. Rev.* **2009**, *38* (5), 1237. https://doi.org/10.1039/b803498m.

(110) Allendorf, M. D.; Bauer, C. A.; Bhakta, R. K.; Houk, R. J. T. Luminescent Metal–Organic Frameworks. *Chem. Soc. Rev.* **2009**, *38* (5), 1330. https://doi.org/10.1039/b802352m.
(111) Willems, T. F.; Rycroft, C. H.; Kazi, M.; Meza, J. C.; Haranczyk, M. Algorithms and Tools for High-Throughput Geometry-Based Analysis of Crystalline Porous Materials. *Microporous Mesoporous Mater.* **2012**, *149* (1), 134–141. https://doi.org/10.1016/j.micromeso.2011.08.020.
(112) Gómez-Gualdrón, D. A.; Moghadam, P. Z.; Hupp, J. T.; Farha, O. K.; Snurr, R. Q. Application of Consistency Criteria To Calculate BET Areas of Micro- And Mesoporous Metal–Organic Frameworks. *J. Am. Chem. Soc.* **2016**, *138* (1), 215–224. https://doi.org/10.1021/jacs.5b10266.
(113) Himanen, L.; Geurts, A.; Foster, A. S.; Rinke, P. Data-Driven Materials Science: Status, Challenges, and Perspectives. *Adv. Sci.* **2019**, 1900808. https://doi.org/10.1002/advs.201900808.
(114) Wei, J.; Chu, X.; Sun, X.; Xu, K.; Deng, H.; Chen, J.; Wei, Z.; Lei, M. Machine Learning in Materials Science. *InfoMat* **2019**, *1* (3), 338–358. https://doi.org/10.1002/inf2.12028.
(115) 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**, acs.jpca.9b03290. https://doi.org/10.1021/acs.jpca.9b03290.
(116) Panella, B.; Hirscher, M.; Roth, S. Hydrogen Adsorption in Different Carbon Nanostructures. *Carbon N. Y.* **2005**, *43* (10), 2209–2214. https://doi.org/10.1016/j.carbon.2005.03.037.
(117) Kiyabu, S.; Lowe, J. S.; Ahmed, A.; Siegel, D. J. Computational Screening of Hydration Reactions for Thermal Energy Storage: New Materials and Design Rules. *Chem. Mater.* **2018**, *30* (6), 2006–2017. https://doi.org/10.1021/acs.chemmater.7b05230.
(118) Moosavi, S. M.; Chidambaram, A.; Talirz, L.; Haranczyk, M.; Stylianou, K. C.; Smit, B. Capturing Chemical Intuition in Synthesis of Metal-Organic Frameworks. *Nat. Commun.* **2019**, *10* (1), 539. https://doi.org/10.1038/s41467-019-08483-9.
(119) Zwillinger, D.; Kokoska, S.; Raton, B.; New, L.; Washington, Y. *Standard Probability and Statistics Tables and Formulae CRC*; 2000.
(120) Oliphant, T. E. Python for Scientific Computing. *Comput. Sci. Eng.* **2007**, *9* (3), 10–20. https://doi.org/10.1109/MCSE.2007.58.
(121) Millman, K. J.; Aivazis, M. Python for Scientists and Engineers. *Comput. Sci. Eng.* **2011**, *13* (2), 9–12. https://doi.org/10.1109/MCSE.2011.36.
(122) Parrt, T.; Turgutlu, K. Rfpimp 1.3.4, Https://Github.Com/Parrt/Random-Forest-Importances.
(123) Machine, P.; With, L. Improve Machine Learning Results with Boosting , Bagging and Blending Ensemble Methods in Weka Start and Practice Machine Learning With Weka. **2017**, 1–12.
(124) Here, S.; Products, B.; Contact, A. Classification And Regression Trees for Machine Learning. **2017**, 1–11.
(125) Witman, M.; Ling, S.; Grant, D. M.; Walker, G. S.; Agarwal, S.; Stavila, V.; Allendorf, M. D. Extracting an Empirical Intermetallic Hydride Design Principle from Limited Data via Interpretable Machine Learning. *J. Phys. Chem. Lett.* **2020**, *11* (1), 40–47. https://doi.org/10.1021/acs.jpclett.9b02971.
(126) Sturluson, A.; Huynh, M. T.; Kaija, A. R.; Laird, C.; Yoon, S.; Hou, F.; Feng, Z.; Wilmer, C. E.; Colón, Y. J.; Chung, Y. G.; et al. The Role of Molecular Modelling and Simulation in the Discovery and Deployment of Metal-Organic Frameworks for Gas Storage and Separation. *Mol. Simul.* **2019**, *45* (14–15), 1082–1121. https://doi.org/10.1080/08927022.2019.1648809.
(127) Barthel, S.; Alexandrov, E. V; Proserpio, D. M.; Smit, B. Distinguishing Metal–Organic Frameworks. **2018**. https://doi.org/10.1021/acs.cgd.7b01663.
(128) Altintas, C.; Avci, G.; Daglar, H.; Nemati Vesali Azar, A.; Erucar, I.; Velioglu, S.; Keskin, S. An Extensive Comparative Analysis of Two MOF Databases: High-Throughput Screening of Computation-Ready MOFs for CH_4 and H_2 Adsorption. *J. Mater. Chem. A* **2019**, *7* (16), 9593–9608. https://doi.org/10.1039/C9TA01378D.
(129) Chen, T.; Manz, T. A. A Collection of Forcefield Precursors for Metal-Organic Frameworks. *RSC Adv.*

2019, *9* (63), 36492–36507. https://doi.org/10.1039/c9ra07327b.
(130) Ahmed, A.; Siegel, D. J. Machine Learning Models for Predicting Hydrogen Storage in Metal-Organic Frameworks. Figshare. Software. Https://Doi.Org/10.6084/M9.Figshare.14173520.V1. 2021. https://doi.org/10.6084/m9.figshare.14173520.
(131) Pinheiro, M.; Martin, R. L.; Rycroft, C. H.; Jones, A.; Iglesia, E.; Haranczyk, M. Characterization and Comparison of Pore Landscapes in Crystalline Porous Materials. *J. Mol. Graph. Model.* **2013**, *44*, 208–219. https://doi.org/10.1016/j.jmgm.2013.05.007.
(132) Pinheiro, M.; Martin, R. L.; Rycroft, C. H.; Haranczyk, M. High Accuracy Geometric Analysis of Crystalline Porous Materials. *CrystEngComm* **2013**, *15* (37), 7531–7538. https://doi.org/10.1039/c3ce41057a.
(133) Ongari, D.; Boyd, P. G.; Barthel, S.; Witman, M.; Haranczyk, M.; Smit, B. Accurate Characterization of the Pore Volume in Microporous Crystalline Materials. **2017**. https://doi.org/10.1021/acs.langmuir.7b01682.
(134) Sarkisov, L.; Bueno-Perez, R.; Sutharson, M.; Fairen-jimenez, D. Material Informatics with PoreBlazer v4.0 and CSD MOF Database. **2020**. https://doi.org/10.26434/CHEMRXIV.12923558.V1.
(135) Chen, Z.; Li, P.; Anderson, R.; Wang, X.; Zhang, X.; Robison, L.; Redfern, L. R.; Moribe, S.; Islamoglu, T.; Gómez-Gualdrón, D. A.; et al. Balancing Volumetric and Gravimetric Uptake in Highly Porous Materials for Clean Energy. *Science (80-.).* **2020**, *368* (6488), 297–303. https://doi.org/10.1126/science.aaz8881.
(136) Camp, J. S.; Stavila, V.; Allendorf, M. D.; Prendergast, D.; Haranczyk, M. Critical Factors in Computational Characterization of Hydrogen Storage in Metal-Organic Frameworks Critical Factors in Computational Characterization of Hydrogen Storage in Metal-Organic Frameworks. **2018**. https://doi.org/10.1021/acs.jpcc.8b04021.
(137) Churchard, A. J.; Banach, E.; Borgschulte, A.; Caputo, R.; Chen, J. C.; Clary, D.; Fijalkowski, K. J.; Geerlings, H.; Genova, R. V.; Grochala, W.; et al. A Multifaceted Approach to Hydrogen Storage. *Physical Chemistry Chemical Physics*. Royal Society of Chemistry October 14, 2011, pp 16955–16972. https://doi.org/10.1039/c1cp22312g.
(138) MacRae, C. F.; Sovago, I.; Cottrell, S. J.; Galek, P. T. A.; McCabe, P.; Pidcock, E.; Platings, M.; Shields, G. P.; Stevens, J. S.; Towler, M.; et al. Mercury 4.0: From Visualization to Analysis, Design and Prediction. *J. Appl. Crystallogr.* **2020**, *53* (1), 226–235. https://doi.org/10.1107/S1600576719014092.
(139) Manos, M. J.; Markoulides, M. S.; Malliakas, C. D.; Papaefstathiou, G. S.; Chronakis, N.; Kanatzidis, M. G.; Trikalitis, P. N.; Tasiopoulos, A. J. A Highly Porous Interpenetrated Metal-Organic Framework from the Use of a Novel Nanosized Organic Linker. *Inorg. Chem.* **2011**, *50* (22), 11297–11299. https://doi.org/10.1021/ic201919q.
(140) Furukawa, H.; Ko, N.; Go, Y. B.; Aratani, N.; Choi, S. B.; Choi, E.; Yazaydin, A. Ö.; Snurr, R. Q.; O'Keeffe, M.; Kim, J.; et al. Ultrahigh Porosity in Metal-Organic Frameworks. *Science (80-.).* **2010**, *329* (5990), 424–428. https://doi.org/10.1126/science.1192160.
(141) Yuan, D.; Zhao, D.; Sun, D.; Zhou, H.-C. An Isoreticular Series of Metal-Organic Frameworks with Dendritic Hexacarboxylate Ligands and Exceptionally High Gas-Uptake Capacity. *Angew. Chemie Int. Ed.* **2010**, *49* (31), 5357–5361. https://doi.org/10.1002/anie.201001009.
(142) Yan, Y.; Telepeni, I.; Yang, S.; Lin, X.; Kockelmann, W.; Dailly, A.; Blake, A. J.; Lewis, W.; Walker, G. S.; Allan, D. R.; et al. Metal-Organic Polyhedral Frameworks: High H2 Adsorption Capacities and Neutron Powder Diffraction Studies. *J. Am. Chem. Soc.* **2010**, *132* (12), 4092–4094. https://doi.org/10.1021/ja1001407.
(143) Karagiaridi, O.; Bury, W.; Tylianakis, E.; Sarjeant, A. A.; Hupp, J. T.; Farha, O. K. Opening Metal-Organic Frameworks Vol. 2: Inserting Longer Pillars into Pillared-Paddlewheel Structures through Solvent-Assisted Linker Exchange. *Chem. Mater.* **2013**, *25* (17), 3499–3503. https://doi.org/10.1021/cm401724v.

(144) Zheng, X.; Huang, Y.; Duan, J.; Wang, C.; Wen, L.; Zhao, J.; Li, D. A Microporous Zn(II)-MOF with Open Metal Sites: Structure and Selective Adsorption Properties. *Dalt. Trans.* **2014**, *43* (22), 8311–8317. https://doi.org/10.1039/c4dt00307a.
(145) He, T.; Pachfule, P.; Wu, H.; Xu, Q.; Chen, P. Hydrogen Carriers. *Nat. Rev. Mater.* **2016**, *1*. https://doi.org/10.1038/natrevmats.2016.59.
(146) Li, G.; Kobayashi, H.; Taylor, J. M.; Ikeda, R.; Kubota, Y.; Kato, K.; Takata, M.; Yamamoto, T.; Toh, S.; Matsumura, S.; et al. Hydrogen Storage in Pd Nanocrystals Covered with a Metal–Organic Framework. *Nat. Mater. 2014 138* **2014**, *13* (8), 802–806. https://doi.org/10.1038/nmat4030.
(147) Barthelemy, H.; Weber, M.; Barbier, F. Hydrogen Storage: Recent Improvements and Industrial Perspectives. *Int. J. Hydrogen Energy* **2017**, *42* (11), 7254–7262. https://doi.org/10.1016/J.IJHYDENE.2016.03.178.
(148) Zhang, X.; Leng, Z.; Gao, M.; Hu, J.; Du, F.; Yao, J.; Pan, H.; Liu, Y. Enhanced Hydrogen Storage Properties of MgH2 Catalyzed with Carbon-Supported Nanocrystalline TiO2. *J. Power Sources* **2018**, *398*, 183–192. https://doi.org/10.1016/J.JPOWSOUR.2018.07.072.
(149) Veenstra, M. Yang, J.; Siegel, D. J.; Ming Y. Ford/BASF-SE/UM Activities in Support of the Hydrogen Storage Engineering Center of Excellence. Https://Www.Hydrogen.Energy.Gov/Pdfs/Progress13/Iv_b_7_veenstra_2013.Pdf (Accessed 2020-09-03), United States Dep.
(150) Rosi, N. L.; Eckert, J.; Eddaoudi, M.; Vodak, D. T.; Kim, J.; O'Keeffe, M.; Yaghi, O. M. Hydrogen Storage in Microporous Metal-Organic Frameworks. *Science (80-.).* **2003**, *300* (5622), 1127–1129.
(151) Song, C.; Wang, P.; Makse, H. A. A Phase Diagram for Jammed Matter. *Nat. 2008 4537195* **2008**, *453* (7195), 629–632. https://doi.org/10.1038/nature06981.
(152) Jaoshvili, A.; Esakia, A.; Porrati, M.; Chaikin, P. M. Experiments on the Random Packing of Tetrahedral Dice. *Phys. Rev. Lett.* **2010**, *104* (18), 185501. https://doi.org/10.1103/PHYSREVLETT.104.185501/FIGURES/4/MEDIUM.
(153) Dutta, A.; Koh, K.; Wong-Foy, A. G.; Matzger, A. J. Porous Solids Arising from Synergistic and Competing Modes of Assembly: Combining Coordination Chemistry and Covalent Bond Formation. *Angew. Chemie Int. Ed.* **2015**, *54* (13), 3983–3987. https://doi.org/10.1002/anie.201411735.
(154) Cai, J.; Rao, X.; He, Y.; Yu, J.; Wu, C.; Zhou, W.; Yildirim, T.; Chen, B.; Qian, G. A Highly Porous NbO Type Metal–Organic Framework Constructed from an Expanded Tetracarboxylate. *Chem. Commun.* **2014**, *50* (13), 1552–1554. https://doi.org/10.1039/C3CC48747D.
(155) Schnobrich, J. K.; Lebel, O.; Cychosz, K. A.; Dailly, A.; Wong-Foy, A. G.; Matzger, A. J. Linker-Directed Vertex Desymmetrization for the Production of Coordination Polymers with High Porosity. *J. Am. Chem. Soc.* **2010**, *132* (39), 13941–13948. https://doi.org/10.1021/ja107423k.
(156) 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* (6759), 276–279. https://doi.org/10.1038/46248.
(157) DeSantis, D.; Mason, J. A.; James, B. D.; Houchins, C.; Long, J. R.; Veenstra, M. Techno-Economic Analysis of Metal-Organic Frameworks for Hydrogen and Natural Gas Storage. *Energy and Fuels* **2017**, *31* (2), 2024–2032. https://doi.org/10.1021/acs.energyfuels.6b02510.
(158) Rowsell, J. L. C.; Yaghi, O. M. Strategies for Hydrogen Storage in Metal-Organic Frameworks. *Angew. Chemie Int. Ed.* **2005**, *44* (30), 4670–4679. https://doi.org/10.1002/anie.200462786.
(159) Koh, K.; Van Oosterhout, J. D.; Roy, S.; Wong-Foy, A. G.; Matzger, A. J. Exceptional Surface Area from Coordination Copolymers Derived from Two Linear Linkers of Differing Lengths. *Chem. Sci.* **2012**, *3* (8), 2429. https://doi.org/10.1039/c2sc20407j.
(160) Park, T. H.; Hickman, A. J.; Koh, K.; Martin, S.; Wong-Foy, A. G.; Sanford, M. S.; Matzger, A. J. Highly Dispersed Palladium(II) in a Defective Metal-Organic Framework: Application to C-H Activation and Functionalization. *J. Am. Chem. Soc.* **2011**, *133* (50), 20138–20141.

https://doi.org/10.1021/JA2094316/SUPPL_FILE/JA2094316_SI_001.PDF.
(161) Koh, K.; Wong-Foy, A. G.; Matzger, A. J. A Crystalline Mesoporous Coordination Copolymer with High Microporosity. *Angew. Chemie Int. Ed.* **2008**, *47* (4), 677–680. https://doi.org/10.1002/ANIE.200705020.
(162) Chae, H. K.; Siberio-Pérez, D. Y.; Kim, J.; Go, Y.; Eddaoudi, M.; Matzger, A. J.; O'Keeffe, M.; Yaghi, O. M. A Route to High Surface Area, Porosity and Inclusion of Large Molecules in Crystals. *Nature* **2004**, *427* (6974), 523–527. https://doi.org/10.1038/nature02311.
(163) Pang, M.; Cairns, A. J.; Liu, Y.; Belmabkhout, Y.; Zeng, H. C.; Eddaoudi, M. Highly Monodisperse M III-Based Soc -MOFs (M = in and Ga) with Cubic and Truncated Cubic Morphologies. *J. Am. Chem. Soc.* **2012**, *134* (32), 13176–13179. https://doi.org/10.1021/JA3049282/SUPPL_FILE/JA3049282_SI_001.PDF.
(164) Umemura, A.; Diring, S.; Furukawa, S.; Uehara, H.; Tsuruoka, T.; Kitagawa, S. Morphology Design of Porous Coordination Polymer Crystals by Coordination Modulation. *J. Am. Chem. Soc.* **2011**, *133* (39), 15506–15513. https://doi.org/10.1021/JA204233Q/SUPPL_FILE/JA204233Q_SI_001.PDF.
(165) The Choice Is Somewhat Arbitrary. There Is No Theory to Guide Optimal Selection of a Randomly Packed Bimodal Distribution of Spheres Much Less Guidance for Cuboid Particles. However, the Use of a Relatively Small Mass Frac-Tion of the Smaller Radius Parti.
(166) Mills, L. A.; Sinka, I. C. Effect of Particle Size and Density on the Die Fill of Powders. *Eur. J. Pharm. Biopharm.* **2013**, *84* (3), 642–652. https://doi.org/10.1016/J.EJPB.2013.01.012.
(167) Zakhvatayeva, A.; Zhong, W.; Makroo, H. A.; Hare, C.; Wu, C. Y. An Experimental Study of Die Filling of Pharmaceutical Powders Using a Rotary Die Filling System. *Int. J. Pharm.* **2018**, *553* (1–2), 84–96. https://doi.org/10.1016/J.IJPHARM.2018.09.067.
(168) Lock, N.; Wu, Y.; Christensen, M.; Cameron, L. J.; Peterson, V. K.; Bridgeman, A. J.; Kepert, C. J.; Iversen, B. B. Elucidating Negative Thermal Expansion in MOF-5. *J. Phys. Chem. C* **2010**, *114* (39), 16181–16186. https://doi.org/10.1021/JP103212Z/SUPPL_FILE/JP103212Z_SI_002.ZIP.
(169) Purewal, J.; Veenstra, M.; Tamburello, D.; Ahmed, A.; Matzger, A. J.; Wong-Foy, A. G.; Seth, S.; Liu, Y.; Siegel, D. J. Estimation of System-Level Hydrogen Storage for Metal-Organic Frameworks with High Volumetric Storage Density, in Press 2019) DOI: 10.1016/j.Ijhydene.2019.04.082. *Int. J. Hydrogen Energy* **2019**. https://doi.org/10.1016/J.IJHYDENE.2019.04.082.

Appendix A
Additional information: Materials Development

Alauddin Ahmed[†] and Donald J. Siegel[†,§,Δ,ß,*]

[†]Mechanical Engineering Department, [§]Materials Science & Engineering, [Δ]Applied Physics Program, and [ß]University of Michigan Energy Institute, University of Michigan, Ann Arbor, MI 48109, United States

Section S1. Details of MOF database reported earlier.[1]

The database is publicly available at the HyMARC Data Hub[2]

Table S1. Database of MOF crystal structures, calculated crystallographic properties, and calculated usable H_2 capacities.

Source[1]	Available in database	Zero accessible surface area	H_2 capacity evaluated empirically	H_2 capacity evaluated with GCMC
UM+CoRE+CSD17	15,235	2,950	12,285	12,799
Mail-Order MOFs	112	4	108	112
In Silico MOFs	2,816	154	2,662	466
In Silico Surface MOFs	8, 885	283	8,602	1,058
MOF-74 Analogs	61	0	61	61
ToBaCCo	13,512	214	13,298	2,854
Zr-MOFs	204	0	204	204
NW Hypothetical MOFs	137,000	30,160	106,840	20,156
UO Hypothetical MOFs	315,615	32,993	291,507	61,247
In-house synthesized via hypothetical design	18	0	18	5
Total	**493,458**	**66,758**	**426,700**	**98,962**

Section S2 Literature review of machine learning for gas storage in MOFs[3–13]

Table S2. Summary of recent studies that use machine learning (ML) to predict gas adsorption in MOFs. ρ_{crys}, vf, gsa, vsa, pv, mpd, lcd, pld represent single crystal density, void fraction, gravimetric surface area, volumetric surface area, pore volume, maximum pore diameter, largest cavity diameter, and pore limiting diameter, respectively. R^2, AUE, and RMSE represent the coefficient of determination, Average Unsigned Error, and Root-Mean-Square Error, respectively. AUC = Area Under the Curve. LASSO: Least Absolute Shrinkage and Selection Operator; MLR: Multi-Linear Regression; SVM: Support Vector Machine; DT: Decision Tress; RF: Random Forest; NN: Nearest Neighbors; GBM: Gradient Boosting Method; RBF: Radial Bias Function; PCA: Principal Component Analysis; ANN: Artificial Neural Network.

Study	Gas	ML Features	ML Method	Properties Predicted	Accuracy
This work	H_2	ρ_{crys}, gsa, vsa, vf, pv, lcd, pld	Extremely Randomized Trees	Deliverable H_2 storage capacity between 5-100 bar at 77 K.	UG at PS: R^2 = 0.997; AUE = 0.14 wt. %; RMSE = 0.18 wt. % UV at PS: R^2 = 0.984; AUE = 0.97 g-H_2 L^{-1}; RMSE = 1.40 g-H_2 L^{-1} UG at TPS: R^2 = 0.997; AUE = 0.16 wt. %; RMSE = 0.23 wt. % UV at TPS: R^2 = 0.967; AUE = 1.32 g-H_2 L^{-1}; RMSE = 1.92 g-H_2 L^{-1}
Anderson et al. (2019)[5]	H_2	Epsilon, temperature, pressure, ρ_{crys}, vf, vsa, mpd, lcd, alchemical catecholate site density, unit cell volume.	Neural network	Total volumetric H_2 for pressures 0.1, 1, 5, 35, 65, and 100 bar at 77, 160, and 295 K	AUE = 0.75 - 2.93 g-H_2 L^{-1}
Bucior et al. (2019)[2]	H_2, CH_4	Energetics of MOF-guest interactions	Multilinear regression with LASSO	H_2: Deliverable capacity 2 and 100 bar at 77 K. CH_4: Deliverable capacity between 5.8 and 65 bar at 298 K	R^2 = 0.96, AUE = 1.4 - 3.4 g/L, RMSE = 3.1 - 4.4 g/L
Anderson et al. (2018)[3]	CO_2	ρ_{crys}, vf, gsa, vsa, mpd, lcd, topology	MLR, SVM, DT, RF, NN, GBM	CO_2 capture	R^2 = 0.601- 0.934
Pardakhti et al (2017)[6]	CH_4	ρ_{crys}, vf, gsa, vsa, mpd, lcd interpenetration capacity, number of interpenetration framework, 19 chemical descriptors	DT, Poisson regression, SVM, and RF	Total at 35 bar and 298 K	R^2 = 0.97
Aghaji et al. (2016)[5]	CO_2, CO_2/CH_4	vf, gsa, lcd	DT, SVM(RBF),	Working capacity for the pressure swing between 1 and 10 atm at 298 K	AUC = 0.889 to 0.953
Fernandez & Barnard (2016)[6]	CO_2, N_2	ρ_{crys}, vf, gsa, vsa, mpd, lcd	PCA, k-means clustering, archetypal analysis, DT, SVM, MLL, ANN, RF	Total at 0.1 and 0.9 bar at 298 K	~94%
Ohno & Mukae (2016)[9]	CH_4	ρ_{crys}, vf, gsa, vsa, mpd, and lcd	GP regression, SVM regression, NN, and LR	Total at 35 bar and 298K.	R^2 = 0.79
Simon e al. (2015)[8]	Xe/Kr	ρ_{crys}, vf, vsa, mpd, dpd, surface density, Voronoi energy	RF	Xe/Kr selectivity	RMSE = 2.21 for 15,000 unitless numbers between 0 and 35 R^2 not Reported
Sezginel et al. (2015)[11]	CH_4	ρ_{crys}, vf, gsa, vsa, mpd, and lcd, pld, Q_{st}	MVL regression	Total at 298 K and pressures in 1 to 65 bar	R^2 =0.3 - 0.9
Fernandez et al. (2014)[10]	CO_2	AP-RDF	SVM classification	Total at P =0.15 & 1 bar at 298 K	94.5% (classification)
Fernandez et al. (2013)[11]	CH_4, CO_2, N_2	AP-RDF	PCA, MLR, and SVM regression	Total at low pressure (0.1-0.9 bar) at 298 K	~70% - ~83%
Fernandez et al. (2013)[12]	CH_4	ρ_{crys}, vf, gsa, vsa, mpd, lcd	DT, MLR, and SVM regression	Uptake at 1, 35, and 100 bar at 298 K	~90% at 1 bar (classification); R^2 (regression) = 0.85 (35bar); R^2 (regression) = 0.93 (100 bar)

Section S3 Grand Canonical Monte Carlo (GCMC) calculations

The pseudo-Feynman-Hibbs interatomic potential parameters of Fischer et al.[14–16] were used to model H_2 molecules. MOF-H_2 interactions were calculated using Lorentz-Berthelot[17,18] combination rules. MOFs were assumed to be rigid and were described using interatomic potential parameters from a generic[19,20] force field. The RASPA package was used to evaluate H_2 uptake via Grand Canonical Monte Carlo (GCMC). All calculations were carried out using a 12 Å cut-off radius with compensating long-range corrections.[21,22] GCMC calculations for a given T,P condition were performed using 1000 initial cycles followed by a 1000 cycle production run. Each cycle consisted of translation, insertion, and deletion moves with equal probabilities.[23] Further details can be found in our recent publication.[1]

Section S4 Benchmarking with experimental data[1,24]

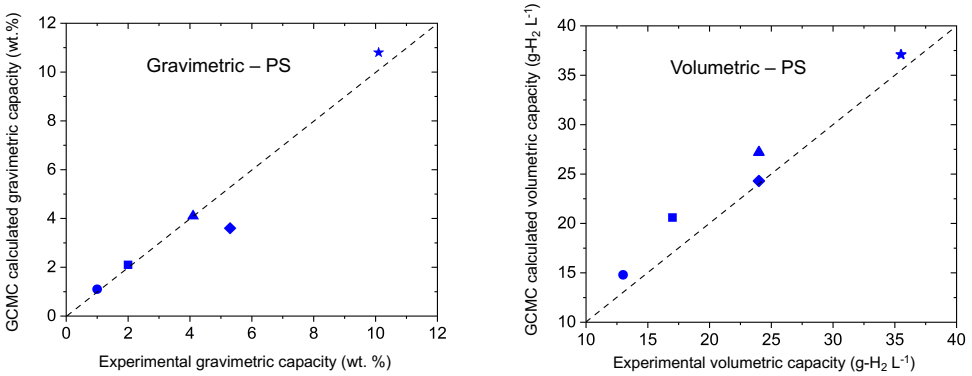


Figure S1. Comparison between experiments and GCMC calculations of H_2 capacities for a benchmark set of open-metal-site MOFs for pressure swing operation: HKUST-1 (■), NOTT-112 (◆), Cu-MOF-74 (●), NU-125 (▲), NU-100/PCN-610 (★).[1,24]

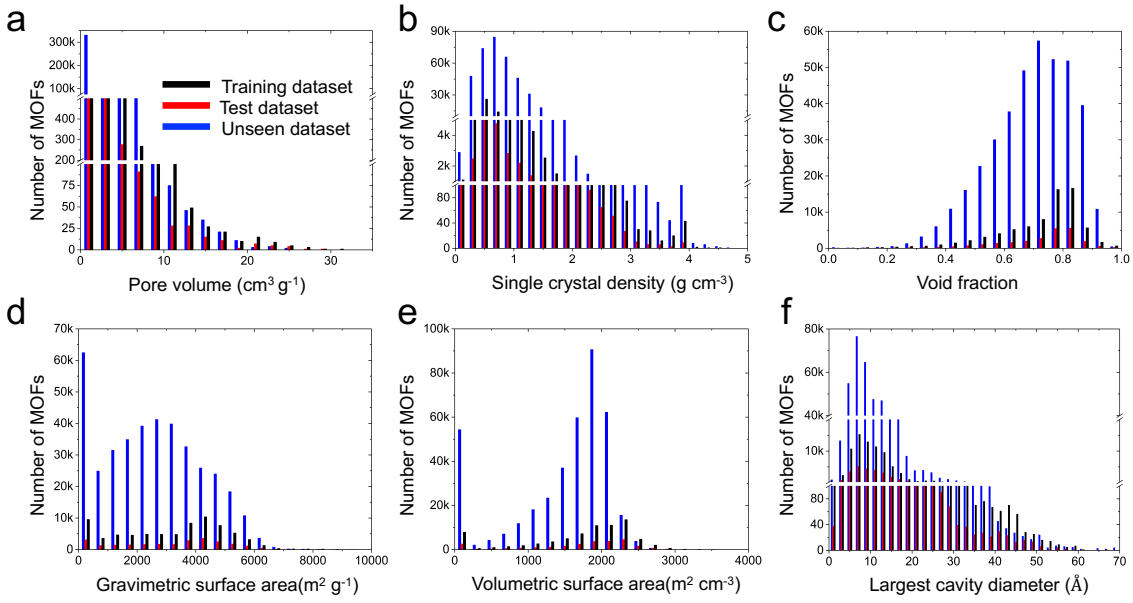


Figure S2. Distribution of 6 crystallographic features in 3 different datasets used in this study. (a) pore volume, (b) single crystal density, (c) void fraction, (d) gravimetric surface area, (e) volumetric surface area, and (f) largest cavity diameter.

Section S5 Description of crystallographic features

Skew and kurtosis were calculated using the scipy.stats module in the SciPy package.[25–27] Skewness is calculated from the ratio of the third moment (m_3) and the cube of the square root of second moment (m_2) of a feature variable, $skew = \mu_3/\mu_2^{3/2}$, where $\mu_i = \left(\sum_{k=1}^{n_{samples}}(x[k] - \bar{x})^i\right)/n_{samples}$ is the *i*-th central moment, and $\bar{x}$ is the mean of the feature variable.[25–27] Kurtosis is the fourth central moment divided by the square of the second moment: $kurtosis = \mu_4/\mu_2^2$.[25–28]

Table S3. H_2 storage capacities for a benchmark set of open metal site (OMS) MOFs. Calculated capacities were predicted using the pseudo-Feynman-Hibbs interatomic potential. Measured H_2 storage data was compiled from García-Holley et al.[24] and from earlier work performed by the present authors.[1] 'Expt.' refers to measured capacities from the literature, 'GCMC' refers to predictions from the present study.

CSD Ref-code	Common name	OMS density Å^{-3}	Usable gravimetric capacity PS conditions (wt. %)		Usable volumetric capacity PS conditions (g-H_2 L^{-1})	
			Expt.[1,23]	GCMC	Expt.[1,23]	GCMC
FQIQCEN	HKUST-1	2.63×10^{-3}	2.0	2.1	17	20.6
FOPFAS	NOTT-112	9.24×10^{-4}	5.3	3.6	24	24.3
LENKIA	Cu-MOF-74	4.91×10^{-3}	1.0	1.1	13	14.8
REWNEO	NU-125	1.09×10^{-3}	4.1	4.1	24	27.2
HABQUY /GAGZEV	NU-100/ PCN-610	4.47×10^{-4}	10.1	10.8	35.5	37.1

Table S4. Statistics for the datasets used in this study.

Feature	Dataset type	Minimum	Maximum	Mean	Median	% zero values	Skew	Kurtosis
d ($g\ cm^{-3}$)	Training	0.03	5.18	0.76	0.62	0	1.84	5.64
	Test	0.03	3.97	0.76	0.61	0	1.79	4.96
	Unseen	0.04	4.7	0.84	0.76	0	1.37	3.81
gsa ($m^2\ g^{-1}$)	Training	0	9750	3112.01	3516	10	-0.16	-0.80
	Test	0	9701	3137.82	3560	10	-0.16	-0.74
	Unseen	0	9671	2530.47	2529	13	0.16	-0.84
vsa ($m^2\ cm^{-3}$)	Training	0	3995	1696.35	1912	10	-1.03	0.23
	Test	0	3966	1703.42	1918	10	-1.04	0.26
	Unseen	0	3482	1473.48	1736	13	-1.10	0.01
vf	Training	0	0.99	0.71	0.76	0	-1.38	2.19
	Test	0.01	0.99	0.71	0.76	0	-1.37	2.18
	Unseen	0	0.98	0.69	0.71	0	-0.70	0.34
pv ($cm^3\ g^{-1}$)	Training	0	35.73	1.34	1.23	0	6.97	91.45
	Test	0.01	29.82	1.37	1.24	0	7.29	89.60
	Unseen	0	24.76	1.18	0.93	0	3.22	30.16
lcd (Å)	Training	0.4	71.6	10.14	9.2	0	2.45	11.94
	Test	0.4	66.2	10.21	9.3	0	2.49	11.95
	Unseen	0.4	69.9	10.41	9.4	0	1.27	3.61
pld (Å)	Training	0	71.5	7.86	7.5	0	2.81	19.54
	Test	0.1	57.7	7.91	7.6	0	2.84	18.43
	Unseen	0	68	7.45	6.9	0	1.21	5.39

Section S6 Machine learning work-flow

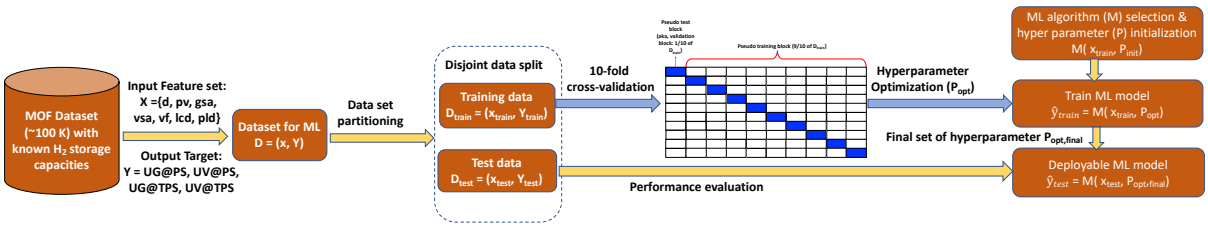


Figure S3. Machine learning work-flow as described in the text.

Section S7 Metrics for ML accuracy

The coefficient of determination (R^2), average unsigned error (AUE), root-mean-squared error (RMSE), and median absolute error (MAE) are used to assess the accuracy of the various ML models with respect to GCMC calculations. If the test/training set contains $n_{samples}$ and $y_{i,gcmc}$ is the GCMC calculated H_2 capacity of i-th sample and $y_{i,ml}$ is the corresponding ML model prediction, then R^2, AUE, RMSE, and MAE are defined as follows:

$$R^2(y_{gcmc}, y_{ml}) = \sqrt{\frac{\sum_{i=1}^{n_{samples}} (y_{i,gcmc} - y_{i,ml})^2}{\sum_{i=1}^{n_{samples}} (y_{i,gcmc} - \overline{y_{gcmc}})^2}}, \quad (1)$$

$$\mathrm{AUE}(y_{gcmc}, y_{ml}) = \frac{\sum_{i=0}^{n_{samples}-1} |y_{i,gcmc} - y_{i,ml}|}{n_{samples}} \quad (2)$$

$$\mathrm{RMSE}(y_{gcmc}, y_{ml}) = \sqrt{\frac{\sum_{i=0}^{n_{samples}-1} (y_{i,gcmc} - y_{i,ml})^2}{n_{samples}}}, \quad (3)$$

$$\mathrm{MAE}(y_{gcmc}, y_{ml}) = median(|y_{1,gcmc} - y_{1,ml}|, \ldots, |y_{n,gcmc} - y_{n,ml}|) \quad (4)$$

where. $\overline{y_{gcmc}} = (\sum_{i=1}^{n_{samples}} y_{i,gcmc})/n_{samples}$.

Kendal τ rank correlation coefficients were calculated using the scipy.stats module[25–27] according to the definition of Kendall τ-b.[29–31]

Section S8 Training set sizes

Table S5. Training set sizes.

100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200,300,400,500,600,700,800,900,1000, 1200, 1300,1400,1500,1600,1700, 1800,1900,2000,3000,4000,5000,6000,7000,8000,9000,10000,11000,12000,13000,14000,15000,16000,17000,18000,19000, 20000,21000,22000,23000,24000,25000,26000,27000,28000,29000,30000,31000,32000,33000,34000,35000,36000,37000, 38000,39000,40000,41000,42000,43000,44000,45000,46000,47000,48000,49000,50000,51000,52000,53000,54000,55000, 56000,57000,58000,59000,60000,61000,62000,63000,64000,65000,66000,67000,68000,69000,70000,71000,72000,73000,74000

Section S9 Performance comparison for ML algorithms

Table S6. Performance of ML models in predicting usable gravimetric capacities under pressure swing conditions. R^2, AUE, RSME, and MAE represent the coefficient of determination, average unsigned error, root-mean-squared error, and median absolute error, respectively.

ML model	Model abbreviation	Feature scaling method	R^2	AUE (wt. %)	RMSE (wt. %)	Kendal τ	EV	MAE
Ada Boost	AB	unscaled	0.975	0.476	0.332	0.910	0.976	0.410
Bagging with Decision Tree	B/DT	unscaled	0.997	0.141	0.037	0.959	0.997	0.110
Bagging with Random Forest	B/RF	unscaled	0.997	0.141	0.037	0.959	0.997	0.110
Boosted Decision Trees	BDT	unscaled	0.997	0.136	0.037	0.963	0.997	0.100
Decision Trees	DT	unscaled	0.995	0.180	0.065	0.949	0.995	0.100
Extremely Randomized Trees	ERT	unscaled	0.997	0.136	0.034	0.961	0.997	0.104
Gradient Boosting	GB	unscaled	0.997	0.158	0.045	0.955	0.997	0.123
K-Nearest Neighbors	K-NN	unscaled	0.983	0.346	0.226	0.900	0.983	0.260
Linear Regression	LR	unscaled	0.987	0.307	0.170	0.915	0.987	0.241
Nu-Support Vector Machine with Radial Basis Function (RBF) Kernel	Nu-SVM/RBF-K	minmax scale	0.986	0.235	0.187	0.958	0.987	0.173
Random Forest	RF	unscaled	0.997	0.141	0.037	0.959	0.997	0.110
Ridge Regression	RR	unscaled	0.987	0.307	0.170	0.915	0.987	0.241
Support Vector Machine Radial Basis Function (RBF) Kernel	SVM/RBF-K	minmax scale	0.986	0.236	0.187	0.958	0.987	0.174
Support Vector Machine with Linear Kernel	SVM/L-K	minmax scale	0.986	0.306	0.187	0.920	0.986	0.224

Table S7. Performance of ML models in predicting usable volumetric capacities under pressure swing condition. R^2, AUE, RSME, and MAE represent the coefficient of determination, average unsigned error, root-mean-squared error, and median absolute error, respectively.

ML model	Model abbreviation	Feature scaling method	R^2	AUE ($g\text{-}H_2\ L^{-1}$)	RMSE ($g\text{-}H_2\ L^{-1}$)	Kendal τ	EV	MAE
Ada Boost	AB	unscaled	0.936	2.258	7.732	0.873	0.938	1.983
Bagging with Decision Tree	B/DT	unscaled	0.982	1.011	2.133	0.918	0.982	0.720
Bagging with Random Forest	B/RF	unscaled	0.983	0.997	2.048	0.919	0.983	0.710
Boosted Decision Trees	BDT	unscaled	0.983	0.979	2.104	0.922	0.983	0.700
Decision Trees	DT	unscaled	0.971	1.298	3.568	0.895	0.971	0.900
Extremely Randomized Trees	ERT	unscaled	0.984	0.967	1.960	0.922	0.984	0.692
Gradient Boosting	GB	unscaled	0.980	1.104	2.454	0.911	0.980	0.829
K-Nearest Neighbors	K-NN	unscaled	0.913	2.378	10.517	0.794	0.913	1.760
Linear Regression	LR	unscaled	0.917	2.403	10.045	0.829	0.917	1.981
Nu-Support Vector Machine with Radial Basis Function (RBF) Kernel	Nu-SVM/RBF-K	minmax scale	0.949	1.899	6.137	0.858	0.951	1.549
Random Forest	RF	unscaled	0.982	1.011	2.156	0.918	0.982	0.720
Ridge Regression	RR	unscaled	0.917	2.404	10.046	0.829	0.917	1.980
Support Vector Machine Radial Basis Function (RBF) Kernel	SVM/RBF-K	minmax scale	0.951	1.836	5.957	0.863	0.954	1.468
Support Vector Machine with Linear Kernel	SVM/L-K	minmax scale	0.910	2.398	10.905	0.846	0.913	1.902

Table S8. Performance of ML models in predicting usable gravimetric capacities under temperature+pressure swing condition. R^2, AUE, RSME, and MAE represent the coefficient of determination, average unsigned error, root-mean-squared error, and median absolute error, respectively.

ML model	Model abbreviation	Feature scaling method	R^2	AUE (wt. %)	RMSE (wt. %)	Kendal τ	EV	MAE
Ada Boost	AB	unscaled	0.970	0.557	0.497	0.939	0.970	0.459
Bagging with Decision Tree	B/DT	unscaled	0.997	0.172	0.055	0.962	0.997	0.130
Bagging with Random Forest	B/RF	unscaled	0.997	0.171	0.054	0.961	0.997	0.130
Boosted Decision Trees	BDT	unscaled	0.997	0.165	0.051	0.963	0.997	0.127
Decision Trees	DT	unscaled	0.994	0.223	0.095	0.951	0.994	0.200
Extremely Randomized Trees	ERT	unscaled	0.997	0.163	0.053	0.966	0.997	0.100
Gradient Boosting	GB	unscaled	0.996	0.199	0.068	0.956	0.996	0.158
K-Nearest Neighbors	K-NN	unscaled	0.993	0.250	0.117	0.943	0.993	0.200
Linear Regression	LR	unscaled	0.992	0.266	0.131	0.947	0.992	0.208
Nu-Support Vector Machine with Radial Basis Function (RBF) Kernel	Nu-SVM/RBF-K	minmax scale	0.991	0.285	0.155	0.952	0.991	0.217
Random Forest	RF	unscaled	0.997	0.173	0.056	0.961	0.997	0.130
Ridge Regression	RR	unscaled	0.992	0.266	0.131	0.947	0.992	0.208
Support Vector Machine Radial Basis Function (RBF) Kernel	SVM/RBF-K	minmax scale	0.991	0.283	0.155	0.952	0.991	0.215
Support Vector Machine with Linear Kernel	SVM/L-K	minmax scale	0.968	0.451	0.535	0.948	0.973	0.345

Table S9. Performance of ML models in predicting usable volumetric capacities under temperature+pressure swing condition. R^2, AUE, RSME, and MAE represent the coefficient of determination, average unsigned error, root-mean-squared error, and median absolute error, respectively.

ML model	Model abbreviation	Feature scaling method	R^2	AUE (wt. %)	RMSE (wt. %)	Kendal τ	EV	MAE
Ada Boost	AB	unscaled	0.911	2.387	9.954	0.752	0.912	1.877
Bagging with Decision Tree	B/DT	unscaled	0.963	1.381	4.147	0.809	0.963	0.940
Bagging with Random Forest	B/RF	unscaled	0.964	1.380	4.042	0.809	0.964	0.940
Boosted Decision Trees	BDT	unscaled	0.965	1.322	3.887	0.819	0.965	0.900
Decision Trees	DT	unscaled	0.936	1.812	7.150	0.755	0.936	1.200
Extremely Randomized Trees	ERT	unscaled	0.967	1.320	3.700	0.819	0.967	0.912
Gradient Boosting	GB	unscaled	0.955	1.572	4.953	0.785	0.955	1.126
K-Nearest Neighbors	K-NN	unscaled	0.926	2.036	8.202	0.710	0.926	1.460
Linear Regression	LR	unscaled	0.913	2.048	9.691	0.764	0.913	1.329
Nu-Support Vector Machine with Radial Basis Function (RBF) Kernel	Nu-SVM/RBF-K	minmax scale	0.913	2.033	9.656	0.767	0.915	1.310
Random Forest	RF	unscaled	0.963	1.383	4.169	0.809	0.963	0.940
Ridge Regression	RR	unscaled	0.913	2.049	9.692	0.764	0.913	1.331
Support Vector Machine Radial Basis Function (RBF) Kernel	SVM/RBF-K	minmax scale	0.913	2.029	9.641	0.768	0.915	1.307
Support Vector Machine with Linear Kernel	SVM/L-K	minmax scale	0.907	2.117	10.404	0.767	0.911	1.390

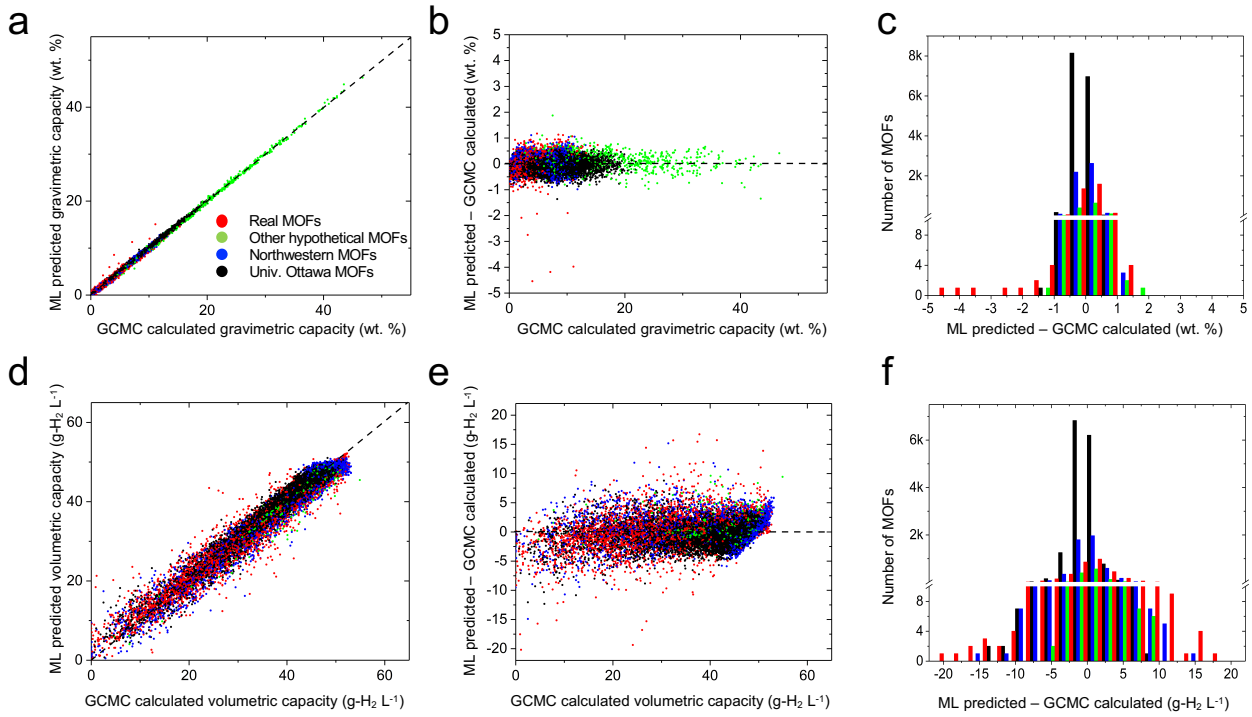


Figure S4. Performance of the Extremely Randomized Trees ML algorithm with respect to GCMC calculations for predicting usable H_2 capacities in MOFs. Data is collected under TPS conditions on a test set of 24,674 MOFs. Different colors represent different categories of MOFs. Top **(a-c)** and bottom **(d-f)** panels illustrate performance for usable gravimetric and volumetric capacities, respectively. **(a, d)**: Agreement between ML and GCMC predictions. **(b, e)**: Difference between ML and GCMC as a function of GCMC capacity. **(c, f)** Distribution of differences in predictions between ML and GCMC.

Section S10 Performance of ML models under TPS conditions

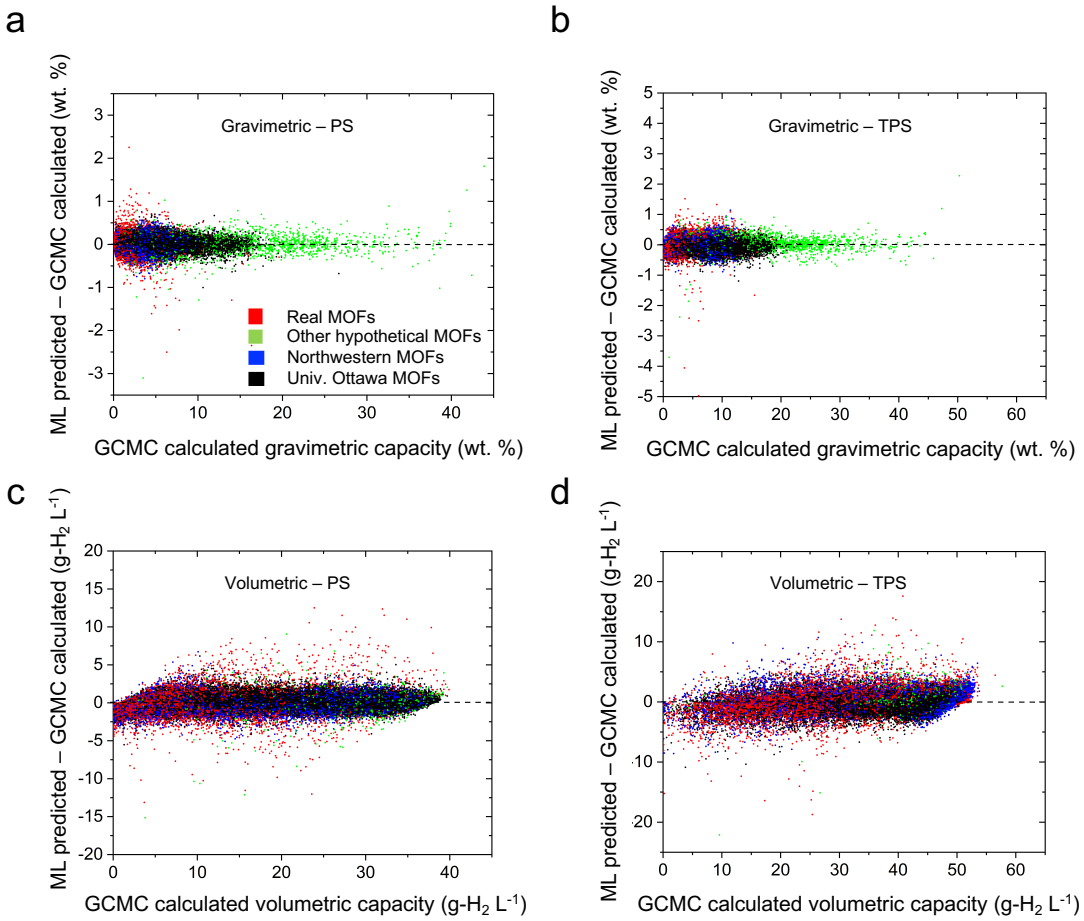


Figure S5. Difference between ML and GCMC as a function of GCMC capacity for the training set of 74,201 MOFs. Performance of the Extremely Randomized Trees ML algorithm with respect to GCMC calculations for predicting usable H_2 capacities in MOFs. Data is collected under PS **(a, c)** and TPS **(b,d).** Different colors represent different categories of MOFs. Top **(a, b)** and bottom **(c,d)** panels illustrate performance for usable gravimetric and volumetric capacities, respectively.

Section S11 Difference between ML and GCMC as a function of GCMC capacity for the training set

Section S12 Effect of training set size on ML model accuracies

Table S10. Parameters of the power-law fit, $\varepsilon(m) = \alpha m^{\beta} + \gamma$, where m is the size of the training dataset and ε represents the metric of accuracy (here average unsigned error or AUE). α, β, and γ are the power-law coefficient, exponent, and constant, respectively.

Condition	β (scaling factor)	α (coefficient)	γ (constant)
UG - PS	-0.43	1.19	0.13
UG - TPS	-0.37	0.92	0.16
UV - PS	-0.23	1.96	0.85
UV - TPS	-0.16	2.10	1.04

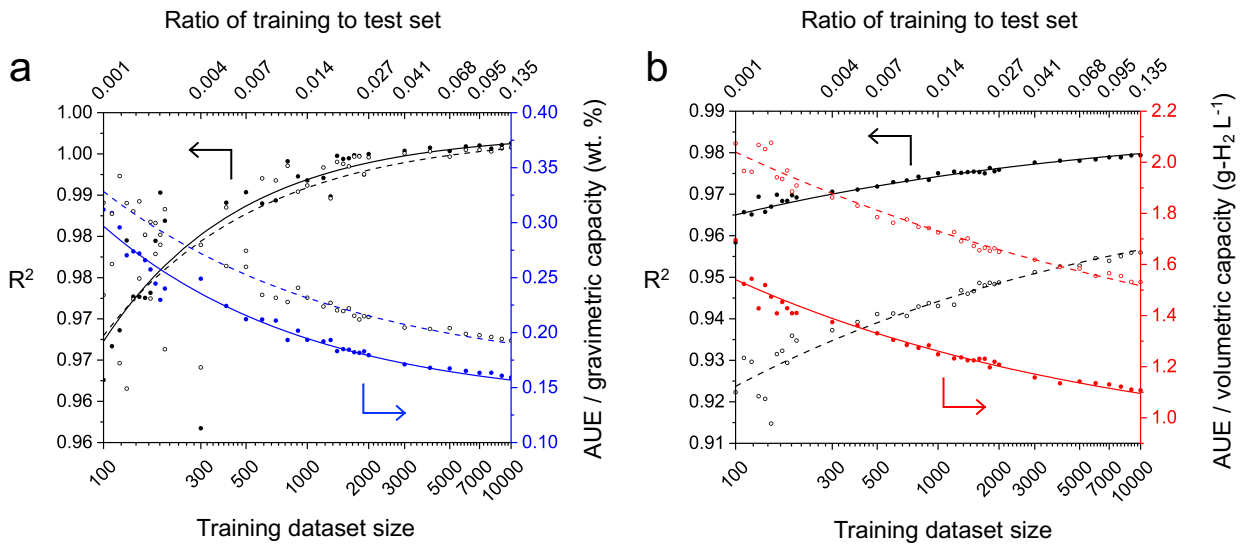


Figure S6. Performance of Extremely Randomized Trees ML models for predicting usable (a) gravimetric and (b) volumetric H_2 capacity as a function of training set size (up to a dataset size of 10,000 MOFs) and the ratio of training to test set size. 100 different training sets ranging in size between 100 and 74,021 MOFs were examined. A common set of 24,674 MOFs was used for testing. Performance is quantified using R^2 (left axis, black) and the average unsigned error, AUE (right axis, blue and red for UG and UV, respectively). Lines represent a power-law fit to the data.

Section 13 Univariate Feature Importance[32,33]

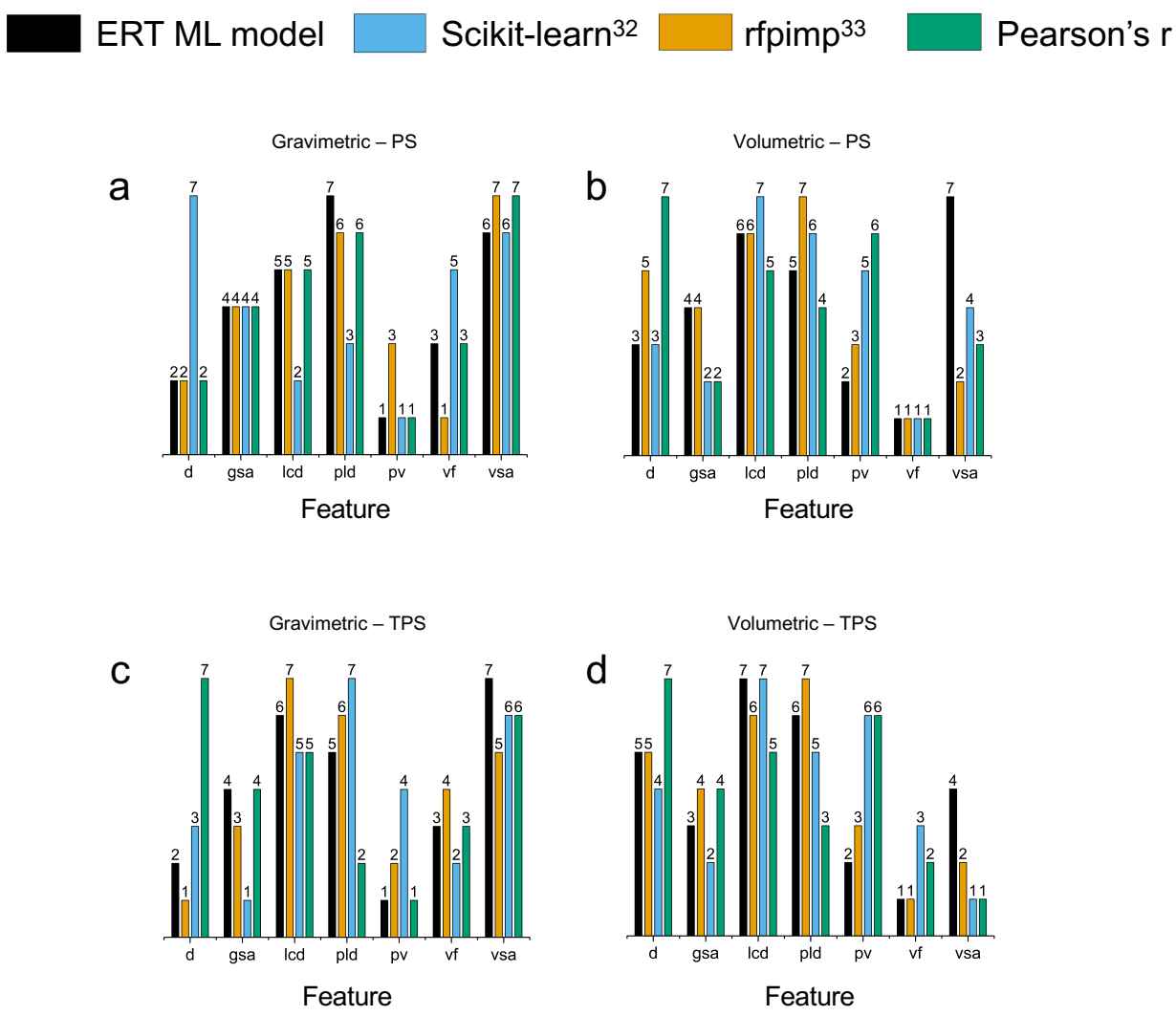


Figure S7. Relative importance of seven features in predicting H_2 storage in MOFs. Features are ranked 1 (most important) through 7 (least important). Four different methods were used: Pearson's correlation coefficient (r), Breiman and Fried-man's tree-based algorithm as implemented in Scikit-learn, and the permutation importance method as implemented in rfpimp package. (a) usable gravimetric and (b) volumetric capacities for PS conditions. (c) usable gravimetric and (d) volumetric capacities for TPS conditions.

Section 14. GCMC verification of ML predictions

Table S11. MOFs predicted by ML to have high capacities under PS condition and whose performance was subsequently verified with GCMC. Here NW and UO represent Northwestern University and University of Ottawa databases.

Name	Source	Density ($g\ cm^{-3}$)	Gravimetric surface area ($m^2\ g^{-1}$)	Volumetric surface area ($m^2\ cm^{-3}$)	Void fraction	Pore volume ($cm^3\ g^{-1}$)	Largest cavity diameter (Å)	Pore limiting diameter (Å)	Usable gravimetric capacity (wt. %)		Usable volumetric capacity ($g\text{-}H_2\ L^{-1}$)	
									GCMC	ML	GCMC	ML
mof_7642	ToBaCCo	0.30	5561	1695	0.89	2.93	12.8	11.8	11.1	10.3	40.5	37.4
mof_7690	ToBaCCo	0.30	5715	1706	0.89	2.98	12.8	12.0	11.3	10.4	40.3	37.3
mof_7594	ToBaCCo	0.40	5070	2031	0.86	2.15	11.2	9.7	8.6	7.9	39.9	37.0
mof_7210	ToBaCCo	0.29	5936	1730	0.89	3.04	13.4	11.7	11.4	10.5	39.8	37.1
mof_7738	ToBaCCo	0.25	6054	1502	0.90	3.64	14.5	13.5	13.0	12.0	39.7	37.0
hypotheticalMOF_5045702_i_1_j_24_k_20_m_2	NW	0.31	5926	1820	0.88	2.87	16.0	11.0	10.9	10.1	39.7	37.2
str_m3_o19_o19_f0_nbo.sym.1.out	UO	0.31	5073	1583	0.90	2.88	17.7	12.9	10.8	10.1	39.7	37.1
hypotheticalMOF_5037315_i_1_j_20_k_12_m_1	NW	0.31	5818	1787	0.88	2.86	16.0	11.0	10.9	10.0	39.7	37.0
hypotheticalMOF_5037467_i_1_j_20_k_12_m_8	NW	0.31	5860	1800	0.88	2.85	16.0	11.0	10.9	10.0	39.7	37.0
str_m3_o5_o20_f0_nbo.sym.1.out	UO	0.39	4772	1882	0.87	2.22	14.1	9.6	8.7	8.1	39.7	37.2
hypothetical MOF_5037563_i_1_j_20_k_12_m_13	NW	0.31	5897	1811	0.88	2.87	16.1	11.0	10.9	10.1	39.7	37.2
hypothetical MOF_5038404_i_1_j_20_k_20_m_15	NW	0.31	5870	1803	0.88	2.87	16.0	11.0	10.9	10.1	39.7	37.2
hypotheticalMOF_5037379_i_1_j_20_k_12_m_4	NW	0.31	5818	1787	0.88	2.86	16.0	11.0	10.9	10.0	39.6	37.0
hypotheticalMOF_5037407_i_1_j_20_k_12_m_5	NW	0.31	5818	1787	0.88	2.86	16.0	11.0	10.9	10.0	39.6	37.0
hypotheticalMOF_5037479_i_1_j_20_k_12_m_9	NW	0.31	5818	1787	0.88	2.86	16.0	11.0	10.9	10.0	39.6	37.0
hypothetical MOF_5055561_i_1_j_28_k_20_m_11	NW	0.31	5874	1804	0.88	2.87	16.0	11.0	10.9	10.1	39.6	37.2
hypotheticalMOF_5037439_i_1_j_20_k_12_m_7	NW	0.31	5858	1799	0.88	2.85	16.0	11.0	10.9	10.0	39.6	37.0
hypothetical MOF_5037499_i_1_j_20_k_12_m_10	NW	0.31	5854	1798	0.88	2.85	16.0	11.0	10.9	10.0	39.6	37.0
hypothetical MOF_5037531_i_1_j_20_k_12_m_11	NW	0.31	5818	1787	0.88	2.86	16.0	11.0	10.9	10.0	39.6	37.0
hypothetical MOF_5037523_i_1_j_20_k_12_m_11	NW	0.31	5857	1799	0.88	2.86	16.0	11.0	10.9	10.0	39.6	37.1

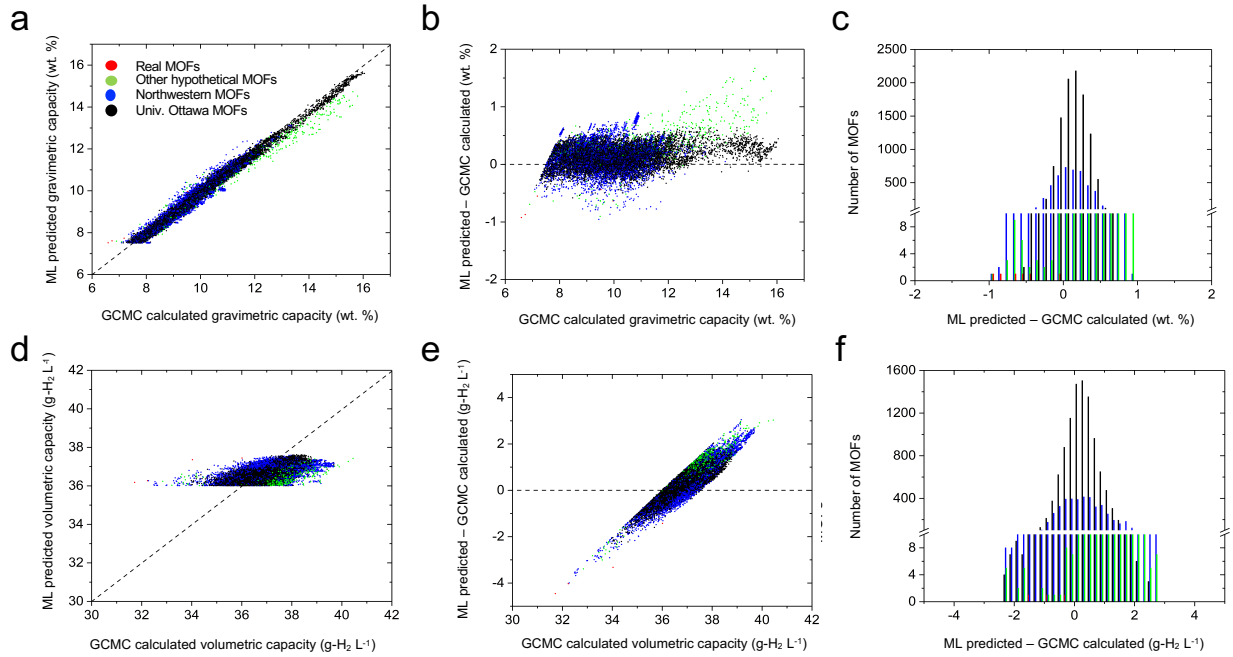


Figure S8. Comparison of GCMC calculations with ML predictions for the 21,700 highest-capacity MOFs predicted by ML for PS conditions. Top (**a**-**c**) and bottom (**d**-**f**) panels illustrate the performance for gravimetric and volumetric capacities, respectively. Left panels (**a**, **d**) show the correlation between GCMC and ML capacities; the diagonal lines indicate perfect correlations. Middle panels (**b**, **e**) show the difference between GCMC and ML, where the horizontal lines represent a zero difference. Right panels (**c**, **f**) show the distribution of differences from plots **b** and **e**.

Table S12. MOFs predicted by ML to have high capacities under TPS condition and whose performance was subsequently verified with GCMC. Here UO represents University of Ottawa database.

Name	Source	Density ($g\,cm^{-3}$)	Gravimetric surface area ($m^2\,g^{-1}$)	Volumetric surface area ($m^2\,cm^{-3}$)	Void fraction	Pore volume ($cm^3\,g^{-1}$)	Largest cavity diameter (Å)	Pore limiting diameter (Å)	Usable gravimetric capacity (wt. %)		Usable volumetric capacity ($g\text{-}H_2\,L^{-1}$)	
									GCMC	ML	GCMC	ML
str_m1_o1_o11_f0_pcu.sym.102.out	UO	0.45	4352	1974	0.84	1.84	12.9	10.1	10.4	9.7	53.1	48.1
str_m1_o1_o11_f0_pcu.sym.117.out	UO	0.47	4162	1977	0.83	1.74	12.8	9.9	9.9	9.0	52.8	48.0
str_m1_o1_o11_f0_pcu.sym.121.out	UO	0.47	4263	2006	0.83	1.76	12.1	10.2	10.0	9.4	52.7	48.1
str_m1_o1_o11_f0_pcu.sym.13.out	UO	0.46	4326	2005	0.83	1.79	12.7	9.9	10.1	9.3	52.6	48.0
str_m1_o1_o11_f0_pcu.sym.159.out	UO	0.58	3703	2138	0.80	1.38	10.4	8.6	8.3	7.6	52.6	48.5
str_m1_o1_o11_f0_pcu.sym.200.out	UO	0.45	4359	1978	0.84	1.84	12.9	10.1	10.3	9.6	52.6	48.1
str_m1_o1_o11_f0_pcu.sym.212.out	UO	0.60	3417	2035	0.83	1.39	12.0	10.1	8.1	7.5	52.5	48.1
str_m1_o1_o11_f0_pcu.sym.51.out	UO	0.46	4330	2007	0.83	1.79	11.9	9.9	10.1	9.3	52.5	48.1
str_m1_o1_o11_f0_pcu.sym.71.out	UO	0.45	4436	1980	0.84	1.87	13.0	10.9	10.4	9.7	52.5	48.1
str_m1_o1_o11_f0_pcu.sym.89.out	UO	0.58	3507	2043	0.83	1.42	12.4	9.8	8.2	7.7	52.5	48.1
str_m1_o1_o17_f0_pcu.sym.1.out	UO	0.46	4283	1985	0.83	1.79	11.9	9.9	10.1	9.4	52.5	48.3
str_m1_o1_o17_f0_pcu.sym.104.out	UO	0.46	4439	2032	0.83	1.82	12.5	11.0	10.2	9.6	52.4	48.2
str_m1_o1_o17_f0_pcu.sym.129.out	UO	0.60	3585	2157	0.83	1.37	14.6	9.2	7.9	7.6	52.3	48.2
str_m1_o1_o17_f0_pcu.sym.132.out	UO	0.60	3438	2048	0.83	1.39	12.7	10.8	8.0	7.8	52.3	48.3
str_m1_o1_o17_f0_pcu.sym.28.out	UO	0.57	3732	2117	0.80	1.41	13.1	10.9	8.4	7.8	52.2	48.1
str_m1_o1_o2_f0_pcu.sym.1.out	UO	0.56	3615	2011	0.83	1.49	13.1	10.8	8.5	7.9	52.2	48.4
str_m1_o1_o2_f0_pcu.sym.101.out	UO	0.56	3549	1978	0.84	1.50	12.9	10.7	8.5	7.7	52.1	48.1
str_m1_o1_o2_f0_pcu.sym.11.out	UO	0.44	4487	1986	0.84	1.89	12.4	10.3	10.4	9.7	52.0	48.2
str_m1_o1_o2_f0_pcu.sym.15.out	UO	0.41	4983	2054	0.84	2.04	12.7	9.1	11.1	10.3	52.0	48.1
str_m1_o1_o2_f0_pcu.sym.2.out	UO	0.47	4179	1977	0.83	1.75	11.9	9.8	9.8	9.0	52.0	48.0
MOF-5									**7.8**		**51.9**	

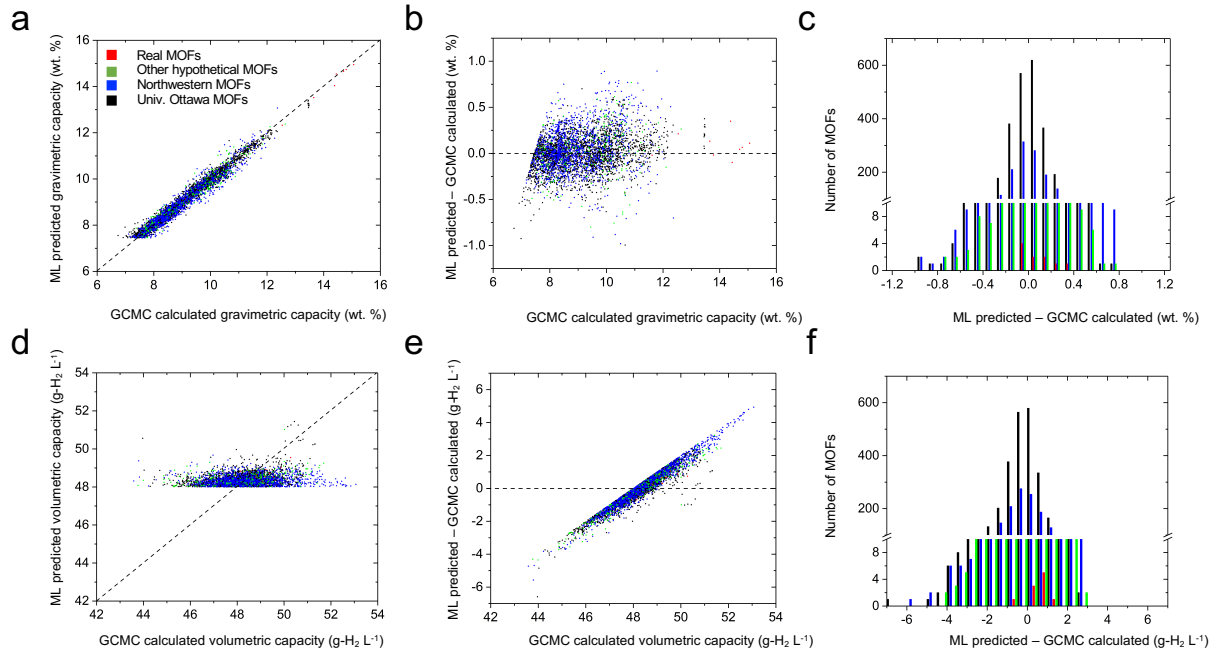


Figure S9. Comparison of GCMC calculations with ML predictions for the 7,901 highest-capacity MOFs predicted by ML for TPS conditions. Top (**a**-**c**) and bottom (**d**-**f**) panels illustrate the performance for gravimetric and volumetric capacities, respectively. Left panels (**a**, **d**) show the correlation between GCMC and ML capacities; the diagonal lines indicate perfect correlations. Middle panels (**b**, **e**) show the difference between GCMC and ML, where the horizontal lines represent a zero difference. Right panels (**c**, **f**) show the distribution of differences from plots **b** and **e**.

Table S13. Differences between ML-predicated and GCMC-calculated H_2 storage capacities of unseen MOFs at PS and TPS conditions. Overprediction and underprediction mean ML predicted values are greater and smaller than those of GCMC calculated actual values, respectively.

Statistics	Pressure swing		Temperature + pressure swing	
	UG (wt. %)	UV (g-H_2 L^{-1})	UG (wt. %)	UV (g-H_2 L^{-1})
Largest overprediction	1.67	3.36	0.94	4.93
Largest underprediction	-0.96	-4.46	-1.0	-6.59
Average unsigned error	0.24	0.66	0.24	1.28
Standard deviation	0.20	0.53	0.17	0.99

Performance for 508 feature combinations

d = Single crystal density (g/cm^3); pv = Pore volume (cm^3/g); vf = void fraction; gsa = Gravimetric surface area (m^2/g); vsa = Volumetric surface area (m^2/cm^3); lcd = Largest cavity diameter (Å); pld = Pore limiting diameter (Å)

Usable gravimetric H_2 capacity (wt.%) under pressure swing (PS) between 100 and 5 bar at 77K					
Combination of features	**Number of features**	**R^2**	**Average Unsigned Error, AUE (wt.%)**	**Root-Mean-Square_Error, RMSE (wt.%)**	**Kendall τ**
pv	1	0.991	0.27	0.35	0.925
d	1	0.981	0.37	0.48	0.891
vf	1	0.971	0.39	0.61	0.910
gsa	1	0.798	0.90	1.52	0.788
pld	1	0.721	1.13	1.82	0.711
lcd	1	0.719	1.32	1.86	0.655
vsa	1	0.073	2.12	3.48	0.314
d, pv	2	0.995	0.19	0.26	0.947
vf, pv	2	0.994	0.19	0.27	0.945
d, vf	2	0.994	0.20	0.27	0.943
pv, lcd	2	0.992	0.24	0.32	0.931
pv, pld	2	0.990	0.26	0.35	0.924
gsa, pv	2	0.988	0.29	0.40	0.914
d, lcd	2	0.987	0.31	0.42	0.908
vsa, pv	2	0.986	0.30	0.41	0.911
d, pld	2	0.982	0.35	0.46	0.894
gsa, vsa	2	0.979	0.38	0.52	0.886
d, gsa	2	0.977	0.41	0.56	0.882
gsa, vf	2	0.976	0.38	0.56	0.900
d, vsa	2	0.975	0.41	0.56	0.878
vf, lcd	2	0.973	0.37	0.58	0.906
vf, pld	2	0.973	0.38	0.61	0.905
vsa, vf	2	0.966	0.43	0.66	0.889
gsa, lcd	2	0.954	0.47	0.76	0.882
gsa, pld	2	0.907	0.69	1.07	0.823
lcd, pld	2	0.838	0.88	1.44	0.777
vsa, lcd	2	0.812	0.97	1.54	0.734
vsa, pld	2	0.803	1.01	1.58	0.731
d, pv, lcd	3	0.996	0.18	0.24	0.949
vf, pv, lcd	3	0.995	0.17	0.24	0.948
vsa, vf, pv	3	0.995	0.19	0.25	0.945
gsa, vsa, pv	3	0.995	0.18	0.26	0.947
d, vf, lcd	3	0.995	0.18	0.25	0.948
d, pv, pld	3	0.995	0.20	0.26	0.944
d, gsa, pv	3	0.995	0.19	0.26	0.944
gsa, vf, pv	3	0.995	0.19	0.26	0.945
d, vf, pv	3	0.995	0.19	0.26	0.946
d, vsa, pv	3	0.995	0.19	0.26	0.944
d, gsa, vf	3	0.994	0.20	0.28	0.942
gsa, vsa, vf	3	0.994	0.19	0.27	0.945
vf, pv, pld	3	0.994	0.20	0.27	0.941
d, vsa, vf	3	0.994	0.20	0.28	0.941
d, vf, pld	3	0.994	0.20	0.27	0.943
vsa, pv, lcd	3	0.993	0.21	0.30	0.936
gsa, pv, lcd	3	0.993	0.22	0.31	0.935
pv, lcd, pld	3	0.993	0.22	0.31	0.932
vsa, pv, pld	3	0.991	0.25	0.34	0.928
gsa, pv, pld	3	0.991	0.25	0.33	0.927
gsa, vsa, lcd	3	0.989	0.27	0.37	0.920
gsa, vf, lcd	3	0.988	0.26	0.38	0.931
d, gsa, lcd	3	0.988	0.28	0.39	0.917
d, vsa, lcd	3	0.987	0.29	0.40	0.916
d, lcd, pld	3	0.987	0.30	0.41	0.909
gsa, vsa, pld	3	0.985	0.32	0.43	0.906
gsa, vf, pld	3	0.985	0.30	0.44	0.916
d, gsa, pld	3	0.985	0.33	0.45	0.903
d, vsa, pld	3	0.984	0.33	0.45	0.900
d, gsa, vsa	3	0.977	0.38	0.52	0.886
gsa, lcd, pld	3	0.975	0.33	0.55	0.916
vsa, vf, lcd	3	0.970	0.40	0.62	0.893
vf, lcd, pld	3	0.969	0.40	0.63	0.893
vsa, vf, pld	3	0.967	0.42	0.65	0.891
vsa, lcd, pld	3	0.927	0.62	0.99	0.835
d, vsa, pv, lcd	4	0.997	0.16	0.21	0.954
gsa, vf, pv, lcd	4	0.997	0.15	0.21	0.955
vsa, vf, pv, lcd	4	0.997	0.15	0.21	0.954
d, gsa, pv, lcd	4	0.997	0.16	0.21	0.955
vsa, vf, pv, pld	4	0.996	0.16	0.22	0.953
gsa, vf, pv, pld	4	0.996	0.16	0.22	0.953
d, pv, lcd, pld	4	0.996	0.16	0.21	0.955
d, gsa, pv, pld	4	0.996	0.17	0.22	0.952
vf, pv, lcd, pld	4	0.996	0.15	0.21	0.955
d, vsa, pv, pld	4	0.996	0.17	0.22	0.950
d, gsa, vf, lcd	4	0.996	0.16	0.22	0.953
gsa, vsa, pv, lcd	4	0.996	0.17	0.24	0.950
d, vf, pv, lcd	4	0.996	0.17	0.23	0.951
gsa, vsa, vf, lcd	4	0.996	0.16	0.24	0.953
d, vsa, vf, lcd	4	0.996	0.16	0.23	0.953
gsa, vsa, vf, pld	4	0.996	0.17	0.23	0.952
gsa, vsa, pv, pld	4	0.995	0.18	0.24	0.949
d, gsa, vf, pv	4	0.995	0.18	0.25	0.946
d, vsa, vf, pld	4	0.995	0.17	0.24	0.951
gsa, vsa, vf, pv	4	0.995	0.18	0.24	0.947
d, vsa, vf, pv	4	0.995	0.18	0.24	0.946
d, gsa, vsa, pv	4	0.995	0.18	0.25	0.947
d, gsa, vf, pld	4	0.995	0.18	0.25	0.949
d, vf, lcd, pld	4	0.995	0.16	0.26	0.954
d, vf, pv, pld	4	0.995	0.19	0.26	0.945
vsa, pv, lcd, pld	4	0.995	0.19	0.27	0.944
d, gsa, vsa, vf	4	0.994	0.19	0.26	0.943
gsa, pv, lcd, pld	4	0.994	0.20	0.27	0.941
gsa, vf, lcd, pld	4	0.992	0.22	0.32	0.943
gsa, vsa, lcd, pld	4	0.992	0.24	0.33	0.929
d, gsa, lcd, pld	4	0.992	0.24	0.33	0.932
d, vsa, lcd, pld	4	0.991	0.24	0.33	0.930
d, gsa, vsa, lcd	4	0.988	0.27	0.38	0.919
d, gsa, vsa, pld	4	0.985	0.32	0.44	0.905
vsa, vf, lcd, pld	4	0.976	0.36	0.56	0.907
gsa, vf, pv, lcd, pld	5	0.997	0.14	0.19	0.959
d, vf, pv, lcd, pld	5	0.997	0.15	0.20	0.956
d, gsa, vf, pv, lcd	5	0.997	0.15	0.20	0.957
d, gsa, pv, lcd, pld	5	0.997	0.15	0.20	0.957
vsa, vf, pv, lcd, pld	5	0.997	0.14	0.20	0.957
d, vsa, vf, pv, lcd	5	0.997	0.15	0.20	0.956
d, gsa, vf, lcd, pld	5	0.996	0.15	0.21	0.957
d, gsa, vsa, pv, lcd	5	0.996	0.16	0.21	0.954
d, vsa, vf, lcd, pld	5	0.996	0.15	0.21	0.956
d, vsa, pv, lcd, pld	5	0.996	0.15	0.21	0.956
gsa, vsa, vf, lcd, pld	5	0.996	0.15	0.22	0.955
d, gsa, vsa, pv, pld	5	0.996	0.17	0.22	0.952
gsa, vsa, vf, pv, lcd	5	0.996	0.15	0.21	0.954
d, vsa, vf, pv, pld	5	0.996	0.15	0.21	0.954
d, gsa, vsa, vf, lcd	5	0.996	0.16	0.22	0.954
d, gsa, vf, pv, pld	5	0.996	0.16	0.22	0.954
d, gsa, vsa, vf, pld	5	0.996	0.16	0.22	0.952
gsa, vsa, vf, pv, pld	5	0.996	0.16	0.24	0.953
gsa, vsa, pv, lcd, pld	5	0.995	0.17	0.23	0.951
d, gsa, vsa, vf, pv	5	0.995	0.18	0.25	0.948
d, gsa, vsa, lcd, pld	5	0.991	0.23	0.32	0.932
d, gsa, vf, pv, lcd, pld	6	0.997	0.14	0.19	0.959
d, gsa, vsa, vf, pv, lcd	6	0.997	0.15	0.21	0.957
gsa, vsa, vf, pv, lcd, pld	6	0.997	0.14	0.20	0.958
d, gsa, vsa, pv, lcd, pld	6	0.997	0.15	0.20	0.957
d, vsa, vf, pv, lcd, pld	6	0.997	0.14	0.20	0.959
d, gsa, vsa, vf, lcd, pld	6	0.996	0.15	0.21	0.956
d, gsa, vsa, vf, pv, pld	6	0.996	0.16	0.22	0.954
d, gsa, vsa, vf, pv, lcd, pld	7	0.997	0.14	0.19	0.959

Usable volumetric H_2 capacity (g/L) under pressure swing (PS) beteween 100 and 5 bar at 77K

Combination of features	Number of features	R^2	Average Unsigned Error, AUE (g/L)	Root-Mean-Square_Error, RMSE (g/L)	Kendall τ
vf	1	0.966	1.45	2.05	0.895
pv	1	0.923	2.28	3.09	0.815
d	1	0.882	2.79	3.78	0.776
gsa	1	0.861	3.14	4.09	0.719
pld	1	0.817	3.61	4.74	0.678
lcd	1	0.716	4.45	5.90	0.643
vsa	1	0.532	5.87	7.57	0.418
vf, lcd	2	0.978	1.15	1.65	0.907
d, pv	2	0.972	1.35	1.84	0.891
vf, pld	2	0.970	1.32	1.88	0.893
vf, pv	2	0.967	1.40	2.01	0.887
d, vf	2	0.966	1.41	2.04	0.888
vsa, vf	2	0.964	1.45	2.10	0.885
gsa, vf	2	0.955	1.65	2.33	0.867
pv, lcd	2	0.934	2.08	2.82	0.819
d, lcd	2	0.923	2.25	3.09	0.806
pv, pld	2	0.923	2.26	3.08	0.804
gsa, lcd	2	0.921	2.28	3.10	0.797
vsa, pv	2	0.905	2.48	3.41	0.789
d, pld	2	0.903	2.51	3.45	0.785
gsa, pv	2	0.901	2.48	3.47	0.792
gsa, vsa	2	0.895	2.67	3.59	0.774
lcd, pld	2	0.891	2.67	3.64	0.763
gsa, pld	2	0.876	2.92	3.90	0.737
d, vsa	2	0.875	2.85	3.91	0.754
d, gsa	2	0.870	2.90	3.98	0.752
vsa, lcd	2	0.864	2.90	4.05	0.716
vsa, pld	2	0.845	3.17	4.32	0.691
vsa, vf, lcd	3	0.976	1.23	1.72	0.900
d, pv, lcd	3	0.975	1.27	1.73	0.895
d, vf, lcd	3	0.974	1.23	1.77	0.898
vf, pv, lcd	3	0.973	1.24	1.82	0.899
gsa, vf, lcd	3	0.973	1.27	1.83	0.898
gsa, vsa, pv	3	0.972	1.33	1.86	0.894
d, vf, pv	3	0.971	1.34	1.88	0.894
vsa, vf, pld	3	0.971	1.31	1.87	0.894
vf, lcd, pld	3	0.971	1.25	1.89	0.898
gsa, vsa, vf	3	0.970	1.33	1.91	0.895
d, gsa, pv	3	0.970	1.40	1.92	0.884
vsa, vf, pv	3	0.969	1.34	1.96	0.893
d, pv, pld	3	0.968	1.41	1.96	0.885
gsa, vf, pld	3	0.968	1.36	1.95	0.890
d, vsa, pv	3	0.968	1.43	1.98	0.882
d, gsa, vf	3	0.968	1.42	2.00	0.886
gsa, vf, pv	3	0.967	1.37	1.99	0.891
d, vsa, vf	3	0.967	1.39	2.00	0.888
vf, pv, pld	3	0.966	1.42	2.02	0.885
d, vf, pld	3	0.966	1.42	2.04	0.886
vsa, pv, lcd	3	0.955	1.72	2.35	0.844
gsa, pv, lcd	3	0.950	1.80	2.46	0.842
gsa, lcd, pld	3	0.950	1.81	2.45	0.838
vsa, lcd, pld	3	0.947	1.86	2.53	0.821
gsa, vsa, lcd	3	0.946	1.86	2.53	0.831
pv, lcd, pld	3	0.945	1.86	2.57	0.837
d, gsa, lcd	3	0.943	1.92	2.64	0.825
d, vsa, lcd	3	0.942	1.93	2.63	0.823
vsa, pv, pld	3	0.938	2.03	2.74	0.821
gsa, pv, pld	3	0.935	2.04	2.83	0.823
d, lcd, pld	3	0.931	2.11	2.90	0.815
gsa, vsa, pld	3	0.930	2.18	2.93	0.810
d, gsa, pld	3	0.926	2.23	3.01	0.800
d, vsa, pld	3	0.924	2.25	3.03	0.798
d, gsa, vsa	3	0.887	2.71	3.71	0.770
gsa, vf, lcd, pld	4	0.981	1.07	1.54	0.913
d, vsa, vf, lcd	4	0.980	1.10	1.58	0.911
d, pv, lcd, pld	4	0.979	1.15	1.60	0.905
vf, pv, lcd, pld	4	0.979	1.10	1.59	0.910
d, vsa, pv, lcd	4	0.978	1.18	1.63	0.902
d, gsa, vf, lcd	4	0.978	1.12	1.63	0.910
vsa, vf, pv, lcd	4	0.978	1.10	1.63	0.912
d, gsa, pv, lcd	4	0.978	1.18	1.64	0.903
vsa, vf, lcd, pld	4	0.978	1.11	1.63	0.909
d, vf, lcd, pld	4	0.977	1.11	1.66	0.910
gsa, vf, pv, lcd	4	0.977	1.13	1.66	0.909
d, gsa, vf, pld	4	0.977	1.17	1.67	0.906
d, vsa, vf, pld	4	0.977	1.17	1.67	0.907
gsa, vf, pv, pld	4	0.977	1.18	1.70	0.905
gsa, vsa, vf, lcd	4	0.976	1.14	1.70	0.907
gsa, vsa, vf, pld	4	0.976	1.18	1.70	0.906
d, vsa, pv, pld	4	0.976	1.22	1.71	0.900
d, gsa, pv, pld	4	0.975	1.22	1.73	0.899
d, vf, pv, lcd	4	0.975	1.20	1.74	0.902
vsa, vf, pv, pld	4	0.975	1.19	1.75	0.906
gsa, vsa, pv, lcd	4	0.974	1.27	1.77	0.894
d, vsa, vf, pv	4	0.973	1.30	1.83	0.897
d, gsa, vsa, pv	4	0.972	1.34	1.86	0.894
gsa, vsa, pv, pld	4	0.971	1.31	1.86	0.894
gsa, vsa, vf, pv	4	0.971	1.29	1.85	0.896
d, gsa, vf, pv	4	0.971	1.31	1.88	0.896
d, vf, pv, pld	4	0.970	1.36	1.92	0.889
d, gsa, vsa, vf	4	0.968	1.35	1.96	0.892
vsa, pv, lcd, pld	4	0.968	1.46	1.98	0.868
gsa, vsa, lcd, pld	4	0.963	1.58	2.14	0.858
d, vsa, lcd, pld	4	0.961	1.59	2.17	0.857
d, gsa, lcd, pld	4	0.960	1.60	2.20	0.854
gsa, pv, lcd, pld	4	0.957	1.63	2.28	0.854
d, gsa, vsa, lcd	4	0.946	1.90	2.58	0.829
d, gsa, vsa, pld	4	0.929	2.19	2.95	0.808
vsa, vf, pv, lcd, pld	5	0.983	1.00	1.45	0.920
d, gsa, vf, lcd, pld	5	0.982	1.01	1.46	0.918
gsa, vf, pv, lcd, pld	5	0.982	1.03	1.48	0.917
gsa, vsa, vf, lcd, pld	5	0.982	1.04	1.50	0.918
d, vsa, vf, pv, lcd	5	0.982	1.04	1.48	0.917
d, vsa, vf, lcd, pld	5	0.982	1.02	1.50	0.916
d, vf, pv, lcd, pld	5	0.981	1.04	1.51	0.916
gsa, vsa, vf, pv, lcd	5	0.981	1.08	1.54	0.912
d, gsa, pv, lcd, pld	5	0.980	1.10	1.54	0.908
d, gsa, vsa, pv, lcd	5	0.980	1.14	1.57	0.906
d, vsa, pv, lcd, pld	5	0.980	1.10	1.56	0.908
d, gsa, vf, pv, lcd	5	0.980	1.08	1.57	0.912
d, gsa, vf, pv, pld	5	0.979	1.12	1.59	0.911
d, gsa, vsa, vf, lcd	5	0.979	1.11	1.61	0.910
d, vsa, vf, pv, pld	5	0.978	1.13	1.63	0.911
gsa, vsa, vf, pv, pld	5	0.977	1.16	1.67	0.908
d, gsa, vsa, vf, pld	5	0.977	1.14	1.67	0.909
d, gsa, vsa, pv, pld	5	0.975	1.23	1.74	0.900
gsa, vsa, pv, lcd, pld	5	0.974	1.27	1.80	0.894
d, gsa, vsa, vf, pv	5	0.974	1.25	1.78	0.900
d, gsa, vsa, lcd, pld	5	0.963	1.56	2.11	0.859
d, vsa, vf, pv, lcd, pld	6	0.983	1.00	1.46	0.918
gsa, vsa, vf, pv, lcd, pld	6	0.982	1.01	1.47	0.918
d, gsa, vsa, vf, lcd, pld	6	0.982	1.04	1.50	0.917
d, gsa, vsa, vf, pv, lcd	6	0.981	1.06	1.52	0.914
d, gsa, vf, pv, lcd, pld	6	0.981	1.03	1.51	0.916
d, gsa, vsa, pv, lcd, pld	6	0.980	1.11	1.58	0.909
d, gsa, vsa, vf, pv, pld	6	0.978	1.13	1.63	0.908
d, gsa, vsa, vf, pv, lcd, pld	7	0.983	0.99	1.43	0.921

Usable gravimetric H_2 capacity (wt.%) under temperature+pressure swing (TPS) between 100bar

Combination of features	Number of features	R^2	Average Unsigned Error, AUE (wt.%)	Root-Mean-Square_Error, RMSE (wt.%)	Kendall τ
pv	1	0.993	0.26	0.34	0.945
d	1	0.986	0.36	0.47	0.926
vf	1	0.950	0.59	0.88	0.871
gsa	1	0.883	0.78	1.31	0.839
pld	1	0.751	1.35	1.99	0.682
lcd	1	0.682	1.69	2.28	0.597
vsa	1	0.169	2.26	3.65	0.352
d, pv	2	0.994	0.23	0.31	0.949
vf, pv	2	0.993	0.23	0.32	0.949
gsa, pv	2	0.993	0.25	0.33	0.944
d, vf	2	0.993	0.24	0.33	0.947
vsa, pv	2	0.993	0.25	0.33	0.943
pv, lcd	2	0.993	0.25	0.35	0.945
pv, pld	2	0.992	0.27	0.36	0.942
d, gsa	2	0.990	0.31	0.40	0.931
d, vsa	2	0.990	0.30	0.41	0.931
d, lcd	2	0.989	0.32	0.43	0.929
d, pld	2	0.987	0.34	0.45	0.925
gsa, vsa	2	0.987	0.32	0.46	0.929
gsa, vf	2	0.972	0.45	0.67	0.897
gsa, lcd	2	0.967	0.48	0.73	0.898
vf, lcd	2	0.957	0.58	0.84	0.867
vf, pld	2	0.956	0.58	0.85	0.864
gsa, pld	2	0.952	0.60	0.90	0.868
vsa, vf	2	0.941	0.67	0.96	0.838
vsa, pld	2	0.844	1.08	1.58	0.728
lcd, pld	2	0.824	1.13	1.66	0.728
vsa, lcd	2	0.819	1.17	1.70	0.693
d, gsa, pv	3	0.995	0.21	0.29	0.951
gsa, vsa, pv	3	0.995	0.21	0.28	0.953
gsa, pv, pld	3	0.995	0.21	0.28	0.952
d, vsa, pv	3	0.995	0.22	0.29	0.949
gsa, vf, pv	3	0.995	0.21	0.29	0.950
vsa, pv, lcd	3	0.995	0.21	0.30	0.952
gsa, pv, lcd	3	0.994	0.22	0.30	0.950
vsa, vf, pv	3	0.994	0.22	0.31	0.950
d, pv, lcd	3	0.994	0.22	0.30	0.951
d, gsa, vf	3	0.994	0.23	0.30	0.949
vsa, pv, pld	3	0.994	0.22	0.30	0.950
gsa, vsa, vf	3	0.994	0.22	0.31	0.950
vf, pv, lcd	3	0.994	0.22	0.30	0.950
d, vf, pv	3	0.994	0.23	0.31	0.949
d, vf, lcd	3	0.994	0.22	0.31	0.950
pv, lcd, pld	3	0.994	0.23	0.32	0.947
d, vsa, vf	3	0.994	0.23	0.31	0.948
d, gsa, lcd	3	0.993	0.26	0.35	0.941
d, pv, pld	3	0.993	0.25	0.34	0.942
d, vsa, pld	3	0.993	0.26	0.35	0.941
d, gsa, pld	3	0.993	0.26	0.34	0.942
vf, pv, pld	3	0.993	0.25	0.35	0.943
d, vf, pld	3	0.992	0.25	0.34	0.943
gsa, vsa, pld	3	0.992	0.26	0.36	0.941
d, vsa, lcd	3	0.992	0.26	0.36	0.940
gsa, vsa, lcd	3	0.992	0.26	0.36	0.940
d, gsa, vsa	3	0.990	0.29	0.39	0.932
d, lcd, pld	3	0.988	0.32	0.43	0.926
gsa, vf, pld	3	0.986	0.35	0.49	0.917
gsa, lcd, pld	3	0.984	0.35	0.51	0.924
gsa, vf, lcd	3	0.983	0.34	0.51	0.920
vf, lcd, pld	3	0.954	0.60	0.87	0.856
vsa, vf, pld	3	0.950	0.62	0.90	0.847
vsa, vf, lcd	3	0.949	0.62	0.89	0.851
vsa, lcd, pld	3	0.905	0.83	1.23	0.802
gsa, vf, pv, lcd	4	0.996	0.18	0.25	0.959
d, gsa, pv, lcd	4	0.996	0.18	0.25	0.959
d, vsa, pv, lcd	4	0.996	0.19	0.26	0.956
gsa, vsa, pv, lcd	4	0.996	0.20	0.26	0.956
vsa, vf, pv, pld	4	0.996	0.19	0.26	0.956
d, vsa, pv, pld	4	0.996	0.20	0.26	0.956
vsa, vf, pv, lcd	4	0.996	0.19	0.26	0.956
vf, pv, lcd, pld	4	0.996	0.19	0.26	0.957
gsa, pv, lcd, pld	4	0.996	0.20	0.27	0.956
gsa, vsa, vf, pv	4	0.996	0.20	0.27	0.953
d, gsa, vf, pld	4	0.996	0.20	0.27	0.956
d, gsa, vf, lcd	4	0.995	0.19	0.27	0.957
d, vsa, vf, lcd	4	0.995	0.19	0.27	0.957
d, vsa, vf, pld	4	0.995	0.20	0.27	0.955
d, gsa, pv, pld	4	0.995	0.19	0.27	0.956
gsa, vsa, pv, pld	4	0.995	0.20	0.27	0.955
d, pv, lcd, pld	4	0.995	0.19	0.28	0.956
vsa, pv, lcd, pld	4	0.995	0.20	0.27	0.956
gsa, vf, pv, pld	4	0.995	0.19	0.27	0.957
d, vsa, vf, pv	4	0.995	0.21	0.28	0.952
d, gsa, vsa, pv	4	0.995	0.21	0.28	0.953
gsa, vsa, vf, lcd	4	0.995	0.20	0.28	0.955
d, vf, pv, lcd	4	0.995	0.21	0.28	0.952
d, vf, lcd, pld	4	0.995	0.20	0.28	0.954
gsa, vsa, vf, pld	4	0.995	0.20	0.29	0.955
d, gsa, vf, pv	4	0.994	0.21	0.30	0.950
d, gsa, vsa, vf	4	0.994	0.22	0.30	0.951
d, vsa, lcd, pld	4	0.994	0.24	0.32	0.947
d, gsa, lcd, pld	4	0.994	0.23	0.31	0.947
d, vf, pv, pld	4	0.993	0.25	0.33	0.944
d, gsa, vsa, pld	4	0.993	0.25	0.33	0.943
gsa, vsa, lcd, pld	4	0.993	0.24	0.34	0.944
d, gsa, vsa, lcd	4	0.992	0.25	0.35	0.942
gsa, vf, lcd, pld	4	0.990	0.29	0.41	0.937
vsa, vf, lcd, pld	4	0.960	0.56	0.79	0.863
gsa, vf, pv, lcd, pld	5	0.997	0.18	0.24	0.960
d, vsa, pv, lcd, pld	5	0.996	0.18	0.24	0.959
d, gsa, vf, lcd, pld	5	0.996	0.18	0.25	0.959
d, gsa, pv, lcd, pld	5	0.996	0.18	0.25	0.960
d, gsa, vf, pv, lcd	5	0.996	0.18	0.25	0.959
gsa, vsa, vf, pv, lcd	5	0.996	0.18	0.25	0.958
d, gsa, vf, pv, pld	5	0.996	0.19	0.25	0.958
gsa, vsa, vf, pv, pld	5	0.996	0.19	0.26	0.958
d, vsa, vf, pv, pld	5	0.996	0.19	0.25	0.957
d, gsa, vsa, pv, lcd	5	0.996	0.19	0.26	0.958
d, vf, pv, lcd, pld	5	0.996	0.19	0.26	0.958
d, vsa, vf, lcd, pld	5	0.996	0.18	0.25	0.960
gsa, vsa, pv, lcd, pld	5	0.996	0.19	0.26	0.957
d, vsa, vf, pv, lcd	5	0.996	0.18	0.26	0.959
d, gsa, vsa, pv, pld	5	0.996	0.19	0.26	0.957
d, gsa, vsa, vf, lcd	5	0.996	0.19	0.26	0.957
d, gsa, vsa, vf, pld	5	0.996	0.20	0.27	0.955
gsa, vsa, vf, lcd, pld	5	0.995	0.19	0.26	0.957
d, gsa, vsa, vf, pv	5	0.995	0.21	0.28	0.952
vsa, vf, pv, lcd, pld	5	0.995	0.18	0.28	0.958
d, gsa, vsa, lcd, pld	5	0.994	0.23	0.31	0.948
d, gsa, vsa, vf, pv, lcd	6	0.996	0.18	0.24	0.960
d, vsa, vf, pv, lcd, pld	6	0.996	0.18	0.24	0.961
d, gsa, vf, pv, lcd, pld	6	0.996	0.17	0.25	0.960
d, gsa, vsa, pv, lcd, pld	6	0.996	0.18	0.24	0.959
d, gsa, vsa, vf, pv, pld	6	0.996	0.19	0.25	0.957
d, gsa, vsa, vf, lcd, pld	6	0.996	0.18	0.25	0.960
gsa, vsa, vf, pv, lcd, pld	6	0.996	0.18	0.27	0.960
d, gsa, vsa, vf, pv, lcd, pld	7	0.997	0.17	0.24	0.961

Usable volumetric H_2 capacity (g/L) under temperature+pressure swing (TPS) between 100bar					
Combination of features	**Number of features**	**R^2**	**Average Unsigned Error, AUE (g/L)**	**Root-Mean-Square_Error, RMSE (g/L)**	**Kendall τ**
vf	1	0.907	2.18	3.25	0.703
pv	1	0.872	2.62	3.78	0.639
gsa	1	0.833	2.93	4.31	0.616
vsa	1	0.819	3.14	4.55	0.663
pld	1	0.788	3.37	4.80	0.539
d	1	0.785	3.23	4.89	0.617
lcd	1	0.689	4.09	5.86	0.412
d, pv	2	0.937	1.86	2.65	0.726
vf, lcd	2	0.933	1.84	2.83	0.768
vf, pv	2	0.929	1.90	2.80	0.727
vf, pld	2	0.924	2.00	2.90	0.710
vsa, vf	2	0.920	2.03	3.02	0.732
d, vf	2	0.907	2.03	3.21	0.725
vsa, pv	2	0.901	2.38	3.38	0.696
gsa, vf	2	0.897	2.38	3.40	0.654
vsa, pld	2	0.894	2.39	3.46	0.694
gsa, pv	2	0.886	2.50	3.57	0.661
vsa, lcd	2	0.883	2.42	3.61	0.689
pv, lcd	2	0.881	2.52	3.66	0.667
gsa, lcd	2	0.879	2.58	3.75	0.675
d, vsa	2	0.877	2.51	3.69	0.682
pv, pld	2	0.877	2.60	3.67	0.636
gsa, pld	2	0.876	2.65	3.70	0.638
d, gsa	2	0.871	2.62	3.88	0.677
gsa, vsa	2	0.850	2.60	4.07	0.688
d, lcd	2	0.843	2.82	4.24	0.669
d, pld	2	0.841	2.84	4.22	0.636
lcd, pld	2	0.828	2.96	4.39	0.626
d, vsa, pv	3	0.943	1.79	2.58	0.747
gsa, vf, pv	3	0.941	1.74	2.54	0.756
vsa, vf, pv	3	0.940	1.75	2.59	0.760
d, vf, lcd	3	0.939	1.78	2.67	0.761
gsa, vf, lcd	3	0.939	1.84	2.67	0.748
d, pv, lcd	3	0.939	1.80	2.61	0.754
d, gsa, pv	3	0.938	1.82	2.66	0.749
d, vf, pv	3	0.938	1.86	2.67	0.722
vsa, vf, pld	3	0.938	1.80	2.61	0.751
vsa, vf, lcd	3	0.935	1.78	2.71	0.753
vf, pv, lcd	3	0.934	1.80	2.75	0.767
d, pv, pld	3	0.933	1.99	2.74	0.698
vf, lcd, pld	3	0.932	1.88	2.82	0.754
gsa, vsa, vf	3	0.930	1.82	2.81	0.761
gsa, vsa, pv	3	0.927	1.87	2.84	0.763
vsa, pv, pld	3	0.926	2.01	2.90	0.735
d, vf, pld	3	0.926	1.98	2.87	0.710
vf, pv, pld	3	0.925	1.99	2.85	0.706
gsa, vf, pld	3	0.922	2.02	2.93	0.700
vsa, pv, lcd	3	0.919	2.02	3.01	0.742
d, gsa, vf	3	0.919	1.90	3.00	0.746
d, vsa, vf	3	0.917	1.95	3.07	0.751
gsa, pv, pld	3	0.917	2.18	3.11	0.719
vsa, lcd, pld	3	0.915	2.00	3.04	0.742
gsa, vsa, lcd	3	0.915	2.08	3.12	0.733
gsa, pv, lcd	3	0.912	2.13	3.09	0.713
d, vsa, pld	3	0.911	2.12	3.21	0.732
gsa, vsa, pld	3	0.911	2.11	3.15	0.733
d, gsa, pld	3	0.903	2.20	3.32	0.724
d, vsa, lcd	3	0.902	2.16	3.33	0.737
d, gsa, lcd	3	0.902	2.14	3.28	0.730
gsa, lcd, pld	3	0.898	2.25	3.31	0.717
pv, lcd, pld	3	0.886	2.42	3.63	0.695
d, gsa, vsa	3	0.877	2.50	3.76	0.688
d, lcd, pld	3	0.865	2.64	3.98	0.675
d, gsa, pv, lcd	4	0.955	1.57	2.29	0.790
d, pv, lcd, pld	4	0.953	1.57	2.29	0.776
d, gsa, vf, lcd	4	0.953	1.53	2.31	0.792
d, gsa, pv, pld	4	0.952	1.61	2.34	0.781
d, vsa, pv, lcd	4	0.952	1.56	2.31	0.785
d, vsa, pv, pld	4	0.951	1.60	2.31	0.779
gsa, vf, pv, lcd	4	0.951	1.55	2.36	0.800
gsa, vsa, vf, pld	4	0.951	1.59	2.34	0.785
gsa, vsa, vf, lcd	4	0.951	1.56	2.34	0.793
vsa, vf, pv, lcd	4	0.951	1.54	2.34	0.796
d, vsa, vf, lcd	4	0.951	1.54	2.35	0.795
vsa, vf, pv, pld	4	0.950	1.60	2.37	0.791
d, gsa, vf, pld	4	0.949	1.63	2.40	0.782
vsa, vf, lcd, pld	4	0.949	1.58	2.36	0.788
gsa, vf, lcd, pld	4	0.948	1.63	2.44	0.788
vf, pv, lcd, pld	4	0.948	1.60	2.38	0.792
d, vsa, vf, pld	4	0.948	1.62	2.41	0.781
d, vsa, vf, pv	4	0.947	1.70	2.46	0.769
d, gsa, vf, pv	4	0.947	1.70	2.48	0.765
gsa, vf, pv, pld	4	0.947	1.62	2.45	0.786
d, vf, lcd, pld	4	0.947	1.59	2.43	0.794
gsa, vsa, vf, pv	4	0.946	1.70	2.45	0.769
d, vf, pv, lcd	4	0.946	1.68	2.44	0.765
d, gsa, vsa, pv	4	0.945	1.70	2.47	0.760
gsa, vsa, pv, pld	4	0.938	1.74	2.61	0.773
gsa, vsa, pv, lcd	4	0.936	1.73	2.69	0.777
d, vf, pv, pld	4	0.934	1.91	2.71	0.709
d, vsa, lcd, pld	4	0.928	1.87	2.89	0.764
vsa, pv, lcd, pld	4	0.927	1.86	2.90	0.766
gsa, pv, lcd, pld	4	0.925	1.98	2.95	0.749
d, gsa, lcd, pld	4	0.923	1.92	2.99	0.761
d, gsa, vsa, vf	4	0.920	1.90	3.04	0.759
gsa, vsa, lcd, pld	4	0.920	1.90	2.97	0.760
d, gsa, vsa, lcd	4	0.909	2.11	3.23	0.732
d, gsa, vsa, pld	4	0.904	2.12	3.26	0.736
vsa, vf, pv, lcd, pld	5	0.961	1.41	2.10	0.814
d, gsa, vf, pv, lcd	5	0.960	1.49	2.19	0.801
d, vf, pv, lcd, pld	5	0.959	1.50	2.19	0.800
d, gsa, pv, lcd, pld	5	0.959	1.46	2.12	0.800
gsa, vf, pv, lcd, pld	5	0.959	1.45	2.15	0.808
d, vsa, vf, pv, lcd	5	0.959	1.50	2.18	0.800
d, gsa, vf, lcd, pld	5	0.958	1.45	2.18	0.803
d, vsa, vf, lcd, pld	5	0.957	1.45	2.18	0.810
d, vsa, pv, lcd, pld	5	0.957	1.52	2.22	0.797
d, gsa, vf, pv, pld	5	0.956	1.54	2.24	0.790
gsa, vsa, vf, pv, lcd	5	0.956	1.51	2.20	0.794
d, vsa, vf, pv, pld	5	0.956	1.54	2.23	0.793
gsa, vsa, vf, lcd, pld	5	0.953	1.48	2.29	0.809
d, gsa, vsa, vf, pld	5	0.951	1.62	2.42	0.792
d, gsa, vsa, pv, lcd	5	0.950	1.56	2.38	0.787
gsa, vsa, vf, pv, pld	5	0.950	1.59	2.42	0.796
d, gsa, vsa, pv, pld	5	0.949	1.62	2.37	0.774
d, gsa, vsa, vf, lcd	5	0.948	1.57	2.46	0.801
d, gsa, vsa, vf, pv	5	0.947	1.69	2.45	0.766
gsa, vsa, pv, lcd, pld	5	0.939	1.70	2.63	0.789
d, gsa, vsa, lcd, pld	5	0.922	1.86	2.94	0.762
d, gsa, vf, pv, lcd, pld	6	0.965	1.41	2.04	0.810
d, gsa, vsa, vf, lcd, pld	6	0.962	1.42	2.12	0.813
d, vsa, vf, pv, lcd, pld	6	0.960	1.42	2.12	0.810
d, gsa, vsa, vf, pv, lcd	6	0.960	1.46	2.12	0.792
gsa, vsa, vf, pv, lcd, pld	6	0.959	1.45	2.15	0.805
d, gsa, vsa, pv, lcd, pld	6	0.958	1.47	2.14	0.791
d, gsa, vsa, vf, pv, pld	6	0.955	1.56	2.29	0.793
d, gsa, vsa, vf, pv, lcd, pld	7	0.962	1.39	2.06	0.809

References

1.Ahmed, A., Seth, S., Purewal, J., Wong-Foy, A.G., Veenstra, M., Matzger, A.J., and Siegel, D.J. (2019). Exceptional hydrogen storage achieved by screening nearly half a million metal-organic frameworks. Nat. Commun. *10*, 1568.
2.Ahmed, A., and Siegel, D.J. HyMARC Sorbent Machine Learning Model: Predicting the hydrogen storage capacity of metal-organic frameworks via machine learning. https://sorbent-ml.hymarc.org/.
3.Bucior, B.J., Bobbitt, N.S., Islamoglu, T., Goswami, S., Gopalan, A., Yildirim, T., Farha, O.K., Bagheri, N., and Snurr, R.Q. Energy-based descriptors to rapidly predict hydrogen storage in metal-organic frameworks. Mol. Syst. Des. Eng. 2018. DOI 10.1039/c8me00050f.
4.Anderson, R., Rodgers, J., Argueta, E., Biong, A., and Go, D.A. (2018). Role of Pore Chemistry and Topology in the CO 2 Capture Capabilities of MOFs: From Molecular Simulation to Machine Learning. Chem. Mater *30*, 11.
5.Anderson, G., Schweitzer, B., Anderson, R., and Gómez-Gualdrón, D.A. (2019). Attainable Volumetric Targets for Adsorption-Based Hydrogen Storage in Porous Crystals: Molecular Simulation and Machine Learning. J. Phys. Chem. C *123*, 120–130.
6.Pardakhti, M., Moharreri, E., Wanik, D., Suib, S.L., and Srivastava, R. (2017). Machine Learning Using Combined Structural and Chemical Descriptors for Prediction of Methane Adsorption Performance of Metal Organic Frameworks (MOFs). ACS Comb. Sci. *19*, 640–645.
7.Aghaji, M.Z., Fernandez, M., Boyd, P.G., Daff, T.D., and Woo, T.K. (2016). Quantitative Structure – Property Relationship Models for Recognizing Metal Organic Frameworks (MOFs) with High CO 2 Working Capacity and CO2/CH4 Selectivity for Methane Purification. 4505–4511.
8.Fernandez, M., and Barnard, A.S. (2016). Geometrical Properties Can Predict CO_2 and N_2 Adsorption Performance of Metal–Organic Frameworks (MOFs) at Low Pressure. ACS Comb. Sci. *18*, 243–252.
9.Ohno, H., and Mukae, Y. (2016). Machine Learning Approach for Prediction and Search: Application to Methane Storage in a Metal–Organic Framework. J. Phys. Chem. C *120*, 23963–23968.
10. Simon, C.M., Kim, J., Gomez-Gualdron, D.A., Camp, J.S., Chung, Y.G., Martin, R.L., Mercado, R., Deem, M.W., Gunter, D., Haranczyk, M., et al. (2015). The materials genome in action: identifying the performance limits for methane storage. Energy Environ. Sci. *8*, 1190–1199.
11. Sezginel, K.B., Uzun, A., and Keskin, S. (2015). Multivariable linear models of structural parameters to predict methane uptake in metal–organic frameworks. Chem. Eng. Sci. *124*, 125–134.
12. Fernandez, M., Woo, T.K., Wilmer, C.E., and Snurr, R.Q. (2013). Large-Scale Quantitative Structure–Property Relationship (QSPR) Analysis of Methane Storage in Metal–Organic Frameworks. J. Phys. Chem. C *117*, 7681–7689.
13. Fernandez, M., Boyd, P.G., Daff, T.D., Aghaji, M.Z., and Woo, T.K. (2014). Rapid and Accurate Machine Learning Recognition of High Performing Metal Organic Frameworks for CO 2 Capture. J. Phys. Chem. Lett. *5*, 3056–3060.
14. Fischer, M., Hoffmann, F., and Fröba, M. (2009). Preferred hydrogen adsorption sites in various MOFs-A comparative computational study. ChemPhysChem *10*, 2647–2657.
15. Feynman, R.P., and Hibbs, A.R. (1965). Quantum mechanics and path integrals (McGraw-Hill).
16. Ahmed, A., Liu, Y., Purewal, J., Tran, L.D., Veenstra, M., Wong-Foy, A., Matzger, A., and Siegel, D. (2017). Balancing Gravimetric and Volumetric Hydrogen Density in MOFs. Energy Environ. Sci. *10*, 2459–2471.
17. Lorentz, H.A. (1881). Ueber die Anwendung des Satzes vom Virial in der kinetischen Theorie der Gase. Ann. Phys. *248*, 127–136.
18. Sandler, S.I. (2006). Chemical, biochemical, and engineering thermodynamics 4th ed. (Wiley).
19. Rappe, A.K., Casewit, C.J., Colwell, K.S., Goddard, W.A., and Skiff, W.M. (1992). UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. J. Am. Chem. Soc. *114*, 10024–10035.
20. Mayo, S.L., Olafson, B.D., and Goddard III, W.A. (1990). DREIDING: A Generic Force Field for Molecular Simulations. J . Phys. Chem *94*, 8897–8909.
21. Allen, M.P., and Tildesley, D.J. (1989). Computer simulation of liquids (Oxford University Press).
22. Sadus, R.J. (1999). Molecular simulation of fluids: theory, algorithms, and object-orientation. (Elsevier).
23. Dubbeldam, D., Calero, S., Ellis, D.E., and Snurr, R.Q. (2016). RASPA: molecular simulation software for adsorption and diffusion in flexible nanoporous materials. Mol. Simul. *42*, 81–101.
24. García-Holley, P., Schweitzer, B., Islamoglu, T., Liu, Y., Lin, L., Rodriguez, S., Weston, M.H., Hupp, J.T., Gómez-Gualdrón, D.A., Yildirim, T., et al. (2018). Benchmark Study of Hydrogen Storage in Metal–Organic Frameworks under Temperature and Pressure Swing Conditions. ACS Energy Lett., 748–754.
25. Zwillinger, D., Kokoska, S., Raton, B., New, L., and Washington, Y. (2000). standard probability and Statistics tables and formulae CRC.
26. Oliphant, T.E. (2007). Python for Scientific Computing. Comput. Sci. Eng. *9*, 10–20.
27. Millman, K.J., and Aivazis, M. (2011). Python for Scientists and Engineers. Comput. Sci. Eng. *13*, 9–12.
28. Abramowitz, M., and Stegun, I.A. (1965). Handbook of mathematical functions, with formulas, graphs, and mathematical tables,

(Dover Publications).
29. Kendall, M.G. (1945). The Treatment of Ties in Ranking Problems. Biometrika *33*, 239–251.
30. Kendall, M.G. (1938). A New Measure of Rank Correlation. Biometrika *30*, 81–93.
31. Press, W.H. (2007). Numerical recipes : the art of scientific computing (Cambridge University Press).
32. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. *12*, 2825–2830.
33. Parrt, T., and Turgutlu, K. rfpimp 1.3.4, https://github.com/parrt/random-forest-importances.

Appendix B
Additional information: Crystal Engineering

Kuthuru Suresh,[1] Darpandeep Aulakh,[1] Justin Purewal,[2] Donald J. Siegel,[3,4,5,6] Mike Veenstra,[2] and Adam J. Matzger[1,7]

[1]Department of Chemistry, University of Michigan, 930 North University Avenue, Ann Arbor, Michigan 48109-1055, United States.

[2]Ford Motor Company, Research and Advanced Engineering, 1201 Village Rd., Dearborn, MI 48121, United States.

[3]Mechanical Engineering Department, University of Michigan, Ann Arbor, MI 48109, United States.

[4]Materials Science & Engineering, University of Michigan, Ann Arbor, MI 48109, United States.

[5]Applied Physics Program, University of Michigan, Ann Arbor, MI 48109, United States.

[6]University of Michigan Energy Institute, University of Michigan, Ann Arbor, MI 48109, United States.

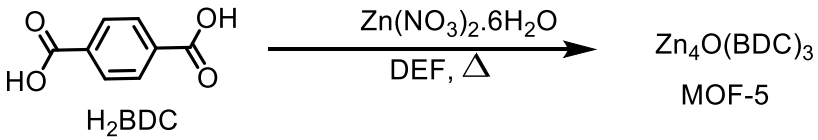
[7]Macromolecular Science and Engineering Program, University of Michigan, Ann Arbor, Michigan 48109-1055, United States.

Correspondence and requests for materials should be addressed to A.J.M. (email: matzger@umich.edu)

Supporting Information

SI 1. MOF-5 cubic morphology size-controlled crystal synthesis and activation procedures

The optimized different metal/ligand molar ratio (3.8:1, 2.3:1, 1.7:1, and 1.7: 1(twofold dilution)) were heated at 100 °C for 24 h, which resulted in four different crystal size distributions (MOF-5(2349), MOF-5(1500), MOF-5(808), and MOF-5(279)) of MOF-5 (see below for synthetic procedures) respectively.



Scheme S1. Synthetic process for MOF-5

MOF-5(2349): H_2BDC (100.0 mg, 0.602 mmol), $Zn(NO_3)_2.6H_2O$ (680.0 mg, 2.286 mmol), and 10 mL of DEF were added to a 20 mL vial. The mixture was sonicated for 15 minutes and heated to 100 °C for 24 hours. The resulting colorless cubic crystals were isolated by decanting the mother liquor and washing with washed with DMF three times (3 × 20 mL) over 24h. The crystals were immersed in CH_2Cl_2 for 3 days, during which time the CH_2Cl_2 was replaced three times (3 × 20 mL). Once solvent exchange was complete, the crystals were isolated by decanting the CH_2Cl_2 and evacuating under dynamic vacuum (0.05 torr) for 24 h at room temperature.

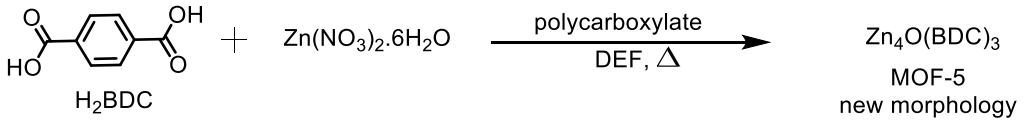
MOF-5(1500): H_2BDC (100.0 mg, 0.602 mmol), $Zn(NO_3)_2.6H_2O$ (400.0 mg, 1.345 mmol), and 10 mL of DEF were added to a 20 mL vial. The mixture was sonicated for 15 minutes and heated to 100 °C for 24 hours. The resulting colorless cubic crystals were isolated by decanting the mother liquor and washing with washed with DMF three times (3 × 20 mL) over 24h. The crystals were immersed in CH_2Cl_2 for 3 days, during which time the CH_2Cl_2 was replaced three times (3 × 20 mL). Once solvent exchange was complete, the crystals were isolated by decanting the CH_2Cl_2 and evacuating under dynamic vacuum (0.05 torr) for 24 h at room temperature.

MOF-5(808): H_2BDC (50.0 mg, 0.301 mmol), $Zn(NO_3)_2.6H_2O$ (150.0 mg, 0.504 mmol), and 10 mL of DEF were added to a 20 mL vial. The mixture was sonicated for 15 minutes and heated to 100 °C for 24 hours. The resulting colorless cubic crystals were isolated by decanting the mother liquor and washing with washed with DMF three times (3 × 20 mL) over 24h. The crystals were immersed in CH_2Cl_2 for 3 days, during which time the CH_2Cl_2 was replaced three times (3 × 20 mL). Once solvent exchange was complete, the crystals were isolated by decanting the CH_2Cl_2 and evacuating under dynamic vacuum (0.05 torr) for 24 h at room temperature.

MOF-5(279): H_2BDC (25.0 mg, 0.151 mmol), $Zn(NO_3)_2.6H_2O$ (75.0 mg, 0.252 mmol), and 10 mL of DEF were added to a 20 mL vial. The mixture was sonicated for 15 minutes and heated to 100 °C for 24 hours. The resulting colorless cubic crystals were isolated by decanting the mother liquor and washing with washed with DMF three times (3 × 20 mL) over 24h. The crystals were immersed in CH_2Cl_2 for 3 days, during which time the CH_2Cl_2 was replaced three times (3 × 20 mL). Once solvent exchange was complete, the crystals were isolated by decanting the CH_2Cl_2 and evacuating under dynamic vacuum (0.05 torr) for 24 h at room temperature.

SI 2. MOF-5 new morphologies synthesis and activation procedures

It was found that introducing a polycarboxylate (See Figure S1) to the initial MOF-5 reaction mixture generates different shaped (octahedral, cuboctahedral, and spherical) crystals in 24 or 48 hours (Scheme S2).



Scheme S2. Synthetic process for different morphologies of MOF-5.

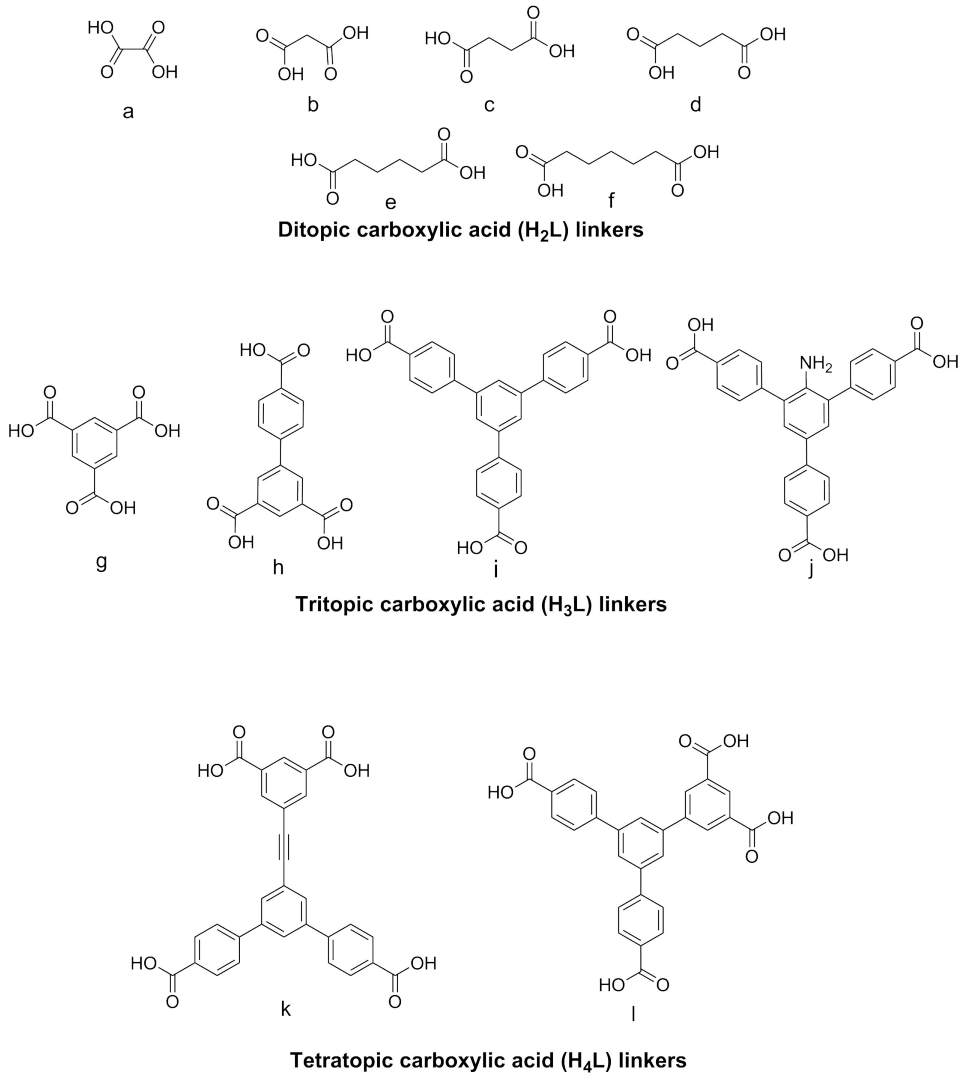


Figure S1. Molecular structures of polycarboxylate linkers studied. oxalic acid (a), malonic acid (b), succinic acid (c), glutaric acid (d), adipic acid (e), suberic acid (f), trimesic acid (g), [1,1'-biphenyl]-3,4',5-tricarboxylic acid (h), 1,3,5-tris(4-carboxyphenyl)benzene (H_3BTB) (i), 2,4,6-tris(4-carboxyphenyl)aniline (NH_2-H_3BTB) (j), 5'-((3,5-dicarboxyphenyl)ethynyl)-[1,1':3',1''-terphenyl]-4,4''-dicarboxylic acid (k), and 5'-(4-carboxyphenyl)-[1,1':3',1''-terphenyl]-3,4'',5-tricarboxylic acid (l).

Ditopic carboxylic acid (H_2L) linkers:

Introducing ditopic carboxylic acids (see Figure S1(a-f) and Table S2) to the MOF-5 initial reaction mixture afforded no change in MOF-5 cubic crystal morphology. It is known that some linear (aromatic) linkers can give rise to new phases incorporating zinc and two linkers,[1] but in all cases examined here the predominant phase was MOF-5.

Table S2. Synthetic conditions of MOF-5 morphologies using ditopic linkers.

S. No.	H_2BDC/ $Zn(NO_3)_2.6H_2O$ (mg (mmol)/ mg(mmol)	Polycarboxylate, mol% (mg)	Solvent (DEF)	Temperature/ time	Observation
1	100.0 (0.602)/ 500.0 (1.68)	oxalic acid, 5.3 (3.0)	10 mL	100 °C/ 24 h	no change in MOF-5 cubic morphology
2	100.0 (0.602)/ 500.0 (1.681)	oxalic acid, 10.6 (6.0)	10 mL	100 °C/ 24 h	no change in MOF-5 cubic morphology
3	100.0 (0.602)/ 500.0 (1.681)	malonic acid, 5.3 (3.5)	10 mL	100 °C/ 24 h	no change in MOF-5 cubic morphology
4	100.0 (0.602)/ 500.0 (1.681)	malonic acid, 10.6 (7.0)	10 mL	100 °C/ 24 h	no change in MOF-5 cubic morphology
5	100.0 (0.602)/ 500.0 (1.681)	succinic acid, 5.3 (4.0)	10 mL	100 °C/ 24 h	no change in MOF-5 cubic morphology
6	100.0 (0.602)/ 500.0 (1.681)	succinic acid, 10.8 (9.0)	10 mL	100 °C/ 24 h	no change in MOF-5 cubic morphology
7	100.0 (0.602)/ 500.0 (1.681)	glutaric acid, 5.4 (4.5)	10 mL	100 °C/ 24 h	no change in MOF-5 cubic morphology
8	100.0 (0.602)/ 500.0 (1.681)	glutaric acid, 10.8 (9.0)	10 mL	100 °C/ 24 h	no change in MOF-5 cubic morphology
9	100.0 (0.602)/ 500.0 (1.681)	adipic acid, 5.4 (5.0)	10 mL	100 °C/ 24 h	no change in MOF-5 cubic morphology
10	100.0 (0.602)/ 500.0 (1.681)	adipic acid, 10.8 (10.0)	10 mL	100 °C/ 24 h	no change in MOF-5 cubic morphology
11	100.0 (0.602)/ 500.0 (1.681)	suberic acid, 5.4 (6.0)	10 mL	100 °C/ 24 h	no change in MOF-5 cubic morphology
12	100.0 (0.602)/ 500.0 (1.681)	suberic acid, 10.8 (12.0)	10 mL	100 °C/ 24 h	no change in MOF-5 cubic morphology

Tritopic carboxylic acid (H_3L) linkers:

The four tritopic linkers (trimesic acid, [1,1'-biphenyl]-3,4',5-tricarboxylic acid, 1,3,5-tris(4-carboxyphenyl)benzene (H_3BTB) and 2,4,6-tris(4-carboxyphenyl)aniline (NH_2-H_3BTB)) (Figure S1g-j) were screened in this study. Among these, the addition of H_3BTB (i) and NH_2-H_3BTB (j) to the initial MOF-5 reagents mixture generated different shaped crystals (see below Table S3).

Table S3. Synthetic conditions for MOF-5-O_h(600), MOF-5-O_C(856), and MOF-5-O_c(575) samples.

Sample name	H_2BDC/ $Zn(NO_3)_2.6H_2O$ (mg (mmol)/ mg(mmol)	Polycarboxylate mol% (mg)	Sol-vent (DEF)	Tempera-ture/ time	Observation
1	(100.0 (0.602)/ 500.0 (1.681)	trimesic acid 10.5 (15.0)	10 mL	100 °C/ 48 h	no change in MOF-5 cubic morphology
2	(100.0 (0.602)/ 500.0 (1.681)	1,1'-biphenyl]-3,4',5-tricarbox-ylic acid, 8.0 (15.0)	10 mL	100 °C/ 48 h	no change in MOF-5 cubic morphology
MOF-5-O_h(600)	(100.0 (0.602)/ 500.0 (1.681)	H_3BTB, 3.6 (10.0)	15 mL	100 °C/ 24 h	octahedral (O_h) morphol-ogy crystals ob-tained
MOF-5-O_C(856)	(100.0 (0.602)/ 500.0 (1.681)	H_3BTB, 1.9 (5.0)	10 mL	100 °C/ 24 h	cuboctahedral (O_c) morphol-ogy crystals ob-tained
MOF-5-O_c(575)	(100.0 (0.602)/ 500.0 (1.681)	NH_2-H_3BTB, 2.7 (7.5)	10 mL	100 °C/ 24 h	cuboctahedral (O_c) morphol-ogy crystals ob-tained

Solvent exchange and activation:

The resulting non-cubic morphology crystals of all three samples were isolated by decanting the mother liquor and were washed with DMF. Crystals were immersed in CH_2Cl_2 for 3 days, during which time the CH_2Cl_2 was replaced three times. Once solvent exchange was complete, the crystals were isolated by decanting the CH_2Cl_2 and evacuation under dynamic vacuum (0.05 torr) for 24 h at room temperature.

^{1}H-NMR Spectroscopy analyses of digested non-cubic morphology samples:

As prepared non-cubic morphology ((MOF-5-O_h(600), MOF-5-O_c(856), and MOF-5-O_c(575)) samples were digested in DCl+DMSO-d_6 solution(Figure S2) and analyzed by NMR spectroscopy. In all three samples, the peaks corresponding to BDC and BTB/BTB-NH_2 were observed (Figure S3a-c). Incorporation of these additives does not yield new phases as confirmed through PXRD (see details and Figure 2 in the main manuscript) and so the linkers must occupy defect sites.

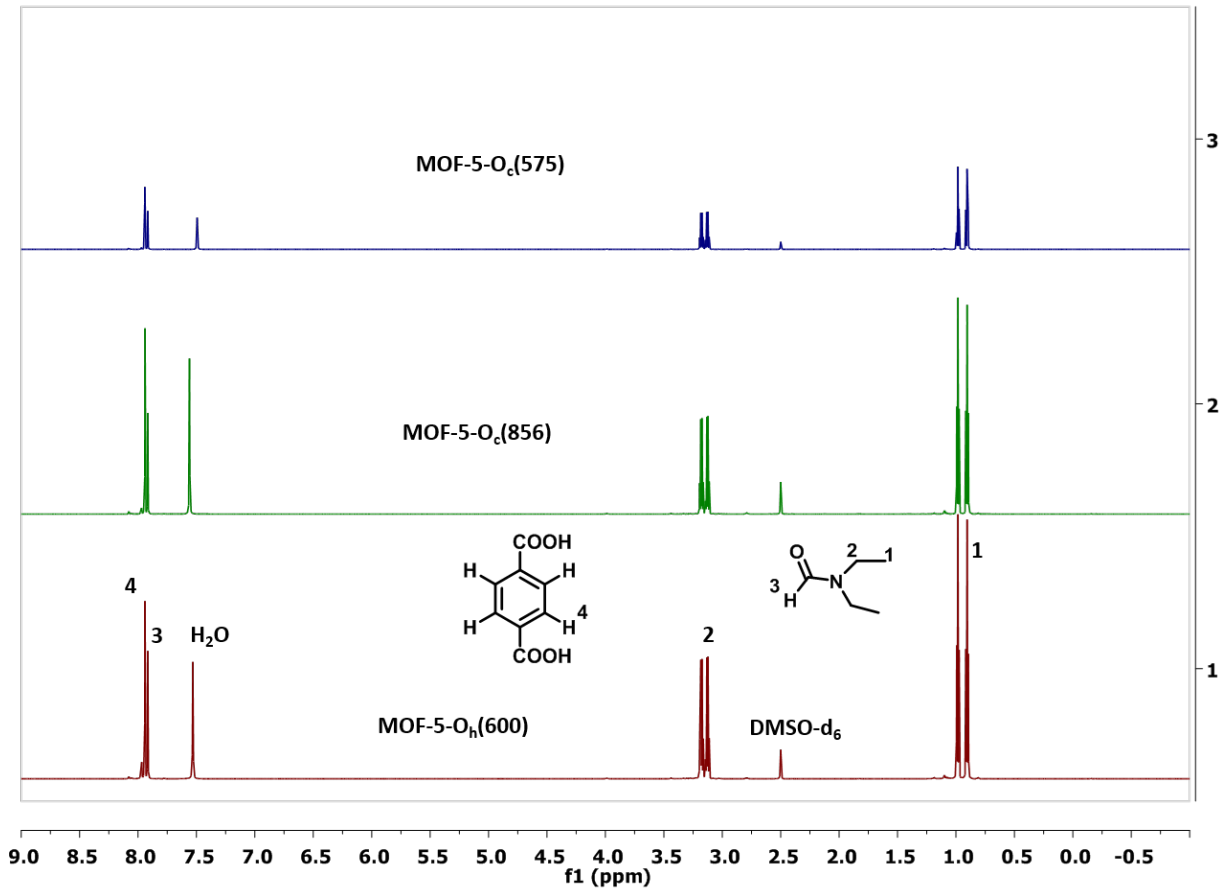


Figure S2. Proton NMR spectrum for as synthesized three different MOF-5 morphology ((MOF-5-O_h(600), MOF-5-O_c(856), and MOF-5-O_c(575)) samples after digesting in DCl+DMSO-d_6 solution.

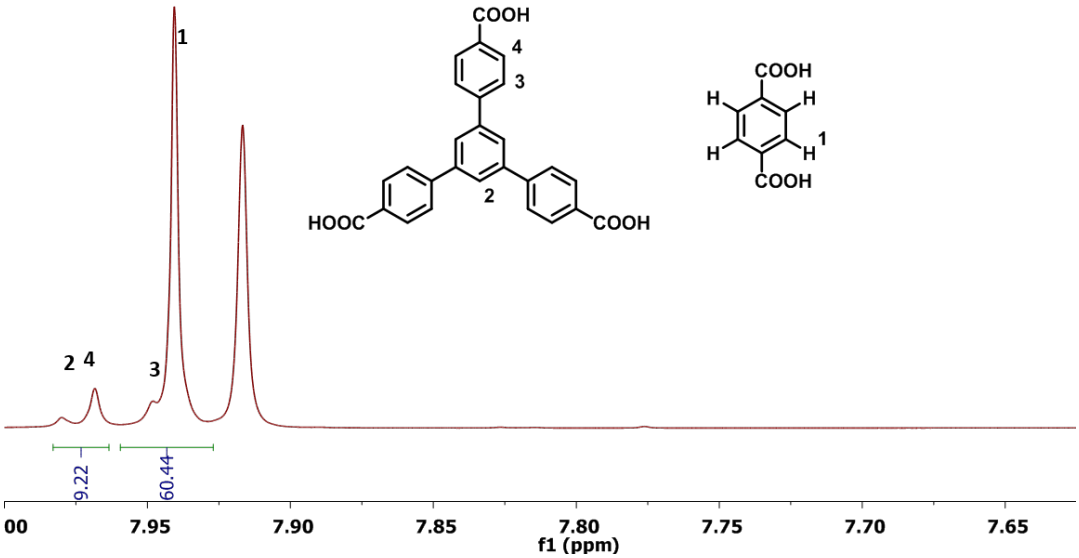


Figure S3a: Aromatic region enlarged ^{1}H-NMR spectrum for an as prepared MOF-5-O_h(600) sample after digesting in DCl+DMSO-d_6 solution.

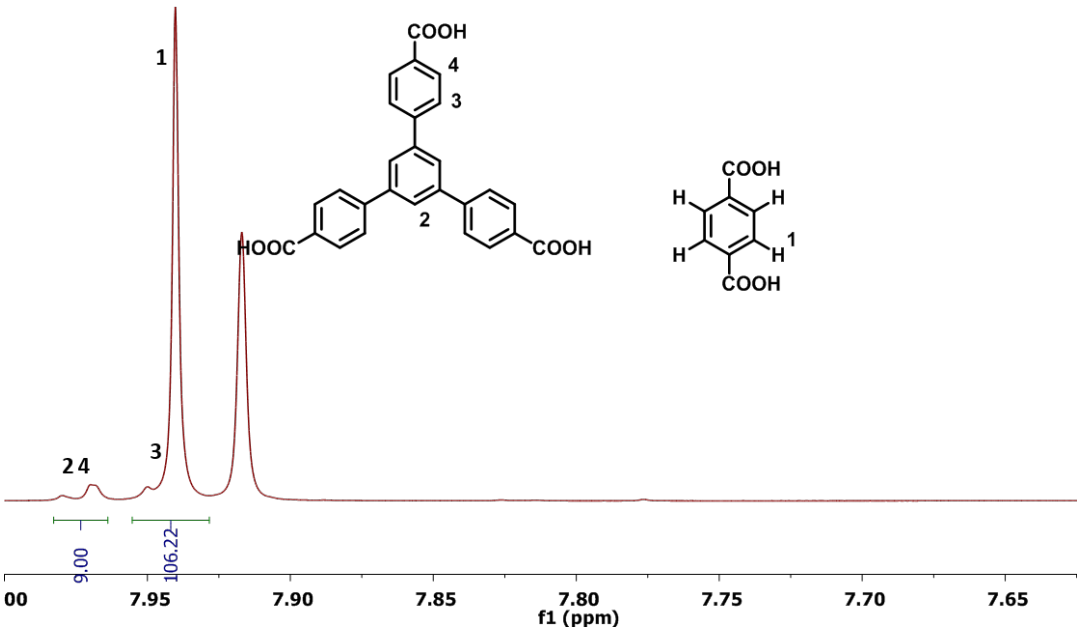


Figure S3b: Aromatic region enlarged ^{1}H-NMR spectrum for an as prepared MOF-5-O$_c$(856) sample after digesting in DCl+DMSO-d$_6$ solution.

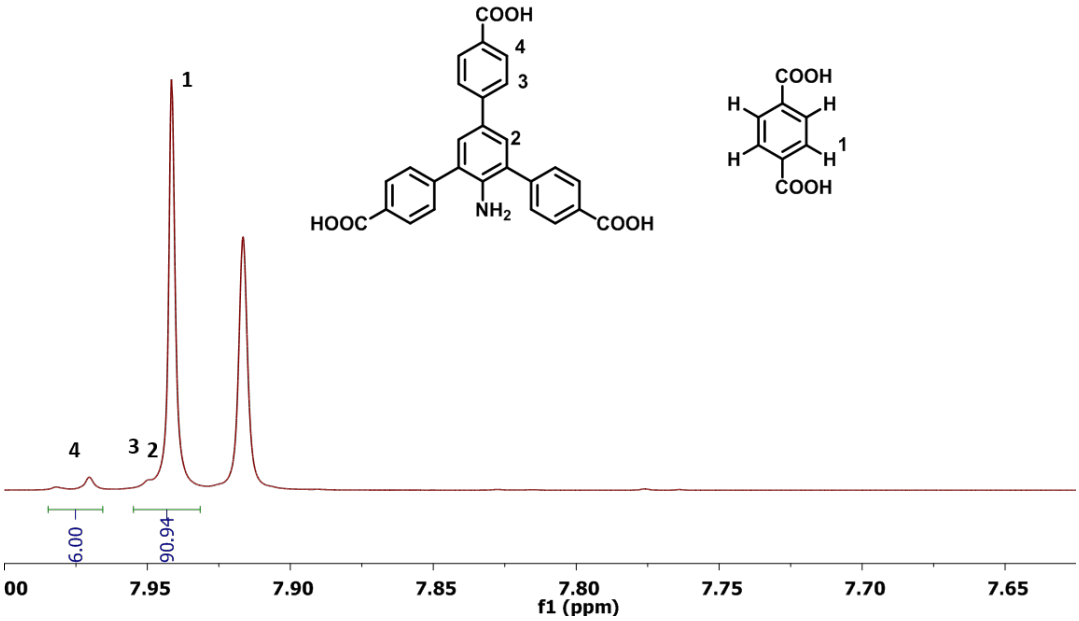


Figure S3c: Aromatic region enlarged ^{1}H-NMR spectrum for an as prepared MOF-5-O$_c$(575) sample after digesting in DCl+DMSO-d$_6$ solution.

Tetratopic carboxylic acid (H_4L) linkers:

The two tetratopic linkers 5'-((3,5-dicarboxyphenyl)ethynyl)-[1,1':3',1''-terphenyl]-4,4''-dicarboxylic acid and 5'-(4-carboxyphenyl)-[1,1':3',1''-terphenyl]-3,4'',5-tricarboxylic acid (Figure S1k and S1l) were screened in this study and inclusion of both additives to the initial MOF-5 reagents mixture generated different shaped crystals (Figure S4).

Table S4. Synthetic conditions for new morphology MOF-5 using H_4L linkers.

Sample name	H_2BDC/ $Zn(NO_3)_2.6H_2O$ (mg (mmol)/ mg(mmol)	Polycarboxylate mol% (mg)	Solvent (DEF)	Temperature/time	Observation
1	100.0 (0.602)/ 500.0 (1.681)	5'-((3,5-dicarboxyphenyl)ethynyl)-[1,1':3',1''-terphenyl]-4,4''-dicarboxylic acid, 1.6 (5.0)	10 mL	100 °C/ 24 h	cuboctahedral (O_c) morphology crystals obtained (See Figure S2 a and b for optical image and PXRD analysis): 2986 m^2/g
2	100.0 (0.602)/ 500.0 (1.681)	5'-((3,5-dicarboxyphenyl)ethynyl)-[1,1':3',1''-terphenyl]-4,4''-dicarboxylic acid, 6.6 (20.5)	10 mL	100 °C/ 24 h	spherical morphology crystals obtained (See Figure S2 a and b for optical image and PXRD analysis): 2445 m^2/g
3	100.0 (0.602)/ 500.0 (1.681)	5'-(4-carboxyphenyl)-[1,1':3',1''-terphenyl]-3,4'',5-tricarboxylic acid, 5.0 (15.5)	10 mL	100 °C/ 15 h	spherical morphology crystals obtained (See Figure S2c and d for optical image and PXRD analysis): 2675 m^2/g

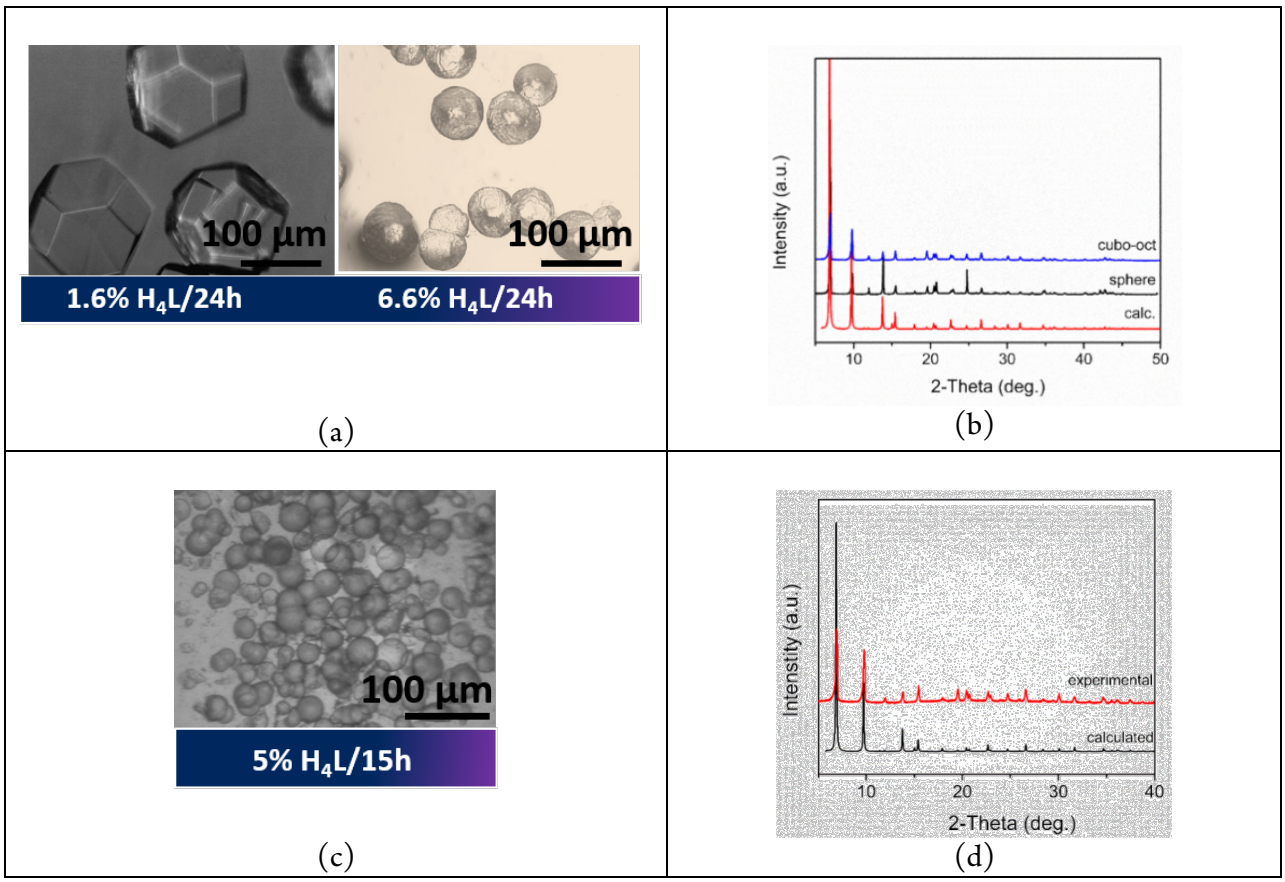


Figure S4. Optical images of different morphologies of MOF-5 crystals obtained by the addition of tetratopic linkers 5'-((3,5-dicarboxyphenyl)ethynyl)-[1,1':3',1''-terphenyl]-4,4''-dicarboxylic acid (a) and 5'-(4-carboxyphenyl)-[1,1':3',1''-terphenyl]-3,4'',5-tricarboxylic acid (b) to the reaction mixture of H_2BDC and $Zn(NO_3)_2{\cdot}6H_2O$. The respective PXRD patterns are shown (b) and (d).

Solvent exchange and activation:

The resulting crystals of all four samples were isolated by decanting the mother liquid and were washed with DMF. Crystals were immersed in CH_2Cl_2 for 3 days, during which time the CH_2Cl_2 was replaced three times. Once solvent exchange was complete, the crystals were isolated by decanting the CH_2Cl_2 and evacuation under dynamic vacuum (0.05 torr) for 24 h at room temperature.

SI 3. Instrumental details

Optical microscopy and scanning electron microscopy:

Inverted Leica DMIL LED and Leica DM2500 LED optical microscopes were used to determine morphologies and suitable images collected and represented in Figure 1, 2 and S3. A JEOL JSM-7800FLV scanning electron microscope operating with an accelerating voltage of 10 kV was used to determine commercially produced MOF-5 morphology (Figure 1). Crystal size distribution analysis graphs were processed using Image-Pro Premier and OriginPro 8 software.

Powder X-ray diffraction

Powder X-ray diffraction (PXRD) data of all samples of MOF-5 were collected on a PANalytical Empyrean diffractometer in Bragg-Brentano geometry using Cu-Kα radiation (λ = 1.54187 Å), operating at 45 kV and 40 mA. The incident beam was equipped with a Bragg-BrentanoHD X-ray optic using fixed slits/soller slits. The detector was a silicon-based linear position sensitive X'Celerator Scientific operating in 1-D scanning mode. Data were

collected from 5 to 50° 2θ using a step size of 0.0083° and a count time of at least 10 s per step. Powder patterns were processed using Data Viewer PANalytical and OriginPro 8 software.

Gas sorption measurements

Sorption experiments were carried out using a NOVA e series 4200 surface area analyzer (Quantachrome Instruments, Boynton Beach, Florida, USA). N2 (99.999%) was purchased from Cryogenic Gases and used as received. For N2 measurements, a glass sample cell was charged with ~30 mg sample and analyzed at 77 K. Sorption isotherms were collected in the NOVAwin software.

^{1}H-NMR measurement

All ^{1}H-NMR measurements were carried out on a Varian Vnmrs 700 (700 MHz (17.6 Tesla) spectrometer. All MOF-5 three different morphologies ((MOF-5-O_h(600), MOF-5-O_c(856), and MOF-5-O_c(575)) were digested in DCl+DMSO-d6 solution(500 μL DMSO-d_6 + 50 μL of 35 wt % DCl in D_2O.)

SI 4. Packing density measurements

A standardized value for the packing or tap density can be measured using DIN, ISO 787 Part II, ISO 3953, or ASTM B 527-93 using a jolting volumeter or tap density measurement instrument. By this method, we measured the tap density for all samples using a custom-built jolting volumeter. Tapping in the volumeter was carried out until the tap density of the sample being studied converged to a constant value. This typically required a minimum of 1,000 taps. The custom-built volumeter has a frequency of 3 taps/sec and included a 35 g weight on the powder within the graduated cylinder for containment and consistent tap density (Figure S5). All tap density measurements were conducted under inert atmosphere in a glove box. There is an insignificant reduction (at most ~150 m^2/g) in surface area observed for all samples after packing density measurements.

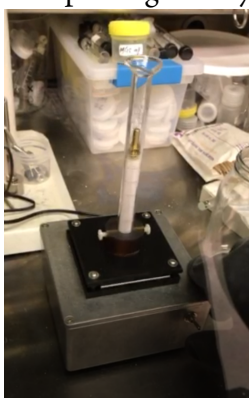
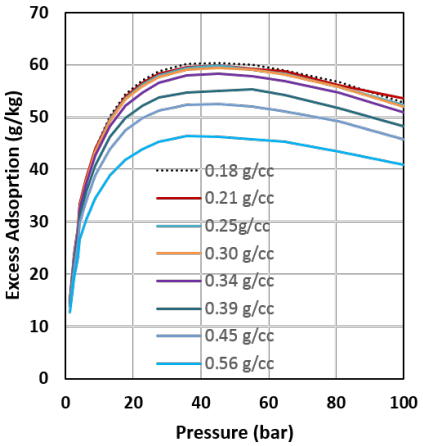


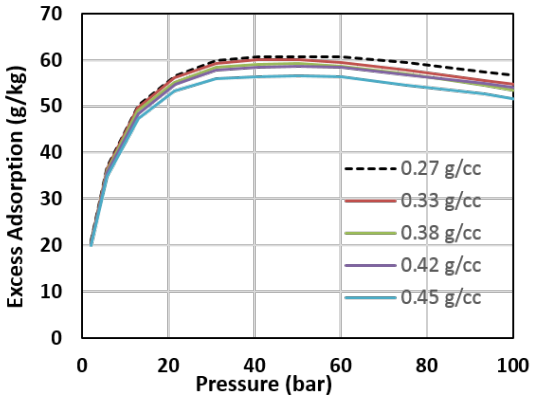
Figure S5. Packing density measurement with custom-built jolting volumeter.

SI 5. Compaction measurements

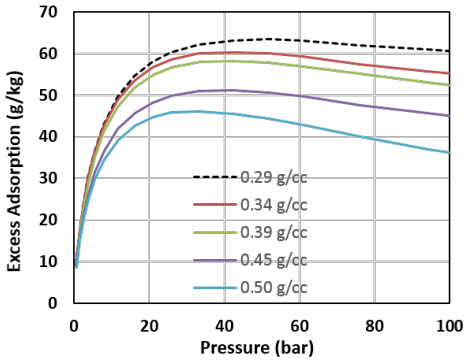
Hydrogen adsorption measurements at variable MOF-5 packing densities were performed by compacting the MOF sample directly within the sample cell to successively higher densities using Parr™ Pellet Press (Stainless steel punch, die holder and die with 1.27 cm dia. x 2.54 cm D die cavity, 17.8 cm L lever arm, 6.2 cm dia. steel cam). Commercial MOF-5, cubic size-controlled, and new morphology of MOF-5 crystals samples were loaded in the sample cell inside the glovebox and compacted to a specified density using a pellet press respectively. Here the applied force was on the order of 50 to 150 lbs., which corresponds to between 10 MPa and 30 MPa pressure, but the force applied will vary based on the press and sample geometry. MOF-5 samples were used for successive measurements without loading fresh MOF samples. The packing densities of the MOF sample were measured before and after each hydrogen sorption measurement and no change in density was detected. Hydrogen adsorption isotherms were measured at 77 K temperature using cryogenic liquid baths. These both MOF sample compaction and hydrogen storage measurements were performed based by the literature method.[2]



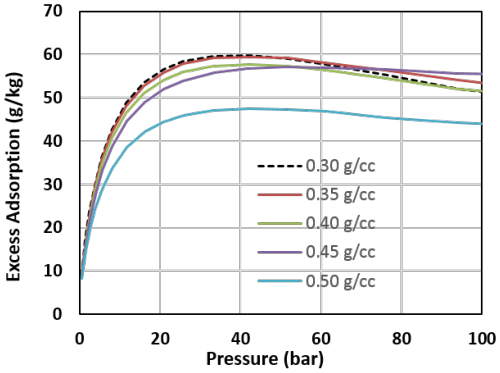
(a) MOF-5 commercial



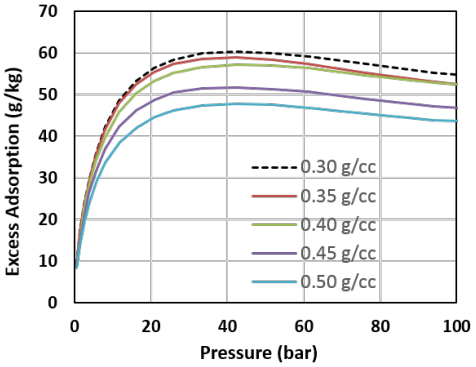
(b) MOF-5(2349)



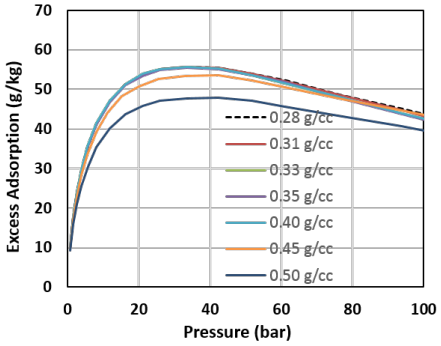
(c) MOF-5(1500)



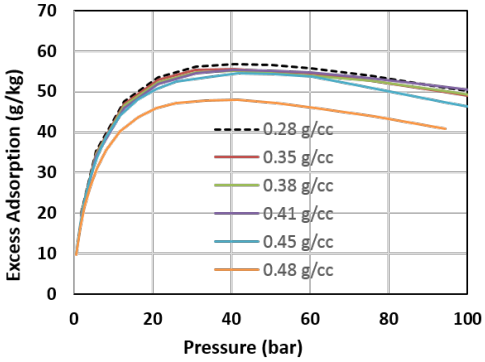
(d)MOF-5(808)



(e) MOF-5(279)



(f) MOF-5(2349):MOF-5(808), 7:1 mixture



(g) MOF-5-O_h(600)

Figure S6. Excess H_2 adsorption isotherms measured at 77 K for commercial MOF-5, MOF(2349), MOF-5(1500), MOF-5(808), MOF-5(279), mixture of MOF-5(2349) and MOF-5(808), and MOF-5-O_h(600) of MOF-5 samples, after compacting to the specified densities.

Table S5. Compacted density, excess H_2 gas adsorption, density fraction and excess fraction values of all samples listed in the below table.

MOF-5 commercial (Crystal density 0.594 g/cc)

Density (g/cc)	Excess (g/kg)	Density Fraction	Excess Fraction
0.18	60.3000	0.3030	1
0.21	59.9100	0.3471	0.9935
0.25	59.8900	0.4209	0.9932
0.30	59.3500	0.5051	0.9842
0.34	58.3200	0.5724	0.9672
0.39	55.0390	0.6566	0.9128
0.45	52.4780	0.7576	0.8703
0.56	46.2480	0.9428	0.7670

MOF-5(2349) (Crystal density 0.594 g/cc)

Density (g/cc)	Excess (g/kg)	Density Fraction	Excess Fraction
0.27	60.715	0.455	1.000
0.33	60.043	0.556	1.000
0.38	59.082	0.640	0.973
0.42	58.478	0.707	0.963
0.45	56.429	0.758	0.929

MOF-5(1500) (Crystal density 0.594 g/cc)

Density (g/cc)	Excess (g/kg)	Density Fraction	Excess Fraction
0.29	63.1190	0.4848	1
0.34	60.3152	0.5724	0.9556
0.40	58.2855	0.6734	0.9234
0.45	51.2119	0.7576	0.8114
0.50	45.4286	0.8418	0.7197

MOF-5(808) (Crystal density 0.594 g/cc)

Density (g/cc)	Excess (g/kg)	Density Fraction	Excess Fraction
0.30	59.8360	0.5051	1.0000
0.35	59.4718	0.5892	0.9939
0.40	57.7560	0.6734	0.9652
0.45	56.7518	0.7576	0.9485
0.50	47.5289	0.8418	0.7943

MOF-5(279) (Crystal density 0.594 g/cc)

Density (g/cc)	Excess (g/kg)	Density Fraction	Excess Fraction
0.30	60.2829	0.5051	1
0.35	58.9213	0.5892	0.9774
0.40	57.0589	0.6734	0.9465
0.45	51.6274	0.7576	0.8564
0.50	47.6547	0.8418	0.7905

MOF-5(2349):MOF-5(808), 7:1 mixture (Crystal density 0.594 g/cc)

Density (g/cc)	Excess (g/kg)	Density Fraction	Excess Fraction
0.28	55.4539	0.4714	1
0.31	55.4456	0.5219	0.9998
0.33	55.2989	0.5556	0.9972
0.35	55.1540	0.5892	0.9946
0.40	55.2989	0.6734	0.9972
0.45	53.5454	0.7576	0.9656
0.50	47.9200	0.8418	0.8641

MOF-5-O_h(600) (Crystal density 0.594 g/cc)

Density (g/cc)	Excess (g/kg)	Density Fraction	Excess Fraction
0.27	56.2580	0.4545	1
0.32	56.3276	0.5387	1.0012
0.35	55.6004	0.5892	0.9883
0.38	55.1372	0.6397	0.9801
0.41	55.3312	0.6902	0.9835
0.45	54.5163	0.7576	0.9690
0.48	48.0933	0.8081	0.8549

SI 6. Notes and References

1. Koh, K.; Van Oosterhout, J. D.; Roy, S.; Wong-Foy, A. G.; Matzger, A. J. Exceptional surface area from coordination copolymers derived from two linear linkers of differing lengths. *Chem. Sci.* 2012, *3*(8), 2429-2432.
2. Purewal, J.; Veenstra, M.; Tamburello, D.; Ahmed, A.; Matzger, A. J.; Wong-Foy, A. G.; Seth, S.; Liu, Y.; Siegel, D. J. Estimation of system-level hydrogen storage for metal-organic frameworks with high volumetric storage density, *Int. J. Hydrog. Energy* 2019, *44* (29), 15135-15145.
3. Note: While this article is believed to contain correct information, Ford Motor Company (Ford) does not expressly or impliedly warrant, nor assume any responsibility, for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, nor represent that its use would not infringe the rights of third parties. Reference to any commercial product or process does not constitute its endorsement. This article does not provide financial, safety, medical, consumer product, or public policy advice or recommendation. Readers should independently replicate all experiments, calculations, and results. The views and opinions expressed are of the authors and do not necessarily reflect those of Ford. This disclaimer may not be removed, altered, superseded or modified without prior Ford permission.