Project ID: ST144

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

Alauddin Ahmed,[1] Darpan Aulakh,[2] Justin Purewal,[3]
Antek Wong-Foy,[2] Adam Matzger,[2] Mike Veenstra,[3] and
**Don Siegel**[1] (PI)

[1]Mechanical Engineering Department and [2]Department of Chemistry, University of Michigan

[3]Ford Motor Company

DOE Annual Merit Review, June 15, 2018, Washington, DC

This presentation does not contain any proprietary, confidential, or otherwise restricted information

# Overview

## Timeline and Budget

**Project Start Date:** Sept 1st, 2017
**Project End Date:** August 31st, 2018

**Total Project Budget:** $1,047,000

**Federal Share:**

**UM**: $807,000
**Ford**: $192,000
**Total:** $999,000

$250,000 (Y1)
$398,000 (Y2)
$351,000 (Y3)

**Cost Share**: $48,000 (Ford)

**Total Funds Spent:*** ~$100,000

*Estimated as of 4/30/18

## Barriers

**Barriers addressed**

- Volumetric Density
- Gravimetric Density

## Partners

**Interactions/collaborations:**

Ford Motor Company, Hydrogen Storage Engineering Center of Excellence (HSECoE)

**Project lead:**

Don Siegel, University of Michigan

# Background (1)

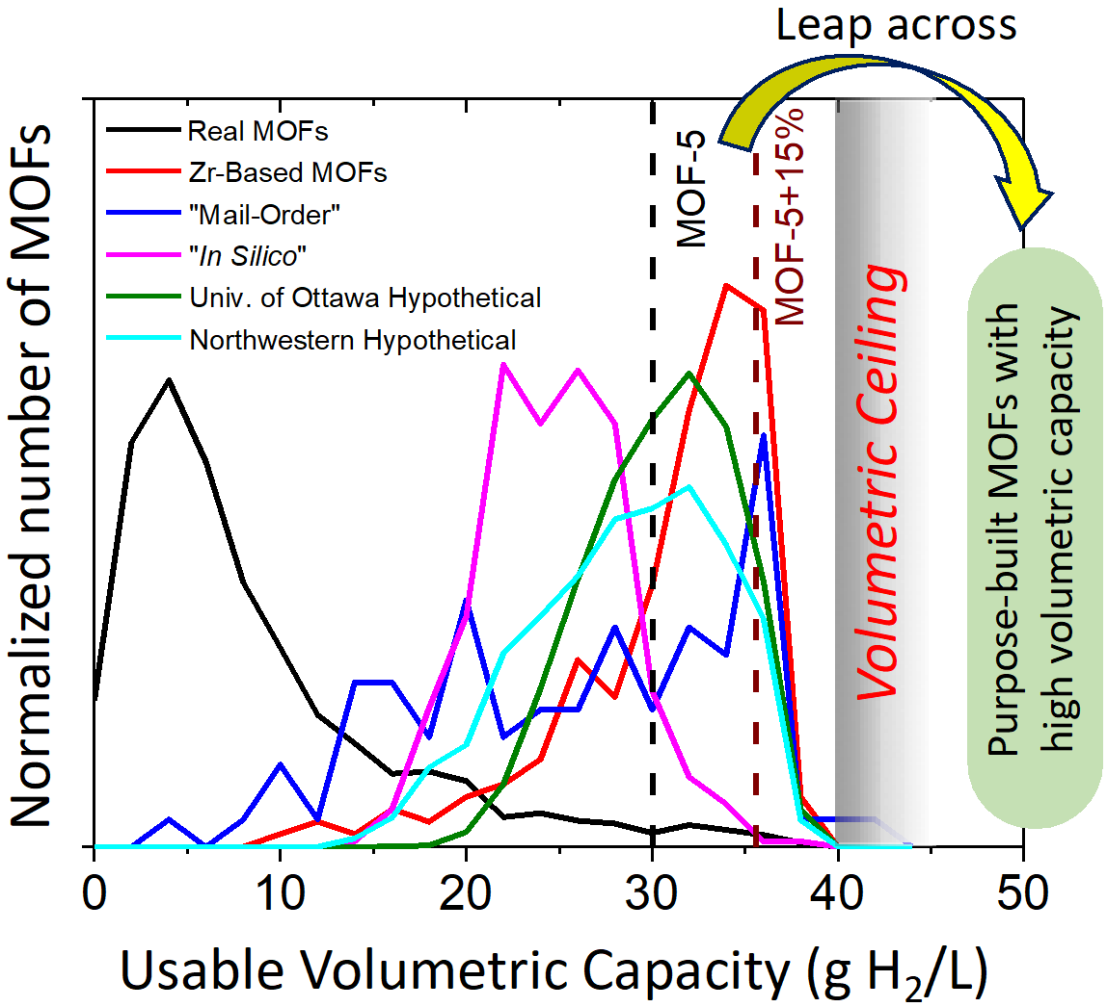
- A high-capacity, low-cost method for storing hydrogen remains one of the primary barriers to the widespread commercialization of fuel cell vehicles
- Storage via adsorption in MOFs is promising due to their fast kinetics, reversibility, and tunable properties
- A viable adsorbent must exhibit a high intrinsic (i.e., materials level) $H_2$ capacity, and pack in a dense fashion at the system level
  - Our prior screening (project ST122) revealed that no known MOF exhibits a usable volumetric capacity exceeding 40 g $H_2$/L (assuming an isothermal pressure swing between 100 and 5 bar at 77 K)
  - Analysis by the HSECoE has shown that inefficient materials packing can result in capacity reductions of more than 60% compared to the single-crystal level. These inefficiencies can negate improvements in volumetric performance achieved at the materials level
  - The present project addresses both of these challenges

**Project goal:** Overcome volumetric limitations associated with physisorptive hydrogen storage at both the materials and systems level in metal-organic frameworks (MOFs)

# Relevance (1)

Screening of ~500,000 MOFs reveals that essentially no compounds exceed 40 g/L usable capacity

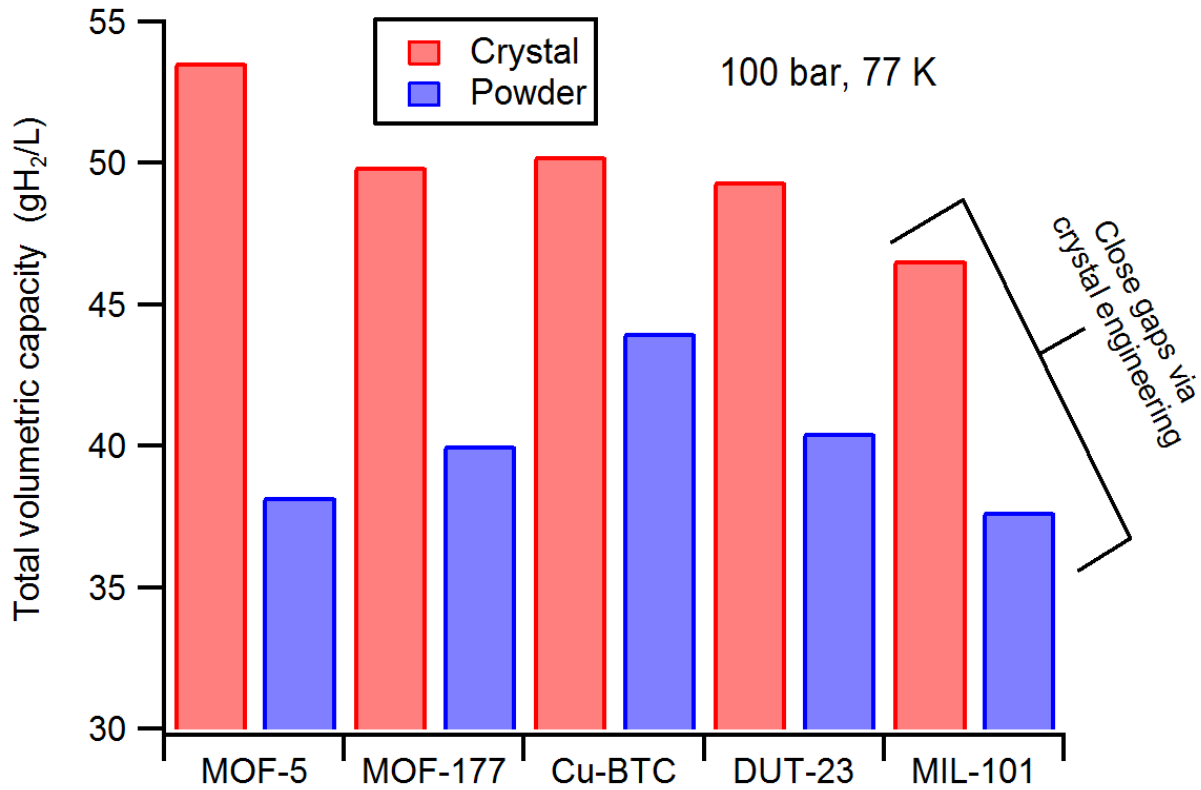
→ New MOFs needed to break through volumetric ceiling



# Relevance (2)

Packing inefficiencies result in significant volumetric penalties in adsorptive hydrogen storage systems

→ Increase packing density via crystal engineering




Data courtesy of Justin Purewal, Ford Motor Company

# Relevance (3)

**Objective 1:** Apply machine learning techniques to identify, design, and demonstrate high-capacity MOFs

- Demonstrate usable volumetric capacities exceeding 50 $H_2$ g/L (single-crystal/pressure swing)
- No compromise to gravimetric capacity, kinetic performance, or reversibility
- If successful, these compounds will set a new high-water mark for $H_2$ density in adsorbents at cryogenic conditions

**Objective 2:** Control MOF crystal morphology and crystallite size distribution to increase packing density

- Increase packing density of target high capacity MOF by at least 30% (compared to its powder tap density)
- Do so with less than 15% decrease in gravimetric performance

# Year 1 Milestones

| Milestone Summary Table | | | | | | | |
|---|---|---|---|---|---|---|---|
| Recipient Name: | | University of Michigan | | | | | |
| Project Title: | | Optimized Hydrogen Adsorbents via Machine Learning and Crystal Engineering | | | | | |
| **Task Number** | **Task or Subtask Title** | **Milestone Type** | **Milestone Number** | **Milestone Description** | **Milestone Verification Process** | **Quarter (from Start)** | **Status** |
| **1.0** | | | | | | | |
| 1.1 | MOF performance from scratch | Milestone | M1.1.1 | Demonstrate ability to predict usable capacity of an arbitrary MOF to within 85% of GCMC capacity using only crystal structure as input | Comparison of machine learning prediction with GCMC calculation | 1 | **Complete** |
| 1.2 | Structure-performance correlations | Milestone | M1.2.1 | Correlate MOF geometric properties with capacity | Random forest, latent variable, or support vector machine analysis of MOF properties | 2 | **Complete** |
| 1.3 | MOF reverse engineering | Go/No-Go | D1 | Identify ranges for 4 MOF crystallographic properties (surface area, density, pore volume, & porosity) consistent with usable volumetric capacity of at least 40 g/L and usable gravimetric capacity of at least 7 wt. % (assuming an isothermal pressure swing between 100 and 5 bar at 77 K) based on single crystal density. Demonstrate that the identified ranges are within the realm of possibility for the development of new MOFs, and thus provide a pathway for meeting the DOE storage targets. | Random forest or SVM analysis of MOF properties and direct GCMC simulation | 4 | **On Track** |
| **2.0** | | | | | | | |
| 2.1 | Morphological engineering | Milestone | M2.1.1 | Identify at least 2 additives capable of controlling morphology from cubes to octahedra | Optical microscopy or SEM measurements | 3 | **In progress – on schedule** |

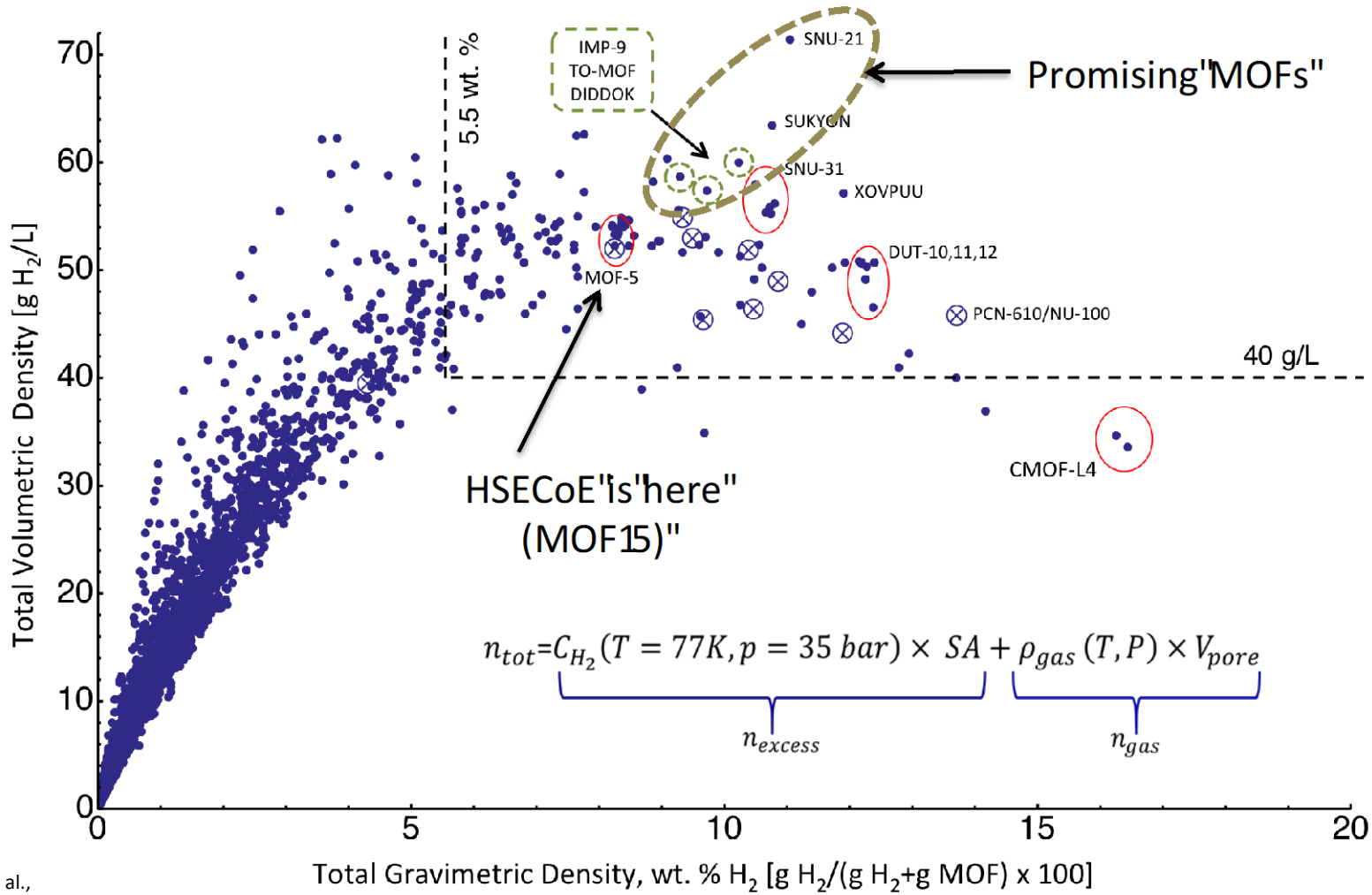
# Approach

**Notes:**

- Unless otherwise stated, all measurements and calculations are performed at T = 77 K.

# High-throughput Screening

**Prior work**: developed a database of MOFs by mining the CSD. Chahine rule and crystal structure were used to predict $H_2$ capacity in thousands of compounds




J. Goldsmith, et al., Chem. Mater., **25**, 3373 (2013).

# MOF Database

Compiled a MOF database of ~500,000 compounds

43,000+ MOFs assessed for temperature+pressure swing storage

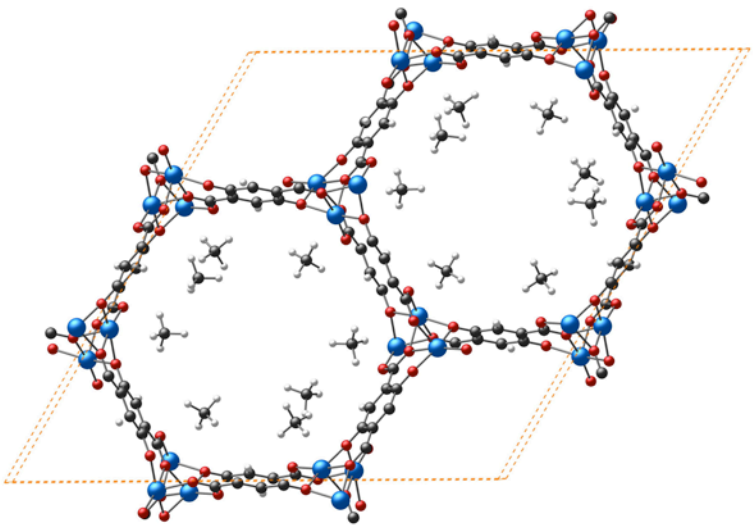
~100,000 MOFs assessed for pressure swing storage

| Source | Available in database | Zero surface area | $H_2$ capacity evaluated empirically | $H_2$ capacity evaluated with GCMC |
|---|---|---|---|---|
| UM+CoRE+CSD17 (RM) | 15,235 | 2,950 | 12,285 | 12,799 |
| Mail-Order MOFs (MO) | 112 | 4 | 108 | 112 |
| In Silico MOFs (IS) | 2,816 | 154 | 2,662 | 466 |
| In Silico Surface MOFs (ISS) | 8, 885 | 283 | 8,602 | 1,058 |
| MOF-74 Analogs (M74) | 61 | 0 | 61 | 61 |
| ToBaCCo (TB) | 13,512 | 214 | 13,298 | 290 |
| Zr-MOFs (ZR) | 204 | 0 | 204 | 204 |
| NW Hypothetical MOFs (NW) | 137,000 | 30,160 | 106,840 | 12,374 |
| UO Hypothetical MOFs (UO) | 324,500 | 32,993 | 291,507 | 16,372 |
| In-house synthesized via hypothetical design | 18 | 0 | 18 | 5 |
| **Total** | **493,458** | **66,758** | **435,585** | **43,741** |

**RM:** (a) **UM:** J.Goldsmith, A. G. Wong-Foy, M. J. Cafarella, and D. J. Siegel, *Chem. Mater.,* 25 , 3373–3382 (2013); (b) **CoRE:** Y. G. Chung, *et al.*, *Chem. Mater.*, 26, 6185–6192 (2014); (c) **CSD17**: P. Z. Moghadam et al., *Chem. Mater.,* 29, 2618–2625 (2017).
**MO:** R. L. Martin, L.-C. Lin, K. Jariwala, B. Smit, M. Haranczyk, *J. Phys. Chem. C* 117, 12159-12167 (2013);
**IS:** Y. Bao, R. L. Martin, M. Haranczyk, M. W. Deem, *J. Phys. Chem. C* 119, 186-195 (2015).
**ISS:** Y. Bao, R. L. Martin, C. M. Simon, M. Haranczyk, B. Smit, M. W. Deem, *Phys. Chem. Chem. Phys.,* 17, 11962-11973 (2015).
**M74:** M. Witman, S. Ling, S. Anderson, L. Tong, K.C. Stylianou, B. Slater, B. Smit, M. Haranczyk, *Chem. Sci.*, 7, 6263-6272 (2016).
**TB:** Y. J. Colón, D. A. Gómez-Gualdrón, and R. Q. Snurr, *Cryst. Growth Des.,* 17, 5801–5810 (2017).
**ZR:** D. A. Gómez-Gualdrón, O.V. Gutov, V. Krungleviciute, B. Borah, J. E. Mondloch, J. T. Hupp, T. Yildirim, O.K. Farha, R.Q. Snurr, *Chem. Mater.* 26, 5632-5639 (2014).
**NW:** C. E. Wilmer, M. Leaf, C. Y. Lee, O. K. Farha, B. G. Hauser, J. T. Hupp, R. Q. Snurr, *Nat. Chem.* 4, 83–89 (2012).
**UO:** M. Z. Aghaji, M. Fernandez, P. G. Boyd, T. D. Daff, and T. K. Woo, *Eur. J. Inorg. Chem.,* 2016, 4505–4511 (2016).

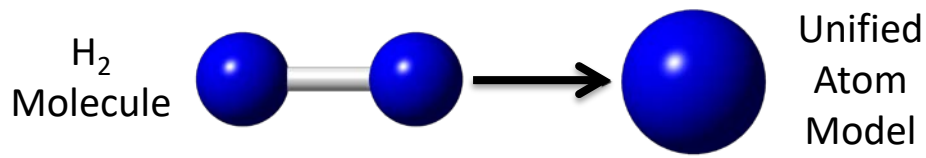
# Grand Canonical Monte Carlo

- GCMC = atomistic method that calculates the total amount of $H_2$ (adsorbed + gas phase) contained within the pore space of a MOF at given T, P
- Does not rely on empirical correlations such as the Chahine-rule



Example GCMC simulation of $CH_4$ adsorption in Ni-DOBDC at 298 K and 35 bar

- Calculations employ the MGS* and the **Pseudo-FH**** unified atom models for $H_2$-MOF interactions
- MOF atoms are fixed




$$U_{ij}(r_{ij}) = 4\varepsilon\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]$$

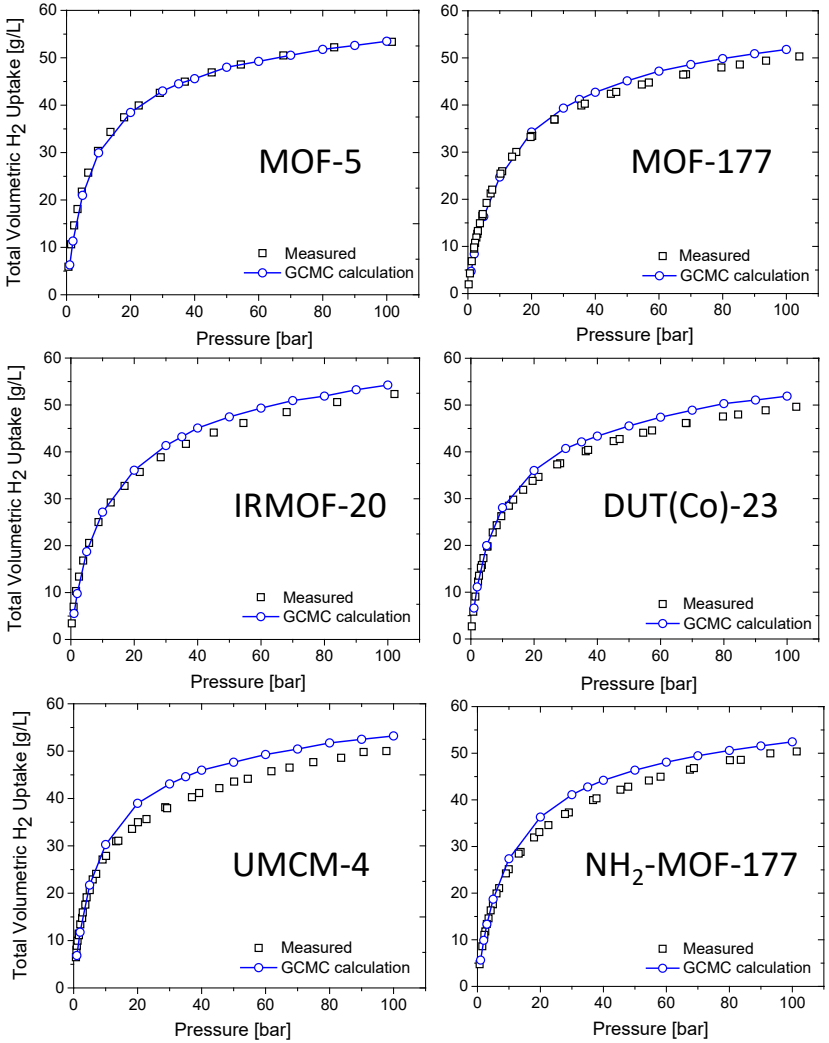
| Force Field | Sigma (Å) | Epsilon/$k_B$ (K) |
|---|---|---|
| MGS | 2.958 | 36.7 |
| Pseudo-FH | 3.064 | 30.1 |

*Michels, de Graaff and Seldam, *Physica*, **1960**, 26, 393; Ryan, Broadbelt, and Snurr, *Chem. Comm*. 2008, 4134
**Fischer, Hoffmann, Fröba, *ChemPhysChem*, **2009**,10, 2647.

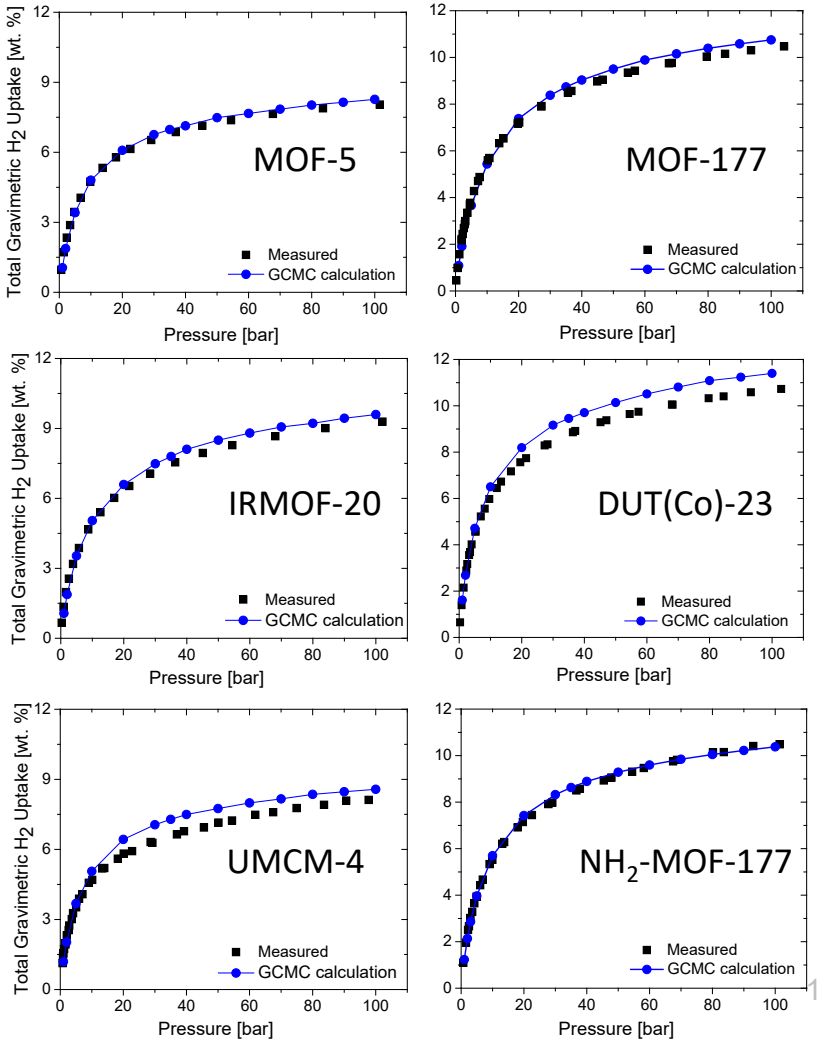
# Examples of Simulated Isotherms

GCMC isotherms calculated with the pseudo-Feynman-Hibbs interatomic potential are in very good agreement with our measured isotherms

## Total Volumetric $H_2$ Uptake



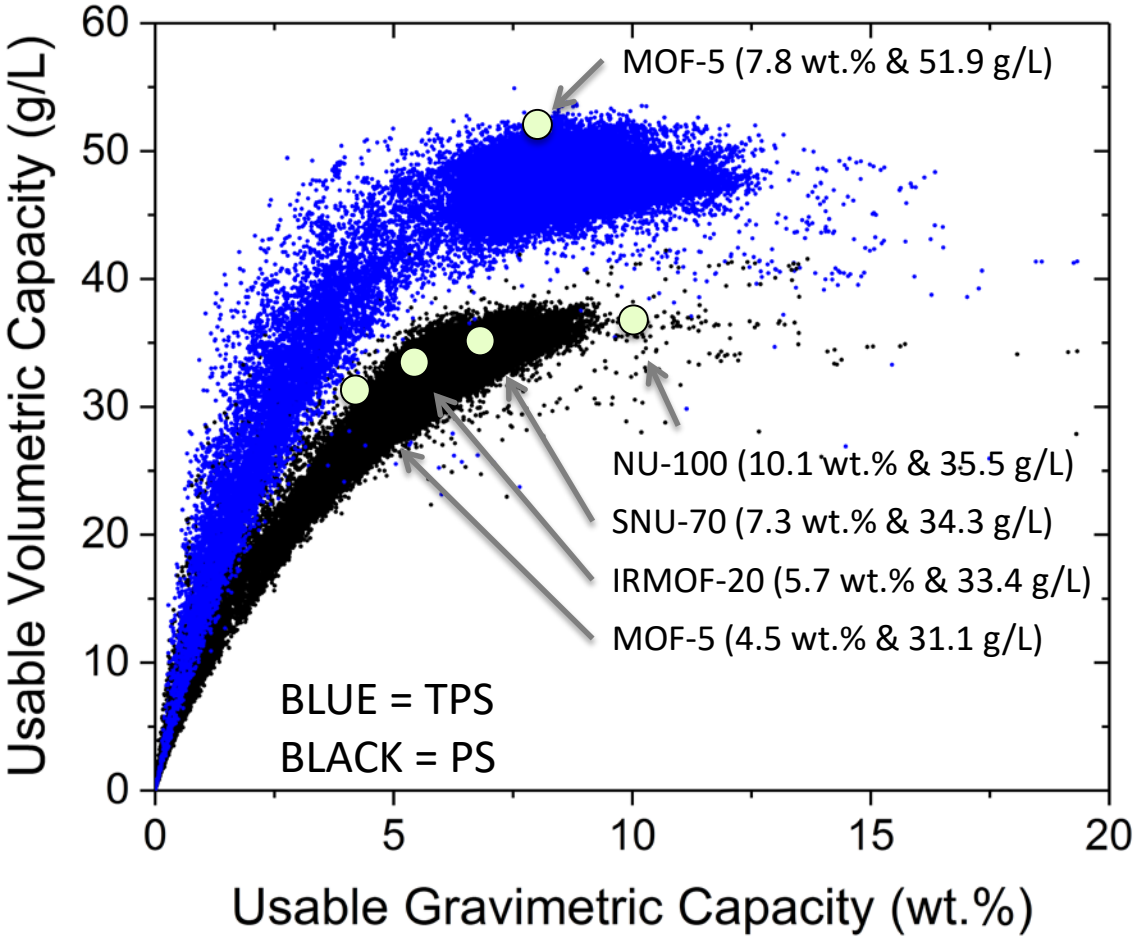

## Total Gravimetric $H_2$ Uptake



# High-Throughput Screening

## Predicted usable $H_2$ capacities for PS and TPS conditions

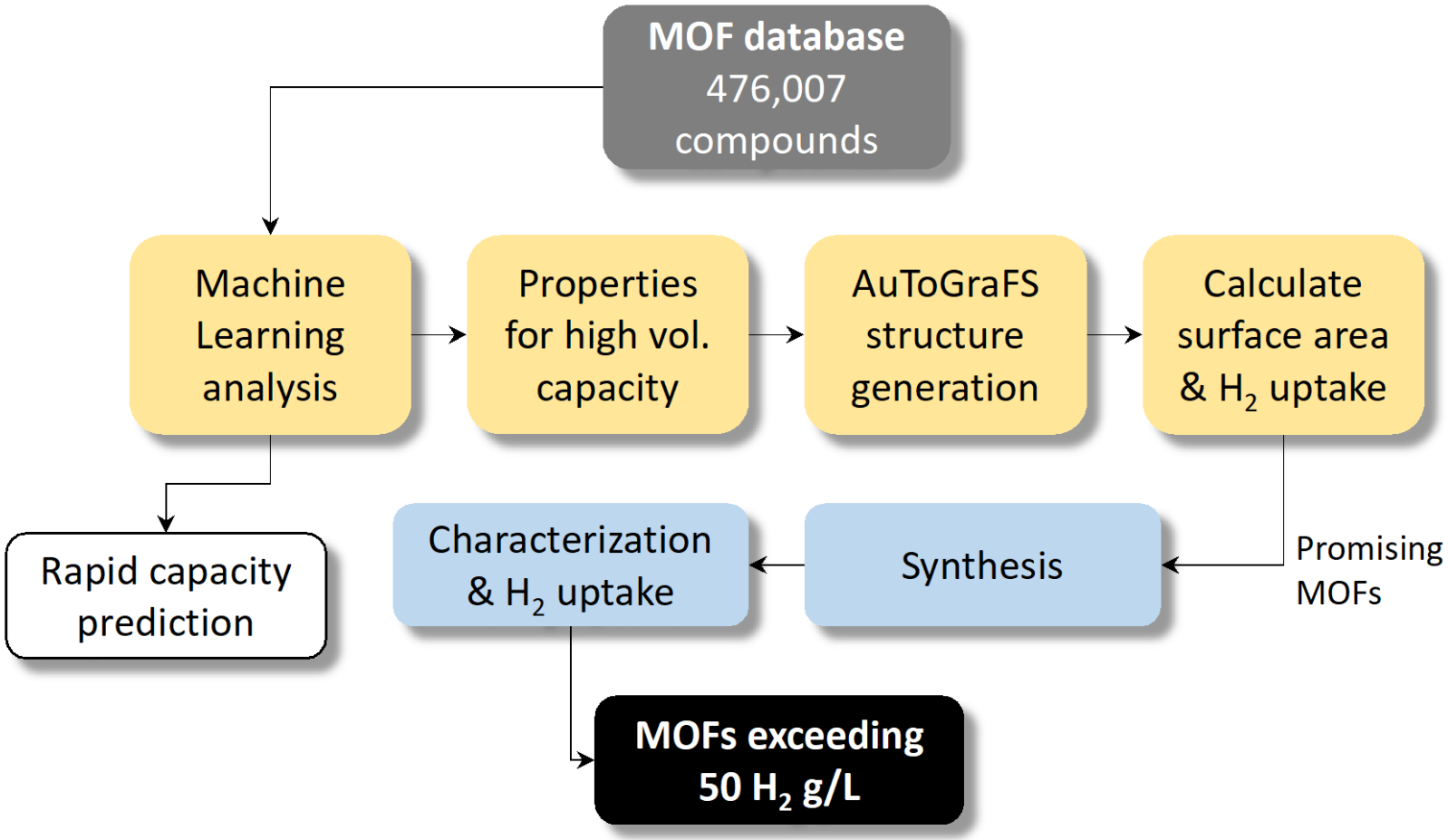
- Pressure swing: $P_{max}$= 100 bar to $P_{min}$= 5 bar at 77 K
- Temp+pressure swing: $T_{min}$= 77 K, $P_{max}$= 100 bar to $T_{max}$= 160 K, $P_{min}$= 5 bar




Only 180 MOFs surpass MOF-5 under TPS conditions.

# Concept

Machine learning will be used to guide the development of MOFs with high volumetric $H_2$ capacities



# Crystal Engineering

- Packing of congruent convex objects indicates that particle morphology and the size distribution are key factors in determining packing efficiency
- We shall vary these properties systematically, leveraging advances in colloid science for the controlled growth of MOFs with various shapes and sizes

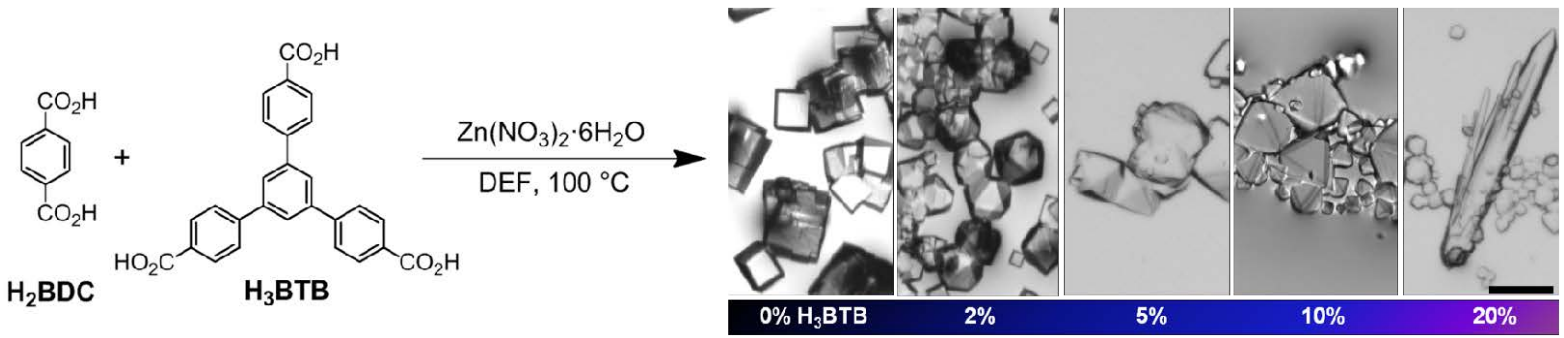
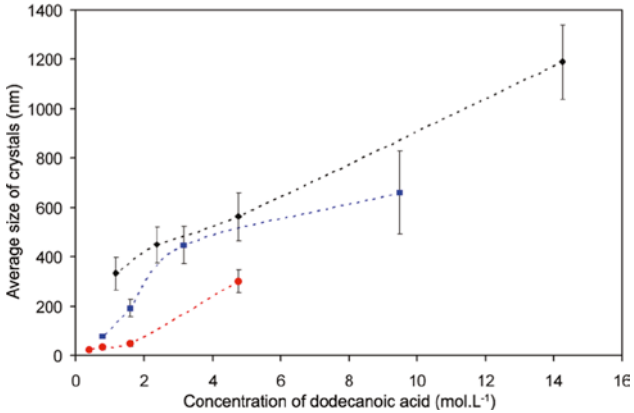



**Fig. 1:** Synthesis of octahedral-shaped MOF-5 crystals by addition of $H_3BTB$ in the reaction mixture of $H_2BDC$ and $Zn(NO_3)_2{\cdot}6H_2O$. Photographs show the dependence of crystal morphology on the percentage of $H_3BTB$ (scale bar: 100 μm). Another phase (needle shaped UMCM-1) appears at 10 mol% $H_3BTB$. From Matzger et al., JACS (2011) **133,** 20138

**Fig. 2:** Average size of HKUST-1 crystals as a function of dodecanoic acid concentration taken at longer and longer times. Colors represent different concentration of dodecanoic acid. From Diring, et al., Chem. Mater., (2010) **22**, 4531



# Accomplishments and Progress

# Input Features & Output Properties

| 7 Crystallographic Features | $H_2$ Adsorption at 10 Conditions |
|---|---|
| **Geometrical Features from Zeo++** | **$H_2$ Capacity from GCMC** |
| 1. **Pore Volume** (PV, $cm^3/g$) | 1. **TG** and **TV** at at 5 bar & 77K. |
| 2. **Void Fraction** (VF) | 2. **TG** and **TV** at 35 bar & 77K |
| 3. **Gravimetric Surface Area** (GSA, $m^2/g$) | 3. **TG** and **TV** at 100 bar & 77K |
| 4. **Volumetric Surface Area** (VSA, $m^2/cm^3$) | 4. **UG** and **UV** for Pressure Swing between 5 and 35 bar at 77 K. |
| 5. **Largest Cavity Diameter** (LCD, Å) | 5. **UG** and **UV** for Pressure Swing between 5 and 100 bar at 77 K. |
| 6. **Pore Limiting Diameter** (PLD, Å) | |
| 7. **Density** (D, $g/cm^3$) | |

**TG & TV** = Total gravimetric and volumetric capacity.
**UG & UV** = Usable gravimetric and volumetric capacity

*Zeo++: T.F. Willems, C.H. Rycroft, M. Kazi, J.C. Meza, and M. Haranczyk, *Algorithms and tools for high-throughput geometry- based analysis of crystalline porous materials*, Microporous and Mesoporous Materials, 149 ,134-141 (2012).

# ML Methods Tested

12 supervised learning methods from 5 different categories were used

**Decision Trees (DT)**

L. Breiman, J. Friedman, R. Olshen, and C. Stone. Classification and Regression Trees. Wadsworth, Belmont, CA, 1984.

**Random Forest (RF)**

L. Breiman, "Random Forests", Machine Learning, 45(1), 5-32, 2001.

**Bagging with DT; Bagging with RF (Bagging)**

L. Breiman, "Bagging predictors", Machine Learning, 24(2), 123-140, 1996.

**Boosted DT; Ada Boost with RF (Ada Boost)**

H. Drucker. "Improving Regressors using Boosting Techniques", 1997

**Gradient Boosting**

J. Friedman, Greedy Function Approximation: A Gradient Boosting Machine, The Annals of Statistics, Vol. 29, No. 5, 2001.

**Extremely Randomized Trees**

P. Geurts, D. Ernst., and L. Wehenkel, "Extremely randomized trees", Machine Learning, 63(1), 3-42, 2006.

Ensemble Methods

**K-Nearest Neighbors (K-NN)**

N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression". The American Statistician. 46(3), 175–185, 1992.

**Support Vector Machine (SVM)**

A. J. Smola, B. Schölkopf, "A Tutorial on Support Vector Regression", Statistics and Computing archive, 14(3), 199-222, 2004.

**Linear Regression; Ridge Regression (Generalized Linear Model)**

T. Hastie, R. Tibshirani and J. Friedman. Elements of Statistical Learning Ed. 2, Springer, 2009.

**Machine Learning Software & Code**: Scikit-learn, R, & in-house code

**Hyperparameter Optimization Method:** Grid search cross validation method

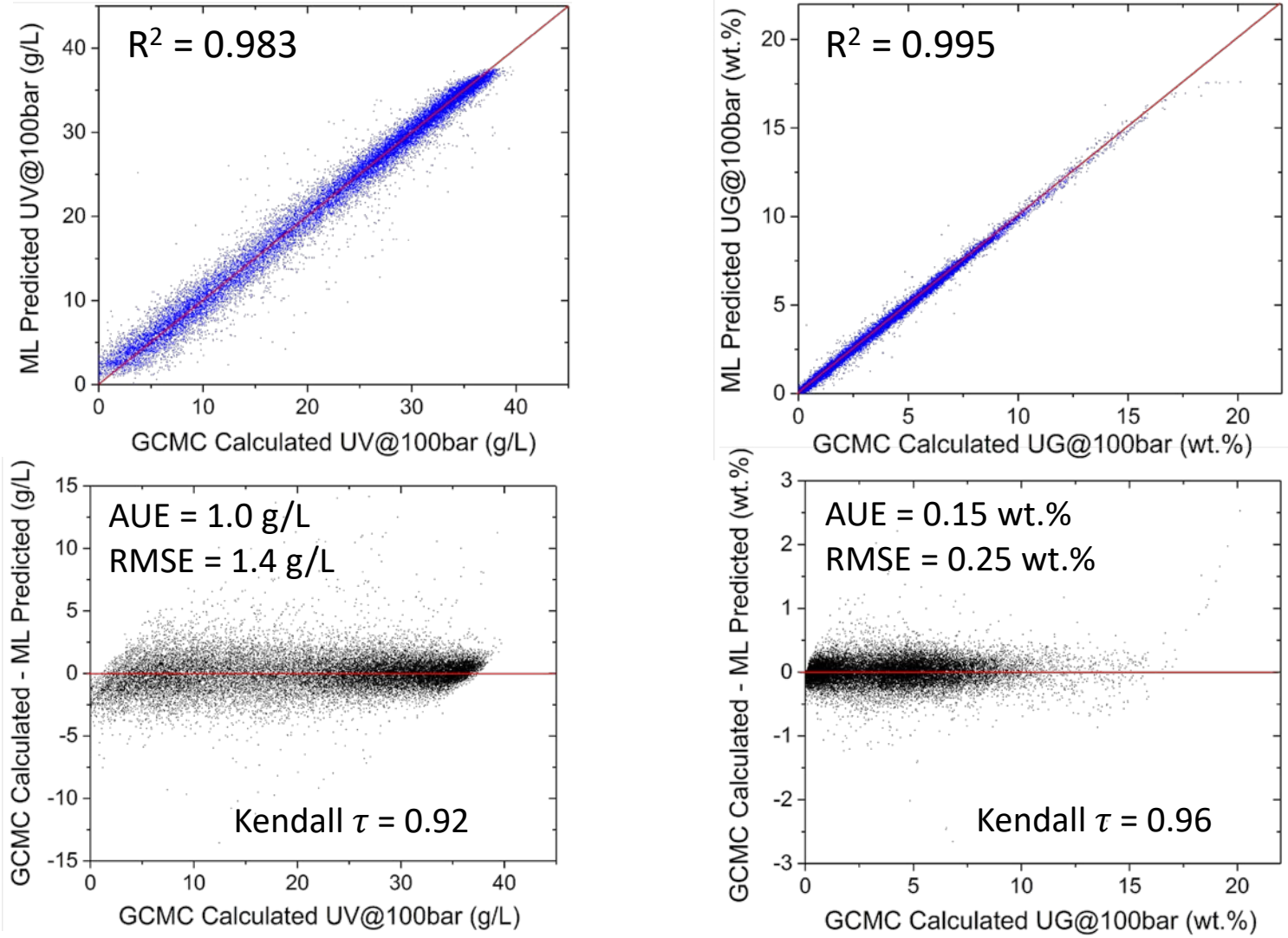
**Training Set: 74, 221; Test Set: 24,741; Unseen Data Set: 394,496**



- Scikit-learn: Pedregosa *et al.*, Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research, 12, 2825-2830, 2011.
- R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

# ML Predictions of Capacities

Comparison between Extremely Randomized Trees ML prediction and GCMC

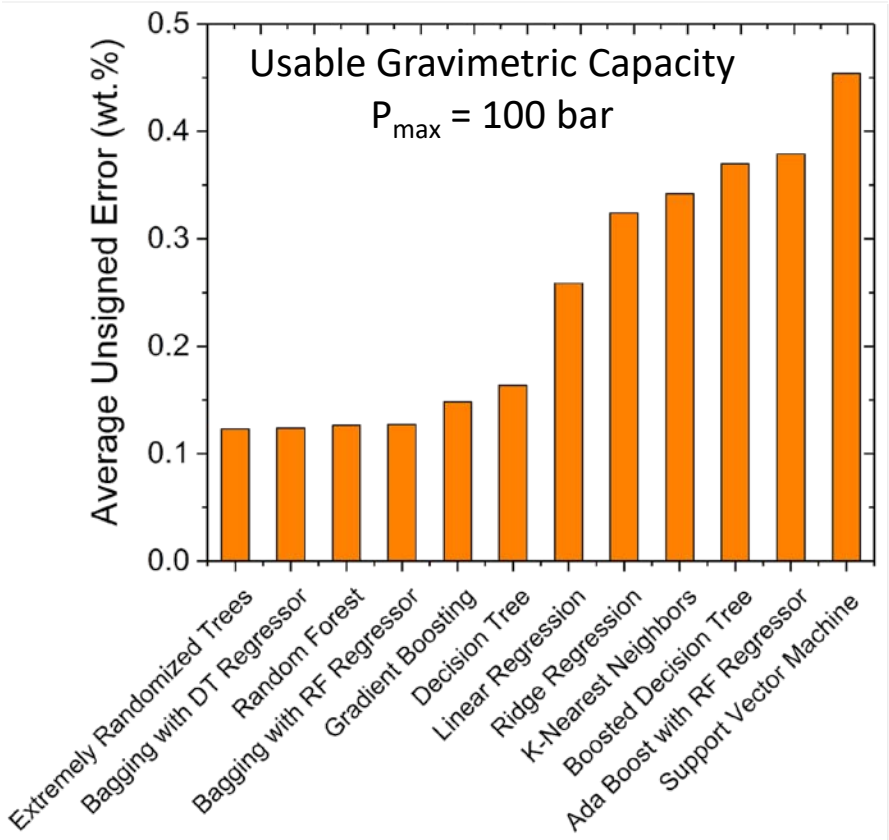


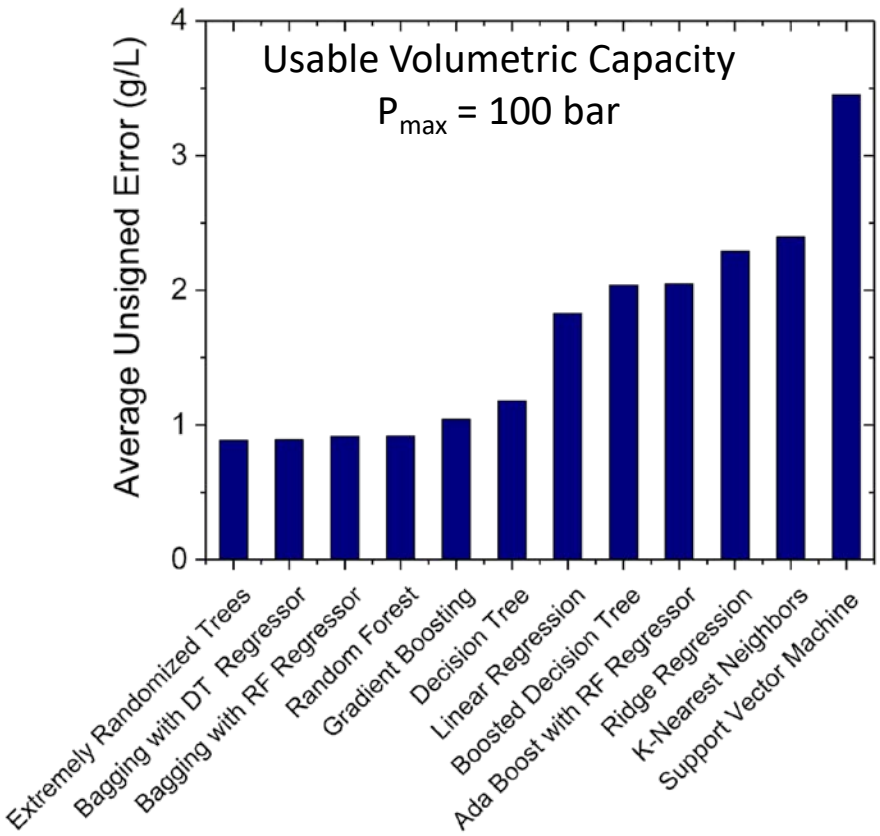
# Benchmarking ML Methods

The Extremely Randomized Trees method is the best performing ML algorithm

# Comparison of ML Methods

The Extremely Randomized Trees method is the best performing ML algorithm

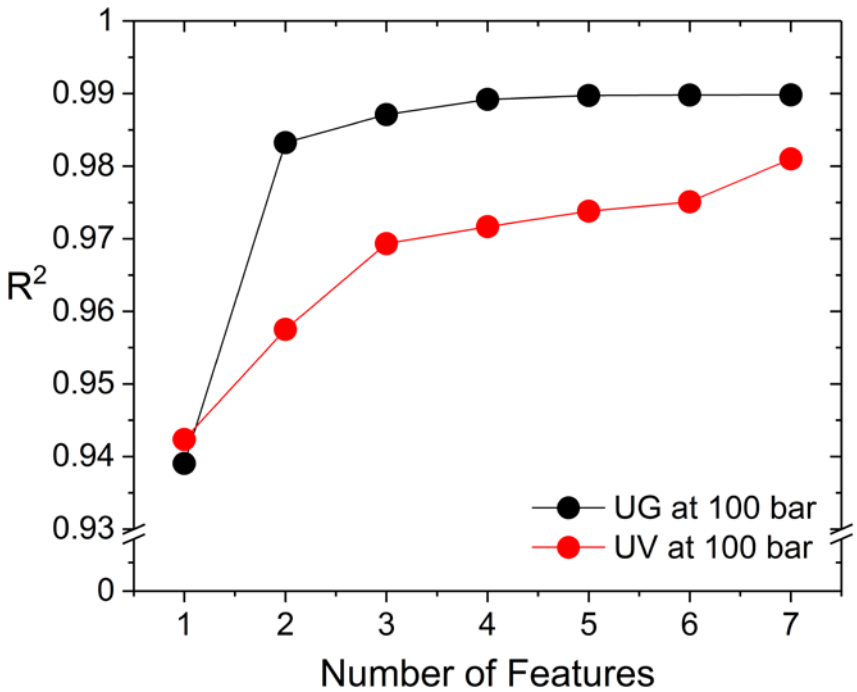


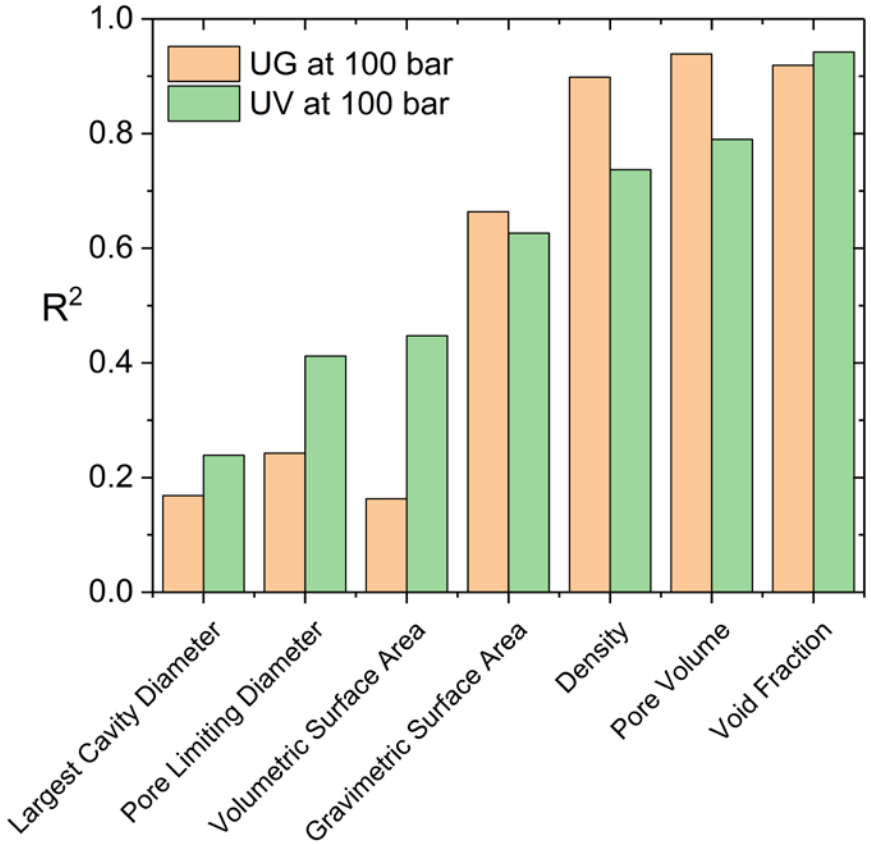
# Structure-Property Correlations

- Void fraction shows the strongest correlation with UV
- Only 4 features needed to predict UV with over 96% accuracy

**Effect of Number of Features**




**Single Feature Performance**

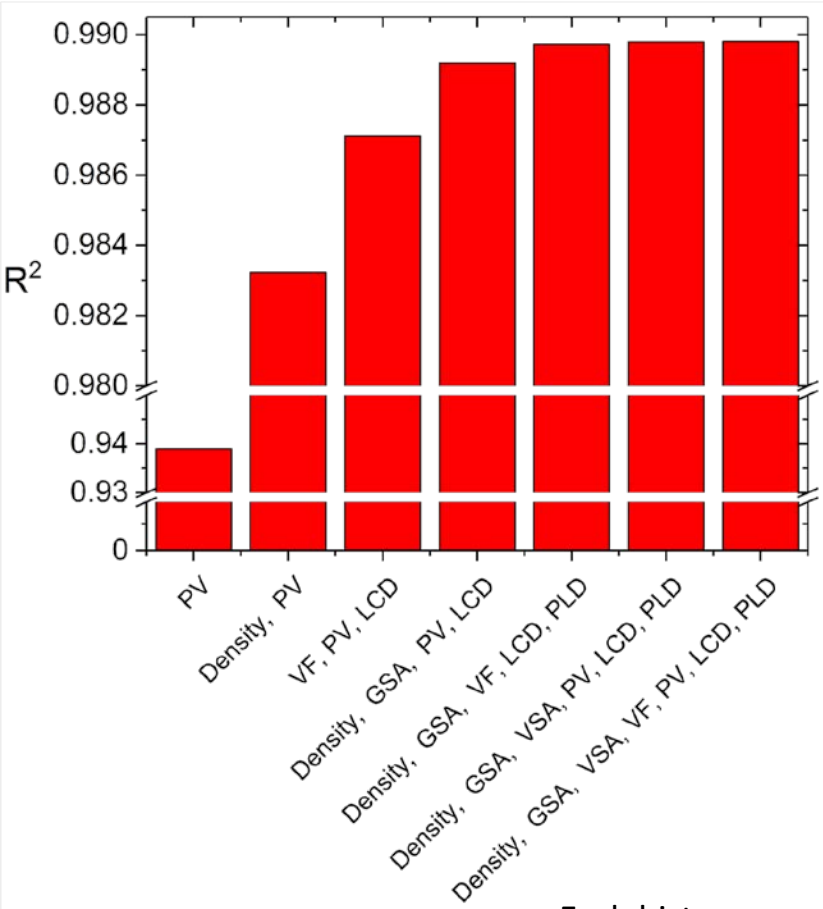



Each point on the plot represents the highest $R^2$ value among all possible($2^n$ -1) combinations of (n = 1,2,3,4,5,6,7) features

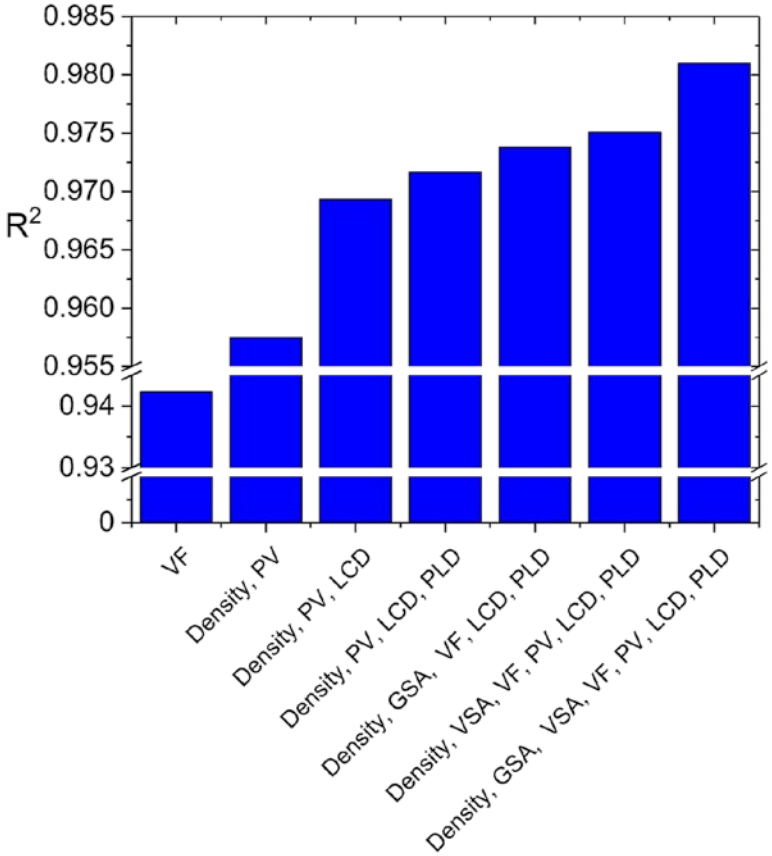
# Structure-Property Correlations

ML models were developed for all possible ($2^7$-1 = 127) combinations of features to identify the optimal feature set

**Usable Gravimetric**



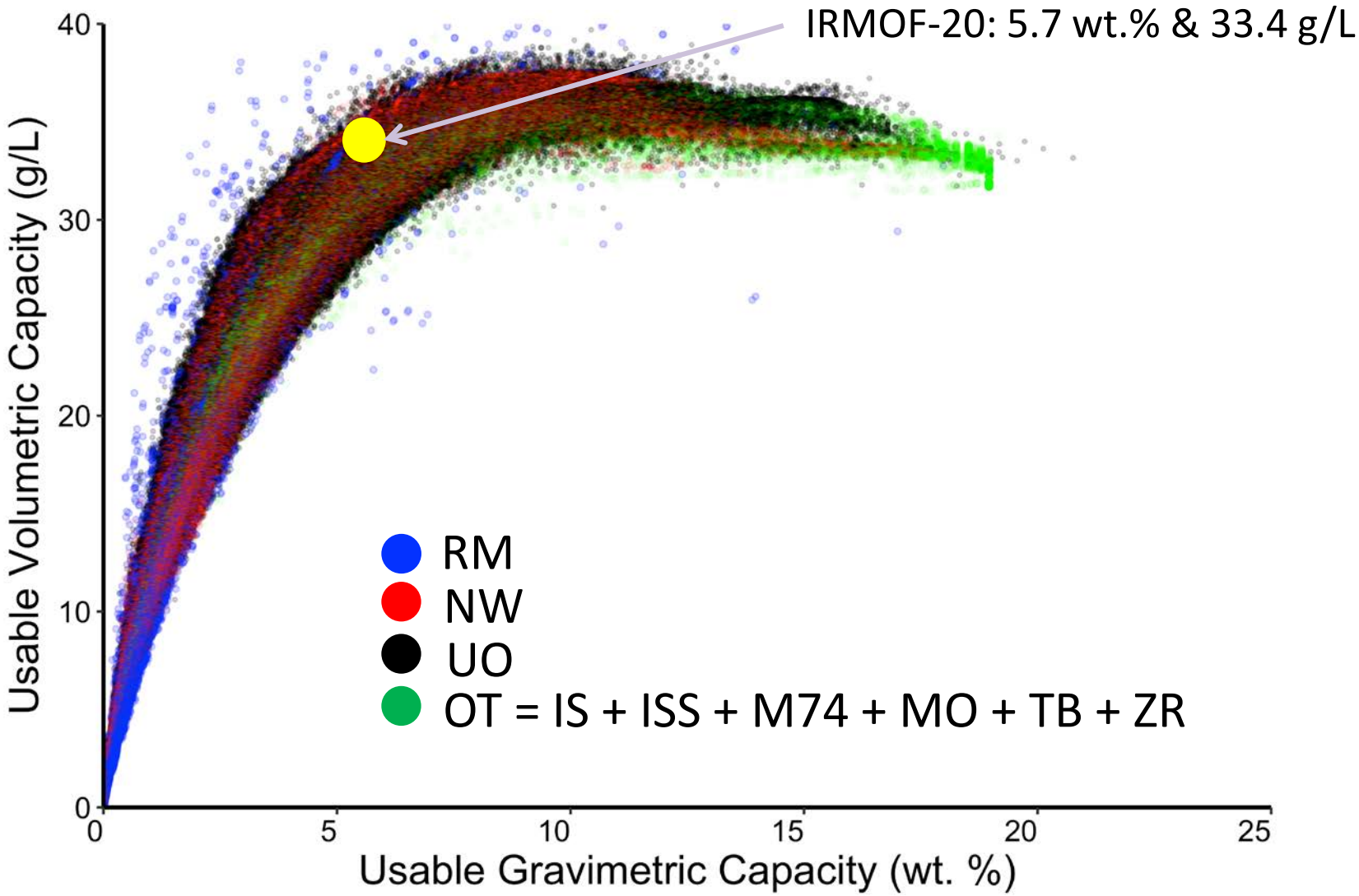

**Usable Volumetric**




Each histogram represents the highest $R^2$ value among all possible combinations of a given number of features.

# $H_2$ Storage in 500,000 MOFs

ML reveals 69,363 MOFs that can potentially out-perform IRMOF-20, the top performing MOF on a volumetric basis



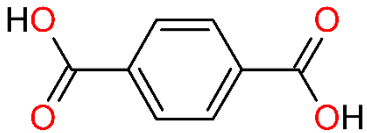
# Crystal Size Control (1)

Successfully developed synthesis protocols that produce MOF-5 with different average sizes

## Synthesis Protocols

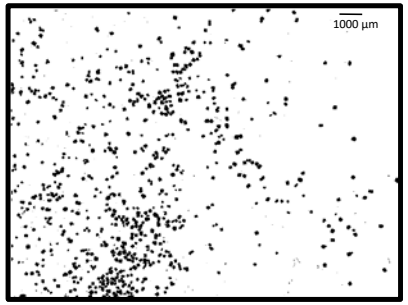
**Varying Metal:Linker ratio**

$Zn(NO_3)_2 \cdot 6H_2O$ +



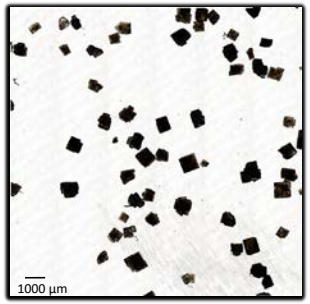

DEF, 100 °C (12hr)

**M:L=1.5**



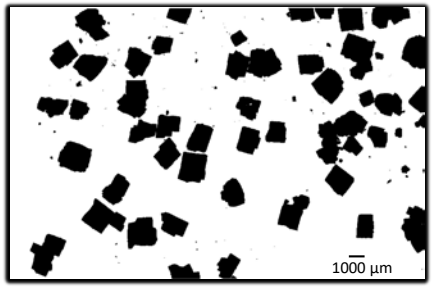

**100-400 µm**

**M:L=2.5**




**500-1000 µm**

**M:L=3.5**



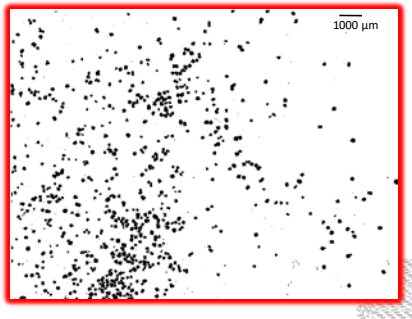

**1300-1900 µm**

**Agitating reaction mixture** — M:L=1.5 →
- 60 rpm → **100-250 µm**
- 100 rpm → **50-150 µm**

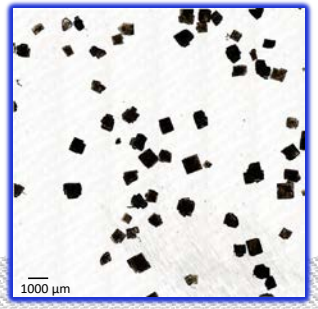
# Crystal Size Control (2)

Crystal size histograms indicate the successful control over crystal size

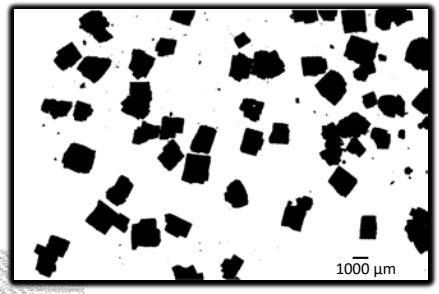
Varying the size distribution of cubic MOF-5 crystallites



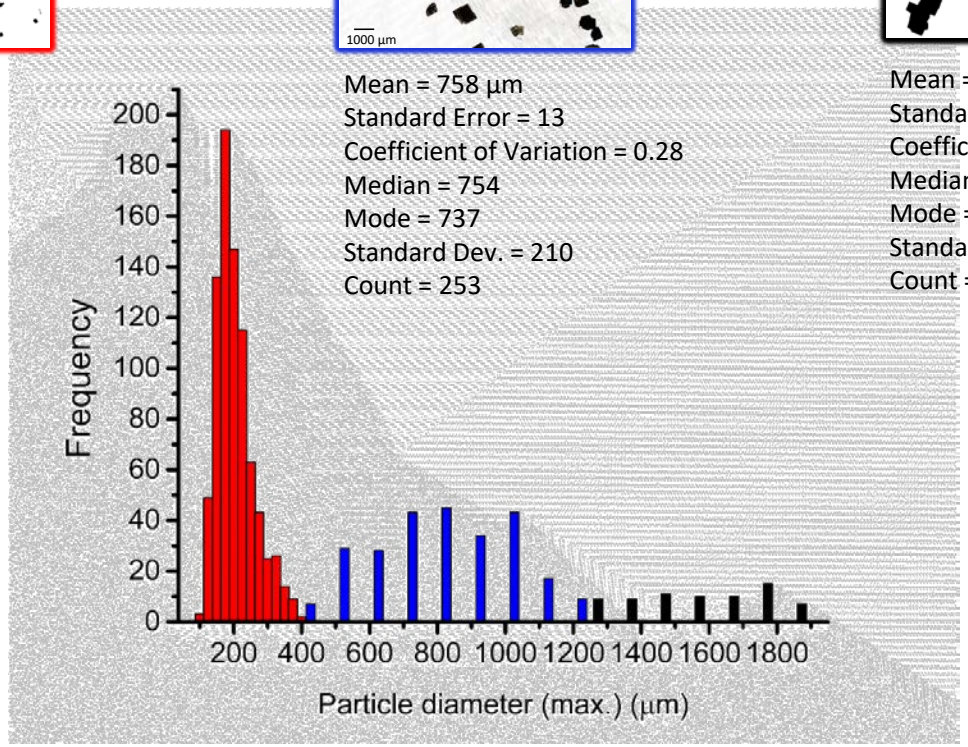

Mean = 192 µm
Standard Error = 1.9
Coefficient of Variation = 0.29
Median = 179
Mode = 174
Standard Dev. = 56
Count = 828




Mean = 758 µm
Standard Error = 13
Coefficient of Variation = 0.28
Median = 754
Mode = 737
Standard Dev. = 210
Count = 253



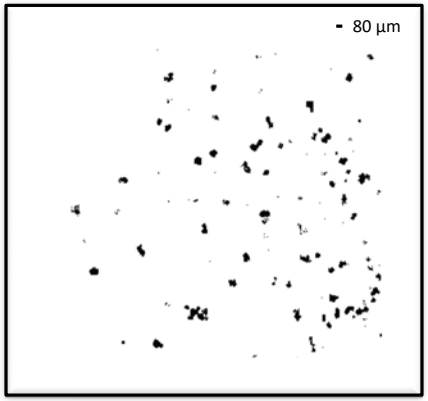

Mean = 1633 µm
Standard Error = 39
Coefficient of Variation = 0.24
Median = 1590
Mode = 1675
Standard Dev. = 389
Count = 97

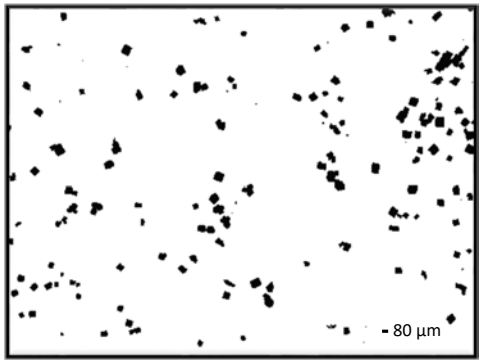


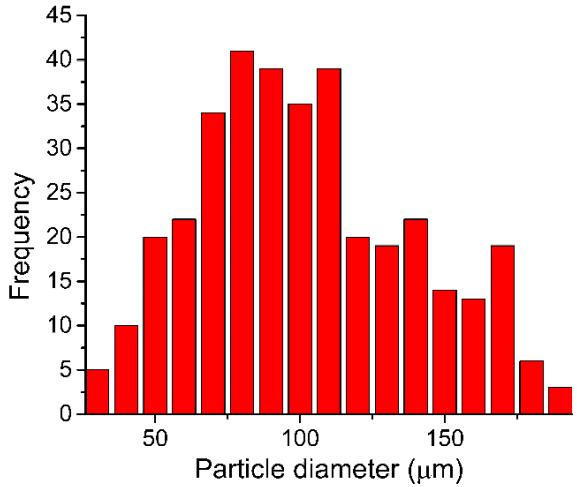
# Crystal Size Control (3)

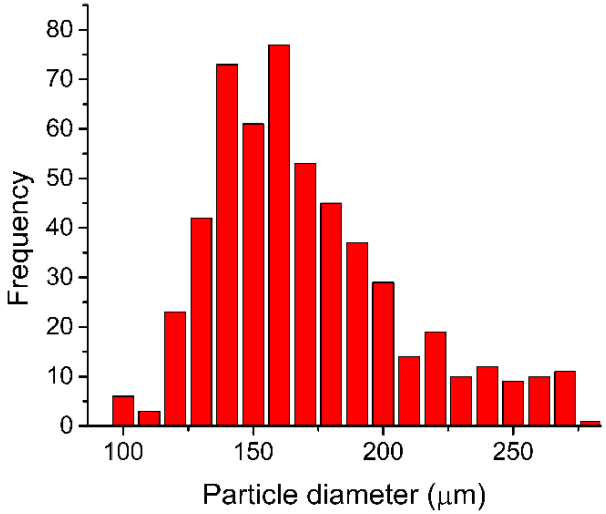
Crystal size histograms indicate the successful control over crystal size

Varying the size distribution of cubic MOF-5 crystallites



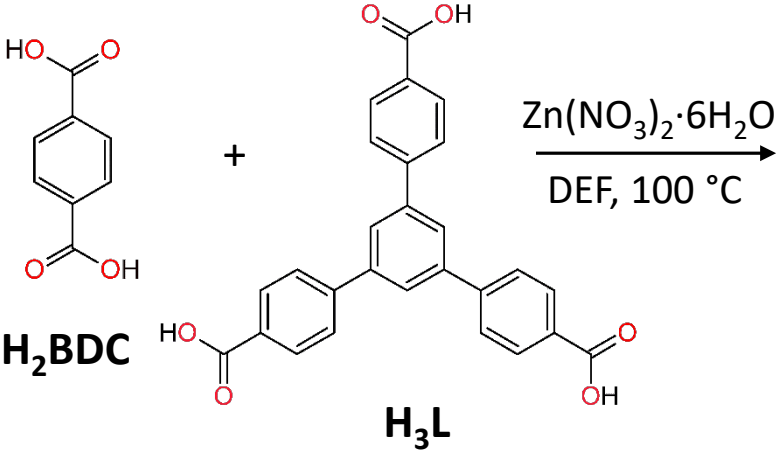



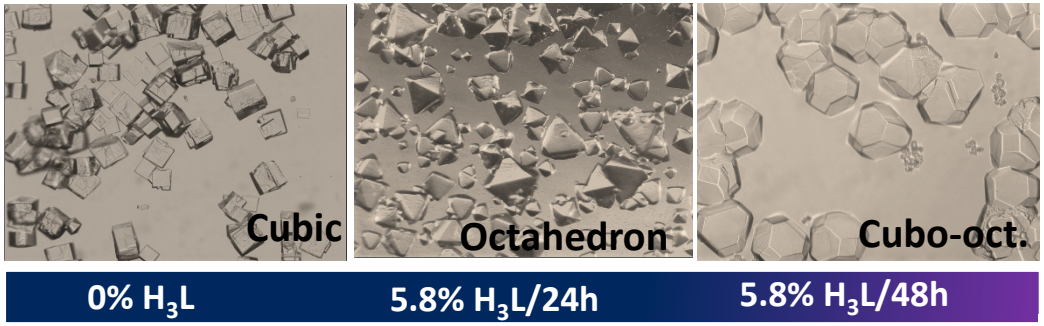



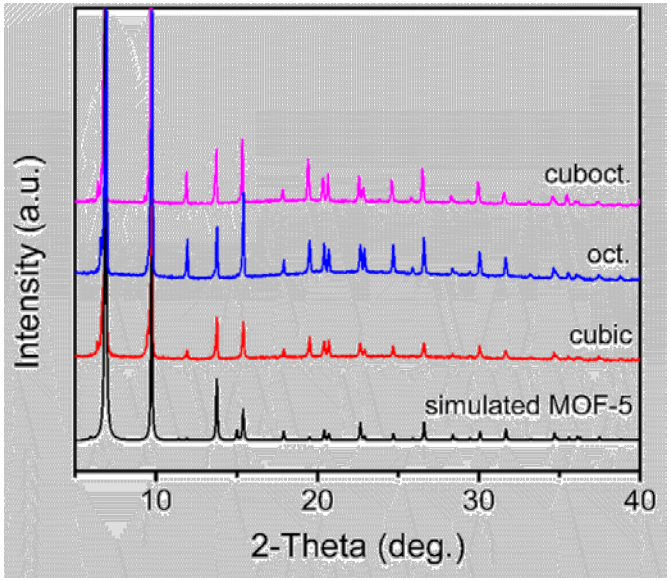
# Morphology Control (1)

Identified additive capable of controlling morphology of MOF-5 crystals



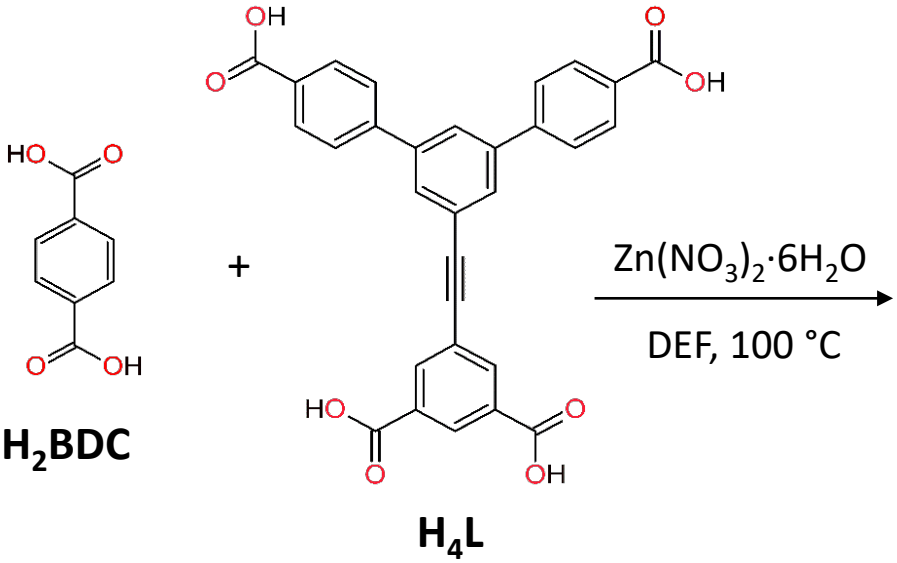



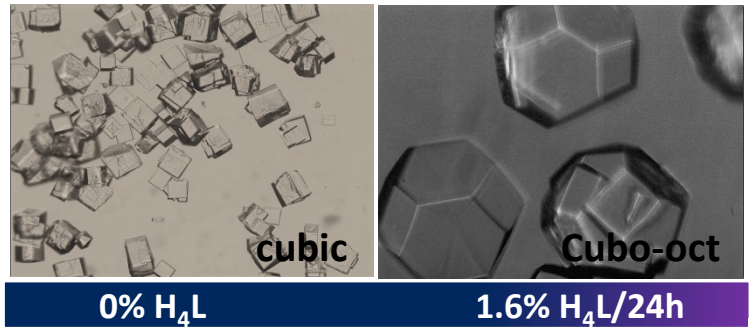

Phase purity

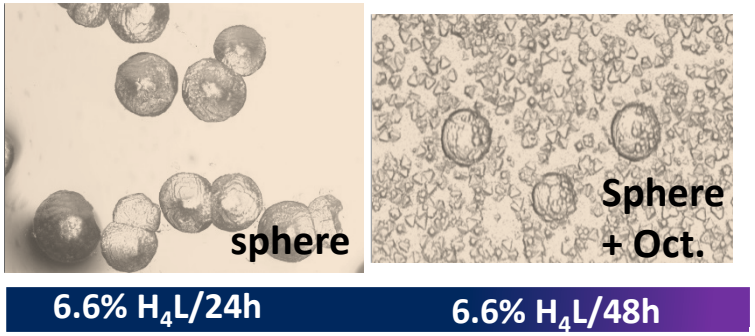


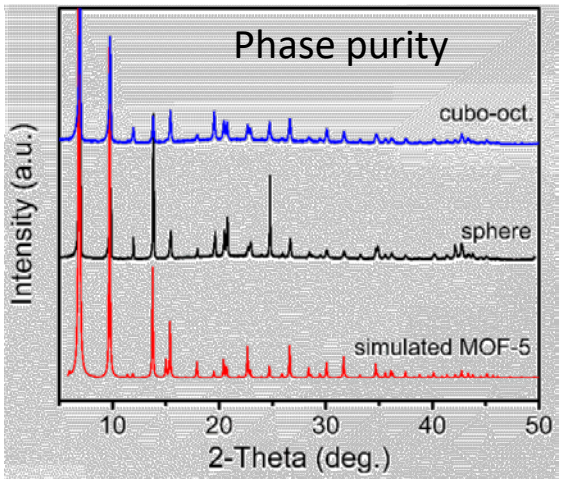
# Morphology Control (2)

Identified additive capable of controlling morphology of MOF-5 crystals




**$H_4L$**=5'-((3,5-dicarboxyphenyl)ethynyl)-[1,1':3',1''-terphenyl]-4,4''-dicarboxylic acid)

# Collaborations

## University of Michigan, Mechanical Engineering

- Atomistic simulation and project management

## University of Michigan, Dept. of Chemistry

- Synthesis and characterization of targeted MOFs

## Ford Motor Company (sub-contractor)

- PCT measurements
- Materials augmentation, characterization, scale-up, and system modeling

## HSECoE/SRNL (unfunded collaborator)

- Assistance with system models (David Tamburello)

# Challenges and Barriers

- Many more compounds identified by computation than can be synthesized
  - Assessment by a human is needed before synthesis can proceed
  - This is a bottleneck
- Structure collapse or incomplete solvent removal during activation
  - “Can it be made?”
  - Failure to achieve expected surface area and porosity
  - Properties that control “synthesizability” are not well-understood
- Incorrect, incomplete, or disordered crystal structure data
  - Garbage in, garbage out
  - False positives in screening

# Potential Future Work

- **Pass 1st go/no-go milestone:** Identify ranges for 4 MOF crystallographic properties consistent with usable volumetric capacity of at least 40 g/L and usable gravimetric capacity of at least 7 wt. %
- **Particle Size Control:** Determine if particle size influences packing efficiency by more than 10% for particles whose size varies by more than 10x
- **Hybrid Approaches to Space Filling:** Identify binary mixtures for a given morphology that yields the highest packing efficiency

Any proposed future work is subject to change based on funding levels

# Summary

- **Goal**: Overcome volumetric limitations associated with physisorptive hydrogen storage at both the materials and systems level in metal-organic frameworks (MOFs)
- **Approach**:
  - Control MOF crystal morphology and crystallite size distribution to increase packing density
  - Apply machine learning techniques to identify, design, and demonstrate high-capacity MOFs
- **Accomplishments:**
  - Benchmarked 12 ML algorithms for their ability to predict $H_2$ storage. The Extra Trees algorithm was found to be highly accurate; used to screen 500,000 MOFs. ML predictions are accurate because the input features are "good"
  - Discovered ~70,000 MOFs capable of outperforming IRMOF-20, the top performing MOF reported to date
  - Successfully developed synthesis protocols that produce MOF-5 with different average sizes. Identified additive capable of controlling morphology of MOF-5 crystals

UNIVERSITY OF MICHIGAN
umich.edu/~djsiege
djsiege@umich.edu

# The Team

# Technical Backup Slides