

Computational Materials Exploration for Capturing Carbon from the Atmosphere

by

Paul Boone

Bachelor of Science, Science and Technology in a Global Environment,
San José State University, 2002

Submitted to the Graduate Faculty of the
Swanson School of Engineering in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

University of Pittsburgh

2022

UNIVERSITY OF PITTSBURGH

SWANSON SCHOOL OF ENGINEERING

This dissertation was presented

by

Paul Boone

It was defended on

July 6, 2022

and approved by

James McKone, Ph.D., Assistant Professor,
Department of Chemical & Petroleum Engineering

Giannis Mpourmpakis, Ph.D., Associate Professor,
Department of Chemical & Petroleum Engineering

Jan Steckel, Ph.D., Research Scientist,
National Energy Technology Laboratory

Dissertation Director: Christopher E. Wilmer, Ph.D., Associate Professor,
Department of Chemical & Petroleum Engineering

Copyright © by Paul Boone

2022

Computational Materials Exploration for Capturing Carbon from the Atmosphere

Paul Boone, PhD

University of Pittsburgh, 2022

Over 1 trillion tons of CO₂ have been emitted into the atmosphere since we passed the concentration threshold for what is widely considered to be livable: 350 ppm CO₂. While numerous negative emissions technologies have been proposed addressing this, in this dissertation I will be focusing solely on extracting CO₂ from the atmosphere, or direct air capture (DAC). In contrast to capturing carbon from a point source, such as the exhaust stream of a fossil fuel power plant, research on DAC technologies has been much less explored and it is unknown what materials and process conditions will be optimal at the scale required. Porous materials, such as MOFs, are potential candidate materials for DAC, but because of the high number of porous materials and possible processes, it is impossible to test all combinations experimentally. To the best of my knowledge, there have been no significant computational screenings of materials or processes specific to DAC and I think that large efficiency gains can be made by comprehensively simulating various material classes under various proposed process conditions. The total number of material / process combinations is still daunting in size so I will also be focusing on how we can screen materials and processes *faster* and *more accurately*.

This work presents three methodology papers that facilitate and accelerate the evaluation of materials for DAC and one applied materials screening. In aggregate, this work includes (1) a correction to the instantaneous heat flux measurement as calculated by LAMMPS, (2) a strategy to automate the exploration of structure-property relationships for new physisorption applications, (3) a new Python package that implements a molecular find and replace operation on periodic

structures, and (4) a computational screening of two MOFs, each with 30 functional group variations, for their ability to be part of a core-shell MOF in a defined DAC process. This work can be built on in the future to address more materials and processes, and, hopefully, will assist the greater academic community in finding better materials and processes for DAC so that it can be feasibly deployed at the necessary scale.

Table of Contents

Preface.....	xix
1.0 Introduction.....	1
2.0 Heat flux for many-body interactions: Corrections to LAMMPS	8
2.1 Introduction	9
2.2 Background	10
2.3 Notes on Implementation in LAMMPS	19
2.4 Results and Discussion	21
2.4.1 Example System 1: Idealized MOF structures.....	23
2.4.2 Example System 2: Propane, Octane and Hexadecane	25
2.5 Conclusion	29
3.0 Towards Comprehensive Exploration of the Physisorption Space in Porous Pseudomaterials Using an Iterative Mutation Search Algorithm	31
3.1 Introduction	31
3.2 Methodology.....	34
3.2.1 Pseudomaterials	34
3.2.2 Overview of Iterative Mutation Exploration (IME)	35
3.2.3 Simulated Structure-Property Space	37
3.2.4 Pseudomaterial Parameters and Run Parameters.....	38
3.2.5 Mutation Strategies.....	39
3.3 Results and Discussion	42
3.3.1 Results for baseline CH ₄ vs VF parameters.....	42

3.3.2 Improvements from Prior Algorithm.....	47
3.3.2.1 Using “Discrete” Mutations Less Often	48
3.3.2.2 Use Fewer Degrees of Freedom	51
3.3.2.3 Unbiased Bounded Mutations	54
3.3.3 Parameter Exploration	57
3.4 Conclusion	59
4.0 MOFUN: a Python package for molecular find and replace	61
4.1 Introduction	61
4.2 MOFUN: Algorithm Details	63
4.3 MOFUN: Optimization and Performance	67
4.4 MOFUN: Usage Details.....	69
4.5 Examples	71
4.5.1 Example 1: Functionalizing linkers in UiO-66	72
4.5.2 Example 2: Adding defects to UiO-66	73
4.5.3 Example 3: Parameterizing UiO-66 with flexible force field terms	75
4.6 Conclusion	77
5.0 Designing optimal core-shell MOFs for direct air capture	79
5.1 Introduction	80
5.2 Methodology.....	84
5.2.1 Overview	84
5.2.2 Idealized Adsorption/Desorption Cycle	85
5.2.3 Experimental	86
5.2.4 Determination of Water Breakthrough Times and Pellet Loadings	87

5.2.5 Scoring of Core-Shell MOF Pairs	90
5.2.6 COMSOL Multiphysics® Modeling	90
5.2.7 Adsorption and diffusion simulations	92
5.3 Results and Discussion	92
5.4 Conclusion	102
Appendix A LAMMPS corrected heat flux for improper potentials	104
Appendix B Supplementary Info to CSMOF paper	106
Appendix B.1 Computational Predictions.....	106
Appendix B.1.1 Preparation of Functionalized MOF files	106
Appendix B.1.2 Diffusion Calculations	107
Appendix B.1.3 Gas Loading Calculations	108
Appendix B.1.4 Custom CO ₂ -NH ₂ Interaction Force-Field Parameters	108
Appendix B.1.5 Simulated Diffusivities.....	110
Appendix B.1.6 Simulated Gas Loadings.....	111
Appendix B.2 Experimental	112
Appendix B.2.1 General Methods.....	112
Appendix B.2.2 Synthesis and characterization of MOF ligands	113
B.2.2.1 Dimethyl 2-nitro-1, 1'-biphenyl-4,4'-dicarboxylate (1).....	113
B.2.2.2 Dimethyl 2-amino-1, 1'-biphenyl-4,4'-dicarboxylate (2).....	113
B.2.2.3 2-Amino-1, 1'-biphenyl-4,4'-dicarboxylic acid (3)	114
B.2.2.4 Dimethyl-2,2'-dinitro-[1,1'-biphenyl]-4,4'-dicarboxylate (4) ¹³⁰ ...	114
B.2.2.5 Dimethyl 2,2'-diamino-[1,1'-biphenyl]-4,4'-dicarboxylate (5)	115
B.2.2.6 2,2'-diamino-[1,1'-biphenyl]-4,4'-dicarboxylic acid (5) ¹³⁰	115

B.2.2.7	Dimethyl 2-methyl-1, 1'-biphenyl-4,4'-dicarboxylate (6) ¹³¹	116
B.2.2.8	2-methyl-1, 1'-biphenyl-4,4'-dicarboxylate (7) ¹³¹	116
B.2.2.9	Dimethyl 2,2'-dimethyl-1, 1'-biphenyl-4,4'-dicarboxylate (8) ¹³¹ ..	117
B.2.2.10	2,2'-dimethyl-1, 1'-biphenyl-4,4'-dicarboxylate (9) ¹³¹	117
Appendix B.2.3 Synthesis of MOFs		118
B.2.3.1	Synthesis of UiO-67	118
B.2.3.2	Synthesis of NH ₂ -UiO-67	118
B.2.3.3	Synthesis of (NH ₂) ₂ -UiO-67	118
B.2.3.4	Synthesis of CH ₃ -UiO-67	119
B.2.3.5	Synthesis of (CH ₃) ₂ -UiO-67	119
Appendix B.2.4 Characterization of MOFs		120
B.2.4.1	Powder X-ray diffraction data	120
B.2.4.2	Thermogravimetric analysis	121
B.2.4.3	Elemental analysis	122
B.2.4.4	Gas sorption data	124
B.2.4.5	Calculation of CO ₂ : N ₂ adsorption selectivity	128
Appendix B.3 COMSOL Multiphysics® model.....		132
Bibliography		134

List of Tables

Table 1: Mutation strategy and ranges for this work and our prior work.	40
Table 2: Unit cell, sigma, epsilon density, adsorbed methane per unit cell and methane loading for selected materials in Figure 10.	47
Table 3: Fixed atom IME runs with applied min and max lattice sizes.	52
Table 4: allowable range for each pseudomaterial parameter for all 12 iterative mutation exploration runs.	56
Table 5: Sigma and epsilon values for UFF, UFF+-25%, and UFF+-50%.	57
Table 6: Top ten core-shell MOFs by greatest improvement over scores of their core and shell individually.	97
Table 7: Top core-shell MOFs by absolute score.	98
Appendix Table 1: Experimental CO₂:N₂ adsorption selectivity.	128
Appendix Table 2: Comparison of experimental and simulated CO₂:N₂ adsorption selectivity.	129
Appendix Table 3: Langmuir adsorption constants (KL), adsorption maximums (CPmax), and diffusivity values for the two MOFs modeled in this paper.	132

List of Figures

- Figure 1: (A) Side view of the layout of a simulation box, where the box is broken up into slabs of width and height $3L$ and depth L . Energy is added to the two center slabs (in orange) at the rate of ΔE per slab and removed from the two end slabs (in blue) at the same rate. The heat flux is measured in the two control regions (in grey). (B),(C), and (D) are snapshots of the cubic, triangular and hexagonal idealized MOFs with both side and end views. 22**
- Figure 2: Per-potential fraction of applied heat flux for both uncorrected LAMMPS and corrected calculations for idealized cubic, triangular and hexagonal MOFs..... 24**
- Figure 3: Simulation box layout for propane, octane and hexadecane..... 26**
- Figure 4: Per-potential fraction of applied heat flux for both uncorrected LAMMPS and corrected calculations for various hydrocarbons. As the length of the hydrocarbon increases, the heat transfer through the many-body potentials increases. $C_{16}H_{34}$ shows LAMMPS-reported heat fluxes of near zero for the angle and dihedral potentials, causing a total error of about 22%..... 27**
- Figure 5: Per-potential fraction of applied heat flux for hydrocarbons from Ohara, et al, which predict increasing reliance on the angle and dihedral many-body terms for heat flux transfer with increasing hydrocarbon length. 28**
- Figure 6: Total heat flux as a fraction of applied heat flux for both the uncorrected LAMMPS calculation (blue) and the corrected LAMMPS calculation (orange), averaged over 1M timesteps, as measured on octane. Prior to 3M timesteps, there is no applied heat flux (noted in grey) and the recorded instantaneous heat flux fluctuates around zero;**

after 3M timesteps, the heat flux is applied and the corrected calculation fluctuates around the correct value but the uncorrected calculation underestimates the real value as well as shows greater magnitude fluctuations. 29

Figure 7: How the iterative mutation exploration strategy works on a simplified 4x4 structure property space X_1 vs X_2 : the first generation of pseudomaterials uses random generation; subsequent generations are created through an iterative select-mutate-simulate strategy where the parents are selected by inverse bin density (the number shown in each bin is the number of materials in that bin). 35

Figure 8: Explored material density bin map (white: unexplored bin; black: 20 or more materials per bin; grey: 1 to 19 materials per bin) of methane loading vs void fraction for both IME and random generation at 50, 500, 5,000 and 25,000 materials..... 42

Figure 9: Number of bins explored vs. the number of materials for IME and random generation. 43

Figure 10: Methane loading vs void fraction for all materials in the reference run of IME, colored by the number of average adsorbed methane per unit cell at 60% opacity, and relatively sized by the lattice size. Example materials are shown on the right, with each atom site colored by atom type epsilon. 45

Figure 11: Material density bin map (black: ≥ 20 materials in the bin, white: 0 materials in the bin) of IME run for 500 generations (25,000 total materials) for every combination of max number of atoms to add or remove (± 1 , ± 4 , ± 16 , \pm unlimited) and mutation frequency (1.0, $M = 0.2$, $M_2 = 0.04$) across methane loading vs void fraction. Bins explored by random generation are in orange. 50

Figure 12: Material density bin map (black: ≥ 20 materials in the bin, white: 0 materials in the bin) of IME run for 500 generations (25,000 total materials) and bins explored by random generation (orange) across methane loading vs. void fraction for 1, 2, 4, 8, 16, 32, 64 and 512 atoms per unit cell.	53
Figure 13: Bins explored vs. number of materials for the pseudomaterial parameter ranges for atom types, atoms, sigma and epsilon as well as the simulation parameter mutation strength.	58
Figure 14: (A) a search pattern for the CH₃ group in Octane with counter-clockwise ordering of hydrogens, (B) a trial match 1 with matching counter-clockwise ordering of hydrogens, (C) a trial match 2 with non-matching clockwise-ordering of hydrogens, and (D) the erroneous result from rotating the directional axis and orientation point into place when starting with the hydrogens ordered in the reverse direction.	65
Figure 15: Time required to find or find + replace all linkers in a UiO-66 unit cell replicated to 2x2x2, 4x4x4, 5x5x5, 6x6x6, 7x7x7, and 8x8x8.....	68
Figure 16: (A) search pattern and search structure (a “slice” of UiO-66), (B) example 1 resulting structure after find and replace using a replacement pattern with an added OH functional group, and (C) example 2 resulting structure after find and replace using a replacement pattern of a defective linker (two capping formate groups), as applied to 25% of the linkers.	71
Figure 17: when parameterizing a structure using two parameterized patterns, for all two-body bond terms $b_1..b_n$ to be replicated, the patterns must share an atom, for all three-body angle terms $a_1..a_n$ to be replicated the patterns must share two atoms, and for	

all four-body dihedral terms $d_1..d_n$ to be replicated the patterns must share three atoms.	75
Figure 18: Overview of the strategy for designing optimal core-shell MOFs for DAC. A library of MOFs are combinatorically assembled into all of the possible core-shell MOF pairs, and then each pair is computationally evaluated to find candidates for experimental synthesis. Optimal designs should prevent H_2O from reaching the core while allowing for significant CO_2 adsorption in the core. Note that we include in our consideration core-shell MOF “pairs” where the core and the shell are the same MOF.	82
Figure 19: UiO-67 (left) and UiO-66 (right) structures showing biphenyl and phenyl ligands.	83
Figure 20: Chemical diagrams of functional groups used to modify UiO-66 and UiO-67 linkers.....	84
Figure 21: Schematic overview of our idealized adsorption/desorption cycle process used to investigate core-shell MOF candidates. We assume 100% evacuation of adsorbed gases during the desorption step, which can be mediated via imposing a vacuum or raising the temperature (or both), but in this idealized model the specific desorption conditions are intentionally ignored.....	86
Figure 22: A 1-D infinite slab model of the outer region of a core-shell MOF used to estimate breakthrough times and fluxes of both H_2O and CO_2.....	88
Figure 23: Setup of COMSOL Multiphysics® model of a single core-shell spherical pellet. Air flows in from the bottom over a 0.453 cm radius core + 0.04 cm thick shell pellet.	91

Figure 24: (a) Gas loadings of CO₂, N₂ and H₂O for selected MOFs based on UiO-67. Vertical lines are the amount of each gas in the atmosphere. (b) Diffusivities of CO₂, N₂ and H₂O for selected MOFs based on UiO-67, with error bars to 95% confidence interval.	93
Figure 25: Scores for all UiO-67-based core-shell MOF combinations (excluding any MOF where all core-shell MOFs derived from it had scores less than 1.0). Black boxes are a guide to highlight the scores for non-core-shell MOFs under the same process. Numbers indicate every core-shell MOF combination where the combination has a higher score than both MOFs that compose it. Bold numbers show combinations with at least a 25% higher score than both MOFs that compose it.	95
Figure 26: CO₂ / H₂O perm-selectivity of shell vs CO₂ / N₂ adsorption selectivity of core.	98
Figure 27: CO₂ concentrations (left) and H₂O concentrations (right) in a simulated amino1-methyl2 spherical pellet at t = 990 s.	100
Appendix Figure 1: (A) atoms IJKL for a typical improper, and (B) atoms IJKL for a dihedral defined using an improper.	104
Appendix Figure 2: Per-potential fraction of applied heat flux for both uncorrected LAMMPS and corrected calculations for octane, as calculated using dihedral or improper potentials.	105
Appendix Figure 3: Simulated CO₂ loadings at 1x, 2x, and 10x of the normal CO₂-amino interaction strength epsilon compared to experimental CO₂ loadings for the two MOFs UiO-67-amino1 and UiO-67-amino2.	109
Appendix Figure 4: Simulated diffusivities of CO₂, N₂, and H₂O for all UiO-66 and UiO-67 functional groups. Error bars to 95% confidence shown for all points.	110

Appendix Figure 5: Simulated gas loading of CO₂, N₂, and H₂O for all UiO-66 and UiO-67 functional groups. Atmospheric CO₂, N₂, and H₂O partial pressures shown by vertical yellow, grey and blue lines. Error bars to 95% confidence shown for all points (though error for most points is too small to be visible).	111
Appendix Figure 6: Simulated PXRD pattern of UiO-67 (black) and experimental PXRD of as synthesized UiO-67 (dark red), NH₂-UiO-67 (blue), (NH₂)₂-UiO-67 (orange), CH₃-UiO-67 (green) and (CH₃)₂-UiO-67 (dark blue). These data confirm the crystallinity and phase purity of the synthesized MOFs.....	120
Appendix Figure 7: TGA curves of UiO-67 (dark red), NH₂-UiO-67 (blue), (NH₂)₂-UiO-67 (orange), CH₃-UiO-67 (green) and (CH₃)₂-UiO-67 (dark blue).	121
Appendix Figure 8: N₂ sorption isotherms of UiO-67 (dark red), NH₂-UiO-67 (blue), (NH₂)₂-UiO-67 (orange), CH₃-UiO-67 (green) and (CH₃)₂-UiO-67 (dark blue) at 77 K. Filled and hollow circles indicate adsorption and desorption respectively. These data were used to calculate the Branauer-Emmett-Teller (BET) surface areas for the MOFs: UiO-67, 2572 m²/g ; NH₂-UiO-67, 2074 m²/g ; (NH₂)₂-UiO-67, 1705 m²/g; CH₃-UiO-67, 2042 m²/g; and (CH₃)₂-UiO-67, 1647 m²/g. These values are consistent with literature data for UiO-67^{133,134} as well as NH₂-UiO-67 and CH₃-UiO-67.¹³⁴ The difunctionalized analogues have lower BET surface areas, as expected due to the additional functional groups.	124
Appendix Figure 9: CO₂ adsorption isotherms of UiO-67 (dark red square), NH₂-UiO-67 (blue circle), (NH₂)₂-UiO-67 (orange left-triangle), CH₃-UiO-67 (green down-triangle) and (CH₃)₂-UiO-67 (dark blue right-triangle) at 298 K.	125

Appendix Figure 10: N₂ adsorption isotherms of UiO-67 (dark red square), NH₂-UiO-67 (blue circle), (NH₂)₂-UiO-67 (orange left-triangle), CH₃-UiO-67 (green down-triangle) and (CH₃)₂-UiO-67 (dark blue right-triangle) at 298 K.	126
Appendix Figure 11: Water adsorption isotherms of UiO-67 (dark red square), NH₂-UiO-67 (blue circle), (NH₂)₂-UiO-67 (orange left-triangle), CH₃-UiO-67 (green down-triangle) and (CH₃)₂-UiO-67 (dark blue right-triangle) at 298 K.	127
Appendix Figure 12: Comparison between experimental and simulated CO₂:N₂ adsorption selectivity of UiO-67, NH₂-UiO-67 and (NH₂)₂-UiO-67: (left) simulation data with default force field parameters for NH₂-CO₂ interaction, (right) simulation data with 10x epsilon force field parameters for NH₂-CO₂ interaction.	129
Appendix Figure 13: Comparison between experimental and simulated CO₂:N₂ adsorption selectivity of UiO-67, CH₃-UiO-67 and (CH₃)₂-UiO-67.	130
Appendix Figure 14: Comparison between experimental and simulated CO₂:N₂ adsorption selectivity of UiO-67, NH₂-UiO-67 and CH₃-UiO-67: (left) simulation data with default force field parameters for NH₂-CO₂ interaction, (right) simulation data with 10x epsilon force field parameters for NH₂-CO₂ interaction.	130
Appendix Figure 15: Comparison between experimental and simulated CO₂:N₂ adsorption selectivity of UiO-67, (NH₂)₂-UiO-67 and (CH₃)₂-UiO-67: (left) simulation data with default force field parameters for NH₂-CO₂ interaction, (right) simulation data with 10x epsilon force field parameters for NH₂-CO₂ interaction.	131
Appendix Figure 16: Experimental vs. modeled Langmuir adsorption isotherms for CO₂ (left), H₂O (middle), and N₂ (right). Experimental values are shown with markers;	

modeled values are shown with solid lines. Values for our simulated core MOF, NH_2 , are shown in blue; values for our simulated shell MOF, $(\text{CH}_3)_2$, are shown in red.133

Preface

I want to thank my advisor and friend, Chris Wilmer, who convinced me to leave industry and get a PhD (“it will be easy” he said). While I am grateful for the academic mentorship—I am acutely aware of how my capacity to comprehend difficult ideas *and* communicate them clearly has grown while working with Chris—what I am most grateful for is him being calm, unflappable, and reasonable at all times. Without these qualities in Chris, I doubt I would have made it through the challenges of becoming a father and living through the era of COVID to reach the finish line of my PhD. Thank you, Chris.

I want to thank Jan Steckel for the many fruitful discussions we’ve had throughout our ongoing struggle to apply new force field constructions to actual problems, and of course the other members of my committee—James McKone and Giannis Mpourmpakis—for their continuing critical feedback and support.

Finally, I want to thank my wife Aigerim and our six-year old son Arlan who regularly confirms with me that I am working on removing CO₂ from the atmosphere so that “we can be kind to planet earth.”

1.0 Introduction

Over 1 trillion tons of CO₂ have been emitted into the atmosphere^{1,2} since we passed the concentration threshold for what is widely considered to be livable, 350 ppm CO₂. Maintaining a livable planet for future generations requires us to do two things: (1) stop emitting CO₂ via rapid decarbonization of all industry, and (2) remove the excess CO₂ via some form of negative emissions technology.³ 1 trillion tons of CO₂ is an extremely large number and the challenges of operating any process at this scale is daunting. If we were to do this in a reasonable amount of time for the climate, or 50 years, then this will require an industry that could remove 20 billion tons of CO₂ / year. Removing this much CO₂ will require building a new industry the size of the fossil fuel industry where almost none exists today.

There are many proposed negative emissions technologies, including biofuel plus sequestration, artificial weathering, enhancing carbon uptake of agricultural land, and what I will be focusing on here: capturing CO₂ directly from the atmosphere, or Direct Air Capture (DAC).⁴ Because of the scale of the problem, likely all these technologies will be used in some manner, but they all have scaling difficulties. Growing biofuel crops, burning them for energy, and capturing and sequestering the carbon, if scaled to 20 billion tons CO₂ / year, would use over 76% of all agricultural land in the world.⁵ Artificial weathering—using crushed limestone or olivine to increase CO₂ uptake of the ocean—would require deployment to the entire coastline of the planet,^a and

^a Estimated CO₂ captured if process extended to entire Netherlands coastline of 451km: 9MtCO₂/yr.⁶
20,000 MtCO₂ * (451km / 9MtCO₂) = ~1e6 km. Coastline of the world is approximately 1.2e6 km.

require mining 65 billion tons of rock per year^b and moving it to the coastline in a carbon neutral manner.^{6,7} For comparison, the most coal ever mined globally in one year was 8 billion tons in 2013.⁸ DAC, with current technology,⁹ would use over 200% of our current global electrical production.^c

Despite these sobering assessments, there is room for optimism due to the scale of possible technology improvements. While there has been some research on DAC, most of it is very early stage.^{9–16} Early work and cost estimates were based on processes and materials used for carbon capture from powerplants,¹⁷ which is not necessarily effective for capturing carbon from an unlimited dilute source.¹⁸ A novel material tailored specifically for the conditions of DAC combined with a complementary and possibly novel process can achieve significant increases in efficiency. And in the worst-case scenario, where energy requirements cannot be dramatically driven down, additional clean energy can be deployed to power the best possible process.

While there are proposed technologies that would use carbon captured from DAC—such as conversion to biofuels, concrete, and building materials—no carbon utilization technology requires CO₂ near the scale of 1 trillion tons so sequestering the CO₂ is necessary.¹⁹ A typical system would have a DAC plant to capture CO₂, which would be transported to a sequestration site, compressed and then pumped underground to an underground reservoir. For this system, the cost can be estimated as \$10/tCO₂ for sequestration,²⁰ \$8/tCO₂^{21,d} to compress to the pressure required, and

^b Approximately 1.3 Mm³ required to remove 1 MtCO₂.⁶ Density of olivine is ~2.5 t/m³.⁷ Olivine required to remove 20 billion tons of CO₂ is: (2.5 Mt/Mm³)*(1.3 Mm³ / 1 MtCO₂)*20,000 MtCO₂ = 65000 Mt olivine.

^c Global electrical production for 2019 was ~27000 TWh. 430 kJ/mol = 2715 kWh/tCO₂. 2715 kWh/tCO₂ * 20B tCO₂ / 27000 TWh = 201%

^d 400 MJ/tCO₂ to reach 11 MPa * \$0.0194/MJ (assumes \$0.07 / kWh)= \$7.76.

the cost of transportation depends on the distance from the DAC facility to the sequestration site but could be eliminated if the DAC facility is sited at the sequestration site. The main unknown cost is the DAC process itself and estimates for DAC have varied widely from \$13-\$1000/tCO₂.^{17,18} This is primarily the energy cost of the separation and if we look at a DAC process from the perspective of its energy cost, there are two reasonable bounds: (1) the energy usage of current DAC pilot plants is an estimated 10,000 MJ/tCO₂,⁹ and (2) the thermodynamic minimum energy is 455 MJ / tCO₂.¹⁸ To achieve the widely accepted goal of capturing CO₂ for \$50/tCO₂ we need to be as close to the thermodynamic minimum energy as possible. To make this efficiency improvement, better materials and better processes are needed.

The separation of CO₂ from the atmosphere is a particularly challenging separation because the concentration of CO₂ in the atmospheric is only 400 ppm—or 0.04%—and it needs to be concentrated to 95% CO₂ for it to be transported as a liquid via pipeline.²² The concentration of water from humidity also must be reduced from about 1-2% to < 500ppm.²² Early analysis predicted that DAC would be infeasible economically based on the low concentration of the input gas and the thermodynamic cost of extracting all the CO₂ from a given input stream.¹⁷ One advantage of DAC, however, is that the input stream can be considered to be unlimited and only the most thermodynamically favorable part of the stream needs to be extracted, based on an energy / cost tradeoff between processing a larger air stream and capturing more CO₂ in the stream.¹⁸ Because of the large concentration change of CO₂, we need a material and process that is highly selective for CO₂. Likely, a multiple stage process will be required to reach the required concentrations.

Currently, there are three companies with pilot-scale DAC plants. Carbon Engineering has a solvent-based plant that can capture 500tCO₂ / year.¹⁰⁻¹² ClimeWorks and Global Thermostat

have amine sorbent-based plants totaling approximately 6000 tCO₂ / year that use waste heat and a partial vacuum to regenerate the sorbent.^{9,13–16} Scaling from the current deployed capacity of approximately 6500 tCO₂ / year to 20 BtCO₂ / year will require scaling up 3,000,000 times current capacity. The materials and processes used will change significantly at this scale; current systems use waste heat—which is great at small scales—but there is not enough waste heat on the planet to power a process at this scale, and the amount of available waste heat will decline since the biggest sources of waste heat are from fossil fuel combustion for power and transport, which will need to be decarbonized.²³ There are also numerous alternative processes that have not reached pilot scale, such as capturing CO₂ from seawater,^{24–26} or moisture swing adsorption,²⁷ which uses the free energy change from the evaporation of water in a dry climate to drive the capture of CO₂. While there are many proposed ideas and pilot plants of varying sizes, there is no consensus as to what materials or process will work best at the scale of 1 trillion tons CO₂.

While there is a large body of research on CO₂ capture from power plant emissions including many computational screenings,^{28–32} the DAC literature is limited to a few research groups working on specific solutions,^{9–16,27,33,34} and, to the best of my knowledge, there have been no significant computational screenings of materials or processes. Because of how unexplored this research area is at present, I think that large efficiency gains can be made by computationally exploring materials and processes for DAC.

My long-term goal is to find order-of-magnitude-level efficiency improvements by comprehensively simulating various material classes under various proposed process conditions. This will include simulating databases of known materials, predicted materials, and simulating materials modified with varying functional groups. The total number of porous materials that could be explored is daunting in size—for example, there are over 90,000 known MOFs, and over 500K

predicted MOFS,³⁵ 100K predicted zeolites,³⁶ 18,000 predicted porous polymer networks,³⁷ over 1,000,000 mixed matrix membranes—before considering an additionally varied number of possible processes. Because exploring everything is not feasible, in addition to targeted explorations of specific materials classes, I will also be focusing on how we can screen materials and processes *faster and more accurately*.

This dissertation is a compilation of four papers, two of which have been peer reviewed in the *Journal of Chemical Theory and Computation* and the *Journal of Chemical Physics*, and the other two which have been submitted to *Digital Discovery* and *Nanoscale*.

A property that is often overlooked when evaluating materials in an adsorption process is the thermal conductivity of the material. For gas adsorption processes, the energy released when a gas adsorbs into a material is released as heat which raises the temperature of the material. If the kinetics of the gas adsorption process are fast enough, the increased temperature of the material may dampen the continued adsorption of the gas, effectively rate-limiting the kinetics of the process. To avoid this problem, a good material needs a sufficiently high thermal conductivity to transfer the heat out of the material. Thermal conductivity calculations are needed when evaluating top-performing materials for DAC, but in our early trials, we discovered that the heat flux command required to calculate thermal conductivities is derived incorrectly in LAMMPS^{38,39}, the open-source software we use for molecular dynamics. The first paper, titled “Heat flux for many-body interactions: Corrections to LAMMPS” (Section 2), published in the *Journal of Chemical Theory and Computation*, describes this error in the instantaneous heat flux as calculated by LAMMPS, and presents an updated software implementation that is correctly derived.

Because the research area of materials and processes for DAC has not been thoroughly explored, there is no consensus on what properties in a material are necessary for the material to

be part of an efficient DAC process. Mapping the structure-property relationships between material properties and the efficiency of DAC processes is a crucial first step to easily filter both materials and processes for more detailed evaluation, and also to find non-intuitive novel processes. The second paper (Section 3), “Towards Comprehensive Exploration of the Physisorption Space in Porous Pseudomaterials Using an Iterative Mutation Search Algorithm,” published in the *Journal of Chemical Physics*, presents a methodology for exploring the structure-property relationships of new physisorption applications. While the paper specifically addresses methane loading as a well-defined application problem to demonstrate the methodology, this approach is designed to be applied to problems that have been largely unexplored to-date, such as DAC.

The third paper (Section 4), “MOFUN: a Python package for molecular find and replace”, submitted to Digital Discovery, presents a new Python package that implements a molecular find and replace operation on periodic structures. This operation is immensely helpful for automating many tasks, such as adding functional groups to a range of structures, adding defects to structures, and parameterizing structures with force field parameters. These tasks are crucial when screening new materials for any application, and I apply them to DAC in the fourth paper (Section 5), “Designing optimal core-shell MOFs for direct air capture” submitted to *Nanoscale*. In this final paper, I computationally screen two MOFs, each with 30 functional group variations, for their ability to be part of a core-shell MOF in a defined DAC process.

In aggregate, this work presents three methodologies that facilitate and accelerate the evaluation of materials for DAC and one applied materials evaluation of a promising new kind of hierarchical material called a core-shell MOF. This work can be built on in the future to address more materials and more processes, and, hopefully, will assist the greater academic community in

finding better materials and processes for DAC so that it can be feasibly deployed at the necessary scale.

2.0 Heat flux for many-body interactions: Corrections to LAMMPS

Paul Boone, Hasan Babaei, and Christopher E. Wilmer

Department of Chemical and Petroleum Engineering, University of Pittsburgh,

3700 O'Hara Street, Pittsburgh, Pennsylvania 15261

The virial stress tensor-based instantaneous heat flux, which is used by LAMMPS, is only valid for the small subset of simulations that contain only pairwise interactions. For systems that contain many-body interactions using 3- or 4-body potentials, a more complete derivation is required. We have created a software patch to LAMMPS that implements the correct heat flux calculation approach for 3- and 4-body potentials, based on the derivation by Torii et al⁴⁰. Using two example systems, the error in the uncorrected code for many-body potential heat flux is shown to be significant and reaches nearly 100% of the many-body potential heat flux for the systems we studied; hence, the error of the total heat flux calculation is proportional to the fraction of the total heat flux transferred through the many-body potentials. This error may have consequences for calculating thermal conductivities calculated using the Green-Kubo method or any NEMD method that uses the instantaneous heat flux. We recommend that all researchers using LAMMPS for heat flux calculations where significant heat is transferred via the many-body potentials adopt the corrected code.

2.1 Introduction

LAMMPS is a commonly-used open-source molecular dynamics package³⁸ and can be used for, among other things, calculating thermal transport properties such as the thermal conductivity.^{41–44} There are four common methods in LAMMPS to measure the thermal conductivity. Three of these are based on non-equilibrium molecular dynamics (NEMD), briefly: (1) enforcing a temperature gradient by thermostating two regions and measuring the molecular heat flux or keeping track of the energy added and removed from the thermostatted regions,⁴⁵ (2) enforcing an energy flux by adding and removing a constant amount of energy to two defined regions and then measuring the resulting temperature gradient,⁴⁶ and (3) defining two regions and swapping the kinetic energy of atoms between the two regions to create a very small temperature gradient, and then measuring the exchanged energy (i.e. the Muller-Plathe method⁴⁷). The fourth approach is the Green-Kubo method,^{48,49} which uses the autocorrelation function of the instantaneous heat flux to calculate the thermal conductivity, under equilibrium conditions. In this paper, we describe an error with how LAMMPS calculates the instantaneous heat flux, which could affect thermal conductivity calculations that employ the instantaneous heat flux, such as some NEMD calculations (subset of case 1 above) as well as the Green-Kubo method. In general, this error will affect any calculation that employs the instantaneous heat flux, such as the calculation of per-potential heat fluxes.^{50,51}

In LAMMPS, the function that calculates the instantaneous heat flux uses a virial stress tensor form. To the best of our knowledge, a derivation for this form of the heat flux has not been published, besides the terse form that exists in the LAMMPS documentation⁵². In systems involving only two-body potentials, this form is valid, but it cannot be extended to systems with many-body potentials. Torii et al⁴⁰ derive general expressions for many-body heat fluxes (which

we will review below in the background section) but do not address differences with the virial stress tensor formulation. Fan et al⁵³ states that the LAMMPS virial stress heat flux applies only to two-body potentials, but does not go into detail about the derivation error in the stress-based form. Additionally, neither paper provides source code corrections to LAMMPS, although Fan et al does have an alternative GPU code available on request which is limited to specific potentials (the Stillinger-Weber and Tersoff potentials as currently publicized).

The purpose of this paper is three-fold: (1) to show how the virial stress tensor formulation used by LAMMPS to calculate heat flux is derived and how it compares to a correctly derived many-body heat flux, (2) to publish publicly-available code for correctly calculating many-body heat flux to LAMMPS, and (3) to demonstrate its importance using different example systems. We analyze an idealized metal-organic framework (MOF) where heat is transferred predominantly via the bond and angle potentials, and the liquid-phase hydrocarbons propane, octane, and hexadecane with bond, angle, dihedral and improper potentials. The scale of the error in the heat flux ranges from significant to inconsequential, depending on the system and how much of the heat flux transfers through the many-body potentials. The magnitude of this effect on thermal conductivities calculated via the Green Kubo method is difficult to generalize and is likely system dependent. The hydrocarbon system is particularly illustrative of the range of error and how the distribution of the heat flux through different potentials affects the total error in the system.

2.2 Background

In LAMMPS, the heat flux is calculated using a virial stress tensor form defined per atom. This definition, when appropriately limited to only two-body potentials, is provably equivalent to

the Irving and Kirkwood⁵⁴ or Hardy⁵⁵ heat fluxes. In this section, we will show a complete derivation of a general heat flux expression for a many-body potential based on the derivation in Torii, et al.⁴⁰, before applying it to two-body potentials only in order to derive the virial stress tensor form of the heat flux used in LAMMPS. We can then compare the virial stress tensor as defined for many-body heat fluxes to our general derivation to see that the per atom virial stress tensor heat flux defined in LAMMPS is not a valid expression for the heat flux.

We start with the definitions of the instantaneous heat flux \mathbf{J} and per-atom energy E_i :⁵⁶

$$E_i = \frac{1}{2} m_i \mathbf{v}_i^2 + U_i \quad (2-1)$$

$$\mathbf{J}V = \frac{d}{dt} \sum_i \mathbf{r}_i E_i \quad (2-2)$$

$$= \sum_i E_i \mathbf{v}_i + \sum_i \mathbf{r}_i \frac{d}{dt} E_i \quad (2-3)$$

Where V is the volume, and m_i , \mathbf{v}_i , U_i , and \mathbf{r}_i are the mass, velocity, potential energy, and position of the i^{th} atom. We can separate equation (3) into a convective term $\mathbf{J}_{cnv} = \sum_i E_i \mathbf{v}_i$ and a potential term $\mathbf{J}_{pot} = \sum_i \mathbf{r}_i \frac{d}{dt} E_i$ where \mathbf{J}_{cnv} represents the heat flux due to the movement of atoms in the system, and \mathbf{J}_{pot} represents the heat flux due to changes in atom potentials. Note that \mathbf{J}_{cnv} is still dependent on the form of the atom potential U_i , which will in general involve a summation over all the potentials defined on the system; in this case: two-body, three-body, and four-body potentials. Here we define, \mathbb{P} , as the set all potentials $\mathbb{P} = \mathbb{P}_2 + \mathbb{P}_3 + \dots + \mathbb{P}_m$ defined on the system, where \mathbb{P}_2 is all two-body potentials, \mathbb{P}_3 is all 3-body potentials, and \mathbb{P}_m is all m-body potentials. Further, let \mathbb{P}_{mi} refer to the set of all m-body potentials which include the atom i , and let U_ϕ be the potential energy for the specific potential ϕ . If we are looking at only two-, three- and four-body potentials, U_i can now be defined as:

$$U_i = \sum_{m=2}^4 \sum_{\phi \in \mathbb{P}_{mi}} \frac{1}{m} U_\phi \quad (2-4)$$

Here, the total potential energy of atom i is the sum of all potentials that include the atom i divided by the number of atoms in each potential. In this way, the energy of each m -body potential is evenly divided up amongst the atoms that constitute it. There are other ways of distributing the potential energy between atoms but this does not affect the resulting aggregate heat flux.^{40,56} We are evenly distributing them in this derivation for simplicity (and because this is what LAMMPS does); interested readers can see a full derivation with arbitrarily distributed potential energies in the paper by Torii, et al.⁴⁰

We have three definitions for forces: (1): let \mathbf{F}_i be the sum of all forces on i , (2) let $\mathbf{F}_{i\phi}$ be the force on atom i due to a specific potential $\phi \in \mathbb{P}$, and (3) let $\mathbf{F}_{ij,\phi}$ be the force on atom i due to atom j as part of a specific potential $\phi \in \mathbb{P}$.

Expanding the J_{pot} term:

$$\begin{aligned} J_{pot} &= \sum_i \mathbf{r}_i \frac{d}{dt} E_i \\ &= \sum_i \mathbf{r}_i \frac{d}{dt} \left(\frac{1}{2} m_i \mathbf{v}_i^2 + U_i \right) \\ &= \sum_i \mathbf{r}_i (m_i \mathbf{a}_i \cdot \mathbf{v}_i) + \sum_i \mathbf{r}_i \frac{d}{dt} \left(\sum_{m=2}^4 \sum_{\phi \in \mathbb{P}_{mi}} \frac{1}{m} U_\phi \right) \\ &= \sum_i \mathbf{r}_i (\mathbf{F}_i \cdot \mathbf{v}_i) + \sum_i \mathbf{r}_i \left(\sum_{m=2}^4 \sum_{\phi \in \mathbb{P}_{mi}} \frac{1}{m} \frac{dU_\phi}{dt} \right) \end{aligned} \quad (2-5abcd)$$

To finish our definition of the J_{pot} term, we introduce some additional notation. First, we define $\mathbf{r}_{jk} = \mathbf{r}_j - \mathbf{r}_k$. We will also use the notation $j \in \phi$ to mean j is one of the atoms that constitutes

the potential ϕ , and if we take all possible pairs of the atoms in ϕ , denoted $[\phi]^2$, then $\{j, k\} \in [\phi]^2$ will mean that $\{j, k\}$ is one of these pairs (e.g., for a two-, three- or four-body potential, there will be 1, 3, and 6 pairs, respectively). We also need to express the derivative $\frac{dU_\phi}{dt}$ in terms of the forces and velocities of its constituent atoms, where $r_j = |\mathbf{r}_j|$:

$$\frac{dU_\phi}{dt} = \sum_{j \in \phi} \frac{dU_\phi}{dr_j} \frac{dr_j}{dt} = - \sum_{j \in \phi} \mathbf{F}_{j\phi} \cdot \mathbf{v}_j \quad (2-6)$$

Now:

$$\begin{aligned} J_{pot} &= \sum_i \mathbf{r}_i (\mathbf{F}_i \cdot \mathbf{v}_i) - \sum_i \mathbf{r}_i \left(\sum_{m=2}^4 \sum_{\phi \in \mathbb{P}_{mi}} \frac{1}{m} \sum_{j \in \phi} \mathbf{F}_{j\phi} \cdot \mathbf{v}_j \right) \\ &= \sum_i \mathbf{r}_i \sum_{m=2}^4 \sum_{\phi \in \mathbb{P}_{mi}} \left(\mathbf{F}_{i\phi} \cdot \mathbf{v}_i - \frac{1}{m} \sum_{j \in \phi} \mathbf{F}_{j\phi} \cdot \mathbf{v}_j \right) \\ &= \sum_{m=2}^4 \sum_{\phi \in \mathbb{P}_m} \sum_{i \in \phi} \mathbf{r}_i \left(\mathbf{F}_{i\phi} \cdot \mathbf{v}_i - \frac{1}{m} \sum_{j \in \phi} \mathbf{F}_{j\phi} \cdot \mathbf{v}_j \right) \\ &= \sum_{m=2}^4 \sum_{\phi \in \mathbb{P}_m} \left(\frac{1}{m} \sum_{i \in \phi} \sum_{j \in \phi} \mathbf{r}_i (\mathbf{F}_{i\phi} \cdot \mathbf{v}_i - \mathbf{F}_{j\phi} \cdot \mathbf{v}_j) \right) \\ &= \sum_{m=2}^4 \sum_{\phi \in \mathbb{P}_m} \left(\frac{1}{m} \sum_{\{i,j\} \in [\phi]^2} [\mathbf{r}_i (\mathbf{F}_{i\phi} \cdot \mathbf{v}_i - \mathbf{F}_{j\phi} \cdot \mathbf{v}_j) + \mathbf{r}_j (\mathbf{F}_{j\phi} \cdot \mathbf{v}_j - \mathbf{F}_{i\phi} \cdot \mathbf{v}_i)] \right) \\ &= \sum_{m=2}^4 \sum_{\phi \in \mathbb{P}_m} \left(\frac{1}{m} \sum_{\{i,j\} \in [\phi]^2} [\mathbf{r}_{ij} (\mathbf{F}_{i\phi} \cdot \mathbf{v}_i - \mathbf{F}_{j\phi} \cdot \mathbf{v}_j)] \right) \quad (2-7abcdef) \end{aligned}$$

We are able to incorporate the $\mathbf{F}_i \cdot \mathbf{v}_i$ term into the outer summations because the total force on i is equivalent to all the forces on i due to all m -body potentials ϕ , i.e. $\mathbf{F}_i = \sum_{m=2}^4 \sum_{\phi \in \mathbb{P}_{mi}} \mathbf{F}_{i\phi} \cdot \mathbf{v}_i$.

We are then able to invert the summations and sum across all atoms i in a potential ϕ for all

potentials, rather than summing all potentials ϕ that atom i is part of. For the inner summation, if we sum $\mathbf{F}_{i\phi} \cdot \mathbf{v}_i$ m times (once for every j in the given m -body potential ϕ), then we would need to divide it by m , i.e. $\mathbf{F}_{i\phi} \cdot \mathbf{v}_i = \frac{1}{m} \sum_{j \in \phi_{atoms}} \mathbf{F}_{i\phi} \cdot \mathbf{v}_i$. Finally, to use the relative positions \mathbf{r}_{ij} , we need to recognize that in equation 7d for every pair $\{i = i_1, j = j_1\}$, there will be a corresponding pair $\{i = j_1, j = i_1\}$ and we can sum them twice per pair if we replace the double sums across i and j with one sum across all pairs that make up the potential.

Our final general expression for the heat flux will therefore be:

$$JV = \sum_i E_i \mathbf{v}_i + \sum_{m=2}^4 \sum_{\phi \in \mathbb{P}_m} \left(\frac{1}{m} \sum_{\{i,j\} \in [\phi]^2} [\mathbf{r}_{ij} (\mathbf{F}_{i\phi} \cdot \mathbf{v}_i - \mathbf{F}_{j\phi} \cdot \mathbf{v}_j)] \right) \quad (2-8)$$

This is a per-potential version of the general heat flux, meaning we sum across the various potentials in the system, and then across all possible pairs of atoms included in the potential. The term inside the outermost parentheses is the heat flux for one specific m -body potential ϕ . To more easily compare to the virial stress heat flux in LAMMPS, we will also need a version of the heat flux that is rearranged to be per-atom, where we sum across all atoms and then across all potentials that include that atom. This rearrangement is straightforward; the only thing to note is that there is an extra factor of 2 due to the summations above being over the pairs $\{i, j\}$ and the one below being over all atoms i and then all $j \in \phi$, yielding two pairs $\{i = i_1, j = j_1\}$ and $\{i = j_1, j = i_1\}$:

$$JV = \sum_i E_i \mathbf{v}_i + \sum_i \sum_{m=2}^4 \frac{1}{2m} \sum_{\phi \in \mathbb{P}_{mi}} \sum_{j \in \phi} [\mathbf{r}_{ij} (\mathbf{F}_{i\phi} \cdot \mathbf{v}_i - \mathbf{F}_{j\phi} \cdot \mathbf{v}_j)] \quad (2-9)$$

We can now show that this equation, when limited to two-body potentials only, is equivalent to the Irving and Kirkwood and Hardy derivations:

$$\begin{aligned}
(JV)_{\mathbb{P}_2} &= \sum_i E_i \mathbf{v}_i + \sum_i \frac{1}{4} \sum_{\phi \in \mathbb{P}_{2i}} \sum_{j \in \phi} [\mathbf{r}_{ij} (\mathbf{F}_{i\phi} \cdot \mathbf{v}_i - \mathbf{F}_{j\phi} \cdot \mathbf{v}_j)] \\
&= \sum_i E_i \mathbf{v}_i + \sum_i \frac{1}{4} \sum_{\phi \in \mathbb{P}_{2i}} \sum_{j \in \phi \setminus \{i\}} [\mathbf{r}_{ij} (\mathbf{F}_{i\phi} \cdot (\mathbf{v}_i + \mathbf{v}_j))] \\
&= \sum_i E_i \mathbf{v}_i + \frac{1}{2} \sum_i \sum_{j>i} [\mathbf{r}_{ij} (\mathbf{F}_{ij} \cdot (\mathbf{v}_i + \mathbf{v}_j))] \tag{2-10abc}
\end{aligned}$$

For a two-body potential, there will be only two j 's for each ϕ ; one of them will be equal to i and can be omitted. Then the summations $\sum_{\phi \in \mathbb{P}_{2i}} \sum_{j \in \phi}$ can be replaced with the simpler summations $\sum_i \sum_{j>i}$ as long as one also replaces the force term, $\mathbf{F}_{i\phi} = \mathbf{F}_{ij,\phi}$ and $\mathbf{F}_{ij} = \sum_{\phi} \mathbf{F}_{ij,\phi}$. This final equation 10c is recognizable as Irving and Kirkwood's definition of the molecular interaction component of the heat flux for a uniform system in absence of fluid flow.⁵⁴ Further, equation 10c is also identical to Hardy's definition of the potential component for heat flux, also for a uniform system in absence of fluid flow⁵⁵.

Continuing with the assumption that there are only two-body potentials defined on the system, we can now derive the LAMMPS virial stress tensor heat flux, starting with showing how a general definition of the stress tensor relates to the per-atom version that LAMMPS uses. The global stress tensor \mathbf{S} is defined to be the ensemble average of a kinetic term summed across all N atoms in the system and the virial tensor,⁵⁷ (or stress) $\mathbf{W}(\mathbf{r}^N)$, which is a function of the N positions \mathbf{r}^N ,

$$\mathbf{SV} = \left\langle \sum_{i=1}^N m_i \mathbf{v}_i \mathbf{v}_i + \mathbf{W}(\mathbf{r}^N) \right\rangle \tag{2-11}$$

where V is the volume, and m_i and v_i are the mass and velocity of atom i . When calculating heat flux, LAMMPS excludes the kinetic term, leaving just the ensemble average of the virial stress.

We can expand the virial term using the positions \mathbf{r}_i and the total force \mathbf{F}_i on atom i :

$$\langle W(\mathbf{r}^N) \rangle = \left\langle \sum_i \mathbf{r}_i \mathbf{F}_i \right\rangle \quad (2 - 12)$$

We can further break this up by recognizing that \mathbf{F}_i will be a summation of the forces caused by all the potentials defined in the system:

$$\langle W(\mathbf{r}^N) \rangle = \left\langle \sum_i \sum_{\phi \in \mathbb{P}_{2i}} \mathbf{r}_i \mathbf{F}_{i\phi} + \sum_i \sum_{\phi \in \mathbb{P}_{3i}} \mathbf{r}_i \mathbf{F}_{i\phi} + \sum_i \sum_{\phi \in \mathbb{P}_{4i}} \mathbf{r}_i \mathbf{F}_{i\phi} \right\rangle = \left\langle \sum_i \sum_{m=2}^4 \sum_{\phi \in \mathbb{P}_{mi}} \mathbf{r}_i \mathbf{F}_{i\phi} \right\rangle \quad (2 - 13)$$

To get the per-atom form of the virial stress tensor that LAMMPS uses to calculate the heat flux, we separate the ensemble average of the virial stress into contributions from each atom, which LAMMPS calls a “per-atom stress tensor” and denoted here by \mathbf{s}_i :

$$\langle W(\mathbf{r}^N) \rangle = \left\langle \sum_i \mathbf{s}_i \right\rangle \quad (2 - 14)$$

where

$$\mathbf{s}_i = \sum_{m=2}^4 \sum_{\phi \in \mathbb{P}_{mi}} \frac{1}{m} \sum_{j \in \phi} \mathbf{r}_j \mathbf{F}_{j\phi} \quad (2 - 15)$$

Equation 13 is equivalent to equations 14-15. In the former, the virial terms $\mathbf{r}_i \mathbf{F}_{i\phi}$ are summed once per atom / potential pair. In the latter, the virial terms $\mathbf{r}_j \mathbf{F}_{j\phi}$ are summed up by potential so each term will appear m times, once for every atom j in the potential ϕ ; the sum of the virial terms is then divided amongst the potential’s constituent atoms. As mentioned above, the potential could be divided amongst the atoms in a different manner, but LAMMPS chooses to divide the potential evenly.

LAMMPS defines the instantaneous heat flux \mathbf{J} as:⁵²

$$(\mathbf{J}\mathbf{V})_{LAMMPS} = \sum_i E_i \mathbf{v}_i - \sum_i \mathbf{s}_i \cdot \mathbf{v}_i \quad (2-16)$$

The \mathbf{J}_{cnv} term $\sum_i E_i \mathbf{v}_i$ is equivalent to our derivation (equations 8-9) so we can focus on showing equivalence for just the \mathbf{J}_{pot} term. Starting with the \mathbf{J}_{pot} term from our general expression (equation 9), and limiting it to two body potentials, we can tie it out to the \mathbf{J}_{pot} of the per-atom virial stress heat flux used in LAMMPS:

$$\begin{aligned} J_{pot, \mathbb{P}_2} &= \sum_i \frac{1}{4} \sum_{\phi \in \mathbb{P}_{2i}} \sum_{j \in \phi \setminus \{i\}} \mathbf{r}_{ij} (\mathbf{F}_{i\phi} \cdot (\mathbf{v}_i + \mathbf{v}_j)) \\ &= \sum_i \frac{1}{4} \sum_{\phi \in \mathbb{P}_{2i}} \sum_{j \in \phi \setminus \{i\}} [(\mathbf{r}_{ij} \mathbf{F}_{i\phi}) \cdot \mathbf{v}_i + (\mathbf{r}_{ij} \mathbf{F}_{i\phi}) \cdot \mathbf{v}_j] \\ &= \sum_i \frac{1}{4} \sum_{\phi \in \mathbb{P}_{2i}} \sum_{j \in \phi \setminus \{i\}} [(\mathbf{r}_{ij} \mathbf{F}_{i\phi}) \cdot \mathbf{v}_i + (\mathbf{r}_{ji} \mathbf{F}_{j\phi}) \cdot \mathbf{v}_j] \\ &= \sum_i \frac{1}{2} \sum_{\phi \in \mathbb{P}_{2i}} \sum_{j \in \phi \setminus \{i\}} (\mathbf{r}_{ij} \mathbf{F}_{i\phi}) \cdot \mathbf{v}_i \\ &= \sum_i \frac{1}{2} \sum_{\phi \in \mathbb{P}_{2i}} \sum_{j \in \phi \setminus \{i\}} (\mathbf{r}_i \mathbf{F}_{i\phi} + \mathbf{r}_j \mathbf{F}_{j\phi}) \cdot \mathbf{v}_i \\ &= \sum_i \frac{1}{2} \left[\sum_{\phi \in \mathbb{P}_{2i}} \sum_{j \in \phi} \mathbf{r}_j \mathbf{F}_{j\phi} \right] \cdot \mathbf{v}_i \\ &= \sum_i [\mathbf{s}_i]_{\phi \in \mathbb{P}_{2i}} \cdot \mathbf{v}_i \quad (2-17abcdefg) \end{aligned}$$

Hence, equation 9, limited to two-body potentials, is equivalent to the LAMMPS heat flux definition, also limited to two-body potentials. This shows that LAMMPS's use of the virial stress tensor form of the heat flux is justified, when limited to two-body potentials.

LAMMPS, however, extends this virial stress beyond two-body potentials to three- and four-body potentials by making the leap that because virial stress can describe the heat flux for two-body potentials, then it can also describe the heat flux for three- and four-body potentials, i.e., if $J_{pot, \mathbb{P}_2} = \sum_i [\mathbf{s}_i]_{\phi \in \mathbb{P}_2} \cdot \mathbf{v}_i$, then $J_{pot, \mathbb{P}} = \sum_i [\mathbf{s}_i]_{\phi \in \mathbb{P}} \cdot \mathbf{v}_i$. This step is not valid. It may be easier to see the difference between the virial stress heat flux and the correct heat flux when comparing the heat flux contribution between the two forms for only one three-body potential $\phi = \{a, b, c\} \in \mathbb{P}_3$. The correct heat flux, starting from equation (8), will be:

$$\begin{aligned}
J_{pot, \phi=\{a,b,c\}} &= \left[\frac{1}{m} \sum_{\{i,j\} \in [\phi]^2} \mathbf{r}_{ij} (\mathbf{F}_{i\phi} \cdot \mathbf{v}_i - \mathbf{F}_{j\phi} \cdot \mathbf{v}_j) \right]_{\phi=\{a,b,c\}} \\
&= \frac{1}{3} [\mathbf{r}_{ab} (\mathbf{F}_{a\phi} \cdot \mathbf{v}_a - \mathbf{F}_{b\phi} \cdot \mathbf{v}_b) + \mathbf{r}_{bc} (\mathbf{F}_{b\phi} \cdot \mathbf{v}_b - \mathbf{F}_{c\phi} \cdot \mathbf{v}_c) + \mathbf{r}_{ac} (\mathbf{F}_{a\phi} \cdot \mathbf{v}_a - \mathbf{F}_{c\phi} \cdot \mathbf{v}_c)] \\
&= \frac{1}{3} [(\mathbf{r}_{ab} + \mathbf{r}_{ac})(\mathbf{F}_{a\phi} \cdot \mathbf{v}_a) + (\mathbf{r}_{bc} - \mathbf{r}_{ab})(\mathbf{F}_{b\phi} \cdot \mathbf{v}_b) + (-\mathbf{r}_{bc} - \mathbf{r}_{ac})(\mathbf{F}_{c\phi} \cdot \mathbf{v}_c)] \\
&= \frac{1}{3} [(\mathbf{r}_{ab} + \mathbf{r}_{ac})\mathbf{F}_{a\phi} \cdot \mathbf{v}_a + (\mathbf{r}_{bc} - \mathbf{r}_{ab})\mathbf{F}_{b\phi} \cdot \mathbf{v}_b + (-\mathbf{r}_{bc} - \mathbf{r}_{ac})\mathbf{F}_{c\phi} \cdot \mathbf{v}_c] \quad (2 - 18abcd)
\end{aligned}$$

Whereas the virial stress heat flux for one three-body potential $\phi = \{a, b, c\} \in \mathbb{P}_3$, starting with equations 15-16:

$$\begin{aligned}
J_{pot, \phi \in \{a,b,c\}} &= \left[\sum_i [\mathbf{s}_i]_{\phi} \cdot \mathbf{v}_i \right]_{\phi=\{a,b,c\}} \\
&= \left[\sum_{i \in \phi} \frac{1}{3} \left(\sum_{j \in \phi} \mathbf{r}_{ij} \mathbf{F}_{j\phi} \right) \cdot \mathbf{v}_i \right]_{\phi=\{a,b,c\}} \\
&= \frac{1}{3} (\mathbf{r}_a \mathbf{F}_{a\phi} + \mathbf{r}_b \mathbf{F}_{b\phi} + \mathbf{r}_c \mathbf{F}_{c\phi}) \cdot (\mathbf{v}_a + \mathbf{v}_b + \mathbf{v}_c) \\
&= \frac{1}{3} (\mathbf{r}_{ab} \mathbf{F}_{a\phi} + \mathbf{r}_{ca} \mathbf{F}_{c\phi}) \cdot (\mathbf{v}_a + \mathbf{v}_b + \mathbf{v}_c) \quad (2 - 19abcd)
\end{aligned}$$

For the virial stress heat flux, the virial stress term is the same for all atoms in the potential; this can therefore be multiplied by the sum of the velocities. For the correct form of the heat flux, however, the terms that multiply the atom velocities are all different from one another and cannot be combined. Curious readers can see the implementation of equation 19 by looking at the referenced LAMMPS source code files^{58,59}.

As shown above, the virial stress heat flux that LAMMPS uses is valid for two body potentials; however, the virial stress heat flux cannot be extended to three- and four-body potentials which may lead to erroneous thermal conductivity predictions. We suspect that the reason why this problem has mostly gone unnoticed is because for many systems, the amount of heat flux being transferred via the three- or four-body potentials is diminutive, so that any error due to the calculation of heat flux for those potentials is rendered largely inconsequential. The error may also be further obscured by compensating factors when using the heat flux for the purpose of calculating thermal conductivity via Green-Kubo. Comparisons of the corrected and virial stress heat flux calculations in example systems follow in the results section.

2.3 Notes on Implementation in LAMMPS

The error described above applies to all many-body potentials in LAMMPS. However, in LAMMPS, there are two different categories of many-body potentials and these different categories require separate code fixes. These two different categories are (1) many-body potentials defined on sets of three or more atoms where the sets are defined in advance, usually due to the bonding structure (i.e., the angle, dihedral, and improper styles in LAMMPS), and (2) many-body potentials where the atoms that interact change over the course of the simulation (i.e. potentials

implemented via the unfortunately named “pair potential” style such as the Tersoff, Brenner, and Stillinger-Weber potentials, and the AGNI and GAP machine learning potentials). The code fix that we have implemented addresses the first category of potentials only. Other research groups are working on the second category of potentials.⁵³

Due to how LAMMPS stores atoms across multiple processors, the velocity for an atom can only be found on the processor the atom is assigned to. In order to support this, rather than dividing the heat flux evenly between the atoms that comprise the potential, we assign the portion of the heat flux that contains the velocity of an atom to that atom. For a single three-body potential $\phi = \{a, b, c\}$, the heat fluxes assigned to the atoms a, b, c will be:

$$\begin{aligned}(JV)_{\phi,a} &= \frac{1}{3}[(\mathbf{r}_{ab} + \mathbf{r}_{ac})\mathbf{F}_{a\phi}] \cdot \mathbf{v}_a \\(JV)_{\phi,b} &= \frac{1}{3}[(\mathbf{r}_{bc} - \mathbf{r}_{ab})\mathbf{F}_{b\phi}] \cdot \mathbf{v}_b \\(JV)_{\phi,c} &= \frac{1}{3}[(-\mathbf{r}_{bc} - \mathbf{r}_{ac})\mathbf{F}_{c\phi}] \cdot \mathbf{v}_c\end{aligned}\tag{2-20abc}$$

These equations are just the individual terms in equation (18d). The total heat flux assigned is unchanged; the only difference is that the velocity the atom a is now only used in the calculation of the heat flux on atom a , and so forth with the velocities and heat fluxes for atoms b and c . Dividing the heat flux in this manner guarantees us the ability to calculate the heat flux, regardless of boundary conditions and number of processors.

We implemented the corrected algorithm described above for angle, dihedral and improper potential styles and it is currently available via our GitHub page at <https://github.com/wilmerlab/lammps>.

2.4 Results and Discussion

For the example systems, we combined a typical non-equilibrium molecular dynamics (NEMD) setup—we enforced a heat flux by adding and removing a fixed amount of energy from two regions—with measurements of the heat flux using the corrected heat flux or the uncorrected LAMMPS virial stress heat flux. The procedure was: a rectangular cuboid simulation box with periodic boundaries was filled with a crystalline solid or a liquid, and after an initial equilibration, we ran an NVE simulation with a fixed heat source applied to the center of the cuboid and a corresponding sink applied to the outer edges (see Figure 1a). After the temperature profiles reached steady-state, we recorded the instantaneous heat fluxes for control volumes on the left and right sides of the box, and averaged them over a time-frame necessary for convergence.

The instantaneous heat fluxes were recorded separately for the convective heat flux as well as any defined pair, bond, angle, dihedral, and improper potentials using both the corrected and uncorrected LAMMPS code. Because we are using the NVE ensemble, and the heat added and removed is fixed, the heat flux code should self-consistently report the applied heat flux, regardless of the force-field parameters, and we can therefore use this as an accurate gauge of measuring the correctness of our implementation. We have included our simulation parameters below for completeness, but in all cases, the calculated heat flux should be a function of the applied energy and the control volume size alone.

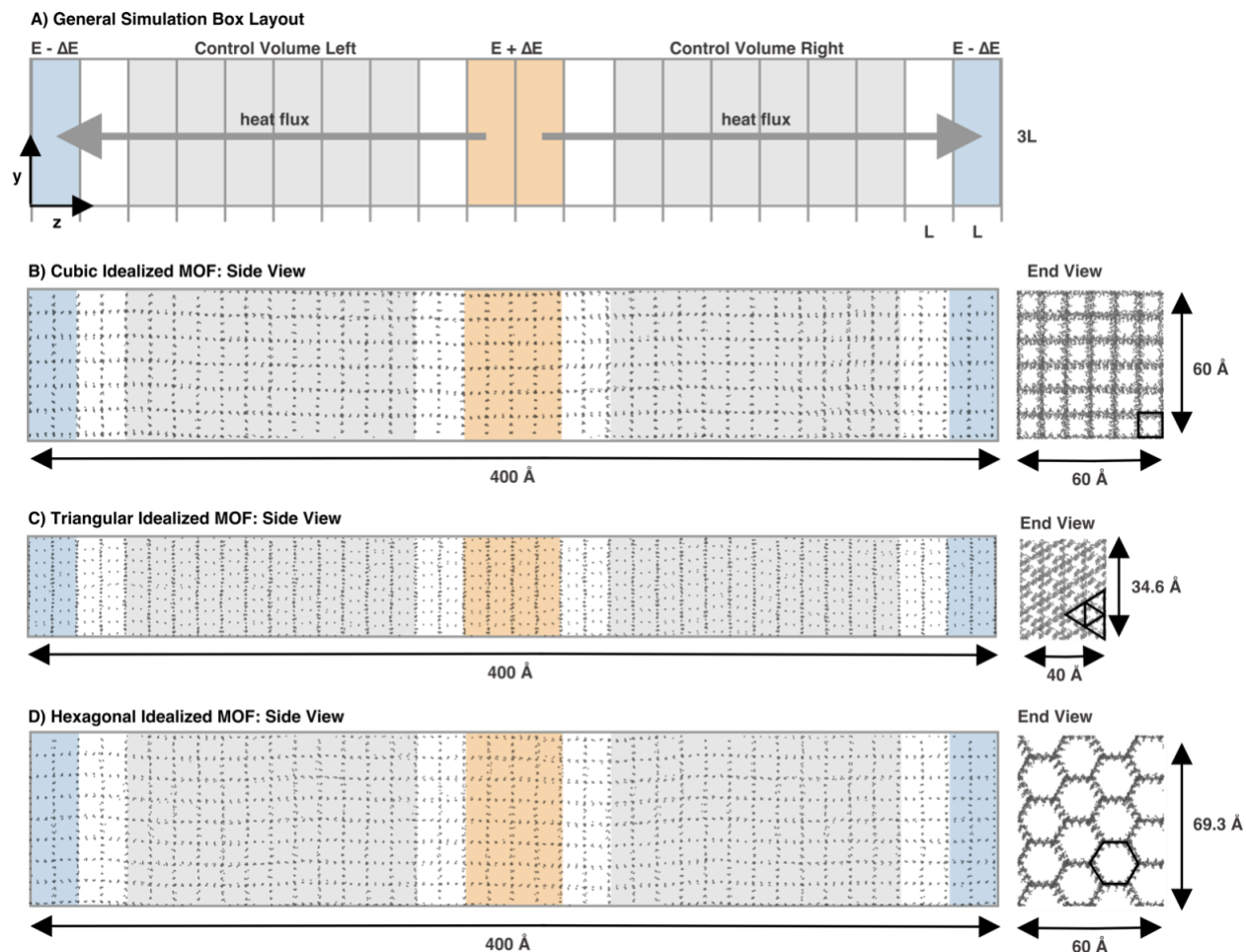


Figure 1: (A) Side view of the layout of a simulation box, where the box is broken up into slabs of width and height $3L$ and depth L . Energy is added to the two center slabs (in orange) at the rate of ΔE per slab and removed from the two end slabs (in blue) at the same rate. The heat flux is measured in the two control regions (in grey). (B),(C), and (D) are snapshots of the cubic, triangular and hexagonal idealized MOFs with both side and end views.

For this simulation setup, the corrected total instantaneous heat flux calculation does not require steady-state to be reached; the recorded instantaneous heat flux may be averaged immediately upon application of the heat source (see Figure 6). This is in contrast to the per-potential averages, which took a significant amount of time to converge after steady-state, from between 40-215M timesteps. Calculating the per-potential heat fluxes is usually unnecessary unless one is studying the distribution of heat flux between different potentials (or writing software to fix it); researchers

should not take the number of timesteps required for our simulations to be indicative of the number required to measure the total heat flux.

2.4.1 Example System 1: Idealized MOF structures

For testing the 3-body potential, we examined cubic, triangular and hexagonal idealized metal-organic framework (MOF) structures (see Figure 1(B)(C)(D)); a full description of these idealized MOF structures, including force field parameters, is covered in a previous paper⁶⁰ and we will present only the briefest description here. These idealized structures are simplified models corresponding to common actual geometries of real MOFs; a simple geometric shape is extended orthogonally to the plane of the shape in a series of channels, forming a lattice that stretches over the whole cuboid and connects across the periodic boundaries. The force field is defined with only bond and angle potentials, with parameters chosen to approximate the thermal conductivity of a typical MOF. After relaxation of the system using NVT / NVE for 500,000 timesteps, an energy source and sink was applied of 0.0040 kcal / mol fs for the cubic structure, 0.0071 kcal / mol fs for the triangular structure, and 0.0213 kcal / mol fs for the hexagonal structure. After a steady-state was reached at 5M timesteps, the per potential heat flux was averaged over 15M timesteps.

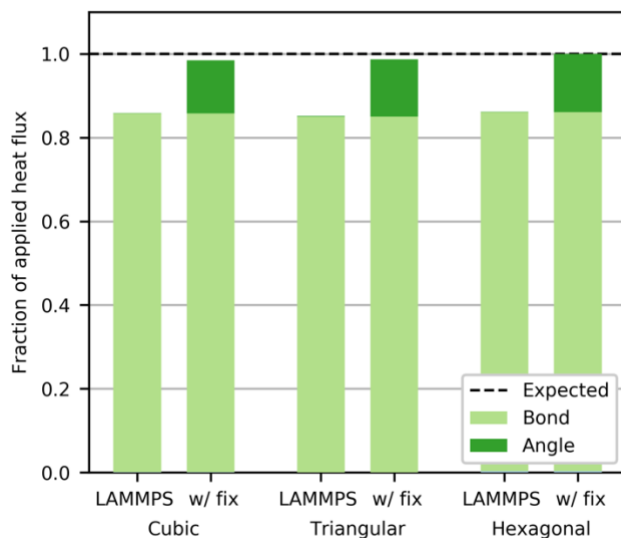


Figure 2: Per-potential fraction of applied heat flux for both uncorrected LAMMPS and corrected calculations for idealized cubic, triangular and hexagonal MOFs.

Because the lattice has zero aggregate momentum, there should be negligible heat transport through convection and all of the heat flux should travel through the bond and angle potentials. As measured for all three systems, the summation of the bond and angle heat fluxes from the corrected code equals the expected heat flux as calculated from the applied energy (see Figure 2). The uncorrected LAMMPS code shows nearly zero heat flux through the angle potential and the error of the total flux is equivalent to the missing angle potential, or about 15% of the expected total flux.

2.4.2 Example System 2: Propane, Octane and Hexadecane

To demonstrate the corrected 3- and 4-body potentials work in conjunction to correctly predict heat flux in a real-life system, we ran simulations for propane, octane and hexadecane. Simulation parameters were adapted from Ohara, et al⁶¹ in order to compare directly with their per-potential results; a brief description of the parameters follows. Each hydrocarbon was defined using the united atom NERD force field⁶² and packed into the rectangular cuboid simulation box using Packmol⁶³ to the density expected at a temperature of 0.7 times the critical temperature (see Figure 3). The simulation box was set to be $3L \times 3L \times 20L$, where L is a hydrocarbon-specific length equal to $1/3$ of the total of the length of the hydrocarbon + a buffer of 3 \AA + the Lennard-Jones cutoff of 13.8 \AA defined by the NERD forcefield. For propane, octane and hexadecane, $3L = 19.38 \text{ \AA}$, 25.84 \AA , and 36.17 \AA , respectively. A timestep of 1 fs was used for all simulations. The system was equilibrated using NVT/NVE for 5M total timesteps; this larger-than-typical equilibration time is because we wanted a longer baseline for statistical averaging for comparing heat flux measurements before and after the application of a heat flux. After equilibration, an energy source and sink were applied of $9.4\text{E-}04 \text{ kcal / mole}$ for propane, $8.3\text{E-}04 \text{ kcal / mole fs}$ for octane, and $6.5\text{E-}04 \text{ kcal / mole fs}$ for hexadecane in order to get an appropriate temperature profile. After 10M timesteps, stable temperature profiles were obtained and then heat flux data was recorded and averaged over 64M timesteps for propane, 132M timesteps for octane, and 234M timesteps for hexadecane.

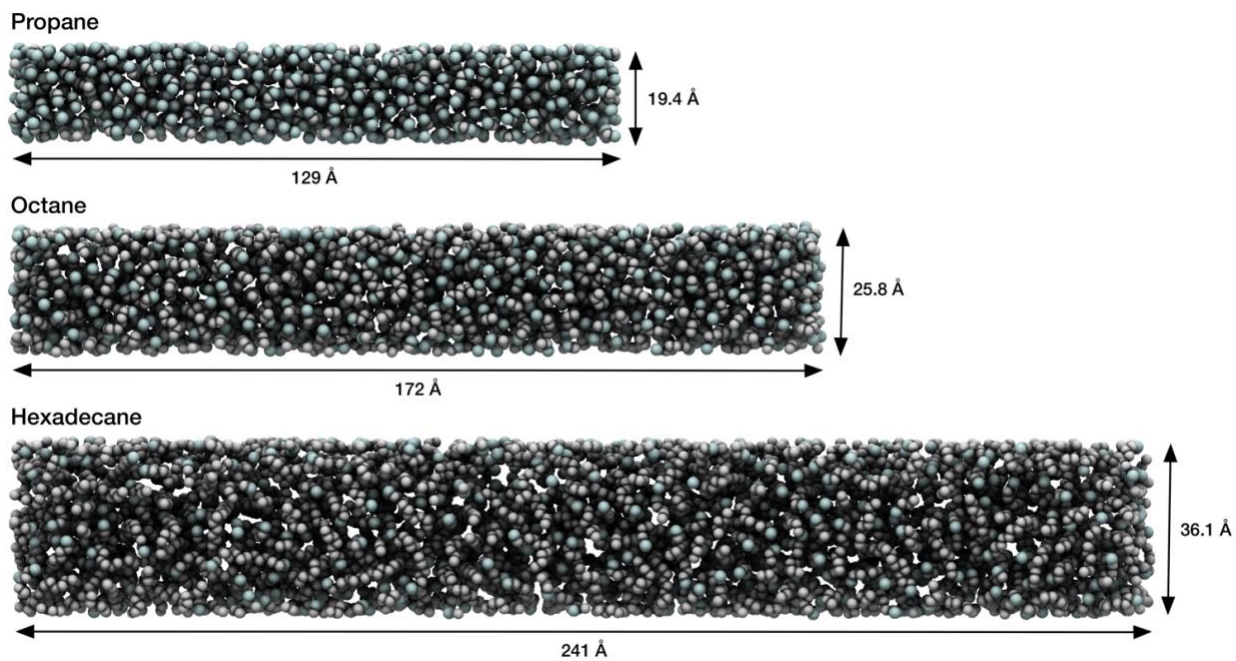


Figure 3: Simulation box layout for propane, octane and hexadecane.

Similar to the case of the idealized MOFs, the summation of all the terms in the corrected heat flux code approximately equals the expected heat flux, regardless of the length of the hydrocarbon (see Figure 4).

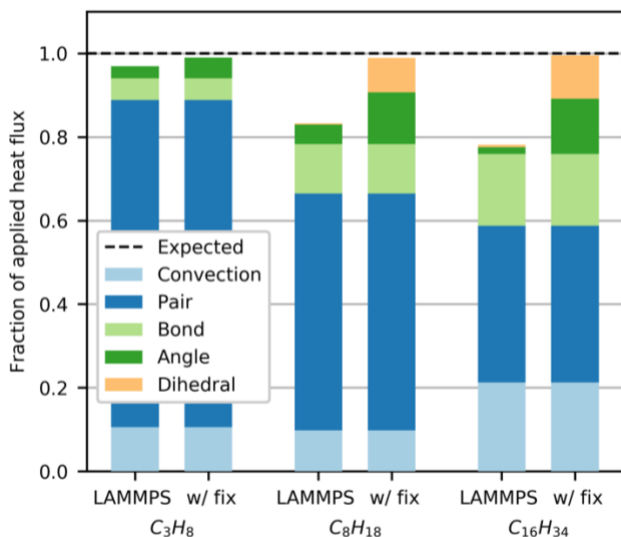


Figure 4: Per-potential fraction of applied heat flux for both uncorrected LAMMPS and corrected calculations for various hydrocarbons. As the length of the hydrocarbon increases, the heat transfer through the many-body potentials increases. $C_{16}H_{34}$ shows LAMMPS-reported heat fluxes of near zero for the angle and dihedral potentials, causing a total error of about 22%.

For the uncorrected code, as the length of the hydrocarbon increases, the amount of heat transfer through the angle and dihedral potentials increases, leading to greater errors with longer hydrocarbons. For propane, the error is minimal, but for octane, the error is greater than 16% and for hexadecane the error reaches 22%. This compares well to the results of Ohara, et al,⁶¹ which roughly predict increasing dependence on the many-body potentials with increasing hydrocarbon length (see Figure 5). A comparable simulation of octane using improper potentials in place of dihedral potentials was also performed; the improper results were comparable to the dihedral results (see Supporting Information).

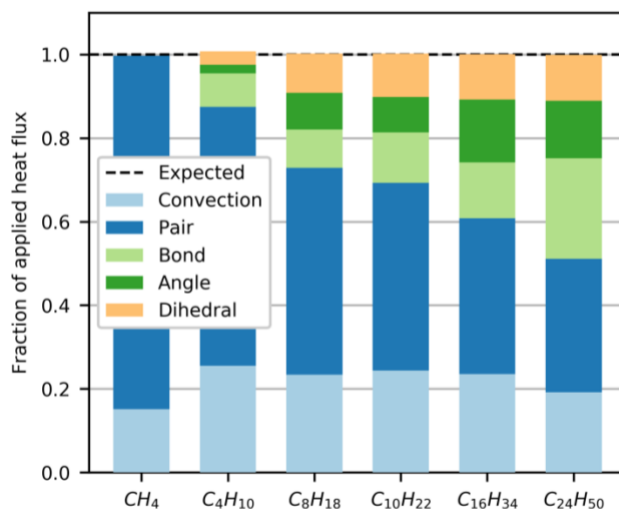


Figure 5: Per-potential fraction of applied heat flux for hydrocarbons from Ohara, et al, which predict increasing reliance on the angle and dihedral many-body terms for heat flux transfer with increasing hydrocarbon length.

In addition to the total error of 16% recorded for octane, the heat flux predicted by uncorrected LAMMPS shows greater swings in magnitude and visually doesn't converge as clearly (see Figure 6).

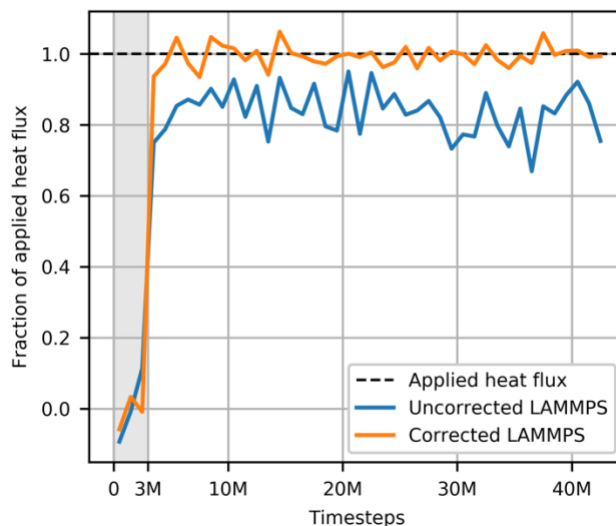


Figure 6: Total heat flux as a fraction of applied heat flux for both the uncorrected LAMMPS calculation (blue) and the corrected LAMMPS calculation (orange), averaged over 1M timesteps, as measured on octane. Prior to 3M timesteps, there is no applied heat flux (noted in grey) and the recorded instantaneous heat flux fluctuates around zero; after 3M timesteps, the heat flux is applied and the corrected calculation fluctuates around the correct value but the uncorrected calculation underestimates the real value as well as shows greater magnitude fluctuations.

Any error in the long-term average of the heat flux, or any error that decorrelates the instantaneous heat flux may have consequences when calculating thermal conductivity via Green-Kubo; more research needs to be done to assess which systems are significantly affected by this error in the heat flux calculations.

2.5 Conclusion

For systems where significant heat transfer occurs within many-body potentials, the error in the uncorrected LAMMPS heat flux code can reach nearly 100% of the many-body potential heat flux, which leads to an erroneous total heat flux calculation. From our example systems, the

largest error in the total heat flux we saw was 22%, though it is not hard to imagine a system where more heat is transferred via the many-body potentials; in that case, the error would be expected to be proportionally higher. Care should be exercised when evaluating prior results to ensure thermal conductivity calculations are not erroneous due to the incorrectly defined heat flux. We have implemented the corrected algorithm into LAMMPS which extends the accuracy of heat flux measurements in LAMMPS to the 3- and 4-body potentials.

For deciding whether it is necessary to adopt the corrected code for your calculations, we recommend evaluating which one of these three cases your work fits into: (1) if you are modeling a system composed of only 2-body potentials, you can use the uncorrected LAMMPS heat flux calculation safely; (2) if you are modeling a system that uses 3- or 4-body potentials implemented via the angle, dihedral or improper potential styles, we recommend you adopt our corrected LAMMPS code available at <https://github.com/wilmerlab/lammps>; and (3) if you are modeling a system with any other many-body potentials, you will need to evaluate whether the potentials have correct or incorrect implementations of the instantaneous heat flux.

3.0 Towards Comprehensive Exploration of the Physisorption Space in Porous Pseudomaterials Using an Iterative Mutation Search Algorithm

Paul Boone and Christopher E. Wilmer

Department of Chemical and Petroleum Engineering, University of Pittsburgh,
3700 O’Hara Street, Pittsburgh, Pennsylvania 15261

We describe an updated algorithm for efficiently exploring structure-property spaces relating to physisorption of gases in porous materials. This algorithm uses previously described “pseudomaterials,” which are crystals of randomly arranged and parameterized Lennard-Jones spheres, and combines it with a new *iterative mutation exploration* method. This algorithm is significantly more efficient at sampling the structure-property space than previously reported methods. For the sake of benchmarking to prior work, we apply this method to exploring methane adsorption at 35 bar (298 K) and void fraction as the main structure-property combination. We demonstrate the effect and importance of the changes that were required to increase efficiency over prior methods. The most important changes were (1) using “discrete” mutations less often, (2) decreasing degrees of freedom, and (3) removing biasing from mutations on bounded parameters.

3.1 Introduction

A reasonable starting point to evaluate a new gas adsorption problem is to simulate a range of relevant materials and observe the resulting structure-property relationships. From these

relationships, certain design rules may present themselves. We can group strategies for finding such structure-property relationships broadly into two categories: screening databases of previously reported materials,^{64,65} and screening novel hypothetical structures generated using some algorithm.^{29,37,66–70}

Screening already known materials to discover new structure-property relationships can present certain challenges: insufficient number of materials, impracticality of preparing suitable simulation input files from the experimental data in the quantity desired, inadequate material diversity, over- or under-representation of certain kinds of materials,³⁵ etc. Additionally, by definition, looking only at known materials cannot capture the properties of materials that have yet to be discovered. For some material classes that have been sufficiently and thoroughly explored, this may not be a problem, but for underexplored material classes, it is likely that materials exist that have properties unlike what has already been discovered.

Generating novel hypothetical materials, however, comes with its own set of (arguably even more daunting) challenges: structures may not have sufficient diversity (pore size, space groups, symmetry, atom types, atom positions, etc.) for the space to be well sampled, certain kinds of structures may be oversampled causing redundant and unnecessary computational evaluations and requiring huge numbers of structures to be generated to find the ones with interesting properties, and it is tremendously difficult to generate materials that are likely to be experimentally synthesizable.

Strategies used to generate hypothetical materials are usually specific to the material classes, and numerous methods in the last decade have been proposed for MOFs,^{66,67,71–73} Zeolites,^{68,69,74} PPNs,³⁷ and ZIFs.²⁹ It is important to note that both the limits of the materials class and the limits of the generating algorithm will determine how much of the space is explored by

any of these algorithms; Simon et al.⁷⁵ performed a meta screening of four databases^{29,37,66,68} of generated materials plus the CoRE MOF database,⁶⁴ which contains only experimentally synthesized MOFs, and found that while the results overlapped somewhat between generating algorithms, each database explored different parts of the structure-property space. For inferring structure-property relationships across all material classes, we either need to look at aggregate data across materials classes, or we need a more general generating strategy.

In the context of the physisorption of non-polar gases (e.g., Xe/Kr, CH₄) in porous crystals, we have previously described such a general approach^{76,77} based on starting with randomly generated porous “pseudomaterials” and mutating them. This approach is general, in principle, for describing any class of crystalline materials and automatically generates pseudomaterials in the lesser-explored parts of a user-defined property space. Although our prior work could be used to find useful structure-property relationships for gas storage applications, the approach was highly inefficient; a large number of pseudomaterials were generated in regions of the structure-property space that were already well explored. In this paper, we update our iterative mutation exploration (IME) methodology for exploring arbitrary structure-property spaces to be dramatically more efficient than prior methods and apply it to a well-studied benchmark application of methane loading at 35 bar and 298 K. Additionally, we analyze how adjusting the algorithm parameters affects both the efficiency and the theoretical range of exploration. The long-term purpose of this work is to automate the rapid exploration of structure-property relationships for new gas adsorption problems.

3.2 Methodology

3.2.1 Pseudomaterials

IME is an algorithm that searches through randomly generated and mutated pseudomaterials, which are abstract representations of rigid crystalline materials. Pseudomaterials are defined by a set of atom positions, atom type assignments for each atom, a unit cell vector, and force field parameters for pairwise forces. A pseudomaterial's atoms may be arranged in space arbitrarily—there is no requirement that the atoms form known bonds, form a stable structure or that the pseudomaterial is synthesizable. All pairwise force field parameters may take any value in a defined continuous range—they do not have to correspond to known atom types or parameter values. Pseudomaterials are, like the crystalline materials they are meant to emulate, periodic. Atoms are held at fixed coordinates during molecular simulations and thus these pseudomaterials do not model flexible structures. When using a sufficiently wide-ranging parameter set, pseudomaterials can be a superset of all possible real rigid crystals.

In this work, for the sake of simplicity, the unit cell vector is limited to be cubic, and the pairwise forces for each atom type are defined by the two parameters from the Lennard-Jones potential: epsilon and sigma.

Unlike real materials, pseudomaterials can be easily randomly generated—within the allowable ranges for their properties—by randomly choosing the number of atoms, atom coordinates x , y and z , Lennard-Jones sigma and epsilon values for each atom type, the unit cell size, and then randomly assigning each atom to an atom type. We refer to this (non-iterative) process as **random generation** and we will use it to benchmark the efficiency of IME.

Pseudomaterials can also be continuously mutated across the allowable ranges for sigma, epsilon, the unit cell vector, and the atom positions. Because these parameters are continuous, we can observe smooth relationships between the pseudomaterial parameters and the simulated properties that interest us. In other words, a small change in the input parameter space will yield small changes in the simulated property space; roughly, similar structures share similar properties. Note that the number of atoms and the atom type assignments cannot change continuously and small changes may result in structures with very different simulated properties, which is discussed in more detail in Section 3.3.2.1.

3.2.2 Overview of Iterative Mutation Exploration (IME)

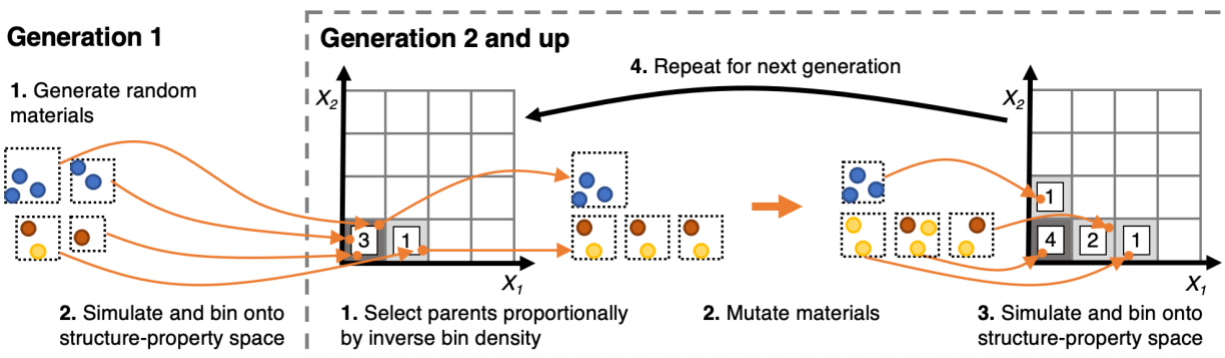


Figure 7: How the iterative mutation exploration strategy works on a simplified 4x4 structure property space

X_1 vs X_2 : the first generation of pseudomaterials uses random generation; subsequent generations are created through an iterative select-mutate-simulate strategy where the parents are selected by inverse bin density (the number shown in each bin is the number of materials in that bin).

We can now automate the search of a gas adsorption structure-property space by first using random generation to create a starting set of pseudomaterials, and then mutating their parameters to adaptively populate under-explored regions of the space.

The high-level overview of IME shown in Figure 7 is identical to our prior work.⁷⁶ The changes we made to improve the efficiency are in the details of how materials are mutated and the allowable ranges of the parameters. For explicitness, we will still describe IME, before discussing the parameter ranges, how mutations are now implemented, and then show example results. After examining the example results for a full run, we will discuss in detail how our new methodology improves upon the efficiency of our prior search algorithm.

Each IME run is governed by run parameters: the number of materials per generation, N , the number of generations, G , the number of bins per simulated property, B , and a mutation strength, M , which determines how much of a change a mutation should make to a pseudomaterial. A set of simulated properties to explore—or structure-property space—is also necessary; each simulated property must be calculable and have defined bounds. The defined range of each property is evenly divided into B property-specific bins. A bin in the structure-property space is defined to be the combination of all property-specific bins. For example, if we have $B=10$ and two simulated properties (a two-dimensional structure-property space), we would have 100 bins; if we had three simulated properties (a three-dimensional structure-property space), we would have 1000 bins.

The first generation of N materials (in this work, $N=50$), or the initial seed population, is created using random generation. Each simulated property is calculated and binned. For subsequent generations, we start by choosing *parent* bins from the structure-property space by inverse population density; we select N parent pseudomaterials by weighting all explored bins by how few materials they contain, then randomly selecting specific pseudomaterials from those bins according to those weights. Then for each selected parent pseudomaterial, we create a child pseudomaterial by mutating its parameters—in this case, the number of atoms, atom positions, atom

type assignments, unit cell size, and all atom type sigma and epsilon values—within their defined ranges. The child pseudomaterials can then be simulated and binned onto the structure-property space, which completes the generation. We repeat this process until we reach the desired number of generations. Note that different convergence criteria could also be used to halt the IME algorithm, but in this work we only considered fixed numbers of generations.

All source code for our IME algorithm is available on github: https://github.com/WilmerLab/htsohm/tree/comprehensive_exploration_methane.

3.2.3 Simulated Structure-Property Space

Our structure-property space in this work is defined by two simulated properties: (1) volumetric methane loading at 298K and 35 bar, and (2) void fraction. We chose these properties because methane loading has been extensively explored,^{66,68,75,78,79} including in our most relevant prior work which used randomly generated materials (no mutation-based exploration) with larger unit cell sizes,⁷⁷ and we wanted to benchmark our results in the context of that large body of extant literature.

We calculated the absolute volumetric loading as reported via RASPA⁸⁰ using the grand Canonical Monte Carlo (GCMC) technique.⁸¹ Methane is modeled using the TraPPE force-field,⁸² which has one united atom for the carbon and four hydrogens. The initial number of cycles for a simulation is 100; if the reported error is more than 5 V/V then an additional 100 cycles is run until either the error is < 5 V/V or the maximum total cycles of 2000 is exceeded. Note that we are able to use such low cycle numbers because our pseudomaterials are relatively simple and methane as defined has no rotational degrees of freedom.

Void fraction is calculated using a simple method developed specifically for this work. The unit cell is divided up into 0.1 \AA cubelets. For every atom in the pseudomaterial, we take the sigma value of the atom type and use it to represent the atomic radius. All the cubelets less than one atomic radius plus a probe size of 1 \AA from the atom center are marked as occupied. The void fraction is calculated as the # of empty cubelets divided by the total number of cubelets. We used this method instead of the helium void fraction method reported in similar prior work, which is intended to mimic the experimental helium adsorption experiment and can give unusual void fraction values (i.e., greater than one) when interactions between the probe and framework are very strong (as is the case for many of the pseudomaterials in our study). Hence, we implemented this simple void fraction routine that ignores energetic interactions altogether and hence always gives results within 0-1.

For both methane loading and void fraction, we use $B = 40$ bins, totaling 1600 bins across both simulated properties. The defined bounds for methane loading are 0-1600 V/V and the defined bounds for void fraction are 0-1.

3.2.4 Pseudomaterial Parameters and Run Parameters

We will describe below many different IME runs using different parameter sets, but they are all variations from a baseline parameter set, which is defined as follows (also listed in Table 1): 1-4 atoms per unit cell, two atom types, sigma and epsilon ranges equal to the total range in UFF,⁸³ a mutation strength of 20%, and a lattice size of 2-16 \AA . The lattice size range of 2-16 \AA was chosen so that a one atom system in the smallest unit cell could achieve a number density greater than the highest number density MOF ($0.107 \text{ atoms} / \text{\AA}^3$) in the CoRE database,⁶⁴ and so that at the opposite range of lattice size, 16 \AA , a four-atom system could achieve a number density

lower than 1/100 of the highest number density— 1.07e^{-3} atoms / \AA^3 —a significantly lower density than the lowest number density MOF (0.016 atoms / \AA^3) in the CoRE database.⁶⁴ The number of atoms was chosen to be small to minimize the degrees of freedom to improve search efficiency (see discussion in Section 3.3.2.2), and the number of atom types is set to two so that there can be some variation in type and still limit the degrees of freedom. All runs use the default *number of materials per generation*, $N=50$, and the *number of generations*, $G=500$, yielding 25,000 materials in total per run. The default mutation strength is 20%. Other possible values for the number of atoms, the atom types, the sigma and epsilon ranges, and mutation strength are explored further in Section 3.3.3.

3.2.5 Mutation Strategies

For this work, we define three mutation strategies: wrapped mutations, infrequent mutations, and bounded mutations.

Wrapped mutations are applied to atom positions. For these mutations, a random point is generated within the unit cell and the shortest vector from the current atom position to this new point is calculated. Then, this vector is multiplied by the mutation strength and added to the original atom position to get the new position. If the new position is outside the unit cell, it is “wrapped” back inside the unit cell. For this strategy, if the mutation strength is 100%, then every atom would effectively get a new position with every mutation. As the mutation strength decreases, the atom positions of a child material will become more localized near its parent’s atom positions. Atom positions are mutated every generation.

When changing atom types or the number of atoms in the unit cell of a pseudomaterial, we use so-called “infrequent” mutations; instead of being mutated every generation, the property is

mutated based on a probability $p = M^2$ for atom type assignments and $p = M$ for the number of atoms, where M is the mutation strength. For each generation, at a default mutation strength of 20%, there is a 20% chance that the number of atoms is mutated, and each atom has a $0.20^2=0.04=4\%$ chance that its type assignment is mutated. When the number of atoms is mutated, there is a 50% chance to add an atom and a 50% chance to remove an atom; if adding an atom would lead to more atoms than allowed, or if removing an atom would lead to fewer atoms than allowed, then no change is made. A newly added atom is randomly assigned a position and an atom type. For atom type assignment mutations, a new atom type is randomly selected from all atom types.

Table 1: Mutation strategy and ranges for this work and our prior work.

	This work		Prior work	
	Mutation strategy	Ranges	Mutation strategy	Ranges
atom positions	wrapped	fractional	wrapped	fractional
# atoms	infrequent (M), ± 1 atom	1 - 4	density	1 - 2701
atom type assignments	infrequent (M^2)	2	biased	4
sigma [\AA]	bounded	2.11 - 4.37	biased	2.11 - 4.37
epsilon [K]	bounded	2.516 - 342.176	biased	2.516 - 342.176
lattice [\AA]	bounded	2.0 - 16.0	biased	25.6 - 51.2

All other parameter mutations—unit cell length, sigma, and epsilon—use the bounded mutation strategy. We randomly choose a trial value from a uniform probability distribution centered around the starting value; if the trial is within the allowable range it is used, otherwise it is bounded to the allowable range:

$$V_{trial} = \text{uniform}\left(V_{parent} \pm \frac{M}{2} * (V_{max} - V_{min})\right) \quad (3 - 1)$$

$$V_{child} = V_{min} \text{ if } V_{trial} < V_{min}$$

$$V_{child} = V_{max} \text{ if } V_{trial} > V_{max}$$

$$V_{child} = V_{trial} \text{ otherwise}$$

Here, V_{parent} is the variable to mutate, V_{max} and V_{min} are the variable's maximum and minimum, V_{trial} is the trial variable and V_{child} is the final child variable. If the mutation strength is 100% and the starting point was in the exact center of the allowable range, then this is equivalent to randomly generating a new value within the range.

The density and biased mutation strategies shown in Table 1 from our prior work are discussed in Section 3.3.2.1 and 3.3.2.3.

3.3 Results and Discussion

3.3.1 Results for baseline CH₄ vs VF parameters

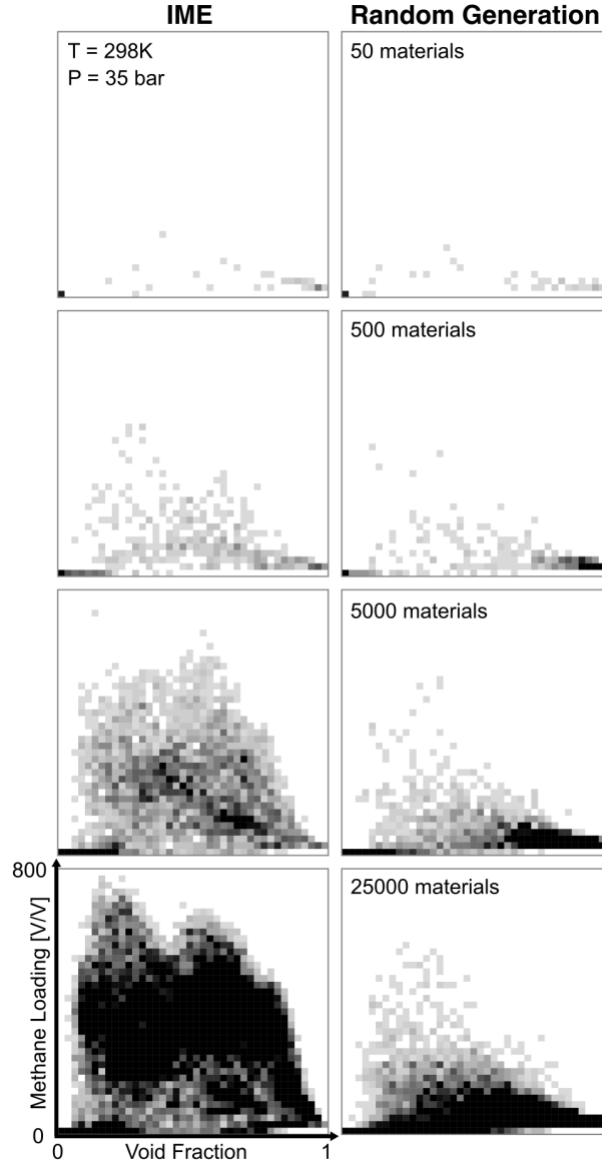


Figure 8: Explored material density bin map (white: unexplored bin; black: 20 or more materials per bin; grey: 1 to 19 materials per bin) of methane loading vs void fraction for both IME and random generation at 50, 500, 5,000 and 25,000 materials.

In Figure 8, we show the methane adsorption vs. void fraction structure-property space at 50, 500, 5,000, and 25,000 materials for both IME and random generation. The first generation of 50 materials is comparable between methods, since the first generation of IME uses random generation. At 500, 5000, and 25,000 materials, IME explores more bins than random generation. While continuing to generate materials using random generation would eventually explore the same bins if given infinite time, most of the generated materials are within a half-elliptical region under 200 V/V and materials farther away from this region are progressively less likely to be generated. Notably, the high methane loading materials found via IME are among the most statistically unlikely to be created using random generation.

We ran random generation until we reached 500K materials (Figure 9). At 25,000 materials, random generation has explored 468 bins; IME explores this number of bins at only 1,786 materials. At 500,000 materials, random generation has explored 732 bins; IME takes only 4,283 materials to reach the same number of bins.

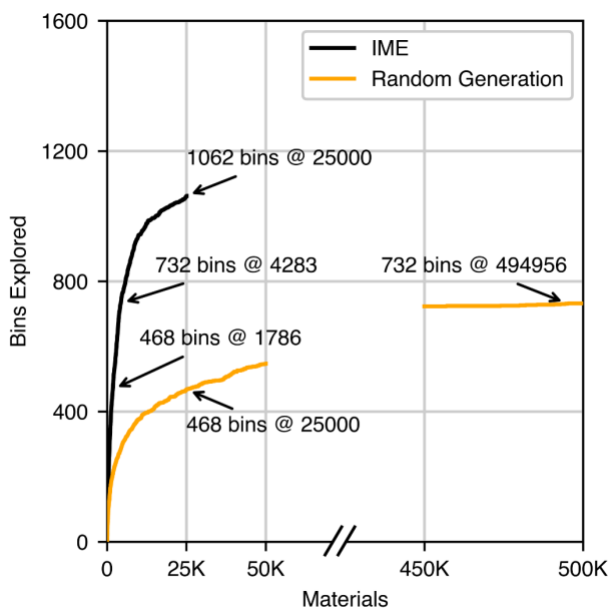


Figure 9: Number of bins explored vs. the number of materials for IME and random generation.

Figure 10 zooms in on the space explored by IME at 25,000 materials (the lower left plot of Figure 8), but shows each material as a separate point instead of showing bin density. Here we see there are two large peaks, at around the void fractions 0.2 and 0.5. The blue peak on the left is formed primarily by materials with small unit cell sizes being loaded by one methane / unit cell. The green peak on the right is from materials with slightly larger unit cells loaded to three methane / unit cell. A smaller third peak (yellow), though not as well resolved, can be seen around a void fraction of 0.8, which may be related to the peak found in prior large-scale screening studies on high pressure methane adsorption.^{64,66,77} The materials with the highest methane loading in the leftmost peak have a noticeable absence of diversity in their structures: they are all high-epsilon, small unit cell, cubic structures and on closer inspection, you can see that there are multiple atoms clustered together in one location, forming an extra-high epsilon “superatom.” The three CH₄ / unit cell (green) peak and the four CH₄ / unit cell (yellow) peak are also similarly constructed from high epsilon superatoms, but at larger lattice sizes so that more methane can be adsorbed per unit cell (even if less per unit of volume).

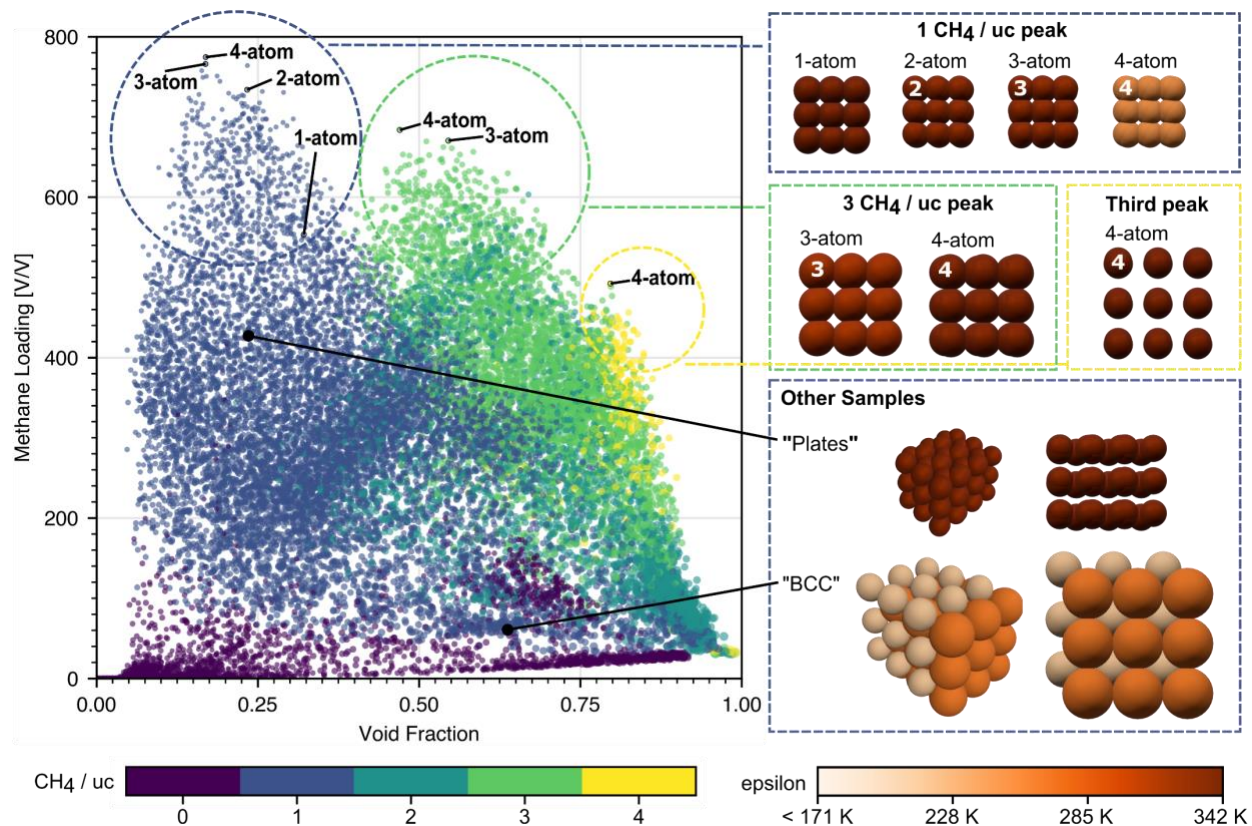


Figure 10: Methane loading vs void fraction for all materials in the reference run of IME, colored by the number of average adsorbed methane per unit cell at 60% opacity, and relatively sized by the lattice size. Example materials are shown on the right, with each atom site colored by atom type epsilon.

For our reference simulation, the number of atoms per unit cell is limited to 1-4 atoms and if we look separately at the materials with the highest methane loading for each of one-, two-, three- and four-atom materials (see Table 2), they all contain overlapped superatoms (with the exception of the one-atom material) where only one methane can be adsorbed per unit cell. The four-atom material with the highest methane loading and the three-atom and two-atom materials below are, except for the number of atoms, otherwise nearly identical materials; all of their atoms share the same sigma (i.e., size), they have nearly the same unit cell dimensions, and they all adsorb, on average, nearly one methane per unit cell. To compare multi-atom materials that have overlapping atoms with single-atom materials, we found it helpful to devise the *epsilon density* metric

(denoted $\frac{\Sigma \epsilon}{V}$ in Table 2), which we defined as the sum of the epsilon for all atoms in the material divided by the unit cell volume. Compared to the four-atom material (lattice size 3.63 Å, epsilon density 20.73 K/Å³), the two-atom material has nearly the same methane loading but a lower epsilon density of 13.47 K/Å³, showing that this epsilon density is still high enough to achieve an average methane per unit cell near one at a lattice size near 3.63 Å. The highest methane-loading one-atom material has an epsilon density of 5.54 K/Å³—the maximum for one atom—and a lattice size of 3.95 Å, which indicates that without having overlapping atoms or defining a higher epsilon range, it is not possible to achieve the higher loadings in the tighter cells.

The highest methane-loading shown in Table 2 is 774 V/V, which is greater than the density of solid methane. If one can choose arbitrarily high epsilon values and an appropriate sigma value, then it is possible to achieve arbitrarily high methane loadings; the attraction between the framework and methane will overcome the methane-methane repulsive forces. Because of the high epsilon superatom clusters, these high methane loading pseudomaterials are also very high density lattices where the epsilon density is high enough to force the adsorbed methane into a configuration denser than solid methane. Models of lattices loaded more densely than solid methane have been reported in more detail by other authors;⁸⁴ this is an expected—though clearly non-physical—result of our model and simulation parameters.

Table 2: Unit cell, sigma, epsilon density, adsorbed methane per unit cell and methane loading for selected materials in Figure 10.

Category	Name	unit cell [\AA]	sigma [\AA]	$\sum \epsilon/V$ [$\text{K}/\text{\AA}^3$]	# CH ₄ / uc	ML [V/V]
1 CH₄ / uc peak	1-atom (id: 3481)	3.95	2.4	5.54	0.91	552
	2-atom (id: 24249)	3.66	2.1	13.47	0.96	734
	3-atom (id: 13499)	3.60	2.1	21.21	0.96	766
	4-atom (id: 14447)	3.63	2.1	20.73	1.00	774
3 CH₄ / uc peak	3-atom (id: 2844)	5.45	3.0	5.88	2.92	670
	4-atom (id: 22900)	5.44	2.7	8.39	2.96	684
Third peak	4-atom (id: 21214)	6.52	2.3	4.93	3.67	492
Other Samples	Plates (id:15134)	5.26	2.2	9.32	1.70	424
	BCC (id: 23030)	8.07	3.0, 4.2	0.92	0.96	68

Amongst all the created pseudomaterials, we can find examples of plate-like materials (see “plates”), and BCC materials (see “BCC2”). These different material geometries show up without any biasing on the part of the algorithm; the algorithm is not specifically searching for materials with diverse structures, although we are interested in doing this in the future.

3.3.2 Improvements from Prior Algorithm

The results above show that our updated IME algorithm can explore the structure-property space much more efficiently than purely random generation and is also capable of finding locally optimized structures, such as shown by the two separate peaks for one CH₄ / unit cell and three CH₄ / unit cell materials. This updated algorithm is much more efficient than our prior work in particular because of several significant changes: (1) reducing the frequency of “discrete”

mutations, (2) reducing the degrees of freedom that define the pseudomaterials, and (3) correcting a bias in mutations on parameters with bounded upper/lower limits.

3.3.2.1 Using “Discrete” Mutations Less Often

In our simulations there are two cases where we apply “discrete” mutations: adding or removing atoms and reassigning atom types. Unlike, for example, mutations that assign atoms to new positions within the pseudomaterial, such “discrete” mutations cannot be made to minimally perturb the properties of the material. Either an atom is there, or it is not, and likewise an atom type can be only one of a select few types. Therefore, applying one of these discrete mutations often leads to a child material not landing in a bin near its parent in the structure-property space. As a consequence, such mutations prevent the IME method from exploring the local bins near a parent material.

For example, in our prior work, we allowed the *atom number density* to mutate. For each generation, the atom number density was mutated and a corresponding number of atoms would be removed or added (this is the density mutation strategy listed in Table 1). Based on the mutation strength we used, this caused a large number of atoms to be either deleted or added with new random positions whenever a child material was created. For example, for a lattice size of 25-50 Å with an atom number density range of $1\text{e-}5$ to $2\text{e-}2$ atoms / Å³, an average pseudomaterial in the middle of both the lattice size and atom number density ranges—i.e. 37.5 Å and $1\text{e-}2$ atoms / Å — would start with 527 atoms. With a 20% mutation strength the worst case scenario would be the lattice size increasing by 10% of the range or 2.5 Å, and the atom number density increasing by 10% of the range or $2\text{e-}3$ atoms / Å³; this would jump the number of atoms from 527 to 768, an addition of 241 atoms with randomly generated positions. Even with lower mutation strengths, where the number of atoms added or removed was small, such changes typically led to child

pseudomaterials far from their parents in the structure-property space. Furthermore, such large changes also caused child materials to regress to the mean, i.e., they would resemble purely randomly generated materials from the initial generation even if their parents were on the outer boundaries (the less explored frontier bins) of the structure-property space.

To quantify the effect of such discrete mutations on exploration efficiency we systematically varied several run parameters and, in each case, counted how many bins were explored over the same total number of generations (see Figure 11). We let the atom number density range from 0.001954 to 0.11 atoms / Å³. We considered three different *mutation frequencies* that were used to mutate the atom number density: 1.0 (every time, which is how frequently our previously reported algorithm did number density mutations), 0.2 (20% of the time, which is the same as the mutation strength M , which is how frequently we do atom number mutations with our new algorithm) and 0.04 (4% of the time or M^2). We also varied the *max number of atoms to add or remove*, which limits how many atoms can be added or removed at once during an atom number density mutation. The possible values here are ± 1 (only one atom can be added or removed at a time, which is how our new algorithm works), ± 4 , ± 16 and \pm unlimited (which means that we add or remove as many atoms as are needed to reach the desired atom number density). For simplicity, we kept the number of atom types at one for all of these runs.

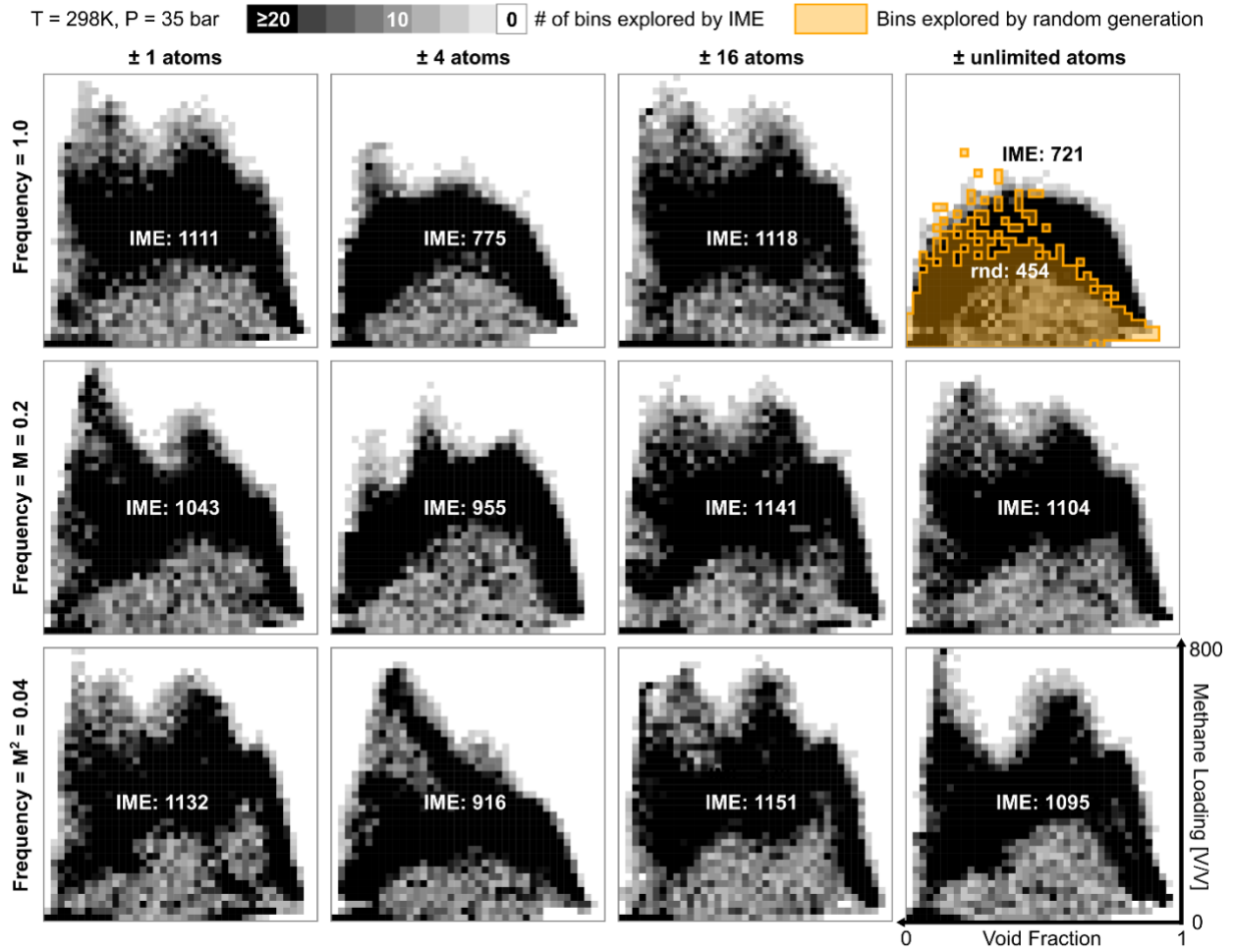


Figure 11: Material density bin map (black: ≥ 20 materials in the bin, white: 0 materials in the bin) of IME run for 500 generations (25,000 total materials) for every combination of max number of atoms to add or remove (± 1 , ± 4 , ± 16 , \pm unlimited) and mutation frequency (1.0, $M = 0.2$, $M^2 = 0.04$) across methane loading vs void fraction. Bins explored by random generation are in orange.

We then ran simulations for every combination of *mutation frequency* and *max number of atoms to add or remove*, as well as a random generation run for comparison. Bin density plots of methane loading vs. void fraction for each parameter are shown in Figure 11.

The parameter set that explores the least amount of space is at \pm unlimited atoms / frequency 1.0, which is the parameter set we used in our prior algorithm. This is because in every generation

significant randomness is introduced by adding or deleting atoms. If you follow the column downwards, at frequency 0.2, the algorithm is able to access the two peaks at one CH₄ / unit cell and three CH₄ / unit cell and significantly more space is explored. At this frequency, 80% of the time there is no change to the number of atoms and the algorithm can optimize the atom positions and the other parameters to explore more space without regressing to more random structures. Farther down at frequency 0.04, the peaks are taller and sharper, implying that since less time is spent changing the number of atoms, local optimizations can be explored more extensively. Higher methane loading materials are found in both peaks compared to frequency 0.2, even though fewer bins are explored. For this example problem of methane loading vs. void fraction, there may be a tradeoff between total bins explored and highest methane loading found that can be adjusted using the mutation frequency.

Similarly, if the max number of atoms is limited to ± 1 , ± 4 , ± 16 , the two peaks can be accessed even if density mutations occur at every generation, suggesting that for this particular system discrete mutations can be managed either by limiting their frequency, or by limiting the extent of the change introduced.

3.3.2.2 Use Fewer Degrees of Freedom

With too many atoms it is difficult, even with iterative mutation schemes, to generate pseudomaterials with the simple lattice configurations shown in Figure 10, which we know are needed to reach some important parts of the methane-adsorption exploration space. While having more atoms increases the theoretically explorable area, there is a significant efficiency cost that limits the actual explored space.

Table 3: Fixed atom IME runs with applied min and max lattice sizes.

# Atoms	Lattice	
	min [Å]	max [Å]
1	2.09	8.00
2	2.63	10.08
4	3.31	12.70
8	4.17	16.00
16	5.26	20.16
32	6.63	25.39
64	8.35	32.00
512	16.70	63.99

To determine the significance of the number of atoms on the exploration efficiency, we ran IME with fixed numbers of atoms and only one atom type, where only the atom positions, the lattice size, epsilon, and sigma were allowed to mutate. We ran separate IME runs for 1, 2, 4, 8, 16, 32, 64, and 512 atoms, as well as an additional series of runs using random generation. Lattice size bounds were chosen for each fixed number of atoms so that each run would explore the same atom number density range of $1.95\text{e-}3\text{-}0.11 \text{ atoms} / \text{\AA}^3$. Parameters for the runs are displayed in Table 3.

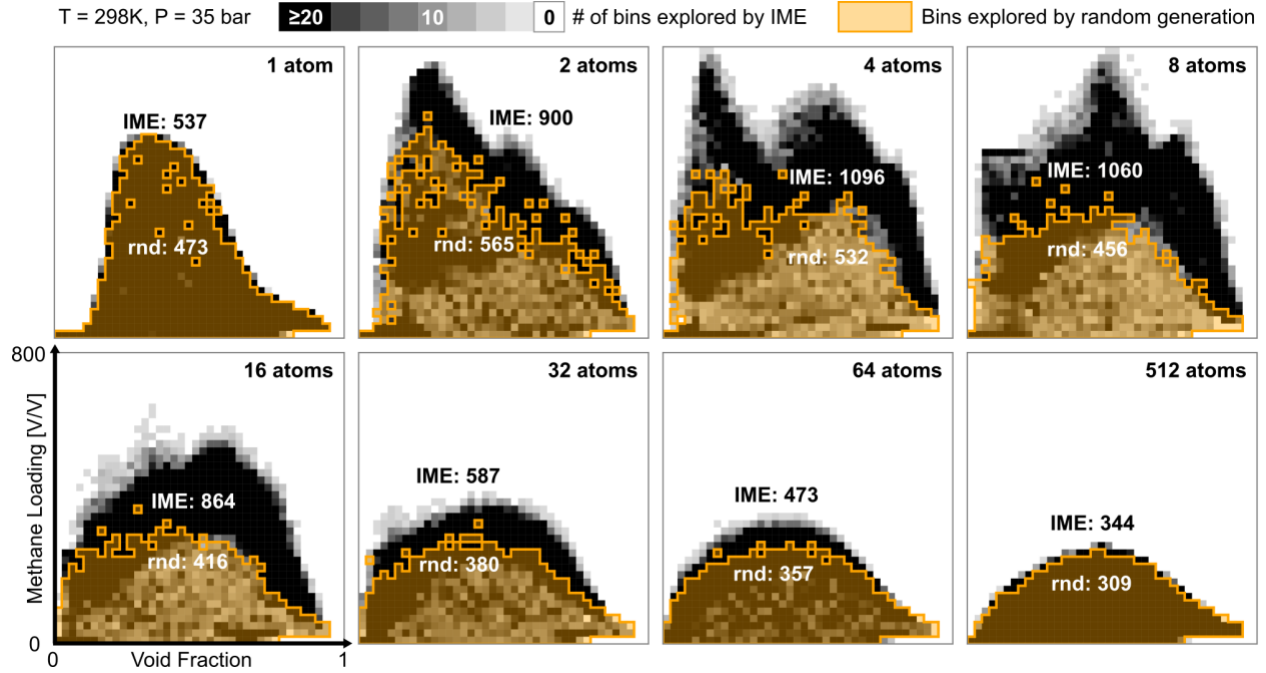


Figure 12: Material density bin map (black: ≥ 20 materials in the bin, white: 0 materials in the bin) of IME run for 500 generations (25,000 total materials) and bins explored by random generation (orange) across methane loading vs. void fraction for 1, 2, 4, 8, 16, 32, 64 and 512 atoms per unit cell.

We can easily see the effect of the number of degrees of freedom on the exploration efficiency in Figure 12. The four atom IME run explores the most bins and as the number of atoms increases, the number of bins explored decreases, showing that the inefficiency of searching with more atoms is overwhelming the theoretically increased explorable area from the additional atoms. This also holds true for random generation.

For both the one-atom and the 512-atom runs, random generation is exploring nearly the same space as IME at 25,000 materials, but the reasons why these techniques look to be converging to the same value are different. For the one-atom run, there are so few degrees of freedom (just lattice size, sigma and epsilon) that it is possible to run a sweep of those variables to get a strict boundary for what is possible; in this case, IME and random are both converging to the actual boundary of the explorable space, just at different speeds due to their computational efficiencies.

While it may appear that IME and random are converging similarly to a boundary for the 512-atom run, in this case we know that the explorable space is larger and looks more like the total space explored in the four-atom run, so here IME and random are running into a computational efficiency limit, not an actual boundary of the space.

The 2- and 8-atom runs show some anomalies that we will briefly address. There are a handful of easy to identify materials that we know must be present in order to have explored the entire structure-property space: a one CH_4 /unit cell structure—corresponding to the leftmost-peak shown in Figure 10—that has a lattice size of about 3.6 Å and very high epsilon density from multiple atoms overlapping at one position, and a three CH_4 /unit cell material—corresponding to the middle peak shown in Figure 10—that has a lattice size of about 5.4 Å created by four overlapped atoms. The two-atom run cannot access the rightmost 5.4 Å lattice size peak since it does not allow for four atoms. The 8-atom run cannot access the leftmost 3.6 Å lattice size peak because its minimum lattice size is 4.17 Å. This peak becomes a “shelf” of 4.17 Å lattice size materials; the range across the void fraction is due to how much space is filled up next to where the methane adsorbs in the middle of the cell.

3.3.2.3 Unbiased Bounded Mutations

In our prior work, we used a simple mutation strategy for bounded parameters that nevertheless led to a subtle inefficiency in the search process. In this strategy, a value V_{rnd} would be chosen from a uniform distribution within the bounded range, and then that value would be averaged (weighted by mutation strength) with the parent’s parameter value to get the new value for the child material:

$$V_{rnd} \sim \text{uniform}(V_{min}, V_{max}) \quad (3 - 2)$$

$$V_{child} = V_{parent} + M(V_{rnd} - V_{parent})$$

This had a side effect we did not intend: as a parameter value got closer to its range limit, almost all mutations would result in a value farther from that limit. To instead get a value closer to the limit, you would have to randomly generate a value closer to the limit than what you already had. This was made somewhat worse by weighting using the mutation strength. Even if you were to randomly generate a value of the exact limit, you would only move towards the limit proportional to the mutation strength. For example, with a mutation strength of 20%, you would only be able to move 1/5 of the way to the limit in the best case. While it is still possible to achieve values very near the bounds of a parameter, it takes many attempts to get there, and the more bounded parameters you have, the more inefficiently the parameter space is explored. Certain areas of the parameter space become statistically hard-to-reach, even if they would easily explore new areas of the structure-property space. For the methane loading vs void fraction structure-property space we are exploring here, this mutation strategy is a second case of an algorithm detail causing child materials to “regress to the mean”; in this case, child materials are more likely to mutate to previously explored areas of the parameter space than to explore new areas, causing significant inefficiency in searching the structure-property space.

To address this regression-to-the-mean bias, we adjusted our mutation strategy for bounded parameters such that regardless of the initial parameter value of the parent material, it is equally likely to mutate higher or lower, the precise details of which are discussed in Section 3.2.5.

Table 4: allowable range for each pseudomaterial parameter for all 12 iterative mutation exploration runs.

Run	# atoms	# atom types	sigma / epsilon	mutation strength	unit cell size
reference system	1-4	2	UFF	20%	2-16 Å
random system	1-4	2	UFF		2-16 Å
# atoms	1	2	UFF	20%	2-16 Å
	1-2	2	UFF	20%	2-16 Å
	1-8	2	UFF	20%	2-16 Å
# atom types	1-4	1	UFF	20%	2-16 Å
	1-4	4	UFF	20%	2-16 Å
sigma / epsilon	1-4	2	UFF±25%	20%	2-16 Å
	1-4	2	UFF±50%	20%	2-16 Å
mutation strength	1-4	2	UFF	5%	2-16 Å
	1-4	2	UFF	10%	2-16 Å
	1-4	2	UFF	40%	2-16 Å

To demonstrate the improvement shown by the new bounded mutation strategy compared to the original biased strategy, we ran an additional IME run of the reference parameters where we used the prior biased mutation strategy. The results can be seen compared to the new mutation with limits strategy for multiple mutation strengths in Figure 13C. At 25,000 materials, the old mutation strategy explored 918 bins compared 1,062 for the new mutation strategy. This is a drop comparable to the difference between running the new mutation strategy at 40% vs. 20% mutation strength.

3.3.3 Parameter Exploration

Finally, to explore the effects that different parameters have on the efficiency of the search and the total explored space, we varied the allowable ranges for the number of atoms, the number of atom types, sigma, epsilon and the mutation strength (ranges for all 12 runs are shown in Table 4). For sigma and epsilon, we test two expanded ranges where the minimums are 25% and 50% lower, and the maximums are 25% and 50% higher than the reference UFF baseline. The epsilon and sigma values for the UFF, UFF+-25% and UFF+-50% are shown in Table 5. For the # of atoms, we test using one atom only, 1-2 atoms, and 1-8 atoms and compare to the 1-4 atoms baseline. For atom types, we test one atom type and four atom types and compare to the two-atom type baseline. For mutation strength we tested 5, 10, and 40% and compare to the 20% reference baseline.

Table 5: Sigma and epsilon values for UFF, UFF+-25%, and UFF+-50%.

	UFF	UFF \pm 25%	UFF \pm 50%
sigma [Å]	2.11-4.37	1.58 - 5.46	1.05-6.55
epsilon [K]	2.516-342.176	1.89 - 427.72	1.258-513.26

In Figure 13, we compare how changing these search parameters affects the efficiency of the search. The number of atom types (Figure 13A) only minimally affected our exploration efficiency. This is because many of the harder-to-reach bins require higher methane loading and will be filled by high epsilon overlapped atom materials; these kinds of materials will be nominally more likely to be generated if there is only one atom type (due to the decrease in degrees of freedom), but due to how infrequent additional atom types are added or removed, having the possibility of more types does not substantially affect the exploration efficiency.

Varying the number of atoms (Figure 13B) shows an increasing number of bins explored from 1 to 1-2 to 1-4 atom. A one atom simulation levels off at around 550 explored bins which is due to the materials having limited epsilon and no positional variation. The two-atom simulation starts leveling off around 900 explored bins for similar reasons, and the four and eight atom simulations access more space but are indistinguishable from each other. As most of the harder-to-reach bins can be accessed by overlapped atom materials, having the possibility of more atoms is theoretically helpful, but for every atom added, it becomes less statistically likely for them to line up near a single point. Simulating more atoms might have an advantage for lower void fraction materials where more space can be filled in such a way that does not significantly affect the methane loading, but according to Figure 13B this does not result in overall more exploration.

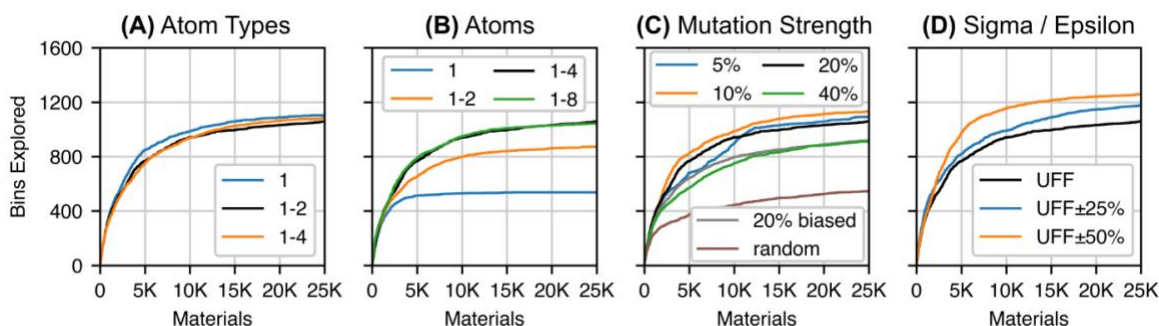


Figure 13: Bins explored vs. number of materials for the pseudomaterial parameter ranges for atom types, atoms, sigma and epsilon as well as the simulation parameter mutation strength.

There is an optimal mutation strength (Figure 13C) for this system. A lower mutation strength of 10% outperforms our default 20%; 5% mutation strength explores fewer bins through 10,000 materials, but after that explores nominally more than our default 20%, and 40% mutation strength performs significantly worse. As the mutation strength approaches 100%, the efficiency will approach random search; as the mutation strength approaches zero, however, child materials become closer and closer replicas of their parents, which also inhibits exploration. While a 10%

mutation strength looks best for this particular set of simulation parameters, our expectation is that this will need to be adjusted on an application-by-application basis, and possibly by starting with higher mutation strengths and decreasing them as the IME run progresses. Also shown is the old perturbation algorithm (labeled 20% biased) as discussed in Section 3.3.2.3. Finally, as expected, if you increase the total range of sigma and epsilon parameters, the explored space increases (Figure 13D).

We have published the data online for the exploration data sets where it can be fully interactively explored here: <http://ch4-pseudomaterials.herokuapp.com/>.

3.4 Conclusion

Our prior mutation algorithm explored a structure-property space less efficiently than merely characterizing pseudomaterials that were randomly generated. With the improvements in this work, our updated IME algorithm is demonstrably more efficient than random generation and can explore areas of the structure-property space that are inaccessible to random generation because of its inefficiency. For our example application of methane loading versus void fraction, random generation of 500K pseudomaterials explored 732 bins, a number reached in only 4,283 pseudomaterials when using IME. This efficiency gain enables us to explore more of the structure-property space, and we can now see examples of previously modeled materials emerging naturally from the IME search process, such as the heavily loaded lattices modeled by other authors.⁸⁴ The efficiency improvements resulted from three algorithm changes: (1) limiting the frequency of discrete mutations, (2) reducing the number of atoms and the degrees of freedom, and (3) removing bias when mutating bounded parameters. With these efficiency updates, this method shows

promise for quickly exploring structure-property spaces for physisorption of various gases in porous materials.

4.0 MOFUN: a Python package for molecular find and replace

Paul Boone and Christopher E. Wilmer

Department of Chemical and Petroleum Engineering, University of Pittsburgh,

3700 O'Hara Street, Pittsburgh, Pennsylvania 15261

MOFUN is an open-source Python package that can *find and replace* molecular substructures in a larger, potentially periodic, system. In the context of molecular simulations, find and replace is a useful operation for adding/swapping functional groups, adding/removing solvent molecules or defect sites, and many other helpful system perturbations. MOFUN can also be used to alter force field terms on certain atoms while leaving the geometry/composition otherwise unchanged. The package is easily automated, which is particularly helpful for preparing input files for large-scale screenings. The package is freely available on GitHub at <https://github.com/WilmerLab/mofun>.

4.1 Introduction

MOFUN is a general purpose, open-source Python package for searching an arbitrary molecular structure for a pattern and replacing any instances of it with a replacement pattern—i.e., find and replace for molecular systems. We built MOFUN, initially, to support our own investigations of metal organic frameworks (MOFs), which are a class of porous materials composed of linkers and metal centers which self-assemble into crystalline structures.⁸⁵ In the

context of MOFs, MOFUN was used for (1) modifying the linkers of MOFs with various functional groups, (2) adding defects to MOFs, and (3) parameterizing MOFs with flexible force field terms. Although these operations can be carried out by editing the underlying files in a text editor or by adding and removing atoms individually in a visual editor such as Avogadro,⁸⁶ this is time-consuming, error-prone and impractical when scaling to greater numbers of structures. We wrote MOFUN to be an automated solution to these problems.

Prior to the development of MOFUN, a less capable molecular search-and-replace program called FunctionalizeThis! was developed by Wilmer.⁸⁷ FunctionalizeThis! did not support finding and replacing bonds, which limited its use particularly when attempting to generate structures that can be used with flexible force fields. More recently, a free and open-source Julia package named PoreMatMod.jl was reported by Henle et al.⁸⁸ Like MOFUN, the highly versatile package by Henle et al. can be used for automating crystal structure modifications and was at least partly motivated to facilitate research on hypothetical MOFs. While there is significant overlap in functionality between MOFUN and PoreMatMod.jl, there are also a few key differences. Whereas MOFUN searches for patterns via comparisons of distances between atoms, PoreMatMod.jl analyzes the molecular graph defined by how a structure's atoms are bonded. Both approaches can handle many common use cases but sometimes one approach is more suitable than the other, in terms of what kinds of patterns can be searched for. For example, when using distances between atoms, it is possible to search for patterns that are not bonded, such as a molecule physisorbed to a binding site or matching defects between two different layers of stacked graphene. In contrast, molecular graph searches are much better suited to searching for substructures in a conformation-invariant manner, such as when looking for hydrocarbon chains that can assume varied configurations in space while their molecular graphs stay the same. We note that MOFUN also supports find and

replace of substructures that come fully parameterized to use with force fields for molecular dynamics (MD) simulations (in particular, for LAMMPS³⁸). By releasing MOFUN as open-source and announcing it here, we hope that other researchers will also benefit from this general-purpose package and can use it to accelerate and expand their research.

MOFUN is available on GitHub at <https://github.com/WilmerLab/mofun> under the open-source MIT license. The version described in this paper is version 1.0. MOFUN can be installed from the GitHub source code, or from PyPI using pip.

4.2 MOFUN: Algorithm Details

In this section, we will discuss the implementation of both the find and the replace parts of the algorithm. We use the word **structure** here to mean a set of atoms with (optional) periodic boundaries that we want to exhaustively examine for instances of a **search pattern**. When we find an instance of the search pattern in the structure, we call this a **match**. Every match found can be replaced with a **replacement pattern**. While we use the word structure here because we predominantly use MOFUN on periodic crystal lattice structures, the software works equally well on molecules or any other aperiodic grouping of atoms.

To find all instances of a search pattern in a structure, we first calculate the distances between all pairs of atoms in the search pattern. For there to be a match of the search pattern in the structure, there must be a set of atoms in the structure that share both the same distances (within a tolerance), and the same atom elements. Let \mathbf{r}_p be an N-length set of all positions in the search pattern. Let \mathbf{r}_s be the set of all atom positions in the structure plus each atom’s periodic images from immediately adjacent unit cells. Let $\mathbf{r}_m \in \binom{\mathbf{r}_s}{N}$ be a subset of positions from the structure of

length N that we will examine as a trial match in the structure. We will refer to specific positions in both \mathbf{r}_p and \mathbf{r}_m as $\mathbf{r}_{p,i}$ and $\mathbf{r}_{m,i}$, where $i \in [1..N]$ refers to the i^{th} element in the set.

The trial match \mathbf{r}_m is a good match if three conditions are met. The first condition is that the distances (or Euclidian norm denoted by $\|\cdot\|$) between all pairs of atoms in the trial pattern must match their corresponding pairs in the search pattern within a specified tolerance δ .

$$\forall i, j \in [1..N], \left| \|\mathbf{r}_{p,i} - \mathbf{r}_{p,j}\| - \|\mathbf{r}_{m,i} - \mathbf{r}_{m,j}\| \right| < \delta \quad (4 - 1)$$

If the distance between any pair of atoms differs from its proposed matching pair by an amount greater than the tolerance, the trial match is not a match. The tolerance can be set higher or lower for cases when a looser or tighter match is appropriate.

The second condition is that the atom elements for the pattern must be the same as the atom elements of the trial match. If we let $E_{p,i}$ be the i th element of the pattern and $E_{m,i}$ be the i th element for the trial match, then:

$$\forall i \in [1..N], E_{p,i} = E_{m,i} \quad (4 - 2)$$

The third condition is that there must exist rotation and translation operations such that when they are applied to the search pattern, the atom positions of the search pattern match the atom positions in the trial match. This condition is necessary to handle cases of symmetry and chirality in the search pattern. For each trial match, we calculate the translation and rotation operations necessary to transform the search pattern to the location of the trial match and then we exclude any matches where the atom positions of the transformed search pattern are not the same as those in the trial match.

To rotate the search pattern into place, we select three points in the search pattern, two to define a direction axis, and the third to use as an orientation point. MOFUN will pick the two atoms in the search pattern that are farthest from each other to define the direction axis and it will

pick the atom that is farthest from the infinite line defined by the direction axis to be the orientation point. If the pattern only contains two atoms, or all the atoms are colinear, then the orientation point can be ignored. Rotating the search pattern to align with the match pattern requires two rotation transformations: (1) we rotate the search pattern so that the two atoms of the direction axis are pointed in the same direction as those same atoms in the match pattern, and (2) we then rotate the search pattern around the directional axis to line up the third atom. Since all the atoms should now be offset by the location of the match in space, we can translate the search pattern by the difference in position between any corresponding pair of atoms between the search pattern and the match pattern.

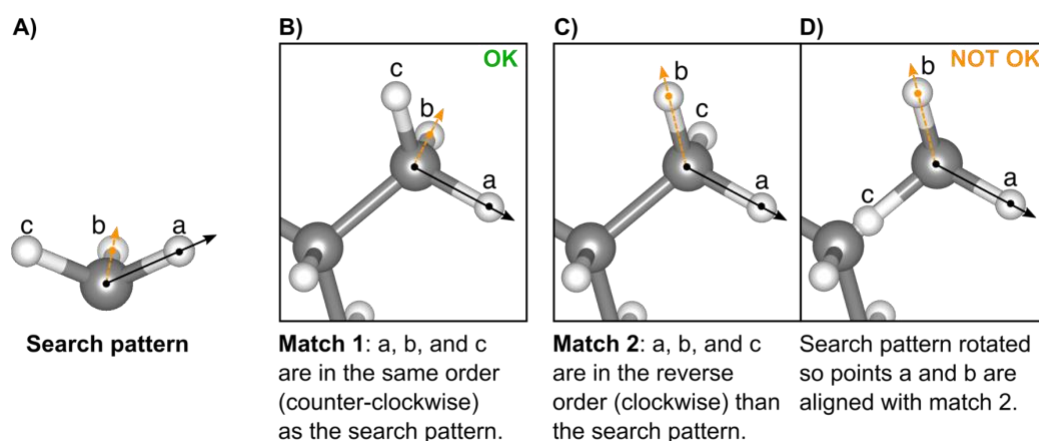


Figure 14: (A) a search pattern for the CH₃ group in Octane with counter-clockwise ordering of hydrogens, (B) a trial match 1 with matching counter-clockwise ordering of hydrogens, (C) a trial match 2 with non-matching clockwise-ordering of hydrogens, and (D) the erroneous result from rotating the directional axis and orientation point into place when starting with the hydrogens ordered in the reverse direction.

To demonstrate the problem posed by symmetry, let us consider searching for the CH₃ group at the ends of an octane molecule. We can define the CH₃ search pattern where the directional axis is the axis from the carbon to the hydrogen labeled 'a', and the orientation point is the hydrogen labeled 'b' (see Figure 14A). If we search an octane for this pattern, we will find six possible matches for each actual CH₃ in the structure (or 12 matches total), one for each possible ordering

of the hydrogens: abc, acb, bac, bca, cab, cba. If we look at the matches that start with the 'a' hydrogen—abc, acb—there is a counter-clockwise ordering of atoms (see Figure 14B) and a clockwise ordering of atoms (see Figure 14C). If the atoms in the match pattern are ordered counter-clockwise like the search pattern then the C-H axis will align the carbon and the 'a' H, the orientation point will line up the 'b' H, and the third H will naturally be in the correct location. However, if the match pattern was numbered clockwise, then the 'b' H in the match pattern is in the location of the 'c' H in our search pattern and aligning it will give us a correctly situated 'b' H, but the 'c' H will be in the wrong location (see Figure 14D). For this example, three of the six possible trial matches have the same counter-clockwise ordering as the search pattern and are good matches, and three trial matches have clockwise ordering and are not matches. Similarly, a chiral search pattern in a structure will match either enantiomorph since the distances between all atoms are the same regardless of which enantiomorph is found, but only an enantiomorph which matches the chirality of the search pattern will be able to be rotated into position of the match pattern.

Thus, the third condition—that the atom positions of a rotated and translated search pattern must be the same as the atom positions of the trial match—removes the bad matches caused by symmetric and chiral search patterns. Once we have all the possible good matches, if we still have multiple matches for the same group of atoms caused by symmetry, then we randomly pick one of the good matches.

At this stage, inserting the replacement atoms is now straightforward. The replacement pattern is defined on the same coordinate system as the search pattern, so to insert the replacement pattern into the structure at the right position and orientation, we take the transformations we calculated above to transform the search pattern into place and apply them to the replacement

pattern. We insert the replacement pattern atoms and topology into the structure, delete the matched atoms and existing topology and our find and replace is complete.

4.3 MOFUN: Optimization and Performance

Here we describe some subtleties to the implementation of MOFUN that were necessary to optimize its performance.

First, depending on the length of the search pattern, we do not search all the atoms in each neighboring unit cell. Searching every atom in a replicated 3x3x3 expanded unit cell could be prohibitively inefficient for larger structures, so we limit the set of atoms searched to only those within a distance d of one of the boundaries of the original unit cell, where d is the length of the search pattern.

Second, we do not generate all trial matches at once as this would lead to running out of memory for all but the smallest systems; instead, we build up trial matches of the search pattern one atom at a time. The first atom in the search pattern is matched by finding all atoms in the structure that share the same element as the first atom in the search pattern. For each of these starting matching structure atoms, we create a set of nearby structure atoms—those atoms that are within a box $x \pm d, y \pm d, z \pm d$ about the matched atom—and precalculate the distances between every pair of atoms in this set. To match the second atom, we find all nearby structure atoms that match the element of the second search atom and select only the atoms where the distance from the second atom to the first atom matches that of the search pattern. Continuing through the search pattern one atom at a time, we match all nearby atoms based on their elements, and select only those atoms where the distances between the atom and all prior atoms match the distances in the

search pattern. At any stage, if there are no viable matches, then we can abort looking for matches using this starting element. If we reach the last atom in the search pattern and have built up one or more complete matches, then these are added to the list of successful matches. Another advantage of generating matches in this manner is that the algorithm can easily be made to operate in parallel, where each starting structure atom is run on a separate core. We have not found it necessary to implement this yet due to MOFUN’s current performance being more-than sufficient for our needs but parallelization is available to us if it becomes necessary.

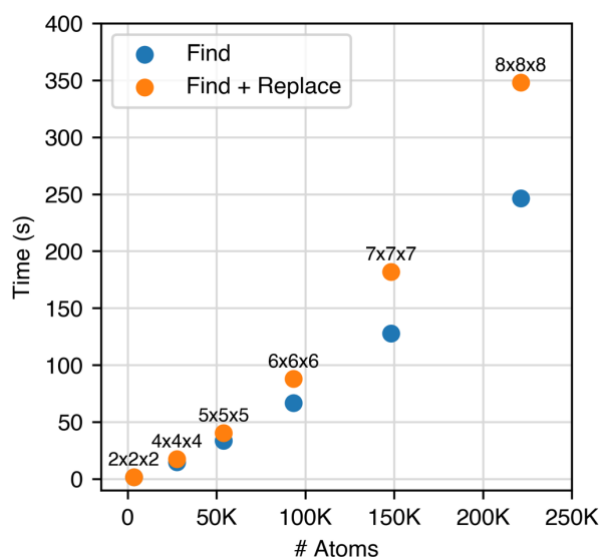


Figure 15: Time required to find or find + replace all linkers in a UiO-66 unit cell replicated to 2x2x2, 4x4x4, 5x5x5, 6x6x6, 7x7x7, and 8x8x8.

The performance of the optimized algorithm is $O(N^2)$, where N is the number of atoms in the system, and the memory usage is $O(M^2)$, where M is the maximum number of atoms within a distance d of any other atom. Practically, this means that system size and CPU speed will limit the kinds of systems that can be run. Our early naïve implementations took over 10 minutes to search for all linkers in a 2x2x2 replicated UiO-66 unit cell—3496 atoms, 192 linkers—and outright failed for systems bigger than that due to running out of memory. With the optimized algorithm, we can search for and replace all linkers in an 8x8x8 replicated UiO-66 unit cell—221,184 atoms and

12,288 linkers—in less than 6 minutes on a single core of an Apple MacBook Pro M1 Pro laptop (see Figure 15). This size system greatly exceeds our lab’s current needs of approximately a 4x4x4 system with 40K atoms for use in thermal conductivity calculations; however, we did run a larger find and replace on a 20x20x20 system—3.5M atoms and 192K linkers—and it completed in 15 hours. Additional optimizations could be implemented if there is a need for a find and replace operation at that scale.

4.4 MOFUN: Usage Details

Structures, search patterns, and replacement patterns can be defined directly in Python, or loaded in from a CML file (for typical output from Avogadro⁸⁶), a P1 CIF file, a LAMMPS³⁸ data file, or from any file format supported by the ASE⁸⁹ package, such as XYZ, PDB, RES, etc. Structures can be defined with either a cubic or triclinic unit cell. MOFUN supports reading and writing LAMMPS data files directly, including the LAMMPS pair, bond, angle, dihedral, and improper styles and all coefficients necessary for defining a flexible force field. When using a parameterized LAMMPS data file as a replacement pattern, MOFUN can insert the appropriate force field terms for all interactions into the resulting structure file. MOFUN also supports optionally replacing only a fraction of the search pattern matches found in a structure. A replacement fraction can be defined so that only a given % of matched search patterns will be replaced. All features can be used via the Python interface, and a command line tool is also included that handles many common use cases.

One of the more advanced features of MOFUN is that it can manage structure topology and force field parameters for flexible force fields and apply them correctly when inserting a

parameterized replacement pattern into a structure. This supports the insertion of two-body pair and bond potentials, three-body angle potentials, and four-body dihedral and improper potentials, as defined by LAMMPS. For each defined potential in the replacement pattern, the atoms that make up the potential and the potential type are inserted into structure alongside the replacement pattern atom positions and types. If the structure already has defined topology, then any topology associated with the search pattern is deleted along with its atoms prior to insertion of the replacement pattern, with one very important exception: if any of the atoms are shared between both the search pattern and the replace pattern (i.e. if the atoms share the same element and position), then any force field potentials defined on these atoms are overridden, rather than deleted and replaced. This is necessary to handle parameterizing a structure like the one shown below in Example 3, or when using find and replace to override the force-field terms in part of a structure while leaving the structure intact. If the replacement pattern defines the potential parameters (i.e. via a “* coeffs” section in the LAMMPS data file), then the potential parameters will also be carried forward into the resulting structure. While this only supports LAMMPS data files (and direct code in Python) at the moment, this is primarily because there is no standardized file format that is commonly-used to define periodic molecular structures along with full topological data and force-field parameters. For LAMMPS users such as ourselves, writing to a LAMMPS data file is extremely convenient as we can immediately simulate systems after find and replace; for users of other simulation packages, the LAMMPS data file should contain sufficient information to be converted into the file formats required by other simulation packages.

4.5 Examples

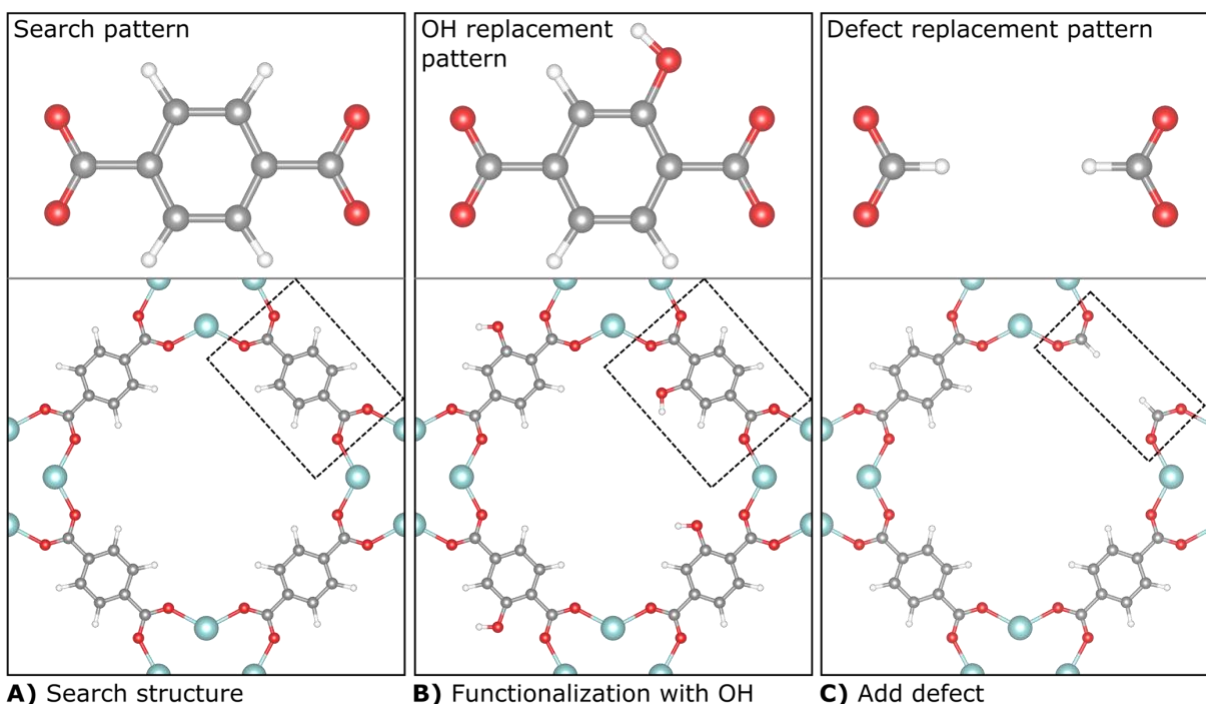


Figure 16: (A) search pattern and search structure (a “slice” of UiO-66), (B) example 1 resulting structure after find and replace using a replacement pattern with an added OH functional group, and (C) example 2 resulting structure after find and replace using a replacement pattern of a defective linker (two capping formate groups), as applied to 25% of the linkers.

We have chosen three examples to demonstrate MOFUN’s capabilities: (1) functionalizing the MOF UiO-66, (2) adding defects to UiO-66, and (3) fully parameterizing UiO-66 across periodic boundary conditions starting with an unparameterized UiO-66 structure and parameterized metal center and linker fragments. These examples are also available online with all supporting files and there is expanded guidance in the software’s documentation. Since MOFUN is living software, the syntax shown below may change and/or additional features may be added in the future. When in doubt, please refer to the online documentation. For each example, we describe the process we

used to prepare files and run the find / replace; we include this level of detail so the example properly illustrates what the task involves, but a user does not have to follow this process exactly.

4.5.1 Example 1: Functionalizing linkers in UiO-66

The first example is how to use find and replace to functionalize a structure. We will take the MOF UiO-66 (Figure 16A) and functionalize all linkers with hydroxyl groups (Figure 16B). We will need a structure file for UiO-66 and files for a standard UiO-66 linker and a linker functionalized with the hydroxyl. To create the UiO-66 linker file, we used Vesta⁹⁰ to pick one linker in the structure, deleted all other atoms, then exported to a file format that Avogadro⁸⁶ can read. We opened the file in Avogadro and saved as CML. The replacement pattern needs to lie in the same coordinate system as the search pattern. The easiest way to do this is to start with the search pattern and simply not move any of the atoms unless you intend to move them with the replacement operation. We took the search pattern CML, replaced one of the hydrogens on the linker with an oxygen atom, and added the attached hydrogen to make the hydroxyl. We used Avogadro's "Fix Selected Atoms" feature to prevent all the atoms from moving except for the newly added ones, then ran optimize structure to let the OH group find a more appropriate position. If you do not fix all the atoms except for the hydroxyls, many of the atoms will move when you optimize and the atoms of your replacement pattern will not correspond to the same atoms in the search pattern. Once all the files are prepared, you can run MOFUN either using the Python interface or the command line interface. For Python:

```
from mofun import Atoms, replace_pattern_in_structure

structure = Atoms.load("uio66.cif")
uio66_linker = Atoms.load("uio66-linker.cml")
uio66_linker_oh = Atoms.load("uio66-linker-oh.cml")

structure_oh = replace_pattern_in_structure(structure, uio66_linker, uio66_linker_oh)
structure_oh.save("uio66-oh.lmpdat")
```

From the command line:

```
mofun uio66.cif uio66-oh.cif --find uio66-linker.cml --replace uio66-linker-oh.cml
```

When we are functionalizing a MOF using find and replace, we are typically replacing a pattern that has fewer atoms with a pattern that has more, and the larger the functional group is in the replacement pattern, the more likely that functional groups from different linkers will overlap. This may not be a problem, for example, when adding hydroxyl groups to the linker in UiO-66, but if one were to add more bulky functional groups overlap would likely occur. The replacement operation inserts the functional group exactly as specified, and the resulting structure may need to be relaxed using molecular dynamics for the functional group to find a more reasonable configuration. When we are adding bulky functional groups to a structure, we create a replacement pattern where the functional groups are tightly placed near the linker, as much parallel to the linker as possible, to limit any overlapping with functional groups on other linkers. While this tight configuration may be high in energy, since we then relax the structure using a flexible force field, the functional groups can relax into a lower energy configuration.

4.5.2 Example 2: Adding defects to UiO-66

While we often assume a MOF is perfectly formed when we evaluate it computationally, it is well known that synthesized MOFs have a variety of defects, typically missing linkers or missing metal centers or both.^{91–93} Missing linker defect rates of 5-20% is normal, depending on

what MOF is being synthesized and the experimental synthesis method used. With MOFUN, we can search for a linker and replace it with a defect site, typically just a pair of capping groups—such as two formates—on the metal centers the linker was formerly connected to. Since MOFUN supports replacing a specified fraction of all instances of a pattern found in a structure, we can create structures with varying defect densities.

For this example, we will introduce defects into UiO-66 by randomly removing 25% of the linkers from the structure. We will first replicate the structure to a 2x2x2 so it fulfills minimum image conventions, which needs to be done before adding defects so that the defects aren't repeated in the structure. We can reuse the structure and search pattern files from example 1, but we will need to create a replacement pattern from the search pattern where the biphenyl ring is removed and replaced with formate caps where the linker would attach to the metal center (see Figure 16C). This replacement pattern can be created in a similar manner to that described in example 1.

To generate a structure with 25% defects, in Python:

```
from mofun import Atoms, replace_pattern_in_structure

structure = Atoms.load("uio66.cif").replicate((2,2,2))
uio66_linker = Atoms.load("uio66-linker.cml")
uio66_linker_defective = Atoms.load("uio66-linker-defective.cml")

defective = replace_pattern_in_structure(structure, uio66_linker, uio66_linker_defective, replace_fraction=0.25)
defective.to_ase().write("uio66-defective.cif")
```

From the command-line:

```
mofun uio66.cif uio66-defective.cif -f uio66-linker.cml -r uio66-linker-defective.cml --replicate 2 2 2 --replace-fraction=0.25
```

4.5.3 Example 3: Parameterizing UiO-66 with flexible force field terms

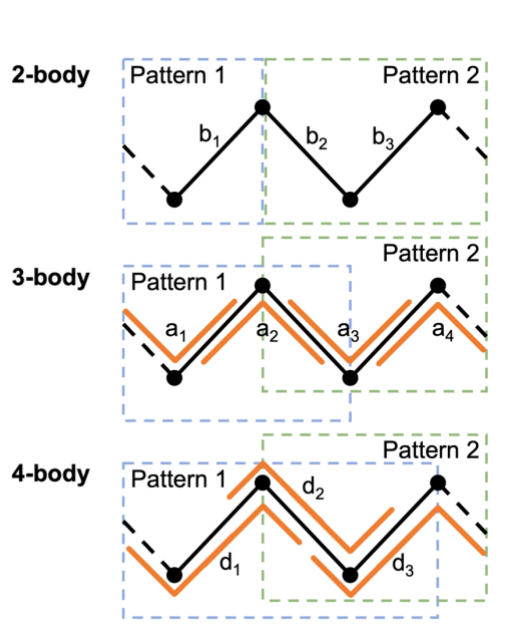


Figure 17: when parameterizing a structure using two parameterized patterns, for all two-body bond terms $b_1..b_n$ to be replicated, the patterns must share an atom, for all three-body angle terms $a_1..a_n$ to be replicated the patterns must share two atoms, and for all four-body dihedral terms $d_1..d_n$ to be replicated the patterns must share three atoms.

For some simulations of gas adsorption in MOFs, it is common to assume the positions of the structure's atoms are fixed and only the adsorbate gases move.⁹⁴ However, we do not always want to assume this (e.g. for thermal conductivity calculations,⁴¹ which require the atoms to move for heat to transfer), and many flexible force-fields have been developed that enable structures to flex and move.^{83,95–99} Despite these force-fields already existing, it can be challenging and time consuming to apply these force-fields to new structures: one needs to define atom types for every atom in the system, all topology required by the force field, and all force field parameters across the entire system, which may be tens of thousands of atoms and topology terms. This is a significant amount of work to do manually. There have been attempts to automate this,¹⁰⁰ but it is

hard to automate this process effectively and still allow for the parameter assignment process to be easily modified so that a researcher can validate and fix any parameter assignment issues when the automated system doesn't assign reasonable parameters. Even if one starts with a fully and correctly parameterized structure, expanding the structure to a larger number of unit cells can also be non-trivial, because bonds that cross periodic boundaries need to be "remapped" across the new periodic boundaries of the larger, expanded unit cell in order to run a LAMMPS simulation.

One tactic to overcome these challenges for structures that can be deconstructed into distinct parts is to assign force-field parameters to the constituent parts of the structure and then use find and replace to apply the force field to the entire structure. In this example, we apply this technique to MOFs: we assign parameters to a MOF's metal center and linker and then replace all unparameterized metal centers and linkers in the full structure with their corresponding parameterized versions.

The patterns for the parameterized linker and the parameterized metal center will need to overlap; every desired force-field term will need to be included fully in at least one of the patterns so some atoms and force-field terms will be defined in both files. For 2-body terms, only the atom that connects the metal center to the linker needs to be shared between the patterns. For 3-body (angle) or 4-body (dihedral / improper) terms there will need to be two or three atoms of overlap, respectively (see Figure 17). Linker and metal center files can be prepared similarly to Example 1 above and parameterized manually, or possibly in an automated manner using the rough UFF parameterizer included in MOFUN (information on the UFF parameterizer is beyond the scope of this paper, but can be found in the documentation online), or with other packages.¹⁰⁰ For Python:

```

from mofun import Atoms, replace_pattern_in_structure

structure = Atoms.load("uio66.cif")
uio66_linker = Atoms.load("uio66-linker-Zr.cml")
uio66_linker_params = Atoms.load("uio66-linker-Zr-parameterized.lmpdat")
uio66_mc = Atoms.load("uio66-metal-center.cml")
uio66_mc_params = Atoms.load("uio66-metal-center-parameterized.lmpdat")

param1 = replace_pattern_in_structure(structure, uio66_mc, uio66_mc_params)
param2 = replace_pattern_in_structure(param1, uio66_linker, uio66_linker_params)
param2.save("uio66-parameterized.lmpdat")

```

From the command-line:

```

mofun uio66.cif uio66-param1.lmpdat --find uio66-metal-center.cml \
--replace uio66-metal-center-parameterized.lmpdat

mofun uio66-param1.lmpdat uio66-parameterized.lmpdat --find uio66-linker-Zr.cml \
--replace uio66-linker-Zr-parameterized.lmpdat

```

4.6 Conclusion

MOFUN is an open-source Python package for generalized molecular find and replace. In our own lab, this is enabling us to quickly screen MOFs with various functional groups at different defect percentages, and easily apply force field parameters to structures. MOFUN is a great tool for automation, but there are some limitations. While MOFUN fully supports force-fields defined in LAMMPS, there is no inherent format support for other molecular packages, except for outputting CIF files containing topology and force-field parameters. When doing a replacement operation, MOFUN places the replacement atoms exactly as specified and does not check if this placement overlaps with other atoms in the system, so using MOFUN requires the researcher to setup the find and replace operation in a reasonable manner and potentially relax the system after the replace operation. At present, MOFUN is primarily optimized for smaller systems (< 40K

atoms), though still works (albeit, more slowly) for larger systems. By making this code available to other labs, we hope that this will enable other labs to perform more ambitious screening and simulation studies.

5.0 Designing optimal core-shell MOFs for direct air capture

Paul Boone^a, Yiwen He^b, Austin R. Lieber^c, Janice A. Steckel^d,
Nathaniel L. Rosi^{ab}, Katherine M. Hornbostel^{ac}, Christopher E. Wilmer^{ae}

^aDepartment of Chemical and Petroleum Engineering, University of Pittsburgh,
Pittsburgh, Pennsylvania 15261

^bDepartment of Chemistry, University of Pittsburgh,
Pittsburgh, Pennsylvania 15260

^cDepartment of Mechanical Engineering & Materials Science, University of Pittsburgh,
Pittsburgh, Pennsylvania 15261

^dUnited States Department of Energy, National Energy Technology Laboratory,
Pittsburgh, PA 15236

^eDepartment of Electrical and Computer Engineering, University of Pittsburgh,
Pittsburgh, Pennsylvania 15261

Metal-organic frameworks (MOFs), along with other novel adsorbents, are frequently proposed as candidate materials to selectively adsorb CO₂ for carbon capture processes. However, adsorbents designed to strongly bind CO₂ nearly always bind H₂O strongly (sometimes even more so). Given that water is present in significant quantities in the inlet streams of most carbon capture processes, a method that avoids H₂O competition for the CO₂ binding sites would be technologically valuable. In this paper, we consider a novel core-shell MOF design strategy, where a high-CO₂-capacity MOF “core” is protected from competitive H₂O-binding via a MOF “shell”

that has very slow water diffusion. We consider a high-frequency adsorption/desorption cycle that regenerates the adsorbents before water can pass through the shell and enter the core. To identify optimal core-shell MOF pairs, we use a combination of experimental measurements, computational modeling, and multiphysics modeling. Our library of MOFs is created from two starting MOFs—UiO-66 and UiO-67—augmented with 30 possible functional group variations, yielding 1,740 possible core-shell MOF pairs. After defining a performance score to rank these pairs, we identified 10 core-shell MOF candidates that significantly outperform any of the MOFs functioning alone.

5.1 Introduction

Negative emissions technologies such as direct air capture (DAC) are necessary to limit planetary warming.³ There are now several companies with DAC pilot plants, such as ClimeWorks, Carbon Engineering and Global Thermostat, whose processes are based on aqueous or solid sorbents that capture CO₂ and a vacuum or temperature swing to regenerate the sorbent that utilizes waste heat.^{9,10,12–16} However, scaling these pilot plants from the current total of 6,500 t CO₂ /year to the required scale of >12 Gt CO₂ / year is a non-trivial process that will strain global resource limitations on water, energy and land.⁵ To make the resource cost of DAC more manageable, there need to be novel breakthroughs in both material design and process design.

DAC technologies typically consist of either solvents or sorbents that remove CO₂ from the atmosphere, where it is present a very low concentration (~400 ppm). Most sorbents require lower regeneration temperatures but larger facilities to obtain the same capture capacity as their solvent counterparts.^{101–104} It is possible to reduce a sorbent-based DAC facility's size and capture

cost by improving the sorbent through chemical functionalization or coupling the sorbent with an additional material, either via surface coatings or impregnation. In prior work, a composite material of the metal-organic framework (MOF) NbOFFIVE-1-Ni@PA affixed to the surface of polyacrylate (PA) led to a CO₂ loading capacity improvement of 10.8% relative to the lone MOF.¹⁰⁵ Additionally, coupling sorbents, which typically have poor thermal conductivities, with unorthodox processes has been shown to lead to lower regeneration duty requirements. For example, microwave-assisted desorption of CO₂ saturated Lewatit VP OC 1065 (benzylamine-functionalized, porous polystyrene particles) showed marked improvement in productivity compared to temperature and/or pressure swing desorption due to the use of radiative heating.¹⁰⁶

Here we consider novel MOF designs to achieve higher performance in a DAC process. MOFs are a promising and very tunable class of materials; inorganic metal centers and organic ligands can be combined in different ways to create porous materials of varying geometries and surface chemistries.⁸⁵ Over 90,000 MOFs have been synthesized to date³⁵ and have demonstrated uses for gas storage, gas separation, catalysis and more.¹⁰⁷ However, it can still be difficult to design a single MOF that fulfills all of the requirements of a challenging process. One particular challenge of using MOFs in a DAC process is that water is present in the atmosphere at higher concentrations than CO₂, and typically adsorption sites that bind strongly to CO₂ bind even more strongly to H₂O, leading to unfavorable competitive adsorption. The presence of water may also negatively affect the stability of the MOF.^{108–110} A MOF with otherwise very high CO₂ / N₂ selectivity may not be viable under humid conditions for a DAC process.

One means of addressing this problem is by constructing a stratified MOF,¹¹¹ the simplest being a “core-shell” MOF consisting of a core MOF surrounded by a shell of another MOF.¹¹² The resulting composite material can exhibit unique properties that neither individual MOF possesses.

The first core-shell MOFs were synthesized in 2009^{113,114} and core-shell MOFs have shown promise for a number of applications¹¹² and specifically for CO₂ separation and capture.^{115–117}

Furthermore, a vast quantity of different stratified MOFs is possible from even a small basis set of individual chemical components. The properties of such MOFs would derive from the compositions of the individual strata *and* the sequence of those strata in the hierarchical structure. It would be time-consuming and impractical to synthesize every possible combination of materials to identify ideal strata compositions and sequences for a specified process and set of properties. We can greatly accelerate this discovery process by computationally screening a wide set of materials to identify promising MOFs and combinations to pursue in the lab. To the best of our knowledge, there have not been any attempts to develop a core-shell MOF process for CO₂ capture, or to attempt to identify promising core-shell MOF candidates computationally.

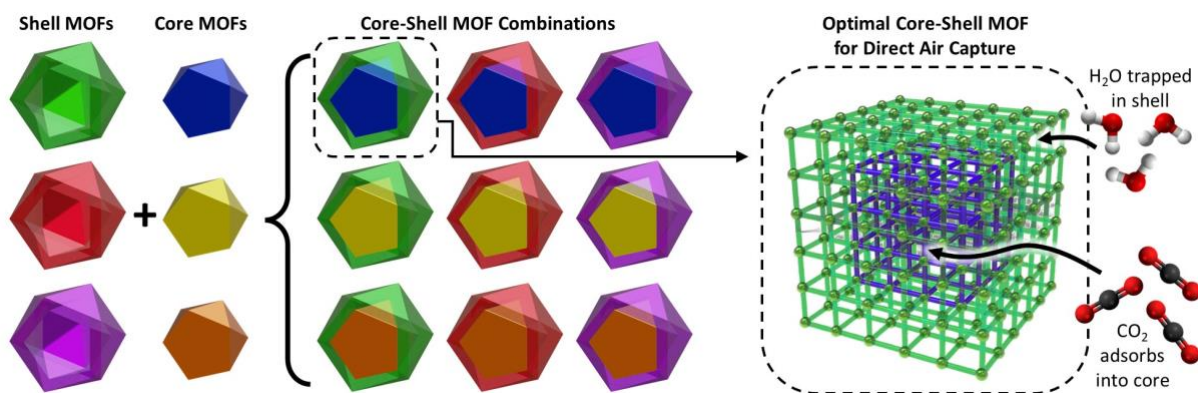


Figure 18: Overview of the strategy for designing optimal core-shell MOFs for DAC. A library of MOFs are combinatorically assembled into all of the possible core-shell MOF pairs, and then each pair is computationally evaluated to find candidates for experimental synthesis. Optimal designs should prevent H₂O from reaching the core while allowing for significant CO₂ adsorption in the core. Note that we include in our consideration core-shell MOF “pairs” where the core and the shell are the same MOF.

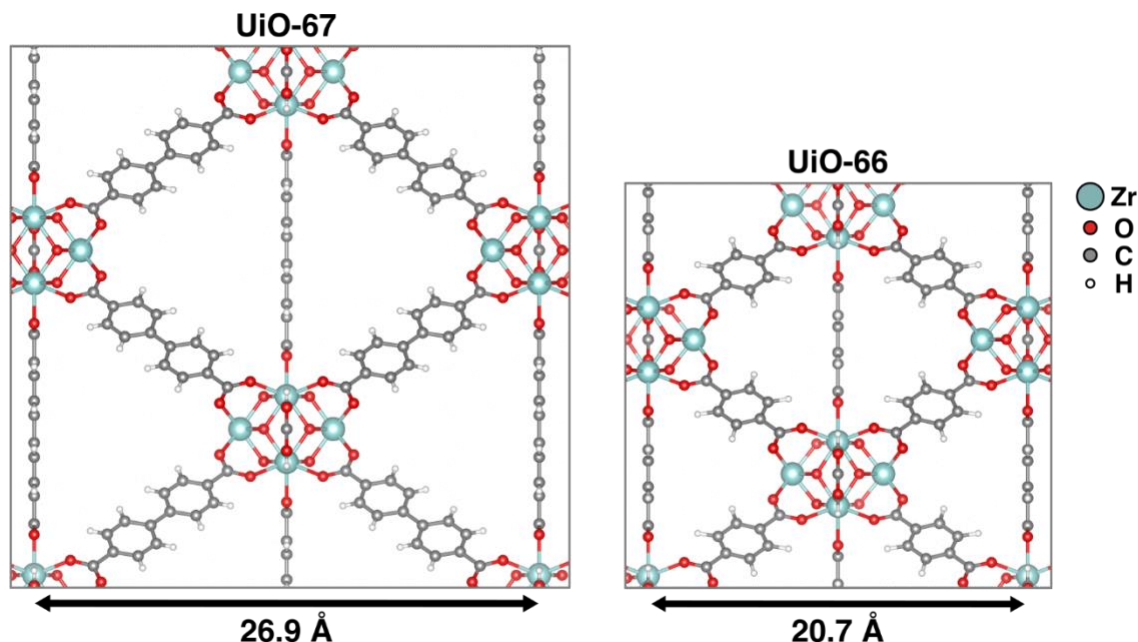


Figure 19: UiO-67 (left) and UiO-66 (right) structures showing biphenyl and phenyl ligands.

The purpose of this work is to identify core-shell MOFs that outperform their constituent MOFs in a DAC process. We define a ranking methodology to score all potential core-shell MOF pairs and identify 10 core-shell MOF pairs that have a performance at least 25% greater than their core or shell individually.

Certain selected MOFs were synthesized experimentally and their single component N_2 , CO_2 , and H_2O isotherms were collected. Sorption selectivities were calculated and compared to predictions to validate the computational approach. One core-shell MOF combination, amino₁◊methyl₂, was then simulated in COMSOL Multiphysics® to demonstrate the core-shell concept at the pellet scale.

5.2 Methodology

5.2.1 Overview

We chose UiO-66 and UiO-67¹¹⁸ (Figure 19) as our base MOFs (original MOFs that will be modified with different functional groups) because they are good candidates for CO₂ capture due to their high adsorption selectivity of CO₂ over N₂.^{109,119} To create our MOF library, we substituted one or more hydrogens on the linkers with 16 functional groups (see Figure 20). For many of the functional groups, we allowed for substitution either once or twice per linker, resulting in 30 different forms of each base MOF. Because the base MOFs UiO-66 and UiO-67 have different unit cell sizes, a core-shell MOF can't be a mix of both MOFs; the core-shell MOF must be composed of one kind or the other. However, 30 functional variations per base MOF makes possible about 30² core-shell MOF combinations per base MOF.


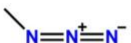
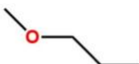
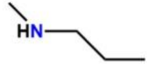



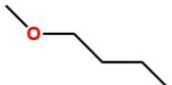
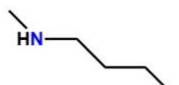
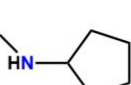


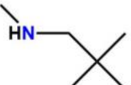
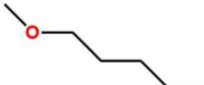
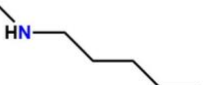
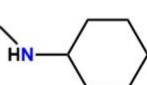
fluoro 		azido 	propoxy 	propanamino 	pyrrolidinyl 
hydroxy 		trifluoromethyl 	butoxy 	butylamino 	cyclopentylamino 
amino 	methyl 	t-butylamino 	pentoxy 	pentylamino 	cyclohexylamino 

Figure 20: Chemical diagrams of functional groups used to modify UiO-66 and UiO-67 linkers.

5.2.2 Idealized Adsorption/Desorption Cycle

In this study, we assume an idealized adsorption/desorption cycle (as shown in Figure 21), where the adsorption step is carried out over a time scale such that CO₂ saturates the core MOF but before water is able to diffuse through the shell MOF. In our model process, we assume that spherical pellets of a core-shell MOF are arranged in a shallow bed reactor, such that every pellet is exposed simultaneously to the input gas stream at the onset of the adsorption step. The input gas stream is assumed to be atmospheric temperature, pressure, and humidity, all of which depend on the time of day, the season of year, the weather, and other factors. For the purposes of this model, we are assuming the input gas stream is 298K, with partial pressures of 42 Pa CO₂, 79 kPa N₂ and 50% relative humidity.

With a shell that allows for faster CO₂ diffusion as compared to water, the CO₂ will reach the core before the water. The results of this process are dependent on the exact timing of the switch from adsorption to regeneration: too early and very little CO₂ reaches the core, too late and both the core and shell reach equilibrium loading (i.e., where the core would be saturated with H₂O). In both cases, there would be no benefit to using a core-shell MOF design. Therefore, a core-shell MOF process requires thoughtful design and timing to be effective.

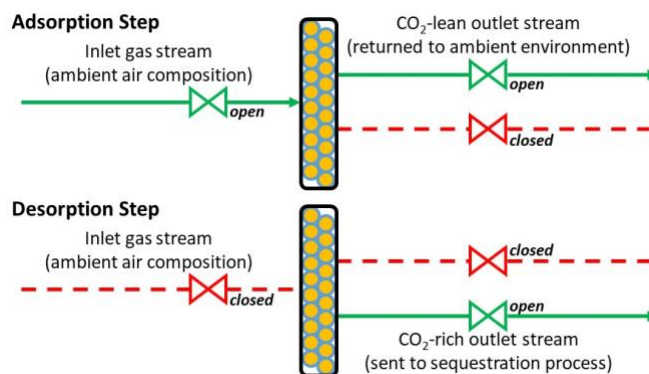


Figure 21: Schematic overview of our idealized adsorption/desorption cycle process used to investigate core-shell MOF candidates. We assume 100% evacuation of adsorbed gases during the desorption step, which can be mediated via imposing a vacuum or raising the temperature (or both), but in this idealized model the specific desorption conditions are intentionally ignored.

In this work we are sizing the pellets so that the water in the input gas stream breaks through into the core at 100 seconds, at which time the pellets are regenerated. For simplicity, we assume 100% evacuation of all gases during the regeneration step, and so in this idealized model we do not specify whether desorption is due to imposing a vacuum, raising of the temperature of the reactor bed, or both.

For this process to be selective for CO₂, the shell MOF of the pellet must be diffusion-selective for CO₂ over H₂O, and the core MOF must be adsorption-selective for CO₂ over N₂. This process is designed to allow different core-shell MOF combinations to be directly compared, and as a proof of concept demonstrating the viability of using a core-shell MOF for direct air capture.

5.2.3 Experimental

To compare with the simulation results, five MOFs, UiO-67, amino₁-UiO-67, amino₂-UiO-67, methyl₁-UiO-67, and methyl₂-UiO-67 were synthesized and characterized. The structures, compositions, and porosities of these MOFs were determined. CO₂, N₂ and water vapor sorption

isotherms at 298 K were collected, and these data were then used to calculate experimental adsorption selectivity (see Appendix B.2).

5.2.4 Determination of Water Breakthrough Times and Pellet Loadings

First, we calculate the breakthrough time of water according to the system shown in Figure 22. Let $x = 0$ be the boundary between the core-shell MOF pellet and the gas stream, let $x = x_0$ be the boundary between the core MOF and the shell MOF, and let $x = x_1$ be an arbitrary limit to the core MOF. The concentration profile of a gas in this system can be calculated using the diffusion equation, a Dirichlet boundary condition at $x = 0$, and a Neumann boundary condition at $x = x_1$:

$$\frac{\partial c_{gas}}{\partial t} = D_{gas} \frac{\partial^2 c_{gas}}{\partial x^2} \quad (5 - 1)$$

$$\text{At } x = 0, c_{gas} = A_{gas,eq} \quad (5 - 2)$$

$$\text{At } x = x_1, \frac{\partial c_{gas}}{\partial x} = 0 \quad (5 - 3)$$

Here, c_{gas} is the concentration of the gas, D_{gas} is the diffusivity of the gas, and $A_{gas,eq}$ is the equilibrium adsorption of the gas in the shell MOF. This partial differential equation can be solved analytically and is used to determine breakthrough times for both CO₂ and H₂O at $x = x_0$.

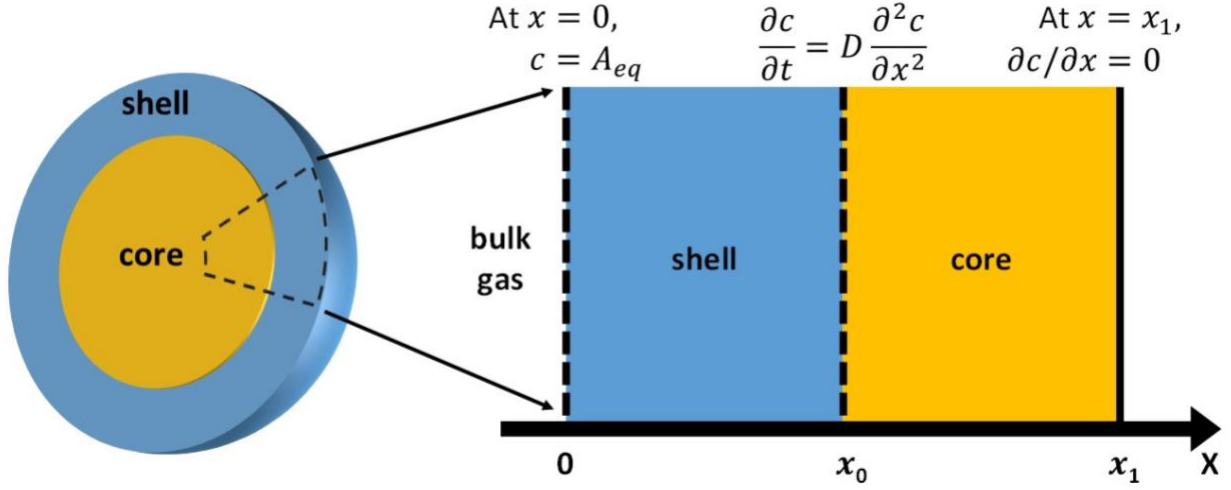


Figure 22: A 1-D infinite slab model of the outer region of a core-shell MOF used to estimate breakthrough times and fluxes of both H₂O and CO₂.

There are two major notes to consider here. The first is that the properties of the core are *not* being considered at this point, and the diffusivity of the shell is applied across the entire system $0 \leq x \leq x_1$. The rationale for including a core region, even though we are only interested in calculating shell properties, is to allow the concentrations and fluxes at the core-shell boundary to vary (i.e., not to be fixed) while maintaining simple boundary conditions elsewhere. For calculating breakthrough times through the shell, the diffusivity of the core should not significantly affect this calculation, and this simplification is necessary in order to evaluate a shell independently from a core. The second note to consider is that we are explicitly using the infinite slab version of the diffusion equation, not the spherical form. This is because we will be sizing the particles based on the results of the calculated breakthrough times so the radius of the core and the thickness of the shell are not known in advance. A more detailed 2D multi-physics model with spherical geometry is described below in Section 4.3.6.

We define the breakthrough time of a gas into the core as the smallest time ($t = \tau_{gas}$) such that $c_{gas}(x = x_0, \tau_{gas}) \geq 0.01 \cdot c_{CO_2,eq}$. In other words, the breakthrough time of a gas (τ_{gas})

occurs when the concentration of the gas (c_{gas}) at the core-shell MOF boundary (x_0) is greater than 1% of the equilibrium loading of CO₂ ($c_{CO_2,eq}$).

For each core-shell MOF, the thickness of the shell (x_0) is chosen so that the breakthrough time of H₂O (τ_{H_2O}) equals 100 seconds. The breakthrough time of CO₂ (τ_{CO_2}) is calculated using this x_0 . We assume a flux of CO₂ through the shell to the core based on the solution diffusion model.

$$j = \frac{\text{permeability}}{\text{shell thickness}} = \frac{A_{CO_2,eq} D_{CO_2}}{x_0} \quad (5 - 4)$$

The core is sized so that at 100s the core will be fully loaded with CO₂, given the flux j and assuming the core-shell MOF is a sphere. At 100s, the total loading of H₂O in moles (M_{H_2O}) is calculated as the surface area of the core with radius (r_{core}) multiplied by the integral of the concentration profile:

$$M_{H_2O} = 4\pi r_{core}^2 \int_0^{x_0} c_{H_2O}(t = 100s) dx \quad (5 - 5)$$

We assume N₂ reaches equilibrium loading in both the shell and core, $M_{N_2} = A_{N_2,eq,shell} \cdot V_{shell} + A_{N_2,eq,core} \cdot V_{core}$, and CO₂ reaches equilibrium loading only in the core: $M_{CO_2} = A_{CO_2,eq,core} \cdot V_{core}$. In this calculation, CO₂ loading of the shell is intentionally neglected because we assume H₂O will out-compete CO₂ for binding sites (note that this is a conservative assumption, as any CO₂ captured in the shell would improve process performance). At 100 seconds the core-shell MOF is regenerated and complete evacuation of all N₂, CO₂ and H₂O in the core-shell MOF is assumed.

5.2.5 Scoring of Core-Shell MOF Pairs

Core-shell MOFs are scored as the output stream CO₂ concentration of the core-shell MOF divided by the output stream CO₂ concentration of standalone UiO-67:

$$score = \frac{\left(\frac{M_{CO_2}}{M_{CO_2} + M_{H_2O} + M_{N_2}} \right)_{C \subset S}}{\left(\frac{M_{CO_2}}{M_{CO_2} + M_{H_2O} + M_{N_2}} \right)_{UiO67}} \quad (5 - 6)$$

C \subset S denotes the core \subset shell MOF (e.g. trifluoromethyl₂amino₁). This gives us a dimensionless number where values are a multiple (or fraction) of the CO₂ concentration in the output stream of a non-functionalized non-core-shell MOF UiO-67 under the same process. The main purpose of the scores is not to predict the absolute CO₂ concentration of the output gas stream but to be able to fairly compare different core-shell MOF pairs and rank them compared to each other and the individual core and shell that they are composed of.

For the core to saturate with CO₂ by the breakthrough time of water, the CO₂ diffusivity of the core must be similar to the CO₂ diffusivity of the shell. To ensure that we are only pairing shells with cores that have comparable CO₂ diffusivity, we set the score to 0 for any core that has CO₂ diffusivity < 1/10 that of the shell.

5.2.6 COMSOL Multiphysics® Modeling

A multiphysics model of a spherical core-shell pellet was developed in COMSOL Multiphysics® to simulate the diffusion and adsorption of CO₂ and H₂O in a macro-scale core-shell MOF. This model is 2D-axisymmetric along the centerline of the pellet, as shown in Figure 23. The core size and shell thickness for a given core-shell MOF was chosen to match the same

properties in our scoring model. For the example amino₁⊂methyl₂, this is a 0.453 cm radius core and a 0.04 cm thickness shell.

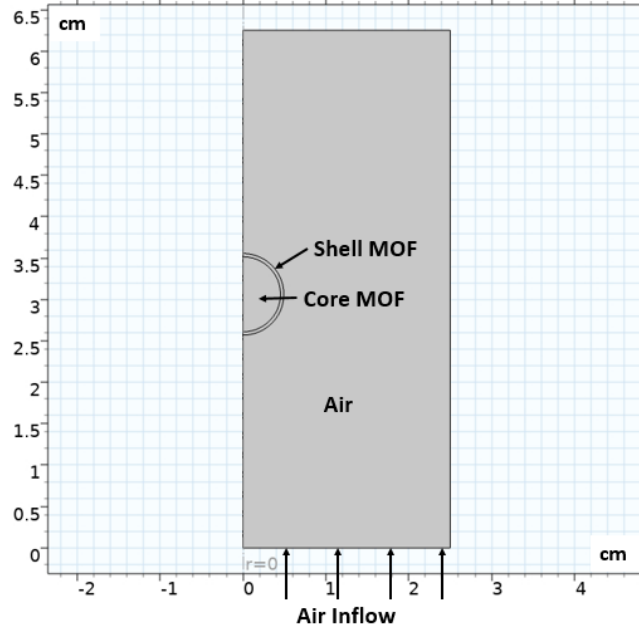


Figure 23: Setup of COMSOL Multiphysics® model of a single core-shell spherical pellet. Air flows in from the bottom over a 0.453 cm radius core + 0.04 cm thick shell pellet.

Adsorption of CO₂, N₂ and H₂O was modeled in COMSOL by curve-fitting the following Langmuir equation to experimental isotherm data:

$$C_{P,i} = \frac{C_{P,max,i} \cdot K_{L,i} \cdot C_i}{1 + K_{L,i} \cdot C_i} \quad (5-7)$$

where $C_{P,i}$ [mol/kg] is the concentration of gas adsorbed, $C_{P,max,i}$ [mol/kg] is the maximum amount of gas the MOF can hold, $K_{L,i}$ [m³/mol] is the Langmuir constant, and C_i [kg/m³] is the concentration of available gas to adsorb. The Langmuir curve fits along with the fitted values for the constants in this equation are provided in Appendix B.3.

5.2.7 Adsorption and diffusion simulations

For every functionalized MOF, we ran molecular dynamics simulations in the NVT ensemble and calculated self-diffusion coefficients for CO₂, N₂, and H₂O. We also performed grand canonical Monte Carlo (GCMC) simulations to calculate adsorption of CO₂ and N₂ and used the Widom insertion method¹²⁰ to determine Henry's constants for H₂O. CO₂ and N₂ were modeled using the TraPPE⁸² force field parameters and H₂O was modeled using TIP4P^{121,122}. Framework charges were calculated using EQeq¹²³ and the framework atoms were modeled with Lennard-Jones parameters from UFF.⁸³ Experimental N₂, CO₂, and H₂O isotherms were collected and the Henry's constant selectivities were calculated for comparison to the computational results. We employed custom force-field parameters for the NH₂-CO₂ interaction to better reflect chemisorption. Full details can be found in Appendix B.1.

5.3 Results and Discussion

Molecular simulations of gas adsorption and diffusion were carried out on all MOFs, followed by calculations using a 1-D infinite slab model to determine water breakthrough times in every shell MOF candidate. These various data were then used to score every core-shell MOF combination in order to rank them from best to worst. In addition to validating the simple 1-D slab model using finite element modeling with COMSOL Multiphysics®, we also synthesized certain MOF combinations and measured the adsorption of CO₂ and N₂.

Calculated gas loadings varied from about 1e⁻² to 1e¹ V/V (cm³ gas STP / cm³ framework) with most functional groups having N₂ loading > H₂O loading > CO₂ loading (see selected MOFs

in Figure 24a). This ordering follows the relative partial pressures of each species in the ambient environment. The fluorinated groups fluoro₈ and trifluoromethyl₂ are notable exceptions, showing very high H₂O loading. Diffusivities varied more widely, from about 1e⁻⁷ to 3e⁻² Å² / fs, with most MOFs having a N₂ diffusivity > CO₂ diffusivity > H₂O diffusivity (see selected MOFs in Figure 24b). There are some MOFs that do not show the same diffusivity ordering, but those have very low diffusivity and very high uncertainty, such as cyclohexylamino₂.

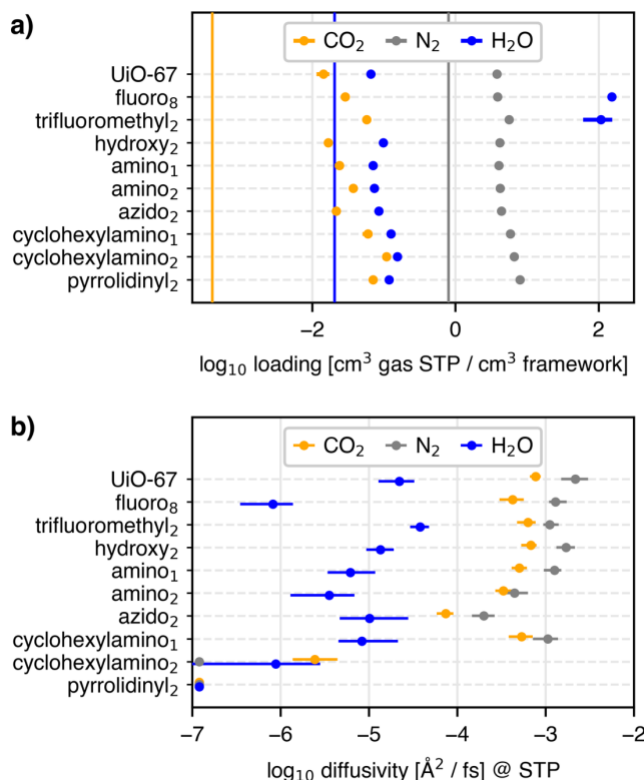


Figure 24: (a) Gas loadings of CO₂, N₂ and H₂O for selected MOFs based on UiO-67. Vertical lines are the amount of each gas in the atmosphere. (b) Diffusivities of CO₂, N₂ and H₂O for selected MOFs based on UiO-67, with error bars to 95% confidence interval.

The simulated gas loadings can be validated by comparing the adsorption selectivity of CO₂/ N₂ calculated using both the predicted gas loadings and the experimentally measured gas loadings (see Appendix B.2.4.5). The predicted selectivities exhibit a similar trend within UiO-67, methyl-UiO-67, methyl₂-UiO-67 and UiO-67, amino-UiO-67, amino₂-UiO-67, respectively (Appendix

Figure 12-Appendix Figure 13). However, when comparing amino- and methyl- functionalized MOFs, the simulation and experimental results do not follow a similar trend (Appendix Figure 14-Appendix Figure 15), prior to our adjustment of the $\text{NH}_2\text{-CO}_2$ interaction force field terms, which is due to chemisorptive effects not being modeled in the non-adjusted simulation model. Overall, general agreement between experimentally and computationally derived selectivities provides confidence that our models can be reasonably used to rank candidate MOF materials.

For UiO-66, there was no observable diffusion in 21 of 28 functionalized structures at the timescales simulated; this is likely because the pore size of UiO-66 is too small to reasonably pack larger functional groups into the empty space, leaving no room for a gas to diffuse through a rigid framework. Of the remaining functionalized structures, only one has a positive diffusive selectivity for CO_2 over H_2O : $\text{fluoro}_4\text{-UiO-66}$. However, this fluorine group has a very high adsorption of water, which will cause the perm-selectivity of $\text{CO}_2/\text{H}_2\text{O}$ to be less than one, making it selective for water over CO_2 . Therefore, none of the screened UiO-66-based MOFs are suitable as candidates for the shell. Since the layers within stratified MOFs should have similar unit cell parameters, we therefore will only be considering and scoring UiO-67 functional groups as potential core-shell MOFs.

For UiO-67, some of the denser functional groups, such as the hydrocarbons with two groups per linker reported no diffusion, likely due to similar causes as UiO-66. For all other groups, we have diffusivity data, and largely all structures are diffusion selective for CO_2 over H_2O . There are many different functionalized UiO-67 structures to choose from for a core-shell MOF. Diffusivities and gas loadings for all functionalized MOFs can be seen in in Appendix Figure 4-Appendix Figure 5.

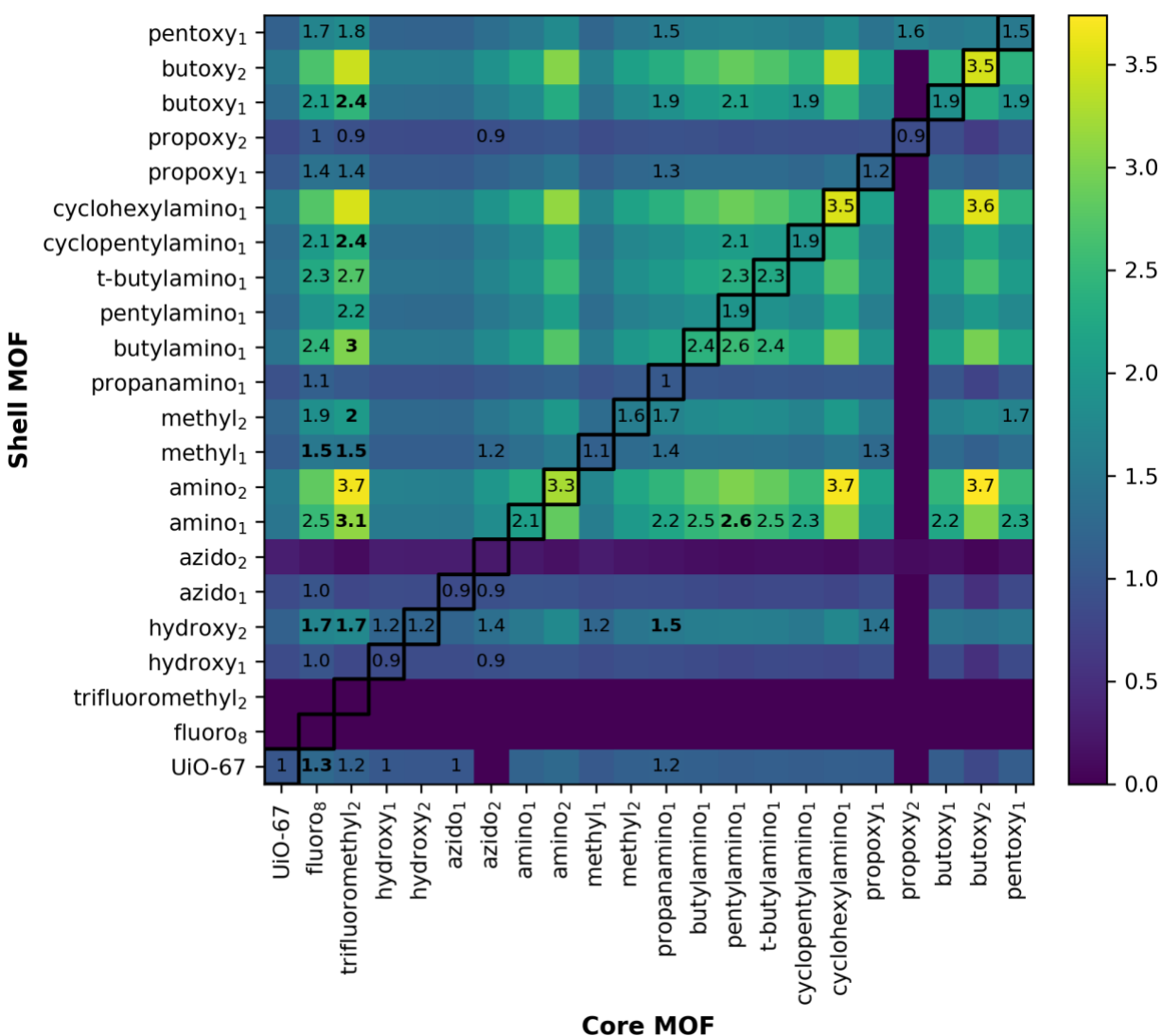


Figure 25: Scores for all UiO-67-based core-shell MOF combinations (excluding any MOF where all core-shell MOFs derived from it had scores less than 1.0). Black boxes are a guide to highlight the scores for non-core-shell MOFs under the same process. Numbers indicate every core-shell MOF combination where the combination has a higher score than both MOFs that compose it. Bold numbers show combinations with at least a 25% higher score than both MOFs that compose it.

When evaluating UiO-67-based core-shell MOF scores, we are looking for two things: (1) a score that is higher than both its individual core or shell under the same process, and (2) a high absolute score. Scores for all core-shell MOF combinations are shown in Figure 25: Scores for all UiO-67-based core-shell MOF combinations (excluding any MOF where all core-shell MOFs derived from

it had scores less than 1.0). Black boxes are a guide to highlight the scores for non-core-shell MOFs under the same process. Numbers indicate every core-shell MOF combination where the combination has a higher score than both MOFs that compose it. Bold numbers show combinations with at least a 25% higher score than both MOFs that compose it, and there are examples of core-shell MOFs that outperform their constituent core and shell, and core-shell MOFs that underperform.

The top 10 core-shell MOFs that most outperform their constituent core and shell are shown in Table 6. The shell MOFs are varied, but the core MOFs are dominated by the two fluorinated groups, trifluoromethyl₂ and fluoros₈. Both fluorinated MOFs are entirely non-viable as a standalone MOF for either a diffusion-based or adsorption-based separation process. Their affinity for water makes them perm-selective for water over CO₂ and hence cannot be used as a membrane or shell, and the water loading also makes them adsorption-selective for water over CO₂ so they cannot be used by themselves in a standalone adsorption process. However, because they have a higher CO₂/N₂ adsorption selectivity than most of the other MOFs in our dataset, they can be paired with almost any other MOF to improve on that MOF's performance. The top three improved core-shell MOFs are trifluoromethyl₂⊂amino₁, trifluoromethyl₂⊂hydroxy₂, and fluoros₈⊂hydroxy₂, all of which show improvement greater than 40% over the score of the standalone shell under the same process. This is a prime example of how pairing two MOFs into a core-shell MOF can make it possible for one to mitigate the negative traits of the other, thereby unlocking its positive traits.

Table 6: Top ten core-shell MOFs by greatest improvement over scores of their core and shell individually.

#	Core MOF	Shell MOF	Score (core-only)	Score (shell-only)	Score (core-shell)	Improvement
1	trifluoromethyl ₂	amino ₁	< 0	2.10	3.11	48%
2	trifluoromethyl ₂	hydroxy ₂	< 0	1.21	1.75	45%
3	fluoro ₈	hydroxy ₂	< 0	1.21	1.69	40%
4	trifluoromethyl ₂	butoxy ₁	< 0	1.87	2.44	31%
5	fluoro ₈	methyl ₁	< 0	1.14	1.49	30%
6	trifluoromethyl ₂	methyl ₂	< 0	1.58	2.04	29%
7	trifluoromethyl ₂	methyl ₁	< 0	1.14	1.47	28%
8	fluoro ₈	UiO-67	< 0	1.00	1.28	28%
9	trifluoromethyl ₂	pentylamino	< 0	1.86	2.38	28%
10	propanamino	hydroxy ₂	1.05	1.21	1.52	26%

The three highest scoring core-shell MOF combinations (see Table 7) have an amine shell and three different cores: butoxy₂, cyclohexylamino₁, and trifluoromethyl₂. All three of these show some improvement over their individual core and shells, from 5-7%. The next seven highest-scoring pairs do not show improvement over their individual core and shells, and in some cases, such as cyclohexylamino₁⊂cyclohexylamino₁, butoxy₂⊂butoxy₂, and amino₂⊂amino₂, the core and shell are the same MOF. All three of these MOFs have high adsorption selectivity for CO₂ / N₂ and high diffusion selectivity for CO₂ / H₂O, making them good candidates for this process when not part of a core-shell MOF. If it is possible to find a MOF with both properties we want, then this will always be a simpler approach than synthesizing a core-shell MOF.

Table 7: Top core-shell MOFs by absolute score.

#	Core MOF	Shell MOF	Score (core-only)	Score (shell-only)	Score (core-shell)	Improvement
1	butoxy ₂	amino ₂	3.50	3.26	3.74	7%
2	cyclohexylamino ₁	amino ₂	3.53	3.26	3.69	5%
3	trifluoromethyl ₂	amino ₂	-	3.26	3.68	13%
4	butoxy ₂	cyclohexylamino ₁	3.50	3.53	3.57	1%
5	cyclohexylamino ₁	cyclohexylamino ₁	3.53	3.53	3.53	0%
6	trifluoromethyl ₂	cyclohexylamino ₁	-	3.53	3.52	0%
7	butoxy ₂	butoxy ₂	3.50	3.50	3.50	0%
8	cyclohexylamino ₁	butoxy ₂	3.53	3.50	3.46	-2%
9	trifluoromethyl ₂	butoxy ₂	-	3.50	3.44	-2%
10	amino ₂	amino ₂	3.26	3.26	3.26	0%

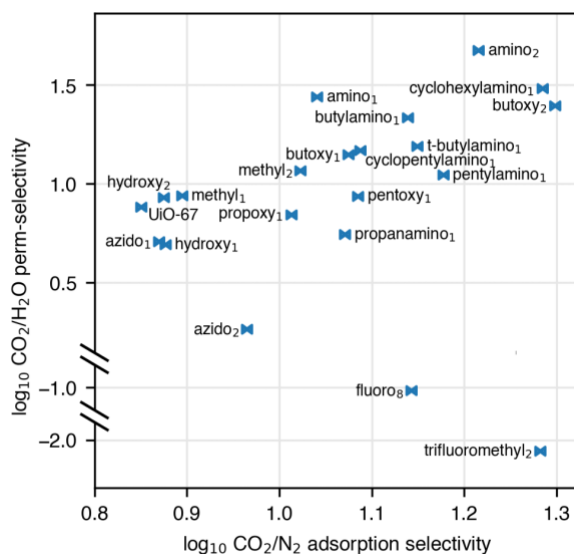


Figure 26: CO₂ / H₂O perm-selectivity of shell vs CO₂ / N₂ adsorption selectivity of core.

As we have defined the system above, the thickness of the shell decreases the better the shell is at separating out the CO₂. Concurrently, the size of the core decreases the better the core is at storing

CO₂. For excellent shells ($\tau_{CO_2} \ll \tau_{H_2O}$, high j_{CO_2}), adsorption in the core-shell MOF is determined primarily by the adsorption of the core, and the resulting CO₂ concentration depends on the adsorption selectivity of CO₂ / N₂. We can plot the perm-selectivity of CO₂ / H₂O vs the adsorption-selectivity of CO₂ / N₂ to rank or identify good candidate core shell MOFs (see Figure 26) without calculating full scores. Note that the perm-selectivity cannot be interpreted as a strict selectivity since this is not a membrane process (i.e. a selectivity of 1 does not divide shells that are selective vs shells that are not selective for this process) but it can be used to rank shells. Using Figure 26, we can arrive at the same conclusions (minus the quantitative metric) as the fully calculated scores. The three highest-performing MOFs—cyclohexylamino₁, butoxy₂ and amino₁—can be readily identified in the upper-right hand corner. Since their properties are superior to every other MOF evaluated, they will not form a significantly improved core-shell MOF with any of the other MOFs. CF₃ has comparable CO₂ / N₂ selectivity as the highest performers, and as a core will improve almost every other MOF, but especially the MOFs with high CO₂ / H₂O perm-selectivity and low CO₂ / N₂ adsorption selectivity (upper left corner). Besides the quantitative comparison, this plot is also missing comparative absolute diffusions, so it is possible to wrongly identify a possible core-shell MOF pair if the diffusions of the MOFs vary widely. However, it is a simple way of validating the calculated scores and understanding the factors that are driving the scores.

Although the 1-D infinite slab model is simple enough to solve analytically, it does not capture many important effects that would take place in a real carbon capture process. In addition to the loss of fidelity from considering a slab vs. a sphere, real fluid flows also experience friction, variations in pressure, turbulence, etc. We primarily expect these factors to significantly affect the timescales over which the gases adsorb/diffuse into/through the core-shell MOF pellets, as opposed to the equilibrium loading capacity, for example.

As a first step to investigate how a core-shell MOF would perform in pellet form under more realistic conditions, we simulated a core-shell MOF pellet in COMSOL Multiphysics®. We selected amino₁⊂methyl₂ as our core-shell MOF system because we had experimental gas sorption isotherms for both MOFs from validating our gas loading calculations. Subsequently, we modeled separate core and shell domains (as opposed to a homogenous core-shell MOF throughout the pellet). The core domain (amino₁) of the spherical pellet had a radius of 0.453 cm, and the surrounding shell domain— methyl₂— had a thickness of 0.04 cm. Figure 27 shows a snapshot of the CO₂ and H₂O concentrations throughout the pellet at t = 990 s. Note that we expect the timescales of adsorption/diffusion to vary from the simplified 1-D slab model, hence the longer breakthrough time than 100 s used elsewhere. As intended, the CO₂ enters the core before H₂O can reach it, which demonstrates that the shell is preventing H₂O from accessing the core over this short time period. Further multiphysics simulations could be performed for more core-shell MOF combinations, however, this case study serves as a proof-of-concept for the basic principle of the core-shell MOF design.

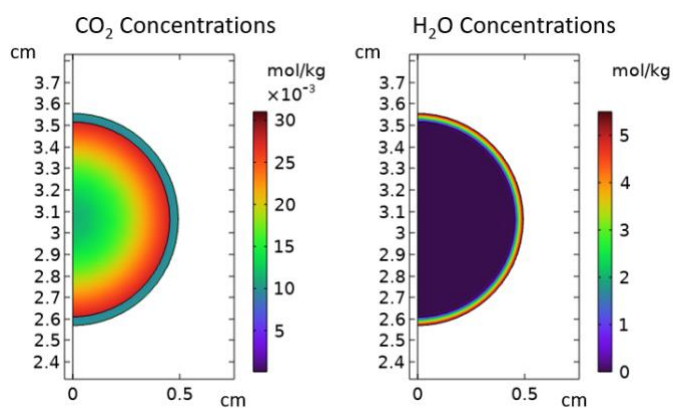


Figure 27: CO₂ concentrations (left) and H₂O concentrations (right) in a simulated amino₁⊂methyl₂ spherical pellet at t = 990 s.

Our methodology for scoring core-shell MOFs is intended to be simple and to efficiently rank core-shell MOF pairs so that top candidates can be scrutinized in more detail. We did not

incorporate multi-gas adsorption simulations or multi-gas diffusion calculations, so cases where CO_2 , H_2O , or N_2 interfere with the adsorption or diffusion of another gas is explicitly not modeled. Since the gas loading of H_2O is derived from its Henry's coefficient, if H_2O is not in the Henry's regime for a specific MOF then the H_2O loading predictions will be high. All simulations are performed on an ideal crystal, when synthesized MOFs typically have varying kinds of defects in their crystal structure which can affect their properties. Out of necessity our models neglect many of the complex details of real materials and processes, and synthesis and testing of core-shell MOFs is required to validate our proposed candidate materials. It is also important to emphasize our idealized adsorption/desorption process, where every MOF pellet is exposed to the input gas stream simultaneously. In future work, more realistic process simulations will be needed to predict the efficacy of these materials in more conventional reactors.

We have only looked at two different base MOFs—UiO-66 and UiO-67—with 30 functional groups, or only 60 total MOFs out of the more than 90,000 MOFs that have been synthesized. An exciting research area could be to search for better core-shell MOF pairs by broadening the search to new base MOFs or new functional groups. Because the best MOFs identified in this work—amino₂, cyclohexylamino₁, butoxy₂—perform well for both the core and shell, any new MOF that would pair nicely with them must either be a significantly superior core or shell. The fluorinated MOFs could be possibilities as core MOFs if their CO_2/N_2 adsorption selectivity can be improved. Regardless, we recommend doing two searches: one for high $\text{CO}_2/\text{H}_2\text{O}$ perm-selectivity materials, and the other for materials with high CO_2/N_2 adsorption selectivity in the absence of water.

5.4 Conclusion

Computational screening of material properties is vital to sift through the vast number of potential stratified MOF combinations, which is exponentially larger than the number of available MOFs themselves. This work represents the first major step in that direction by identifying MOFs that could be good shells or good cores as part of a core-shell MOF used to separate CO₂ from the atmosphere.

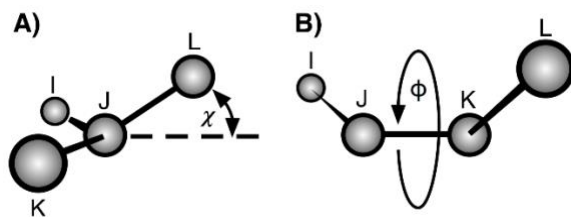
We have looked at the MOFs UiO-66 and UiO-67 augmented with 16 different functional groups (leading to 30 functional group variations) and experimentally tested gas sorption on five UiO-67 analogues to verify computational predictions. All functionalized UiO-66 MOFs were eliminated from further consideration as none of them were sufficiently selective for CO₂ or well-suited for acting as a shell. For UiO-67, we identified multiple possible combinations where a core-shell MOF was better than either of the component MOFs in isolation. Notably, when the fluorine-based functional groups—fluoro₈ and trifluoromethyl₂—were used as the core, they almost always resulted in an improved core-shell MOF. Hence, a result from our study with potentially broader applications is that a MOF that is selective for an undesirable gas in a standard adsorption or diffusive process may still be high-performing when used as a core in a core-shell MOF.

We also found that the MOFs amino₂, cyclohexylamino₁ and butoxy₂ had the best separation when *not* part of a core-shell MOF—this was due to them having good properties for being both a core and a shell. Finding a core-shell MOF where the core and shell serve two distinct needs therefore requires that (1) there must not already be a single MOF that has superior characteristics across both needs, and (2) there must be two distinct MOFs that individually fulfill one need but do not fulfill the other need.

Finally, A multiphysics case study of a core-shell amino₁⊂methyl₂ pellet was performed to demonstrate that the core-shell MOF design can be applied to the pellet-scale to effectively block water from the core while it loads with CO₂. This paper provides a framework for computationally screening MOF combinations for a given application and lays the foundation for a novel approach to hybrid solid sorbent materials optimization.

Appendix A LAMMPS corrected heat flux for improper potentials

In LAMMPS, improper potentials are four-body potentials and unlike the dihedral four-body potentials, the atoms making up an improper potential do not need to be connected via bonds. Any four atoms IJKL may be used, and the improper potential is defined using the angle between the planes IJK and JKL. As defined, this potential may be used for both traditional improper definitions (see Appendix Figure 1(A)) or in our case for replicating a dihedral potential using an improper potential (see Appendix Figure 1(B)).



Appendix Figure 1: (A) atoms IJKL for a typical improper, and (B) atoms IJKL for a dihedral defined using an improper.

There is no improper potential defined in LAMMPS that directly corresponds to the dihedral potential OPLS, but we were able to parameterize the CVFF potential to get a reasonable comparison so we could compare per-term energies. For OPLS, the potential is:

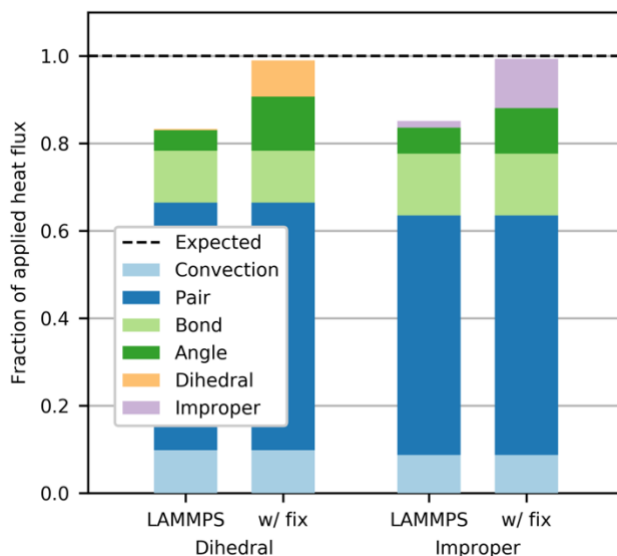
$$E = K_1[1 + \cos(\phi)] + K_2[1 - \cos(2\phi)] + K_3[1 + \cos(3\phi)] + K_4[1 - \cos(4\phi)]$$

The parameters for octane from the NERD FF⁶² are $K_1 = 1.411074$ Kcal / mol, $K_2 = -0.271015$ Kcal / mol, $K_3 = 3.145743$ Kcal / mol, and $K_4 = 0$.

For CVFF, the potential is:

$$E = K[1 + d \cos(n\phi)]$$

If we choose the parameters, $K = K_3$, $d = 1$, $n = 3$, the CVFF potential will be equivalent to the third term of the OPLS potential, which represents the highest mode and dominates the energy. This should give a good approximation to the OPLS potential used for the dihedral. The specifics here are mostly unnecessary, since as we mention in the full paper, the total heat flux should self-consistently match the applied heat flux, regardless of the parameters we choose. As can be seen from Appendix Figure 2, the improper results closely match the dihedral results, demonstrating that our implementation of the improper potential in LAMMPS is correct.



Appendix Figure 2: Per-potential fraction of applied heat flux for both uncorrected LAMMPS and corrected calculations for octane, as calculated using dihedral or improper potentials.

Appendix B Supplementary Info to CSMOF paper

Appendix B.1 Computational Predictions

Appendix B.1.1 Preparation of Functionalized MOF files

To create functionalized variations of each base MOF, we used the find and replace operation of MOFUN¹²⁴ to find all linkers in each base MOF and replace them with functionalized linkers. For both the UiO-66 and UiO-67 MOFs, we built functionalized forms of the linkers in Avogadro⁸⁶ for each of the 30 functional group variations. We assigned Universal Force Field (UFF)⁸³ atom types and parameters to each of the functionalized linkers using MOFUN's rule-based UFF typer and parameterizer. For the larger functional groups, we kept the functional group atoms as near as possible to the linker to avoid overlapping with other functional groups once the functionalized form was incorporated into the MOF. These functionalized and parameterized linkers were then substituted into UiO-66 or UiO-67. As the functional groups were placed in a relatively compressed configuration to avoid overlap, we ran a short NVT molecular dynamics simulation in LAMMPS³⁸ to relax the functional group into a reasonable configuration. The atoms of the MOF's metal center and parent ligand were kept fixed; only the functional group atoms were allowed to move based on their assigned UFF parameters. We ran 2,000 timesteps of an NVT molecular dynamics simulation for each of 1e-5, 1e-4, 1e-3, and 1e-2 fs and finished with 12,000 timesteps at 0.1 fs. For some of the denser structures, the UFF parameters occasionally resulted in the two hydrogens from an H-C-H unit in an alkane chain being unreasonably close to each other, and sometimes overlapping on the same point. For structures with this problem, we increased the

force constant of the H-C-H angle from ~75 kcal / mol to 200 kcal /mol and reran the NVT simulation. After the functional groups were relaxed, the final coordinates of all the atoms were stored for use in fixed framework diffusion and adsorption simulations.

Appendix B.1.2 Diffusion Calculations

Diffusion calculations were run in LAMMPS³⁸ with the NVT ensemble at 298K. Because the framework was modeled as fixed for computational efficiency, NVT was required to keep the gas molecules at standard temperature. For each UiO-67-based MOF and gas, 10 molecules of the gas were randomly inserted into the MOF using Packmol⁶³ for five independent simulations, giving 50 trajectories per gas total. The trajectories for the centers of each gas molecule were averaged together to calculate the mean squared displacement (MSD). The diffusivity was calculated by attempting fits to MSD vs time at various intervals from 0.1 to 0.5 of the total simulation time and selecting the fit with the highest R^2 . Per the procedure described in Maginn, et al,¹²⁵ the uncertainty was estimated by generating 500 random subsets (of 50 trajectories each) randomly selected with replacement from the original 50 trajectories and estimating the diffusivities of each subset. Upper and lower bounds on the diffusivities were calculated using a 95% confidence interval.

The same procedure was used for UiO-66-based MOFs and gases, but only 10 molecules total of the gas were simulated or 10 trajectories total. We eliminated UiO-66-based MOFs for consideration based on the results of the diffusivities calculated with the 10 initial trajectories and did not run the additional 40 trajectories. The two CH₃ functional groups were added to our simulation list after we disqualified UiO-66 so we only report CH₃ for UiO-67.

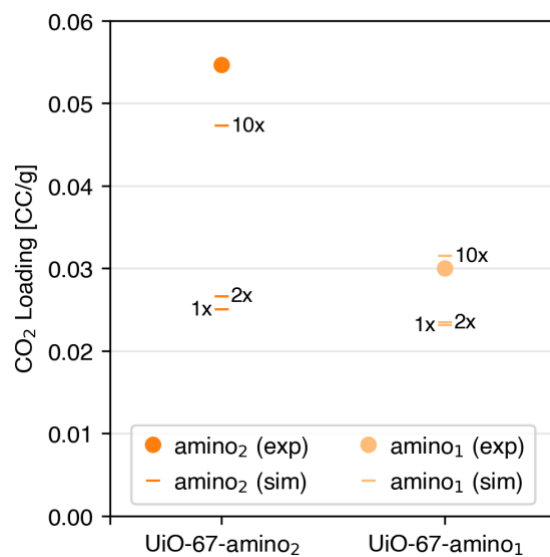
Appendix B.1.3 Gas Loading Calculations

Single component gas adsorption calculations were run in RASPA³¹ using GCMC³² to get absolute gas loadings in V/V (cm^3 gas [STP] / cm^3 framework) at typical atmospheric concentrations: 78% N_2 (79,033 Pa), 400 ppm CO_2 (42.18 Pa), and 50% relative humidity H_2O . It can be difficult and computationally expensive to calculate H_2O adsorption this way;^{126,127} so instead, we calculated Henry's constants for H_2O , and estimated loading at 50% humidity by multiplying the Henry's constant by the partial pressure of 50% of the saturation loading of H_2O of the TIP4P model, or 2050 Pa.

Appendix B.1.4 Custom CO_2 - NH_2 Interaction Force-Field Parameters

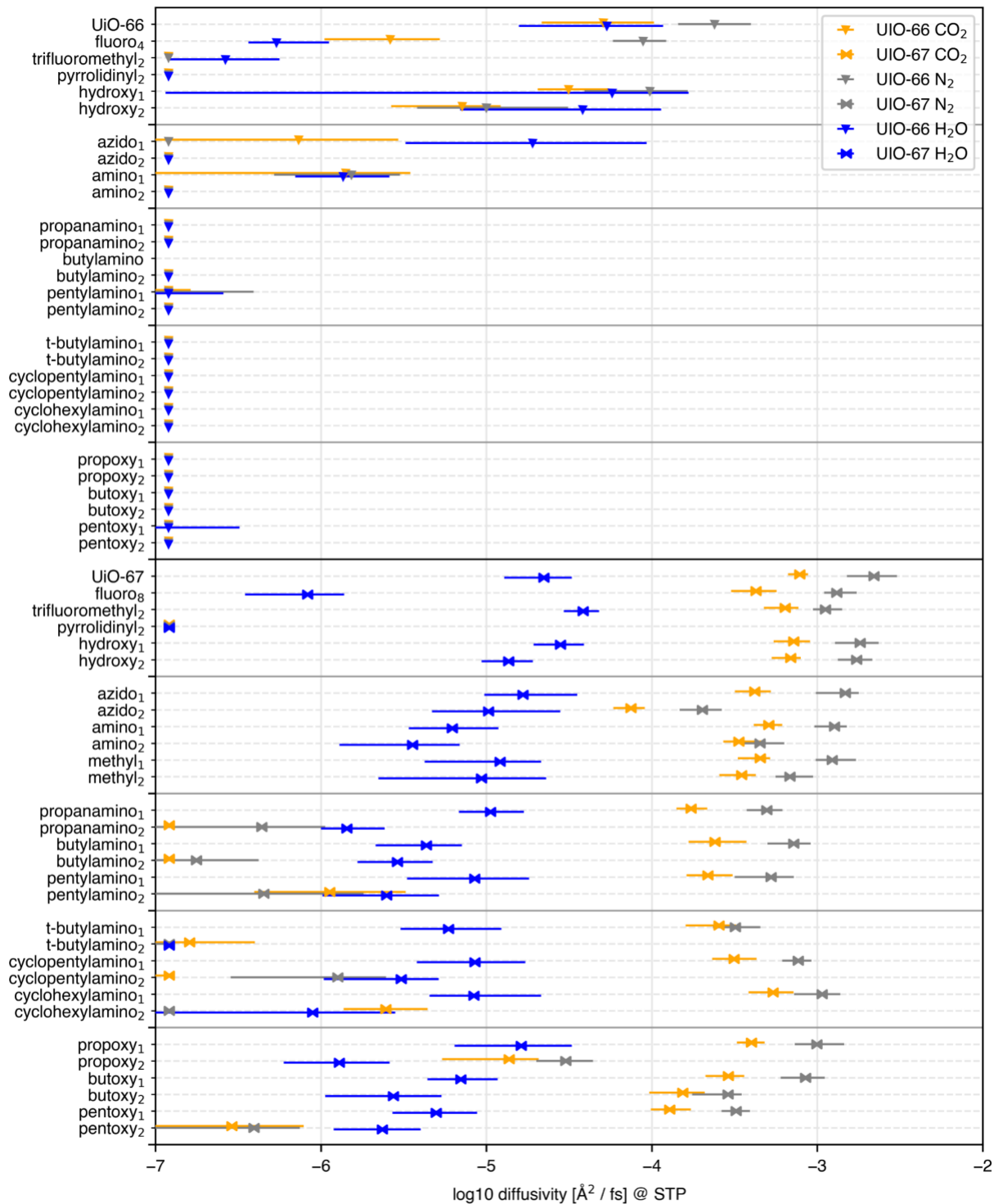
Our GCMC simulations of gas loading do not allow bonds to break and reform (i.e., for chemical reactions to take place). For most functional groups we have chosen to study, only physisorption of gases is expected. However, CO_2 is expected to chemically react with the amino groups (for further details on the likely mechanism, see reference ¹²⁸). Modeling the full reaction pathway is complex and beyond the scope of our investigation, where we are primarily interested in the *amount* of CO_2 that loads into the pores. Thus, to emulate such chemisorption behavior, we simply adjusted the strength of the CO_2 -framework interactions by amplifying the Lennard-Jones parameter for CO_2 -amino interactions from the UFF default values. We ran four sets of adsorption simulations for the MOFs UiO-67-amino₁ and UiO-67-amino₂ with different epsilon strengths: 1x, 2x, 10x and 100x of normal. These simulation results were compared to experimental values (see Appendix Figure 3) and the epsilon strength that is closest to the experimental values for amino₁ and amino₂ is 10x. The 100x simulations reported gas loadings above that shown in Appendix

Figure 3. The 10x epsilon parameters were used for both the diffusion and adsorption calculations for the amino₁ and amino₂ MOFs.



Appendix Figure 3: Simulated CO₂ loadings at 1x, 2x, and 10x of the normal CO₂-amino interaction strength epsilon compared to experimental CO₂ loadings for the two MOFs UiO-67-amino₁ and UiO-67-amino₂.

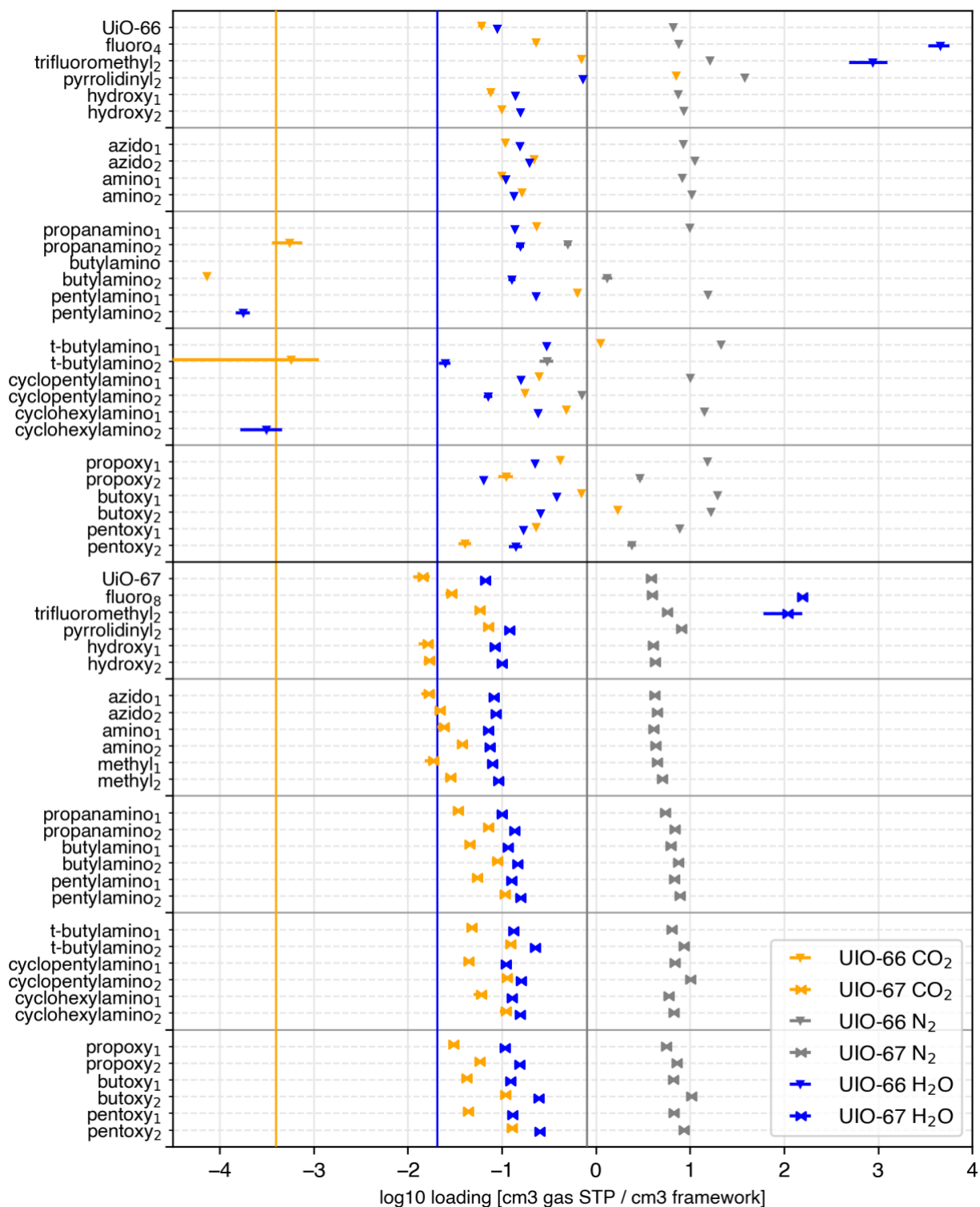
Appendix B.1.5 Simulated Diffusivities



Appendix Figure 4: Simulated diffusivities of CO₂, N₂, and H₂O for all UiO-66 and UiO-67 functional groups.

Error bars to 95% confidence shown for all points.

Appendix B.1.6 Simulated Gas Loadings



Appendix Figure 5: Simulated gas loading of CO₂, N₂, and H₂O for all UiO-66 and UiO-67 functional groups.

Atmospheric CO₂, N₂, and H₂O partial pressures shown by vertical yellow, grey and blue lines. Error bars to 95% confidence shown for all points (though error for most points is too small to be visible).

Appendix B.2 Experimental

Appendix B.2.1 General Methods

All purchased chemicals were used without further purification. Powder X-ray diffraction patterns were collected using a Bruker AXS D8 Discover powder diffractometer at 40 kV, 40 mA for Cu K α ($\lambda = 1.5406 \text{ \AA}$) with a scan speed of 0.20 sec/step from 5 to 30° at a step size of 0.02°. The data were analyzed using the EVA program from the Bruker Powder Analysis Software package. The simulated powder patterns were calculated using Mercury 3.8 based on MOF crystal structures.

Thermogravimetric analysis (TGA) was conducted on a TGA Q500 thermal analysis system under a constant N₂ UHP flow from room temperature to 800 °C at a rate of 1 °C/min.

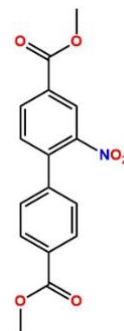
Gas adsorption isotherms were collected on a Micromeritics 3-flex gas adsorption analyzer. As-synthesized MOF crystals soaking in DMF were exchanged with 10 mL of dry methanol three times a day at 65 °C for one day. Then, the crystals were dried under a N₂ stream until they became a free-flowing powder. Approximately 40-60 mg of each sample were added into a pre-weighed sample analysis tube that had been evacuated and backfilled with He before massing. The samples were degassed at 298 K under vacuum for ~24 hours until the pressure change rate was no more than 3.5 mTorr/min. A liquid N₂ bath was used for the N₂ adsorption experiments at 77 K. A water/ethylene glycol bath was used for isotherms collected at 298 K. UHP grade N₂ and CO₂ gas adsorbates (99.999 %) were used in this study.

¹H NMR spectra were obtained using Bruker Avance III 400 MHz spectrometers. Chemical shifts are in parts per million (ppm) using the residual solvent peak (DMSO-d₆ or D₂O) as references. MOF samples were digested with DMSO-d₆ and a small amount of K₃PO₄ and D₂O.

Appendix B.2.2 Synthesis and characterization of MOF ligands

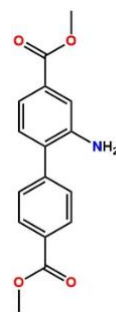
B.2.2.1 Dimethyl 2-nitro-1, 1'-biphenyl-4,4'-dicarboxylate (1)

Compound 1 was synthesized according to literature conditions.¹²⁹



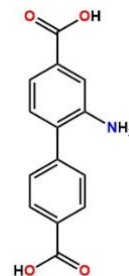
B.2.2.2 Dimethyl 2-amino-1, 1'-biphenyl-4,4'-dicarboxylate (2)

To a 100 mL Schlenk flask equipped with a stir bar were added compound **1** (710 mg, 2 mmol), 10 wt. % palladium on carbon (70 mg) and ethyl acetate (30 mL). The Schlenk flask was quickly evacuated on a vacuum line and then backfilled with argon gas. This evacuation and backfill process was repeated 3 times. The Schlenk flask was then evacuated and attached to a H₂ balloon. The reaction mixture was stirred at room temperature under H₂ atmosphere and monitored via thin layer chromatography (TLC). After 6 hours, the reaction was stopped by removing Pd catalyst via vacuum filtration through a celite cake. The filtrate was concentrated in vacuo to yield light yellow solid compound **2** (570 mg, 93%). Compound **2** was used without further purification. ¹H NMR (400 MHz, CDCl₃) δ 8.15 (m, 2H), 7.53 (m, 2H), 7.49 (dd, J = 7.9, 1.4 Hz, 1H), 7.47 (d, J = 1.4 Hz, 1H), 7.22 (d, J = 7.9 Hz, 1H), 3.97 (s, 3H), 3.94 (s, 3H), 3.88 (s, 2H).



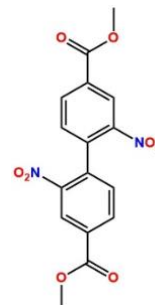
B.2.2.3 2-Amino-1, 1'-biphenyl-4,4'-dicarboxylic acid (**3**)

Compound **3** was synthesized using compound **2** as starting material based on literature conditions.¹²⁹ ¹H NMR (400 MHz, DMSO-d₆) δ 12.83 (s, 2H), 8.03 (d, J = 8.5 Hz, 2H), 7.63 (d, J = 8.5 Hz, 2H), 7.43 (d, J = 1.5 Hz, 1H), 7.22 (dd, J = 7.5, 1.5 Hz, 1H), 7.16 (d, J = 7.5 Hz, 1H), 5.25 (s, 2H).



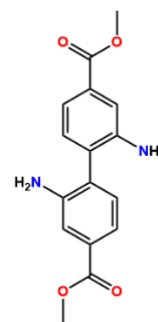
B.2.2.4 Dimethyl-2,2'-dinitro-[1,1'-biphenyl]-4,4'-dicarboxylate (**4**)¹³⁰

To a solution of commercial dimethyl-biphenyl-4,4'-dicarboxylate (10 g, 37 mmol) in 100 mL of concentrated H₂SO₄ at ~278 K a mixture of nitric acid (56%, 12 mL, 74 mmol) in 15 mL of concentrated sulfuric acid was added dropwise. The reaction mixture was stirred vigorously for 2 h at ~278 K and then was carefully poured onto ice (300 g). The pale yellow precipitated was filtered, washed with abundant cold water until neutral pH and air-dried to obtain 9.8 g of compound **4**. ¹H NMR (400 MHz, DMSO-d₆) δ : 8.77 (s, 2H), 8.23 (dd, J = 7.8, 1.7 Hz, 2H), 7.35 (d, J = 7.8 Hz, 2H), 3.98 (s, 6H).



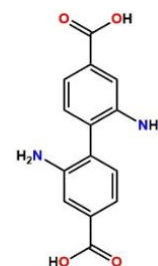
B.2.2.5 Dimethyl 2,2'-diamino-[1,1'-biphenyl]-4,4'-dicarboxylate (5)

A 250-mL three-necked round-bottomed flask was charged with 2.0 g of 10 wt.% Pd/C, 5 g of **4** (13.8 mmol), and 165 mL of THF. The flask was then evacuated and attached to a H₂ balloon. The reaction mixture was stirred at room temperature under H₂ atmosphere and monitored via thin layer chromatography (TLC). After filtration over Celite, the solvent was removed in vacuo affording 3.3 g of light yellow compound **5**. ¹H NMR (400 MHz, DMSO-d₆) δ : 7.42 (d, *J* = 1.3 Hz, 2H), 7.26 (dd, *J* = 7.8, 1.3 Hz, 2H), 7.09 (d, *J* = 7.8 Hz, 2H), 4.95 (s, 4H), 3.82 (s, 6H).



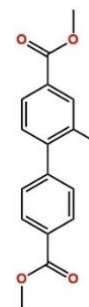
B.2.2.6 2,2'-diamino-[1,1'-biphenyl]-4,4'-dicarboxylic acid (**5**)¹³⁰

Compound **4** (3.75 g, 12.5 mmol) was dissolved in a mixture of 50:50 v/v THF/5% KOH (total volume 200 mL). The mixture was stirred overnight at 353 K. The aqueous layer was separated, then concentrated HCl was added until acid pH to give a yellowish solid. The solid was filtered, washed with abundant cold water and air-dried to obtain the desired product (3.01 g). ¹H NMR (400 MHz, DMSO-d₆) δ : 7.47 (d, *J* = 1.4 Hz, 2H), 7.28 (dd, *J* = 7.8, 1.4 Hz, 2H), 7.12 (d, *J* = 7.9 Hz, 2H).



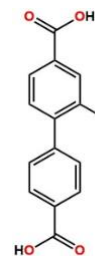
B.2.2.7 Dimethyl 2-methyl-1, 1'-biphenyl-4,4'-dicarboxylate (6)¹³¹

Ethylene glycol dimethyl ether (DME, 200 mL) was bubbled with nitrogen for about one hour before introduced into nitrogen-protected solid mixture of methyl 4-iodo-3-methylbenzoate (5.0 g, 18.1 mmol), 4-methoxycarbonylphenylboronic acid (3.9 g, 21.7 mmol), potassium carbonate (7.5 g, 54.25 mmol) and tetrakis (triphenylphosphine) palladium (0.3 g, 0.26 mmol). The mixture was allowed to reflux for 3 days under nitrogen protection. After cooling to room temperature, the solvent was evaporated to dryness. The residue was washed with a large amount of water followed by acetone. After removing the solvent, the residue was purified with column chromatography (silica gel, CH₂Cl₂) to give the ester as a white solid (3.4 g). ¹H NMR (400 MHz, DMSO-d₆) δ 8.05 (s, 2H), 7.94 (s, 1H), 7.86 (s, 1H), 7.56 (s, 1H), 7.41 (s, 1H), 3.89 (s, 6H), 2.30 (s, 3H).



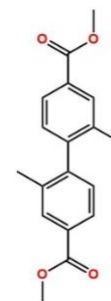
B.2.2.8 2-methyl-1, 1'-biphenyl-4,4'-dicarboxylate (7)¹³¹

Compound **5** (3.0 g, 10.6 mmol) was suspended in a mixture of THF/MeOH (50 mL, v/v = 1/1,) and 30 mL aqueous solution of 2 M KOH. The resultant mixture was stirred and refluxed overnight. After cooling to room temperature, organic solvents were removed, and the remaining solution was acidified with 1 M HCl to give a precipitate, which was collected and washed with water. Dried in the oven to produce 2.0 g of compound **6**. ¹H NMR (DMSO-d₆, 400 MHz) δ: 12.98 (s, 2H), 8.02 (s, 2H), 7.91 (s, 1H), 7.86 (s, 1H), 7.53 (s, 2H), 7.38 (s, 1H), 2.30 (s, 3H).



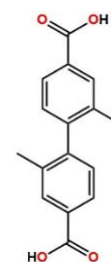
B.2.2.9 Dimethyl 2,2'-dimethyl-1, 1'-biphenyl-4,4'-dicarboxylate (8)¹³¹

Methyl 4-iodo-3-methylbenzoate (5.0 g, 18.1 mmol) and copper powder (20.0 g, 314.7 mmol) were thoroughly mixed under argon atmosphere. The resulting mixture was heated up to 250 °C overnight. After cooling, the reaction mixture was extracted with chloroform (30 mL x 3). The combined extracts were evaporated to dryness. Flash chromatography with ethyl acetate/hexanes (5% - 10%) as eluent afforded dimethyl 2,2'-dimethylbiphenyl-4,4'-dicarboxylate (1.4 g). ¹H NMR (CDCl₃, 400 MHz) δ : 7.90 (d, 2H), 7.89 (dd, 2H), 7.10 (d, 2H), 3.87 (s, 6H), 2.0 (s, 6H).



B.2.2.10 2,2'-dimethyl-1, 1'-biphenyl-4,4'-dicarboxylate (9)¹³¹

Compound 7 (1.0, 3.3 mmol) was suspended in a mixture of THF/MeOH (50 mL, v/v = 1/1, THF = tetrahydrofuran, MeOH = methanol), to which an aqueous solution of 2 M KOH (20 mL) was added. The resulting mixture was stirred and refluxed overnight. After cooling to room temperature, organic solvents were evaporated, and the remaining aqueous solution was acidified with 1 M HCl to give a precipitate, which was collected by filtration, washed with water, and dried in the oven to produce 0.6 g of compound 8. ¹H NMR (DMSO-d₆, 400 MHz) δ : 12.97 (s, 2H), 7.92 (d, 2H), 7.86 (dd, 2H), 7.23 (d, 2H), 2.05 (s, 6H).



Appendix B.2.3 Synthesis of MOFs

B.2.3.1 Synthesis of UiO-67

To a 20 mL Pyrex vial was added ZrCl_4 (9.8 mg, 0.04 mmol), DMF (10 mL), CH_3COOH (0.5 mL) and 1, 1'-biphenyl-4,4'-dicarboxylic acid (H_2 -BPDC) (9.3 mg, 0.04 mmol). After sonication for 5 min, the vial was placed in a 100 °C for 24 hours. The reaction suspension was then centrifuged at 10,000 rpm for 3 min to obtain white precipitate. The precipitate was washed with fresh DMF (16 mL, 4x) and dispersed in DMF (5 mL).

B.2.3.2 Synthesis of NH_2 -UiO-67

To a 20 mL Pyrex vial was added ZrCl_4 (9.8 mg, 0.04 mmol), DMF (10 mL), CH_3COOH (0.5 mL) and 2-amino-1, 1'-biphenyl-4,4'-dicarboxylic acid (H_2 - NH_2 -BPDC) (9.9 mg, 0.04 mmol). After sonication for 5 min, the vial was placed in a 100 °C for 24 hours. The reaction suspension was then centrifuged at 10,000 rpm for 3 min to obtain white precipitate. The precipitate was washed with fresh DMF (16 mL, 4x) and dispersed in DMF (5 mL).

B.2.3.3 Synthesis of $(\text{NH}_2)_2$ -UiO-67

To a 20 mL Pyrex vial was added ZrCl_4 (9.8 mg, 0.04 mmol), DMF (10 mL), CH_3COOH (0.5 mL) and 2,2'-diamino-1, 1'-biphenyl-4,4'-dicarboxylic acid (H_2 -2 NH_2 -BPDC) (10.5 mg, 0.04 mmol). After sonication for 5 min, the vial was placed in a 100 °C for 24 hours. The reaction

suspension was then centrifuged at 10,000 rpm for 3 min to obtain white precipitate. The precipitate was washed with fresh DMF (16 mL, 4x) and dispersed in DMF (5 mL).

B.2.3.4 Synthesis of CH₃-UiO-67

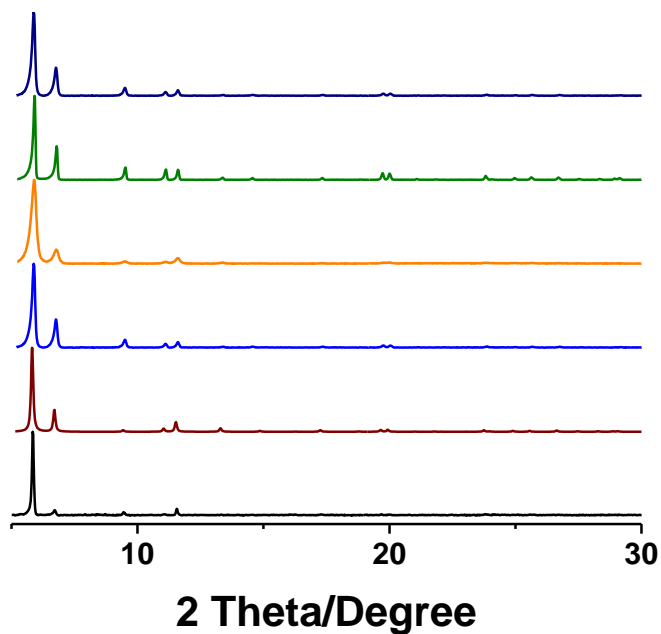
To a 20 mL Pyrex vial was added ZrCl₄ (9.8 mg, 0.04 mmol), DMF (10 mL), CH₃COOH (0.5 mL) and 2-methyl-1, 1'-biphenyl-4,4'-dicarboxylic acid (H₂-Me-BPDC) (9.8 mg, 0.04 mmol). After sonication for 5 min, the vial was placed in a 100 °C for 24 hours. The reaction suspension was then centrifuged at 10,000 rpm for 3 min to obtain white precipitate. The precipitate was washed with fresh DMF (16 mL, 4x) and dispersed in DMF (5 mL).

B.2.3.5 Synthesis of (CH₃)₂-UiO-67

To a 20 mL Pyrex vial was added ZrCl₄ (9.8 mg, 0.04 mmol), DMF (10 mL), CH₃COOH (0.5 mL) and 2,2'-dimethyl-1, 1'-biphenyl-4,4'-dicarboxylic acid (H₂-Me₂-BPDC) (10.4 mg, 0.04 mmol). After sonication for 5 min, the vial was placed in a 100 °C for 24 hours. The reaction suspension was then centrifuged at 10,000 rpm for 3 min to obtain white precipitate. The precipitate was washed with fresh DMF (16 mL, 4x) and dispersed in DMF (5 mL).

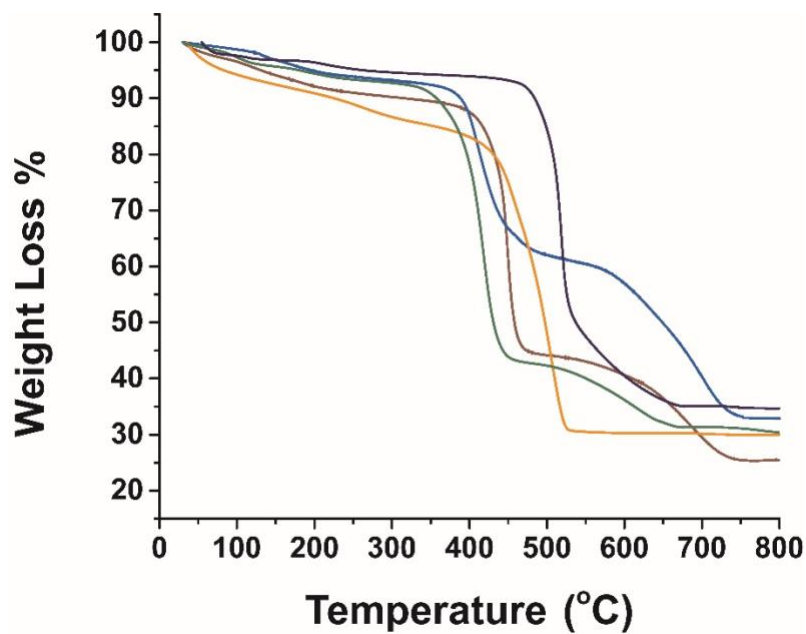
Appendix B.2.4 Characterization of MOFs

B.2.4.1 Powder X-ray diffraction data



Appendix Figure 6: Simulated PXRD pattern of UiO-67 (black) and experimental PXRD of as synthesized UiO-67 (dark red), NH₂-UiO-67 (blue), (NH₂)₂-UiO-67 (orange), CH₃-UiO-67 (green) and (CH₃)₂-UiO-67 (dark blue). These data confirm the crystallinity and phase purity of the synthesized MOFs.

B.2.4.2 Thermogravimetric analysis



Appendix Figure 7: TGA curves of UiO-67 (dark red), NH₂-UiO-67 (blue), (NH₂)₂-UiO-67 (orange), CH₃-UiO-67 (green) and (CH₃)₂-UiO-67 (dark blue).

B.2.4.3 Elemental analysis

Elemental and thermogravimetric analyses were used to determine the MOF molecular formulas. The theoretical molecular formulas $\text{Zr}_6\text{O}_4(\text{OH})_4(\text{X-BPDC})_{n1} \cdot n2\text{DMF}$ were determined based on TGA analysis according to literature method.¹³² TGA analysis of MOFs revealed an initial loss of ~4% between 90-200 °C corresponding to the DMF molecules within the framework (Appendix Figure 7). In TGA analysis, we denoted the mass right before the MOF samples started to decompose (~450 °C) as **a**, which represented $\text{Zr}_6\text{O}_4(\text{OH})_4(\text{X-BPDC})_n$. The final mass in TGA analysis after reaching plateau at ~500-750 °C was the remaining ZrO_2 . This was then used to calculate mass **b** for $\text{Zr}_6\text{O}_4(\text{OH})_4$ in the corresponding MOFs. The difference between mass **a** and **b** was attributed to ligands and used to calculate the number of ligand X-BPDC per molecule.

UiO-67

$\text{Zr}_6\text{O}_4(\text{OH})_4(\text{BPDC})_{5.7} \cdot \text{DMF}$, weight loss % in TGA by 200 °C: 3.7%

Calcd.: C, 46.57; H, 2.67; N, 0.66. Found: C, 46.99; H, 2.58; N, 0.34

NH₂-UiO-67

$\text{Zr}_6\text{O}_4(\text{OH})_4(\text{NH}_2\text{-BPDC})_{5.4} \cdot 0.8\text{DMF}$, weight loss % in TGA by 200 °C: 2.8%

Calcd.: C, 45.17; H, 2.80; N, 4.19. Found: C, 45.23; H, 2.89; N, 4.21

(NH₂)₂-UiO-67

$\text{Zr}_6\text{O}_4(\text{OH})_4(2\text{NH}_2\text{-BPDC})_{5.6} \cdot 1.4\text{DMF}$, weight loss % in TGA by 200 °C: 4.6%

Calcd.: C, 43.34; H, 3.12; N, 7.89. Found: C, 43.39; H, 3.07; N, 7.83

CH₃-UiO-67

Zr₆O₄(OH)₄(CH₃-BPDC)_{5.4}•1.2DMF, weight loss % in TGA by 200 °C: 4.2%

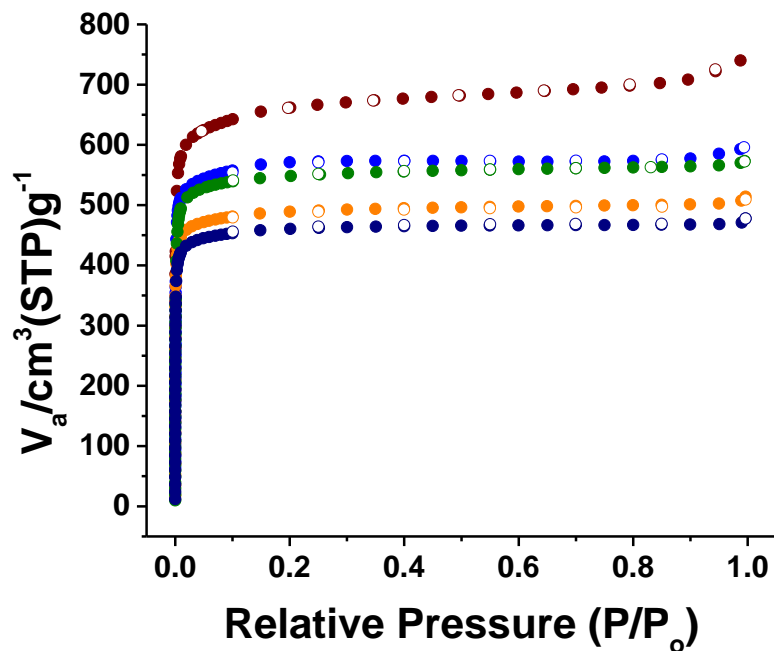
Calcd.: C, 46.57; H, 3.18; N, 0.80. Found: C, 46.53; H, 3.12; N, 0.74

(CH₃)₂-UiO-67

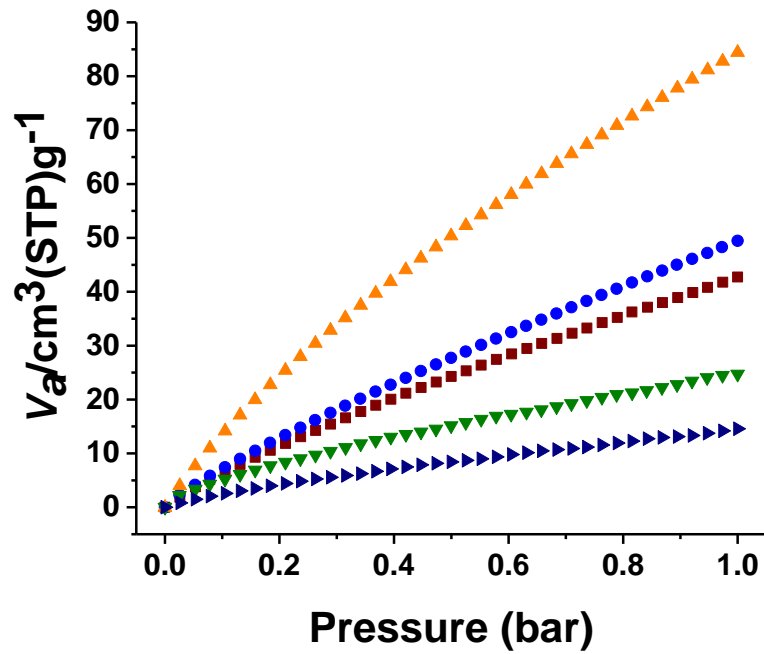
Zr₆O₄(OH)₄((CH₃)₂-BPDC)_{5.8}•1.1DMF, weight loss % in TGA by 200 °C: 3.5%

Calcd.: C, 51.12; H, 3.60; N, 0.68. Found: C, 50.99; H, 3.58; N, 0.64

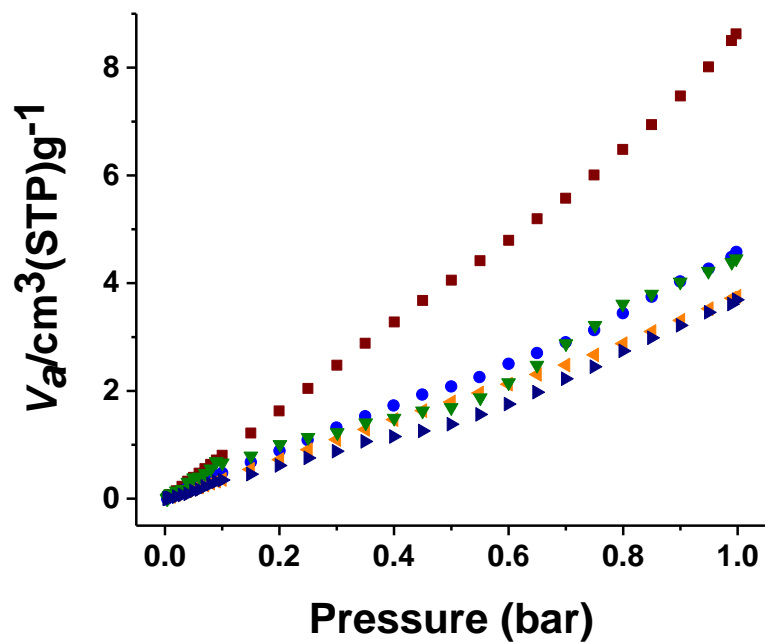
B.2.4.4 Gas sorption data



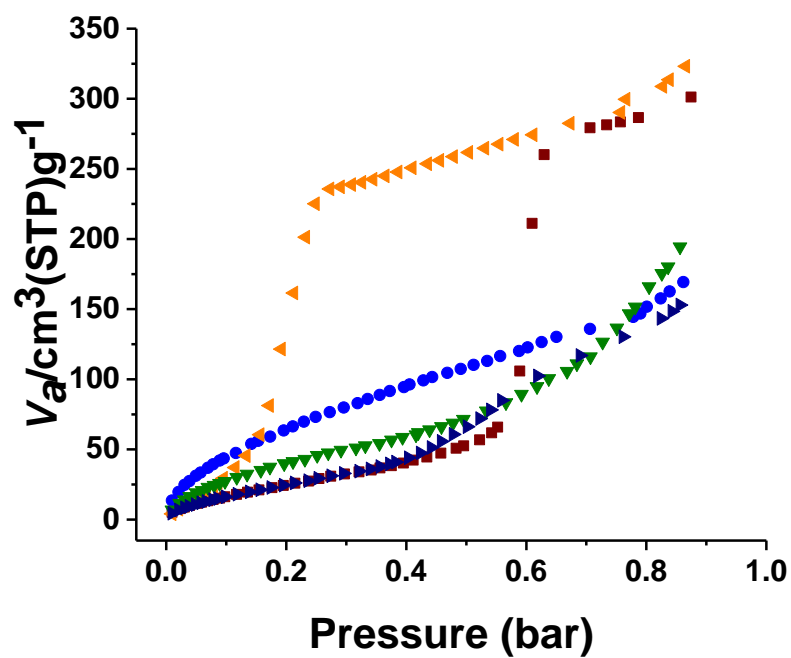
Appendix Figure 8: N₂ sorption isotherms of UiO-67 (dark red), NH₂-UiO-67 (blue), (NH₂)₂-UiO-67 (orange), CH₃-UiO-67 (green) and (CH₃)₂-UiO-67 (dark blue) at 77 K. Filled and hollow circles indicate adsorption and desorption respectively. These data were used to calculate the Branauer-Emmett-Teller (BET) surface areas for the MOFs: UiO-67, 2572 m²/g ; NH₂-UiO-67, 2074 m²/g ; (NH₂)₂-UiO-67, 1705 m²/g; CH₃-UiO-67, 2042 m²/g; and (CH₃)₂-UiO-67, 1647 m²/g. These values are consistent with literature data for UiO-67^{133,134} as well as NH₂-UiO-67 and CH₃-UiO-67.¹³⁴ The difunctionalized analogues have lower BET surface areas, as expected due to the additional functional groups.



Appendix Figure 9: CO₂ adsorption isotherms of UiO-67 (dark red square), NH₂-UiO-67 (blue circle), (NH₂)₂-UiO-67 (orange left-triangle), CH₃-UiO-67 (green down-triangle) and (CH₃)₂-UiO-67 (dark blue right-triangle) at 298 K.



Appendix Figure 10: N₂ adsorption isotherms of UiO-67 (dark red square), NH₂-UiO-67 (blue circle), (NH₂)₂-UiO-67 (orange left-triangle), CH₃-UiO-67 (green down-triangle) and (CH₃)₂-UiO-67 (dark blue right-triangle) at 298 K.



Appendix Figure 11: Water adsorption isotherms of UiO-67 (dark red square), NH₂-UiO-67 (blue circle), (NH₂)₂-UiO-67 (orange left-triangle), CH₃-UiO-67 (green down-triangle) and (CH₃)₂-UiO-67 (dark blue right-triangle) at 298 K.

B.2.4.5 Calculation of CO₂: N₂ adsorption selectivity

Appendix Table 1: Experimental CO₂:N₂ adsorption selectivity

	CO ₂ -298 K (cm ³ /g) @ 42 Pa	N ₂ -298 K (cm ³ /g) @ 79 kPa	CO ₂ /N ₂ Adsorption Selectivity
UiO-67	0.029	6.29	9.33
NH ₂ -UiO-67	0.032	3.38	17.8
(NH ₂) ₂ -UiO-67	0.058	2.84	38.4
CH ₃ -UiO-67	0.033	3.54	17.5
(CH ₃) ₂ -UiO-67	0.013	2.68	9.42

CO₂:N₂ adsorption selectivity was calculated using following equation:

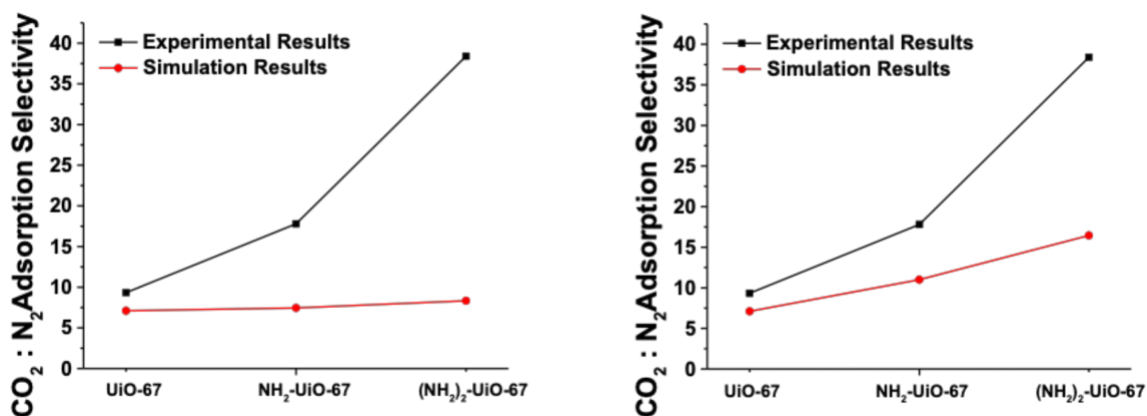
$$Adsorption\ selectivity = \frac{\frac{a}{42}}{\frac{b}{7.9} * 10^4}$$

a: CO₂ loading at 42 Pa, 298 K

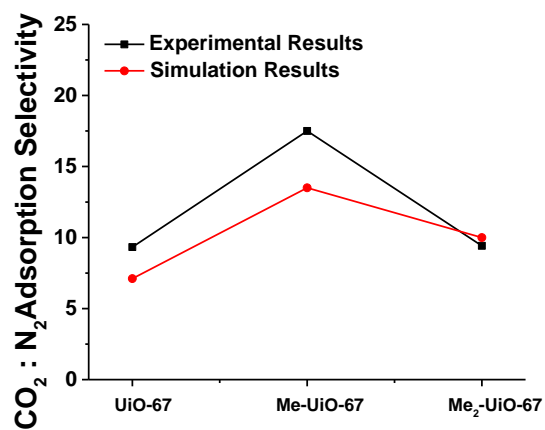
b: N₂ loading at 79 KPa, 298 K

Appendix Table 2: Comparison of experimental and simulated CO₂:N₂ adsorption selectivity.

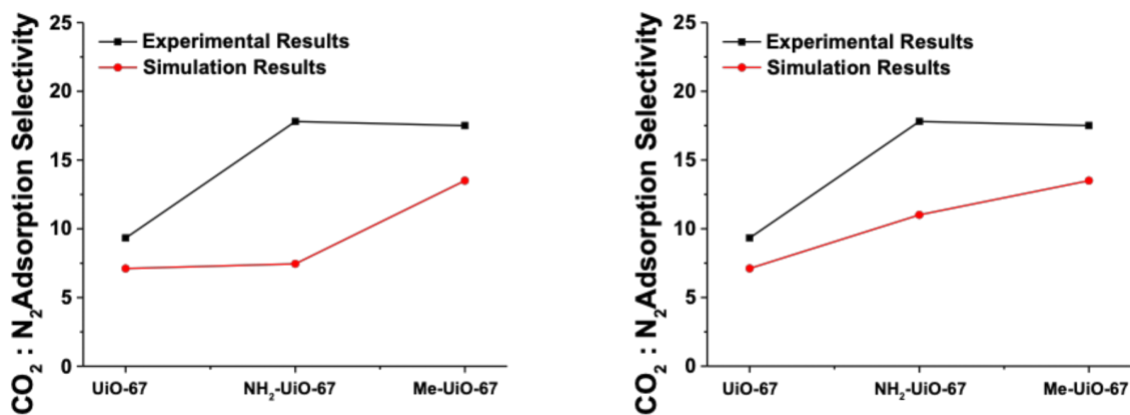
	CO ₂ /N ₂ Adsorption Selectivity	Simulation (default epsilon / 10x epsilon)
UiO-67	9.33	7.11
NH ₂ -UiO-67	17.8	7.45 / 11.01
(NH ₂) ₂ -UiO-67	38.4	8.33 / 16.45
CH ₃ -UiO-67	17.5	13.5
(CH ₃) ₂ -UiO-67	9.42	10.0



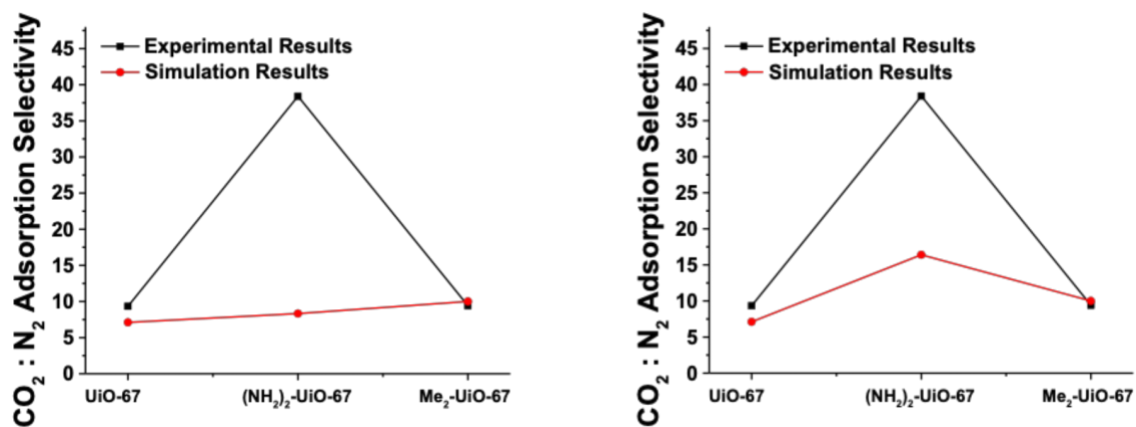
Appendix Figure 12: Comparison between experimental and simulated CO₂:N₂ adsorption selectivity of UiO-67, NH₂-UiO-67 and (NH₂)₂-UiO-67: (left) simulation data with default force field parameters for NH₂-CO₂ interaction, (right) simulation data with 10x epsilon force field parameters for NH₂-CO₂ interaction.



Appendix Figure 13: Comparison between experimental and simulated CO₂:N₂ adsorption selectivity of UiO-67, CH₃-UiO-67 and (CH₃)₂-UiO-67.



Appendix Figure 14: Comparison between experimental and simulated CO₂:N₂ adsorption selectivity of UiO-67, NH₂-UiO-67 and CH₃-UiO-67: (left) simulation data with default force field parameters for NH₂-CO₂ interaction, (right) simulation data with 10x epsilon force field parameters for NH₂-CO₂ interaction.



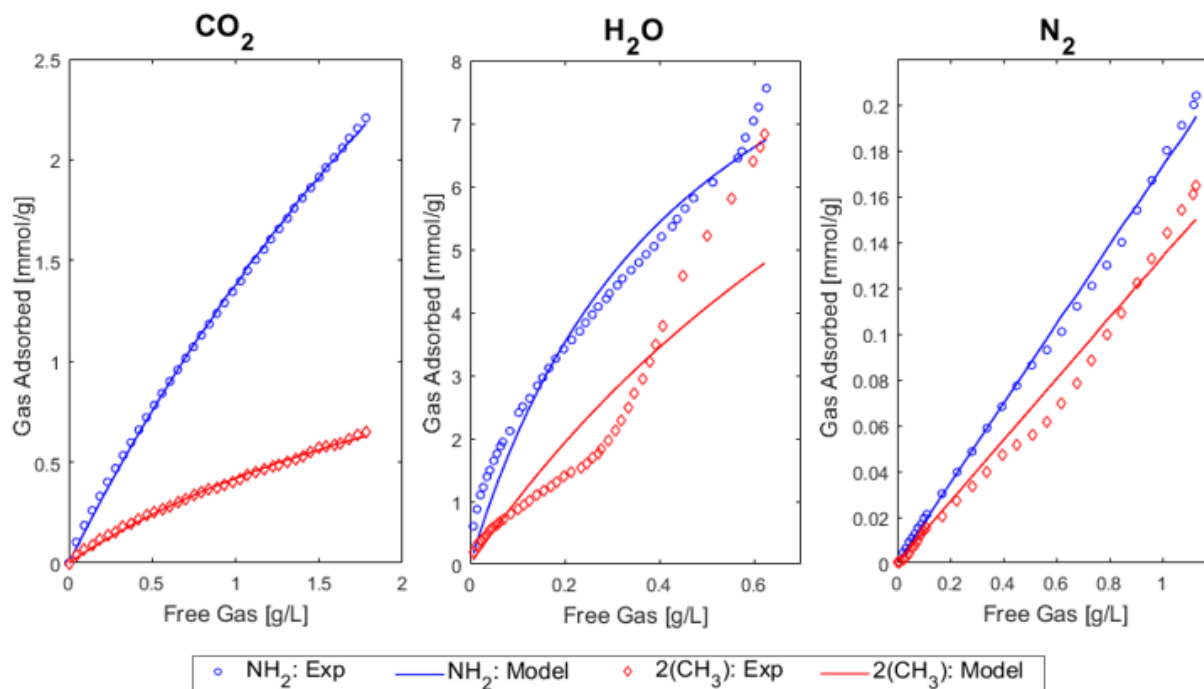
Appendix Figure 15: Comparison between experimental and simulated $\text{CO}_2:\text{N}_2$ adsorption selectivity of UiO-67, $(\text{NH}_2)_2\text{-UiO-67}$ and $(\text{CH}_3)_2\text{-UiO-67}$: (left) simulation data with default force field parameters for $\text{NH}_2\text{-CO}_2$ interaction, (right) simulation data with 10x epsilon force field parameters for $\text{NH}_2\text{-CO}_2$ interaction.

Appendix B.3 COMSOL Multiphysics® model

For the COMSOL Multiphysics® model case study, the Langmuir equation (Equation X) was fitted to experimental data in order to simulate gas adsorption. The fitted constants from that equation (K_L and C_{Pmax}) are presented for CO_2 , H_2O and N_2 in Appendix Table 3 for the two MOFs simulated in this paper: NH_2 and $(CH_3)_2$. The diffusivity values of CO_2 , H_2O and N_2 for these two MOFs are also reported in this table. Appendix Figure 16 shows our modeled Langmuir curve fits to experimental data for CO_2 , H_2O and N_2 for NH_2 and $(CH_3)_2$.

Appendix Table 3: Langmuir adsorption constants (K_L), adsorption maximums (C_{Pmax}), and diffusivity values for the two MOFs modeled in this paper.

	Constant	NH_2	$(CH_3)_2$
CO_2	K_L [m^3/mol]	0.197871	0.321812
	C_{Pmax} [mol/kg]	8.375795	1.72746
	D_{CO_2} [m^2/s]	5.05E-09	3.45E-09
H_2O	K_L [m^3/mol]	2.121584	0.691886
	C_{Pmax} [mol/kg]	11.81449	15.87641
	D_{H_2O} [m^2/s]	6.15E-11	9.24E-11
N_2	K_L [m^3/mol]	0.006575	0.00263
	C_{Pmax} [mol/kg]	26.45345	50.87282
	D_{H_2O} [m^2/s]	1.26E-08	6.75E-09



Appendix Figure 16: Experimental vs. modeled Langmuir adsorption isotherms for CO_2 (left), H_2O (middle), and N_2 (right). Experimental values are shown with markers; modeled values are shown with solid lines. Values for our simulated core MOF, NH_2 , are shown in blue; values for our simulated shell MOF, $(\text{CH}_3)_2$, are shown in red.

Bibliography

- (1) Joint Research Centre (European Commission); Monforti-Ferrario, F.; Muntean, M.; Vignati, E.; Crippa, M.; Solazzo, E.; Schaaf, E.; Olivier, J. G. J.; Guizzardi, D. *Fossil CO₂ and GHG Emissions of All World Countries: 2020 Report*; Publications Office of the European Union: LU, 2020.
- (2) Crippa, M.; Muntean, M.; Guizzardi, D.; Schaaf, E.; Monforti-Ferrario, F.; Olivier, J.; Vignati, E. Emissions Database for Global Atmospheric Research, Version v5.0_FT_2019 (Fossil CO₂ Time-Series), 2020. <http://data.europa.eu/89h/f806b247-2f28-4baa-8416-f8386a711ebf>.
- (3) Rogelj, J.; Shindell, D.; Jiang, K.; Fifita, S.; Forster, P.; Ginzburg, V.; Handa, C.; Kobayashi, S.; Kriegler, E.; Mundaca, L.; Séférián, R.; Vilariño, M. V. Mitigation Pathways Compatible with 1.5°C in the Context of Sustainable Development. In *Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty*; 2018.
- (4) de Coninck, H.; Revi, A.; Babiker, M.; Bertoldi, P.; Buckeridge, M.; Cartwright, A.; Dong, W.; Ford, J.; Fuss, S.; Hourcade, J.-C.; Ley, D.; Mechler, R.; Newman, P.; Revokatova, A.; Schultz, S.; Steg, L.; Sugiyama, T. Strengthening and Implementing the Global Response. In *Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty*; 2018.
- (5) Smith, P.; Davis, S. J.; Creutzig, F.; Fuss, S.; Minx, J.; Gabrielle, B.; Kato, E.; Jackson, R. B.; Cowie, A.; Kriegler, E.; van Vuuren, D. P.; Rogelj, J.; Ciais, P.; Milne, J.; Canadell, J. G.; McCollum, D.; Peters, G.; Andrew, R.; Krey, V.; Shrestha, G.; Friedlingstein, P.; Gasser, T.; Grubler, A.; Heidug, W. K.; Jonas, M.; Jones, C. D.; Kraxner, F.; Littleton, E.; Lowe, J.; Moreira, J. R.; Nakicenovic, N.; Obersteiner, M.; Patwardhan, A.; Rogner, M.; Rubin, E.; Sharifi, A.; Torvanger, A.; Yamagata, Y.; Edmonds, J.; Yongsung, C. Biophysical and Economic Limits to Negative CO₂ Emissions. *Nat. Clim. Change* **2016**, 6 (1), 42–50. <https://doi.org/10.1038/nclimate2870>.
- (6) Montserrat, F.; Renforth, P.; Hartmann, J.; Leermakers, M.; Knops, P.; Meysman, F. J. R. Olivine Dissolution in Seawater: Implications for CO₂ Sequestration through Enhanced Weathering in Coastal Environments. *Environ. Sci. Technol.* **2017**, 51 (7), 3960–3972. <https://doi.org/10.1021/acs.est.6b05942>.

- (7) Basu, P. *Biomass Gasification, Pyrolysis and Torrefaction: Practical Design and Theory*, 3rd edition.; Academic Press: London, United Kingdom ; San Diego, CA, United States, 2018.
- (8) IEA. *Coal Information: Overview*. IEA. <https://www.iea.org/reports/coal-information-overview> (accessed 2021-08-05).
- (9) Wurzbacher, J. A.; Gebald, C.; Steinfeld, A. Separation of CO₂ from Air by Temperature-Vacuum Swing Adsorption Using Diamine-Functionalized Silica Gel. *Energy Environ. Sci.* **2011**, 4 (9), 3584. <https://doi.org/10.1039/c1ee01681d>.
- (10) Keith, D. W.; Holmes, G.; St. Angelo, D.; Heidel, K. A Process for Capturing CO₂ from the Atmosphere. *Joule* **2018**, 2 (8), 1573–1594. <https://doi.org/10.1016/j.joule.2018.05.006>.
- (11) Holmes, G.; Keith, D. W. An Air-Liquid Contactor for Large-Scale Capture of CO₂ from Air. *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* **2012**, 370 (1974), 4380–4403. <https://doi.org/10.1098/rsta.2012.0137>.
- (12) Holmes, G.; Nold, K.; Walsh, T.; Heidel, K.; Henderson, M. A.; Ritchie, J.; Klavins, P.; Singh, A.; Keith, D. W. Outdoor Prototype Results for Direct Atmospheric Capture of Carbon Dioxide. *Energy Procedia* **2013**, 37, 6079–6095. <https://doi.org/10.1016/j.egypro.2013.06.537>.
- (13) Wurzbacher, J. A.; Gebald, C.; Piatkowski, N.; Steinfeld, A. Concurrent Separation of CO₂ and H₂O from Air by a Temperature-Vacuum Swing Adsorption/Desorption Cycle. *Environ. Sci. Technol.* **2012**, 46 (16), 9191–9198. <https://doi.org/10.1021/es301953k>.
- (14) Gebald, C.; Wurzbacher, J. A.; Tingaut, P.; Zimmermann, T.; Steinfeld, A. Amine-Based Nanofibrillated Cellulose As Adsorbent for CO₂ Capture from Air. *Environ. Sci. Technol.* **2011**, 45 (20), 9101–9108. <https://doi.org/10.1021/es202223p>.
- (15) Choi, S.; Drese, J. H.; Eisenberger, P. M.; Jones, C. W. Application of Amine-Tethered Solid Sorbents for Direct CO₂ Capture from the Ambient Air. *Environ. Sci. Technol.* **2011**, 45 (6), 2420–2427. <https://doi.org/10.1021/es102797w>.
- (16) Li, W.; Choi, S.; Drese, J. H.; Hornbostel, M.; Krishnan, G.; Eisenberger, P. M.; Jones, C. W. Steam-Stripping for Regeneration of Supported Amine-Based CO₂ Adsorbents. *ChemSusChem* **2010**, 3 (8), 899–903. <https://doi.org/10.1002/cssc.201000131>.
- (17) House, K. Z.; Baclig, A. C.; Ranjan, M.; van Nierop, E. A.; Wilcox, J.; Herzog, H. J. Economic and Energetic Analysis of Capturing CO₂ from Ambient Air. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, 108 (51), 20428–20433. <https://doi.org/10.1073/pnas.1012253108>.
- (18) Lackner, K. S. The Thermodynamics of Direct Air Capture of Carbon Dioxide. *Energy* **2013**, 50, 38–46. <https://doi.org/10.1016/j.energy.2012.09.012>.
- (19) Hepburn, C.; Adlen, E.; Beddington, J.; Carter, E. A.; Fuss, S.; Mac Dowell, N.; Minx, J. C.; Smith, P.; Williams, C. K. The Technological and Economic Prospects for CO₂

- Utilization and Removal. *Nature* **2019**, 575 (7781), 87–97. <https://doi.org/10.1038/s41586-019-1681-6>.
- (20) Rubin, E. S.; Davison, J. E.; Herzog, H. J. The Cost of CO₂ Capture and Storage. *Int. J. Greenh. Gas Control* **2015**, 40, 378–400. <https://doi.org/10.1016/j.ijggc.2015.05.018>.
 - (21) National Academies of Sciences, Engineering, and Medicine. *Negative Emissions Technologies and Reliable Sequestration: A Research Agenda*; National Academies Press: Washington, D.C., 2019.
 - (22) Brownsort, D. P. A. *Briefing on Carbon Dioxide Specifications for Transport. 1st Report of the Thematic Working Group on: CO₂ Transport, Storage and Networks*; 2019.
 - (23) Firth, A.; Zhang, B.; Yang, A. Quantification of Global Waste Heat and Its Environmental Effects. *Appl. Energy* **2019**, 235, 1314–1334. <https://doi.org/10.1016/j.apenergy.2018.10.102>.
 - (24) Willauer, H. D.; DiMascio, F.; Hardy, D. R.; Lewis, M. K.; Williams, F. W. Development of an Electrochemical Acidification Cell for the Recovery of CO₂ and H₂ from Seawater II. Evaluation of the Cell by Natural Seawater. *Ind. Eng. Chem. Res.* **2012**, 51 (34), 11254–11260. <https://doi.org/10.1021/ie301006y>.
 - (25) Willauer, H. D.; DiMascio, F.; Hardy, D. R.; Lewis, M. K.; Williams, F. W. Development of an Electrochemical Acidification Cell for the Recovery of CO₂ and H₂ from Seawater. *Ind. Eng. Chem. Res.* **2011**, 50 (17), 9876–9882. <https://doi.org/10.1021/ie2008136>.
 - (26) de Lannoy, C.-F.; Eisaman, M. D.; Jose, A.; Karnitz, S. D.; DeVaul, R. W.; Hannun, K.; Rivest, J. L. B. Indirect Ocean Capture of Atmospheric CO₂: Part I. Prototype of a Negative Emissions Technology. *Int. J. Greenh. Gas Control* **2018**, 70, 243–253. <https://doi.org/10.1016/j.ijggc.2017.10.007>.
 - (27) Wang, T.; Lackner, K. S.; Wright, A. B. Moisture-Swing Sorption for Carbon Dioxide Capture from Ambient Air: A Thermodynamic Analysis. *Phys Chem Chem Phys* **2013**, 15 (2), 504–514. <https://doi.org/10.1039/C2CP43124F>.
 - (28) Budhathoki, S.; Ajayi, O.; Steckel, J. A.; Wilmer, C. E. High-Throughput Computational Prediction of the Cost of Carbon Capture Using Mixed Matrix Membranes. *Energy Environ. Sci.* **2019**, 12 (4), 1255–1264. <https://doi.org/10.1039/C8EE02582G>.
 - (29) Lin, L.-C.; Berger, A. H.; Martin, R. L.; Kim, J.; Swisher, J. A.; Jariwala, K.; Rycroft, C. H.; Bhowan, A. S.; Deem, M. W.; Haranczyk, M.; Smit, B. In Silico Screening of Carbon-Capture Materials. *Nat. Mater.* **2012**, 11 (7), 633–641. <https://doi.org/10.1038/nmat3336>.
 - (30) Tong, M.; Lan, Y.; Yang, Q.; Zhong, C. High-Throughput Computational Screening and Design of Nanoporous Materials for Methane Storage and Carbon Dioxide Capture. *Green Energy Environ.* **2018**, 3 (2), 107–119. <https://doi.org/10.1016/j.gee.2017.09.004>.

- (31) Sharma, A.; Huang, R.; Malani, A.; Babarao, R. Computational Materials Chemistry for Carbon Capture Using Porous Materials. *J. Phys. Appl. Phys.* **2017**, *50* (46), 463002. <https://doi.org/10.1088/1361-6463/aa87e9>.
- (32) Dzubak, A. L.; Lin, L.-C.; Kim, J.; Swisher, J. A.; Poloni, R.; Maximoff, S. N.; Smit, B.; Gagliardi, L. Ab Initio Carbon Capture in Open-Site Metal–Organic Frameworks. *Nat. Chem.* **2012**, *4* (10), 810–816. <https://doi.org/10.1038/nchem.1432>.
- (33) Gíslason, S. R.; Sigurdardóttir, H.; Aradóttir, E. S.; Oelkers, E. H. A Brief History of CarbFix: Challenges and Victories of the Project’s Pilot Phase. *Energy Procedia* **2018**, *146*, 103–114. <https://doi.org/10.1016/j.egypro.2018.07.014>.
- (34) Gutknecht, V.; Snæbjörnsdóttir, S. Ó.; Sigfússon, B.; Aradóttir, E. S.; Charles, L. Creating a Carbon Dioxide Removal Solution by Combining Rapid Mineralization of CO₂ with Direct Air Capture. *Energy Procedia* **2018**, *146*, 129–134. <https://doi.org/10.1016/j.egypro.2018.07.017>.
- (35) Moosavi, S. M.; Nandy, A.; Jablonka, K. M.; Ongari, D.; Janet, J. P.; Boyd, P. G.; Lee, Y.; Smit, B.; Kulik, H. J. Understanding the Diversity of the Metal–Organic Framework Ecosystem. *Nat. Commun.* **2020**, *11* (1), 4068. <https://doi.org/10.1038/s41467-020-17755-8>.
- (36) Treacy, M. M. J.; Rivin, I.; Balkovsky, E.; Randall, K. H.; Foster, M. D. Enumeration of Periodic Tetrahedral Frameworks. II. Polynodal Graphs. *Microporous Mesoporous Mater.* **2004**, *74* (1), 121–132. <https://doi.org/10.1016/j.micromeso.2004.06.013>.
- (37) 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>.
- (38) Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **1995**, *117* (1), 1–19.
- (39) Thompson, A. P.; Aktulga, H. M.; Berger, R.; Bolintineanu, D. S.; Brown, W. M.; Crozier, P. S.; in ’t Veld, P. J.; Kohlmeyer, A.; Moore, S. G.; Nguyen, T. D.; Shan, R.; Stevens, M. J.; Tranchida, J.; Trott, C.; Plimpton, S. J. LAMMPS - a Flexible Simulation Tool for Particle-Based Materials Modeling at the Atomic, Meso, and Continuum Scales. *Comput. Phys. Commun.* **2022**, *271*, 108171. <https://doi.org/10.1016/j.cpc.2021.108171>.
- (40) Torii, D.; Nakano, T.; Ohara, T. Contribution of Inter- and Intramolecular Energy Transfers to Heat Conduction in Liquids. *J. Chem. Phys.* **2008**, *128* (4), 044504. <https://doi.org/10.1063/1.2821963>.
- (41) Sezginel, K. B.; Asinger, P. A.; Babaei, H.; Wilmer, C. E. Thermal Transport in Interpenetrated Metal–Organic Frameworks. *Chem. Mater.* **2018**, *30* (7), 2281–2286. <https://doi.org/10.1021/acs.chemmater.7b05015>.

- (42) Henry, A.; Chen, G. High Thermal Conductivity of Single Polyethylene Chains Using Molecular Dynamics Simulations. *Phys. Rev. Lett.* **2008**, *101* (23), 235502. <https://doi.org/10.1103/PhysRevLett.101.235502>.
- (43) Xu, X.; Pereira, L. F. C.; Wang, Y.; Wu, J.; Zhang, K.; Zhao, X.; Bae, S.; Tinh Bui, C.; Xie, R.; Thong, J. T. L.; Hong, B. H.; Loh, K. P.; Donadio, D.; Li, B.; Özyilmaz, B. Length-Dependent Thermal Conductivity in Suspended Single-Layer Graphene. *Nat. Commun.* **2014**, *5*, 3689. <https://doi.org/10.1038/ncomms4689>.
- (44) Lv, W.; Henry, A. Phonon Transport in Amorphous Carbon Using Green – Kubo Modal Analysis. *Appl. Phys. Lett.* **2016**, *108* (18), 181905. <https://doi.org/10.1063/1.4948605>.
- (45) Shiomi, J. Nonequilibrium Molecular Dynamics Methods for Lattice Heat Conduction Calculations. *Annu. Rev. Heat Transf.* **2014**, *17*, 177–203. <https://doi.org/10.1615/AnnualRevHeatTransfer.2014007407>.
- (46) Schelling, P. K.; Phillpot, S. R.; Keblinski, P. Comparison of Atomic-Level Simulation Methods for Computing Thermal Conductivity. *Phys. Rev. B* **2002**, *65* (14), 144306. <https://doi.org/10.1103/PhysRevB.65.144306>.
- (47) Müller-Plathe, F. A Simple Nonequilibrium Molecular Dynamics Method for Calculating the Thermal Conductivity. *J. Chem. Phys.* **1997**, *106* (14), 6082–6085. <https://doi.org/10.1063/1.473271>.
- (48) Kubo, R. Statistical-Mechanical Theory of Irreversible Processes. I. General Theory and Simple Applications to Magnetic and Conduction Problems. *J. Phys. Soc. Jpn.* **1957**, *12* (6), 570–586. <https://doi.org/10.1143/JPSJ.12.570>.
- (49) Green, M. S. Markoff Random Processes and the Statistical Mechanics of Time-Dependent Phenomena. II. Irreversible Processes in Fluids. *J. Chem. Phys.* **1954**, *22* (3), 398–413. <https://doi.org/10.1063/1.1740082>.
- (50) Ohara, T.; Chia Yuan, T.; Torii, D.; Kikugawa, G.; Kosugi, N. Heat Conduction in Chain Polymer Liquids: Molecular Dynamics Study on the Contributions of Inter- and Intramolecular Energy Transfer. *J. Chem. Phys.* **2011**, *135* (3), 034507. <https://doi.org/10.1063/1.3613648>.
- (51) Rashidi, V.; Coyle, E. J.; Sebeck, K.; Kieffer, J.; Pipe, K. P. Thermal Conductance in Cross-Linked Polymers: Effects of Non-Bonding Interactions. *J. Phys. Chem. B* **2017**, *121* (17), 4600–4609. <https://doi.org/10.1021/acs.jpcc.7b01377>.
- (52) *compute heat/flux command — LAMMPS documentation.* http://lammps.sandia.gov/doc/compute_heat_flux.html (accessed 2018-02-13).
- (53) Fan, Z.; Pereira, L. F. C.; Wang, H.-Q.; Zheng, J.-C.; Donadio, D.; Harju, A. Force and Heat Current Formulas for Many-Body Potentials in Molecular Dynamics Simulations with Applications to Thermal Conductivity Calculations. *Phys. Rev. B* **2015**, *92* (9), 094301. <https://doi.org/10.1103/PhysRevB.92.094301>.

- (54) Irving, J. H.; Kirkwood, J. G. The Statistical Mechanical Theory of Transport Processes. IV. The Equations of Hydrodynamics. *J. Chem. Phys.* **1950**, *18* (6), 817–829. <https://doi.org/10.1063/1.1747782>.
- (55) Hardy, R. J. Formulas for Determining Local Properties in Molecular-dynamics Simulations: Shock Waves. *J. Chem. Phys.* **1982**, *76* (1), 622–628. <https://doi.org/10.1063/1.442714>.
- (56) McGaughey, A. J. H.; Kaviani, M. Phonon Transport in Molecular Dynamics Simulations: Formulation and Thermal Conductivity Prediction. In *Advances in Heat Transfer*; Greene, G. A., Hartnett†, J. P., Bar-Cohen, A., Cho, Y. I., Eds.; Elsevier, 2006; Vol. 39, pp 169–255. [https://doi.org/10.1016/S0065-2717\(06\)39002-8](https://doi.org/10.1016/S0065-2717(06)39002-8).
- (57) Thompson, A. P.; Plimpton, S. J.; Mattson, W. General Formulation of Pressure and Stress Tensor for Arbitrary Many-Body Interaction Potentials under Periodic Boundary Conditions. *J. Chem. Phys.* **2009**, *131* (15), 154107. <https://doi.org/10.1063/1.3245303>.
- (58) *lammps/compute_heat_flux.cpp* at master · WilmerLab/lammps. https://github.com/WilmerLab/lammps/blob/master/src/compute_heat_flux.cpp (accessed 2019-05-07).
- (59) *lammps/angle.cpp* at master · WilmerLab/lammps. <https://github.com/WilmerLab/lammps/blob/master/src/angle.cpp> (accessed 2019-05-07).
- (60) Babaei, H.; McGaughey, A. J. H.; Wilmer, C. E. Effect of Pore Size and Shape on the Thermal Conductivity of Metal-Organic Frameworks. *Chem. Sci.* **2017**, *8* (1), 583–589. <https://doi.org/10.1039/c6sc03704f>.
- (61) Ohara, T.; Chia Yuan, T.; Torii, D.; Kikugawa, G.; Kosugi, N. Heat Conduction in Chain Polymer Liquids: Molecular Dynamics Study on the Contributions of Inter- and Intramolecular Energy Transfer. *J. Chem. Phys.* **2011**, *135* (3), 034507. <https://doi.org/10.1063/1.3613648>.
- (62) Nath, S. K.; Escobedo, F. A.; de Pablo, J. J. On the Simulation of Vapor–Liquid Equilibria for Alkanes. *J. Chem. Phys.* **1998**, *108* (23), 9905–9911. <https://doi.org/10.1063/1.476429>.
- (63) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. PACKMOL: A Package for Building Initial Configurations for Molecular Dynamics Simulations. *J. Comput. Chem.* **2009**, *30* (13), 2157–2164. <https://doi.org/10.1002/jcc.21224>.
- (64) 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* (21), 6185–6192. <https://doi.org/10.1021/cm502594j>.
- (65) Yeo, B. C.; Kim, D.; Kim, H.; Han, S. S. High-Throughput Screening to Investigate the Relationship between the Selectivity and Working Capacity of Porous Materials for

- Propylene/Propane Adsorptive Separation. *J. Phys. Chem. C* **2016**, *120* (42), 24224–24230. <https://doi.org/10.1021/acs.jpcc.6b08177>.
- (66) 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>.
- (67) 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>.
- (68) Simon, C. M.; Kim, J.; Lin, L.-C.; Martin, R. L.; Haranczyk, M.; Smit, B. Optimizing Nanoporous Materials for Gas Storage. *Phys. Chem. Chem. Phys.* **2014**, *16* (12), 5499. <https://doi.org/10.1039/c3cp55039g>.
- (69) Pophale, R.; Cheeseman, P. A.; Deem, M. W. A Database of New Zeolite-like Materials. *Phys. Chem. Chem. Phys.* **2011**, *13* (27), 12407–12412. <https://doi.org/10.1039/C0CP02255A>.
- (70) Boyd, P. G.; Lee, Y.; Smit, B. Computational Development of the Nanoporous Materials Genome. *Nat. Rev. Mater.* **2017**, *2* (8), 17037. <https://doi.org/10.1038/natrevmats.2017.37>.
- (71) Draznieks, C. M.; Newsam, J. M.; Gorman, A. M.; Freeman, C. M.; Férey, G. De Novo Prediction of Inorganic Structures Developed through Automated Assembly of Secondary Building Units (AASBU Method). *Angew. Chem. Int. Ed.* **2000**, *39* (13), 2270–2275. [https://doi.org/10.1002/1521-3773\(20000703\)39:13<2270::AID-ANIE2270>3.0.CO;2-A](https://doi.org/10.1002/1521-3773(20000703)39:13<2270::AID-ANIE2270>3.0.CO;2-A).
- (72) 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>.
- (73) 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>.
- (74) Earl, D. J.; Deem, M. W. Toward a Database of Hypothetical Zeolite Structures. *Ind. Eng. Chem. Res.* **2006**, *45* (16), 5449–5454. <https://doi.org/10.1021/ie0510728>.
- (75) Simon, C. M.; Kim, J.; Gomez-Gualdrón, D. A.; Camp, J. S.; Chung, Y. G.; Martin, R. L.; Mercado, R.; Deem, M. W.; Gunter, D.; Haranczyk, M.; Sholl, D. S.; Snurr, R. Q.; Smit, B. The Materials Genome in Action: Identifying the Performance Limits for Methane Storage. *Energy Environ. Sci.* **2015**, *8* (4), 1190–1199. <https://doi.org/10.1039/C4EE03515A>.
- (76) Kaija, A. R.; Wilmer, C. E. Efficiently Mapping Structure–Property Relationships of Gas Adsorption in Porous Materials: Application to Xe Adsorption. *Faraday Discuss.* **2017**, *201*, 221–232. <https://doi.org/10.1039/C7FD00038C>.

- (77) Kaija, A. R.; Wilmer, C. E. High-Pressure Methane Adsorption in Porous Lennard-Jones Crystals. *J. Phys. Chem. Lett.* **2018**, *9* (15), 4275–4281. <https://doi.org/10.1021/acs.jpclett.8b01421>.
- (78) Gómez-Gualdrón, D. A.; Wilmer, C. E.; Farha, O. K.; Hupp, J. T.; Snurr, R. Q. Exploring the Limits of Methane Storage and Delivery in Nanoporous Materials. *J. Phys. Chem. C* **2014**, *118* (13), 6941–6951. <https://doi.org/10.1021/jp502359q>.
- (79) Bracco, S.; Piga, D.; Bassanetti, I.; Perego, J.; Comotti, A.; Sozzani, P. Porous 3D Polymers for High Pressure Methane Storage and Carbon Dioxide Capture. *J. Mater. Chem. A* **2017**, *5* (21), 10328–10337. <https://doi.org/10.1039/C7TA00934H>.
- (80) Dubbeldam, D.; Calero, S.; Ellis, D. E.; Snurr, R. Q. RASPA: Molecular Simulation Software for Adsorption and Diffusion in Flexible Nanoporous Materials. *Mol. Simul.* **2016**, *42* (2), 81–101. <https://doi.org/10.1080/08927022.2015.1010082>.
- (81) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Clarendon Press: New York, NY, USA, 1987.
- (82) Martin, M. G.; Siepmann, J. I. Transferable Potentials for Phase Equilibria. 1. United-Atom Description of n-Alkanes. *J. Phys. Chem. B* **1998**, *102* (14), 2569–2577.
- (83) Rappé, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard Iii, W. A.; Skiff, W. M. UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **1992**, *114* (25), 10024–10035.
- (84) Gómez-Gualdrón, D. A.; Simon, C. M.; Lassman, W.; Chen, D.; Martin, R. L.; Haranczyk, M.; Farha, O. K.; Smit, B.; Snurr, R. Q. Impact of the Strength and Spatial Distribution of Adsorption Sites on Methane Deliverable Capacity in Nanoporous Materials. *Chem. Eng. Sci.* **2017**, *159*, 18–30. <https://doi.org/10.1016/j.ces.2016.02.030>.
- (85) 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>.
- (86) Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R. Avogadro: An Advanced Semantic Chemical Editor, Visualization, and Analysis Platform. *J. Cheminformatics* **2012**, *4* (1), 17. <https://doi.org/10.1186/1758-2946-4-17>.
- (87) Bae, Y.-S.; Liu, J.; Wilmer, C. E.; Sun, H.; Dickey, A. N.; Kim, M. B.; Benin, A. I.; Willis, R. R.; Barpaga, D.; LeVan, M. D.; Snurr, R. Q. The Effect of Pyridine Modification of Ni–DOBDC on CO₂ Capture under Humid Conditions. *Chem Commun* **2014**, *50* (25), 3296–3298. <https://doi.org/10.1039/C3CC44954H>.
- (88) Henle, E. A.; Gantzler, N.; Thallapally, P. K.; Fern, X. Z.; Simon, C. M. PoreMatMod.Jl: Julia Package for in Silico Postsynthetic Modification of Crystal Structure Models. *J. Chem. Inf. Model.* **2022**, *62* (3), 423–432. <https://doi.org/10.1021/acs.jcim.1c01219>.

- (89) Larsen, A. H.; Mortensen, J. J.; Blomqvist, J.; Castelli, I. E.; Christensen, R.; Du\lak, M.; Friis, J.; Groves, M. N.; Hammer, B.; Hargus, C.; Hermes, E. D.; Jennings, P. C.; Jensen, P. B.; Kermode, J.; Kitchin, J. R.; Kolsbjerg, E. L.; Kubal, J.; Kaasbjerg, K.; Lysgaard, S.; Maronsson, J. B.; Maxson, T.; Olsen, T.; Pastewka, L.; Peterson, A.; Rostgaard, C.; Schiøtz, J.; Schütt, O.; Strange, M.; Thygesen, K. S.; Vegge, T.; Vilhelmsen, L.; Walter, M.; Zeng, Z.; Jacobsen, K. W. The Atomic Simulation Environment—a Python Library for Working with Atoms. *J. Phys. Condens. Matter* **2017**, 29 (27), 273002. <https://doi.org/10.1088/1361-648X/aa680e>.
- (90) Momma, K.; Izumi, F. VESTA: A Three-Dimensional Visualization System for Electronic and Structural Analysis. *J. Appl. Crystallogr.* **2008**, 41 (3), 653–658. <https://doi.org/10.1107/S0021889808012016>.
- (91) Wu, H.; Chua, Y. S.; Krungleviciute, V.; Tyagi, M.; Chen, P.; Yildirim, T.; Zhou, W. Unusual and Highly Tunable Missing-Linker Defects in Zirconium Metal–Organic Framework UiO-66 and Their Important Effects on Gas Adsorption. *J. Am. Chem. Soc.* **2013**, 135 (28), 10525–10532. <https://doi.org/10.1021/ja404514r>.
- (92) Gutov, O. V.; Hevia, M. G.; Escudero-Adán, E. C.; Shafir, A. Metal–Organic Framework (MOF) Defects under Control: Insights into the Missing Linker Sites and Their Implication in the Reactivity of Zirconium-Based Frameworks. *Inorg. Chem.* **2015**, 54 (17), 8396–8400. <https://doi.org/10.1021/acs.inorgchem.5b01053>.
- (93) Al-Janabi, N.; Fan, X.; Siperstein, F. R. Assessment of MOF’s Quality: Quantifying Defect Content in Crystalline Porous Materials. *J. Phys. Chem. Lett.* **2016**, 7 (8), 1490–1494. <https://doi.org/10.1021/acs.jpcllett.6b00297>.
- (94) Rouquerol, J.; Rouquerol, F.; Llewellyn, P.; Maurin, G.; Sing, K. S. W. *Adsorption by Powders and Porous Solids: Principles, Methodology and Applications*, 2nd edition.; Academic Press: Amsterdam, 2013.
- (95) Mayo, S. L.; Olafson, B. D.; Goddard, W. A. DREIDING: A Generic Force Field for Molecular Simulations. *J. Phys. Chem.* **1990**, 94 (26), 8897–8909. <https://doi.org/10.1021/j100389a010>.
- (96) Sun, H. COMPASS: An Ab Initio Force-Field Optimized for Condensed-Phase Applications Overview with Details on Alkane and Benzene Compounds. *J. Phys. Chem. B* **1998**, 102 (38), 7338–7364.
- (97) Addicoat, M. A.; Vankova, N.; Akter, I. F.; Heine, T. Extension of the Universal Force Field to Metal–Organic Frameworks. *J. Chem. Theory Comput.* **2014**, 10 (2), 880–891. <https://doi.org/10.1021/ct400952t>.
- (98) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, 117 (19), 5179–5197. <https://doi.org/10.1021/ja00124a002>.

- (99) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102* (18), 3586–3616. <https://doi.org/10.1021/jp973084f>.
- (100) 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>.
- (101) Lebling, K.; Leslie-Bole, H.; Byrum, Z.; Bridgwater, E. 6 Things to Know About Direct Air Capture. **2022**.
- (102) *Direct Air Capture: 6 Things To Know | World Resources Institute*. <https://www.wri.org/insights/direct-air-capture-resource-considerations-and-costs-carbon-removal> (accessed 2022-06-01).
- (103) Sanz-Pérez, E. S.; Murdock, C. R.; Didas, S. A.; Jones, C. W. Direct Capture of CO₂ from Ambient Air. *Chem. Rev.* **2016**, *116* (19), 11840–11876. <https://doi.org/10.1021/acs.chemrev.6b00173>.
- (104) McQueen, N.; Gomes, K. V.; McCormick, C.; Blumanthal, K.; Pisciotta, M.; Wilcox, J. A Review of Direct Air Capture (DAC): Scaling up Commercial Technologies and Innovating for the Future. *Prog. Energy* **2021**, *3* (3), 032001. <https://doi.org/10.1088/2516-1083/abf1ce>.
- (105) Guo, M.; Wu, H.; Lv, L.; Meng, H.; Yun, J.; Jin, J.; Mi, J. A Highly Efficient and Stable Composite of Polyacrylate and Metal–Organic Framework Prepared by Interface Engineering for Direct Air Capture. *ACS Appl. Mater. Interfaces* **2021**, *13* (18), 21775–21785. <https://doi.org/10.1021/acsami.1c03661>.
- (106) van Schagen, T. N.; van der Wal, P. J.; Brilman, D. W. F. Development of a Novel, through-Flow Microwave-Based Regenerator for Sorbent-Based Direct Air Capture. *Chem. Eng. J. Adv.* **2022**, *9*, 100187. <https://doi.org/10.1016/j.cej.2021.100187>.
- (107) Furukawa, H.; Cordova, K. E.; O’Keeffe, M.; Yaghi, O. M. The Chemistry and Applications of Metal-Organic Frameworks. *Science* **2013**, *341* (6149). <https://doi.org/10.1126/science.1230444>.
- (108) Kumar, A.; Madden, D. G.; Lusi, M.; Chen, K.-J.; Daniels, E. A.; Curtin, T.; Perry, J. J.; Zaworotko, M. J. Direct Air Capture of CO₂ by Physisorbent Materials. *Angew. Chem. Int. Ed.* **2015**, *54* (48), 14372–14377. <https://doi.org/10.1002/anie.201506952>.
- (109) Madden, D. G.; Scott, H. S.; Kumar, A.; Chen, K.-J.; Sanii, R.; Bajpai, A.; Lusi, M.; Curtin, T.; Perry, J. J.; Zaworotko, M. J. Flue-Gas and Direct-Air Capture of CO₂ by Porous Metal–

- Organic Materials. *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* **2017**, 375 (2084), 20160025. <https://doi.org/10.1098/rsta.2016.0025>.
- (110) Burtch, N. C.; Jasuja, H.; Walton, K. S. Water Stability and Adsorption in Metal–Organic Frameworks. *Chem. Rev.* **2014**, 114 (20), 10575–10612. <https://doi.org/10.1021/cr5002589>.
- (111) Luo, T.-Y.; Liu, C.; Gan, X. Y.; Muldoon, P. F.; Diemler, N. A.; Millstone, J. E.; Rosi, N. L. Multivariate Stratified Metal–Organic Frameworks: Diversification Using Domain Building Blocks. *J. Am. Chem. Soc.* **2019**, 141 (5), 2161–2168. <https://doi.org/10.1021/jacs.8b13502>.
- (112) Hong, D. H.; Shim, H. S.; Ha, J.; Moon, H. R. MOF-on-MOF Architectures: Applications in Separation, Catalysis, and Sensing. *Bull. Korean Chem. Soc.* **2021**, 42 (7), 956–969. <https://doi.org/10.1002/bkcs.12335>.
- (113) Furukawa, S.; Hirai, K.; Nakagawa, K.; Takashima, Y.; Matsuda, R.; Tsuruoka, T.; Kondo, M.; Haruki, R.; Tanaka, D.; Sakamoto, H.; Shimomura, S.; Sakata, O.; Kitagawa, S. Heterogeneously Hybridized Porous Coordination Polymer Crystals: Fabrication of Heterometallic Core–Shell Single Crystals with an In-Plane Rotational Epitaxial Relationship. *Angew. Chem. Int. Ed.* **2009**, 48 (10), 1766–1770. <https://doi.org/10.1002/anie.200804836>.
- (114) Koh, K.; Wong-Foy, A. G.; Matzger, A. J. MOF@MOF: Microporous Core–Shell Architectures. *Chem. Commun.* **2009**, No. 41, 6162–6164. <https://doi.org/10.1039/B904526K>.
- (115) Li, T.; Sullivan, J. E.; Rosi, N. L. Design and Preparation of a Core–Shell Metal–Organic Framework for Selective CO₂ Capture. *J. Am. Chem. Soc.* **2013**, 135 (27), 9984–9987. <https://doi.org/10.1021/ja403008j>.
- (116) He, Y.; Sun, M.; Zhao, Q.; Shang, J.; Tian, Y.; Xiao, P.; Gu, Q.; Li, L.; Webley, P. A. Effective Gas Separation Performance Enhancement Obtained by Constructing Polymorphous Core–Shell Metal–Organic Frameworks. *ACS Appl. Mater. Interfaces* **2019**, 11 (33), 30234–30239. <https://doi.org/10.1021/acsami.9b08592>.
- (117) Song, Z.; Qiu, F.; Zaia, E. W.; Wang, Z.; Kunz, M.; Guo, J.; Brady, M.; Mi, B.; Urban, J. J. Dual-Channel, Molecular-Sieving Core/Shell ZIF@MOF Architectures as Engineered Fillers in Hybrid Membranes for Highly Selective CO₂ Separation. *Nano Lett.* **2017**, 17 (11), 6752–6758. <https://doi.org/10.1021/acs.nanolett.7b02910>.
- (118) Cavka, J. H.; Jakobsen, S.; Olsbye, U.; Guillou, N.; Lamberti, C.; Bordiga, S.; Lillerud, K. P. A New Zirconium Inorganic Building Brick Forming Metal Organic Frameworks with Exceptional Stability. *J. Am. Chem. Soc.* **2008**, 130 (42), 13850–13851. <https://doi.org/10.1021/ja8057953>.
- (119) Usman, M.; Helal, A.; Abdelnaby, M. M.; Alloush, A. M.; Zeama, M.; Yamani, Z. H. Trends and Prospects in UiO-66 Metal–Organic Framework for CO₂ Capture, Separation, and Conversion. *Chem. Rec.* **2021**, 21 (7), 1771–1791. <https://doi.org/10.1002/tcr.202100030>.

- (120) Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd edition.; Academic Press: San Diego, 2001.
- (121) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79* (2), 926–935. <https://doi.org/10.1063/1.445869>.
- (122) Horn, H. W.; Swope, W.; Pitera, J.; Madura, J.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. Development of an Improved Four-Site Water Model for Biomolecular Simulations: TIP4P-Ew. *J. Chem. Phys.* **2004**, *120* (20), 9665–9678. <https://doi.org/10.1063/1.1683075>.
- (123) Wilmer, C. E.; Kim, K. C.; Snurr, R. Q. An Extended Charge Equilibration Method. *J. Phys. Chem. Lett.* **2012**, *3* (17), 2506–2511. <https://doi.org/10.1021/jz3008485>.
- (124) Boone, P.; Wilmer, C. E. MOFUn: A Python Package for Molecular Find and Replace. *Manuscr. Submitt. Publ.*
- (125) Maginn, E. J.; Messerly, R. A.; Carlson, D. J.; Roe, D. R.; Elliott, J. R. Best Practices for Computing Transport Properties 1. Self-Diffusivity and Viscosity from Equilibrium Molecular Dynamics [Article v1.0]. *Living J. Comput. Mol. Sci.* **2019**, *1* (1). <https://doi.org/10.33011/livecoms.1.1.6324>.
- (126) Zhang, H.; Snurr, R. Q. Computational Study of Water Adsorption in the Hydrophobic Metal–Organic Framework ZIF-8: Adsorption Mechanism and Acceleration of the Simulations. *J. Phys. Chem. C* **2017**, *121* (43), 24000–24010. <https://doi.org/10.1021/acs.jpcc.7b06405>.
- (127) Datar, A.; Witman, M.; Lin, L.-C. Improving Computational Assessment of Porous Materials for Water Adsorption Applications via Flat Histogram Methods. *J. Phys. Chem. C* **2021**, *125* (7), 4253–4266. <https://doi.org/10.1021/acs.jpcc.0c11082>.
- (128) Said, R. B.; Kolle, J. M.; Essalah, K.; Tangour, B.; Sayari, A. A Unified Approach to CO₂–Amine Reaction Mechanisms. *ACS Omega* **2020**, *5* (40), 26125–26133. <https://doi.org/10.1021/acsomega.0c03727>.
- (129) Liu, C.; Li, T.; Rosi, N. L. Strain-Promoted “Click” Modification of a Mesoporous Metal–Organic Framework. *J. Am. Chem. Soc.* **2012**, *134* (46), 18886–18888. <https://doi.org/10.1021/ja307713q>.
- (130) Gil-San-Millan, R.; López-Maya, E.; Hall, M.; Padial, N. M.; Peterson, G. W.; DeCoste, J. B.; Rodríguez-Albelo, L. M.; Oltra, J. E.; Barea, E.; Navarro, J. A. R. Chemical Warfare Agents Detoxification Properties of Zirconium Metal–Organic Frameworks by Synergistic Incorporation of Nucleophilic and Basic Sites. *ACS Appl. Mater. Interfaces* **2017**, *9* (28), 23967–23973. <https://doi.org/10.1021/acsami.7b06341>.
- (131) Yuan, S.; Zou, L.; Li, H.; Chen, Y.-P.; Qin, J.; Zhang, Q.; Lu, W.; Hall, M. B.; Zhou, H.-C. Flexible Zirconium Metal–Organic Frameworks as Bioinspired Switchable Catalysts.

Angew. Chem. Int. Ed. **2016**, *55* (36), 10776–10780.
<https://doi.org/10.1002/anie.201604313>.

- (132) Valenzano, L.; Civalleri, B.; Chavan, S.; Bordiga, S.; Nilsen, M. H.; Jakobsen, S.; Lillerud, K. P.; Lamberti, C. Disclosing the Complex Structure of UiO-66 Metal Organic Framework: A Synergic Combination of Experiment and Theory. *Chem. Mater.* **2011**, *23* (7), 1700–1718. <https://doi.org/10.1021/cm1022882>.
- (133) Katz, M. J.; Brown, Z. J.; Colón, Y. J.; Siu, P. W.; Scheidt, K. A.; Snurr, R. Q.; Hupp, J. T.; Farha, O. K. A Facile Synthesis of UiO-66, UiO-67 and Their Derivatives. *Chem. Commun.* **2013**, *49* (82), 9449–9451. <https://doi.org/10.1039/C3CC46105J>.
- (134) Ruffley, J. P.; Goodenough, I.; Luo, T.-Y.; Richard, M.; Borguet, E.; Rosi, N. L.; Johnson, J. K. Design, Synthesis, and Characterization of Metal–Organic Frameworks for Enhanced Sorption of Chemical Warfare Agent Simulants. *J. Phys. Chem. C* **2019**, *123* (32), 19748–19758. <https://doi.org/10.1021/acs.jpcc.9b05574>.