# UC Berkeley
## UC Berkeley Electronic Theses and Dissertations

**Title**
Molecular Simulation Techniques for Studying Nanoporous Materials

**Permalink**
https://escholarship.org/uc/item/7w00h3qw

**Author**
Braun, Efrem Daniel

**Publication Date**
2018

Peer reviewed|Thesis/dissertation

# Molecular Simulation Techniques for Studying Nanoporous Materials

by

Efrem Braun

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Chemical Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Berend Smit, Chair
Professor Jeffrey A. Reimer
Professor Phillip L. Geissler

Spring 2018

# Molecular Simulation Techniques for Studying Nanoporous Materials

**Abstract**

Molecular Simulation Techniques for Studying Nanoporous Materials

by

Efrem Braun

Doctor of Philosophy in Chemical Engineering

University of California, Berkeley

Professor Berend Smit, Chair

Broadly speaking, molecular simulations are used for two different purposes. First, zooming out, it allows for high-throughput screening of a much larger chemical space than is feasible by experimental work. Second, zooming in, it serves as a tool similar to a microscope, allowing scientists to understand atomic-level phenomena that underlie chemical properties. These two uses serve a variety of applications, from drug discovery to catalyst development to semiconductor processing. In this work, we illustrate the steps necessary to apply molecular simulations for its two purposes, using nanoporous materials as the example application.

First, we consider the selection of the molecular simulation method itself. Common methods are Monte Carlo and molecular dynamics (MD) simulations. Though the fundamentals of these methods have existed for well over half a century, techniques are still being developed to overcome the small system sizes and timescales to which we are limited. Thermostats are one such technique, frequently used in MD simulations to allow sampling of the canonical ensemble without requiring the large computational expense of simulating a heat bath. Many thermostats have been proposed, and several are in common use. The computational chemist's choice and parameterization of thermostat is not trivial. We have shown that some of the most common thermostats in use today do not sample their intended ensembles, and that this can bring about large errors in a simulation. Fortunately, we have also found that alternative thermostats exist which do not exhibit these errors, and we advise molecular simulation practitioners to use them.

Having selected the molecular simulation technique, a computational chemist that seeks to perform a high-throughput screening must also find a library of materials on which the technique can be performed. Some databases of synthesized nanoporous materials exist, such as the International Zeolite Association (IZA) database and the Computation-Ready Experimental Metal-Organic Framework database, and multiple databases of hypothetical materials are also available. However, if a novel material class is desirable for exploration, the library must be developed. We became interested in zeolite-templated carbons (ZTCs), a material class in which two-dimensional graphene sheets are assembled in a three-dimensional scaffold, but we found no computational library of ZTCs existed. We developed a Monte Carlo

technique which generates a ZTC for a given zeolite template, and we performed this *in silico* synthetic procedure using zeolites taken from the IZA database and the hypothetical zeolite databases. We then found that we could use the mathematical concept of minimal surfaces to describe the ZTCs, and in so doing, we established a link between experimentally-known ZTCs and schwarzites, which had so far been purely hypothetical materials. Schwarzites are negatively-curved carbons, and with their establishment as experimentally-known materials, the triumvirate of two-dimensional nanocarbons (along with positively-curved fullerenes and nanotubes and flat graphene sheets) is completed.

With the simulation method and material library at hand, the computational chemist is ready to perform the high-throughput screening. One important application of nanoporous materials is for adsorptive separations, which can be more energy efficient than distillation. When designing adsorbents for particular separations, understanding how the molecular structure affects gas adsorption is important. We screened tens of thousands of hypothetical zeolites to fill in the gaps that remain in our understanding of the natural gas purification process. Through this screening, we were able to find which adsorbent properties were most correlated with the adsorbent's separation performance. Furthermore, we were able to test the validity of the commonly-used Ideal Adsorbed Solution Theory (IAST) to predict mixture isotherms from pure-component data.

Finally, we examine a particular material in-depth, illustrating the second broad purpose of molecular simulations. Continuing with the theme of using nanoporous materials for separations, we studied the behavior of benzene and xylenes adsorbed in MOF-5, a prototypical metal-organic framework (MOF). We found that the adsorbates separated into liquid and vapor phases that extended over multiple unit cells of MOF-5. This result was surprising because condensation is not generally found in materials with pore size below 2 nm, as the confinement decreases the number of neighbors an adsorbate can interact with, suppressing the energetic benefit of the liquid phase over the more entropically-favorable vapor phase. However, the limiting pore size that had been found in prior studies assumed a one-dimensional capillary-like pore structure, whereas MOF-5 has a three-dimensional pore structure that less restricts the number of neighboring adsorbate molecules. Using NMR, our collaborators were able to find experimental evidence that further attested to our phase separation hypothesis.

# Contents

# Acknowledgments

First and foremost, I am grateful to my adviser, Berend Smit, for having given me the pleasure of working in his laboratory for the past five years. During my first two years, when I was not capable of developing project ideas unaided, Berend handed me two projects on silver platters; during my last three years, Berend challenged me to develop ideas on my own. This progression of burgeoning independence was an appropriate framework for intellectual development, and I thank Berend for having provided me with it. I appreciate his guidance and encouragement, as well as the patience he has given me during my rebellious years of scientific adolescence.

The MolSim group has been a terrific learning environment. The postdocs and older graduate students in the group early on were selfless teachers to whom I owe much. Sondre Schnell deserves particular thanks for his mentorship; without him, I would probably still be trying to figure out why my first simulated isotherm of methane in silicalite didn't match literature results. Bess Vlaisavljevich's thoughts on scientific philosophy were inspirational in guiding my own, and they helped stave off a turn to total cynicism in the face of finding inevitable non-idealities about academic science. Cory Simon was a model of a curiosity-driven scientist for me, from whom I learned much about best practices for scientific programming. It is not possible to give proper acknowledgements to all members of the group from whom I have been privileged to learn and with whom I have formed some deep friendships. Thank you Alex Zurhelle, Amber Mace, Daniele Ongari, Davide Tiana, Gloria Capano, Johanna Huck, Katie Deeg, Lennart Joos, Li-Chiang Lin, Matt Witman, Michelle Liu, Mohamad Moosavi, Nakul Rampal, Özge Kadioglu, Pete Boyd, Rocío Mercado, Sai Sanigepalli, Senja Barthel, Shachi Katira, Sudi Jawahery, Wouter Thijssen, and Yongjin Lee. Some of you deserve extra thanks for having valiantly put up with my humor over the years.

Thanks to my external collaborators, particularly in Jeff Reimer's group. I consider Jeff to be a model researcher who cares more about doing good science than receiving recognitions, and I thank him for his generosity in sharing credit for our joint work. Thank you also Davide Proserpio, Igor Baburin, Jihan Kim, Joseph Chen, Joshua Thompson, Tom Osborn Popp, and Velencia Witherspoon.

No student of molecular simulations could ask to be in better surroundings than Berkeley's Pitzer Center for Theoretical Chemistry. Thanks to the researchers who have made it such an awe-inspiring place. It is no secret that David Chandler is my scientific hero, and he is dearly missed.

Thank you Carlet Altamirano, Kelley McDonald, and Kristin Stangl for having kept the computers running and the red tape at bay. Thanks to my advisers from Penn for having given a chance to my confused undergraduate self. I benefited greatly from a select group of dedicated teachers at the high school level and earlier, who deserve more praise than is possible. Thank you to my friends and to the Jewish community of Berkeley for emotional and spiritual support.

Finally, I must thank my family for reasons that cannot be enumerated. I dedicate this dissertation to my dad, mom, sisters, and aunt, and to the memory of my grandmother.

# Chapter 1

# Introduction

## 1.1 Molecular simulation methods

Computational chemists have at their disposal a seemingly dizzying array of methods for studying chemical matter. At their core, almost all of these methods are designed to circumvent one key problem, which is that the size and timescale of atomic systems that can be simulated on contemporary computers is astoundingly small. From the perspective of a chemist unconcerned with nuclear or relativistic effects, chemistry operates following MD in the microcanonical ensemble. However, on even the most powerful of our supercomputers, 23,558 atom systems can be simulated under these dynamics for only 85 µs per day of real time (using inexpensive empirical force fields at that).[1] Experimental chemists operate on molar quantities, and for computational chemists to contribute to understanding relevant systems, methods to accelerate simulations are necessary.

Molecular simulation methods can be divided into those concerned with accelerating the sampling of relevant sizes and timescales, and those concerned with accelerating the calculation of system energies and forces. We will briefly review these methods here.[2–5]

MD is the primary tool used for the study of system dynamics. By integrating the classical Newtonian equations of motion, MD simulations naturally sample the microcanonical (NVE) ensemble due to conservation laws.[4,6] For comparison with experiment, it is often desirable to sample constant-temperature ensembles such as the canonical (NVT) or isothermal-isobaric (NPT) ensembles. In analogy with experiment, these ensembles could be generated by sampling a subspace of a much larger microcanonical system that serves as a heat and momenta bath, but such an approach is usually too computationally-expensive to implement in practice. Instead, various thermostatting and barostatting algorithms are typically applied to change the dynamics in a manner such that the intended ensemble is sampled. Common thermostatting methods are discussed in detail in Chapter 2.

Monte Carlo methods use the laws of statistical thermodynamics to overcome many limitations of MD. For example, the amount of gas molecules adsorbed by a porous system can be simulated using microcanonical MD using a large particle bath, but the system size

would be required to be enormous to overcome interfacial effects. However, it is known that when two systems are held in chemical equilibrium, both will have the same chemical potential ($\mu$); by attempting particle insertion and deletion moves and accepting these moves in accordance with rules intended to ensure the sampling of the grand-canonical ($\mu$VT) ensemble, Monte Carlo simulations can overcome the limits of MD. We used grand-canonical Monte Carlo simulations to obtain adsorption isotherms in Chapters 4 and 5.

Similarly, when static—rather than dynamic—properties are of interest, Monte Carlo methods may be preferable to MD. For example, obtaining the vapor-liquid phase diagram of a fluid using MD is possible, but it requires the simulation of a large system to avoid interfacial effects, and rare events may require the simulation of unattainably long timescales. Monte Carlo moves can be designed to overcome timescale barriers while sampling the intended ensemble, such as allowing particles to regrow in a new section of the simulation box.

MD and Monte Carlo methods require the calculation of system energies and/or forces. These quantities can be calculated using quantum mechanical methods, but simplifying assumptions are required to obtain them for even a single configuration on reasonable computational timescales.[5] Compromises between accuracy and computational expense are necessary, and for materials science application, density functional theory (DFT) is oftentimes considered the most expensive method that can be used for realistically-complex systems; we demonstrate its use in Chapter 3. For most systems studied here, even DFT was too expensive, and we used simple empirical force fields elsewhere in Chapters 2–5.

## 1.2   Nanoporous materials and their applications

This dissertation is primarily concerned with the application of molecular simulation methods to the study of nanoporous materials. Many categories of nanoporous materials exist, including carbons, zeolites, and MOFs. These materials are widely used for adsorption, catalysis, electrochemistry, and myriad other applications.[7]

To simulate nanoporous materials, a starting crystal structure is necessary, as the use of periodic boundary conditionss (PBCs) implies crystallinity. Fortunately, zeolites and MOFs are highly crystalline materials (though crystal defects can greatly affect their properties), and their crystal structures can be obtained experimentally with X-ray diffraction (XRD). Porous carbons tend to be less crystalline, so efforts are needed to overcome this limitation. Early studies on porous carbons assumed slit pore geometries,[8] and there have been more recent efforts to simulate the carbons as quasi-amorphous materials using representative atomistic models.[9,10] Ordered but non-crystalline mesoporous carbons were developed in the late 1990s by the template carbonization process (more detail on this process is given in Chapter 3), which was soon followed by the development of ordered microporous carbons using the same technique with zeolite templates.[11] Chapter 3 is concerned with developing a computational library of ordered microporous carbons that can be attained via zeolite templating.

As one example of an application of nanoporous materials, we study natural gas purification in depth in Chapter 4. Natural gas, composed largely of methane ($CH_4$), is a large and growing part of the domestic and global energy portfolio, with total domestic production having grown by $35\%$ from 2005 to 2013 and now representing $28\%$ of total American energy consumption.[12] Although this growth occurred for largely economic reasons as a response to the development of shale gas resources,[12] since the carbon dioxide ($CO_2$) emissions per unit of electricity are about half as large from natural gas power plants as from coal power plants, a shift from coal to natural gas may be an effective way to reduce anthropogenic greenhouse gas emissions during the transitory period toward a sustainable energy economy,[13] though the practical achievement of these environmental benefits requires adequate controls of fugitive gas emissions[14,15]. Additionally, since natural gas power plants have lower cycling times and costs than coal power plants, they can adjust more quickly to fluctuations in grid electricity demands:[16] an advantage of increasing importance given the intermittency of many renewable energy sources such as wind and solar power.

Prior to use, the natural gas must be treated to remove contaminants including acid gases such as $CO_2$ and hydrogen sulfide, inert gases such as nitrogen ($N_2$) and helium, and other species such as water and heavier hydrocarbons.[17] $CO_2$ is one of the most common impurities, and in addition to its presence decreasing the natural gas's heating value, it can cause problems during transportation by corroding pipelines or forming solids in cryogenic tanker trucks, necessitating purity specifications like $<2\,\text{mol}\,\%$ $CO_2$ for a natural gas pipeline.[17,18] Nanoporous solids can be used for the separation of $CH_4$ and $CO_2$ via the processes of pressure-swing adsorption (PSA), temperature-swing adsorption (TSA), or a hybrid process (PTSA), in all of which the adsorption columns undergo a cycle of adsorption followed by a regenerating desorption cycle at conditions of decreased pressure and/or increased temperature.

# Chapter 2

# Method development: anomalous effects of velocity rescaling algorithms[*]

The flying ice cube effect is a molecular dynamics simulation artifact in which the use of velocity rescaling thermostats sometimes causes the violation of the equipartition theorem, affecting both structural and dynamic properties. The reason for this artifact and the conditions under which it occurs have not been fully understood. Since the flying ice cube effect was first demonstrated, a new velocity rescaling algorithm (the CSVR thermostat) has been developed and become popular without its effects on the equipartition theorem being truly known. Meanwhile, use of the simple velocity rescaling and Berendsen thermostat algorithms has not abated but has actually continued to grow. Here, we have calculated the partitioning of the kinetic energy between translational, rotational, and vibrational modes in simulations of diatomic molecules to explicitly determine whether the equipartition theorem is violated under different thermostats and while rescaling velocities to different kinetic energy distributions. We have found that the underlying cause of the flying ice cube effect is a violation of balance leading to systematic redistributions of kinetic energy under simple velocity rescaling and the Berendsen thermostat. When velocities are instead rescaled to the canonical ensemble's kinetic energy distribution, as is done with the CSVR thermostat, the equipartition theorem is not violated, and we show that the CSVR thermostat satisfies detailed balance. The critical necessity for molecular dynamics practitioners to abandon the use of popular yet incorrect velocity rescaling algorithms is underscored with an example demonstrating that the main result of a highly-cited study is entirely due to artifacts resulting from the study's use of the Berendsen thermostat.

---

[*]This chapter is based on Braun, Moosavi, and Smit [19].

## 2.1 Introduction

As mentioned in Chapter 1, thermostats are necessary to simulate the canonical ensemble. Many thermostatting algorithms have been proposed, and some of the more well known choices include:

- Simple velocity rescaling, pioneered by Woodcock[20] for thermal equilibration, rescales the velocities of all particles at the end of each timestep (it can also be conducted with a less frequent time rescaling period) by a factor $\lambda$ to achieve a target instantaneous temperature: $\lambda = \left(\frac{K_{\text{target}}}{K}\right)^{\frac{1}{2}}$ with $K_{\text{target}} = \frac{1}{2}N_{\text{DOF}}k_{\text{B}}T_{\text{target}}$, where $N_{\text{DOF}}$ is the number of degrees of freedom in the system.

- The Gaussian thermostat supplements Newton's second law with a force intended to keep the kinetic energy constant:[21–23] $\dot{\mathbf{p}}_i = -\nabla U_i - \alpha\mathbf{p}_i$, where $\alpha$ is a Lagrange multiplier determined using Gauss' principle of least constraint to be $\alpha = \left(\sum_{i=1}^{N}\mathbf{F}_i \cdot \mathbf{p}_i/m_i\right)\Big/\left(\sum_{i=1}^{N}\mathbf{p}_i^2/m_i\right)$.

- Langevin dynamics supplements Newton's second law with terms describing Brownian motion:[24] $\dot{\mathbf{p}}_i = -\nabla U_i - \gamma\mathbf{p}_i + \eta$, where $\gamma$ represents a frictional dissipative force and $\eta(t, T, \gamma, m_i)$ is a stochastic term representing random collisions.

- The Berendsen thermostat takes the Langevin equation, removes the stochastic term, and modifies the frictional dissipative force to yield similar temperature time dependence as with the stochastic term present:[25] $\dot{\mathbf{p}}_i = -\nabla U_i - \gamma\mathbf{p}_i\left(\frac{K_{\text{target}}}{K} - 1\right)$, where $K_{\text{target}} = \frac{1}{2}N_{\text{DOF}}k_{\text{B}}T_{\text{target}}$. In practice, this is implemented as a smoother version of the simple velocity rescaling technique, in which the velocities of all particles are rescaled at the end of each timestep by a factor $\lambda$, with $\lambda = \left[1 + \frac{\Delta t}{\tau_T}\left(\frac{K_{\text{target}}}{K} - 1\right)\right]^{\frac{1}{2}}$. $\tau_T$ represents a time damping constant; if it is set equal to the timestep, the Berendsen algorithm recovers simple velocity rescaling, and as the time damping constant approaches infinity, the Berendsen algorithm recovers conventional microcanonical dynamics.

- The canonical sampling through velocity rescaling (CSVR) thermostat is a velocity rescaling algorithm in which the velocities of all particles are rescaled at the end of each timestep by a factor $\lambda$ designed such that the kinetic energy exhibits the distribution of the canonical ensemble.[26,27] To this end, $\lambda = \left(\frac{K_{\text{target}}}{K}\right)^{\frac{1}{2}}$, where $K_{\text{target}}$ is stochastically drawn from the probability density function $P(K_{\text{target}}) \propto K_{\text{target}}^{N_{\text{DOF}}/2-1}e^{-\beta K_{\text{target}}}$. This algorithm can be adjusted to yield a smoother evolution in a similar manner as the Berendsen algorithm smoothes simple velocity rescaling.[26]

- The Nosé-Hoover thermostat extends the classical Lagrangian to include the additional coordinate $s$ and its time-derivative:[28,29] $\mathcal{L} = s^2\sum_{i=1}^{N}\frac{\mathbf{p}_i^2}{2m_i} - U + \frac{1}{2}Q\dot{s}^2 - k_{\text{B}}T_{\text{target}}L\ln s$,

where $Q$ is the effective "mass" associated with $s$ and $L$ is set by the number of degrees of freedom. A single Nosé-Hoover thermostat may be used, or chains of thermostats may be implemented to improve ergodicity and to take into account additional conservation laws.[30]

There exist numerous additional thermostats (e.g., the Andersen thermostat[31]), and small changes can be made to the listed thermostats, such as implementing the originally global Nosé-Hoover thermostat in a local "massive" manner by pairing a separate Nosé-Hoover thermostat to each degree of freedom.[32] The reader is referred to a non-comprehensive list of reviews and textbooks for additional information.[4,33–35]

Simple velocity rescaling and the Gaussian thermostat aim to sample the isokinetic ensemble (NVK). However, they are often presented as equivalent to the canonical ensemble with respect to position-dependent equilibrium properties, with justification for this based on the argument that the configurational part of the isokinetic ensemble's partition function is exactly equal to that of the canonical ensemble's.[22,36–39] Meanwhile, the Berendsen thermostat does not correspond to a known ensemble but is rather supposed to sample a configurational phase space intermediate to the canonical and microcanonical ensembles.[25,40,41]

In the 1990s, it was found that the simple velocity rescaling and Berendsen thermostat algorithms introduce an artifact:[42,43] the "flying ice cube effect," as coined by Harvey *et al.*[43], describes a violation of the equipartition theorem observed when using these algorithms in which kinetic energy drains from high-frequency modes such as bond stretching into low-frequency modes such as center of mass (COM) translation. This was shown to affect systems' structural, thermodynamic, and dynamic properties.[43] As it can be proven that the equipartition theorem holds in the canonical ensemble, microcanonical ensemble, and isokinetic ensemble (see section 2.5),[44–48] a simulation exhibiting the flying ice cube effect is not ergodically sampling any of these ensembles, neither in configurational phase space nor in momentum phase space.

Nonetheless, simple velocity rescaling and the Berendsen thermostat continue to be commonly used,[34,49] with Cooke and Schmidler[49] stating, "By far the most commonly used algorithm for constant temperature MD of biomolecules is the Berendsen heat bath, due to its ease of implementation and availability in standard software packages." Use of the Berendsen thermostat can be approximated by tracking citations of its canonical reference,[25] which have continued to grow over time (Fig. 2.1).

Some technical aspects of the flying ice cube effect are as of yet still unclear. Since Harvey *et al.*[43], there has been continued discussion about whether the flying ice cube effect may occur with other thermostats.[50,51] The CSVR thermostat rescales velocities to yield the canonical ensemble's distribution of kinetic energies, similar to how simple velocity scaling yields the isokinetic ensemble's distribution of kinetic energies and the Berendsen thermostat yields a kinetic energy distribution intermediate to the two ensembles. If all velocity rescaling algorithms always lead to the flying ice cube effect, then it may be suspected that the same flying ice cube artifact occurs when using the CSVR thermostat,[52] which would be worrisome because the CSVR thermostat has been quickly adopted into widespread use (Fig. 2.1). In
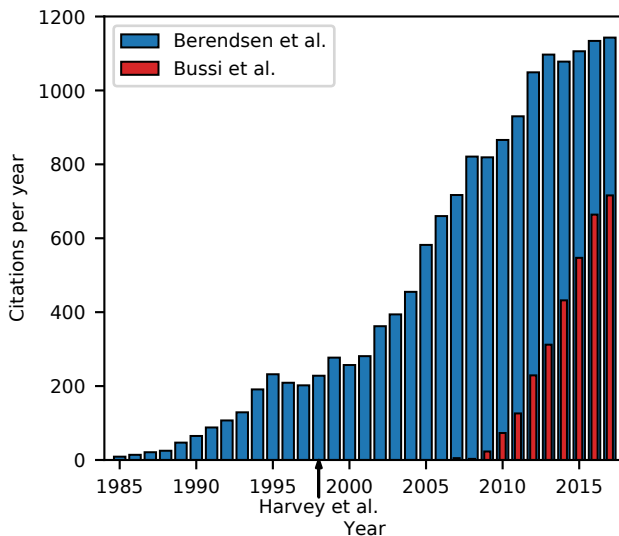
Figure 2.1: Citations of Berendsen *et al.*[25] and Bussi *et al.*[26] over time. Data provided by Web of Science, extracted on January 21, 2018.

addition, since the Gaussian thermostat has been shown to be similar to simple velocity rescaling,[53] it may be suspected that the Gaussian thermostat exhibits the artifact as well. Given the wide-spread use of these algorithms in MD simulations, more understanding is warranted, and we will show that neither the CSVR thermostat nor the Gaussian thermostat bring about the flying ice cube effect.

In the present work we refer to the flying ice cube effect as the term was originally used to describe the violation of the equipartition theorem as caused by velocity rescaling procedures.[43] Other MD simulation methods that fail to conserve energy in the microcanonical ensemble can also bring about equipartition theorem violations.[50] These methods include approximate treatment of long-range electrostatic interactions, certain multiple timestep algorithms, constraining molecular geometries with too loose of a tolerance, not updating neighbor lists frequently enough, and using too large of a timestep.[50,54,55] In some cases these issues are also referred to as flying ice cube effects,[56–58] but these are not related to the artifact with which we are concerned.

In this work, we have revisited the simple model system of united-atom diatomic ethane molecules that Harvey *et al.*[43] first used to illustrate the flying ice cube effect. By explicitly calculating the partitioning of kinetic energies between translational, rotational, and vibrational degrees of freedom, we are able to determine which thermostats and conditions lead to the violation of equipartition, as well as the manner and degree to which they do so. We go on to rationalize these findings by illustrating how simple velocity rescaling violates balance, while the CSVR thermostat satisfies detailed balance. We end by illustrating some severe errors that are directly caused by these subtleties related to thermostatting.

## 2.2   Simulation details

Diatomic ethane molecule simulations were conducted with the open-source LAMMPS code.[59] We used the November 17, 2016 release of LAMMPS to conduct our simulations. The Gaussian thermostat was not implemented in LAMMPS, so we wrote an extension that integrates the equations of motion given by Minary *et al.*[38]. This extension was later incorporated into the LAMMPS code and made publicly available starting with the January 6, 2017 update as part of the "fix nvk" command.

Except where stated otherwise, the simulations consisted of cubic simulation boxes with PBC, setup by placing the ethane molecules on a simple cubic lattice, equilibrated with a Langevin thermostat for at least 50 ns, switched to the target thermostat for at least a further 50 ns of equilibration, and finally ran with the target thermostat for at least 50 ns of production. We verified that all simulations were conducted for sufficient time periods for the energies to equilibrate and be well sampled. The velocities of the particles in microcanonical simulations were rescaled once after Langevin equilibration such that the total energy was equal to the average total energy seen in the Langevin simulation. For the simulations in which the COM linear momentum was fixed to zero (stated in the figure captions), the system's linear momentum was zeroed every timestep, followed by a rescaling of velocities to maintain the same total kinetic energy as before the zeroing had occurred to prevent energy leakage. The equations of motion were integrated with a standard Velocity Verlet algorithm using half-step velocity calculations. The timestep used was 0.5 fs, which was found to give adequate energy conservation in the microcanonical ensemble.

Thermostat parameters were as follows, except where stated otherwise. Simple velocity rescaling was done every timestep. The Nosé-Hoover chain consisted of three thermostats. The Berendsen, Nosé-Hoover, and CSVR thermostats were used with time damping constants ($\tau_T$) of 100 fs, and the Nosé-Hoover thermostat used effective thermostat masses of $Q_1 = N_{\mathrm{DOF}} k_{\mathrm{B}} T \tau_T{}^2$ and $Q_{i>1} = k_{\mathrm{B}} T \tau_T{}^2$.[30] When doing simulations in the microcanonical ensemble, the total energy was set such that a simulation temperature equal to the canonical ensemble simulations' target temperature was achieved. The target simulation temperature was set to 350 K, well above the critical temperature of ethane.[60]

Kinetic energies of each diatomic molecule were partitioned into translational, rotational, and vibrational kinetic energies. In one dimension, this was done as:

$$
\begin{aligned}
K &= \frac{1}{2} m_1 v_{1,x}^2 + \frac{1}{2} m_2 v_{2,x}^2 \\
&= \frac{1}{2} (m_1 + m_2) \left( \frac{m_1 v_{1,x} + m_2 v_{2,x}}{m_1 + m_2} \right)^2 + \frac{1}{2} \left( \frac{m_1 m_2}{m_1 + m_2} \right) (v_{2,x} - v_{1,x})^2 \\
&= \underbrace{\frac{1}{2} (m_1 + m_2) v_{\mathrm{trans},x}^2}_{K_{\mathrm{trans}}} + \underbrace{\frac{1}{2} \mu v_{\mathrm{vib},x}^2}_{K_{\mathrm{vib}}}
\end{aligned}
$$

where $v_{\mathrm{trans},i} = \frac{m_1 v_{1,i} + m_2 v_{2,i}}{m_1 + m_2}$, $v_{\mathrm{vib},i} = v_{2,i} - v_{1,i}$, and $\mu = \frac{m_1 m_2}{m_1 + m_2}$, giving one translational and one vibrational degrees of freedom for the molecule. In three dimensions, this was similarly

done as:

$$
\begin{aligned}
K &= \frac{1}{2}m_1 \left( v_{1,x}^2 + v_{1,y}^2 + v_{1,z}^2 \right) + \frac{1}{2}m_2 \left( v_{2,x}^2 + v_{2,y}^2 + v_{2,z}^2 \right) \\
&= \underbrace{\frac{1}{2} \left( m_1 + m_2 \right) \left( v_{\text{trans},x}^2 + v_{\text{trans},y}^2 + v_{\text{trans},z}^2 \right)}_{K_{\text{trans}}} + \underbrace{\frac{1}{2}\mu \left( \left( v_{2,x} - v_{1,x} \right)^2 + \left( v_{2,y} - v_{1,y} \right)^2 \right)}_{K_{\text{rot}}} \\
&\quad + \underbrace{\frac{1}{2}\mu \left( v_{2,z} - v_{1,z} \right)^2}_{K_{\text{vib}}}
\end{aligned}
$$

with an arbitrary coordinate axis aligned with the bond vector (we chose to label the equation above such that the $z$-axis was aligned), giving three translational, two rotational, and one vibrational degrees of freedom for the molecule. The code to calculate these partitioned energies was incorporated into the open-source LAMMPS code and made publicly available starting with the September 13, 2016 update as part of the "compute bond/local" command.

In all figures that plot kinetic energies, the error bars shown represent $\pm 1$ standard error of the mean. This was calculated by dividing the data from the production timesteps into 20 consecutive blocks, averaging the data for each block, and computing the standard error over the 20 data values.[4] Error bars are not shown when they would be smaller than the symbols or the line widths.

Bonded parameters for the united-atom ethane molecule were taken from Harvey *et al.*[43] (harmonic bond potential $U(r) = k(r - r_0)^2$ with $r_0 = 1.54$ Å and $k = 240$ kcal mol$^{-1}$ Å$^{-2}$) and non-bonded parameters were taken from Martin and Siepmann[60] (Lennard-Jones potential with $\epsilon = 0.195$ kcal mol$^{-1}$, $\sigma = 3.75$ Å, truncated and shifted at 14 Å, and no charges).

Simulations of benzene in MOF-5 were conducted with the Tinker package,[61] version 7.1, for the purposes of using the force field of Tafipolsky *et al.*[62] to compare results with Amirjalayer *et al.*[63]. Tinker input scripts are available in the Supplementary Information†. The force field of Tafipolsky *et al.*[62] was used with the modification of using point charges instead of bond dipoles since—to the best of our knowledge—computing Ewald summations with the latter is not implemented in Tinker (it is unclear how Ewald summations were computed in Tafipolsky *et al.*[62] and Amirjalayer *et al.*[63]); we used the atomic charges that Tafipolsky *et al.*[62] used to parameterize their force field, as given in their Table 2. We strived to keep conditions as similar to those used by Amirjalayer *et al.*[63] as possible; we used a timestep of 1 fs, a Lennard-Jones potential cutoff of 12 Å, Ewald summations with default Tinker 7.1 parameters, the formerly default Berendsen thermostat time damping constant of 100 fs (except for where we noted that we used the currently default time damping constant of 200 fs), and Nosé-Hoover default Tinker parameters. Simulations were run for at least 2 ns of equilibration and at least 100 ns of production, which was found to be sufficiently long for the mean-squared displacement (MSD) to become a linear function of time. Simulations were conducted in a simulation box consisting of a single cubic unit cell taken from a minimized structure described in Tafipolsky *et al.*[62] and kindly supplied to us by Rochus Schmid, consisting of 424 atoms and with a lattice constant of 25.9457 Å. The 10 benzene molecules

were distributed through the MOF-5 crystal by running a 200 ps MD simulation at 1,000 K with the MOF atoms frozen prior to equilibration.

## 2.3   Results and discussion

### Examining equipartition under different thermostats

It is instructive to reconsider the simple case previously examined by Harvey *et al.*[43]: that of a single ethane molecule moving in one-dimensional space along its bond axis. In the microcanonical ensemble under perfect energy conservation, the translational kinetic energy will remain constant at its set initial energy and the vibrational kinetic energy will oscillate. In the canonical ensemble, equipartition states that the translational and vibrational degree of freedom should each have an average kinetic energy of $\frac{1}{2}k_{\mathrm{B}}T$. As expected, the Langevin thermostat satisfies the equipartition theorem (see Fig. 2.2). In agreement with the work of Harvey *et al.*[43], we find that simple velocity rescaling and the Berendsen thermostat bring about a violation of equipartition in the kinetic degrees of freedom, with all kinetic energy flowing to translational motion, in the plainest illustration of the flying ice cube effect. We find that the CSVR thermostat correctly partitions the energies.

We next consider the more complex case of a large number of ethane molecules interacting in three dimensions with anharmonic Lennard-Jones potentials. Each diatomic ethane molecule now has three translational modes, two rotational modes, and one vibrational modes, so the equipartition theorem states that these modes' kinetic energies should be equal to $\frac{3}{2}k_{\mathrm{B}}T$, $\frac{2}{2}k_{\mathrm{B}}T$, and $\frac{1}{2}k_{\mathrm{B}}T$ respectively, with a correction of $\frac{3}{2}k_{\mathrm{B}}T/N_{\mathrm{molecs}}$ to the translational kinetic energy in cases where the COM momentum is constrained. In Fig. 2.3, we show that the Langevin, Nosé-Hoover, CSVR, and Gaussian thermostats all exhibit correctly equipartitioned energies, as does the microcanonical ensemble. As in the case of the single ethane molecule in one dimension, the simple velocity rescaling and Berendsen thermostat algorithms lead to a violation of equipartition, with translational and rotational modes having too much kinetic energy and vibrational modes having too little.

### Equivalence of simple velocity rescaling and the Gaussian thermostat

Since the thermostatting under simple velocity rescaling does not take place within the equations of motion, this ad hoc temperature control algorithm was initially difficult to investigate theoretically, and its validity was considered questionable.[22,23] The algorithm's use was justified on the basis of empirical arguments, such as that simple velocity rescaling and the Gaussian thermostat give similar static and dynamic properties for the Lennard-Jones fluid.[36] It was eventually proven that simple velocity rescaling is analytically equivalent to the Gaussian thermostat within an error of $\mathcal{O}$ (timestep) when the velocity rescaling time
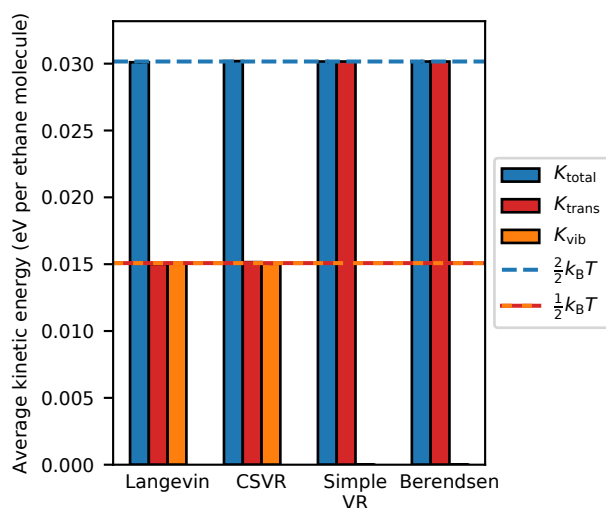
Figure 2.2:   Partitioning of the kinetic energies obtained from one-dimensional MD simulations of a single ethane molecule using various thermostats.  Both atoms were given a starting velocity of $100\,\mathrm{m\,s^{-1}}$ along the same direction as the bond vector.  For the thermostats shown, the same energy partitionings were observed regardless of initial bond length and initial COM momentum.  The microcanonical, Nosé-Hoover thermostat, and Gaussian thermostat results are not shown here since we found that the energy partitionings are dependent on the initial conditions, indicative of these thermostats' well-known lacks of ergodicity that are more manifest for small systems.[4,22,30,34,53,64–66]
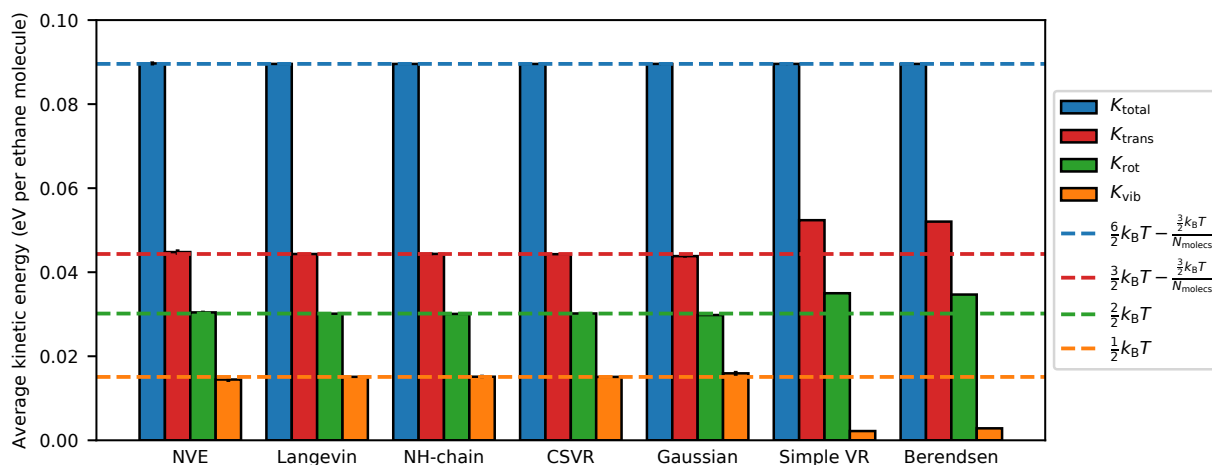


Figure 2.3:   Partitioning of the kinetic energies obtained from MD simulations of 50 ethane molecules in a $30\,\text{Å}$ cubic simulation box using various thermostats.  In all simulations shown, the COM momentum was fixed to zero.
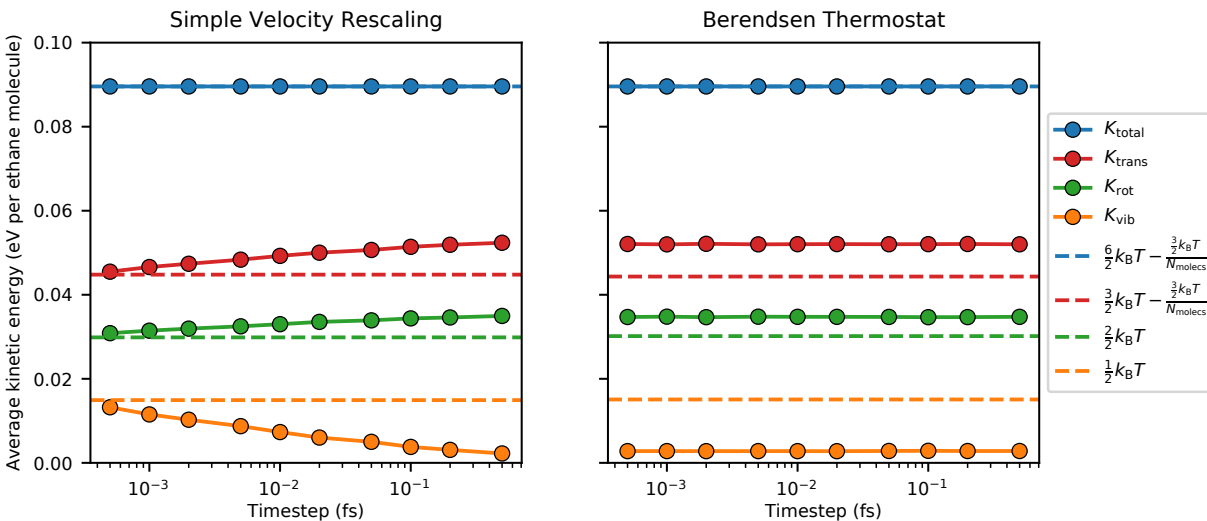
Figure 2.4: Partitioning of the kinetic energies obtained from MD simulations performed under the same conditions as in Fig. 2.3 but changing the timestep, using (left) simple velocity rescaling and (right) the Berendsen thermostat with the time damping constant maintained at 100 fs. Lines are a guide to the eye.

period is set equal to the timestep,[53] which gave support for the legitimacy of using simple velocity rescaling to sample the isokinetic ensemble.

However, we have shown that the Gaussian thermostat exhibits correct energy equipartitioning while simple velocity rescaling does not. We prove in section 2.5 that the isokinetic ensemble should satisfy the equipartition theorem. Thus, it is clear that simple velocity rescaling does not actually sample the isokinetic ensemble.

The equivalence of simple velocity rescaling and the Gaussian thermostat under small timesteps leads to the expectation that the flying ice cube effect will gradually disappear under simple velocity rescaling as the timestep is decreased. We demonstrate confirmation of this expectation in Fig. 2.4. However, Fig. 2.4 shows that the timestep needs to be reduced by over three orders of magnitude from typical simulation timesteps before the flying ice cube effect is no longer discerned. Thus, although the argument that simple velocity rescaling can be a legitimate method for sampling the isokinetic ensemble is true, this argument is not a proper justification for the current usage of the method. We also note that no dependence of energy partitioning on the timestep can be observed in Fig. 2.4 when using the Berendsen thermostat, likely indicating the presence of an additional error term introduced by the time damping.
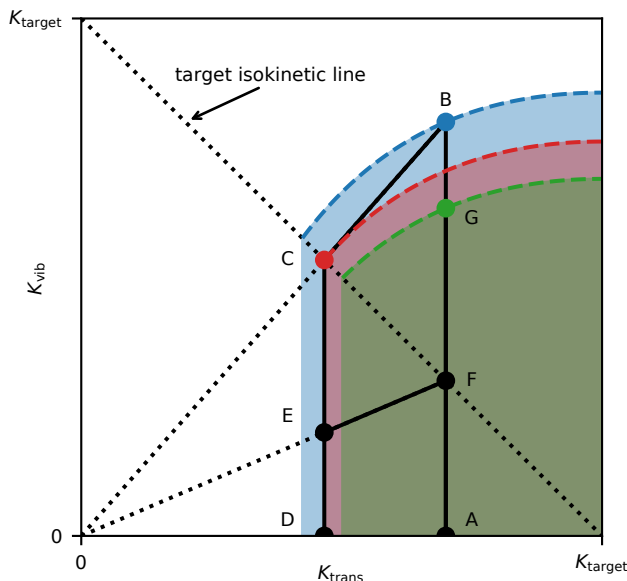
Figure 2.5: Kinetic phase space of a single ethane molecule moving in one-dimensional space along its bond axis under simple velocity rescaling. $K_{\text{target}} = k_{\text{B}}T_{\text{target}}$, $K_{\text{trans}} = \frac{1}{2}(m_1 + m_2)\left(\frac{m_1 v_{1,x} + m_2 v_{2,x}}{m_1 + m_2}\right)^2$, and $K_{\text{vib}} = \frac{1}{2}\left(\frac{m_1 m_2}{m_1 + m_2}\right)(v_{2,x} - v_{1,x})^2$. Solid lines show a particular path in phase space between labeled points, referred to in the text. Dotted lines are guides useful to understanding the velocity rescaling moves. Dashed lines show the boundaries of phase space accessible by any sequence of MD and velocity rescalings from lines $\overline{AB}$, $\overline{CD}$, and $\overline{AG}$, with the accessible phase spaces shaded.

## Violation of balance causes the flying ice cube effect

The mechanism underlying the flying ice cube effect can be elucidated graphically for the first test case we examined, that of a single ethane molecule. In Fig. 2.5, we show this system's phase space, putting translational kinetic energy on the $x$-axis and vibrational kinetic energy on the $y$-axis.

During microcanonical MD, the system can only explore phase space on a vertical line between $y = 0$ and $y = U_{\text{max}}$ because a constant total energy and translational kinetic energy is maintained, with energy exchanges only allowed between vibrational kinetic energy and potential energy. Consider a MD simulation initially on such a vertical line in phase space, $\overline{AB}$. Under simple velocity rescaling, if a rescaling move is conducted at point $B$, the system will move to point $C$; this occurs because the translational and vibrational energies are both scaled by the same factor $\lambda^2$ such that their sum is equal to the target kinetic energy, moving the system to the intersection of the lines $y = \frac{y_B}{x_B}x$ and the target isokinetic line ($y = -x + K_{\text{target}}$). Since points $B$ and $C$ have the same configuration with zero potential energy, MD will now explore line $\overline{CD}$.

Let us examine whether we can reach point $B$ by rescaling from line $\overline{CD}$ back to a line with the same translational energy of line $\overline{AB}$. With a single rescaling, we would need to rescale from point $E$ to point $F$. From point $F$, MD will explore phase space on line $\overline{AG}$, where the lengths of lines $\overline{FG}$ and $\overline{CE}$ are equal, with both representing the stored potential energy of the system prior to the rescaling. Obviously, line $\overline{EF}$ must have a smaller slope than line $\overline{BC}$; accordingly, $y_G$ will necessarily be smaller than $y_B$. Hence, with a single velocity rescaling, point $B$ cannot be reached. Multiple velocity rescalings from line $\overline{CD}$ allows us to reach a point with greater vibrational kinetic energy than point $G$. However, all phase space reachable by any number of velocity rescalings from line $\overline{CD}$ is bounded by the red dashed line in Fig. 2.5, which can be derived as follows.

Consider a MD simulation initially on line $\overline{CD}$ in phase space (Fig. 2.5). By being infinitesimally close to point $C$ when rescaling, the rescaling line will have a slope of $\frac{K_{\text{target}} - x_C}{x_C}$. To remain moving in phase space in the direction of increasing translational kinetic energy, rescaling must continue to occur below the target isokinetic line; thus, the greatest slope that can continue to be achieved is $\frac{K_{\text{target}} - x}{x}$. We see that the red dashed line can therefore be derived by solving the differential equation $\frac{dy_{\max}}{dx} = \frac{K_{\text{target}} - x}{x}$ with boundary condition $y_{\max}(x_C) = y_C = K_{\text{target}} - x_C$, which results in $y_{\max}(x) = K_{\text{target}} \ln\left(\frac{x}{x_C}\right) - x + K_{\text{target}}$. A similar situation occurs when moving in phase space in the direction of decreasing translational kinetic energy. The velocity rescaling line with smallest slope is achieved when rescaling just above the target isokinetic line, e.g., just above point $F$ when rescaling from line $\overline{AB}$. The same differential equation must be solved to derive the blue dashed line, only changing the boundary condition to $y_{\max}(x_B) = y_B$, which results in $y_{\max}(x) = K_{\text{target}} \ln\left(\frac{x}{x_B}\right) - x + x_B + y_B$. By replacing $x_B$ and $y_B$ in this equation with the beginning positions $x_0$ and $y_0 = U_{\max}(x_0)$, one obtains the general expression $y_{\max}(x) = K_{\text{target}} \ln\left(\frac{x}{x_0}\right) - x + x_0 + U_{\max}(x_0)$.

Hence, continuing to rescale will continue to shrink the volume of accessible phase space, as rescaling from lines $\overline{AB}$ to $\overline{CD}$ to $\overline{AG}$ lowers the boundary from the blue to the red to the green dashed lines; eventually, accessible phase space will be confined only to the point with all kinetic energy in the translational mode.

Notably, the decrease in accessible phase space becomes smaller as velocity rescaling occurs closer to the isokinetic line. In a simulation, this occurs when the timestep between velocity rescalings is reduced. This explains why the flying ice cube effect is reduced under simple velocity rescaling by decreasing the timestep (Fig. 2.4).

## Monte Carlo perspective

We can view the combination of MD and velocity scaling moves as a Monte Carlo simulation. Hence, our previous example shows that simple velocity rescaling violates the condition of balance.[4,67]

In contrast, the CSVR thermostat can explicitly be proven to sample the desired distribution by considering the condition of detailed balance. Let us assume that we do a large

and random number of MD steps between velocity rescaling moves. We define $A$ as the set of all configurations of the system with a total energy $E_A$. The flow of configurations from set $A$ to set $B$ is given by:

$$K(A \rightarrow B) =$$
$$P(E_A) \sum_{\mathbf{r}_1^n} \sum_{\mathbf{p}_1^n} \sum_{\mathbf{r}_2^n} \sum_{\mathbf{p}_2^n} p(\mathbf{r}_1^n, \mathbf{p}_1^n | E_A) \delta(E(\mathbf{r}_1^n, \mathbf{p}_1^n) - E_A) \alpha(\mathbf{r}_1^n, \mathbf{p}_1^n \rightarrow \mathbf{r}_2^n, \mathbf{p}_2^n) \delta(E(\mathbf{r}_2^n, \mathbf{p}_2^n) - E_B)$$
$$(2.1)$$

where $\mathbf{r}_1^n, \mathbf{p}_1^n$ is the configuration with position vector $\mathbf{r}_1^n$ and momentum vector $\mathbf{p}_1^n$, $p(\mathbf{r}_1^n, \mathbf{p}_1^n | E_A)$ is the probability to find the configuration $\mathbf{r}_1^n, \mathbf{p}_1^n$ from all configurations with energy $E_A$ during MD, and $\alpha(\mathbf{r}_1^n, \mathbf{p}_1^n \rightarrow \mathbf{r}_2^n, \mathbf{p}_2^n)$ is the *a priori* probability to velocity rescale from configuration $\mathbf{r}_1^n, \mathbf{p}_1^n$ to configuration $\mathbf{r}_2^n, \mathbf{p}_2^n$. Recognizing that velocity rescaling does not alter positions:

$$K(A \rightarrow B) =$$
$$P(E_A) \sum_{\mathbf{r}^n} \sum_{\mathbf{p}_1^n} \sum_{\mathbf{p}_2^n} p(\mathbf{r}^n, \mathbf{p}_1^n | E_A) \delta(E(\mathbf{r}_1^n, \mathbf{p}_1^n) - E_A) \alpha(\mathbf{r}^n, \mathbf{p}_1^n \rightarrow \mathbf{r}^n, \mathbf{p}_2^n) \delta(E(\mathbf{r}^n, \mathbf{p}_2^n) - E_B)$$
$$(2.2)$$

Next, recognizing that velocity rescaling can only give one configuration in momentum space with $E(\mathbf{r}^n, \mathbf{p}_2^n) = E_B$ from starting configuration $\mathbf{r}^n, \mathbf{p}_1^n$, and that the acceptance probabilities only involve the kinetic energy:

$$K(A \rightarrow B) =$$
$$P(E_A) \sum_{\mathbf{r}^n} \sum_{\mathbf{p}^n} p(\mathbf{r}^n, \mathbf{p}^n | E_A) \delta(E(\mathbf{r}_1^n, \mathbf{p}_1^n) - E_A) \alpha(K = E_A - U(\mathbf{r}^n) \rightarrow E_B - U(\mathbf{r}^n)) \quad (2.3)$$

where $\alpha(K = E_A - U(\mathbf{r}^n) \rightarrow E_B - U(\mathbf{r}^n))$ is the *a priori* probability to velocity rescale to the configuration having kinetic energy $K = E_B - U(\mathbf{r}^n)$ given we start with a configuration having kinetic energy $K = E_A - U(\mathbf{r}^n)$. Then, recognizing that momentum and position are decoupled, i.e., the number of possible states in momentum space only depends on the total kinetic energy but does not depend on the details of the potential energy surface, and each of these possible states in momentum space are equally likely:

$$K(A \rightarrow B) = P(E_A) \sum_{\mathbf{r}^n} \omega(E_A - U(\mathbf{r}^n)) p(\mathbf{r}^n, \mathbf{p}^n | E_A) \alpha(K = E_A - U(\mathbf{r}^n) \rightarrow E_B - U(\mathbf{r}^n))$$
$$(2.4)$$

where $\omega(K)$ is the number of configurations in momentum space for a given kinetic energy $K$ (equivalent to the ideal gas microcanonical partition function). Finally, by making the substitutions $p(\mathbf{r}^n, \mathbf{p}^n | E_A) = \Omega_{NVE_A}^{-1}$ and $P(E_A) = \frac{e^{-\beta E_A} \Omega_{NVE_A}}{Z_{NVT}}$:

$$K(A \rightarrow B) = \frac{e^{-\beta E_A}}{Z_{NVT}} \sum_{\mathbf{r}^n} \omega(E_A - U(\mathbf{r}^n)) \alpha(K = E_A - U(\mathbf{r}^n) \rightarrow E_B - U(\mathbf{r}^n)) \quad (2.5)$$

The two flows, $K(A \to B)$ and $K(B \to A)$, are equal if we impose as condition for the *a priori* probabilities:

$$
\frac{\alpha\left(K = E_A - U\left(\mathbf{r}^n\right) \to E_B - U\left(\mathbf{r}^n\right)\right)}{\alpha\left(K = E_B - U\left(\mathbf{r}^n\right) \to E_A - U\left(\mathbf{r}^n\right)\right)} = \frac{e^{-\beta E_B}\omega\left(E_B - U\left(\mathbf{r}^n\right)\right)}{e^{-\beta E_A}\omega\left(E_A - U\left(\mathbf{r}^n\right)\right)}
$$
$$
= \frac{e^{-\beta(E_B - U(\mathbf{r}^n))}\left(E_B - U\left(\mathbf{r}^n\right)\right)^{N_{\mathrm{DOF}}/2-1}}{e^{-\beta(E_A - U(\mathbf{r}^n))}\left(E_A - U\left(\mathbf{r}^n\right)\right)^{N_{\mathrm{DOF}}/2-1}} \tag{2.6}
$$

in which we used the known expression for the ideal gas microcanonical partition function.[35] Eq. 2.6 is satisfied by the CSVR thermostat, which rescales velocities to the target kinetic energy distribution given by the gamma distribution:

$$
P(K) = \frac{e^{-\beta K}K^{N_{\mathrm{DOF}}/2-1}}{\int_0^\infty dK\, K^{N_{\mathrm{DOF}}/2-1}e^{-\beta K}} = \frac{e^{-\beta K}K^{N_{\mathrm{DOF}}/2-1}}{\beta^{-N_{\mathrm{DOF}}/2}\Gamma\left(N_{\mathrm{DOF}}/2\right)} \tag{2.7}
$$

Hence, the CSVR thermostat satisfies detailed balance.

**Velocity rescaling to other kinetic energy distributions**

We have seen that simple velocity rescaling violates balance and brings about the flying ice cube effect, while the CSVR thermostat satisfies detailed balance and does not exhibit the artifact. One key difference between these algorithms is that simple velocity rescaling restricts the rescaling factor ($\lambda$) to be less than one when the system's instantaneous temperature is greater than the target temperature and greater than one when the instantaneous temperature is less than the target temperature. It is this restriction which allowed us to show graphically that simple velocity rescaling moves decrease accessible phase space. It is instructive to consider the effects of relaxing this restriction while rescaling velocities to a non-canonical kinetic energy distribution. This procedure would not render any areas of phase space inaccessible, but the rescaling would be to a distribution that is not necessarily invariant under Hamiltonian dynamics.[31,67]

To change the target kinetic energy distribution, we modified the CSVR thermostat's value of $N_{\mathrm{DOF}}$ in Eq. 2.7 from the actual number of degrees of freedom ($N_{\mathrm{DOF},0}$) while simultaneously adjusting $\beta$ from its initial value ($\beta_0$) such that $\beta = \beta_0 \frac{N_{\mathrm{DOF}}}{N_{\mathrm{DOF},0}}$ in order to maintain a constant average kinetic energy. The resulting kinetic energy distributions are shown in the top of Fig. 2.6 and include distributions that are sharper ($N_{\mathrm{DOF}} > N_{\mathrm{DOF},0}$) and broader ($N_{\mathrm{DOF}} < N_{\mathrm{DOF},0}$) than the canonical distribution. In the limit of $N_{\mathrm{DOF}} \to \infty$, this method closely approximates simple velocity rescaling or the Berendsen thermostat, depending on the time damping constant used.

The energy partitionings that resulted from setting these target kinetic energy distributions are shown for simulations in the bottom of Fig. 2.6. It can be seen that with sharper distributions, the flying ice cube effect is observed, with more kinetic energy partitioned in low-frequency modes and less in high-frequency modes. Interestingly, the opposite effect is
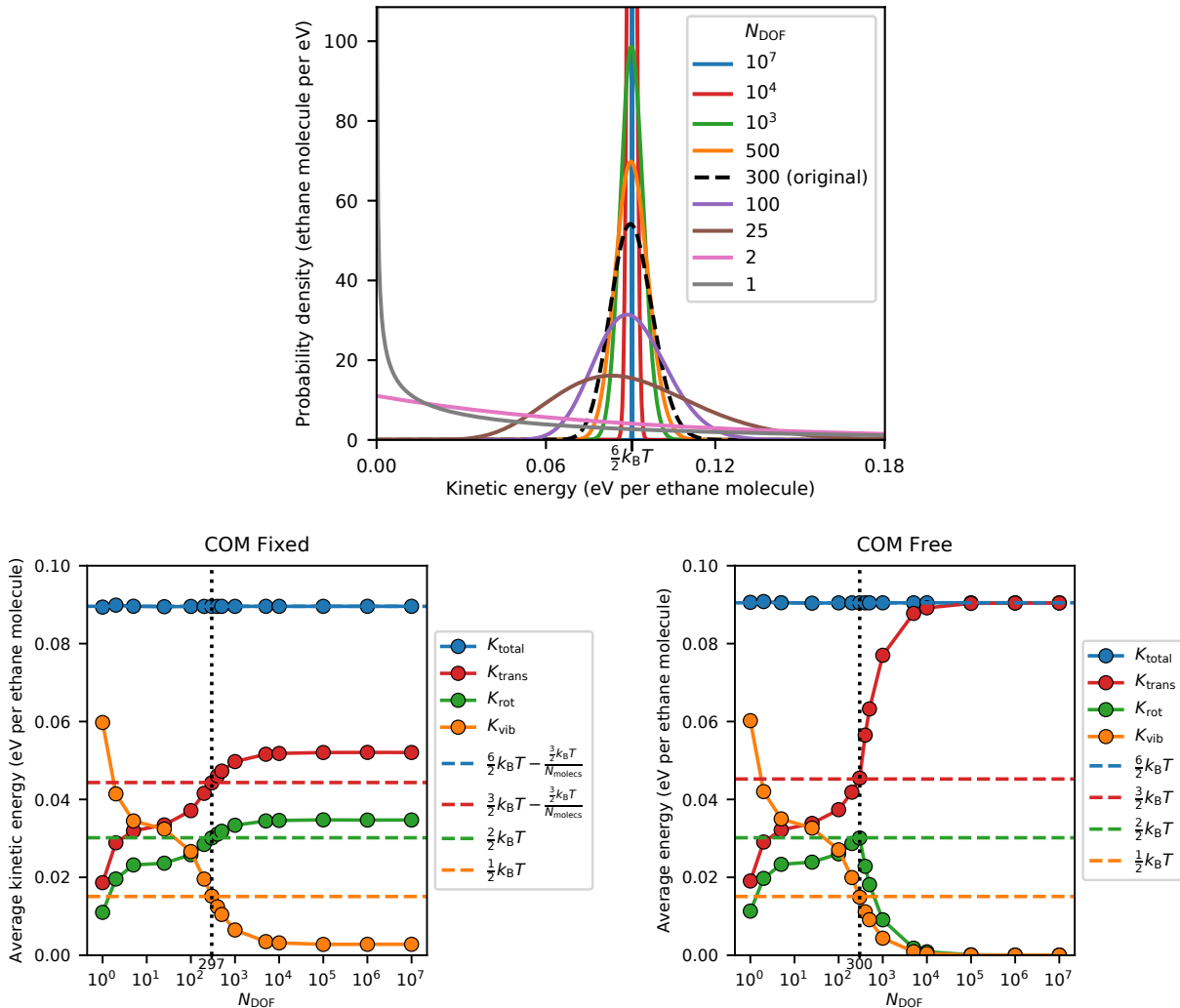
Figure 2.6: (top) Probability density function of kinetic energies following $P(K) = \frac{e^{-\beta K} K^{N_{\mathrm{DOF}}/2-1}}{\beta^{-N_{\mathrm{DOF}}/2} \Gamma(N_{\mathrm{DOF}}/2)}$, where $\beta$ is chosen such that the average kinetic energy (temperature) is the same for all choices of $N_{\mathrm{DOF}}$ via $\beta = \beta_0 \frac{N_{\mathrm{DOF}}}{N_{\mathrm{DOF},0}}$, $N_{\mathrm{DOF},0} = 300$, and $\beta_0 = (k_{\mathrm{B}} \times 350\,\mathrm{K})^{-1}$. (bottom) Partitioning of the kinetic energies obtained from MD simulations of 50 ethane molecules in a $30\,\text{Å}$ cubic simulation box using the CSVR thermostat, modified such that the target distribution of kinetic energies was set to those shown in the top part of the figure for the proper $N_{\mathrm{DOF},0}$ value. (bottom left) Here, the COM momentum was fixed at zero and $N_{\mathrm{DOF},0}$ was set to 297. (bottom right) Here, the COM momentum was not fixed after the Langevin thermostat equilibration, allowing the COM momentum to drift, and $N_{\mathrm{DOF},0}$ was set to 300. Lines are a guide to the eye.

observed with broader distributions, with more kinetic energy partitioned in high-frequency modes and less in low-frequency modes. When the COM momentum is not constrained to zero, a more drastic effect is observed, such that rotational kinetic energy decreases both with decreasing $N_{\mathrm{DOF}}$ as energy flows to the higher-frequency vibrational modes and with increasing $N_{\mathrm{DOF}}$ as almost all energy flows to the lower-frequency translational modes. Only at the canonical kinetic energy distribution ($N_{\mathrm{DOF}} = 297$ and $N_{\mathrm{DOF}} = 300$ for the constrained and not-constrained COM momentum simulations, respectively) is proper equipartitioning observed.

## Conditions affecting the flying ice cube effect's conspicuousness

Artifacts relating to the flying ice cube effect do not always appear when the simple velocity rescaling or Berendsen thermostat algorithms are used.[52,68,69] Indeed, when the flying ice effect was first found,[42,43] fewer alternatives to these thermostatting algorithms were available than at present, e.g., the CSVR thermostat had not yet come into popular use, and so protective measures were recommended to lower the likelihood of the artifact occurring under these faulty thermostats.[43] Here, we investigate these recommendations and other conditions which we found affect the conspicuousness of the flying ice cube effect for our system of interacting diatomic ethane molecules.

One recommendation given in Harvey *et al.*[43] was to lower the thermostat's coupling strength, either by less frequent rescaling under simple velocity rescaling or by increasing the time damping constant under the Berendsen thermostat. Decreasing the coupling strength allows for the system's natural dynamics to bring about energy equipartitioning faster than the thermostat can disturb it. In Fig. 2.7, we show that this recommendation does indeed reduce the violation of equipartition. However, the flying ice cube artifact was not fully resolved until these time parameters were larger than 100 ps, a value much greater than the 0.5 ps time damping constant above which Berendsen *et al.*[25] showed that energy fluctuations under the Berendsen thermostat are similar to energy fluctuations in the microcanonical ensemble and thus concluded that the thermostat has little influence on the dynamics. This discrepancy may be partially explained by the use of the rigid SPC water model[70] to evaluate the Berendsen thermostat in Berendsen *et al.*[25], as a rigid molecule lacks the high-frequency vibrational modes that lead most directly to the flying ice cube effect. Meanwhile, we found that energy equipartitioning held under the CSVR thermostat regardless of the value of the time damping constant. At the weakest coupling strengths shown in Fig. 2.7, it can be seen that the desired temperature was not well established in these 100 ns simulations.

Varying the coupling strength does not come without its risks. Fig. 2.7 shows an anomalous data point when simple velocity rescaling is performed every 500 fs. Further investigation allowed us to characterize this anomaly as a resonance effect associated with bond vibration. The characteristic period of the $\mathrm{CH_3-CH_3}$ harmonic bond is 38.4 fs. When the time rescaling period is set close to an integer multiple of half this characteristic period, large amplitude bond vibrations occur, becoming stronger when the time rescaling period more exactly matches the multiple. These resonance effects become weaker as the multiple grows,
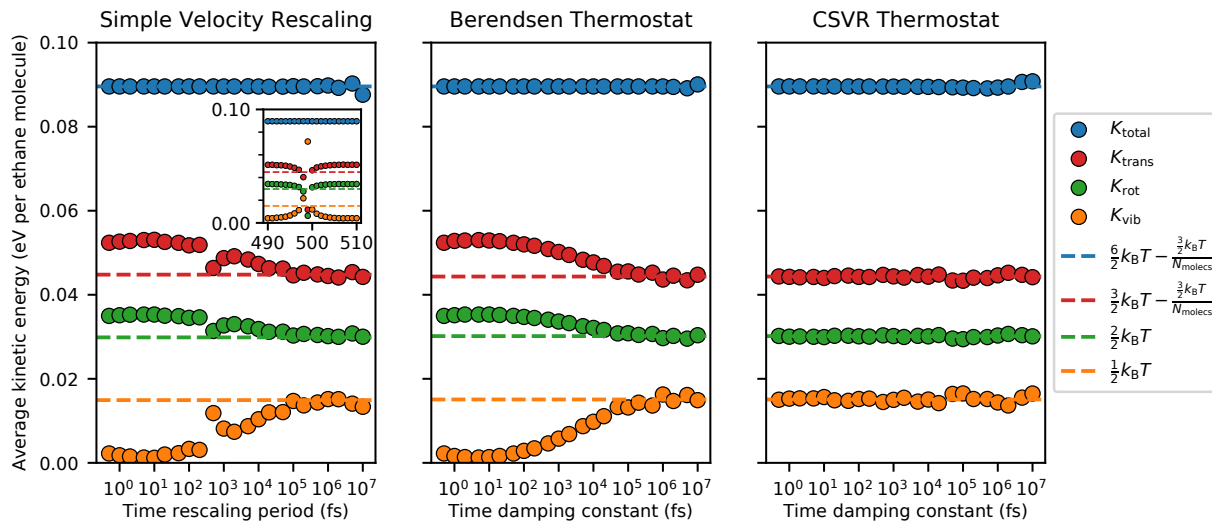
Figure 2.7:    Partitioning of the kinetic energies obtained from MD simulations performed under the same conditions as in Fig. 2.3, but changing (left) the time rescaling period for simple velocity rescaling, (middle) the time damping constant for the Berendsen thermostat, and (right) the time damping constant for the CSVR thermostat, all three with the timestep maintained at 0.5 fs. The inset shown in the simple velocity rescaling graph shows additional data near the time rescaling period of 500 fs, at which point a resonance artifact associated with the $CH_3-CH_3$ bond's characteristic vibrational frequency can be observed.

which explains why the vibrational energy at the time rescaling period of 1,000 fs is greater than at 2,000 fs. We observed resonance effects when rescaling close to other multiples of half the bond's characteristic period that we also tested. We will shortly show that altering the coupling strength can bring about resonance effects under the Berendsen thermostat as well.

Another precautionary measure recommended in Harvey *et al.*[43] was to periodically zero the COM momentum, as it represents the lowest-frequency degree of freedom into which most kinetic energy flows. The Newtonian equations of motion preserve COM momentum, but numeric errors cause this preservation to be inexact. Constraint of the COM momentum to zero is oftentimes used to safeguard against these numeric errors: a safeguard we used throughput this paper except where stated. In Fig. 2.8, we show that releasing this constraint does indeed significantly worsen the flying ice cube effect, though equipartition is violated both with and without the constraint. We further explored the effects of allowing the COM momentum to vary by replacing the PBC with reflecting walls, which we found gets rid of the flying ice cube effect completely, with no violation of the equipartition theorem. In both of these cases, COM momentum is not conserved, but with opposite results observed (though in the former case, COM momentum can build-up, while in the latter case, it cannot), We hypothesize that reflecting walls void the flying ice cube effect because the
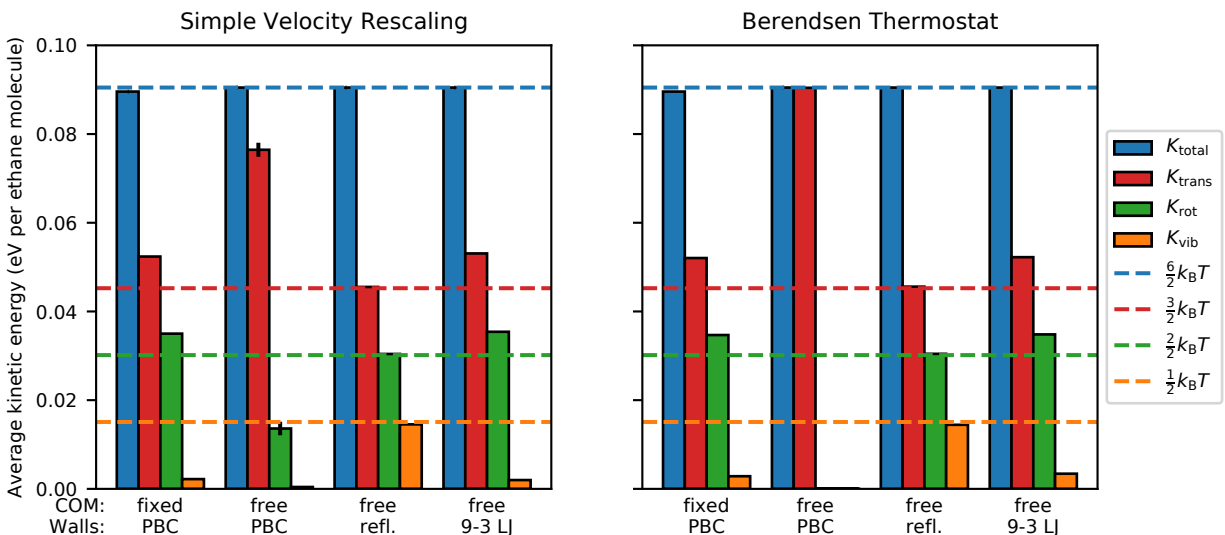
Figure 2.8:   Partitioning of the kinetic energies obtained from MD simulations of 50 ethane molecules in a 30 Å cubic simulation box under different conditions using (left) simple velocity rescaling and (right) the Berendsen thermostat. In each, the first simulation from left is the same simulation as shown in Fig. 2.3 and provides a basis for comparison. The second simulation shows the effects of letting the COM momentum drift (COM: free) as opposed to fixing it to zero (COM: fixed). The third and fourth simulations show the effects of hard (PBC: reflecting) and soft (PBC: 9-3 Lennard-Jones) wall boundaries, respectively, as opposed to PBC (Walls: PBC). Note that the dashed lines meant as a guide to the eye do not include the COM momentum constraint correction of $\frac{\frac{3}{2}k_{\mathrm{B}}T}{N_{\mathrm{molecs}}}$ that is reflected in the first simulation.

additional collisions with the walls give additional opportunities for energy to be transferred between kinetic modes, which acts more quickly than the Berendsen thermostat works to incorrectly partition the energy. To test this hypothesis, we made the walls softer so that a smaller redistribution of intramolecular kinetic energy would take place upon collision. Instead of reflecting walls, we used wall-particle interactions with a softer 9-3 Lennard-Jones potential,[71] $U(r) = \epsilon \left[ \frac{2}{15} \left( \frac{\sigma}{r} \right)^9 - \left( \frac{\sigma}{r} \right)^3 \right]$ with arbitrary $\epsilon$ and $\sigma$ values of $0.195\,\mathrm{kcal\,mol^{-1}}$ and $3.75\,\text{Å}$, respectively, and a shifted cutoff of $14\,\text{Å}$. We found that with this softer wall, energy equipartitioning holds less well than with the harder wall, giving some support to our hypothesis. We note further that the presence of the reflecting wall did not significantly change the distribution of total kinetic energies, i.e., the wall did not bring about equipartition indirectly through bringing about a more proper kinetic energy distribution.

Finally, we found that increasing the size of the simulation box reduces the flying ice cube effect, as can be seen in Fig. 2.9. As with decreasing the timestep (Fig. 2.4), here too we find that simple velocity rescaling recovers equipartition more easily than the Berendsen
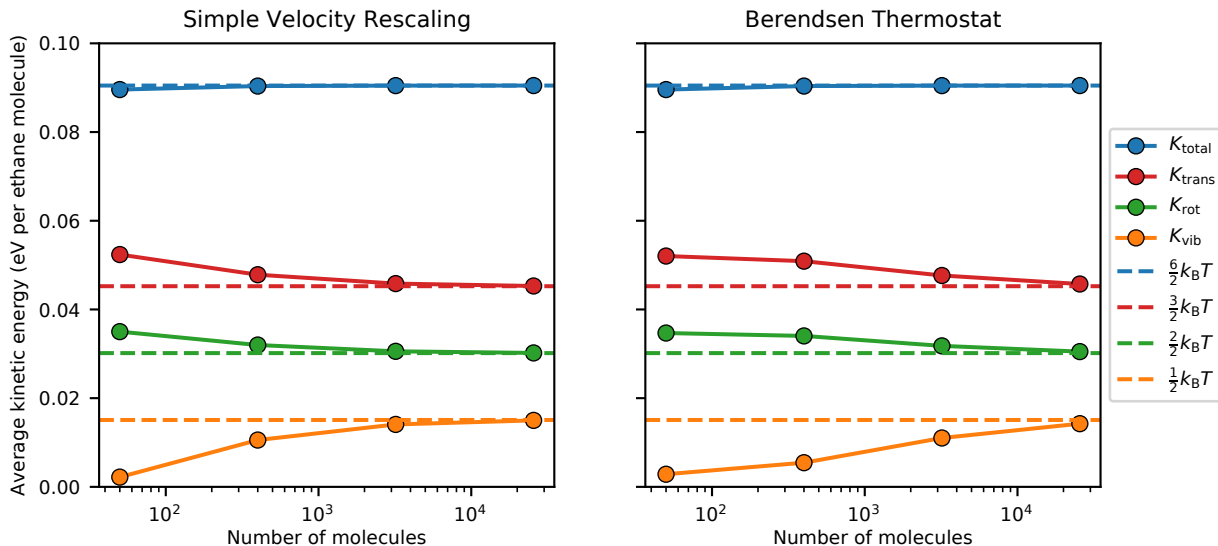
Figure 2.9: Partitioning of the kinetic energies obtained from MD simulations performed under the same conditions as in Fig. 2.3 but changing the number of ethane molecules, using (left) simple velocity rescaling and (right) the Berendsen thermostat. The simulation with 50 ethane molecules took place in a $30\,\text{Å}$ cubic simulation box, and the other simulations had their simulation boxes enlarged to maintain the same density. Note that the dashed lines meant as a guide to the eye do not include the COM momentum constraint correction of $\frac{\frac{3}{2}k_{\mathrm{B}}T}{N_{\mathrm{molecs}}}$, which is responsible for the slight deviation of the total kinetic energy from $\frac{6}{2}k_{\mathrm{B}}T$ that is more evident for the simulations with less molecules.

thermostat. We conjecture that this finite size effect occurs because the canonical ensemble's distribution of kinetic energy becomes more sharply peaked with increasing number of particles, i.e., the ratio of the standard deviation to the mean of the canonical kinetic energy distribution (the gamma distribution given in Eq. 2.7) scales as $\mathcal{O}\left(\frac{1}{\sqrt{N_{\mathrm{DOF}}}}\right)$ at constant temperature. Thus, as the number of particles increases, simple velocity rescaling and the Berendsen thermostat become more similar to the CSVR thermostat.

## Sampling configurational degrees of freedom

So far, we have exclusively used kinetic degrees of freedom to show that the simple velocity rescaling and Berendsen thermostat algorithms cause the violation of equipartition. These methods are sometimes used only to sample configurational degrees of freedom, justified on the grounds that the isokinetic ensemble samples the same configurational phase space as the canonical ensemble.[22,36–39] Since we have proven that the violation of equipartition is incommensurate with sampling the isokinetic ensemble, it follows that this justification is invalid. We now wish to show this explicitly. To do so, we will examine the radial distribution

function (RDF), which is solely dependent on configurational degrees of freedom.

In Fig. 2.10 (top-left), we show the RDFs of the supercritical ethane simulations whose kinetic energy partitionings are shown in Fig. 2.3. The Nosé-Hoover, CSVR, Langevin, and Gaussian thermostat simulations exhibit identical RDFs, but the simple velocity rescaling and Berendsen thermostat simulations show a subtly different RDF. Although the difference is slight, it is sufficient to demonstrably disprove the claims that simple velocity rescaling samples the same configurational phase space as the canonical ensemble and that the Berendsen thermostat samples a configurational phase space intermediate between the canonical and microcanonical ensembles.[40,41]

We next turn to saturated liquid phase ethane simulations, for which we show RDFs under various thermostats in Fig. 2.10 (top-right). The Nosé-Hoover, Langevin, CSVR, and Gaussian thermostats all give identical results typical of a simple diatomic liquid.[2] The simple velocity rescaling algorithm once again shows a subtle difference, but the Berendsen thermostat shows a very different RDF more reminiscent of the solid phase than the liquid phase,[2] and visualization of the Berendsen thermostat system shows that the ethane molecules have indeed packed into a volume smaller than available in the simulation box. Examination of the kinetic energy partitionings in Fig. 2.10 (bottom) shows that most of the kinetic energy is in vibrational modes, which is unexpected since that is the opposite of the usual flying ice cube result. The Berendsen thermostat's results are heavily dependent on the choice of time damping constant, with the RDF indicating a solid-like phase for time damping constants approximately from 10 to 150 fs (Fig. 2.11). This effect of intermediate time damping constants giving larger deviations than small or large ones has been observed before in simulations of bulk water, where the effect was attributed to the intermediate time constant matching a characteristic time scale on which dynamical correlations are most pronounced.[69] It appears clear that the Berendsen thermostat is not immune to the resonance artifacts that we have also seen with simple velocity rescaling (Fig. 2.7).

## Contemporary use of the simple velocity rescaling and Berendsen thermostat algorithms

Ours is not the first publication to warn against the use of simple velocity rescaling and the Berendsen thermostat.[43,49,72] Nonetheless, as we have stated, these algorithms continue to be widely used (Fig. 2.1). As we have just shown, for some systems the improper velocity rescaling algorithms may not greatly affect the system properties, and there are a slew of studies in which these thermostats are tested for specific systems, with some showing artifacts and others showing indistinguishability.[52,68,69,73–76] However, slight changes to a system could introduce artifacts in an unpredictable fashion. Rather than testing for the correctness of simple velocity rescaling or the Berendsen thermostat in every specific system, we advocate for the cessation of their use. We find no reason to use simple velocity rescaling or the Berendsen thermostat instead of the CSVR thermostat given their similar ease of implementation, likely similar speeds of equilibration,[77] and our study's finding that the
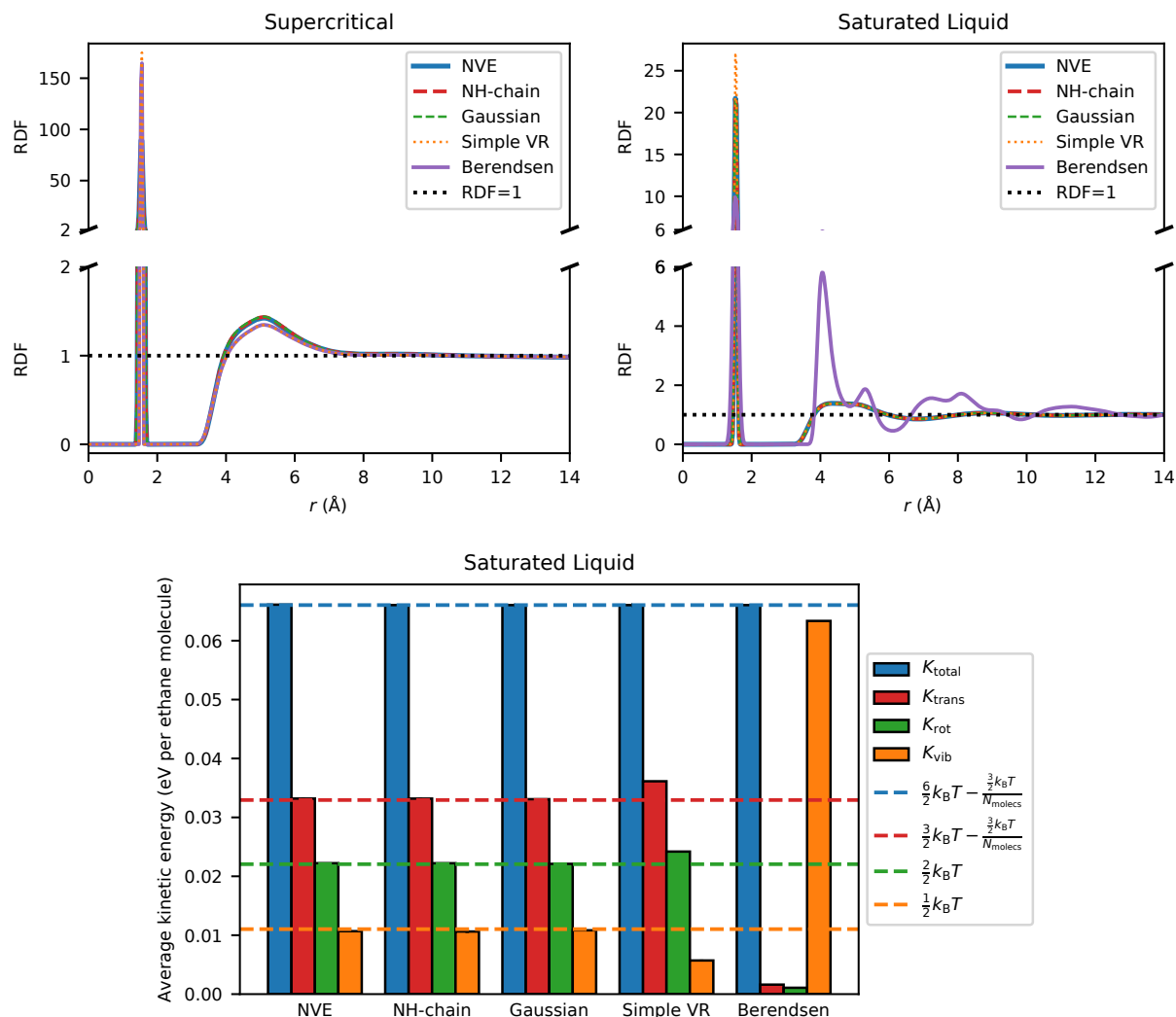
Figure 2.10:   (top-left) Radial distribution function (RDF) of the $CH_3-CH_3$ distance obtained from the MD simulations of 50 ethane molecules in a 30 Å cubic simulation box with a target temperature set to 350 K using various thermostats. These simulations were the same as the ones whose kinetic energy partitionings are shown in Fig. 2.3. (top-right) RDF of the $CH_3-CH_3$ distance obtained from MD simulations of 235 ethane molecules in a 30 Å cubic simulation box with a target temperature set to 256 K using various thermostats. These conditions were chosen such that the simulation would take place under saturated liquid conditions.[60] For both sets of simulations, COM momentum was fixed to zero throughout. The RDFs of both sets of simulations done using the Langevin and CSVR thermostats were indistinguishable from the RDF using the Nosé-Hoover thermostat within the line width. (bottom) Partitioning of the kinetic energies obtained from the saturated liquid simulations. The results of the simulations using the Langevin and CSVR thermostats were indistinguishable from the dashed lines of equipartition within the line width.
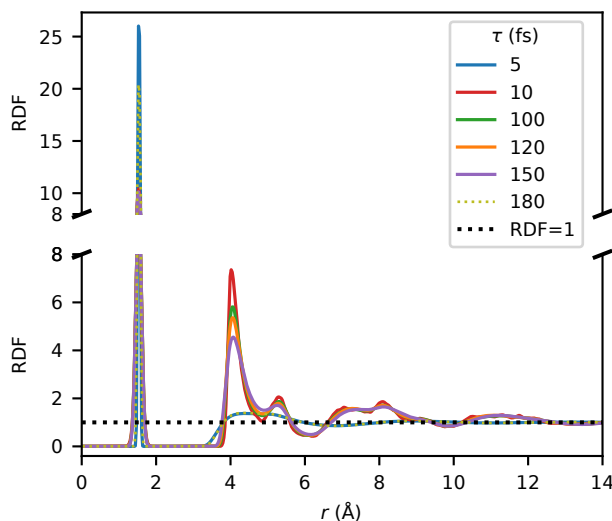
Figure 2.11: Similar to Fig. 2.10, radial distribution function (RDF) of the $CH_3$-$CH_3$ distance obtained from MD simulations of 235 ethane molecules in a 30 Å cubic simulation box with a target temperature set to 256 K using the Berendsen thermostat with different values of the time damping constant. RDFs for simulations with the time damping constant set to 0.5 and 2 fs look similar to the 5 fs case, and RDFs for simulations with the time damping constant set to 200 and 1,000 fs look similar to the 180 fs case.

CSVR thermostat does not lead to the flying ice cube effect, As a case study on the dangers of continuing to use these thermostat algorithms, we examine a highly-cited study in depth, the replication of which initially led us to examine the flying ice cube phenomenon.

In 2007, a flexible force field intended for use with MOF-5 was parameterized,[62] and it was shortly thereafter used to study the confined transport of guest molecules within the framework.[63] The authors were able to replicate the experimental diffusion coefficient of confined benzene, but they found that this replicability only held when the MOF was allowed to be flexible; when the MOF atoms were held rigid, the benzene diffusion coefficient increased by an order of magnitude. The conclusions of this manuscript are often evoked to question the validity of the rigid framework assumption that is commonly used in many MOF molecular simulation studies.

The finding continues to be accepted since it is known that the effect of framework flexibility on guest diffusion is complex,[78] though surprise has been expressed[79] since a rigid lattice more typically leads to a decrease in the diffusion coefficient for tight fitting molecules.[78] In addition, using a different flexible force field for MOF-5,[80] it was found that flexibility had little effect on the diffusion coefficient, increasing it by less than a factor of 1.5.[81]

As the reader now anticipates, Amirjalayer *et al.*[63] used the Berendsen thermostat, which
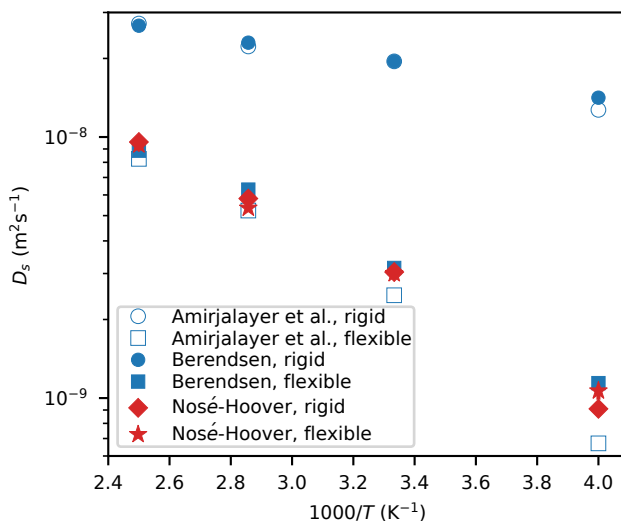
Figure 2.12:   Self-diffusion coefficient of benzene in MOF-5 at a loading of 10 molecules per unit cell as a function of inverse temperature.  Data are shown for flexible and rigid frameworks, and using the Berendsen and Nosé-Hoover chain thermostats (use of the CSVR thermostat gives diffusion coefficients that are statistically indistinguishable from use of the Nosé-Hoover thermostat).  With the Berendsen thermostat, it appears that the framework flexibility has a large effect on the calculated diffusion coefficient, replicating the main finding of Amirjalayer $et$ $al.$[63].  However, it is seen that this result is a flying ice cube artifact, as no flexibility effect is seen with the Nosé-Hoover thermostat. Error bars represent $\pm 1$ standard error of the mean using block averaging,[4] and are not shown for the data from Amirjalayer $et$ $al.$[63] or if they would be smaller than the symbol size.

was the default option in the Tinker simulation package at the time (the default has since been changed to the CSVR thermostat).[61] As we show in Fig. 2.12, the result of Amirjalayer $et$ $al.$[63] was completely an artifact of the Berendsen thermostat. Using the same force field, no dependence of the benzene diffusion coefficient on the framework flexibility is observed when a Nosé-Hoover or CSVR thermostat is used. Apparently, when the Berendsen thermostat is thermostatted to fewer degrees of freedom during rigid framework simulations, the flying ice cube effect becomes more noticeable and kinetic energy is drawn into the translational modes of the guest benzene molecules, accounting for the result observed by Amirjalayer $et$ $al.$[63]. We also found that changing the time damping constant of the Berendsen thermostat had a large effect on the diffusion coefficient (Fig. 2.13).

As an aside, it is now known that bulk-like vapor and liquid phases of benzene exist in MOF-5 below a critical temperature.[82] It is actually improper to calculate the diffusion coefficient at a loading that is within the vapor-liquid phase envelope, e.g., 3 to 67 molecules per unit cell at $300\,K$ in this system,[82] since there is not a single homogeneous phase present
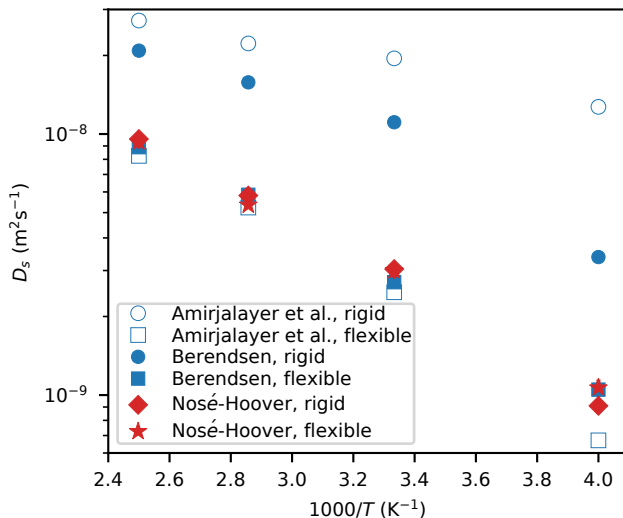
Figure 2.13:   Identical to Fig. 2.12, except the Berendsen thermostat was used with a time damping constant of 200 fs instead of 100 fs.

at these conditions. Here, we are not attempting to calculate correct diffusion coefficients of benzene in MOF-5, but rather to compare results with the prior work of Amirjalayer *et al.*[63], which conducted the simulations at a loading of 10 molecules per unit cell. The importance of framework flexibility on the simulated diffusion coefficient is expected to be independent of the choice of loading.

Other errors, varying in severity, are likely present in many of the thousands of studies that have used simple velocity rescaling or the Berendsen thermostat. Occasionally, one of these errors is explicitly pointed out,[83,84] but negative replications are not commonly published,[85] so the extent to which these articles contain data contaminated by the flying ice cube artifact cannot be estimated.

## 2.4   Concluding remarks

In this work, we have shown that rescaling velocities to a non-canonical distribution of kinetic energies, as is done with the simple velocity rescaling and Berendsen thermostat algorithms, causes the flying ice cube effect whereby the equipartition theorem is violated. Thus, simple velocity rescaling does not sample the isokinetic ensemble, and the Berendsen thermostat does not sample a configurational phase space intermediate between the canonical and microcanonical ensembles; justifications for their use do not hold. The flying ice cube effect is brought about by a violation of balance causing systematic redistributions of kinetic energy; this violation is lessened as the timestep between simple velocity rescalings is decreased, eventually making simple velocity rescaling equivalent to the Gaussian thermostat. Equipartition

violation is completely avoided when velocities are rescaled to the canonical distribution of kinetic energies, as is done under the CSVR thermostat, because detailed balance is obeyed.

We have identified several simulation parameters which affect the prominence of the flying ice cube effect under simple velocity rescaling and the Berendsen thermostat. These include the timestep, the thermostat's coupling strength, the frequency of collisions within the simulation (e.g., with a wall), and the system size. However, most of these parameters cannot be adjusted in a manner that eliminates the flying ice cube effect without making simulations prohibitively expensive for relevant systems of contemporary interest. Another reason not to attempt to tune these simulation parameters to allow the use of incorrect thermostatting algorithms is the existence of additional resonance artifacts that occur when the thermostat coupling strengths are set to particular values that are difficult to predict *a priori*.

Finally, we have demonstrated several severe simulation artifacts that the flying ice cube effect can bring about to the system's structural and dynamic properties. These include incorrect RDFs, phase properties, and diffusion coefficients. We have highlighted one case in which the flying ice cube effect has been wholly responsible for the main finding of a highly-cited study. Many more such cases are likely present in the literature.

We strongly advocate for discontinuing use of the simple velocity rescaling and Berendsen thermostat algorithms in all MD simulations for both equilibration and production cycles. The results of past studies that have used these two algorithms should be treated with caution unless they are shown to be replicable with a more reliable thermostat. In situations where velocity rescaling methods are desirable, such as for fast equilibration of a system,[86] the CSVR thermostat should be used instead.

## 2.5    Appendix: equipartition in the isokinetic ensemble

To the best of our knowledge, it has not been shown that the equipartition theorem need necessarily apply in the isokinetic ensemble, and it is not immediately clear that it must. When additional constraints are added to the system, such as the constraint of a constant COM momentum that is typical in MD simulations with PBC or the constraint of a constant kinetic energy in the isokinetic ensemble, the change to the partition function can bring about a changed type of energy partitioning.[46,47]

To illustrate, we can briefly examine the former constraint of constant COM momentum, which has been analyzed before.[45–47] One might naively think that the equipartition theorem for degrees of freedom related to the constraint (in this case, kinetic degrees of freedom, $p_i$) would simply change to:

$$\langle H_{p_i} \rangle = \frac{1}{2} k_{\mathrm{B}} T \frac{N-1}{N} \tag{2.8}$$

However, this is incorrect. Instead, it can be shown that for the canonical ensemble with its COM momentum constrained to zero, the principle of energy equipartitioning is violated for

kinetic degrees of freedom.[47] The system instead obeys the equation:

$$\langle H_{p,i} \rangle = \frac{1}{2} k_{\mathrm{B}} T \frac{M_{\mathrm{total}} - m_i}{M_{\mathrm{total}}} \tag{2.9}$$

with a similar expression for the microcanonical ensemble when the COM momentum is constrained to zero.[47] For a system of equally-sized particles, the naive expression of Eq. 2.8 is recovered and equipartition holds, but for a system with particles of difference masses, the particles will have different amounts of kinetic energy; for a system containing massive tracer particles, the difference between the expressions can be severe.[48] In the thermodynamic limit, the constraint of constant COM momentum does not affect the equipartition theorem.

Our equipartition theorem analysis of the isokinetic ensemble very closely follows the work of Uline $et\ al.$[47] for the momentum-constrained canonical ensemble. The system to be analyzed is described by the Hamiltonian:

$$H(\mathbf{r}^N, \mathbf{p}^N) = \sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i} + U(\mathbf{r}^N) \tag{2.10}$$

The configurational part of the isokinetic ensemble's partition function is not interesting since it is equivalent to that of the canonical ensemble's. We will focus on the integral over momenta, or the "translational" partition function of the isokinetic ensemble:[38,47]

$$Q_{\mathrm{trans}}(N, V, T, K) = \int d\mathbf{p}^N \exp\left(-\beta \sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i}\right) \delta\left(K - \sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i}\right) \tag{2.11}$$

To solve this expression, we will take the Laplace transform, integrate, and then take the inverse Laplace transform. We Laplace transform $Q_{\mathrm{trans}}$ from $K \to s$ to obtain:

$$Q_{\mathrm{trans}}(N, V, T, s) = \int d\mathbf{p}^N \exp\left(-(\beta + s)\sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i}\right) \tag{2.12}$$

Integrating over all $\mathbf{p}_i$ gives:

$$Q_{\mathrm{trans}}(N, V, T, s) = (2\pi)^{\frac{dN}{2}} \left(\prod_{i=1}^{N} m_i^{\frac{d}{2}}\right) (s + \beta)^{-\frac{dN}{2}} \tag{2.13}$$

The inverse Laplace transform from $s \to K$ yields:

$$Q_{\mathrm{trans}}(N, V, T, K) = (2\pi)^{\frac{dN}{2}} \left(\prod_{i=1}^{N} m_i^{\frac{d}{2}}\right) \frac{\exp(-\beta K) K^{\frac{dN}{2}-1}}{\Gamma\left(\frac{dN}{2}\right)} \tag{2.14}$$

This translational partition function is then used to generate the probability distribution $f$ for a single kinetic degree of freedom, $p_1$:

$$f(N, V, T, K, p_1) = \frac{\int dp^{dN-1} \exp\left(-\beta \sum_{i=1}^{dN} \frac{p_i^2}{2m_i}\right) \delta\left(K - \sum_{i=1}^{dN} \frac{p_i^2}{2m_i}\right)}{Q_{\mathrm{trans}}(N, V, T, K)} \tag{2.15}$$

As before, we Laplace transform $(K \to s)$, integrate, and inverse Laplace transform $(s \to K)$ to obtain:

$$f(N,V,T,K,p_1) = \frac{(2\pi)^{\frac{dN-1}{2}} \frac{\left(\prod_{i=1}^{N} m_i^{\frac{d}{2}}\right)}{m_1^{\frac{1}{2}}} \frac{\exp(-\beta K)\left(K - \frac{p_1^2}{2m_1}\right)^{\frac{dN-1}{2}-1} \Theta\left(K - \frac{p_1^2}{2m_1}\right)}{\Gamma\left(\frac{dN-1}{2}\right)}}{Q_{\text{trans}}(N,V,T,K)}$$

$$= (2\pi m_1)^{-\frac{1}{2}} \frac{\left(K - \frac{p_1^2}{2m_1}\right)^{\frac{dN-1}{2}-1}}{K^{\frac{dN}{2}-1}} \frac{\Gamma\left(\frac{dN}{2}\right)}{\Gamma\left(\frac{dN-1}{2}\right)} \Theta\left(K - \frac{p_1^2}{2m_1}\right) \qquad (2.16)$$

The presence of the Heaviside step function is a consequence of the impossibility of satisfying the kinetic energy constraint if the kinetic energy of a single degree of freedom is greater than the set total kinetic energy. The function can be integrated over $p_1$ by setting the integration bounds as $p_1 = \pm\sqrt{2m_1 K}$ to remove the Heaviside step function from the integral. It can be verified that the integral of $f(N,V,T,K,p_1)$ over $p_1$ is unity.

The average kinetic energy of a kinetic degree of freedom is then:

$$\frac{<p_1^2>}{2m_1} = \int dp_1 \frac{p_1^2}{2m_1} f(N,V,T,K,p_1)$$

$$= \frac{K}{dN} \qquad (2.17)$$

which indicates equipartition for every kinetic degree of freedom, regardless of the value of $K$. If $K$ is set to the average kinetic energy for a particular temperature, i.e., $K = \frac{d}{2}Nk_{\text{B}}T$:

$$\frac{<p_1^2>}{2m_1} = \frac{1}{2}k_{\text{B}}T \qquad (2.18)$$

so every degree of freedom will have the same average kinetic energy as in the canonical or microcanonical ensembles.

Thus, the equipartition theorem must hold in the isokinetic ensemble.

# Chapter 3

# Generating material libraries: carbon schwarzites via zeolite-templating*

ZTCs comprise a relatively recent material class synthesized via the chemical vapor deposition of a carbon-containing precursor on a zeolite template, followed by the removal of the template. We have developed a theoretical framework to generate a ZTC model from any given zeolite structure, which we show can successfully predict the structure of known ZTCs. We use our method to generate a library of ZTCs from all known zeolites, to establish criteria for which zeolites can produce experimentally-accessible ZTCs, and to identity over ten ZTCs that have never before been synthesized. We show that ZTCs partition space into two disjoint labyrinths which can be described by a pair of interpenetrating nets. Since such a pair of nets also describes a triply periodic minimal surface (TPMS), our results establish the relationship between ZTCs and schwarzites—carbon materials with negative Gaussian curvature that resemble TPMSs—linking the research topics and demonstrating that schwarzites should no longer be thought of as purely hypothetical materials.

## 3.1   Introduction

The search for new forms of carbon has not stopped with the discovery of fullerenes and graphene. Indeed, there are over 500 unique triply periodic hypothetical carbon structures present in the Samara Carbon Allotrope Database, with the vast majority having been proposed in the last decade.[88] Many of these hypothetical carbon allotropes take the form of schwarzites, which are carbon structures that mimic a TPMS.[88–93] The interest in synthesizing these novel materials is not only motivated by the scientific beauty of these surfaces, but also by the predictions that they have unique electronic, magnetic, and optical properties that may make them useful for applications such as supercapacitors, battery electrodes, catalysis, gas storage, and separations.[94–98]

---

*This chapter is based on Braun, Lee, Moosavi, Barthel, Mercado, Baburin, Proserpio, and Smit [87].

The synthesis of these materials has been challenging.[97,99] A true breakthrough has been the development of the template carbonization process, in which a carbon-containing precursor is introduced into a porous structured template and carbonized, followed by removal of the template.[11,100–103] Varying the template structure can bring about a wide variety of carbon materials, including one-dimensional nanotubes and nanorods, two-dimensional graphene stacks, and three-dimensional ordered mesoporous or microporous carbons.[11,103]

Of interest for this work is when a zeolite is used as a template, with the resulting liberated carbon material referred to as a ZTC.[11,102] A motivation of this work is the discussion in the literature about whether a ZTC can be seen as a schwarzite.[104–108] Indeed, the experimental properties of ZTCs are exactly those which have been predicted for schwarzites, and hence ZTCs are considered as promising materials for the same applications.[11,102] As we will show, the similarity between ZTCs and schwarzites is striking, and we explore this similarity to establish whether the theory of schwarzites/TPMSs is a useful concept to understand ZTCs. In particular, we will focus on the role of the template in determining whether a stable schwarzite will form. At present, the selection of zeolite templates for this process has been based on trial and error, and in this work we provide a rationale for why the three commonly used zeolite templates—FAU, EMT, and beta—have been successful.[102,109,110] In addition, our approach yields suggestions for over ten additional zeolite structures that can be used for schwarzites that have not yet been synthesized.

## 3.2  *In silico* generation of ZTC structures

Unlike zeolites, ZTCs are not crystalline; instead, they are non-periodic orderings of atoms on periodic surfaces. Consequently, we cannot simply take the crystal structure as a starting point for a molecular simulation. The *in silico* generation of a sufficiently reliable ZTC structure for a given zeolite template is therefore an essential first step, for which two methods have so far been proposed. Nishihara and coworkers assembled carbon atoms inside the pores of FAU to match the ZTC-FAU carbon loading determined experimentally.[106,108,111] Roussel and co-workers instead used a grand-canonical Monte Carlo (GCMC) scheme that results in a ZTC model with a carbon loading dependent on the choice of chemical potential;[104,105,112] they subsequently tuned the chemical potential such that a structure was obtained with carbon neither undertemplating the zeolite surface nor filling in the zeolite's void space.

To avoid some of the limitations of the existing methods, we developed a Monte Carlo (MC) algorithm that aims to fully template the zeolite surface, providing a well-defined upper limit of the experimentally-achievable carbon loading. Our method closely mimics the synthetic process, in which the carbon precursor catalytically decomposes upon contact with the zeolite surface, forming a new surface of $sp^2$-hybridized carbon atoms.[102,107,110] Accordingly, our algorithm inserts $sp^2$-hybridized carbon atoms next to unsaturated carbons on the zeolite surface, and it performs additional MC moves to allow the added carbons to find their optimal positions. Proposed moves are accepted or rejected using an energy criterion. The MC process stops when all carbons are $sp^2$-saturated and no more surface binding
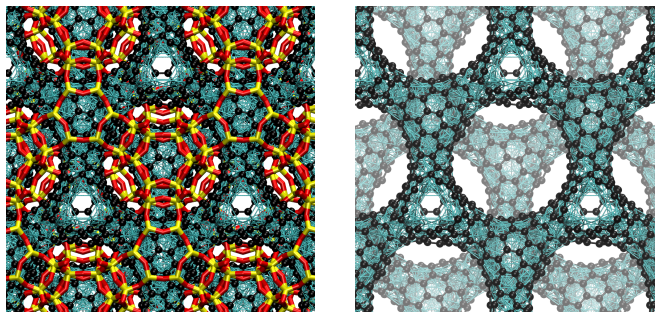
Figure 3.1: **Similarity between ZTC-FAU synthesized *in silico* and experiment.** Shown is our atomistic model (black carbon atoms), the parent FAU zeolite (red oxygen and yellow silicon atoms), and the XRD-derived model from Kim *et al.*[110] (cyan lines connecting partially-occupied carbon atom positions). The right view has the zeolite atoms removed for clarity.

sites are available. The advantage of this approach is that it does not require experimental data and it is relatively computationally inexpensive, so it can be readily extended to other zeolite templates to enable the high-throughput generation of novel ZTCs. More details on the algorithm can be found in the Methods section. Next we focus on a justification of our method by comparison with the experimental information reported on the (small) subset of ZTCs that have been synthesized so far.

Kim *et al.*[110] and Parmentier *et al.*[107] studied the ZTCs templated by the FAU and EMT zeolites using XRD. They found that the carbon-decorated surfaces were well-ordered, but they found disorder in the positions of the carbon atoms. To express this disorder, Kim *et al.*[110] introduced a model in which ZTC-FAU is described as a network of partially-occupied atomic sites. In our ZTC-FAU model, we obtain a similar type of disorder if we compare carbon atom positions over different units cells or over structures generated using different sets of random numbers. We also observe that the carbon atoms span a consistent surface. In Fig. 3.1, we compare our ZTC-FAU model with the model of Kim *et al.*[110], revealing that the carbon atoms of the two models lie on the same surface. Interestingly, Kim *et al.*[110] found the electron densities in the narrow nanotube-like necks to be more disordered than elsewhere, which they modeled by locating some partially-occupied carbon atoms inside the nanotube surfaces, while the rest of the ZTC is completely hollow. In our ZTC-FAU model, we find that some of the necks form fully cylindrical nanotubes while other necks are partially collapsed; in the model of Kim *et al.*[110], these two structural features are superimposed. In addition, we find that the computed XRD pattern of both our ZTC-FAU and ZTC-EMT structures match well with those of experiment,[109,110] exhibiting one strong peak for ZTC-FAU and three strong peaks for ZTC-EMT (Fig. 3.2).

Two further illustrative materials are ZTC-LTA and ZTC-LTL, templated by two of the smaller-pore zeolites used by Kim *et al.*[110]. Kim *et al.*[110] found that although these ZTCs exhibited zeolite-like crystal morphology and pore order, they could be easily crushed by hand
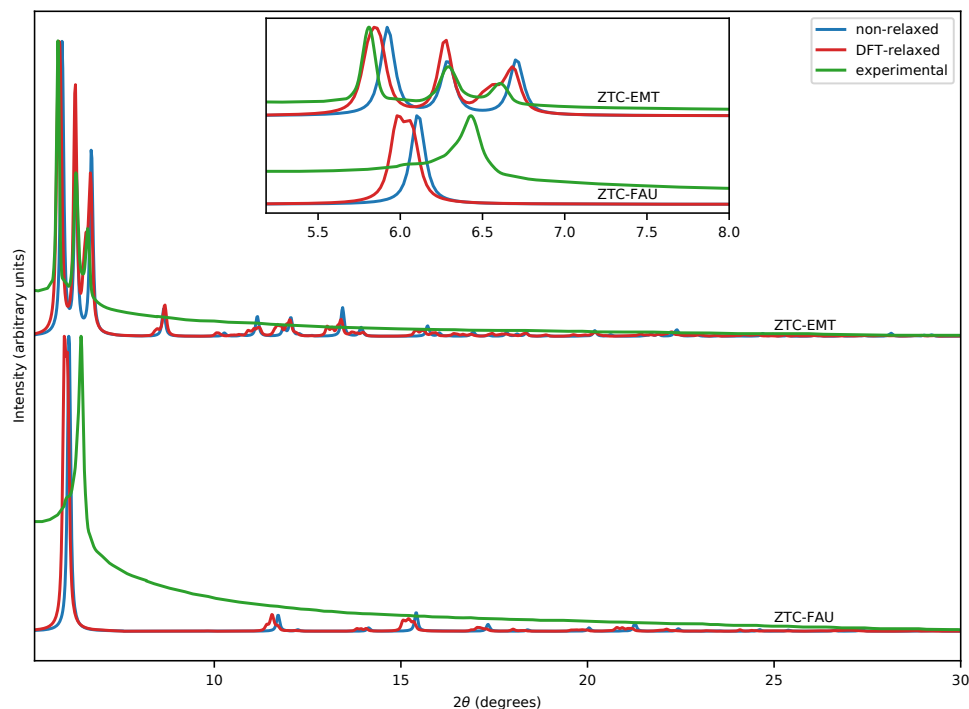
Figure 3.2: **Computed powder XRD diffraction patterns of the ZTC-FAU and ZTC-EMT structures.** Patterns are shown for the crystal structures both before and after DFT relaxation, along with experimental data from Kim *et al.*[110]. Inset is a zoomed image.

rubbing and they exhibited photoluminescence when dissolved in *N*-methylpyrrolidone; this was taken as evidence that the carbon was obtained as quantum dots or nanotubes rather than as a strongly bound triply periodic framework. Our ZTC-LTA structure consists of a network of fullerene-type structures linked by single atom bridges (Fig. 3.3). These single atom bridges may explain the physical characteristics experimentally observed, as single atom bridges are unlikely to be strong enough to withstand forces caused by compression or solvation. Similarly, our ZTC-LTL structure consists of arrays of carbon nanotubes. These observations suggest that zeolite templating might be an alternative synthetic method for producing pure fullerenes—which more typically require post-synthetic purification methods[113]—in much the same way that aluminum oxide templating has been used for producing carbon nanotubes.[11]

## Carbon loading in ZTC-FAU

Experimentally, carbon loadings in ZTC-FAU have been reported to range widely, extending from 0.12 to $0.65 \, \mathrm{g_{carbon} \, g_{zeolite}^{-1}}$ within a single study.[114] However, at high chemical vapor deposition (CVD) temperatures carbon deposits on the external surface of zeolites, and so the reported loadings for those samples should not be used for comparison to atomistic

models.[100,115–118] One group corrected for this both by conducting the deposition at low temperature, which resulted in a carbon loading of $0.29\,\mathrm{g_{carbon}\,g_{zeolite}^{-1}}$,[111] and by proposing a method to estimate the fraction of carbon deposited on the external surface, which led to the finding that the internal carbon loading could be increased up to $0.54\,\mathrm{g_{carbon}\,g_{zeolite}^{-1}}$ by varying the CVD conditions.[106]

Since Kim *et al.*[110] were able to perform low-temperature deposition by pre-loading the zeolite with metal catalysts, external surface deposition did not occur, and CVD could be carried out until the carbon loading saturated with time, which was found to occur at $0.31$ to $0.32\,\mathrm{g_{carbon}\,g_{zeolite}^{-1}}$ via thermogravimetric analysis (TGA). It was estimated that only $83\,\%$ of the zeolite pores were replicated with carbon,[110] so the pores filled with carbon actually contained $0.38\,\mathrm{g_{carbon}\,g_{zeolite}^{-1}}$. don't lanthanum, sodium, and oxygen, the mass ratio is in the range of $0.47$ to $0.51\,\mathrm{g_{carbon}\,g_{siliceous\ zeolite}^{-1}}$. (A zeolite composition of $\mathrm{La_{23.52}Na_{14.27}O_{26.69}[T(Si_A l)O_2]_{192}}$ was found by Kim *et al.*[110] via XRD analysis, but per a personal correspondence with the authors of that study, a smaller amount of lanthanum was incorporated into the zeolite crystal used for the TGA measurement. ICP-AES conducted on the TGA crystal found a composition of $\mathrm{La_{14}Na_{17}[Si_{0.7}Al_{0.3}O_2]_{192}}$ with an unknown amount of extra-framework oxygen, which our calculation assumed to be equal to that in the XRD crystal.)

Among the four ZTC-FAU models existing in the literature, the mass ratio is $0.65\,\mathrm{g_{carbon}\,g_{zeolite}^{-1}}$ in the model of Roussel *et al.*[104] (stated incorrectly as being $0.60\,\mathrm{g_{carbon}\,g_{zeolite}^{-1}}$, but later corrected to $0.65\,\mathrm{g_{carbon}\,g_{zeolite}^{-1}}$)[119], $0.28\,\mathrm{g_{carbon}\,g_{zeolite}^{-1}}$ in the incompletely-templated model of Nishihara *et al.*[111], $0.5\,\mathrm{g_{carbon}\,g_{zeolite}^{-1}}$ in the model of Nueangnoraj *et al.*[106], and $0.3\,\mathrm{g_{carbon}\,g_{zeolite}^{-1}}$ in the model of Nishihara *et al.*[108]. The latter three structures were generated specifically to fit experimental carbon loadings.

We obtained an average ZTC-FAU carbon loading of $0.73\,\mathrm{g_{carbon}\,g_{zeolite}^{-1}}$. This is similar to the carbon loading found by Roussel *et al.*[104], and it represents an upper limit of what can be experimentally achieved by the surface-templating of a pristine zeolite crystal in the absence of diffusion limitations.

We note that the present work concerns ZTCs formed from zeolites that have been fully-templated with carbon. We thus made sure to compare our ZTC models to experimental studies which similarly attempted to synthesize ZTCs by fully-templating the parent zeolites, such as Kim *et al.*[110] and Parmentier *et al.*[107]. We caution that comparisons of gas adsorption isotherms between simulation and experiment require similar rigor. However, fully-templated structures with higher carbon loading have lower specific surface areas than partially-templated structures with lower carbon loading, since under-templating a zeolite can result in a structure consisting of two-sided sheets rather than a closed schwarzite-like surface that has only one side accessible to external adsorbates.[106] Hence, many of the experimentally-obtained gas adsorption isotherms available in the literature cannot be simply compared to simulated data conducted on model ZTCs, since many experimental studies focus on optimizing ZTCs for applications requiring high surface areas, such as gas storage and separations, and hence under-templated materials are more likely to have been used. This may explain why prior ZTC studies have shown disagreement between simulated and

experimental gas adsorption isotherms.[104,119]

## 3.3 Results and discussion

In the previous section we have shown that our ZTC-generating method gives very reasonable structural descriptions of the known ZTCs; in this section we explore novel ZTCs.

### Experimental accessibility of ZTCs

We show a small subset of the ZTCs we generated using zeolite templates taken from the IZA database[120] in Fig. 3.3, demonstrating the wide diversity of ZTC topologies possible. From ZTC-LTA and ZTC-LTL, we know that not all zeolite templates will yield ZTCs that remain triply periodic once subject to stress due to the absence of linkages in all three dimensions or due to the presence of weak single atom bridges formed in zeolite pores of insufficiently large diameter. In addition, zeolites with small pores will be difficult to template since large channels are necessary for the CVD precursor to diffuse throughout the zeolite.

We found that ZTCs formed from a parent zeolite with a largest triply periodic free sphere diameter ($D_{f,3p}$) (which we define as the size of the largest spherical probe that can travel through the zeolite's channels in three directions)[121] smaller than about 5 Å either lack linkages in some dimensions, resulting in one- or two-dimensional carbon structures, or the linkages contain some sort of defect, such as single atom bridges or flat sheet connectors (Fig. 3.3, top two rows). Of the ZTCs formed from a parent zeolite with a $D_{f,3p}$ larger than about 5 Å, we were able to find a three-periodic ZTC model which contained none of these defects (Fig. 3.3, bottom two rows).

We established the finding of a limiting $D_{f,3p}$ around 5 Å by having made ZTCs from a test set of the IZA zeolites, and we then went on to generate ZTCs for nearly all current IZA zeolites with a $D_{f,3p}$ greater than 4.8 Å, going slightly below the limiting value to ensure completeness. We also generated a ZTC for beta polymorph B because zeolite beta is one of the few zeolites that has already been shown to generate a ZTC that retains the parent zeolite's structural regularity,[102,109,110] and zeolite beta is composed of two intergrown structures, polymorphs A and B (the three-letter IZA code of the former structure is BEA, and we refer to the latter structure as BEB, though it lacks an IZA assignment)[122]. The only IZA zeolites with a $D_{f,3p}$ greater than 4.8 Å for which we could not produce ZTCs were -CLO, -IFT, and -ITV, with the dash denoting an interrupted framework; these interrupted frameworks have under-coordinated oxygen atoms that extend into the void space of the zeolite and give a less smooth surface for carbon atoms to wrap around.

Next, we relaxed the zeolite-free three-periodic ZTC structures to their local energetic minima using DFT, and we compared their per-atom energies to diamond using the formula $E_{ZTC,relative} = \frac{E_{ZTC}}{N_{atoms,ZTC}} - \frac{E_{diamond}}{N_{atoms,diamond}}$, where $E$ is the unit cell energy and $N_{atoms}$ is the number of atoms in the unit cell (Fig. 3.4). We found that the ZTCs free of defects tend to
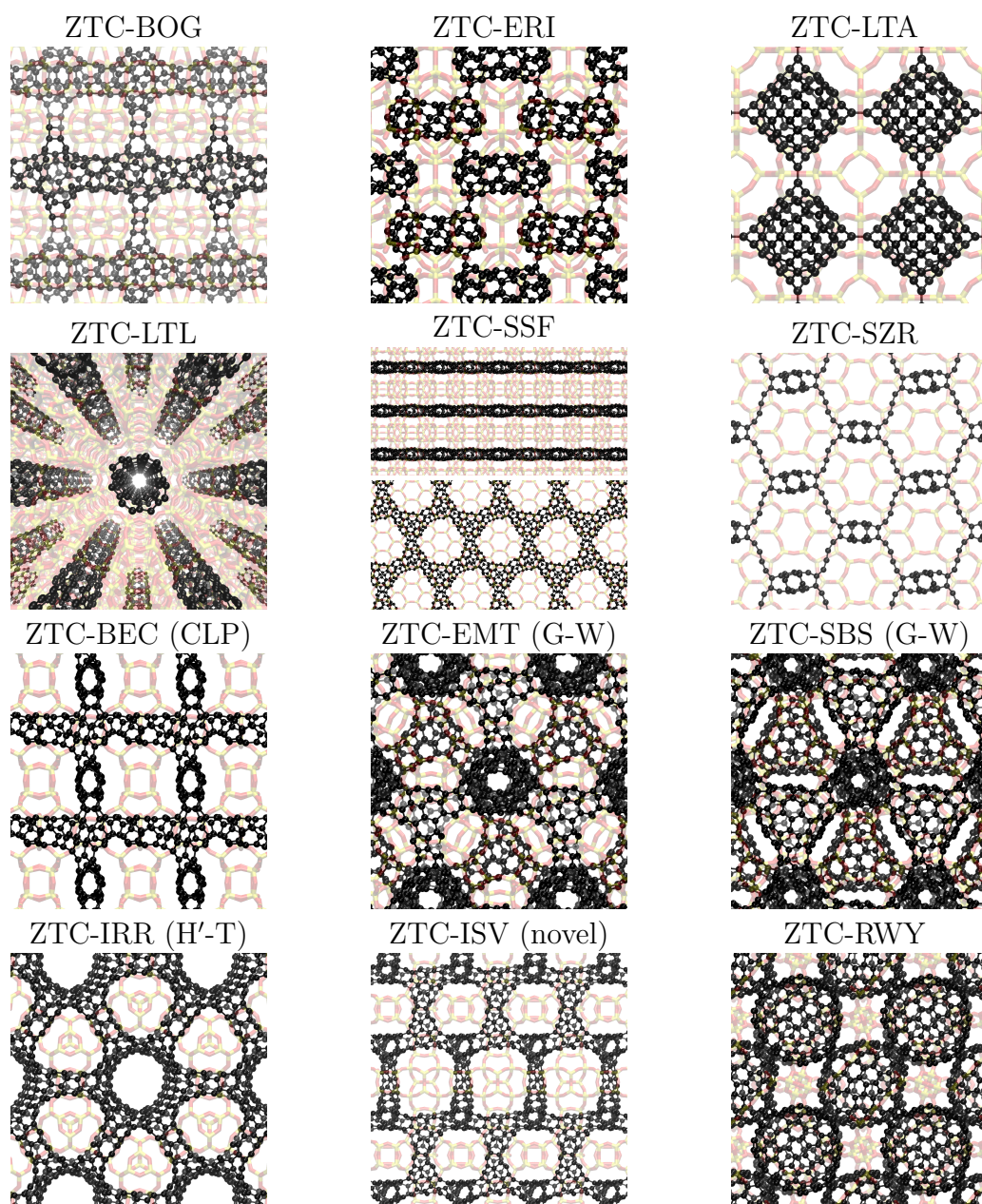
Figure 3.3:   **Library of ZTCs synthesized *in silico*.** Shown is a subset of the ZTC structures (black carbon atoms) and their parent zeolites (red oxygen and yellow silicon atoms). The top two rows of ZTCs were generated from parent zeolites with $D_{f,3p}$ less than $5\,\text{Å}$, and they either lack linkages in some dimensions or they contain linkages with defects; we consider these ZTCs either unlikely to be experimentally synthesizable or unlikely to be stable subject to stress. The bottom two rows of ZTCs were generated from parent zeolites with $D_{f,3p}$ greater than $5\,\text{Å}$, which we propose as experimentally synthesizable and stable structures. The TPMS each structure resembles is given in parentheses. ZTC-RWY does not resemble a TPMS, but rather consists of a body-centered cubic packing of fullerenes.
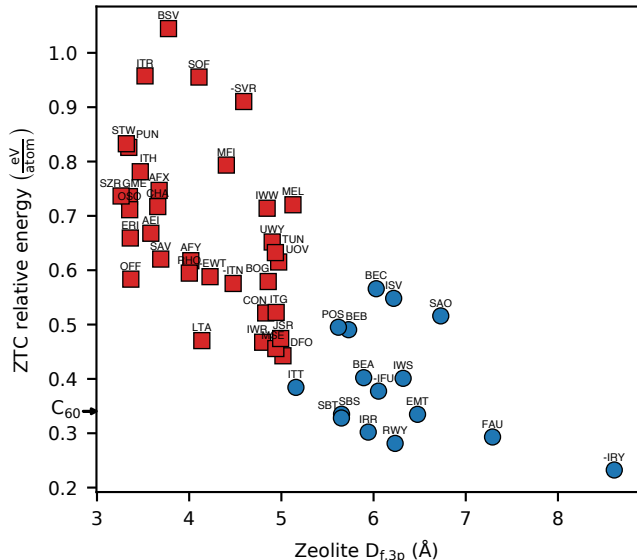
Figure 3.4:  **Per-atom energies of DFT-relaxed three-periodic ZTC models relative to diamond as a function of the parent zeolite's largest triply periodic free sphere diameter ($D_{f,3p}$).** The energy of an isolated buckminsterfullerene molecule ($C_{60}$) is shown as a reference. Blue circles represent ZTCs free of defects such as single atom bridges and flat sheet connectors, while red squares represent ZTCs containing these features. Table A.1 contains the information shown in this figure.

have lower energies, supporting our structural analysis. Several of these ZTCs have energies close to or less than the energy of an isolated buckminsterfullerene molecule.

Finally, we assessed the finite-temperature stability of the ZTCs by performing MD simulations, for which we used the AIREBO force field to allow dynamic breaking and formation of carbon-carbon bonds.[123] We summarize the results of these simulations in Table A.1. We found that many of the ZTCs containing defects collapsed, broke bonds, or exhibited significant changes to the unit cell parameters. However, almost all of the ZTCs free of defects maintained their structures for several nanoseconds, even at temperatures as elevated as 1,000 K; the only exception was ZTC-IFU, formed from a zeolite with an interrupted framework which, as stated earlier, provides a less smooth surface.

We expect the ZTCs free of defects and which are stable under finite-temperature MD to be experimentally-accessible as triply periodic frameworks. The exhaustive list of IZA zeolites which satisfy these thermodynamic and kinetic stability criteria comprises BEA, BEB, BEC, EMT, FAU, IRR, -IRY, ISV, ITT, IWS, POS, RWY, SAO, SBS, and SBT (Table 3.1). It is encouraging to see that we correctly predict those ZTCs that have already been shown to retain the parent zeolite's structural regularity: namely, ZTC-FAU, ZTC-EMT, and ZTC-beta.[102,109,110] The remaining zeolites therefore warrant closer attention as templates of potentially synthesizable structures.
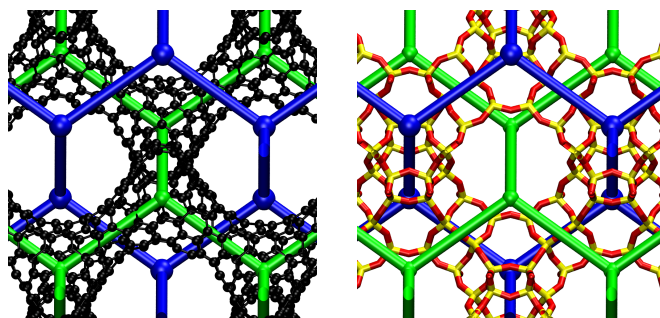
Figure 3.5: **Labyrinth nets of ZTC-FAU.** (left) One of our ZTC-FAU models (black carbon atoms), and (right) its parent FAU zeolite (red oxygen and yellow silicon atoms), both shown with the ZTC's **dia** zeolite labyrinth net (blue) and the ZTC's **dia** void labyrinth net (green).[129]

## Comparison of experimentally accessible ZTCs with TPMSs

In prior theoretical work, schwarzites were created by decorating TPMSs with carbon atoms, and so it was known by definition which TPMS a schwarzite resembled.[89–93] We must instead associate a ZTC with a TPMS *a posteriori.* The carbon atoms of a ZTC lie on a surface[124]— the ZTC surface—and here we find the TPMS that most closely resembles this surface.

Of the 15 ZTCs we suggest as experimentally accessible, the majority form surfaces that partition space into two disjoint labyrinths which can be described by the nets running through them, as is common in the study of TPMSs (Fig. 3.5).[125–127] Only ZTC-RWY does not form a surface that bounds two labyrinths, with the structure instead consisting of a body-centered cubic packing of fullerenes (Fig. 3.3); we focus on the remaining 14 ZTCs. We associate each of these 14 ZTCs with the TPMS that could be traced by the same labyrinth nets in a matching lattice system. We obtained the two ZTC labyrinth nets in a manner that provides an unambiguous assignment (see Methods section for details). Depositing the carbon atoms on the zeolite surface will leave the zeolite atoms in one of the two labyrinths defined by that surface.[128] We refer to the ZTC labyrinth that contained the zeolite atoms prior to their removal as the zeolite labyrinth, with the other labyrinth referred to as the void labyrinth (Fig. 3.5).

One way to illustrate the similarity between a ZTC surface and its associated TPMS is by performing a geometrical transformation, numerically minimizing the area of the ZTC surface subject to constant volume constraints. By gradually varying the volumes of the two labyrinths until their ratio matches that of the TPMS, the relation between the surfaces becomes apparent. We demonstrate this process in Fig. 3.6 for ZTC-FAU and in Fig. 3.7 for ZTC-EMT.

We show the TPMSs we associate with each of the experimentally-accessible ZTCs in Table 3.1, where it can be seen that many of the ZTCs resemble known TPMSs. We identify ZTC-FAU as resembling the Schwarz D TPMS and ZTC-EMT as resembling the Schoen
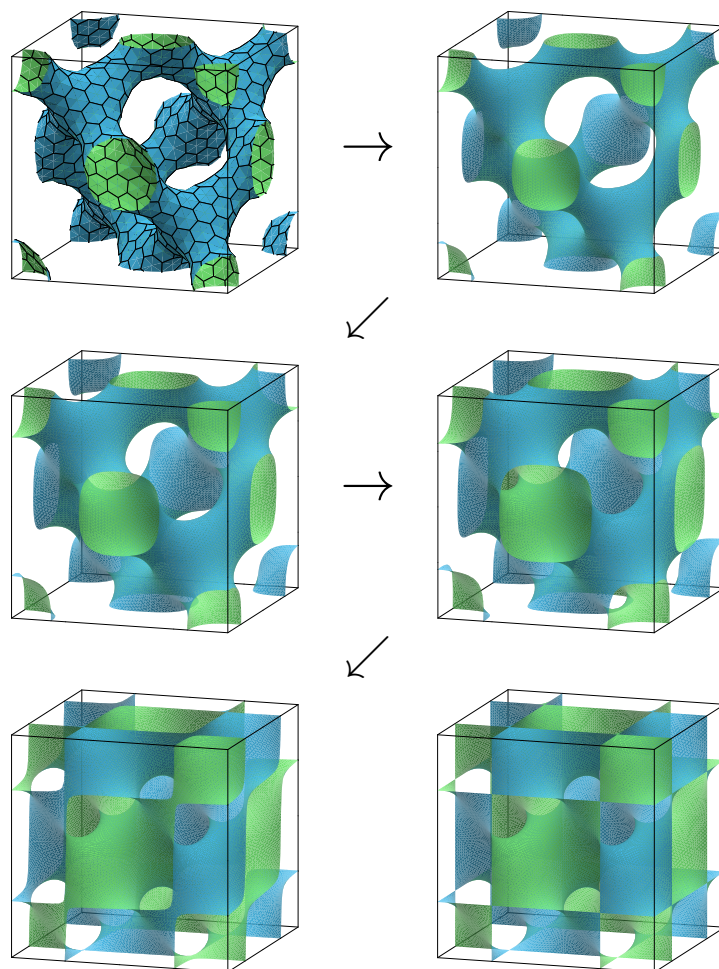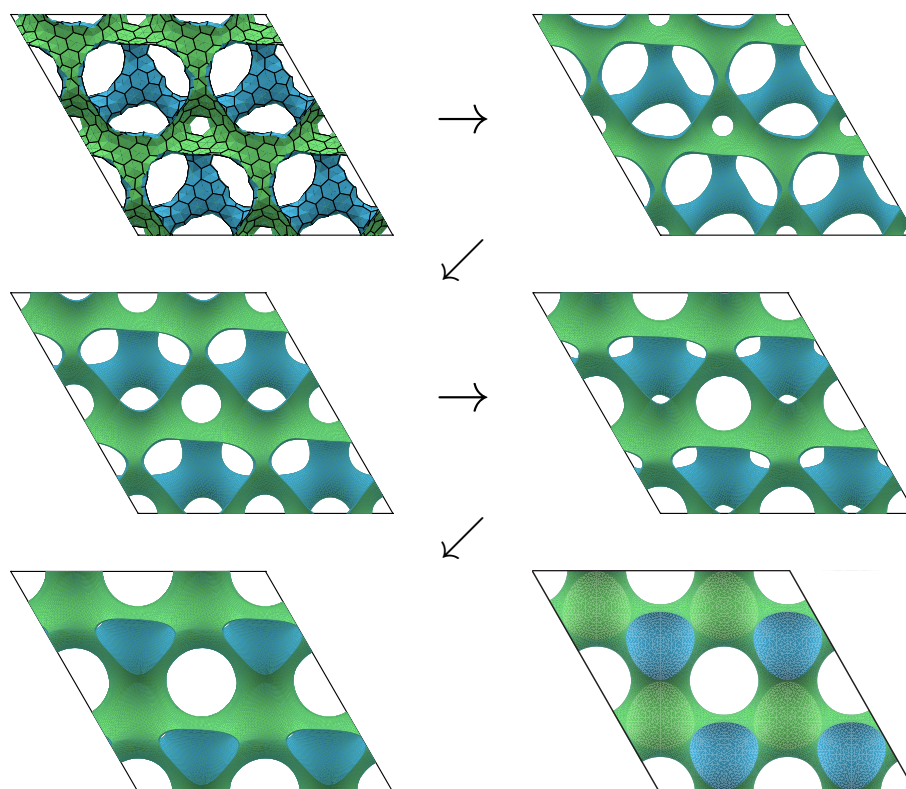
Figure 3.6:   **Numerical minimization of ZTC-FAU's surface area.** The process begins with the vertices of the surface defined by the carbon atoms and triangulation of the faces (first image), followed by further refinement of the triangulation by subdividing the original triangles and minimizing the surface area subject to a constant volume constraint (second image). The ratio of the two labyrinths' volumes in the constraint is slowly varied while continuing to minimize the surface area (third and fourth images) until it is unity (fifth image). The resulting surface is visually very similar to the Schwarz D TPMS (final image). The side of the surface touching the ZTC's zeolite labyrinth is colored blue and the side touching the ZTC's void labyrinth is colored green.

Figure 3.7: **Numerical minimization of ZTC-EMT's surface area.** The process begins with the vertices of the surface defined by the carbon atoms and triangulation of the faces (first image), followed by further refinement of the triangulation by subdividing the original triangles and minimizing the surface area subject to a constant volume constraint (second image). The ratio of the two labyrinths' volumes in the constraint is slowly varied while continuing to minimize the surface area (third and fourth images) until it is approximately equal to that of a particular member of the Schoen G-W family of TPMS (fifth image). The resulting surface is visually very similar to that member of the Schoen G-W family of TPMS (final image). The side of the surface touching the ZTC's zeolite labyrinth is colored blue and the side touching the ZTC's void labyrinth is colored green.

Table 3.1: Topology and symmetry of the experimentally-accessible ZTCs and their parent zeolites.

| zeolite | zeolite's space group | ZTC's zeolite labyrinth net | ZTC's void labyrinth net | labyrinth nets' genus | TPMS resembling ZTC |
|---|---|---|---|---|---|
| RWY | $Im\bar{3}m$ | **nbo**[a] | **bcu**[a] | 4[a] | I-WP[a] |
| FAU | $Fd\bar{3}m$ | **dia** | **dia** | 3 | D |
| EMT | $P6_3/mmc$ | **lon** | **gra** | 5 | G-W |
| -IRY | $P6_3/mmc$ | **lon** | **gra** | 5 | G-W |
| SBS | $P6_3/mmc$ | **lon** | **gra** | 5 | G-W |
| IRR | $P6/mmm$ | **bnn** | **hex** | 4 | H′-T |
| ITT | $P6/mmm$ | **bnn** | **hex** | 4 | H′-T |
| SBT | $R\bar{3}m$ | **dia**[b] | **dia**[b] | 3 | rPD |
| IWS | $I4/mmm$ | 4,6T585 | **crb** | 5 | novel |
| POS | $P4_2/mnm$ | **cds**[b] | **cds**[b] | 3 | CLP[c] |
| BEC | $P4_2/mmc$ | **cds** | **cds** | 3 | CLP |
| ISV | $P4_2/mmc$ | sqc145 | **tfi** | 5 | novel |
| SAO | $I\bar{4}m2$ | **dia**[b] | **tfa** | 3 | tD |
| BEA | $P4_122$ | **ths**[b] | **ths**[b] | 3 | tD |
| BEB | $C2/c$ | **ths**[b] | **ths**[b] | 3 | mDCLP |

[a] The model ZTC's connecting tunnels are pinched off, disrupting the labyrinth and forming a structure consisting of a body-centered cubic packing of fullerenes. Here we describe the surface that would have formed if the pinching off had not occurred and the cages were instead connected.

[b] Not the maximum-symmetry embedding of the net.

[c] The CLP TPMS has the $P4_2/mmc$ space group, which is a minimal supergroup of the POS zeolite's $P4_2/mnm$ space group. All other identified TPMSs in the table have the same space group as the parent zeolite.

G-W TPMS (originally referred to as g-g′),[125] with the latter TPMS also resembling ZTC-IRY and ZTC-SBS. ZTC-EMT and ZTC-SBS are illustrated in Fig. 3.3, where the structural similarities between these two ZTCs are readily apparent, making evident the reason for their associations with the same TPMS. Nonetheless, it is clear that the two ZTCs are not identical: the two surfaces have different labyrinth volume ratios and unit cell parameter to carbon-carbon bond length ratios, and the carbon networks are distinct, e.g., they are composed of unequal fractions of hexagonal, heptagonal, and octagonal rings with varying degrees of strain. These disparities will lead to distinctive electrical and mechanical properties, giving materials with different performances for select applications.

Other TPMSs we have found to resemble ZTCs include Schwarz CLP,[130] Schoen H′-T,[125] Schoen tD,[125,131] rPD of Koch and Fischer[131], and mDCLP of Fogden and Hyde[132], the last three of which are lower-symmetry variants of the Schwarz D TPMS. To the best

of our knowledge, TPMSs with the labyrinth nets of ZTC-IWS and ZTC-ISV have not been reported in the literature, which may be because they are of the less well-studied non-balanced surfaces.[133,134] Since we have made atomistic models of ZTCs from all known non-interrupted zeolites with sufficiently large $D_{f,3p}$, to the best of our knowledge the TPMSs discussed in this work represent an exhaustive list of all schwarzites that can be made by templating the presently-known zeolites with carbon.

## Rational design of schwarzite templates

So far, we have shown how to determine which schwarzite will be generated by templating a given zeolite. Equally interesting is the inverse problem: determining which template will produce a schwarzite resembling a given TPMS. If a TPMS of, say, type CLP is desired, Table 3.1 can be consulted to find that zeolites POS and BEC can be used as templates. However, if the TPMS is not in Table 3.1, a novel material would be needed. Fortunately, new zeolites are continuously being synthesized, and the ability to rationally design synthetic methods to achieve desired zeolites is beginning to become available.[135,136] Furthermore, one need not restrict oneself to zeolites as multiple new material classes are being developed for use as hard-templates.[101,103]

To obtain some insights into the types of new structures that can be generated, it is interesting to consider the databases of hypothetical zeolites that have been developed to guide the discovery of new zeolites. In this work we used Treacy and Foster's silver hypothetical zeolite database, which contains 1,270,921 structures, and Deem's SLC hypothetical zeolite database, with 331,172 structures.[137,138] In principle we can solve the inverse problem by generating the ZTCs for all these hypothetical zeolites and tabulating the corresponding TPMSs, similarly to our approach for the known zeolites, but the large number of structures makes this impractical.

To avoid this enormous computational effort, we can significantly filter down the hypothetical zeolite databases. Namely, we can restrict the search to zeolites with $D_{f,3p}$ larger than 5 Å and which have the same space group as the target TPMS. The latter filter comes from an expectation that the space group of the TPMS associated with a ZTC should either be identical with or a supergroup of the space group of the parent zeolite: an expectation that is consistent with our findings given in Table 3.1. (We note that when a TPMS divides space into two congruent labyrinths, it is known as a balanced TPMS and it can be described by a group-subgroup pair of space groups;[133] we refer to the subgroup which describes the symmetry of the oriented surface, i.e., the space group that distinguishes the surface's two sides.)

However, a space group does not uniquely define a TPMS, and so the target TPMS would not be the only one that could form from zeolites with the same space group. To understand which structures are more likely to be formed, we must consider the topology of TPMSs in addition to their symmetry. A useful measure of a surface's topology is its genus, which is the number of handles the surface contains. For example, known TPMSs with space group Pm$\bar{3}$m include the P, C(P), C(P)b, C(P)a, Pb, and Pa TPMSs of genus 3, 9, 15, 19, 21, and

25, respectively.[134] Higher genus surfaces have more tunnels in the unit cell, the formation of which require correspondingly more channels in the unit cell of the zeolite template. The bond length and angle constraints of zeolites prevent a large number of channels without requiring a larger unit cell, yet zeolites tend to be very regular materials with high symmetry and small unit cells. Thus, we would expect zeolites to be more likely to form low genus TPMSs. Indeed, we observe that the TPMSs from all known zeolites have a low genus of 3, 4, or 5 (Table 3.1). Hence, provided that the target TPMS is of sufficiently low genus, we expect to be able to solve the inverse problem within the zeolite material class.

To illustrate these points, let us attempt to solve the inverse problem for a set of target TPMSs. A natural choice of targets are the "minimal minimal surfaces," so called because they have genus 3, the smallest possible genus for a TPMS.[127] Only five minimal minimal surfaces are known; these are the Schwarz D, CLP, H, and P TPMSs, and the Schoen G TPMS (deformations of which admit lower-symmetry variants like the tD, mDCLP, and rPD TPMSs).[125,127,130] By examining Table 3.1, we see that there already exist zeolites that can generate the D and CLP TPMSs, but corresponding zeolites are not present for the H, P, and G TPMSs. It is an interesting question as to whether all of the minimal minimal surfaces can be generated using zeolite templates, since these surfaces play a relevant role in the science of materials,[126,127] and they have been examined as schwarzites in the most depth.[89–93] As a motivating example, Qin *et al.*[139] have shown that a schwarzite resembling the Schoen G TPMS has interesting properties, but they could not suggest how to synthesize it; here, we demonstrate how to find a material that would template it.

The first step is to seek zeolites with a space group matching the desired surface, e.g., P$\bar{6}$m2 for the H TPMS, which also must have sufficiently large $D_{f,3p}$. Out of the 1.6 million hypothetical zeolite structures, we found 476 with the P$\bar{6}$m2 space group, of which 23 had a $D_{f,3p}$ greater than 5.0 Å. We evaluated the first 6 materials for which our ZTC-generating algorithm converged on final structures, and we found that 5 of these ZTCs had **hms** labyrinth nets and thus resembled the H TPMS, while 1 instead resembled the G-W TPMS (Fig. 3.8 and Table A.1). The ZTC resembling the G-W TPMS exhibits the same increase in symmetry from zeolite to TPMS as was observed with ZTC-POS (the G-W TPMS has space group P6$_3$/mmc, which is a supergroup of P$\bar{6}$m2); this result demonstrates that it is expedient to first determine the likely labyrinth nets by generating the accessible zeolite surface using inexpensive geometric methods[121] prior to generating the full atomistic ZTC structure.

Similarly, we can find corresponding zeolites for the remaining P and G TPMSs by matching their respective space groups, Pm$\bar{3}$m and I4$_1$32. Out of the hypothetical zeolite structures, we found 23 and 4 with these space groups, of which 4 and 2 had a $D_{f,3p}$ greater than 5.0 Å. From these, the first materials for which our ZTC-generating algorithm converged on a final structure did resemble the P and G TPMSs, having **pcu** and **srs** labyrinth nets (Fig. 3.8 and Table A.1). Thus, we see that we are indeed capable of solving the inverse problem.
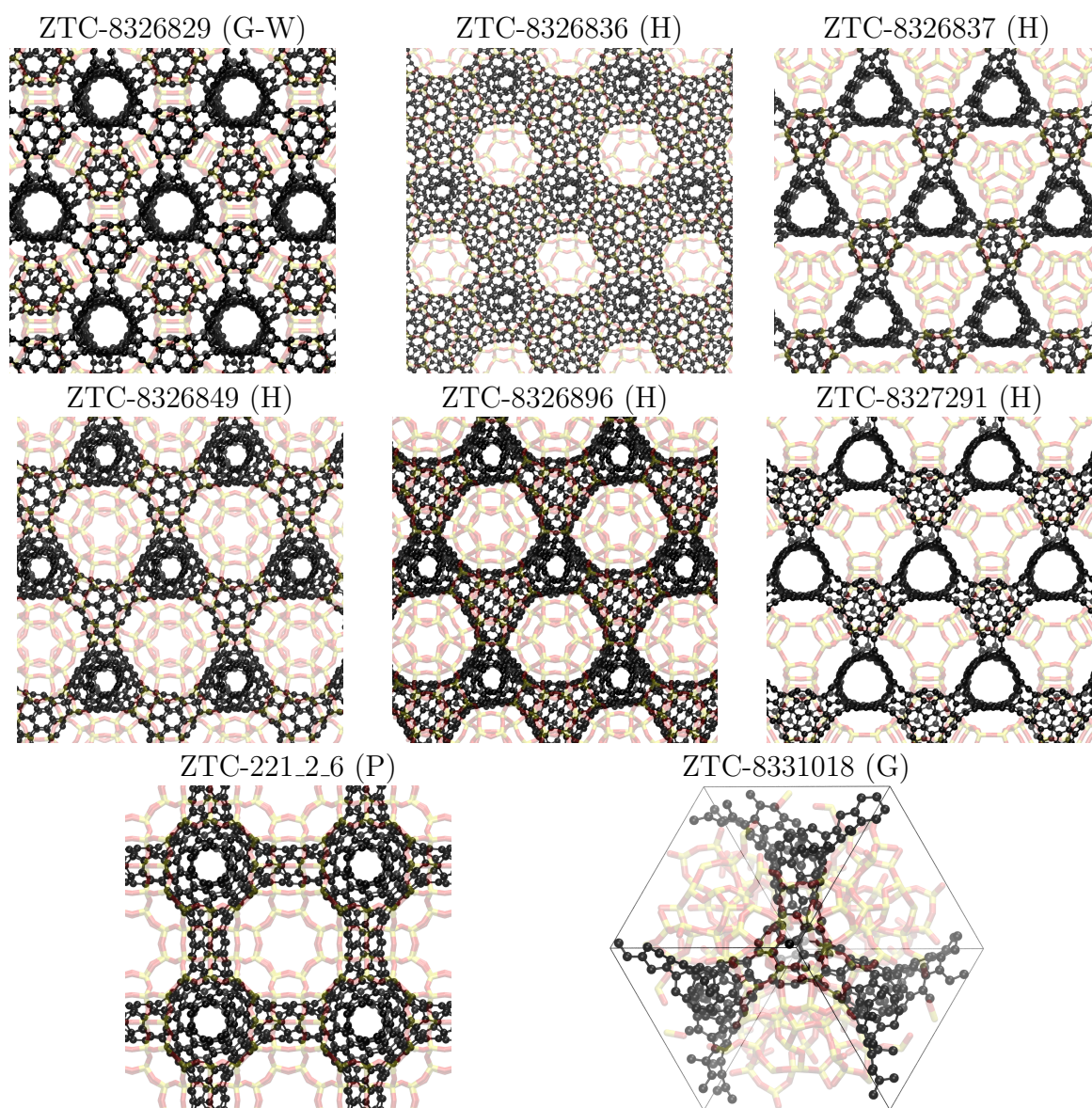
Figure 3.8: **ZTCs synthesized *in silico* from hypothetical zeolites with the same space groups as the H, P, and G minimal minimal surfaces.** Shown are the ZTC structures (black carbon atoms) along with their parent zeolites (red oxygen and yellow silicon atoms). The TPMS each structure resembles is given in parentheses. The top two rows of ZTCs were generated from zeolites with space group P$\bar{6}$m2, ZTC-221_2_6 was generated from a zeolite with space group Pm$\bar{3}$m, and ZTC-8331018 was generated from a zeolite with space group I4$_1$32. Zeolite 221_2_6 was taken from Treacy and Foster's silver hypothetical zeolite database, and the other zeolites were taken from Deem's SLC hypothetical zeolite database.[137,138]

## 3.4 Concluding remarks

In this work we have developed an *in silico* method to generate ZTCs from any zeolite structure. Our method correctly describes the structures of the known ZTCs and allows us to present a complete library of ZTCs that ought to be synthesizable from known zeolites; this library should serve to guide experimentalists and to provide computational scientists with realistic atomistic models. Furthermore, we have shown how ZTCs can be associated with TPMSs, linking the research topics of ZTCs and schwarzites. Finally, we have demonstrated how to find template materials that would yield a desired schwarzite.

An interesting observation is that all known zeolites form low genus surfaces, while in the databases of hypothetical zeolites one can find structures that give surfaces with relatively high genus. For example, Treacy and Foster's hypothetical zeolite 225_6_1852 has one of the largest unit cell volumes,[137] and the genus of the corresponding ZTC surface would be 11. In fact, one can envision generating hypothetical zeolite structures with even higher genus by simply replicating an existing zeolite's unit cell and then breaking the symmetry. A tempting speculation is that a large genus ZTC surface might itself be an indication of the difficulty of synthesizing its parent zeolite. Some connection might be made here between ZTCs and the structure-directing agents (SDAs) that frequently direct zeolite synthesis. These SDAs sometimes act as soft templates around which zeolites crystallize,[140] making ZTCs a sort of inverse-inverse material. The self-assembly of surfactants has also been described using the TPMS concept.[126] Thus, the synthesis of high genus TPMSs might require the synthesis of zeolites which in turn require the use of SDAs that self-assemble into complicated structures.

By focusing on the striking similarities between ZTCs and schwarzites, we have found that theories about TPMSs give insight into ZTCs, useful for such problems as how to select a zeolite template to obtain a microporous ordered carbon with a particular pore topology. It is important to recall that a schwarzite has been a purely hypothetical concept of a novel form of carbon, while our work suggests that ZTCs are schwarzites incarnate. One conceivable difference between these material classes is that the carbon atoms of ZTCs deviate from the TPMS surface, whereas schwarzites were generated by placing carbon atoms exactly on the TPMS surface. However, this difference is in quantity rather than in quality, as energetic relaxation of the schwarzites also causes a deviation from the true TPMS.[90,92] A second difference is that past works hypothesized carbon schwarzites without regard to the method by which they can be synthesized, assuming that structures with low energies, high stability, and high symmetry will eventually be found. In contrast, our ZTC structures have been made with a direct synthetic pathway in mind, which may lead to more success in guiding experimental efforts.

## 3.5 Methods

*ZTC generation* was performed using an in-house MC code. All known zeolite structures were taken from the IZA database[120]; two exceptions to this were FAU, which was taken

as the zeolite atoms in the XRD-derived structure of the carbon-loaded zeolite from Kim *et al.*[110] to allow for better visual comparison with their experimental model, and BEB, which was taken from the ICSD[141] database entry of Martínez-Iñesta *et al.*[142]. The hypothetical zeolite structures were taken from Treacy and Foster's silver hypothetical zeolite database and Deem's SLC hypothetical zeolite database.[137,138]

The generation of a ZTC is started with a seed of four carbon atoms (a central atom bonded to three neighbors) placed in the zeolite's pore space. Two types of MC moves were then attempted: moves that insert/delete carbon atoms and moves that change the carbon bonding network. Insertion moves were performed by first selecting an existing under-coordinated carbon atom, and, depending on the coordination of the selected carbon atom, adding either one or two new carbon atoms. The new carbon atoms were placed in an empty space in a shell around the selected carbon atom and were bonded to this atom. Additional bonds were added between the inserted atoms and other neighboring under-coordinated carbon atoms. Deletion moves were done by first selecting an existing under-coordinated carbon atom and removing the atom and its one or two bonds; if the deleted atom had two bonds, a new bond was created between the two atoms it was formerly bonded to.

Moves that change the bonding network include bond propagation, bond disassociation, and bond switching. The bond propagation move involves three carbon atoms, two that are bonded and a non-bonded neighbor. The bond between the two carbon atoms is deleted, and a new bond is placed between one of them and the neighbor. The bond disassociation move is deletion of a bond. The bond switching move involves four neighboring carbon atoms, between which the bonding network is reorganized.

In our MC procedure, we randomly select to insert, delete, propagate, or disassociate, and each trial move includes a subsequent energetic relaxation of all carbon atoms. This relaxation involves finding the optimum bonding network by bond switching moves, followed by a conjugate-gradient relaxation of the carbon atomic positions. The overall trial move is then accepted with probability $\min\left[1, \exp\left(-\beta\Delta E\right)\right]$, where $\Delta E$ is the energy difference between the old structure and the relaxed trial structure. In order to surmount energetic barriers, it is important to set the effective temperature sufficiently high;[104] we found $2{,}000\,\mathrm{K}$ to work well. The MC simulation was ended when all carbon atoms had three bonds and no significant reduction in energy was being observed. Changing the temperature and other variables in our algorithm does not affect the overall surface formed.

The valence force field of Lee and Hwang[143] was used to model carbon-carbon interactions. The exact parameters are from a slightly improved set provided by Lee and Hwang: $r_0 = 1.41239\,\text{Å}$, $\theta_0 = 120°$, $k_r = 20.559\,\mathrm{eV}$, $k_\theta = 3.5125\,\mathrm{eV}$, $k_\phi = 0.561735\,\mathrm{eV}$, and $k_\varphi = 0.0081\,\mathrm{eV}$. This force field was chosen because it was designed specifically for $sp^2$-hybridized carbon and it features an energetic penalty for misalignment of neighboring $\pi$ orbitals, making it both more descriptive of ZTC carbon atoms and less computationally expensive than the general-purpose bond order force field AIREBO.[123,143,144] (We initially attempted to use the AIREBO force field to model carbon-carbon interactions in a GCMC simulation, but we found that with no disincentive against the formation of a non-planar structure, and with the carbon bonding strength orders of magnitude greater than the weak

physisorption interactions between the carbon atoms and the zeolite, there was a very small range of acceptable chemical potentials in which carbon neither undertemplates the zeolite surface nor fills in the zeolite's void space.) Carbon-framework interactions were described by Lennard-Jones interactions using Universal force field parameters with Lorentz-Berthelot mixing rules and a simply truncated cutoff of 7.0 Å.[145]

Our criteria for considering a ZTC experimentally-accessible included the absence of defects such as single atom bridges and flat sheet connectors. For some zeolites, we noticed that a ZTC model would contain defective features on some generations and be free of them on others, indicating a free energy barrier to assembly. Thus, we chose our criteria to be that at least one model free of these features had to be obtained for a ZTC to be considered experimentally-accessible, and we generated additional ZTC models in borderline cases. For example, zeolites IRR, IWS, and SBT all gave defect-free ZTCs on a second attempt, while we could not obtain defect-free ZTCs for zeolites JSR, SVR, and TUN over several attempts. Zeolites with $D_{f,3p}$ close to 5 Å for which we were unable to find a defect-free ZTC model might be obtainable with a large number of trials, but since ZTCs would be difficult to synthesize from these templates due to CVD precursor diffusion limitations, we considered them less likely to be experimentally-accessible, and so we chose to focus on ZTCs formed from zeolites with $D_{f,3p}$ above 5 Å.

*DFT calculations* were performed with CP2K version 4.1,[146,147] using the PBE density functional[148] and Grimme's D3 dispersion correction.[149] Integration over the Brillouin zone was carried out over a Γ-centered Monkhorst-Pack grid, where the number of subdivisions along each reciprocal lattice vector was given by $\max\left(1, \text{floor}\left(25|b_i| + 0.5\right)\right)$, where $|b_i|$ is the norm of the $i$th reciprocal lattice vector; this formula gave a 2×2×2 grid for ZTC-FAU, which was found to be sufficient for k-point convergence. We used double-$\zeta$ shorter range Gaussian basis sets,[150] plane-wave basis sets cutoff at 300 Ry, and the Goedecker-Teter-Hutter pseudopotentials.[151,152] The wavefunction energy convergence criterion was set to $1 \times 10^{-6}$, and the atomic positions, unit cell shape, and unit cell volume were optimized simultaneously until all forces were smaller than $4.5\, E_h\, a_0{}^{-1}$. The buckminsterfullerene energy calculation was conducted by placing a single molecule in a cubic unit cell of length 40 Å to approximate isolation.

*MD simulations* were conducted with the August 11, 2017 release of LAMMPS.[59] The AIREBO[123] force field was used with no charges and a cutoff of 10.2 Å. The ZTC unit cells were replicated to ensure that the three perpendicular widths of the simulation boxes were all greater than 20.4 Å. The timestep used was 1 fs. The equations of motion were integrated with a standard velocity Verlet algorithm using half-step velocity calculations. Following minimization of the system energy by adjusting atomic coordinates, the MD simulation was run for 1 ns in the NVT ensemble, an additional 1 ns in the NPT ensemble with only the three lattice constants allowed to change, and a final 1 ns in the NPT ensemble with the three lattice constants and the three angles between them allowed to change. A Nosé-Hoover thermostat[28-30] was used with a time damping constant of 100 fs and a total of three chained thermostats. A Nosé-Hoover barostat[153-155] was used with a time damping constant of 1,000 fs, and it was thermostatted with a chain of three thermostats.

*Topological analysis* was conducted with ToposPro version 5.3.0.2.[156] The two nets that describe the ZTCs's two labyrinths were obtained by taking the natural tiling of the parent zeolite,[157] determining the dual net of this tiling using ToposPro,[156] and decomposing the dual net into two interpenetrated subnets by deleting edges that cross the ZTC surface. In some cases this construction leads to a labyrinth net that is disconnected, and the missing edges were added. Subsequently, both labyrinth nets were simplified in the following manners. When an edge was found to be incident to 2-coordinated vertices, the edge was contracted. When a net was found to contain a strong ring (a cycle that is not a sum of smaller cycles)[156] that bounds a face wholly contained within its labyrinth, the ring was replaced by a single vertex located at the ring's centroid, which was connected to the vertices that were formerly connected to the ring. Similarly, when a net was found to contain several strong rings with shared edges that each bounds a face wholly contained within their labyrinth, the set of rings was replaced by a single vertex as described above; for example, six strong rings making up the edges of a cube were replaced by the cube's centroid, and two strong rings whose sum forms the boundary of a quadrilateral were replaced by the quadrilateral's centroid. The resulting nets were assigned names according to the Reticular Chemistry Structure Resource (RCSR) nomenclature when available (e.g., **dia**, **lon**),[129] Epinet nomenclature otherwise (sqc145),[158] and TTD Topos database nomenclature when neither of the first two were available (4,6T585).[159] Our method provides an unambiguous assignment of the nets, and it captures the nets' symmetry embeddings. In general, different nets can be used to trace a given labyrinth,[127] so although our method gives a valid choice of nets which captures the geometry of the labyrinths well for our purpose, other choices can equally well lie in the labyrinth, e.g., ZTC-SAO's void labyrinth net can be properly traced by a $I\bar{4}m2$ embedding of both the **tfa** and **dia** nets, and the multiple labyrinths described by the **ths** net can also be traced by the **dia** net.

The $D_{f,3p}$ of zeolites were calculated using Zeo++[121] version 0.3 with silicon and oxygen atomic radii both set to 1.35 Å for consistency with the sphere diameters calculated in the IZA database.[120] Powder XRD patterns were calculated with Mercury version 3.9 using the Cu $K\alpha_1$ wavelength of 1.54056 Å.[160] Numerical surface area minimization was conducted with the Surface Evolver version 2.70,[161] for which input files of the Schwarz D and Schoen G-W TPMSs were taken from Ken Brakke's website.[162]

# Chapter 4

# High-throughput screening: finding optimal zeolites for carbon dioxide capture from natural gas[*]

With the growth of natural gas as an energy source, upgrading $CO_2$-contaminated supplies has become increasingly important. Here we develop a single metric that captures how well an adsorbent performs the separation of $CH_4$ and $CO_2$, and we then use this metric to computationally screen tens of thousands of all-silica zeolites. We show that the most important predictors of separation performance are the $CO_2$ heat of adsorption ($Q_{st,CO_2}$) and the $CO_2$ saturation loading capacity. We find that a higher-performing material results when the absolute value of the $CH_4$ heat of adsorption ($Q_{st,CH_4}$) is decreased independently of $Q_{st,CO_2}$, but a correlation that exists between $Q_{st,CH_4}$ and $Q_{st,CO_2}$ in all-silica zeolites leads to incongruity between the objectives of optimizing $Q_{st,CO_2}$ and minimizing $Q_{st,CH_4}$, rendering $Q_{st,CH_4}$ nonpredictive of separation performance. We also conduct a large-scale analysis of IAST by comparing results obtained using directly-generated mixture isotherms to those obtained using IAST; IAST appears adequate for the purposes of establishing performance trends and structure-property relationships in a high-throughput manner, but it must be tested for validity when analyzing individual adsorbents in detail since it can produce significant errors for materials in which there is site segregation of the adsorbate species.

## 4.1 Introduction

As discussed in Chapter 1, natural gas must have its major contaminants—such as $CO_2$—removed prior to transportation. In order to find the optimal adsorbents for this and other gas separations, metrics are necessary to quantify an adsorbent's separation performance. Two of the most widely used metrics are the adsorption selectivity, defined as the ratio of the loading mole fractions divided by the ratio of the gas phase mole fractions, and the working

---

[*]This chapter is based on Braun, Zurhelle, Thijssen, Schnell, Lin, Kim, Thompson, and Smit [163].

capacity for each of the adsorbates, defined as the difference in loading of each adsorbate at adsorption and desorption conditions. These metrics are easily calculated from mixture adsorption isotherms, and many adsorbent screening projects have examined these metrics individually to evaluate an adsorbent at given adsorption and desorption conditions.[164–168] However, the selectivity and working capacities are not necessarily representative of the economic drivers that chemical process designers actually consider;[169,170] in fact, they can even be deceptive, as increasing the selectivity of an already highly selective material may provide no true improvement in performance, and recent process modeling studies have found no significant correlation between selectivity and process cost.[171,172]

Other metrics have been proposed as being more appropriate for the comparison of adsorbents. Rege and Yang[173] and Wiersum *et al.*[174] suggested multiplying the selectivities and working capacities in different manners to obtain metrics they called the Sorbent Selection Parameter ($S$) and the Adsorbent Performance Indicator (API). As we will show, these metrics can lead to undesirable results since they directly include the selectivity, which is unbound and can approach infinity. Some studies have created metrics that aim to provide a more direct representation of the process economics. For example, in the $CO_2/N_2$ gas separation being investigated for carbon capture and storage (CCS) from the flue gas of fossil-fuel powered power plants, the economic driver is the energy cost of the separation, which has been estimated to parasitically reduce the net output of power plants by $30\%$.[169] This parasitic energy has been simply estimated as a sum of the thermal energy required to desorb the adsorbate and of the compression work by Lin *et al.*[169], and Maring and Webley[170] has developed a simplified process model to estimate the power required for the vacuum pump. Of course, a full process modeling analysis provides the most realistic cost estimates, but such detailed process modeling can be computationally expensive and thus unfeasible for a high-throughput screening study.[171,172] Additionally, such process modeling requires expertise beyond what many researchers studying adsorption science possess.[170]

In this contribution, we first develop a single simple metric that is representative of the economic drivers behind the $CH_4/CO_2$ gas separation required for upgrading various natural gas sources to pipeline quality, which we call the Separation Performance Parameter (SPP). We then use this metric to explore the separation performance and structure-property relationships of tens of thousands of all-silica zeolites taken from the IZA database[120] and the predicted crystallography open database (PCOD) of hypothetical zeolites,[138] using a high-performance GPU GCMC code to directly-generate mixture isotherms.[4,175–179] Finally, we take the opportunity afforded by the large amount of data collected in this study to evaluate the accuracy of IAST[180], a prominent technique to generate mixture adsorption isotherms.

We considered three different sources of natural gas, all at a temperature of $300\,K$ and assumed to be a binary mixture of $CH_4$ and $CO_2$: (1) landfill gas (LFG), produced by microorganisms acting upon municipal wastes, assumed to contain $60\,mol\%$ $CH_4$ at a total pressure of $1\,bar$, (2) associated petrolum gas (APG), found in oil reservoirs, assumed to contain $90\,mol\%$ $CH_4$ at a total pressure of $7\,bar$, and (3) non-associated gas (NAG), found apart from oil, assumed to contain $90\,mol\%$ $CH_4$ at a total pressure of $70\,bar$.[17] These pressures and temperatures were used as the adsorption conditions. As different materials perform

optimally at different conditions, we evaluated each material at desorption conditions that optimize its SPP to allow for fair comparison.

## 4.2 Methods

The all-silica zeolites frameworks were selected from the 110 orthogonal structures in the IZA database[120] and the orthogonal structures in the predicted crystallography open database (PCOD)[138]. The PCOD database[138] was reduced to a set of 121,966 structures by removing those with a largest free-sphere diameter below 2.25 Å as these are less accessible to $CO_2$ (Figs. A.13–A.14); of these, a randomly chosen subset of 81,526 structures was screened. Note that some of the results for the IZA zeolites appear outside the range of the hypothetical zeolite results, which may be due to the energy minimization step used in the construction of the hypothetical zeolite database.[138] The geometric structure descriptors of largest included sphere diameter, largest free sphere diameter, and accessible surface area was obtained using the open-source software Zeo++,[121] using its high accuracy setting,[181], a $CH_4$ probe of 1.625 Å radius, a silicon atom radius of 2.10 Å, and an oxygen atom radius of 1.52 Å.

Mixture adsorption isotherms were produced both directly by GCMC simulation[4] and from the simulated pure-component isotherms using IAST.[180] Except for where the two results were compared, the directly-generated mixture isotherms were used throughout this paper. All isotherms were generated with a GPU GCMC code that has been described elsewhere,[175,176] which uses a parallel flood fill algorithm to find blocked pockets inaccessible from the gas phase[175] and uses density-biased sampling to accelerate convergence.[176] The force field developed by García-Pérez *et al.*[177] was used, which consists of Lennard-Jones and Coulombic terms for guest-guest and guest-host interactions. The host framework atoms were assumed to be rigid, and the number of simulated unit cells was chosen such that the simulation box extended at least twice the cutoff radius of 12 Å. Our work exclusively used units of absolute loading.

*Directly generating mixture isotherms:* The GPU GCMC code was adapted to allow for multiple adsorbate species, and blocked pockets were separately found and applied to $CH_4$ and $CO_2$. For all zeolites, the number of equilibration and production steps, respectively, were set to 10,000,000 and 1,000,000 for the APG and NAG processes for which isotherms were simulated up to 100 bar, and 5,000,000 and 1,000,000 for the LFG process for which isotherms were simulated up to 5 bar. These numbers of steps were validated as being sufficient for isotherm convergence[176] at the pressures involved for screening the three processes studied in this paper by comparison to isotherms created using a conventional CPU-based code (Figs. A.24–A.29).

*Generating mixture isotherms by IAST:* Pure-component isotherms for the IZA zeolites were generated using 5,000,000 equilibration and 3,000,000 production steps, pure-component isotherms for the hypothetical zeolites with a largest free-sphere diameter less than 3.75 Å were generated using 1,250,000 equilibration and 500,000 production steps, and pure-component isotherms for the hypothetical zeolites with a largest free-sphere diame-

ter greater than $3.75\,\text{Å}$ were taken from previous work.[168,182] The number of steps used to generate pure-component isotherms for the hypothetical zeolites with a largest free-sphere diameter less than $3.75\,\text{Å}$ were chosen to achieve a similar degree of convergence as the pure-component isotherms taken from previous work. The pure-component isotherms were then fit via the method of least squares to single- or dual-site Langmuir isotherms for each adsorbate based on which fit gave a larger adjusted $R^2$ value for that adsorbate. The spreading pressure was then calculated via analytical integration of equation 19 of Myers and Prausnitz[180]. As this procedure sometimes requires the fitted isotherms to be evaluated at pressures above the highest pressure point of the simulated isotherms, we only allowed such extrapolation for structures which had fitted saturation loading capacities less than ten times the amount of the loading at the highest simulated pressure point; the remainder were not used in comparing IAST results to directly-generated mixture isotherm results. The fitted saturation loading capacities of these pure-component adsorption isotherms were used in plots in this paper as stated (fitted Henry coefficients were not used in plots as we instead used values obtained from Widom insertions; see below). When a dual-site Langmuir isotherm was used, the sum of the two saturation loading capacities was used.

The Peng-Robinson equation of state[178] was used to convert between pressure and fugacity, with the critical temperatures, critical pressures, and acentric factors of CH$_4$ and CO$_2$ being $190.6\,\text{K}$, $46.0\,\text{bar}$, $0.008$, $304.2\,\text{K}$, $73.76\,\text{bar}$, and $0.225$, respectively. For mixture isotherms generated directly by GCMC, the known total pressure and mole fractions were converted to component fugacities using the van der Waals mixing rules with a binary interaction parameter of $0.0919$, which were then used as inputs to the simulations. For pure-component isotherms used for IAST, the component's fugacity was input directly to the simulation and then converted to pressure prior to IAST calculations; however, in Fig. 8 fugacity was not converted to pressure prior to the IAST calculation to ensure consistency of units on the x-axis.

Widom insertions were performed by the GPU code[175,176] to calculate isosteric heats of adsorption, Henry coefficients, and helium void fractions, all at $300\,\text{K}$. Some zeolites have no enthalpically favorable adsorption sites for CH$_4$, either intrinsically or due to blocking; $Q_{\text{st,CH}_4}$ for these materials was set to $0\,\text{kJ}\,\text{mol}^{-1}$ for the purposes of plotting and for calculating the objective function. Isosteric heats of adsorption and Henry coefficients for all materials were calculated using 200,000 Widom insertions. For all materials, helium void fractions were calculated using 100,000 Widom insertions and the force field of Talu and Myers[179].

As different materials perform optimally at different conditions, we evaluated each material at its own optimal desorption conditions to allow for fair comparison. For each zeolite, the optimal desorption pressure (PSA), temperature (TSA), or both (PTSA) was found by minimization of the objective function value (OFV). We note that future studies may need not strictly find each adsorbent's optimal desorption conditions if desorption conditions can be well-approximated prior to the study's commencement, as we found that the SPP values for the adsorbents undergoing the three PSA processes at set desorption pressures are well-correlated with the SPP values for the adsorbents undergoing the three PSA processes at each adsorbent's optimal desorption pressure (Fig. A.30). For our study, we used the opti-

mal desorption conditions for all analyses except for the data presented in Fig. A.30 and for the comparison of the various metrics presented in Figs. 1 and A.1, since we found that the Sorbent Selection Parameter and Adsorbent Performance Indicator could not be used to find the optimal desorption conditions. SciPy version 0.15.1 was used to perform the minimization.[183] The adsorption pressures were set by the process, and the adsorption temperature was always 300 K. The desorption temperature was constrained to be greater than or equal to 300 K for TSA and PTSA and set to 300 K for PSA, while the desorption pressure was constrained to be between 0.1 bar and 1 bar for PSA and PTSA and set to 1 bar for TSA. The mole fractions were set by the process, and the same mole fraction was used for both adsorption and desorption. For all materials, mixture isotherms generated at 300 K were used, and for the IZA zeolites, upon which TSA and PTSA were also performed, mixture isotherms were also generated at higher temperatures going up in 5 K increments. Continuous optimization was applied to find the optimal desorption pressure, applying Akima cubic spline interpolation[184] as implemented by SciPy[183] to calculate loadings between the pressures that were simulated. Discrete optimization was applied to find the optimal desorption temperature in 5 K increments. For PTSA, pressure optimization was performed for all desorption temperatures, and the temperature that gave the lowest OFV was then chosen. When mixture isotherms gave a negative working capacity for either adsorbate (possible since mixture isotherms are not necessarily monotonically increasing functions), the working capacity for that adsorbate was set to $0\,\mathrm{mol\,kg^{-1}}$ prior to calculation of the metrics that go into the objective function.

To implement the random forest of decision trees regression algorithm and calculate the importances of the individual descriptors, we used open-source scikit-learn version 0.14.1-2.[185] Our forest included 1,000 trees, which has been shown to be an adequate number for random forest accuracy.[186] Nodes were expanded until all leaves were pure.

We have placed the computer code we developed for the screening online with the Open Science Framework, along with the pure and mixture isotherms and the screening results.[187] The code is capable of computing the SPP, Sorbent Selection Parameter, or Adsorbent Performance Indicator at optimized or set desorption conditions. Mixture isotherms can be input directly as loadings at multiple pressures, in which case the code will perform interpolation to calculate loadings, or mixture isotherms will be calculated by the code using IAST with pure-component isotherms input either as loadings at multiple pressures or as already-fitted dual-site Langmuir isotherms.

## 4.3 Metric development

We chose to model the natural gas upgrading process as one in which $CO_2$ is selectively adsorbed for several reasons. As $CO_2$ tends to adsorb onto zeolites more strongly than $CH_4$ due to its quadrupole moment, more zeolites will be capable of performing this separation, with about 90 % of the hypothetical zeolites having a $CO_2$ selectivity greater than 1 at conditions of 1 bar total pressure, 300 K, and 90 mol % $CH_4$. Furthermore, in making the raffinate the

high-value stream, there is no energetic cost to performing PSA down to atmospheric pressure since there is no need to compress the product,[17] unlike in CCS and other adsorption processes where the adsorbed gas is the desired product. Finally, in adsorbing the minor component in the processes we analyzed, cycle times and energetic costs are expected to be less.

Upgrading natural gas requires the examination of multiple costs to evaluate the economics of different adsorbents. The fixed costs vary with the size of the adsorption column and thus the amount of adsorbent required for the separation. The variable costs include the energy required for operating the adsorption column. A high fractional CH$_4$ recovery is also desirable to reduce the variable costs of feed natural gas as well as to reduce the loss of CH$_4$ to the atmosphere with the waste CO$_2$. If equilibrium adsorption and desorption are assumed, i.e., the mass transfer zone length approaches 0, these metrics can be individually calculated given mixture adsorption isotherms, as shown below. They can also be grouped into a single objective function to form a multiple-input single-output optimization problem, with the inputs being the desorption pressure (PSA), desorption temperature (TSA), or both (PTSA), and the output being the OFV. By varying the inputs to minimize the OFV, one finds the optimal desorption conditions for a given adsorbent.[183,184] The better-performing adsorbents will have a lower OFV, which leads us to term the inverse of the OFV the Separation Performance Parameter (SPP), which we use to evaluate the screening results.

It would seem that the objective function that most appropriately captures the process economics would be a summation of the capital and operating expenditures of the separations plant. However, our goal of creating a simple and nonephemeral metric (the value of which is not dependent on present market conditions) precludes such a calculation. Instead, we will assume that the mass of adsorbent, energy, and fractional CH$_4$ recovery are equally important economic drivers, and we will then go on to show that this assumption results in a metric that closely agrees with the economic results of a more detailed process simulation study. Thus we use as a definition of the OFV and SPP:

$$\text{OFV} = \frac{1}{\text{SPP}} = \frac{\left(\frac{M_{\text{ads}}}{M_{\text{CH}_4,\text{raff}}}\right) \times \left(\frac{E}{M_{\text{CH}_4,\text{raff}}}\right)}{\left(\frac{M_{\text{CH}_4,\text{raff}}}{M_{\text{CH}_4,\text{feed}}}\right)} \tag{4.1}$$

where $M_{\text{ads}}$ is the mass of the adsorbent, $M_{i,k}$ is the moles of species $i$ in stream $k$, $E$ is the total energy required for the separation, "feed" is the adsorption column inlet stream, and "raff" (raffinate) is the adsorption column outlet stream consisting of the gas that did not adsorb which contains the recovered CH$_4$. Note that here and throughout this paper, we use units of mass for the adsorbent and units of moles for the adsorbates. The three terms in Eq. 4.1 are given below, followed by their derivation.

The first term in Eq. 4.1 is the mass of adsorbent required per mole of CH$_4$ captured in one batch adsorption-desorption cycle:

$$\frac{M_{\text{ads}}}{M_{\text{CH}_4,\text{raff}}} = \frac{y_{\text{CH}_4,\text{raff}} - y_{\text{CH}_4,\text{feed}}}{y_{\text{CH}_4,\text{raff}} \left[y_{\text{CH}_4,\text{feed}} \left(\Delta q_{\text{CH}_4} + \Delta q_{\text{CO}_2}\right) - \Delta q_{\text{CH}_4}\right]} \tag{4.2}$$

where $y_{i,k}$ is the mole fraction of species $i$ in stream $k$ and $\Delta q_i$ is the working capacity for species $i$. Since this is not the way that adsorption columns are sized in practice, $\frac{M_{ads}}{M_{CH_4,raff}}$ can alternatively be thought of as a proxy for the cycle time of a batch when comparing equally-sized columns. Eq. 4.2 may be further understood in the context of two extremes. A perfectly selective adsorbent will have $\Delta q_{CH_4} = 0$, and if $y_{CH_4,raff}$ is set to 1, then $\frac{M_{ads}}{M_{CH_4,raff}}$ is simply equal to $\frac{y_{CO_2,feed}}{y_{CH_4,feed}\Delta q_{CO_2}}$. At the other extreme, as a material's ratio of $\frac{\Delta q_{CH_4}}{\Delta q_{CO_2}}$ increases, approaching the ratio of $\frac{y_{CH_4,feed}}{y_{CO_2,feed}}$, less and less separation will be performed and $\frac{M_{ads}}{M_{CH_4,raff}}$ will approach infinity. Materials with a working capacity ratio exceeding this limit will have a required mass of adsorbent that is negative, and cannot perform the given separation.

If $y_{CH_4,raff}$ is set to 1, one obtains the mass of adsorbent required to capture a mole of $CH_4$ in an adsorption column operated until the column is completely saturated. Alternatively, one may set $y_{CH_4,raff}$ to the purity required for a specific application, which corresponds to an adsorption column with a mass transfer zone length approaching 0 that is operated beyond initial breakthrough in order to obtain more product which on average still meets its specifications. The latter option was chosen for this study, and $y_{CH_4,raff}$ was set to $98 \, mol \, \%$ $CH_4$.

The second term in Eq. 4.1 is the energy required per mole of $CH_4$ captured:

$$\frac{E}{M_{CH_4,raff}} = \frac{M_{ads}}{M_{CH_4,raff}} [C_p(T_{des} - T_{ads})$$
$$+ \Delta q_{CH_4} Q_{st,CH_4} + \Delta q_{CO_2} Q_{st,CO_2}] + \frac{W_{vac}}{M_{CH_4,raff}} \qquad (4.3)$$

where $C_p$ is the heat capacity of the adsorbent (assumed to be $0.75 \, kJ \, kg^{-1} \, K^{-1}$ for all zeolites)[188], $T_{des}$ is the desorption temperature, $T_{ads}$ is the adsorption temperature, $Q_{st,i}$ is the isosteric heat of adsorption at zero loading of species $i$, and $W_{vac}$ is the work required to pull vacuum. Here and throughout this work, $Q_{st}$ refers to the negative value of the enthalpy of adsorption such that positive values are shown. The energy term in Eq. 4.3 consists of the sensible energy to raise the temperature of the bed during a temperature swing, the energy to desorb adsorbates, and the work to pull vacuum on the adsorption column. We approximate the work required to pull vacuum below 1 bar as the energy required to compress the desorbed vapor from its vacuum pressure to atmospheric pressure using the formula for isentropic compression:[170,171,189]

$$\frac{W_{vac}}{M_{CH_4,raff}} = \begin{cases} \frac{1}{\eta} RT_{des} \frac{\gamma}{\gamma-1} \left[ \left( \frac{P_{atm}}{P_{des}} \right)^{\frac{\gamma-1}{\gamma}} - 1 \right] (\Delta q_{CH_4} + \Delta q_{CO_2}) \frac{M_{ads}}{M_{CH_4,raff}}, \\ \qquad \text{if } 0.1 \text{ bar} < P_{des} < 1 \text{ bar} \\ 0, \\ \qquad \text{if } 1 \text{ bar} < P_{des} \end{cases} \qquad (4.4)$$

where $\eta$ is the vacuum pump efficiency (assumed to be 75 %), $R$ is the gas constant, $\gamma$ is the heat capacity ratio (assumed to be 1.3 for both $CH_4$ and $CO_2$), $P_{atm}$ is atmospheric pressure

(approximated as 1 bar), and $P_{\text{des}}$ is the desorption pressure.[190] Eq. 4.4 was considered valid down to a desorption pressure of 0.1 bar; at lower pressures, the equation may fail as the efficiency falls below 75 %.[189]

The third term in Eq. 4.1 is the fractional $CH_4$ recovery:

$$\frac{M_{CH_4,\text{raff}}}{M_{CH_4,\text{feed}}} = \frac{y_{CH_4,\text{raff}}\left[y_{CH_4,\text{feed}}(\Delta q_{CH_4} + \Delta q_{CO_2}) - \Delta q_{CH_4}\right]}{y_{CH_4,\text{feed}}\left[y_{CH_4,\text{raff}}(\Delta q_{CH_4} + \Delta q_{CO_2}) - \Delta q_{CH_4}\right]} \tag{4.5}$$

Using Eqs. 4.2–4.5 we can calculate the OFV and SPP in Eq. 4.1 at a given desorption pressure and desorption temperature. The components of the objective function given by these equations can be individually evaluated, but they cannot necessarily be individually used as the full objective function to find the optimal desorption conditions because this can result in desorption conditions that are at the most extreme values allowed. For example, if one were to attempt to minimize the mass of adsorbent metric for a PSA process, one would arrive at a desorption pressure of 0.1 bar for every material as there would be no energetic penalty.

We have placed the computer code we developed for the screening online with the Open Science Framework, along with the pure-component and mixture isotherms and the screening results.[187] The code is capable of computing the SPP at optimized or set desorption conditions either by using input mixture isotherms or by using input pure-component isotherms to calculate mixture isotherms with IAST.

## Metric derivation

The derivation of the terms in the SPP begins by calculating the mole fraction of $CH_4$ in the raffinate stream:

$$
\begin{aligned}
y_{CH_4,\text{raff}} &= \frac{M_{CH_4,\text{raff}}}{M_{CH_4,\text{raff}} + M_{CO_2,\text{raff}}} \\[8pt]
&= \frac{M_{CH_4,\text{feed}} - \Delta q_{CH_4} M_{\text{ads}}}{(M_{CH_4,\text{feed}} - \Delta q_{CH_4} M_{\text{ads}}) + (M_{CO_2,\text{feed}} - \Delta q_{CO_2} M_{\text{ads}})} \\[8pt]
&= \frac{1 - \Delta q_{CH_4}\frac{M_{\text{ads}}}{M_{CH_4,\text{feed}}}}{(1 - \Delta q_{CH_4}\frac{M_{\text{ads}}}{M_{CH_4,\text{feed}}}) + \left(\frac{y_{CO_2,\text{feed}}}{y_{CH_4,\text{feed}}} - \Delta q_{CO_2}\frac{M_{\text{ads}}}{M_{CH_4,\text{feed}}}\right)} \\[8pt]
&= \frac{1 - \Delta q_{CH_4}\frac{M_{\text{ads}}}{M_{CH_4,\text{feed}}}}{\frac{1}{y_{CH_4,\text{feed}}} - \frac{M_{\text{ads}}}{M_{CH_4,\text{feed}}}(\Delta q_{CH_4} + \Delta q_{CO_2})}
\end{aligned}
$$

Rearranging, we obtain:

$$\frac{M_{\text{ads}}}{M_{CH_4,\text{feed}}} = \frac{y_{CH_4,\text{raff}} - y_{CH_4,\text{feed}}}{y_{CH_4,\text{feed}}\left[y_{CH_4,\text{raff}}(\Delta q_{CH_4} + \Delta q_{CO_2}) - \Delta q_{CH_4}\right]} \tag{4.6}$$

We then calculate the moles of recovered $CH_4$ per moles of feed $CH_4$:

$$
\begin{aligned}
\frac{M_{CH_4,raff}}{M_{CH_4,feed}} &= \frac{M_{CH_4,feed} - \Delta q_{CH_4} M_{ads}}{M_{CH_4,feed}} \\
&= 1 - \Delta q_{CH_4} \frac{M_{ads}}{M_{CH_4,feed}} \\
&= 1 - \Delta q_{CH_4} \left( \frac{y_{CH_4,raff} - y_{CH_4,feed}}{y_{CH_4,feed} \left[ y_{CH_4,raff} \left( \Delta q_{CH_4} + \Delta q_{CO_2} \right) - \Delta q_{CH_4} \right]} \right) \\
&= \frac{y_{CH_4,raff} \left[ y_{CH_4,feed} \left( \Delta q_{CH_4} + \Delta q_{CO_2} \right) - \Delta q_{CH_4} \right]}{y_{CH_4,feed} \left[ y_{CH_4,raff} \left( \Delta q_{CH_4} + \Delta q_{CO_2} \right) - \Delta q_{CH_4} \right]}
\end{aligned}
\tag{4.7}
$$

Finally:

$$
\begin{aligned}
\frac{M_{ads}}{M_{CH_4,raff}} &= \frac{M_{ads}}{M_{CH_4,feed}} \div \frac{M_{CH_4,raff}}{M_{CH_4,feed}} \\
&= \frac{y_{CH_4,raff} - y_{CH_4,feed}}{y_{CH_4,raff} \left[ y_{CH_4,feed} \left( \Delta q_{CH_4} + \Delta q_{CO_2} \right) - \Delta q_{CH_4} \right]}
\end{aligned}
\tag{4.8}
$$

Since $\frac{M_{CH_4,raff}}{M_{CH_4,feed}}$ is included as part of $\frac{M_{ads}}{M_{CH_4,raff}}$, which in turn is included in the equation for SPP, one could reasonably exclude the fractional $CH_4$ recovery term from appearing directly in the equation for SPP. However, a high fractional $CH_4$ recovery is desirable for more than direct economic costs, as any $CH_4$ not captured will likely be emitted and can be considered to be an environmental cost. Thus, we choose to directly include the fractional $CH_4$ recovery term in our equation for SPP.

## Separation Performance Parameter validity

At this point it is important to mention that a requirement imposed during the development of the SPP was that it be practical to compute the metric for thousands of materials, which precludes a full process design for all zeolites. We intended for the SPP to be representative of the most important economic drivers behind the $CH_4/CO_2$ gas separation. However, the particular form of the SPP (a product of factors) instead of the conventional sum of capital and operating costs assumes that all costs scale with the amount of adsorbent. In some important cases a significant fraction of the capital costs is independent of the amount of material, and if these capital costs dominate, the material cost can become irrelevant. In addition, our study used isothermal working capacities to calculate the SPP for PSA processes whereas industrial adsorption columns operate under adiabatic conditions, which leads to lower working capacities.[170] It is therefore important to compare the SPP with the separation costs calculated by a more detailed process engineering analysis which used adiabatic working capacities. First *et al.*[172] conducted just such a study. From a database of 199 IZA zeolites, the authors selected 86 that topological analysis suggested would be viable for the separation, and adsorption isotherms indicated that 22 of these were highly

selective for $CO_2$. These 22 zeolites then underwent PSA process modeling and optimization to minimize the $CH_4/CO_2$ separation cost. Of the processes First $et\ al.$[172] analyzed, the most similar to one of the three processes we analyzed was one which had adsorption pressures of 3 to 5 bar, desorption pressures of approximately 0.1 bar, and 90 mol % $CH_4$: conditions quite similar to the APG process. The authors found that 8 zeolites could most feasibly perform the separation (ABW, AEN, AHT, APC, BIK, JBW, MON, and WEI), with the separation cost being practically identical among them. Of these top 8 performing zeolites, the SPP identified 4 as the top 4 materials of the 110 IZA zeolites we investigated, and the remaining 4 as being within the top 20. These results indicate that the SPP is a reasonable indicator of separation performance.

To further justify the need to use the SPP, we can compare it to alternative metrics that have been given earlier, such as the Sorbent Selection Parameter ($S$) of Rege and Yang[173]:

$$S = \frac{\alpha^2_{\text{CO2,CH4,ads}}}{\alpha_{\text{CO2,CH4,des}}} \times \frac{\Delta q_{\text{CO}_2}}{\Delta q_{\text{CH}_4}} \tag{4.9}$$

where $\alpha_{i,j} = \frac{q_i/q_j}{y_{i,\text{feed}}/y_{j,\text{feed}}}$ is the selectivity for species $i$ over species $j$ at adsorption or desorption conditions. The Sorbent Selection Parameter seemingly has the advantage of being simpler than the SPP in Eq. 4.1, though the only additional data required by Eq. 4.1 are the heats of adsorption and an estimate for the specific heat of the adsorbent (with the latter being unnecessary for the evaluation of a PSA process). Indeed, we found that the Sorbent Selection Parameter compared as favorably with the results of First $et\ al.$[172] as the SPP, with the 8 top performing zeolites all being within the top 20 zeolites as ranked by the Sorbent Selection Parameter (2 were tied for first with 11 zeolites that had infinite selectivity for $CO_2$, and 6 were ranked within 12–20), so a justification for the value of the SPP is wanted.

Fig. 4.1 shows the relationship between the SPP developed in this work and the Sorbent Selection Parameter for the APG process carried out with PSA at a set desorption pressure, which is representative of the results for the other two processes as well. A close correlation between the two metrics can be observed, though the metrics begin to differ for the top-performing materials, with the Sorbent Selection Parameter being more variable than the SPP due to the selectivity being unbound. We examine a subset of materials selected for their deviating SPP and Sorbent Selection Parameter values in Table 4.1. Here, the undesirability of having a performance metric dependent on the unbound selectivity is demonstrated as all the materials in Table 4.1 have sufficiently low $\Delta q_{\text{CH}_4}$ values such that additional decreases in $\Delta q_{\text{CH}_4}$ should have little effect on separation performance, yet the Sorbent Selection Parameter values continue to increase as $\Delta q_{\text{CH}_4}$ is negligibly decreased despite much more significant decreases in $\Delta q_{\text{CO}_2}$; for example, compare PCOD8294501 and PCOD8310046 in Table 4.1: both have $\Delta q_{\text{CH}_4}$ two orders of magnitude lower than $\Delta q_{\text{CO}_2}$ and thus effectively zero, yet the Sorbent Selection Parameter of PCOD8294501 is greater than that of PCOD8310046 despite the latter's much larger $\Delta q_{\text{CO}_2}$. Furthermore, for materials which do not load any $CH_4$ at all, the selectivity for $CO_2$ goes to infinity and yields Sorbent Selection Parameter values that are uncomparable. Overall, the use of the Sorbent Selec-
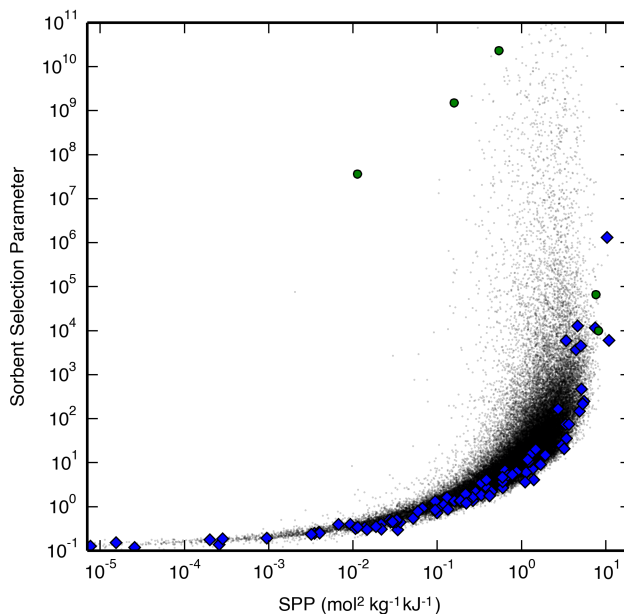
Figure 4.1: A correlation between the SPP and the Sorbent Selection Parameter for the APG process carried out with PSA at set desorption conditions of 1 bar and 300 K. Hypothetical zeolites are shown as black dots, IZA zeolites are shown as blue diamonds, and the hypothetical zeolites in Table 4.1 are shown as green circles.

tion Parameter can thus mislead scientific efforts focused on finding the highest-performing materials.[170]

In addition, the Sorbent Selection Parameter cannot be used to find the optimal desorption conditions for a given adsorbent,[170] as we found that the ratio of working capacities was almost always more sensitive to a change in desorption conditions than the selectivity at desorption conditions (the selectivity at adsorption conditions does not change as a function of desorption conditions). For example, with PSA the Sorbent Selection Parameter was almost always maximized at either the lowest or the highest allowed desorption pressure, simply depending on whether the relative change in $\Delta q_{CH_4}$ or $\Delta q_{CO_2}$ is greater as a function of desorption pressure (it was very rare for the relative changes to become equal in the desorption pressure range allowed). This can be unrelated to the adsorbent's performance, e.g., for an adsorbent very selective for $CO_2$ one may find that the Sorbent Selection Parameter is maximized at a high desorption pressure because negligible changes in $\Delta q_{CH_4}$ have a greater impact on the working capacity ratio than significant changes in $\Delta q_{CO_2}$. Thus, the Sorbent Selection Parameter cannot be used if one is interested in finding optimal desorption conditions.

We similarly compared the SPP to the alternative metric of the Adsorbent Performance

Table 4.1: A selection of hypothetical zeolites with deviating SPP and Sorbent Selection Parameter ($S$) values, shown in order of increasing $\Delta q_{CO_2}$. The first three zeolites have smaller SPP values but larger Sorbent Selection Parameter values than the last two zeolites. The values shown are for materials undergoing the same process conditions as in Fig. 4.1, where these zeolites are circled in green. Note that PCOD8238989 exhibits a higher SPP than PCOD8310046 despite having a higher $\Delta q_{CH_4}$ and a lower $\Delta q_{CO_2}$, which is due to it having a higher $Q_{st,CO_2}$.

| Adsorbent | $\Delta q_{CH_4}$ $(mol\,kg^{-1})$ | $\Delta q_{CO_2}$ $(mol\,kg^{-1})$ | SPP $(mol^2\,kg^{-1}\,kJ^{-1})$ | $S$ |
|---|---|---|---|---|
| PCOD8294501 | $1 \times 10^{-5}$ | 0.005 | 0.01 | $3.6 \times 10^7$ |
| PCOD8164653 | $8 \times 10^{-6}$ | 0.063 | 0.16 | $1.5 \times 10^9$ |
| PCOD8306302 | $4 \times 10^{-6}$ | 0.192 | 0.54 | $2.3 \times 10^{10}$ |
| PCOD8238989 | $7 \times 10^{-2}$ | 2.359 | 8.18 | $9.9 \times 10^3$ |
| PCOD8310046 | $3 \times 10^{-2}$ | 2.550 | 7.71 | $6.6 \times 10^4$ |

Indicator (API) given by Wiersum *et al.*[174]:

$$\text{API} = \frac{(\alpha_{CO2,CH4,ads} - 1)\,\Delta q_{CO_2}}{Q_{st,CO_2}} \tag{4.10}$$

However, the Adsorbent Performance Indicator is highly correlated with the Sorbent Selection Parameter, and suffers from much the same problems (Fig. A.1).

## 4.4 Results and discussion

### Screening

We first investigated the relative merits of PSA and TSA by examining the optimal desorption pressure and desorption temperature for the IZA zeolites undergoing PTSA adsorption. For all three process, less than $10\,\%$ of the zeolites displayed an optimal desorption temperature above $305\,K$, with most materials benefiting more from pulling vacuum than from a temperature-swing (Fig. A.2). For the few zeolites that had a higher optimal desorption temperature, PTSA only negligibly increased the SPP from its value with PSA, so the temperature-swing does not appear to be truly beneficial for any material; conversely, PTSA significantly increased the SPP of several materials compared to that obtained with TSA, demonstrating that pulling vacuum can be truly beneficial (Fig. A.3). In industrial practice as well, PSA is generally preferable to TSA due to a shorter cycle time and lower complexity.[17] It was thus concluded that only PSA processes need be analyzed for the larger datasets studied herein.

PSA screening was then performed for the three processes using all IZA and hypothetical zeolites. For the hypothetical zeolites, we plot the SPP as a function of $Q_{st,CO_2}$ and the $CO_2$
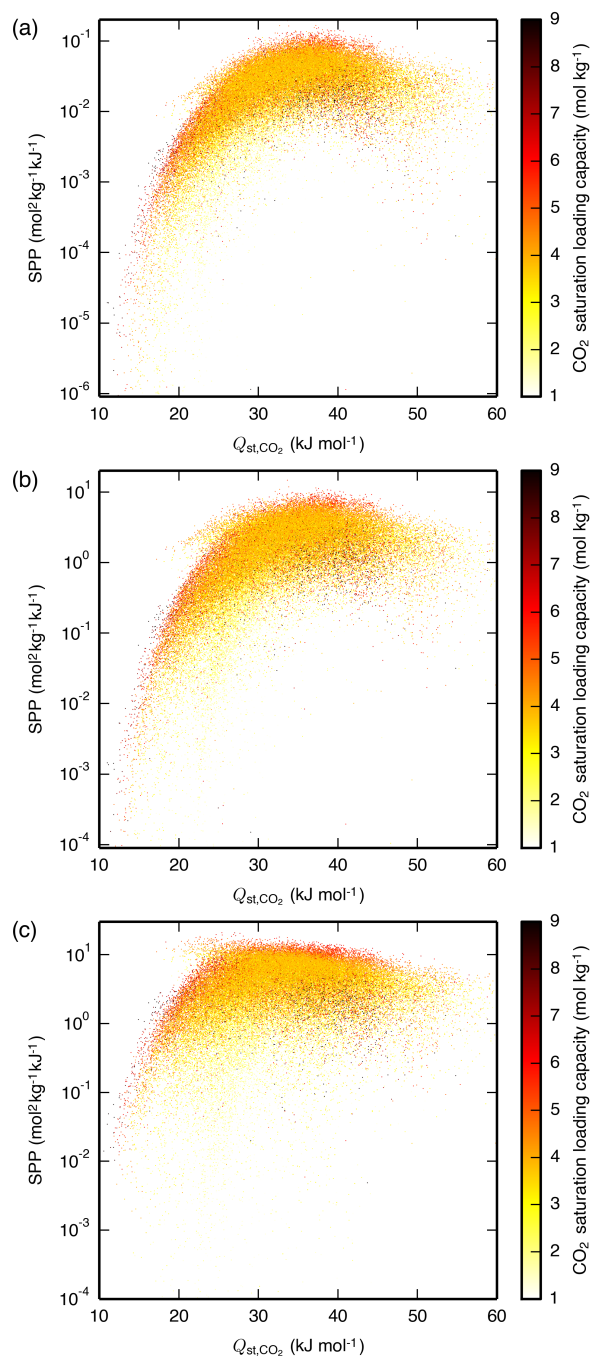
Figure 4.2:   SPP of the hypothetical zeolites as a function of $Q_{st,CO_2}$ and the $CO_2$ saturation loading capacity for the (a) LFG, (b) APG, and (c) NAG processes, all carried out with PSA. The materials are plotted in random order such that the data shown are representative of the materials hidden due to having similar $Q_{st,CO_2}$ and SPP. For the same plot with the IZA zeolites overlaid, see Fig. A.5. Similar results are also seen when using the $CO_2$ Henry coefficient in place of $Q_{st,CO_2}$ (Fig. A.6).
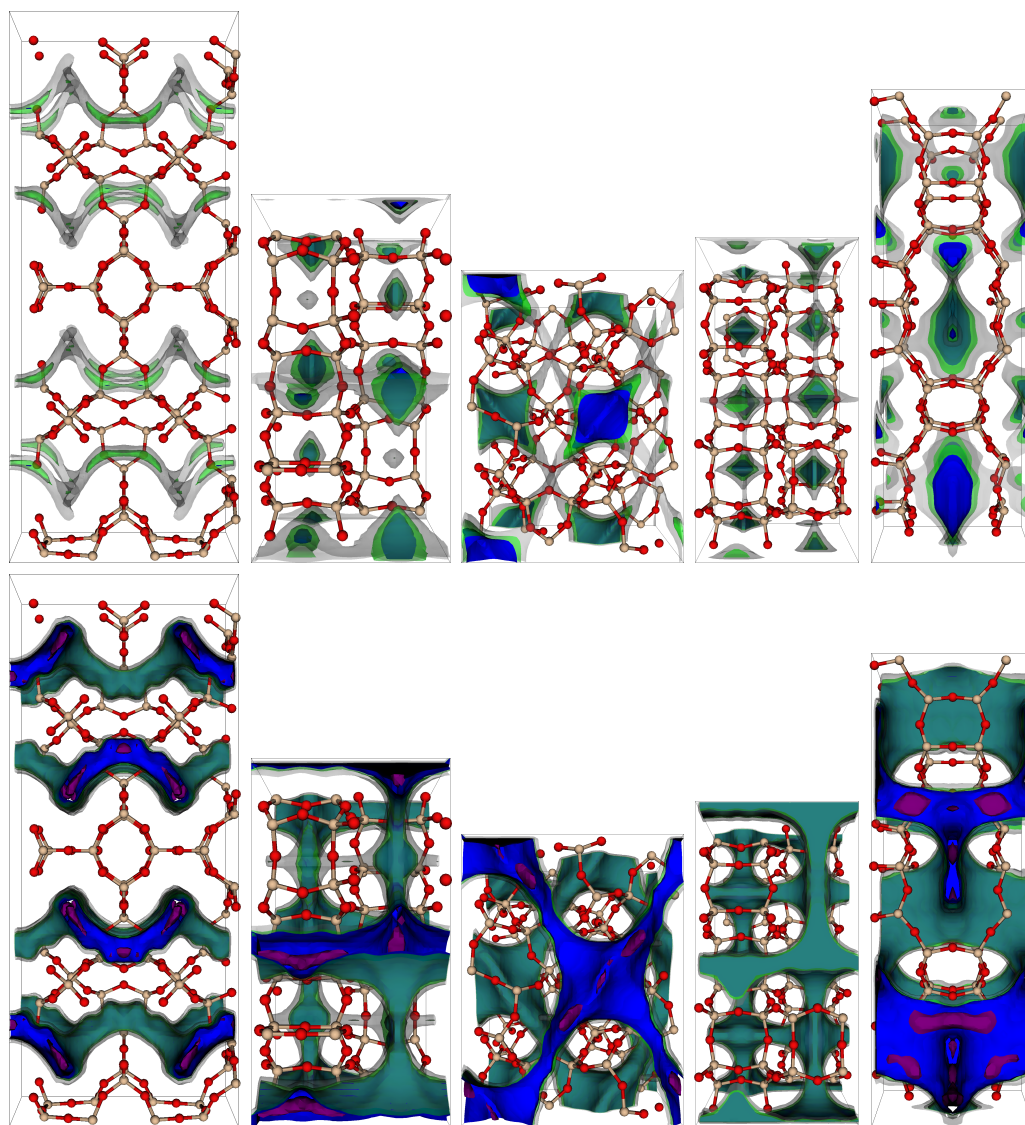
Figure 4.3: Visualization of the (top) $CH_4$ and (bottom) $CO_2$ potential energy surfaces of some of the top-performing hypothetical zeolites for the APG PSA process. From left to right, the zeolites shown are PCOD8284104, PCOD8246225, PCOD8184913, PCOD8246205, and PCOD8310046. Contours correspond to potential energies of $-30\,\mathrm{kJ\,mol^{-1}}$ (purple), $-10\,\mathrm{kJ\,mol^{-1}}$ (blue), $0\,\mathrm{kJ\,mol^{-1}}$ (green), and $37.4\,\mathrm{kJ\,mol^{-1}}$ (gray), with the last representing the limits of accessible channels, approximately defined as $15R \times 300\,\mathrm{K}$, which our GPU GCMC code uses to determine blocked pockets inaccessible from the gas phase, which are then eliminated from use as possible adsorption sites.[175] The second and fourth zeolites from the left demonstrate some $CH_4$ blocked pockets. Both cage- and channel-containing topologies are seen in these top-performing materials. Framework silicon and oxygen atoms are beige and red. The $CO_2$ potential energy shown at each grid point is a Boltzmann-weighted average of 1,000 random rotations of the $CO_2$ molecule about the carbon atom which remains stationary at the grid point.[175,176]

saturation loading capacity (as obtained by fitting the pure-component isotherm) in Fig. 4.2, and we illustrate topologies and potential energy surfaces of some of the highest-performing frameworks in Fig. 4.3. The LFG, APG, and NAG processes exhibit different ranges of optimal $Q_{st,CO_2}$ values, centered around approximately 37, 35, and $32\,kJ\,mol^{-1}$ respectively, decreasing with an increase in adsorption pressure as predicted by theory.[191] Although these optimal $Q_{st,CO_2}$ values are different from each other, they are not so different that the zeolites' SPP values for different processes are not well-correlated with each other (Fig. A.4). However, this finding of zeolite performance being correlated between processes may be specific to our choice of processes, which exhibit some overlap in the $CO_2$ partial pressures over which they cycle; processes with a smaller overlap may in fact exhibit unique optimal adsorbents.[164,167] Our optimal $Q_{st,CO_2}$ values are larger than those found in a screening of a large database of hypothetical MOFs by Wilmer *et al.*[167], where processes similar to the LFG and APG processes showed optimal Sorbent Selection Parameter values with materials having $Q_{st,CO_2}$ values centered around approximately 29 and $31\,kJ\,mol^{-1}$, respectively. This is likely due in part to the different performance metric, pressures, and mole fractions used between our studies, and in part due to zeolites being more confining materials than MOFs, with typically smaller pores and void fractions[192] that provide a larger decrease in entropy upon adsorption of $CO_2$ and thus a larger optimal $Q_{st,CO_2}$.[191]

Within the optimal ranges of $Q_{st,CO_2}$, the best materials are those with a sufficiently large $CO_2$ saturation loading capacity. The range of the SPP for all materials spans several orders of magnitude; by selecting only those materials with near-optimal $Q_{st,CO_2}$ values, the range is reduced but still spans orders of magnitude, and by filtering out the materials with too low a $CO_2$ saturation loading capacity the variability is reduced further. As others have found in large database screenings, bounds on the optimal performance of materials appear to follow simple functions of descriptors, but a given material's performance cannot be perfectly predicted with these same descriptors.[167,169]

## Performance prediction

One wishes to find material descriptors that allow one to predict a zeolite's performance. For this purpose there exist several well-established geometric descriptors of porous media including void fraction (as measured using a helium atom probe), crystal density, accessible surface area, largest included sphere diameter (largest sphere that can fit inside the material), and largest free sphere diameter (largest sphere that can diffuse through the material).[121,181] In addition, the pure-component isotherm descriptors $Q_{st}$ and saturation loading capacity are available for each adsorbate. To determine which of these descriptors are the most important predictors of an adsorbent's performance, we used a random forest machine learning algorithm,[185,193] a method that Simon *et al.*[186] used to screen similar materials for a Xe/Kr separation process and which provides a quantitative measure of a descriptor's importance to improving the quality of prediction. Since it has been shown that the helium void fraction and the crystal density are strongly negative correlated while the largest included sphere diameter and the largest free sphere diameter are strongly positively correlated,[186] we did
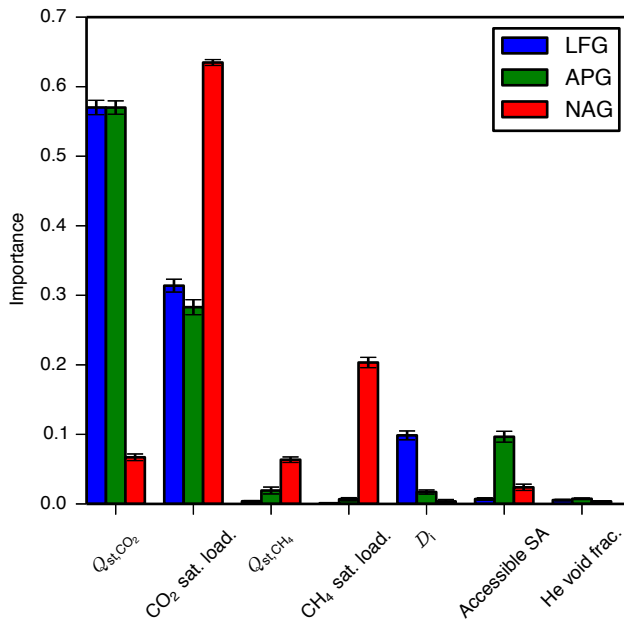
Figure 4.4: Importance of geometric and isotherm descriptors to the SPP of the hypothetical zeolites undergoing PSA processes, as determined using a random forest of decision trees. The importance of a descriptor is calculated by summing the reductions in mean squared error brought about at each node where that descriptor splits a decision tree, averaging over all decision trees, and normalizing.[194] Here, sat. load. is the saturation loading capacity, $D_i$ is the largest included sphere diameter, Accessible SA is the accessible surface area, and He void frac. is the helium void fraction.

not include the crystal density and the largest free diameter as descriptors in the regression; we verified that switching these choices did not give qualitatively different results.

Results from performing the random forest regression on all three processes are shown in Fig. 4.4. The CO$_2$ isotherm descriptors show the greatest influence. Figs. 4.2 and 4.4 both show that for the high-pressure NAG process, the SPP becomes less dependent on $Q_{st,CO_2}$ and more on the CO$_2$ saturation loading capacity, which can be easily rationalized based on the shape of a Langmuir isotherm, which is more sensitive to the Henry coefficient (and thus $Q_{st}$) at low pressures and saturation loading capacity at high pressures.

That the isotherm descriptors proved more important than the geometric descriptors is not altogether surprising. Separation performance is directly dictated by a material's adsorption isotherms, which in turn are determined by a zeolite's topology; thus, the geometric descriptors are at least one step further from separation performance than isotherm descriptors. Less intuitive is the seemingly minor role taken by the CH$_4$ isotherm descriptors. The low importance of the CH$_4$ isotherm descriptors cannot be attributed to the presence of a correlation with the CO$_2$ isotherm descriptors corrupting the regression results, as scatter plots show no clear relationship between CH$_4$ isotherm descriptors and the SPP, while

the $CO_2$ isotherm descriptors are seen to create upper and lower limits on an adsorbent's performance (Figs. A.7–A.9).

The effect of the $CH_4$ isotherm descriptors' influence on separation performance is complicated by the fact that the $CH_4$ and $CO_2$ isotherm descriptors are correlated (Fig. 4.5). While saturation loading capacities of the two species may be difficult to decouple due their heavy dependence on an adsorbent's topology, $Q_{st,CH_4}$ and $Q_{st,CO_2}$ can be separately tuned by taking advantage of the molecules' difference in electrostatics. It would therefore be useful to observe the effect of removing this correlation and then varying the $Q_{st}$ of one adsorbate while keeping the $Q_{st}$ of the other adsorbate constant. We accomplished this by varying the guest-host force field parameters for one adsorbate at a time, i.e. to see the individual effect of $Q_{st,CH_4}$, the $CH_4$-$O_{zeo}$ epsilon parameter of the Lennard-Jones potential was varied above and below its original value, while the individual effect of $Q_{st,CO_2}$ was observed by varying the $C_{CO_2}$-$O_{zeo}$ and $O_{CO_2}$-$O_{zeo}$ epsilon parameters. The resulting $Q_{st}$ values are shown along with the SPP of the PSA APG process for a subset of the IZA zeolites in Fig. 4.6. It can be seen that lowering $Q_{st,CH_4}$ independently of $Q_{st,CO_2}$ increases performance for a range of zeolites nearly uniformly, with the SPP values approaching limiting values where no $CH_4$ adsorbs at all. Conversely, $Q_{st,CO_2}$ exhibits an optimal value unique to each zeolite and process, with performance getting worse as $Q_{st,CO_2}$ decreases or increases from this optimal value. For several of the zeolites, the $Q_{st,CO_2}$ graph is broadly peaked, with the SPP remaining near its maximum value over a range of about $10 \, \text{kJ mol}^{-1}$. The top-performing zeolites seem to exhibit a $Q_{st,CO_2}$ near their optimal values, but there clearly exist materials which would perform poorly even at their optimal $Q_{st,CO_2}$ due to a low $CO_2$ saturation loading capacity. Overall, an ideal material with uncorrelated heats of adsorption for $CH_4$ and $CO_2$ would exhibit a very small $Q_{st,CH_4}$ and an appropriately-sized $Q_{st,CO_2}$.

To understand these trends, consider the argument of Bhatia and Myers[191] that the the working capacity of a single-component gas in a porous material is maximized at a particular $Q_{st}$, which is simply demonstrated by finding the analytical maximum of the difference between loadings at two pressures of a Langmuir isotherm, and positing that the entropy of adsorption is similar in many materials. One expects a similar argument to hold for $CO_2$ in our multicomponent separation, and one might also expect that to maximize the SPP, one would like a $Q_{st,CH_4}$ that is either much larger or much smaller than the $Q_{st,CH_4}$ that maximizes its single-component working capacity. Since a large $Q_{st,CH_4}$ takes adsorption sites away from $CO_2$, it instead makes sense that the SPP increases with a decreasing $Q_{st,CH_4}$. Indeed, even for the zeolite FER, in which SPP is seen to increase with increasing $Q_{st,CH_4}$ beyond a certain point, the $CO_2$ working capacity is decreasing, and the improved SPP is due only to the working capacity of $CH_4$ decreasing at an even faster rate than that of $CO_2$. Notably, the zeolites do not exhibit a single optimal $Q_{st,CO_2}$ but rather a range of optimal $Q_{st,CO_2}$ values, e.g., for the APG PSA process, the zeolites STI, WEI, and MOR exhibit optimal $Q_{st,CO_2}$ values of 32, 41, and $>50 \, \text{kJ mol}^{-1}$, respectively, reflecting the negation of the assumption of similar entropies of adsorption across all materials, which Simon *et al.*[195] has recently demonstrated.

Since $Q_{st,CH_4}$ and $Q_{st,CO_2}$ are positively correlated, it becomes difficult to obtain an
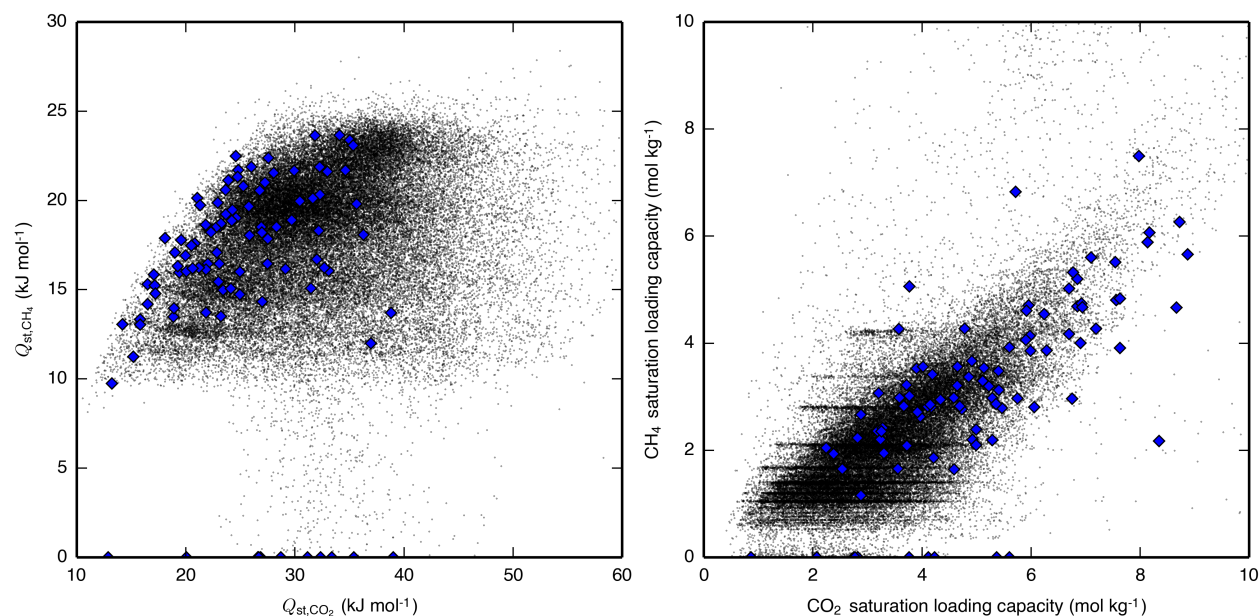
Figure 4.5:   Correlations between (left) $Q_{st,CH_4}$ and $Q_{st,CO_2}$ and (right) saturation loading capacities of $CH_4$ and $CO_2$. Hypothetical zeolites are shown as black dots and IZA zeolites are shown as blue diamonds. The stripes seen in the $CH_4$ saturation loading capacity are due to that variable being more likely to be integer values of $CH_4$ molecules per unit cell (Fig. A.10).

optimal $Q_{st,CO_2}$ while minimizing $Q_{st,CH_4}$, so $Q_{st,CH_4}$ becomes difficult to relate to a zeolite's performance, which explains the low importance of $Q_{st,CH_4}$ obtained from the regression. This understanding suggests that the ability to tune the $CO_2$-host interaction potential independently of the $CH_4$-host interaction potential can lead to higher-performing materials; this can be achieved in practice by changing the charges of framework atoms, more easily done with different material classes like cation-exchanged zeolites or MOFs.

## Structure-property relationships

To aid in the design of new materials, it is preferable to be able to predict a material's performance based solely on geometric descriptors (e.g., helium void fraction, crystal density, accessible surface area, largest included sphere diameter, and largest free sphere diameter), and so we would like to eliminate the use of isotherm descriptors. Within a material class, one would expect this to be possible in theory, particularly for purely-siliceous zeolites as they have identical stoichiometric makeup. Since we have shown that isotherm descriptors are well-capable of predicting a zeolite's performance, if geometric proxies can be found that have a strong relationship with the important isotherm descriptors, the necessity for the isotherm descriptors should be able to be eliminated. However, while it is known that satu-
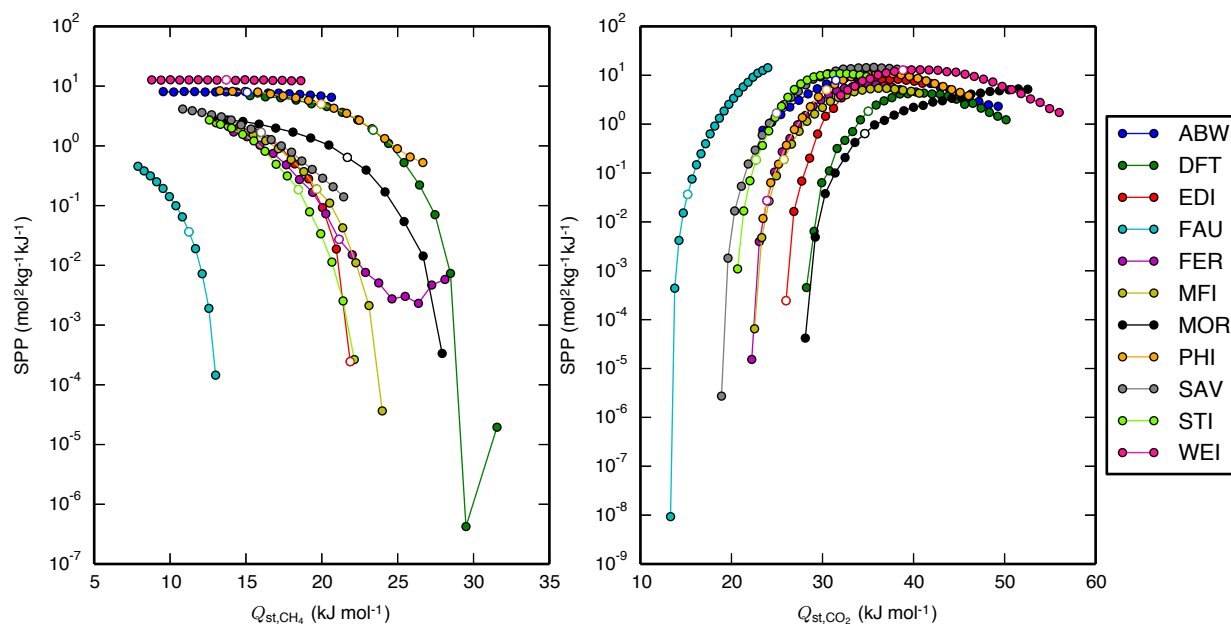
Figure 4.6: SPP of a subset of IZA zeolites undergoing the APG PSA process as a function of (left) $Q_{st,CH_4}$ and (right) $Q_{st,CO_2}$. The points with white marker color represent the data with original guest-host epsilon parameters of the Lennard-Jones potential (115 K for $CH_4$-$O_{zeo}$, 50.2 K for $C_{CO_2}$-$O_{zeo}$, and 84.93 K for $O_{CO_2}$-$O_{zeo}$). (left) The $CH_4$-$O_{zeo}$ epsilon value was varied between 75 K and 155 K in increments of 5 K (4.35 % of the original value), with larger values resulting in a larger $Q_{st,CH_4}$. (right) The $C_{CO_2}$-$O_{zeo}$ epsilon value was varied between 32.128 K and 88.603 K in increments of 2.259 K while the $O_{CO_2}$-$O_{zeo}$ epsilon value was concurrently varied between 54.354 K and 149.904 K in increments of 3.822 K (both 4.50 % of the original values), with larger values resulting in a larger $Q_{st,CO_2}$. When $Q_{st,CH_4}$ was brought too high or $Q_{st,CO_2}$ was brought too low, the material would become unfit for the separation, so these points are not shown. Note that SPP of the zeolite WEI does not change with $Q_{st,CH_4}$ because $CH_4$ loading is negligibly low at all $Q_{st,CH_4}$ values. Qualitatively similar results for the LFG and NAG processes are shown in Figs A.11 and A.12.

ration loading capacities of both $CO_2$ and $CH_4$ can be predicted by several simple geometric descriptors, and past studies have shown a clear relationship between a material's largest sphere diameter and $CH_4$ adsorption energetics,[196] more complex descriptors are necessary to predict $Q_{st,CO_2}$[197,198] (Figs A.13 and A.14). Since we lack a simple geometric proxy for $Q_{st,CO_2}$, we will be unable to remove it from use. Indeed, although past screening studies have shown that purely geometric descriptors can predict performance for single-component gas storage of simple molecules like $CH_4$,[199] for more complicated multicomponent gas separation processes involving $CO_2$ it has been found necessary for geometric parameters to be supplemented by $Q_{st,CO_2}$[167] or by both $Q_{st}$ values[200]. It appears that the inclusion of $Q_{st,CO_2}$ (or a similar isotherm descriptor such as the Henry coefficient) as a performance predictor will be necessary until simple geometric descriptors that are well-correlated with $Q_{st,CO_2}$ are found.

Conversely, one would expect to find the $CO_2$ saturation loading capacity unnecessary because it seems to have strong relationships with several geometric descriptors. However, the relationship between the geometric descriptors and the SPP is more ambiguous than that between the $CO_2$ saturation loading capacity and the SPP, so a simple heuristic cannot be deduced (Figs. A.15–A.20), though the descriptor is less crucial for machine learning prediction.

## IAST validity

As we have generated both pure and mixture adsorption isotherms for over 80,000 zeolites, we can study the validity of the IAST assumption for nanoporous materials at a larger scale than has been attempted before. Many studies have evaluated the validity of IAST for a small number of materials.[201–210] Of these, the ones that have focused on mixtures of $CO_2$ and $CH_4$ in zeolites and MOFs have generally found good agreement between IAST and mixture isotherms at low pressures but some divergence emerging at pressures above about 1 bar, with the various studies showing that divergence can manifest itself in several ways.[201–205] One reason commonly cited for the violation of IAST is its assumption that the same adsorbent surface area is available to all adsorbate molecules,[206,207] an assumption that Myers and Prausnitz[180] acknowledged is invalid for molecular sieves, and which tends to be less valid at higher pressures as adsorbate molecules compete for adsorption sites and adsorbates become accommodated at less energetically favorable adsorption sites. Another reason commonly cited for the violation of IAST is its assumption of adsorbed-phase ideality,[206] which also becomes less valid at higher pressures as higher adsorbate loadings are induced.

We compared mixture isotherm loadings obtained directly with those obtained through IAST, with results shown for a 60 mol % $CH_4$ mixture in Fig. 4.7 and qualitatively similar results shown for a 90 mol % $CH_4$ mixture in Fig. A.21. For $CH_4$, it can be seen that at low pressures the IAST error is centered around 0, but as the pressure increases IAST tends to underpredict the loading, with errors becoming considerable by around 5 bar. For $CO_2$, IAST brings about less bias at all pressures investigated. These divergence trends are similar to what was seen in Krishna and van Baten[203], where it was found that different adsorption sites

(e.g., windows versus cages) contain different proportions of the two adsorbates; this leads IAST to anticipate greater competition between the adsorbates than actually occurs, and the loading of the weaker-adsorbing species is underpredicted. This type of error brought about by IAST due to site segregation can be eliminated by using a segregated IAST model.[207]

For both adsorbates and at all pressures, there exist materials in which IAST introduces large errors. One of the materials with the largest introduced error, PCOD8205017, was investigated in further detail. Pure-component and mixture isotherms of $CH_4$ and $CO_2$ in this material are shown in Fig. 4.8, where it can be seen that the directly-generated mixture isotherms are similar to the pure-component isotherms, and that IAST drastically underpredicts $CH_4$ loading. The snapshots in Fig. 4.8 show that the $CH_4$ and $CO_2$ adsorbates occupy very distinct adsorption sites; these same sites are also preferentially occupied by each adsorbate during pure-component adsorption, so during mixture adsorption there is a lack of competition between adsorbates for adsorption sites until a pressure is reached at which one of the two sites becomes saturated. This extreme case of site segregation illustrates the dangers of not verifying the validity of IAST, as IAST results lead to the prediction that PCOD8205017 is a much higher-performing material than it actually is.

By further examining the PCOD8205017 isotherms in Fig. 4.8, it can be seen that at certain points the uptakes are even larger in the directly-generated mixture isotherms than in the pure-component isotherms. This is due to the $CH_4$-$C_{CO_2}$ and $CH_4$-$O_{CO_2}$ adsorption site distances both being approximately 1.1 times their Lennard-Jones sigma values, near the potential's minimum. Thus, rather than following the IAST assumption that adsorbates compete for the same adsorption sites, this guest-guest nonideality instead makes the presence of the second adsorbate species conducive to the uptake of the first. That only small deviations from ideality occur despite such positioning of adsorption sites lends support to the ideality assumption of IAST and segregated IAST, as well as to a previous finding that guest-host interactions give rise to nonideality to a greater extent than do guest-guest interactions.[211]

For the purposes of high-throughput materials screening, a small number of outliers is less important than the introduction of systematic errors. To test whether IAST brought about systematic errors, we redid our earlier analysis of descriptors using the IAST results, and we obtained qualitatively-similar results, with the higher-pressure NAG process associated with a decrease in the optimal $Q_{st,CO_2}$ values and an increased dependence on the $CO_2$ saturation loading capacity (Figs. A.22 and A.23). Thus, IAST appears to be adequate for the purposes of performing high-throughput materials screening to find trends that relate structure to performance, but it must be tested for validity when performing detailed analyses of individual adsorbents.

# 4.5   Concluding remarks

In this work, we have developed the SPP metric, which avoids the pitfalls associated with other adsorption metrics such as the Sorbent Selection Parameter and the Adsorbent Perfor-
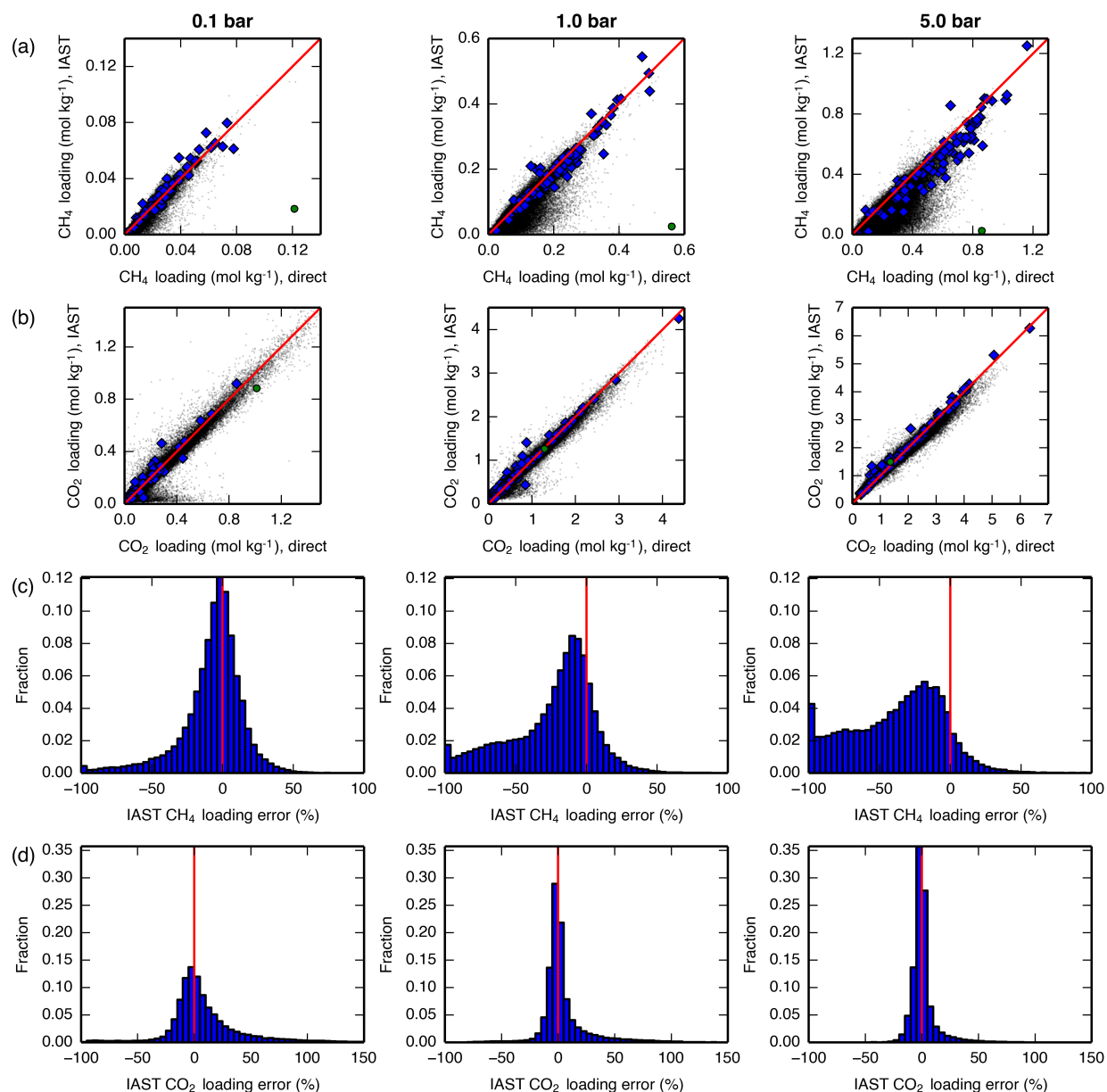
Figure 4.7: Loadings of (a) CH$_4$ and (b) CO$_2$ at various total pressures, 300 K, and 60 mol % CH$_4$. Data on the x-axis are taken from directly-simulated mixture isotherms, while data on the y-axis are taken from applying IAST to pure-component isotherms. A line is drawn at $y = x$ for reference. Hypothetical zeolites are shown as black dots, IZA zeolites are shown as blue diamonds, and hypothetical zeolite PCOD8205017 is shown as a green circle. Histograms of the fractional IAST error $\left( \frac{\text{IAST loading} - \text{direct loading}}{\text{direct loading}} \times 100\% \right)$ of the hypothetical zeolites only for (c) CH$_4$ and (d) CO$_2$ at the same conditions. A line is drawn at $x = 0$ for reference. All plots in the left-column are at 0.1 bar, all plots in the middle-column are at 1 bar, and all plots in the right-column are at 5 bar. A similar plot showing results at 90 mol % CH$_4$ is given in Fig. A.21.
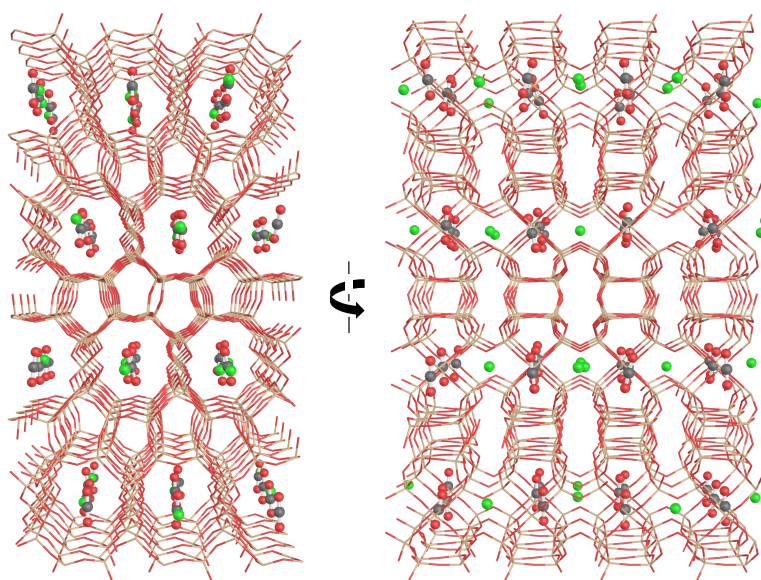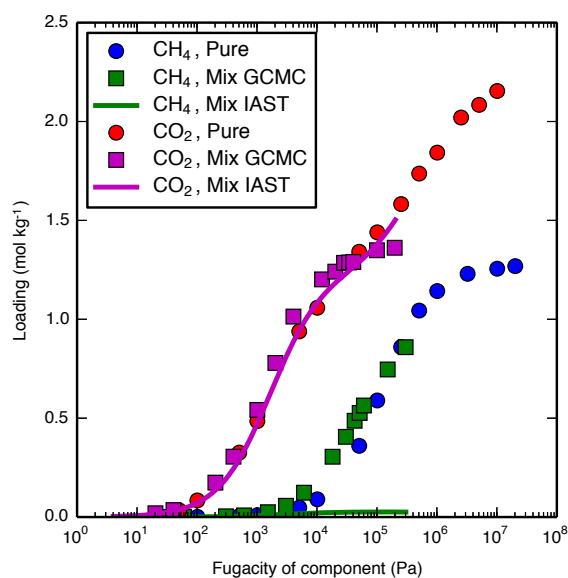
Figure 4.8: (top) Pure-component and mixture isotherms of CH$_4$ (60 mol %) and CO$_2$ (40 mol %) in the hypothetical zeolite PCOD8205017 at 300 K. Mixture isotherms are shown both as directly-generated from GCMC and as calculated from the pure-component isotherms using IAST.[180] Unlike the isotherms used in the screening, these used the ideal gas equation of state to ensure consistency of units on the x-axis. (bottom) A snapshot of CH$_4$ and CO$_2$ in hypothetical zeolite PCOD8205017 taken from the mixture isotherm simulation conducted at CH$_4$ and CO$_2$ fugacities of 0.6 and 0.4 bar, respectively, shown at viewing angles orthogonal to each other. Framework silicon and oxygen atoms are beige and red, CH$_4$ united atoms are green, and CO$_2$ carbon and oxygen atoms are gray and red.

mance Indicator; namely that they can overemphasize negligible gains in selectivity at the expense of more important changes in working capacity. Although the SPP was developed specifically for a process in which CO$_2$ is adsorbed from a CH$_4$/CO$_2$ mixture, it can be easily adapted for other adsorption processes.

The screening work that we then undertook with the SPP is complementary to several recent screening analyses of materials for CO$_2$ capture.[164,167–169,182,200,212,213] Similarly to several of these studies[164,167,169] we found that $Q_{st,CO_2}$ should be neither too low nor too high to make an optimal material. We built on this finding to show that the CO$_2$ saturation loading capacity is another important parameter to consider when designing materials, particularly for high-pressure processes. We then extended the theory of Bhatia and Myers[191], initially developed for storage of a single-component gas, to a multicomponent gas separation case by demonstrating that $Q_{st,CH_4}$ should be as low as possible to optimize separations, but that because $Q_{st,CH_4}$ is positively correlated with $Q_{st,CO_2}$, it misleadingly appears to be an unimportant parameter.

Since we chose to directly generate mixture isotherms rather than rely on the predictions of IAST, we were also able to perform a large-scale evaluation of the validity of IAST for siliceous zeolites. We found that IAST works reliably for the purposes of using high-throughput screening to find performance trends, but above about 5 bar IAST begins to underpredict the loading of CH$_4$. At all pressures there exist zeolites for which IAST does not work, which we showed can be due to site segregation.

# Chapter 5

# Studying one material in-depth: vapor-liquid coexistence in a metal-organic framework[*]

Molecular simulations and NMR relaxometry experiments demonstrate that pure benzene or xylene confined in isoreticular metal-organic frameworks (IRMOFs) exhibit true vapor-liquid phase equilibria where the effective critical point may be reduced by tuning the structure of the MOF. Our results are consistent with vapor and liquid phases extending over many MOF unit cells. These results are counterintuitive since the MOF pore diameters are approximately the same length scale as the adsorbate molecules. As applications of these materials in catalysis, separations, and gas storage rely on the ability to tune the properties of adsorbed molecules, we anticipate that the ability to systematically control the critical point, thereby preparing spatially inhomogeneous local adsorbate densities, could add a new design tool for MOF applications.

## 5.1 Introduction

There is a rich emerging literature on MOFs owing to the ability to tune local chemical structure and geometry by joining various metal-containing units with a panoply of organic linkers to create nanoporous crystalline materials.[214] Judicious choice of linker-metal combinations yields MOFs exhibiting ultrahigh porosity in combination with thermal and chemical stability. These are widely explored for applications ranging from separations, storage, catalysis, etc.[215]

Confinement of molecules within pores affects their collective properties, such as phase behavior. A well-known example is capillary condensation where interactions between adsorbed molecules and the pore wall induce quasi one-dimensional vapor-liquid coexistence at conditions where the bulk fluid is singularly present as a vapor.[8] In the literature, capillary

---

[*]This chapter is based on Braun, Chen, Schnell, Lin, Reimer, and Smit[82].

condensation is often associated with a single pore, and if the dimensions of this pore are of similar size as the adsorbed molecules, capillary condensation is suppressed; thus, the adsorbed fluid is present as a single phase at all conditions. As the pores of most MOFs are on the order of nanometers, capillary condensation will not occur.

Phase transitions induced by adsorption have been observed in MOFs and are associated with mechanisms other than capillary condensation. For example, in flexible MOFs such as MIL-53 adsorption-induced phase transitions are observed in the host MOF crystals.[216,217] Other phase transitions observed in nanoporous materials are governed by commensurate-incommensurate transitions in the geometric packing of the molecules in the pores.[218] Our work is motivated by the surprising behavior of benzene and related molecules in IRMOF-1 (MOF-5). Previous NMR and molecular dynamics studies of benzene motion in IRMOF-1 yielded diffusivities associated with two domains of differing mobilities.[219–221] Adsorption experiments showed features suggesting multiple surface energies.[222,223] Finally, grand-canonical Monte Carlo simulations detailed a narrow hysteresis loop for cyclohexane in IRMOF-1, which was reluctantly associated with capillary condensation even though the authors noted that the pores of IRMOF-1 are too small to support the effect.[8,224] In contrast, others proposed that the step in the $CO_2$ adsorption isotherm in IRMOF-1 is related to a vapor-liquid transition.[225] Other studies, however, do not observe the hysteresis and conclude that attractive electrostatic interactions between $CO_2$ molecules are responsible for the unusual shape of the adsorption isotherms,[226] or explain such unusual behavior in terms of clustering.[227,228]

## 5.2 Results and discussion

We used molecular simulations to examine the adsorption of benzene and xylenes in a representative set of IRMOFs as shown in Fig. 5.1. Fig. 5.2 illustrates the key result of our Monte Carlo simulations: ordinary vapor-liquid coexistence of benzene in these materials. We observe liquid and vapor phases that are similar to bulk phase coexistence that extend over many unit cells. Fig. 5.3 shows the corresponding phase diagram; the densities indicated by the phase envelope are consistent with loadings found from the hysteresis loop portion in the adsorption isotherms as obtained from grand-canonical Monte Carlo simulations (Fig. 5.4). These results are consistent with the conclusions reported for $CO_2$ in IRMOF-1.[225] The critical point of benzene in IRMOF-1 was found to be 375 K, about 200 K below the bulk critical point. This decrease in the critical point stems from the frustration of benzene packing in the liquid phase by the ligands of the MOF.

Interestingly, we can leverage the native structural and chemical flexibility afforded by the modular construction of MOFs to tune this critical point. Phase diagrams were simulated for benzene in the IRMOF series. We found that as the pore size increases, the critical temperature of benzene increases, approaching its bulk value (Fig. 5.3). For IRMOFs with the same linker length, such as IRMOF-1 and IRMOF-7 or IRMOF-10 and IRMOF-14, the addition of a side group lowers the critical temperature. We surmise that this is because
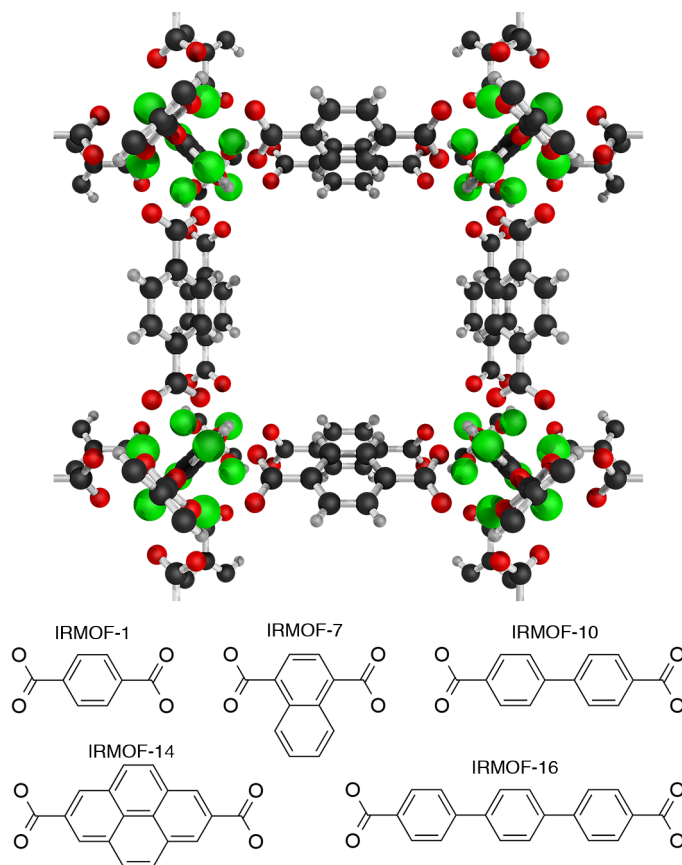
Figure 5.1: (Top) Crystal structure of IRMOF-1, which consists of $Zn_4O$ complexes connect by 1,4-benzene dicarboxylate linkers. (Bottom) Linker molecules of all IRMOFs used in the molecular simulations.

these side groups further frustrate the packing in the liquid phase. Additionally, we repeated our simulations with xylene isomers in IRMOF-1 and found a similar reduction in the critical temperatures for all three isomers (Figs. A.31–A.34).

## Experimental

Experimental confirmation of vapor-liquid phase transitions lies in the detection of spatially extended domains of vapor-liquid coexistence. Towards that end we turned to NMR relaxometry, a technique that allows us to probe subtle changes in molecular motion. The differing densities of the vapor and liquid phases coexisting within the MOFs are expected to create dissimilar magnetic environments that are distinguishable using NMR spin-spin relaxation measurements. We measured the distribution of spin-spin ($T_2$) relaxation times using Carr-Purcell-Meiboom-Gill (CPMG) experiments combined with a Laplace inversion numerical analysis algorithm to disentangle the multiple relaxation time constants.[230–232]
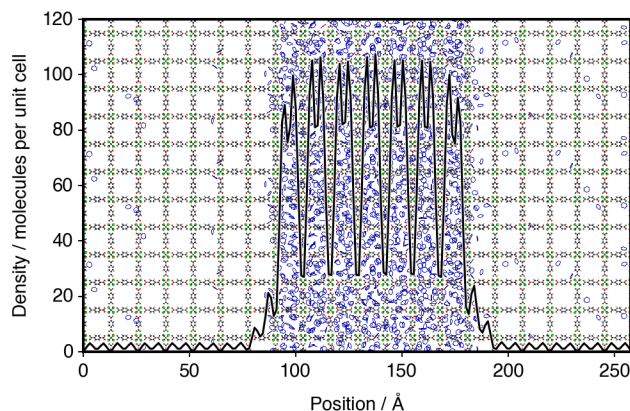
Figure 5.2: Overlay of the density profile of benzene molecules in IRMOF-1 at 270 K with a snapshot of the system as obtained from NVT Monte Carlo simulations; the snapshot has been scaled to exactly match the density profile's axes. The distribution of this density profile between the two different types of cages present in IRMOF-1 is shown in SI Figure 6. Because of the periodic boundary conditions we observe an infinite slab with two interfaces. The sizes of the liquid and gas slabs depend on the total size of our simulation cell.
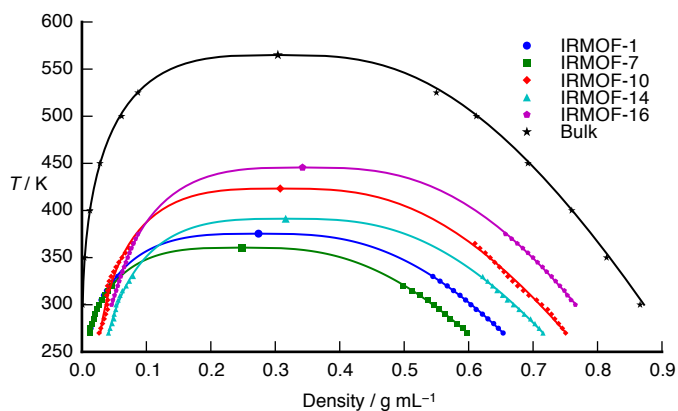


Figure 5.3: Vaporliquid coexistence curves for benzene in various IRMOFs. The density scaling law (with the three-dimensional Ising critical exponent $\beta = 0.32$) and the law of rectilinear diameters were used to estimate the critical point and to interpolate the data. Note that the error bars of these calculations are smaller than the symbols. Densities were converted to mass per void space volume using void fractions of 0.832, 0.782, 0.910, 0.914, and 0.939 for IRMOFs 1, 7, 10, 14, and 16, respectively.
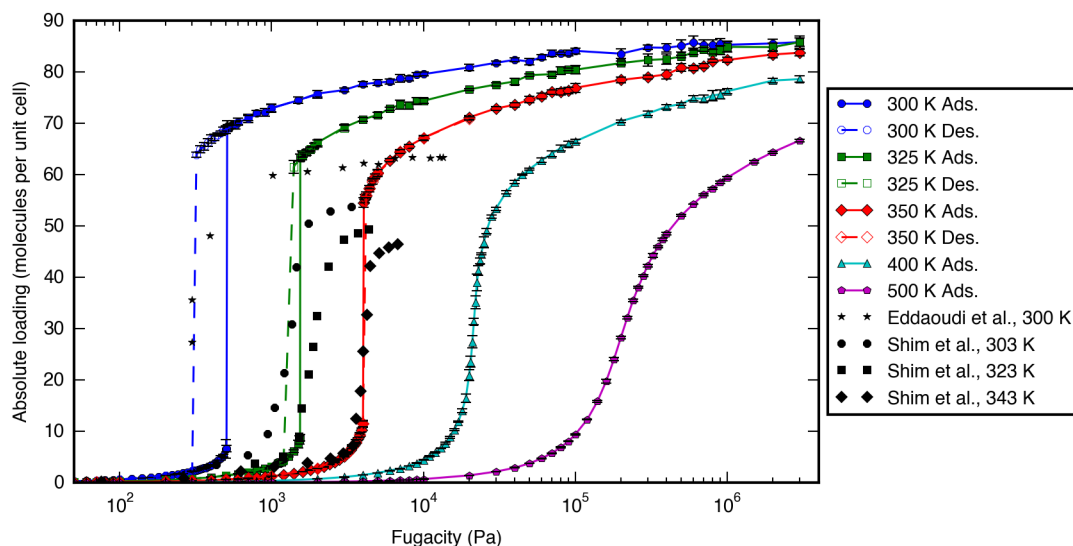
Figure 5.4:   Adsorption isotherms of benzene in IRMOF-1. Solid lines and dashed lines are guides to the eye showing adsorption and desorption branches, respectively. Black symbols are experimental data taken from the literature.[222,229] Although the experimental data shown are actually excess loadings as a function of pressure, the experimental conditions were such that excess loading is approximately equal to absolute loading and pressure is approximately equal to fugacity.

Fig. 5.5a illustrates the distribution of $T_2$ relaxation times obtained from benzene in IRMOF-1 at different temperatures. A single thermodynamic phase of benzene would produce a single relaxation time, yet we observe three separate peaks in the $T_2$ distributions, illustrating that benzene resides in distinct environments. For example, at 295 K we observe a dominant peak at intermediate relaxation times (120 ms) as well as a small peak at long relaxation times (870 ms). We associate the dominant peak with benzene in the liquid phase and the smaller peak with benzene in the vapor phase for two reasons. First, spin-spin relaxation times of adsorbed fluids increase with molecular mobility,[233] hence the shorter $T_2$ emanates from the liquid phase. Secondly, the experiments are conducted under conditions where the majority of molecules (70 %) reside in the liquid phase. The third small peak (5 ms) is assigned to molecules near the vapor-liquid interface. As the temperature increases the peak associated with the liquid phase (intermediate-$T_2$ relaxation times) decreases in size, corresponding to a decrease in the mole fraction of liquid, while the peak associated with the vapor phase (long-$T_2$ relaxation times) correspondingly increases in size. This trend continues such that at 343 K the number of molecules associated with each relaxation peak is roughly equal, and at 388 K the vapor phase dominates. The distinct changes in the peak intensities are convincingly commensurate with the simulated vapor-liquid coexistence curve (Fig. 5.3). Indeed, at higher temperatures more benzene molecules are in the vapor phase.
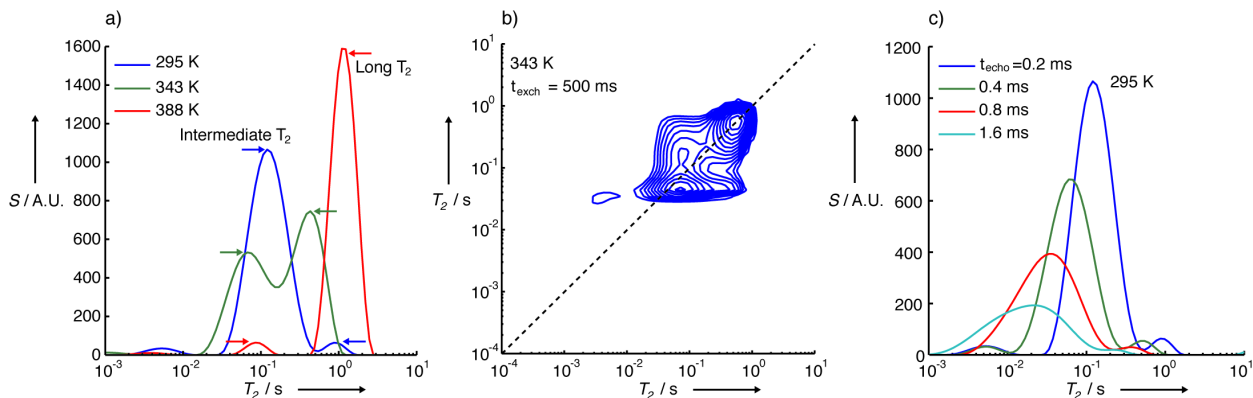
Figure 5.5: NMR relaxation and exchange of benzene in IRMOF-1. (a) $T_2$ relaxation distributions of benzene in IRMOF-1 at a loading of $9 \pm 1$ molecules of benzene per unit cell at different temperatures, where the right-facing arrows indicate the intermediate $T_2$ population and the left-facing arrows indicate the long $T_2$ population. (b) A contour plot of the $T_2$-$T_2$ relaxation exchange distribution at 343 K. The off-diagonal intensities illustrate the presence of molecular exchange between the intermediate (liquid) and long (vapor) relaxation time environments during the timescale set by the experiment; in these data $t_{exch} = 500$ ms. (c) $T_2$ relaxation distributions at 295 K with varying echo period ($t_{echo}$) indicating the presence of magnetic field gradients arising from the vaporliquid interface.

A unique and defining feature of this coexistence is the extension of the vapor and liquid phases over many unit cells (Fig. 5.2). The corresponding length scales may be probed directly by two-dimensional $T_2$-$T_2$ relaxation exchange experiments where the appearance of cross-peaks demonstrates that during the measurement time ($t_{exch}$) molecules have moved from a liquid phase to a vapor phase. Fig. 5.5b shows the $T_2$-$T_2$ exchange plots for $t_{exch} = 500$ ms when benzene is adsorbed in IRMOF-1. Peaks centered on the diagonal represent benzene molecules that have not exchanged between the liquid and vapor phases during $t_{exch}$, whereas peaks that are centered symmetrically on either side of the diagonal are associated with molecules that have exchanged between these phases. Exchange peaks are prominent in Fig. 5.5b, confirming that the molecules are exchanging between the liquid and vapor phases within a timescale of $t_{exch} = 100$ to 500 ms. This time scale, along with a representative diffusion coefficient of benzene in IRMOF-1 ($10^9$ m$^2$ s$^{-1}$),[219,220] gives us ca. $10^5$ m as the characteristic length scale over which a molecule diffuses between both phases. This 10-micron length scale shows that indeed the vapor and liquid phases extend over many unit cells.

A third experimental observation supporting our assignment of two distinct phases is seen in the dependence of spin-spin relaxation times ($T_2$) on the echo period of the CPMG sequence. An interface such as the one shown in Fig. 5.2 creates spatially-localized magnetic field gradients owing to the differences in magnetic susceptibility between the vapor and liquid phases.[234] The magnitude of these field gradients increases with proximity to the

interface. We directly probe the diffusion of molecules through the vapor-liquid interface by varying the echo period ($t_{echo}$) in the CPMG experiment.[235,236] By varying this echo period we tune the window over which we monitor molecular diffusion. With short echo periods, the diffusion length is short, and the effect of the gradient on $T_2$ is minimal; increasing the echo period increases the effect of the field gradient on $T_2$. Furthermore, increasing the echo period (and the associated diffusion length) also increases the number of molecules that are influenced by the interface. Both of these effects explain the shortening relaxation times depicted in Fig. 5.5c.

## 5.3 Concluding remarks

It is important to distinguish the present phenomenon from capillary condensation. Capillary condensation cannot support a true phase transition in cylindrical pores because the correlation length associated with density fluctuations can only grow in one dimension.[8] Conversely, the IRMOF-type frameworks studied herein exhibit relatively open structures, and hence correlation lengths can grow in any direction, with phase transitions likely belonging to the universal class of the three-dimensional Ising model.[225] In this context it is interesting to compare these materials to aerogels, which are also open structures and change the vapor-liquid equilibrium curve of adsorbed molecules.[237,238] In aerogels, however, the porosity is so high that the critical point is only changed by a few mK.

The present phenomenon is also distinct from previous reports of adsorbate clustering in which phase coexistence does not play a role.[227] In several zeolites and MOFs, previous workers found that as temperatures are lowered below the adsorbate bulk critical temperature, the height of the first peak in the adsorbate radial distribution function (equivalently, the height of the first peak in the probability density function of finding pairs of adsorbate molecules a given distance apart) increases, which was interpreted as an increase in the degree of molecular clustering.[227] We examined methane adsorption in IRMOF-1, confirmed that our simulations reproduced these previously reported effects at 125 K (Fig. A.36), and did not find phase coexistence (Fig. A.37). We did find that phase coexistence was present when the system's temperature was lowered to 50 K (Fig. A.37)—a temperature below the range previously investigated—in accordance with a phase transition experimentally observed in this system at 60 K,[239] and which further demonstrates that the phase coexistence phenomenon is not limited to aromatic adsorbates.

We conclude that even materials with pore sizes on the order of nanometers can support vapor-liquid coexistence provided they possess the correct topology. This provides a natural explanation for the anomalous behavior reported for benzene and related molecules in IRMOF-1.[219–223] More broadly, we anticipate that this phenomenon is not limited to vapor-liquid equilibria. Similar effects, for example, can be expected for liquid-liquid mixtures. The ability to substantially tune the critical point of a pure fluid sheds new light on a very classic phenomenon, portending a revisiting of pure and multicomponent phase behavior in nanoporous materials. The phenomenon of mesoscopic domains of differing adsorbate densi-

ties/compositions in MOFs introduces heterogeneity which, similar to multivariate MOFs,[240] has potential applications in catalysis, separations, and storage technologies.

## 5.4   Methods

We focus here on the simulation methods used. For details on the experimental methods, the reader is referred to Braun *et al.*[82].

### Thermodynamics

Before discussing the different ensembles that we have used in this work it is instructive to discuss the experimental conditions. Experimental adsorption isotherms are obtained by imposing an external pressure (chemical potential) and temperature and measuring the number of adsorbed molecules. At these conditions one would not expect to see coexistence in the adsorbed phase as the system can minimize the free energy by avoiding a gas-liquid interface and hence be entirely in the gas or liquid phase. To mimic this experiment we use grand-canonical ($\mu$VT) Monte Carlo simulations.

The situation is different, however, for the NMR experiments. Here a system is prepared with a fixed number of particles in a sealed tube. Hence the number of particles and volume (MOF and tube) is fixed. Unlike the experimental adsorption isotherm system, the free volume is small and cannot act as a reservoir. To see how in this system one can achieve co-existence, one can envision the following thought experiment. We load the MOF with a fixed number of particles and place this in the NMR tube at such low pressure and temperature that the number of gas molecules in the tube is negligible compared to the number of gas molecules in the MOF. We now seal the NMR tube and bring the system to a temperature where coexistence is expected. At this point molecules will leave the MOF and the pressure in the tube will increase; this process will stop once the chemical potential of the adsorbate molecules in the tube is equal to the chemical potential of the adsorbate molecules in the MOF. The difference with the adsorption experiment is that because the NMR tube is small and has a fixed volume it cannot act as a reservoir and hence the total number of molecules that can go from the MOF into the gas phase is fixed. In this situation, at conditions where we have coexistence, the system can no longer increase or decrease the number of adsorbed molecules to have only a liquid or gas phase in equilibrium with the gas in the tube, but the constraint of a fixed volume and number of particles forces the system to stay in coexistence where the relative amounts of gas and liquid are consistent within these constraints. Hence, to mimic the NMR system we use canonical (NVT) Monte Carlo simulations.

### Canonical Monte Carlo

NVT simulations were used to observe vapor-liquid coexistence of guest molecules inside the IRMOF frameworks. Liquid-vapor equilibrium was established by selecting a density

Table 5.1: Canonical Monte Carlo simulation parameters.

| Structure | Simulation box dimensions (unit cells) | Loading (molecules) | N highest density unit cells used for liquid phase | M lowest density unit cells used for vapor phase |
| --- | --- | --- | --- | --- |
| IRMOF-1 | $2 \times 2 \times 10$ | 1072 benzene | 2 | 3 |
| IRMOF-1 | $2 \times 2 \times 10$ | 768 xylene | 3 | 3 |
| IRMOF-7 | $2 \times 2 \times 10$ | 800 benzene | 2 | 3 |
| IRMOF-10 | $1 \times 1 \times 8$ | 700 benzene | 2 | 3 |
| IRMOF-14 | $1 \times 1 \times 7$ | 700 benzene | 1 | 2 |
| IRMOF-16 | $1 \times 1 \times 10$ | 1400 benzene | 1 | 4 |

intermediate to that of the liquid and vapor phases as estimated by the adsorption isotherms generated by the grand-canonical Monte Carlo simulations (Table 5.1). The liquid-vapor interface was minimized by choosing a high aspect ratio simulation box which caused the two phases to only form interfaces in the small-area cross-sectional direction.

The density profile shown in Fig. 5.2 was generated by dividing the simulation box into 20 bins per unit cell in the elongated dimension and averaging the number of CH united atoms (and C and $CH_3$ united atoms for the xylenes) in each bin over the length of the simulation. To examine the effects of the two different types of cages that make up an IRMOF unit cell, the two cross-sectional directions were further subdivided into 2 bins per unit cell (1 bin per cage), and every other cage was averaged in the cross-sectional dimensions to produce separate density profiles for the A and B cages as shown in Fig. A.35.

At high temperatures, the phases were not stationary but rather moved in the elongated direction, so these density profiles could not be used to produce phase diagrams. Instead, the phase diagrams were produced by dividing the simulation box into 1 bin per unit cell in the elongated dimension; each time data were collected, the liquid phase density was calculated by averaging the N unit cells with the highest densities and the vapor phase density was calculated by averaging the M unit cells with the lowest densities, where N and M are integers specific to the structure. These densities were then averaged over the length of the simulation. We selected values of N and M (Table 5.1) to ensure that we neglected regions close to or containing an interface while including as much of the pure phases as possible. This algorithm will necessarily sometimes discard the lowest-density pure liquid bin from being included in the liquid phase density calculation and vice-versa for the vapor phase, which may lead to an overestimation of the liquid density and an underestimation of the vapor density. To estimate the error inherent in this method, we recalculated the density of the IRMOF-10 benzene vapor phase in a worst-case manner, using the 2 lowest-density bins rather than 3, which ensures that at least 1 pure vapor bin will always be discarded from the vapor density calculation; over all temperatures, this gave an average decrease in the calculated vapor density of $12\,\%$, which resulted in an increase of the fitted critical temperature by $2\,K$ (less than $1\,\%$) and a decrease in the fitted critical density of $2\,\%$, which

serve as approximate error bounds on the method.

Critical temperatures and densities were obtained by fitting to the density scaling law (with the three-dimensional Ising critical exponent $\beta = 0.32$) and the law of rectilinear diameters.[4] The maximum temperature at which data were collected was selected by stopping at a temperature when droplets of liquid were visually observed to nucleate in the vapor phase. Data were collected every 10 cycles over at least 200,000 cycles following at least 200,000 equilibration cycles.

## Grand-canonical Monte Carlo

$\mu VT$ simulations were conducted to generate gas adsorption isotherms of benzene and the xylenes in IRMOF-1. Data were collected over at least 50,000 cycles following at least 50,000 equilibration cycles. The IRMOF-1 simulation box consisted of 2 by 2 by 2 unit cells.

## Void fractions

Void fractions were obtained using Widom insertions of helium at 298 K as per Talu and Myers[179]. These void fractions were used to convert the densities of the liquid and vapor phases from units of molecules per unit cell to units of mass per volume of void space for Fig. 5.3 and Fig. A.34.

## Force field

The Lennard-Jones potential was used to describe dispersive interactions with Universal Force Field parameters used for the framework atoms,[145,241] TraPPE model parameters for the adsorbate molecules,[60,242] and helium-helium parameters as given by Talu and Myers[179] Lorentz-Berthelot mixing rules were used with a simple truncation at 14 Å. As the TraPPE models do not have partial charges, Coulombic interactions were not used. The adsorbate-framework potentials were pretabulated and interpolated over a 0.1 Å grid. The frameworks were assumed to be rigid and the experimental structures were used.[243]

Our choice of force field was validated by comparison to experimental adsorption isotherms (Figs. 5.4–A.33).[222,229,244] The benzene isotherms of Shim *et al.*[222] are of the IUPAC IV type which is associated with condensation, but Eddaoudi *et al.*[229] do not provide sufficient resolution at low pressure to show an initial plateau. For benzene, the simulated condensation pressures well match both sets of experimental data, and the disparity in saturation loadings may be explained by the simulations' use of a perfectly crystalline framework. The simulated xylene isotherms do not match well with the data of Gu *et al.*[244] that was taken at what our simulations showed to be a supercritical temperature; this may be indicative of poor crystal quality, which would also explain the small BET surface area of the MOF-5 of Gu *et al.*[244].

# Chapter 6

# Outlook

In this dissertation, many of the ways in which molecular simulations can be used for the study of nanoporous materials have been illustrated. Here, some personal thoughts on these studies are given. These are given in chronological order rather than in chapter order.

Chapter 4 was my first project in the MolSim group. I did not perform a full review of the literature prior to beginning the project, and I found during the project's write-up that most of the results had already been observed by Wilmer et al.[167]. The Separation Performance Parameter was developed as a performance metric without my being aware of previous metric such as the Sorbent Selection Parameter, and although I found the Separation Performance Parameter to be slightly preferable to the Sorbent Selection Parameter, its complexity makes it unlikely to be used by others (or even by myself in future studies). The most helpful section of Braun et al.[163] may be on the validity of IAST; although other studies have been conducted in a similar manner, few have been on such a large scale. The examples given of the materials in which the IAST assumptions are grossly violated may be useful as test cases for competing theories developed in the future. More likely, Braun et al.[163] will be of little future influence. Nonetheless, I learned a lot while conducting the project and writing the manuscript, and my failure to have conducted a proper literature review was not a mistake that I have repeated or one that I will forget.

Chapter 5 found that benzene exhibited vapor-liquid coexistence in MOF-5 despite its microporous confinement. After the publication of Braun et al.[82], Sondre and I attempted to find ways in which the condensation phenomenon could be exploited for a storage or separation application, as materials exhibiting a stepped isotherm are particularly attractive for these applications.[245] The effect of the confinement is to lower a fluid's vapor pressure at a particular temperature (or alternatively, to increase a fluid's boiling temperature at a particular pressure), so we reasoned that the most likely industrial applications were for fluids with reasonably high vapor pressure at room temperature which necessitated high-pressure liquid storage. We found the storage of ammonia to be a promising application, as ammonia has a vapor pressure of approximately $10\,\text{bar}$ at $298\,\text{K}$, and the safe storage of ammonia is an important industrial problem. Unfortunately, ammonia degrades MOF-5 (and many other MOFs), so finding a stable MOF is a technical challenge that needs to be overcome

for this application. Of course, it was already widely known that capillary condensation in mesoporous materials decreases a fluid's vapor pressure, and this was already being exploited for gas storage applications.[246] Whether the particular kind of bulk-like (three-dimensional) confined condensation we observed will be useful for applications remains to be seen.

Chapter 2 showed that two common thermostats—simple velocity rescaling and the Berendsen thermostat—result in simulation artifacts which can be extreme enough to nullify a study's conclusions. Fortunately, the CSVR thermostat avoids these artifacts with negligible extra computational costs. The CSVR thermostat had been growing in popularity prior to the publication of Braun *et al.*[19] (Fig. 2.1), and while the Berendsen thermostat had also been growing in popularity, its growth was seeming to plateau (perhaps due to such studies as Shirts[72], which empirically showed some of the Berendsen thermostat's ill effects). It is clear from the results of Braun *et al.*[19] that the simple velocity rescaling and Berendsen thermostat algorithms should be entirely discontinued; their only acceptable future use should be in replicating results of previous studies. Seeing the speed with which these thermostats' use declines will be interesting from a sociological perspective since it is not frequent that a common scientific method is so thoroughly discredited. The speed will be dictated by a combination of forces, including the dissemination of Braun *et al.*[19], the denouncement of the thermostats by textbook authors, and the availability of thermostats provided by MD software distributors. I look forward to seeing how Fig. 2.1 looks in a few years.

Chapter 3 was primarily concerned with showing that ZTCs can be viewed as an experimental realization of schwarzites, which had been thought to be a purely hypothetical material. Negatively-curved schwarzites complete the triumvirate of two-dimensionally curved carbons, along with positively-curved fullerenes and zero-curved graphene. Although ZTCs have been explored for applications, I am most interested in their use as a research platform for studying porous carbons. Porous carbons are one of the most common porous materials in use today for adsorption, electrochemical, and catalytic applications, but their amorphous nature makes it difficult to study structure-property relationships. Having an array of ordered microporous carbons experimentally available in the form of ZTCs could lead to many advances in understanding these materials. At present, the largest impediment to realization of this goal is the synthesis of novel ZTCs. Many of the zeolites which we predicted could template experimentally-accessible ZTCs are not available off-the-shelf, but have only been synthesized a handful of times by select groups with high expertise in zeolite synthesis, so obtaining these zeolites will be an experimental challenge. Furthermore, the template carbonization process requires the zeolites to be stable at elevated temperatures, which not all of the available zeolites are without post-processing. A separate synthetic challenge lies in fully-templating the zeolite surface. The carbon loading that has been achieved in zeolites is only half of what we have found possible, likely due to diffusion limitations of the carbon precursor through the small zeolite pores. Decreasing the size of the carbon precursor will improve the templating procedure, and the ultimate limit may be found when using atomized carbon as a precursor. I look forward to seeing how the field of synthesizing schwarzites via zeolite-templating advances over the coming years.

# Bibliography

[1]     D. E. Shaw, J. P. Grossman, J. A. Bank, B. Batson, J. A. Butts, J. C. Chao, M. M. Den-
        eroff, R. O. Dror, A. Even, C. H. Fenton, A. Forte, J. Gagliardo, G. Gill, B. Greskamp,
        C. R. Ho, D. J. Ierardi, L. Iserovich, J. S. Kuskin, R. H. Larson, T. Layman, L.-S.
        Lee, A. K. Lerer, C. Li, D. Killebrew, K. M. Mackenzie, S. Y.-H. Mok, M. A. Moraes,
        R. Mueller, L. J. Nociolo, J. L. Peticolas, T. Quan, D. Ramot, J. K. Salmon, D. P.
        Scarpazza, U. B. Schafer, N. Siddique, C. W. Snyder, J. Spengler, P. T. P. Tang,
        M. Theobald, H. Toma, B. Towles, B. Vitale, S. C. Wang and C. Young, Proceedings
        of the International Conference for High Performance Computing, Networking, Storage
        and Analysis, 2014, pp. 41–53.

[2]     D. Chandler, *Introduction to Modern Statistical Mechanics*, Oxford University Press,
        1987.

[3]     M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*, Oxford University
        Press, Oxford, 1989.

[4]     D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to
        Applications*, Elsevier Science, 2002.

[5]     C. J. Cramer, *Essentials of Computational Chemistry: Theories and Models*, John
        Wiley & Sons, 2004.

[6]     B. Leimkuhler and C. Matthews, *Molecular Dynamics With Deterministic and Stochas-
        tic Numerical Methods*, Springer, 2015.

[7]     M. E. Davis, *Nature*, 2002, **417**, 813–821.

[8]     L. D. Gelb, K. E. Gubbins, R. Radhakrishnan and M. Sliwinska-Bartkowiak, *Rep.
        Prog. Phys.*, 1999, **62**, 1573–1659.

[9]     T. J. Bandosz, M. J. Biggs, K. E. Gubbins, Y. Hattori, T. Iiyama, K. Kaneko, J. Piku-
        nic and K. T. Thomson, in *Chemistry and Physics of Carbon*, ed. L. R. Radovic, Marcel
        Dekker, 2003, vol. 28, ch. 2, pp. 41–228.

[10]    M. J. Biggs and B. A., *Mol. Simul.*, 2006, **32**, 579–593.

[11] H. Nishihara and T. Kyotani, *Adv. Mater.*, 2012, **24**, 4473–4498.

[12] *Annual Energy Outlook 2015*, U. S. Energy Information Administration technical report, 2015.

[13] S. Pacala and R. Socolow, *Science*, 2004, **305**, 968–972.

[14] R. W. Howarth, R. Santoro and A. Ingraffea, *Clim. Change*, 2011, **106**, 679–690.

[15] A. R. Brandt, G. A. Heath, E. A. Kort, F. O'Sullivan, G. Pétron, S. M. Jordaan, P. Tans, J. Wilcox, A. M. Gopstein, D. Arent, S. Wofsy, N. J. Brown, R. Bradley, G. D. Stucky, D. Eardley and R. Harriss, *Science*, 2014, **343**, 733–735.

[16] N. Kumar, P. Besuner, S. Lefton, D. Agan and D. Hilleman, *Power Plant Cycling Costs*, NREL subcontract report prepared by Intertek APTECH, 2012.

[17] M. Tagliabue, D. Farrusseng, S. Valencia, S. Aguado, U. Ravon, C. Rizzo, A. Corma and C. Mirodatos, *Chem. Eng. J.*, 2009, **155**, 553–566.

[18] S. Cavenati, C. A. Grande and A. E. Rodrigues, *J. Chem. Eng. Data*, 2004, **49**, 1095–1101.

[19] E. Braun, M. S. Moosavi and B. Smit, *Submitted*, 2018.

[20] L. V. Woodcock, *Chem. Phys. Lett.*, 1971, **10**, 257–261.

[21] D. J. Evans, W. G. Hoover, B. H. Failor, B. Moran and A. J. C. Ladd, *Phys. Rev. A*, 1983, **28**, 1016–1021.

[22] S. Nosé, *J. Chem. Phys.*, 1984, **81**, 511–519.

[23] D. J. Evans and G. P. Morriss, *Statistical Mechanics of Nonequilibrium Liquids*, Academic Press, London, 1990.

[24] T. Schneider and E. Stoll, *Phys. Rev. B*, 1978, **17**, 1302–1322.

[25] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684–3690.

[26] G. Bussi, D. Donadio and M. Parrinello, *J. Chem. Phys.*, 2007, **126**, 014101.

[27] D. M. Heyes, *Chem. Phys.*, 1983, **82**, 285–301.

[28] S. Nosé, *Mol. Phys.*, 1984, **52**, 255–268.

[29] W. G. Hoover, *Phys. Rev. A*, 1985, **31**, 1695–1697.

[30] G. J. Martyna, M. L. Klein and M. Tuckerman, *J. Chem. Phys.*, 1992, **97**, 2635–2643.

[31]  H. C. Andersen, *J. Chem. Phys.*, 1980, **72**, 2384–2393.

[32]  D. J. Tobias, G. J. Martyna and M. L. Klein, *J. Phys. Chem.*, 1993, **97**, 12959–12966.

[33]  G. P. Morriss and C. P. Dettmann, *Chaos*, 1998, **8**, 321–336.

[34]  P. H. Hünenberger, in *Advanced Computer Simulation*, ed. C. Holm and K. Kremer, Springer, 2005, vol. 173, pp. 104–149.

[35]  M. Tuckerman, *Statistical Mechanics: Theory and Molecular Simulation*, Oxford University Press, Oxford, 2010.

[36]  J. M. Haile and S. Gupta, *J. Chem. Phys.*, 1983, **79**, 3067–3076.

[37]  D. J. Evans and G. Morriss, *Phys. Lett. A*, 1983, **98**, 433–436.

[38]  P. Minary, G. J. Martyna and M. E. Tuckerman, *J. Chem. Phys.*, 2003, **118**, 2510–2526.

[39]  P. Collins, G. S. Ezra and S. Wiggins, *J. Chem. Phys.*, 2010, **133**, 014105.

[40]  T. Morishita, *J. Chem. Phys.*, 2000, **113**, 2976–2982.

[41]  T. Morishita, *J. Chem. Phys.*, 2003, **119**, 7075–7082.

[42]  A. S. Lemak and N. K. Balabaev, *Mol. Simul.*, 1994, **13**, 177–187.

[43]  S. C. Harvey, R. K.-Z. Tan and T. E. Cheatham, *J. Comput. Chem.*, 1998, **19**, 726–740.

[44]  H. B. Callen, *Thermodynamics and an Introduction to Thermostatistics*, John Wiley & Sons, 2nd edn., 1985.

[45]  T. Çağin and J. R. Ray, *Phys. Rev. A*, 1988, **37**, 4510–4513.

[46]  R. B. Shirts, S. R. Burt and A. M. Johnson, *J. Chem. Phys.*, 2006, **125**, 164102.

[47]  M. J. Uline, D. W. Siderius and D. S. Corti, *J. Chem. Phys.*, 2008, **128**, 124301.

[48]  N. H. Siboni, D. Raabe and F. Varnik, *Phys. Rev. E*, 2013, **87**, 030101.

[49]  B. Cooke and S. C. Schmidler, *J. Chem. Phys.*, 2008, **129**, 164112.

[50]  M. Lingenheil, R. Denschlag, R. Reichold and P. Tavan, *J. Chem. Theory Comput.*, 2008, **4**, 1293–1306.

[51]  N. Goga, A. J. Rzepiela, A. H. de Vries, S. J. Marrink and H. J. C. Berendsen, *J. Chem. Theory Comput.*, 2012, **8**, 3637–3649.

[52]  J. E. Basconi and M. R. Shirts, *J. Chem. Theory Comput.*, 2013, **9**, 2887–2899.

[53] S. Nosé, *Prog. Theor. Phys. Suppl.*, 1991, **103**, 1–46.

[54] S. Chiu, M. Clark, S. Subramaniam and E. Jakobsson, *J. Comput. Chem.*, 2000, **21**, 121–131.

[55] M. P. Eastwood, K. A. Stafford, R. A. Lippert, M. Ø. Jensen, P. Maragakis, C. Predescu, R. O. Dror and D. E. Shaw, *J. Chem. Theory Comput.*, 2010, **6**, 2045–2058.

[56] C. Sagui and T. A. Darden, *Annu. Rev. Biophys. Biomol. Struct.*, 1999, **28**, 155–179.

[57] J. R. Wagner, G. S. Balaraman, M. J. M. Niesen, A. B. Larsen, A. Jain and N. Vaidehi, *J. Comput. Chem.*, 2013, **34**, 904–914.

[58] L. Yan, C. Sun and H. Liu, *Adv. Manuf.*, 2013, **1**, 160–165.

[59] S. Plimpton, *J. Comput. Phys.*, 1995, **117**, 1–19.

[60] G. M. Martin and J. I. Siepmann, *J. Phys. Chem. B*, 1998, **102**, 2569–2577.

[61] J. W. Ponder and F. M. Richards, *J. Comput. Chem.*, 1987, **8**, 1016–1024.

[62] M. Tafipolsky, S. Amirjalayer and R. Schmid, *J. Comput. Chem.*, 2007, **28**, 1169–1176.

[63] S. Amirjalayer, M. Tafipolsky and R. Schmid, *Angew. Chem., Int. Ed.*, 2007, **46**, 463–466.

[64] S. Toxvaerd and O. H. Olsen, *Phys. Scr.*, 1990, **1990**, 98–101.

[65] M. E. Tuckerman, Y. Liu, G. Ciccotti and G. J. Martyna, *J. Chem. Phys.*, 2001, **115**, 1678–1702.

[66] S. Hess, *Z. Naturforsch. A*, 2003, **58**, 377–391.

[67] V. I. Manousiouthakis and M. W. Deem, *J. Chem. Phys.*, 1999, **110**, 2753–2756.

[68] P. Mark and L. Nilsson, *J. Comput. Chem.*, 2002, **23**, 1211–1219.

[69] A. Mudi and C. Chakravarty, *Mol. Phys.*, 2004, **102**, 681–685.

[70] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren and J. Hermans, Intermolecular Forces, Dordrecht, 1981, pp. 331–342.

[71] W. A. Steele, *Surf. Sci.*, 1973, **36**, 317–352.

[72] M. R. Shirts, *J. Chem. Theory Comput.*, 2013, **9**, 909–926.

[73] A. Cheng and K. M. Merz, *J. Phys. Chem.*, 1996, **100**, 1927–1937.

[74] A. Mor, G. Ziv and Y. Levy, *J. Comput. Chem.*, 2008, **29**, 1992–1998.

[75] E. Rosta, N. V. Buchete and G. Hummer, *J. Chem. Theory Comput.*, 2009, **5**, 1393–1399.

[76] Y. G. Spill, S. Pasquali and P. Derreumaux, *J. Chem. Theory Comput.*, 2011, **7**, 1502–1510.

[77] G. Bussi and M. Parrinello, *Comput. Phys. Commun.*, 2008, **179**, 26–29.

[78] B. Smit and T. L. M. Maesen, *Chem. Rev.*, 2008, **108**, 4125–4184.

[79] K. Seehamart, T. Nanok, R. Krishna, J. M. van Baten, T. Remsungnen and S. Fritzsche, *Microporous Mesoporous Mater.*, 2009, **125**, 97–100.

[80] D. Dubbeldam, K. S. Walton, D. E. Ellis and R. Q. Snurr, *Angew. Chem. Int. Ed.*, 2007, **46**, 4496–4499.

[81] D. C. Ford, D. Dubbeldam and R. Q. Snurr, *diffusion-fundamentals.org*, 2009, **11**, 1–8.

[82] E. Braun, J. J. Chen, S. K. Schnell, L.-C. Lin, J. A. Reimer and B. Smit, *Angew. Chem. Int. Ed.*, 2015, **54**, 14349–14352.

[83] J.-M. Leyssale and G. L. Vignoles, *Chem. Phys. Lett.*, 2008, **454**, 299–304.

[84] J. Wong-ekkabut, M. S. Miettinen, C. Dias and M. Karttunen, *Nat. Nanotechnol.*, 2010, **5**, 555–557.

[85] M. Baker, *Nature*, 2016, **533**, 452–454.

[86] Y. Hu and S. B. Sinnott, *J. Comput. Phys.*, 2004, **200**, 251–266.

[87] E. Braun, Y. Lee, M. S. Moosavi, S. Barthel, R. Mercado, I. A. Baburin, D. M. Proserpio and B. Smit, *Submitted*, 2018.

[88] R. Hoffmann, A. A. Kabanov, A. A. Golov and D. M. Proserpio, *Angew. Chem., Int. Ed.*, 2016, **55**, 10962–10976.

[89] A. L. Mackay and H. Terrones, *Nature*, 1991, **352**, 762.

[90] T. Lenosky, X. Gonze, M. Teter and V. Elser, *Nature*, 1992, **355**, 333–335.

[91] D. Vanderbilt and J. Tersoff, *Phys. Rev. Lett.*, 1992, **68**, 511–513.

[92] H. Terrones and A. L. Mackay, *Carbon*, 1992, **30**, 1251–1260.

[93] M. O'Keeffe, G. B. Adams and O. F. Sankey, *Phys. Rev. Lett.*, 1992, **68**, 2325–2328.

[94] M. Z. Huang, W. Y. Ching and T. Lenosky, *Phys. Rev. B*, 1993, **47**, 1593–1606.

[95] H. Terrones and M. Terasaki, *New J. Phys.*, 2003, **5**, 126.

[96] N. Park, M. Yoon, S. Berber, J. Ihm, E. Osawa and D. Tománek, *Phys. Rev. Lett.*, 2003, **91**, 237204.

[97] A. Lherbier, H. Terrones and J.-C. Charlier, *Phys. Rev. B*, 2014, **90**, 125434.

[98] D. Odkhuu, D. H. Jung, H. Lee, S. S. Han, S. H. Choi, R. S. Ruoff and N. Park, *Carbon*, 2014, **66**, 39–47.

[99] C. P. Ewels, X. Rocquefelte, H. W. Kroto, M. J. Rayson, P. R. Briddon and M. I. Heggie, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, 15609–15612.

[100] Y. Xia, Z. Yang and R. Mokaya, *Nanoscale*, 2010, **2**, 639–659.

[101] M. Inagaki, H. Orikasa and T. Morishita, *RSC Adv.*, 2011, **1**, 1620–1640.

[102] H. Nishihara and T. Kyotani, in *Novel Carbon Adsorbents*, ed. J. M. Tascón, Elsevier, 2012, ch. 10, pp. 295–322.

[103] V. Malgras, Q. Ji, Y. Kamachi, T. Mori, F. K. Shieh, K. C.-W. Wu, K. Ariga and Y. Yamauchi, *Bull. Chem. Soc. Jpn.*, 2015, **88**, 1171–1200.

[104] T. Roussel, A. Didion, R. J.-M. Pellenq, R. Gadiou, C. Bichara and C. Vix-Guterl, *J. Phys. Chem. C*, 2007, **111**, 15863–15876.

[105] T. Roussel, C. Bichara, K. E. Gubbins and R. J.-M. Pellenq, *J. Chem. Phys.*, 2009, **130**, 174717.

[106] K. Nueangnoraj, H. Nishihara, K. Imai, H. Itoi, T. Ishii, M. Kiguchi, Y. Sato, M. Terauchi and T. Kyotani, *Carbon*, 2013, **62**, 455–464.

[107] J. Parmentier, F. O. M. Gaslain, O. Ersen, T. A. Centeno and L. A. Solovyov, *Langmuir*, 2014, **30**, 297–307.

[108] H. Nishihara, H. Fujimoto, H. Itoi, K. Nomura, H. Tanaka, M. T. Miyahara, P. A. Bonnaud, R. Miura, A. Suzuki, N. Miyamoto, N. Hatakeyama, A. Miyamoto, K. Ikeda, T. Otomo and T. Kyotani, *Carbon*, 2018, **129**, 854–862.

[109] F. O. M. Gaslain, J. Parmentier, V. P. Valtchev and J. Patarin, *Chem. Commun.*, 2006, 991–993.

[110] K. Kim, T. Lee, Y. Kwon, Y. Seo, J. Jong, J. K. Park, H. Lee, J. Y. Park, H. Ihee, S. J. Cho and R. Ryoo, *Nature*, 2016, **535**, 131–135.

[111] H. Nishihara, Q. H. Yang, P. X. Hou, M. Unno, S. Yamauchi, R. Saito, J. I. Paredes, A. Martínez-Alonso, J. M. D. Tascón, Y. Sato, M. Terauchi and T. Kyotani, *Carbon*, 2009, **47**, 1220–1230.

[112] T. Roussel, C. Bichara and R. J.-M. Pellenq, *Adsorption*, 2005, **11**, 709–714.

[113] D. Jariwala, V. K. Sangwan, L. J. Lauhon, T. J. Marks and M. C. Hersam, *Chem. Soc. Rev.*, 2013, **42**, 2824–2860.

[114] Z. Ma, T. Kyotani and A. Tomita, *Carbon*, 2002, **40**, 2367–2374.

[115] C. Ducrot-Boisgontier, J. Parmentier and J. Patarin, *Microporous Mesoporous Mater.*, 2009, **126**, 101–106.

[116] N. Alam and R. Mokaya, *Energy Environ. Sci.*, 2010, **3**, 1773–1781.

[117] Y. Xia, R. Mokaya, D. M. Grant and G. S. Walker, *Carbon*, 2011, **49**, 844–853.

[118] N. Alam and R. Mokaya, *Microporous Mesoporous Mater.*, 2011, **144**, 140–147.

[119] S. Builes, T. Roussel, C. M. Ghimbeu, J. Parmentier, R. Gadiou, C. Vix-Guterl and L. F. Vega, *Phys. Chem. Chem. Phys.*, 2011, **13**, 16063–16070.

[120] C. Baerlocher and L. B. McCusker, *Database of Zeolite Structures*, `http://www.iza-structure.org/databases`.

[121] T. F. Willems, C. H. Rycroft, M. Kazi, J. C. Meza and M. Haranczyk, *Microporous Mesoporous Mater.*, 2012, **149**, 134–141.

[122] M. M. J. Treacy and J. M. Newsam, *Nature*, 1988, **332**, 249–251.

[123] S. J. Stuart, A. B. Tutein and J. A. Harrison, *J. Chem. Phys.*, 2000, **112**, 6472–6486.

[124] A. L. Mackay, *Curr. Sci.*, 1995, **69**, 151–161.

[125] A. H. Schoen, *Inifinite periodic minimal surfaces without self-intersections*, NASA Electronics Research Center Technical Report NASA TN D-5541, C-98, 1970.

[126] S. Andersson, S. T. Hyde, K. Larsson and S. Lidin, *Chemical Reviews*, 1988, **88**, 221–242.

[127] L. de Campo, O. Delgado-Friedrichs, S. T. Hyde and M. O'Keeffe, *Acta Crystallogr. Sect. A*, 2013, **A69**, 483–489.

[128] S. Brenner, L. B. McCusker and C. Baerlocher, *J. Appl. Crystallogr.*, 1997, **30**, 1167–1172.

[129] M. O'Keeffe, M. A. Peskov, S. J. Ramsden and O. M. Yaghi, *Acc. Chem. Res.*, 2008, **41**, 1782–1789.

[130] H. A. Schwarz, *Gesammelte Mathematische Abhandlungen*, Springer-Verlag, Berlin, 1890.

[131] E. Koch and W. Fischer, *Z. Kristallogr.*, 1988, **183**, 129–152.

[132] A. Fogden and S. T. Hyde, *Acta Crystallogr. Sect. A*, 1992, **A48**, 575–591.

[133] W. Fischer and E. Koch, *Phil. Trans. R. Soc. A*, 1996, **354**, 2105–2142.

[134] E. A. Lord and A. L. Mackay, *Curr. Sci.*, 2003, **85**, 346–362.

[135] M. Moliner, F. Rey and A. Corma, *Angew. Chem. Int. Ed.*, 2013, **52**, 13880–13889.

[136] N. E. R. Zimmermann and M. Haranczyk, *Cryst. Growth Des.*, 2016, **16**, 3043–3048.

[137] M. M. J. Treacy, I. Rivin, E. Balkovsky, K. H. Randall and M. D. Foster, *Microporous Mesoporous Mater.*, 2004, **74**, 121–132.

[138] R. Pophale, P. A. Cheeseman and M. W. Deem, *Phys. Chem. Chem. Phys.*, 2011, **13**, 12407.

[139] Z. Qin, G. S. Jung, M. J. Kang and M. J. Buehler, *Sci. Adv.*, 2017, **3**, e1601536.

[140] *Zeolites in Catalysis: Properties and Applications*, ed. J. Čejka, R. E. Morris and P. Nachtigall, Royal Society of Chemistry, 2017.

[141] A. Belsky, M. Hellenbrandt, V. L. Karen and P. Luksch, *Acta Crystallogr. Sect. B*, 2002, **58**, 364–369.

[142] M. M. Martínez-Iñesta, I. Peral, T. Proffen and R. F. Lobo, *Microporous Mesoporous Mater.*, 2005, **77**, 55–66.

[143] S. Lee and G. S. Hwang, *J. Appl. Phys.*, 2011, **110**, 093524.

[144] V. Perebeinos and J. Tersoff, *Phys. Rev. B*, 2009, **79**, 241409.

[145] A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard and W. M. Skiff, *J. Am. Chem. Soc.*, 1992, **114**, 10024–10035.

[146] J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing and J. Hutter, *Comput. Phys. Commun.*, 2005, **167**, 103–128.

[147] J. Hutter, M. Iannuzzi, F. Schiffmann and J. VandeVondele, *WIREs Comput. Mol. Sci.*, 2014, **4**, 15–25.

[148] J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865–3868.

[149] S. Grimme, S. Ehrlich and L. Goerigk, *J. Comput. Chem.*, 2011, **32**, year.

[150] J. VandeVondele and J. Hutter, *J. Chem. Phys.*, 2007, **127**, 114105.

[151] S. Goedecker, M. Teter and J. Hutter, *Phys. Rev. B*, 1996, **54**, 1703–1710.

[152] M. Krack, *Theor. Chem. Acc.*, 2005, **114**, 145–152.

[153] G. J. Martyna, D. J. Tobias and M. L. Klein, *J. Chem. Phys.*, 1994, **101**, 4177–4189.

[154] W. Shinoa, M. Shiga and M. Mikami, *Phys. Rev. B*, 2004, **69**, 134103.

[155] M. E. Tuckerman, J. Alejandre, R. López-Rendón, A. L. Jochim and G. J. Martyna, *J. Phys. A: Math. Gen.*, 2006, **39**, 5629–5651.

[156] V. A. Blatov, A. P. Shevchenko and D. M. Proserpio, *Cryst. Growth Des.*, 2014, **14**, 3576–3586.

[157] V. A. Blatov, O. Delgado-Friedrichs, M. O'Keeffe and D. M. Proserpio, *Acta Crystallogr. Sect. A*, 2007, **A63**, 418–425.

[158] S. J. Ramsden, V. Robins and S. T. Hyde, *Acta Crystallogr. Sect. A*, 2009, **A65**, 81–108.

[159] E. V. Alexandrov, V. A. Blatov, A. V. Kochetkov and D. M. Proserpio, *CrystEngComm*, 2011, **13**, 3947–3958.

[160] C. F. Macrae, P. R. Edgington, P. McCabe, E. Pidcock, G. P. Shields, R. Taylor, M. Towler and J. van de Streek, *J. Appl. Crystallogr.*, 2006, **39**, 453–457.

[161] K. A. Brakke, *Exper. Math.*, 1992, **1**, 141–165.

[162] K. A. Brakke, *Triply Periodic Minimal Surfaces*, `http://facstaff.susqu.edu/brakke/evolver/examples/periodic/periodic.html`.

[163] E. Braun, A. F. Zurhelle, W. Thijssen, S. K. Schnell, L.-C. Lin, J. Kim, J. A. Thompson and B. Smit, *Mol. Syst. Des. Eng.*, 2016, **1**, 175–188.

[164] J. M. Simmons, H. Wu, W. Zhou and T. Yildirim, *Energy Environ. Sci.*, 2011, **4**, 2177–2185.

[165] Y. Bae and R. Q. Snurr, *Angew. Chem., Int. Ed.*, 2011, **50**, 11586–11596.

[166] R. Krishna and J. M. van Baten, *Phys. Chem. Chem. Phys.*, 2011, **13**, 10593–10616.

[167] C. E. Wilmer, O. K. Farha, Y. Bae, J. T. Hupp and R. Q. Snurr, *Energy Environ. Sci.*, 2012, **5**, 9849–9856.

[168] J. Kim, A. Maiti, L.-C. Lin, J. K. Stolaroff, B. Smit and R. D. Aines, *Nat. Commun.*, 2013, **4**, 1694.

[169] L.-C. Lin, A. H. Berger, R. L. Martin, J. Kim, J. A. Swisher, K. Jariwala, C. H. Rycroft, A. S. Bhown, M. W. Deem, M. Haranczyk and B. Smit, *Nat. Mater.*, 2012, **11**, 633–641.

[170] B. J. Maring and P. A. Webley, *Int. J. Greenhouse Gas Control*, 2013, **15**, 16–31.

[171] M. M. F. Hasan, E. L. First and C. A. Floudas, *Phys. Chem. Chem. Phys.*, 2013, **15**, 17601–17618.

[172] E. L. First, M. M. F. Hasan and C. A. Floudas, *AIChE J.*, 2014, **60**, 1767–1785.

[173] S. Rege and R. Yang, *Sep. Sci. Technol.*, 2001, **36**, 3355–3365.

[174] A. D. Wiersum, J.-S. Chang, C. Serre and P. L. Llewellyn, *Langmuir*, 2013, **29**, 3301–3309.

[175] J. Kim, R. L. Martin, O. Rübel, M. Haranczyk and B. Smit, *J. Chem. Theory Comput.*, 2012, **8**, 1684–1693.

[176] J. Kim and B. Smit, *J. Chem. Theory Comput.*, 2012, **8**, 2336–2343.

[177] E. García-Pérez, J. B. Parra, C. O. Ania, A. García-Sánchez, J. M. van Baten, R. Krishna, D. Dubbeldam and S. Calero, *Adsorption*, 2007, **13**, 469–476.

[178] D. Peng and D. B. Robinson, *Ind. Eng. Chem. Fundamen.*, 1976, **15**, 59–64.

[179] O. Talu and A. L. Myers, *AIChE J.*, 2001, **47**, 1160–1168.

[180] A. L. Myers and J. M. Prausnitz, *AIChE J.*, 1965, **11**, 121–127.

[181] M. Pinheiro, R. L. Martin, C. H. Rycroft and M. Haranczyk, *CrystEngComm*, 2013, **15**, 7531–7538.

[182] J. Kim, M. Abouelnasr, L.-C. Lin and B. Smit, *J. Am. Chem. Soc.*, 2013, **135**, 7545–7552.

[183] E. Jones, T. Oliphant, P. Peterson *et al.*, *SciPy: Open source scientific tools for Python*, 2001–, `http://www.scipy.org/`, Online; accessed Apr. 8, 2015.

[184] H. Akima, *J. ACM*, 1970, **17**, 589–602.

[185] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, *J. Mach. Learn. Res.*, 2011, **12**, 2825–2830.

[186] C. M. Simon, R. Mercado, S. K. Schnell, B. Smit and M. Haranczyk, *Chem. Mater.*, 2015, **27**, 4459–4475.

[187] E. Braun, A. F. Zurhelle, W. Thijssen, S. K. Schnell, L.-C. Lin, J. Kim, J. A. Thompson and B. Smit, *High-Throughput Computational Screening of Nanoporous Adsorbents for CO2 Capture from Natural Gas*, Open Science Framework, 2016, `http://doi.org/10.17605/OSF.IO/KTBY4`, Online.

[188] J. Boerio-Goates, R. Stevens, B. K. Hom, B. F. Woodfield, P. M. Piccione, M. E. Davis and A. Navrotsky, *J. Chem. Thermodynamics*, 2002, **34**, 205–227.

[189] S. Krishnamurthy, V. R. Rao, S. Guntuka, P. Sharratt, R. Haghpanah, A. Rajendran, M. Amanullah, I. A. Karimi and S. Farooq, *AIChE J.*, 2014, **60**, 1830–1842.

[190] J. M. Smith, H. C. Van Ness and M. M. Abbott, *Introduction to Chemical Engineering Thermodynamics*, McGraw-Hill, 7th edn., 2005.

[191] S. K. Bhatia and A. L. Myers, *Langmuir*, 2006, **22**, 1688–1700.

[192] C. M. Simon, J. Kim, D. A. Gomez-Gualdron, J. S. Camp, Y. G. Chung, R. L. Martin, R. Mercado, M. W. Deem, D. Gunter, M. Haranczyk, D. S. Sholl, R. Q. Snurr and B. Smit, *Energy Environ. Sci.*, 2015, **8**, 1190–1199.

[193] L. Breiman, *Machine Learning*, 2001, **45**, 5–32.

[194] L. Breiman, J. Friedman, C. J. Stone and R. Olshen, *Classification and Regression Trees*, Chapman & Hall/CRC, 1984.

[195] C. M. Simon, J. Kim, L.-C. Lin, R. L. Martin, M. Haranczyk and B. Smit, *Phys. Chem. Chem. Phys.*, 2014, **16**, 5499–5513.

[196] E. Haldoupis, S. Nair and D. S. Sholl, *Phys. Chem. Chem. Phys.*, 2011, **13**, 5053–5060.

[197] H. Amrouche, B. Creton, F. Siperstein and C. Nieto-Draghi, *RSC Adv.*, 2012, **2**, 6028–6035.

[198] E. J. García, J. Pérez-Pellitero, C. Jallut and G. D. Pirngruber, *Phys. Chem. Chem. Phys.*, 2013, **15**, 5648–5657.

[199] M. Fernandez, T. K. Woo, C. E. Wilmer and R. Q. Snurr, *J. Phys. Chem. C*, 2013, **117**, 7681–7689.

[200] D. Wu, Q. Yang, C. Zhong, D. Liu, H. Huang, W. Zhang and G. Maurin, *Langmuir*, 2012, **28**, 12094–12099.

[201] Z. Babarao, R.and Hu, J. Jiang, S. Chempath and S. I. Sandler, *Langmuir*, 2007, **23**, 659–666.

[202] H. Chen and D. S. Sholl, *Langmuir*, 2007, **23**, 6431–6437.

[203] R. Krishna and J. M. van Baten, *Sep. Purif. Technol.*, 2008, **61**, 414–423.

[204] Y.-S. Bae, K. L. Mulfort, H. Frost, P. Ryan, S. Punnathanam, L. J. Broadbelt, J. T. Hupp and R. Q. Snurr, *Langmuir*, 2008, **24**, 8592–8598.

[205] S. E. Jee and D. S. Sholl, *J. Am. Chem. Soc.*, 2009, **131**, 7896–7904.

[206] N. F. Cessford, N. A. Seaton and T. Düren, *Ind. Eng. Chem. Res.*, 2012, **51**, 4911–4921.

[207] J. A. Swisher, L.-C. Lin, J. Kim and B. Smit, *AIChE J.*, 2013, **59**, 3054–3064.

[208] A. Goj, D. S. Sholl, E. D. Akten and D. Kohen, *J. Phys. Chem. B*, 2002, **106**, 8367–8375.

[209] E. D. Akten, R. Siriwardane and D. S. Sholl, *Energy Fuels*, 2003, **17**, 977–983.

[210] M. Murthi and R. Q. Snurr, *Langmuir*, 2004, **20**, 2489–2497.

[211] M. Sakuth, J. Meyer and J. Gmehling, *Chem. Eng. Process. Process Intensif.*, 1998, **37**, 267–277.

[212] E. Haldoupis, S. Nair and D. S. Sholl, *J. Am. Chem. Soc.*, 2012, **134**, 4313–4323.

[213] J. Kim, L.-C. Lin, J. A. Swisher, M. Haranczyk and B. Smit, *J. Am. Chem. Soc.*, 2012, **134**, 18940–18943.

[214] H. Furukawa, K. E. Cordova, M. O'Keeffe and O. M. Yaghi, *Science*, 2013, **341**, 1230444.

[215] A. U. Czaja, N. Trukhan and U. Müller, *Chem. Soc. Rev.*, 2009, **38**, 1284–1293.

[216] G. Férey and C. Serre, *Chem. Soc. Rev.*, 2009, **38**, 1380–1399.

[217] A. Ghysels, L. Vanduyfhuys, M. Vandichel, M. Waroquier, V. Van Speybroeck and B. Smit, *J. Phys. Chem. C*, 2013, **117**, 11540–11554.

[218] B. Smit and T. L. M. Maesen, *Nature*, 1995, **374**, 42–44.

[219] F. Stallmach, S. Gröger, V. Künzel, J. Kärger, O. M. Yaghi, M. Hesse and U. Müller, *Angew. Chem., Int. Ed.*, 2006, **45**, 2123–2126.

[220] S. Hertel, M. Wehring, S. Amirjalayer, M. Gratz, J. Lincke, H. Krautscheid, R. Schmid and F. Stallmach, *Eur. Phys. J. Appl. Phys.*, 2011, **55**, 20702.

[221] D. C. Ford, D. Dubbeldam, R. Q. Snurr, V. Künzel, M. Wehring, F. Stallmach, J. Kärger and U. Müller, *J. Phys. Chem. Lett.*, 2012, **3**, 930–933.

[222] W. Shim, K. Hwang, J. Chung, Y. Baek, S. Yoo, S. Kim, H. Moon and J. Lee, *Adv. Powder Technol.*, 2012, **23**, 615–619.

[223] W. Makowski, M. Mańko, P. Zabierowski, K. Mlekodaj, D. Majda, J. Szklarzewicz and W. Łasocha, *Thermochim. Acta*, 2014, **587**, 1–10.

[224] L. Sarkisov, T. Düren and R. Q. Snurr, *Mol. Phys.*, 2004, **102**, 211–221.

[225] M. De Toni, P. Pullumbi, F. Coudert and A. H. Fuchs, *J. Phys. Chem. C*, 2010, **114**, 21631–21637.

[226] K. S. Walton, A. R. Millward, D. Dubbeldam, H. Frost, J. J. Low, O. M. Yaghi and R. Q. Snurr, *J. Am. Chem. Soc.*, 2008, **130**, 406–407.

[227] R. Krishna and J. M. van Baten, *Langmuir*, 2010, **26**, 3981–3992.

[228] K. Takakura, T. Ueda, K. Miyakubo and T. Eguchi, *Phys. Chem. Chem. Phys.*, 2013, **15**, 279–290.

[229] M. Eddaoudi, H. Li and O. M. Yaghi, *J. Am. Chem. Soc.*, 2000, **122**, 1391–1397.

[230] H. Y. Carr and E. M. Purcell, *Phys. Rev.*, 1954, **94**, 630–638.

[231] S. Meiboom and D. Gill, *Rev. Sci. Instrum.*, 1958, **29**, 688–691.

[232] L. Venkataramanan, Y. Song and M. D. Hürlimann, *IEEE Trans. Signal Process.*, 2002, **50**, 1017–1026.

[233] Y. Song, *J. Magn. Reson.*, 2013, **229**, 12–24.

[234] V. Vaidyanathan, *Phys. Rev.*, 1927, **30**, 512–515.

[235] M. D. Hürlimann, *J. Magn. Reson.*, 1998, **131**, 232–240.

[236] J. Mitchell, T. C. Chandrasekera, M. L. Johns, L. F. Gladden and E. J. Fordham, *Phys. Rev. E*, 2010, **81**, 026101.

[237] A. P. Y. Wong and M. H. W. Chan, *Phys. Rev. Lett.*, 1990, **65**, 2567–2570.

[238] J. P. Donley and A. J. Liu, *Phys. Rev. E*, 1997, **55**, 539–543.

[239] H. Wu, W. Zhou and T. Yildirim, *J. Phys. Chem. C*, 2009, **113**, 3029–3035.

[240] X. Kong, H. Deng, F. Yan, J. Kim, J. A. Swisher, B. Smit, O. M. Yaghi and J. A. Reimer, *Science*, 2013, **341**, 882–885.

[241] M. A. Addicoat, N. Vankova, I. F. Akter and T. Heine, *J. Chem. Theory Comput.*, 2014, **10**, 880–891.

[242] C. D. Wick, M. G. Martin and J. I. Siepmann, *J. Phys. Chem. B*, 2000, **104**, 8008–8016.

[243] M. Eddaoudi, J. Kim, N. Rosi, D. Vodak, J. Wachter, M. O'Keeffe and O. M. Yaghi, *Science*, 2002, **295**, 469–472.

[244] Z. Gu, D. Jiang, H. Wang, X. Cui and X. Yan, *J. Phys. Chem. C*, 2010, **114**, 311–316.

[245] T. M. McDonald, J. A. Mason, X. Kong, E. D. Bloch, D. Gygi, A. Dani, V. Crocellà, F. Giordanino, S. O. Odoh, W. S. Drisdell, B. Vlaisavljevich, A. Dzubak, R. Poloni, S. Schnell, N. Planas, K. Lee, T. Pascal, L. F. Wan, D. Prendergast, J. B. Neaton, B. Smit, J. B. Kortright, L. Gagliardi, S. Bordia, J. A. Reimer and J. R. Long, *Nature*, 2015, **519**, 303–308.

[246] E. Kockrick, S. Schrage, L. Borchardt, N. Klein, M. Rose, I. Senkovska and S. Kaskel, *Carbon*, 2010, **48**, 1707–1717.

# Appendix A

# Supplemental figures and tables
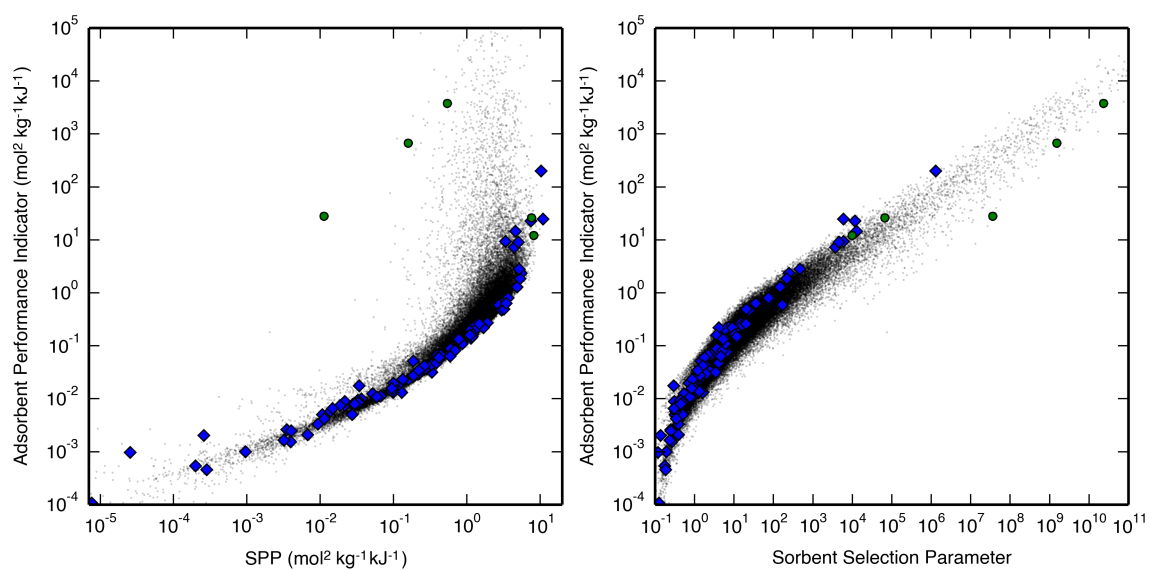
Figure A.1: (left) A correlation between the SPP and the Adsorbent Performance Indicator and (right) between the Sorbent Selection Parameter and the Adsorbent Performance Indicator, both for the APG process carried out with PSA. Hypothetical zeolites are shown as black dots, IZA zeolites are shown as blue diamonds, and the hypothetical zeolites shown in Table 4.1 are shown as green circles.
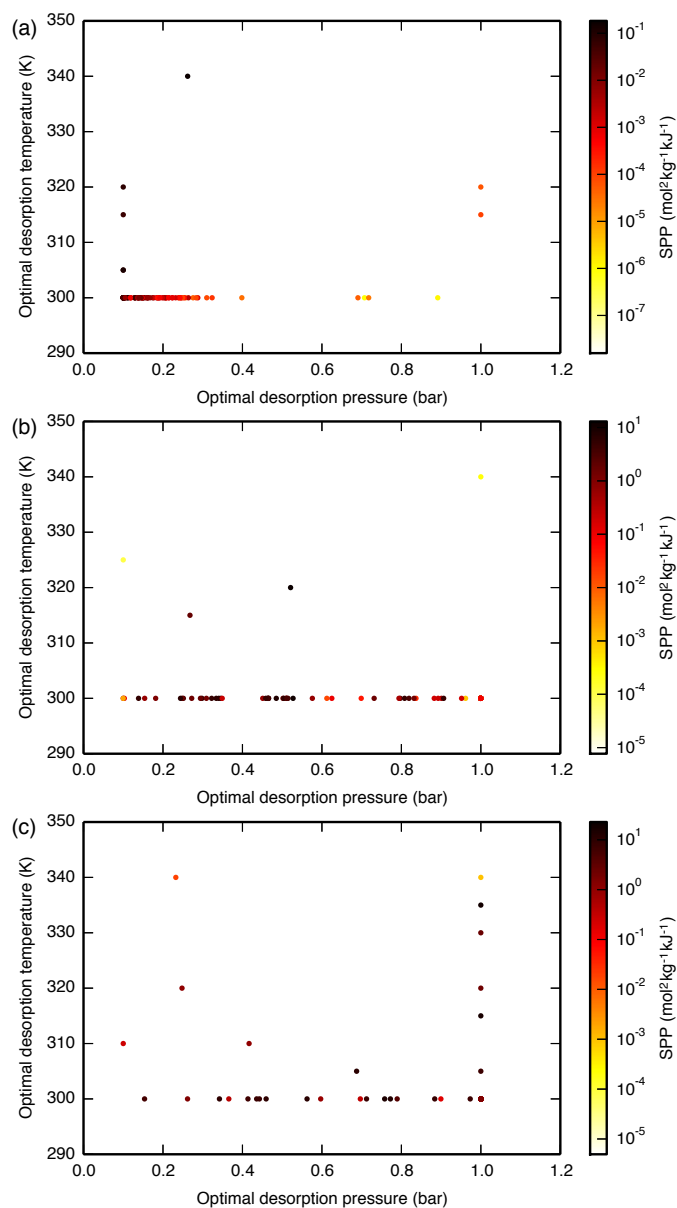
Figure A.2: SPP of the IZA zeolites for the (a) LFG, (b) APG, and (c) NAG processes carried out with PTSA as a function of the optimal desorption pressure and temperature. Note that the APG and NAG processes have several data points overlapping at 1 bar and 300 K. For the LFG process, a clear correlation exists between the desorption conditions and the SPP: zeolites with lower optimized desorption pressures are also better-performing. By comparing the SPP values of these materials at their optimal desorption conditions with the SPP values of these materials at set desorption conditions of 0.1 bar and 300 K (Fig. A.30) it becomes clear that it was not the lowered desorption pressure that caused some materials to perform better than others; rather, top-performing materials have more to gain by pulling additional vacuum than do the poor-performing materials.
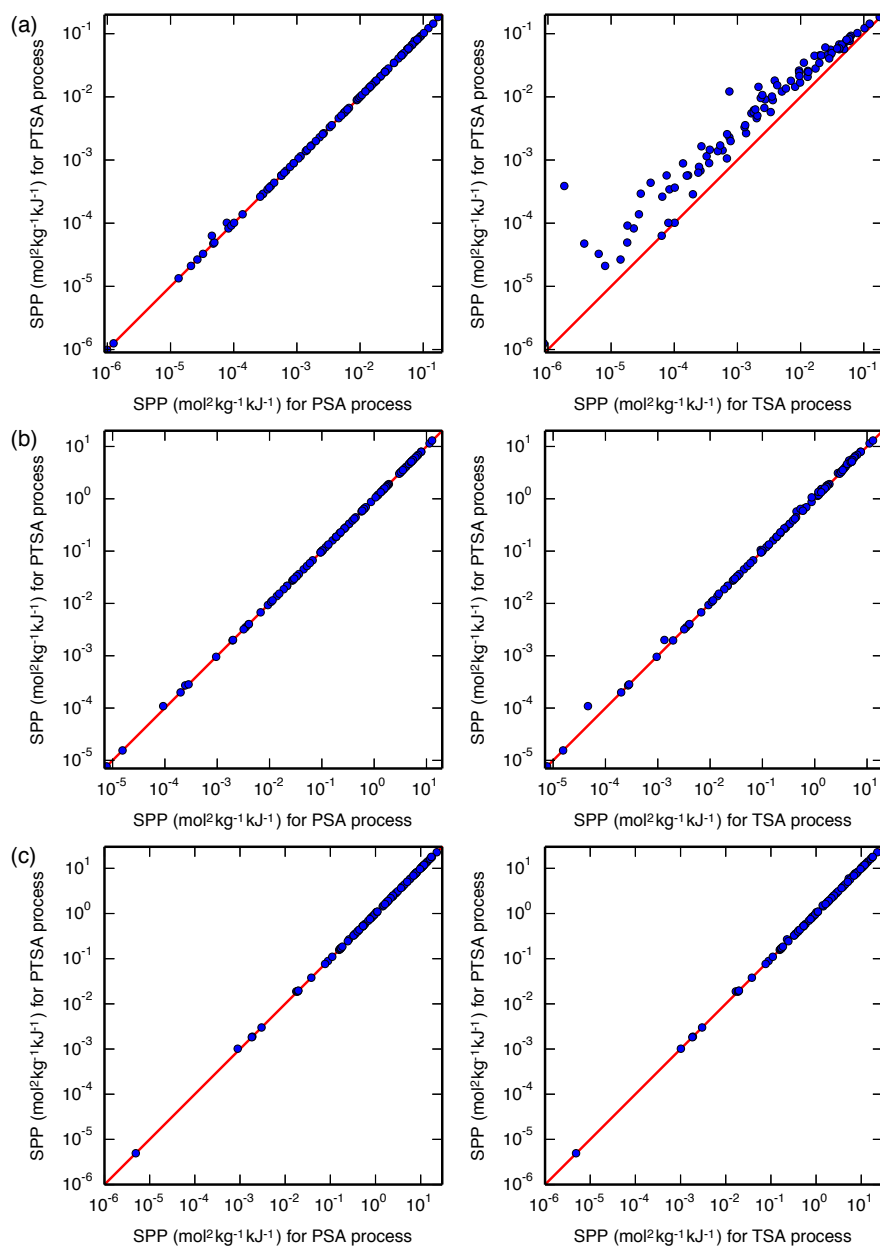
Figure A.3: SPP of the IZA zeolites for the (a) LFG, (b) APG, and (c) NAG processes carried out with PTSA at optimal desorption temperatures and pressures is plotted (left) against the SPP of the same process carried out with PSA at a set desorption temperature of 300 K and an optimal desorption pressure or (right) against the SPP of the same process carried out with TSA at an optimal desorption temperature and a set desorption pressure of 1 bar. Of the TSA processes, only the LFG process appears to have benefited from pulling vacuum, while the two higher-pressure processes received a "free" pressure-swing down to 1 bar, and thus do not benefit as much from the additional vacuum. A line is drawn at $y = x$ for reference.
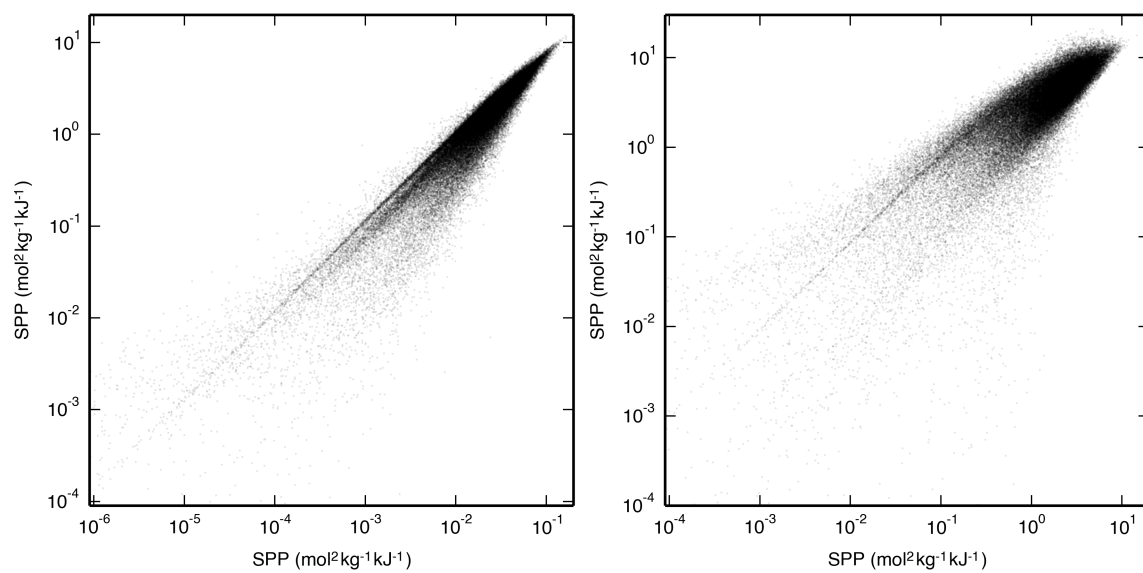
Figure A.4:   SPP of the hypothetical zeolites for the (a) LFG and APG processes and for the (b) APG and NAG processes, all carried out with PSA.
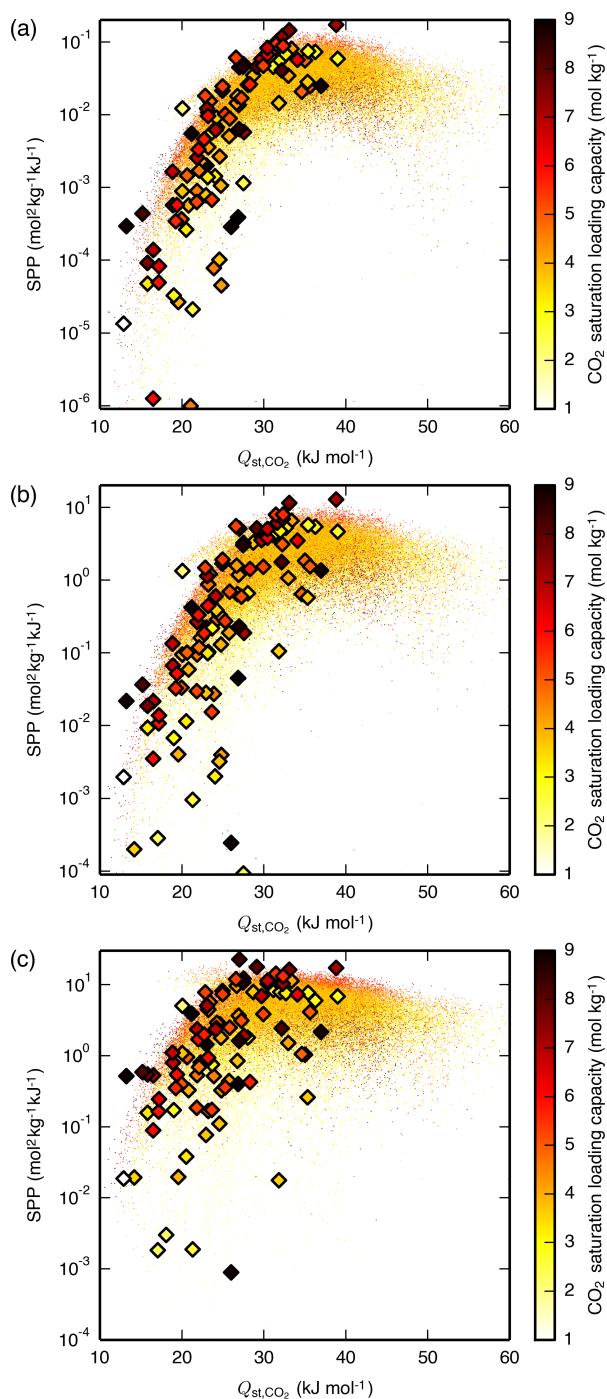
Figure A.5: SPP as a function of $Q_{st,CO_2}$ and the $CO_2$ saturation loading capacity, for the (a) LFG, (b) APG, and (c) NAG processes, all carried out with PSA. The materials are plotted in random order such that the data shown are representative of the materials hidden due to having similar $Q_{st,CO_2}$ and SPP. IZA zeolites are shown as diamonds. This plot is the same as in Fig. 4.2, but with the IZA zeolites overlaid.
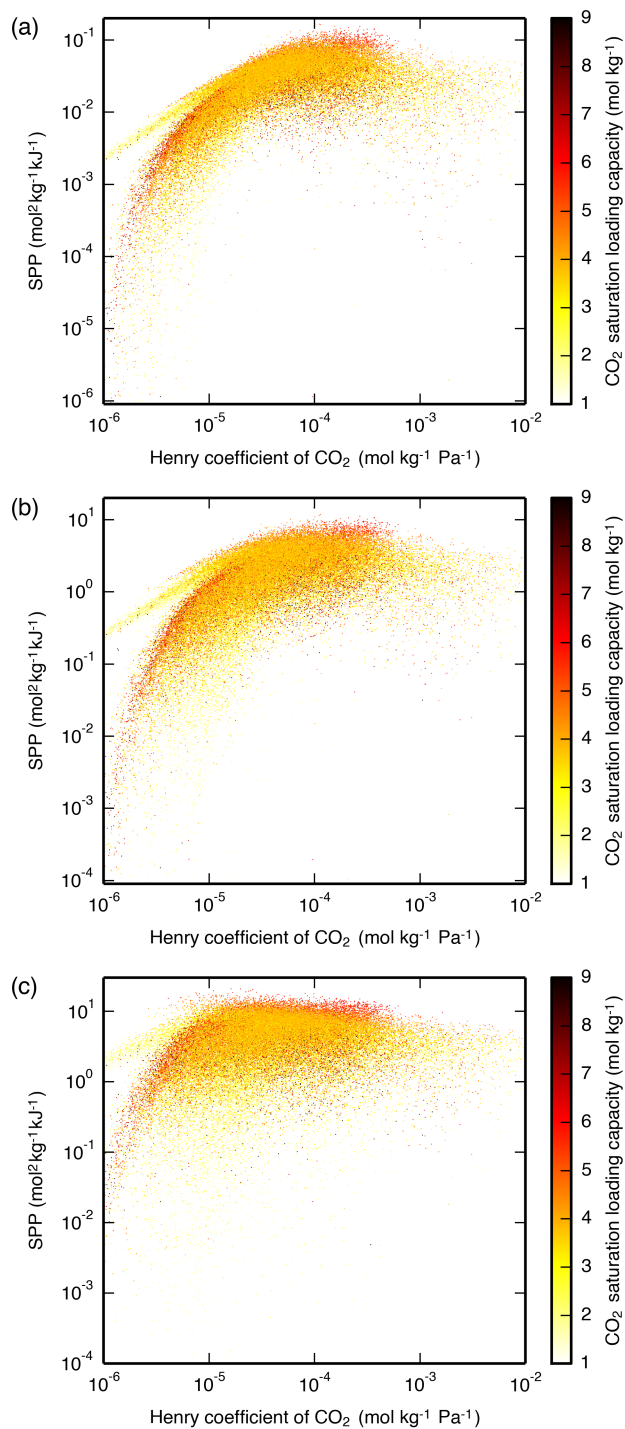
Figure A.6: SPP of the hypothetical zeolites as a function of the Henry coefficient and $CO_2$ saturation loading capacity, for the (a) LFG, (b) APG, and (c) NAG processes, all carried out with PSA. The materials are plotted in random order such that the data shown are representative of the materials hidden due to having similar $Q_{st,CO_2}$ and SPP.
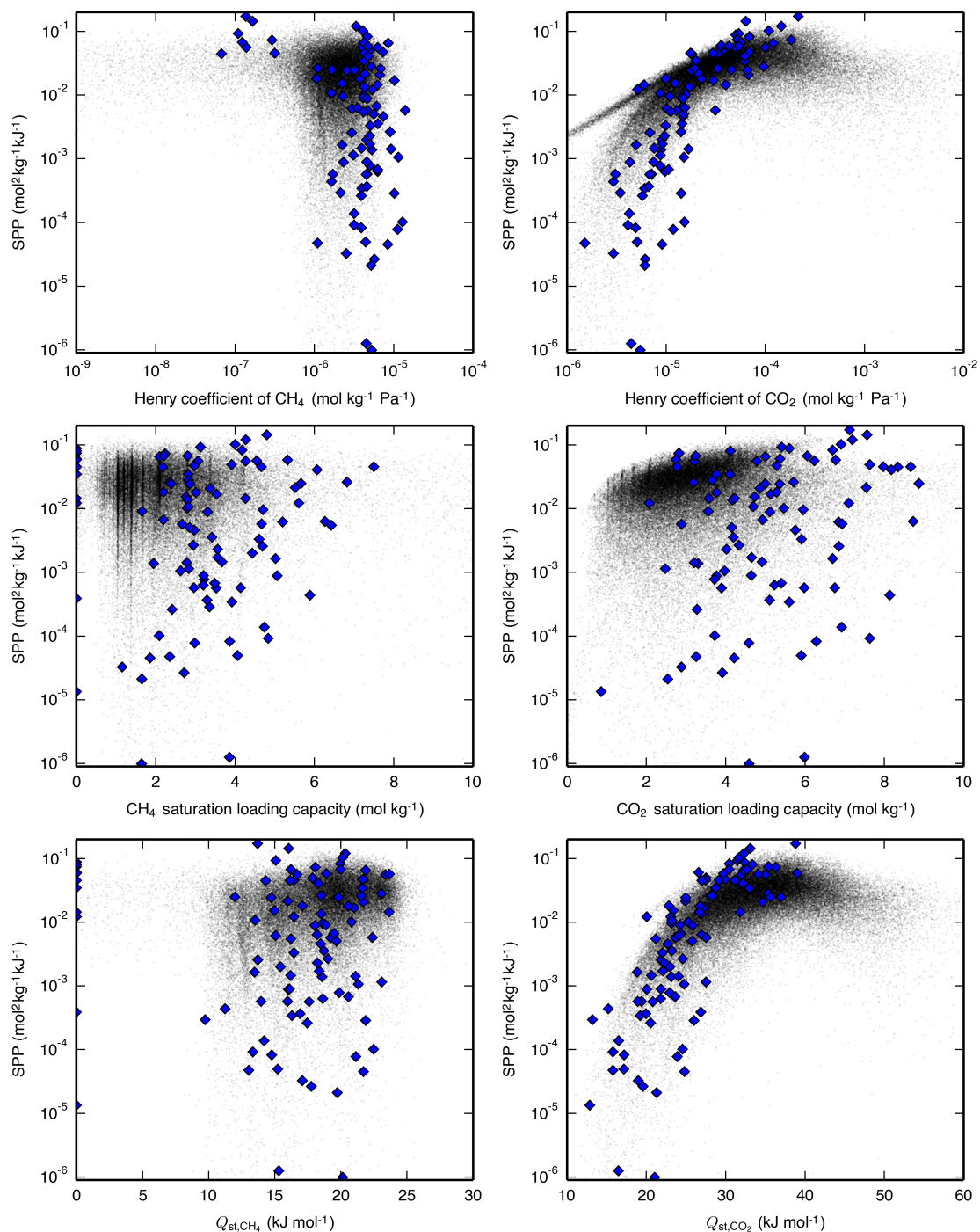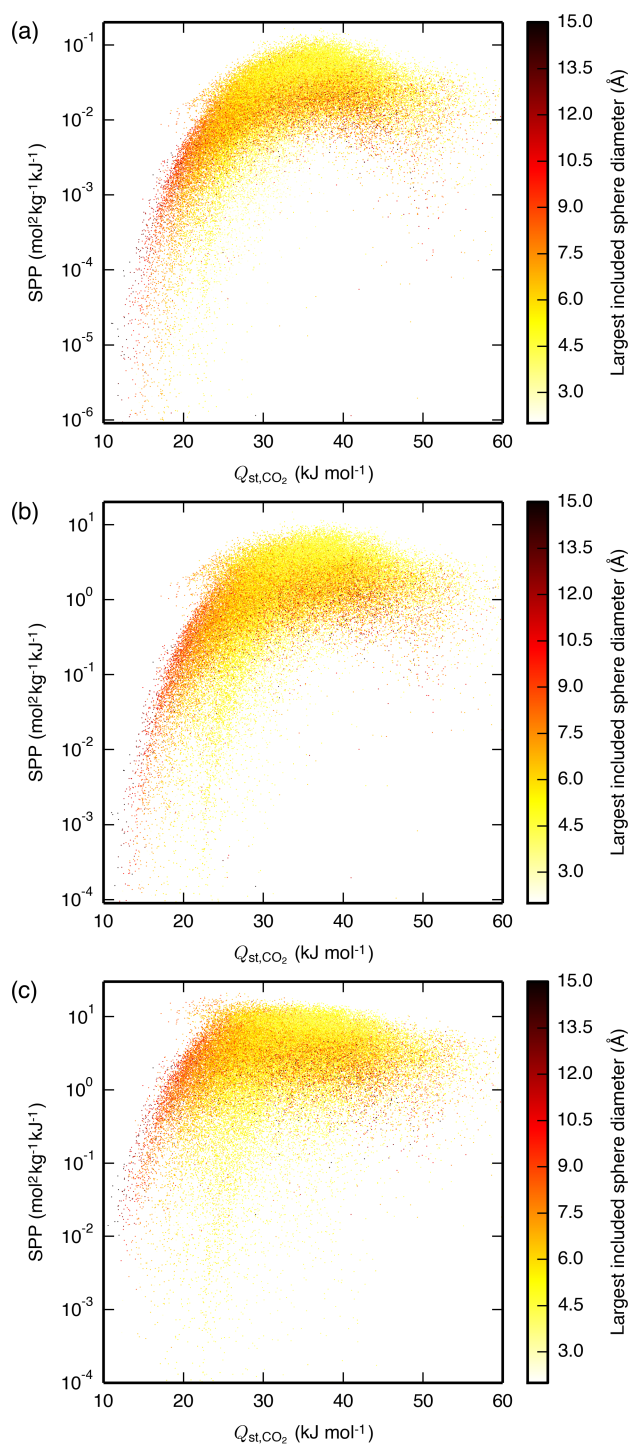
Figure A.7: SPP as a function of Henry coefficients, saturation loadings, and $Q_{st}$ for the LFG process carried out with PSA. Hypothetical zeolites are shown as black dots, and IZA zeolites are shown as blue diamonds. The stripes seen in plots containing $CH_4$ saturation loading capacity are due to that variable being more likely to be integer values of $CH_4$ molecules per unit cell (Fig. A.10).
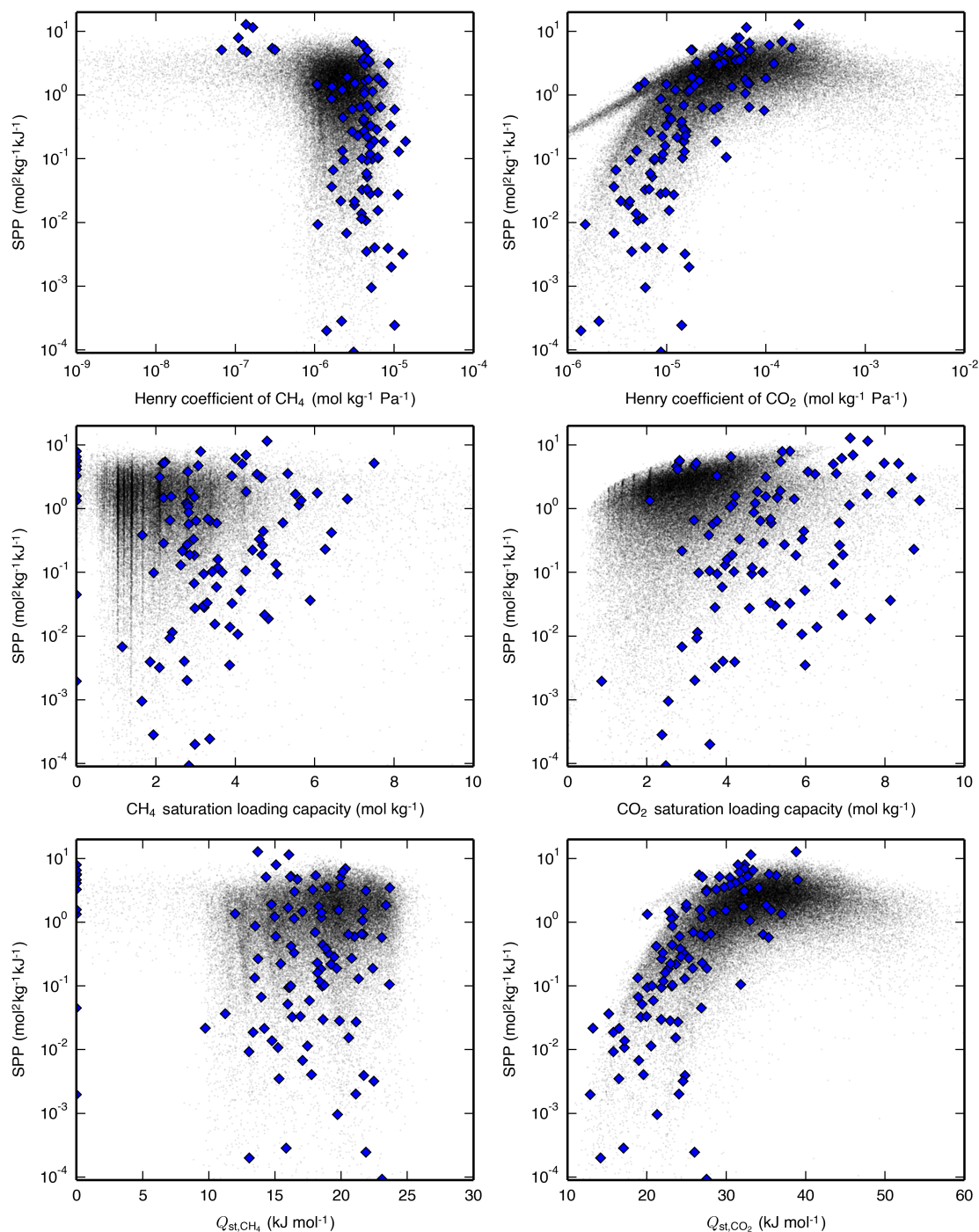
Figure A.8:   SPP as a function of Henry coefficients, saturation loadings, and $Q_{st}$ for the APG process carried out with PSA. Hypothetical zeolites are shown as black dots, and IZA zeolites are shown as blue diamonds. The stripes seen in plots containing $CH_4$ saturation loading capacity are due to that variable being more likely to be integer values of $CH_4$ molecules per unit cell (Fig. A.10).
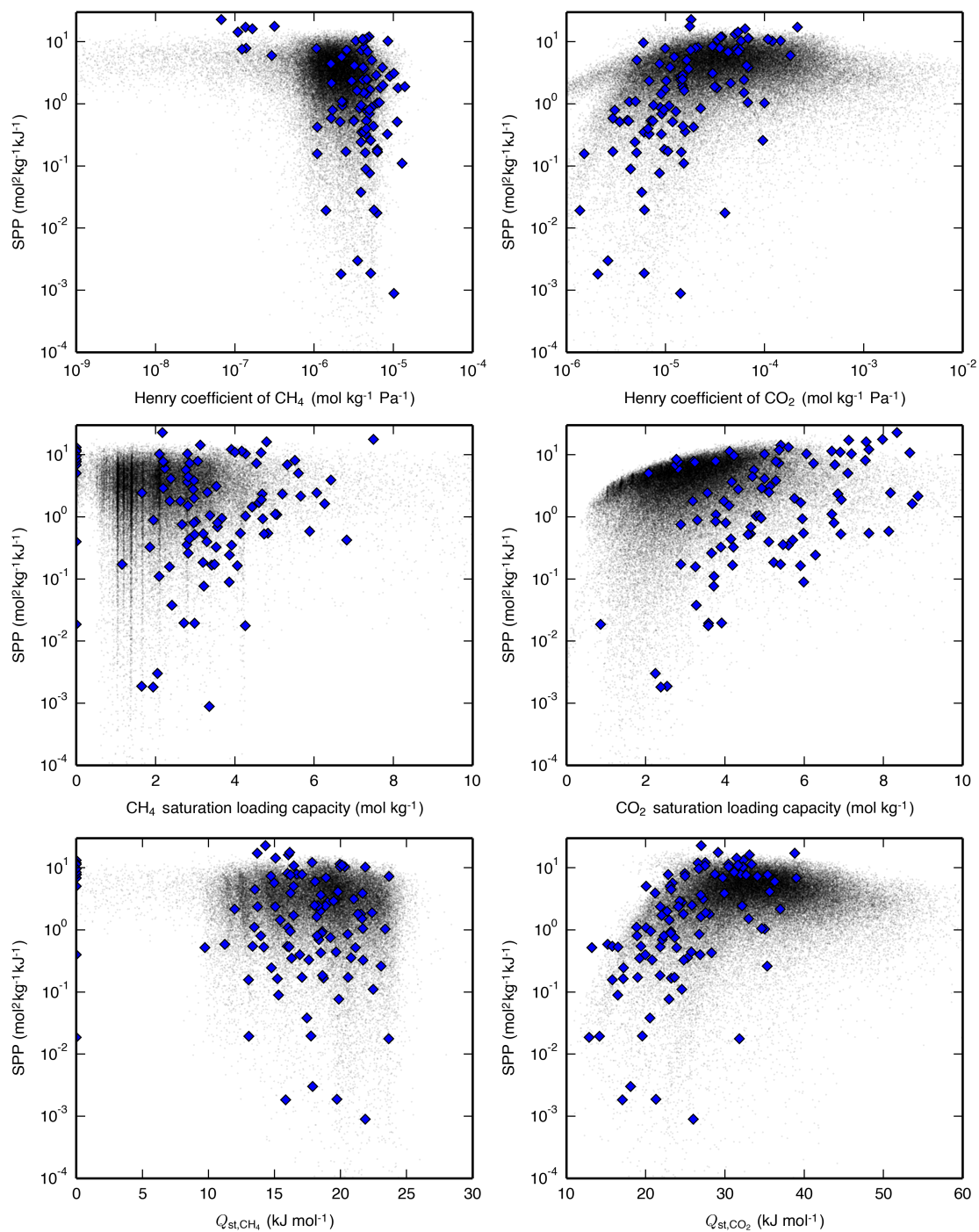
Figure A.9:  SPP as a function of Henry coefficients, saturation loadings, and $Q_{st}$ for the NAG process carried out with PSA. Hypothetical zeolites are shown as black dots, and IZA zeolites are shown as blue diamonds. The stripes seen in plots containing $CH_4$ saturation loading capacity are due to that variable being more likely to be integer values of $CH_4$ molecules per unit cell (Fig. A.10).
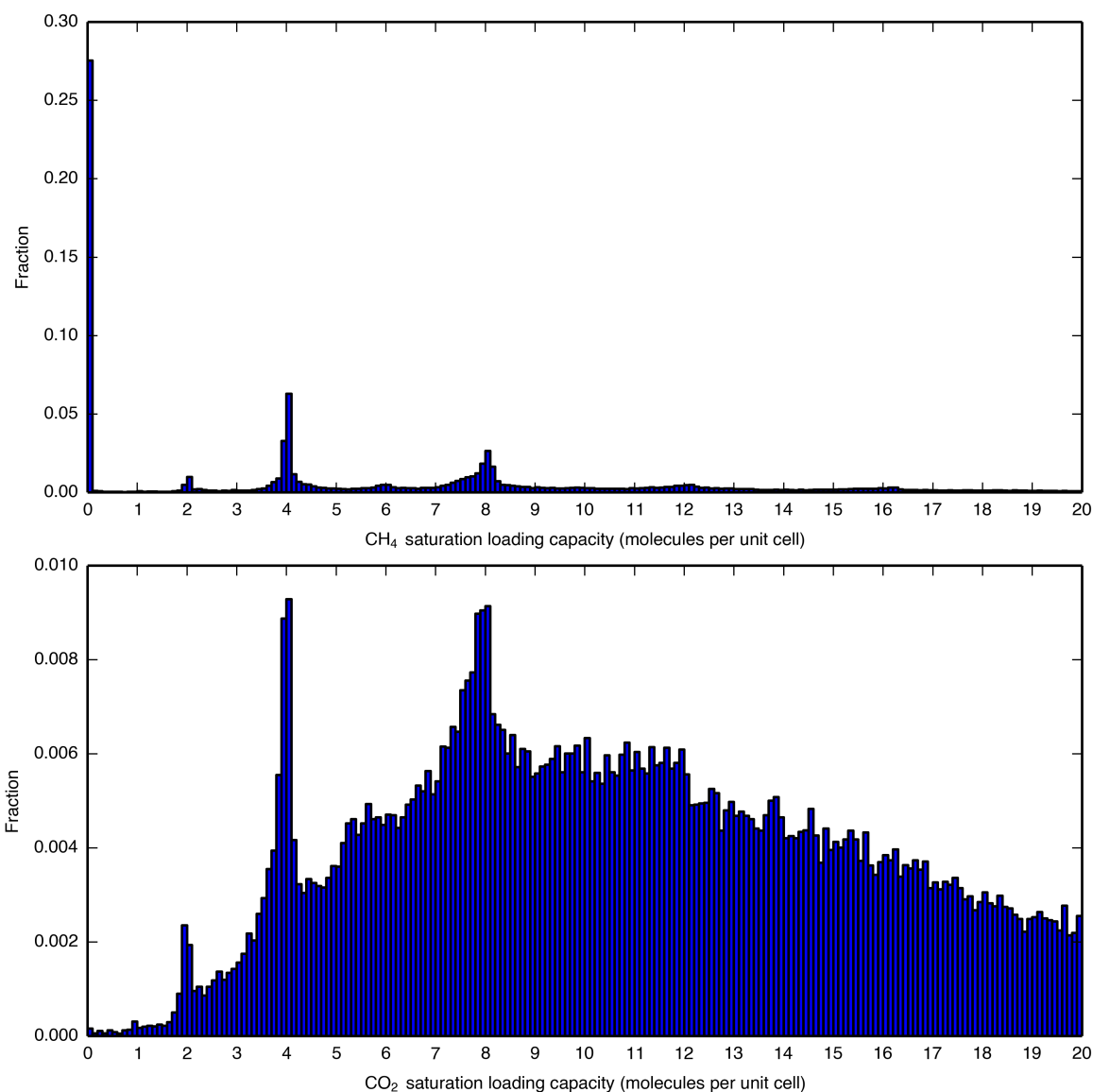
Figure A.10: The distribution of (top) $CH_4$ saturation loading capacities for the hypothetical zeolites shows that even integer values in units of molecules per unit cell are most common, a trend which is less noticeable for (bottom) $CO_2$ saturation loading capacities. This explains the stripes seen in Figs. 4.5, A.7–A.9, and A.13 in the plots containing $CH_4$ saturation loading capacity. Both histograms have been truncated at 20 molecules per unit cell.
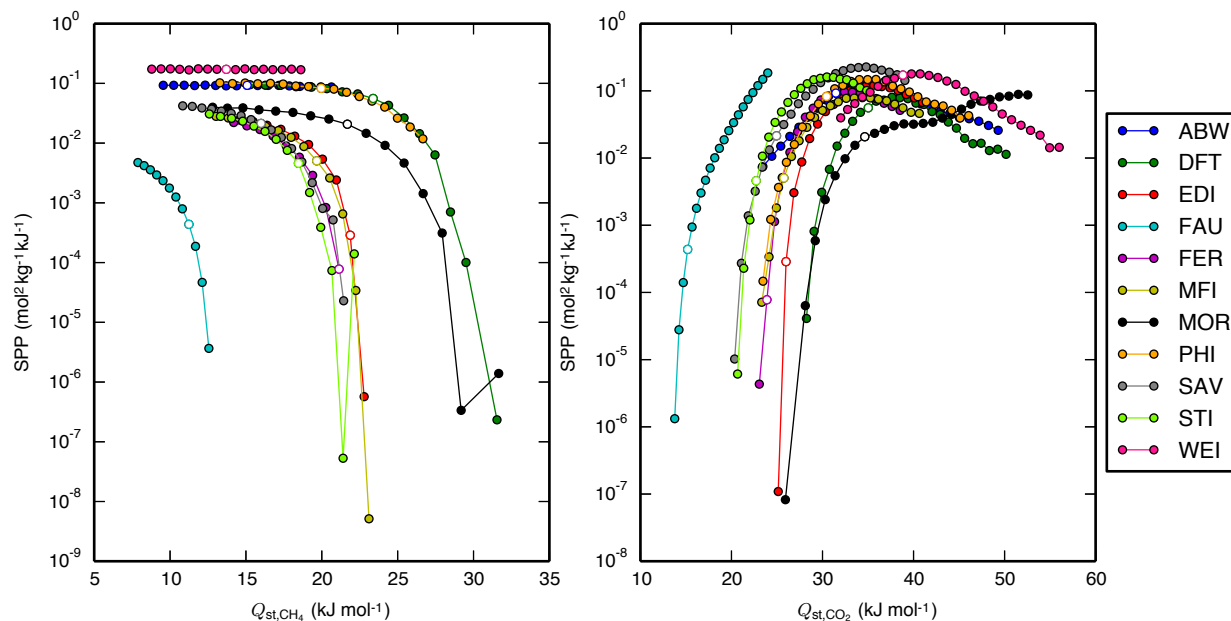
Figure A.11:   SPP of a subset of IZA zeolites undergoing the LFG PSA process as a function of (left) $Q_{\text{st,CH}_4}$ and (right) $Q_{\text{st,CO}_2}$. The points with white marker color represent the data with original guest-host epsilon parameters of the Lennard-Jones potential (115.00 K for $CH_4$-$O_{\text{zeo}}$, 50.20 K for $C_{CO_2}$-$O_{\text{zeo}}$, and 84.93 K for $O_{CO_2}$-$O_{\text{zeo}}$). (left) The $CH_4$-$O_{\text{zeo}}$ epsilon value was varied between 75 K and 155 K in increments of 5 K (4.35 % of the original value), with larger values resulting in a larger $Q_{\text{st,CH}_4}$.   (right) The $C_{CO_2}$-$O_{\text{zeo}}$ epsilon value was varied between 32.128 K and 88.603 K in increments of 2.259 K while the $O_{CO_2}$-$O_{\text{zeo}}$ epsilon value was concurrently varied between 54.354 K and 149.904 K in increments of 3.822 K (both 4.50 % of the original values), with larger values resulting in a larger $Q_{\text{st,CO}_2}$. When $Q_{\text{st,CH}_4}$ was brought too high or $Q_{\text{st,CO}_2}$ was brought too low, the material would become unfit for the separation, so these points are not shown. Note that SPP of the zeolite WEI does not change with $Q_{\text{st,CH}_4}$ because $CH_4$ loading is negligibly low at all $Q_{\text{st,CH}_4}$ values.

Figure A.12: SPP of a subset of IZA zeolites undergoing the NAG PSA process as a function of (left) $Q_{st,CH_4}$ and (right) $Q_{st,CO_2}$. The points with white marker color represent the data with original guest-host epsilon parameters of the Lennard-Jones potential (115.00 K for $CH_4$-$O_{zeo}$, 50.20 K for $C_{CO_2}$-$O_{zeo}$, and 84.93 K for $O_{CO_2}$-$O_{zeo}$). (left) The $CH_4$-$O_{zeo}$ epsilon value was varied between 75 K and 155 K in increments of 5 K (4.35 % of the original value), with larger values resulting in a larger $Q_{st,CH_4}$. (right) The $C_{CO_2}$-$O_{zeo}$ epsilon value was varied between 32.128 K and 88.603 K in increments of 2.259 K while the $O_{CO_2}$-$O_{zeo}$ epsilon value was concurrently varied between 54.354 K and 149.904 K in increments of 3.822 K (both 4.50 % of the original values), with larger values resulting in a larger $Q_{st,CO_2}$. When $Q_{st,CH_4}$ was brought too high or $Q_{st,CO_2}$ was brought too low, the material would become unfit for the separation, so these points are not shown. Note that SPP of the zeolite WEI does not change with $Q_{st,CH_4}$ because $CH_4$ loading is negligibly low at all $Q_{st,CH_4}$ values.

Figure A.13: Relationships between a zeolite's geometric descriptors and its saturation loading capacities. Hypothetical zeolites are shown as black dots, and IZA zeolites are shown as blue diamonds. The stripes seen in plots containing $CH_4$ saturation loading capacity are due to that variable being more likely to be integer values of $CH_4$ molecules per unit cell (Fig. A.10).
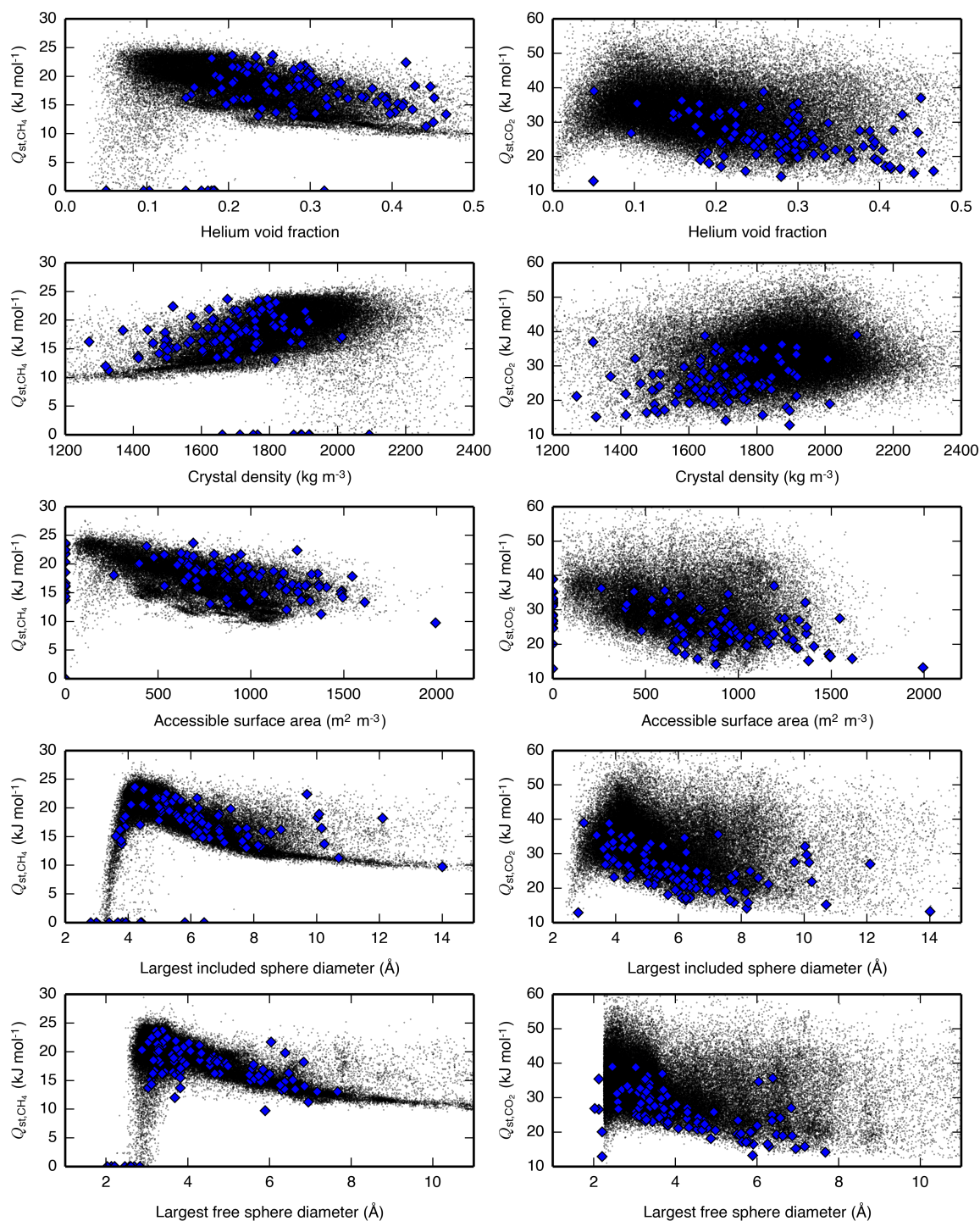
Figure A.14: Relationships between a zeolite's geometric descriptors and $Q_{st}$. Hypothetical zeolites are shown as black dots, and IZA zeolites are shown as blue diamonds.
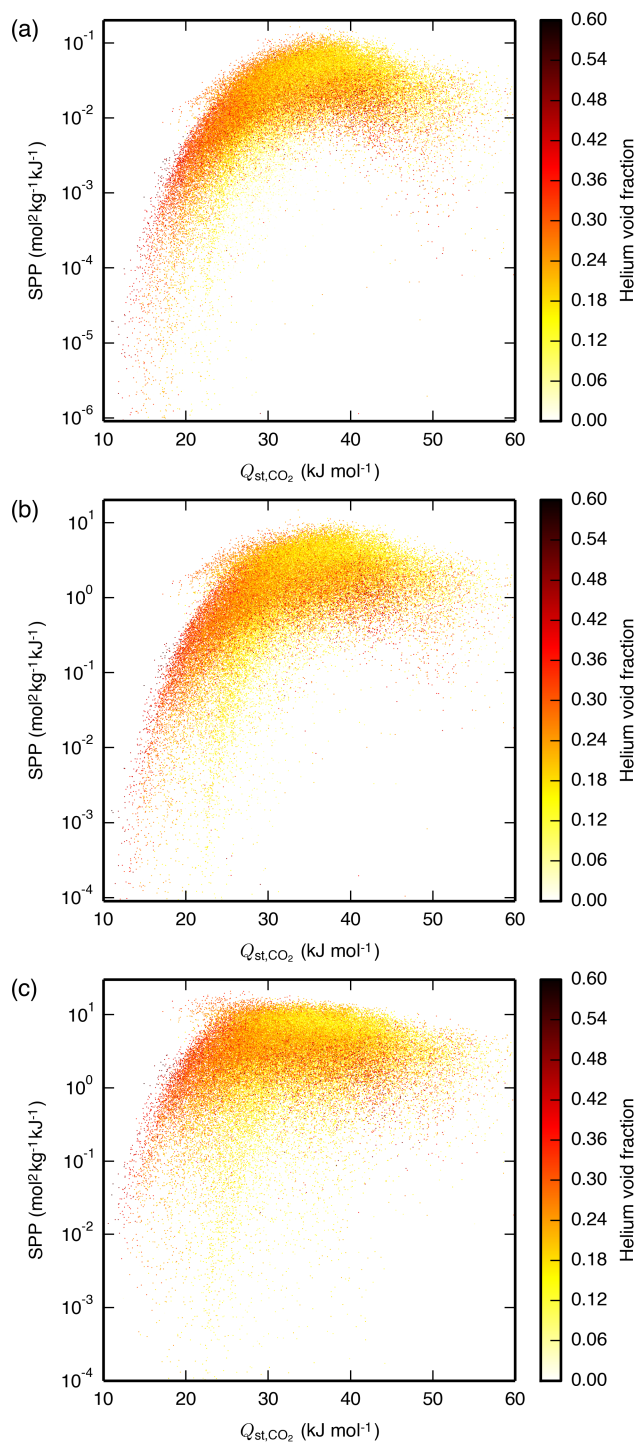
Figure A.15: SPP of the hypothetical zeolites as a function of $Q_{\text{st,CO}_2}$ and helium void fraction, for the (a) LFG, (b) APG, and (c) NAG processes, all carried out with PSA. The materials are plotted in random order such that the data shown are representative of the materials hidden due to having similar $Q_{\text{st,CO}_2}$ and SPP.
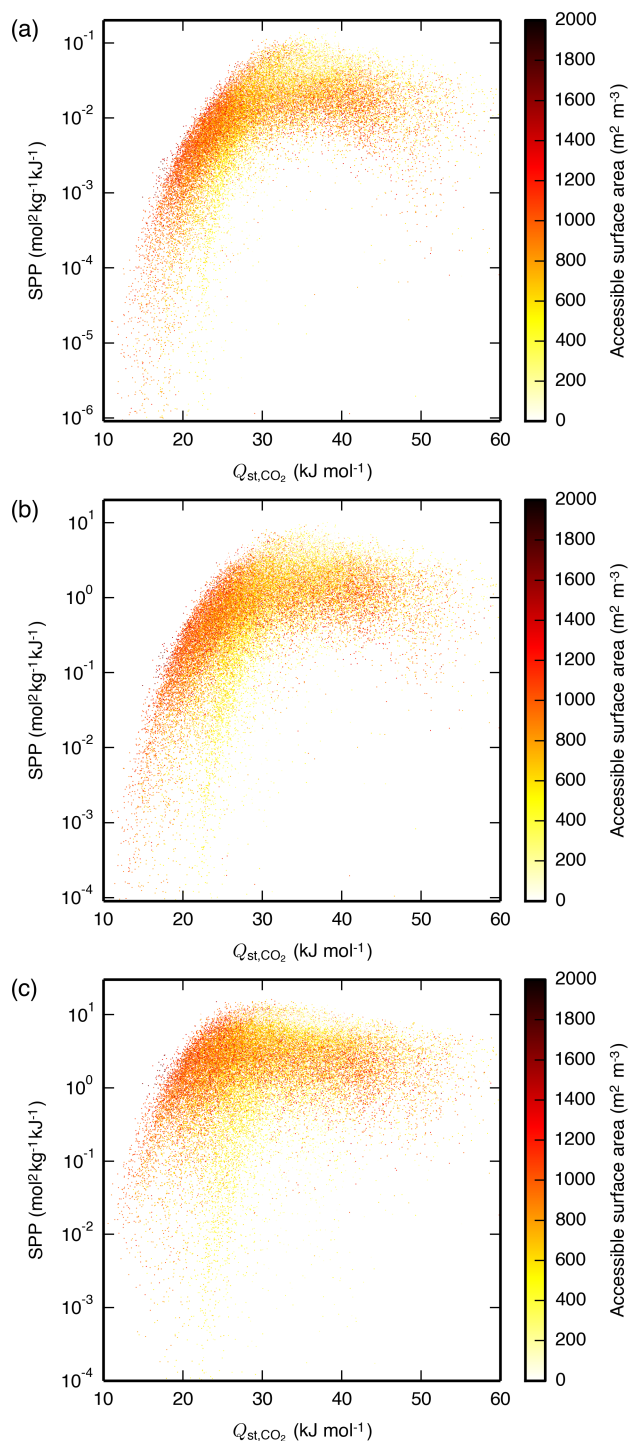
Figure A.16: SPP of the hypothetical zeolites as a function of $Q_{st,CO_2}$ and accessible surface area, for the (a) LFG, (b) APG, and (c) NAG processes, all carried out with PSA. The materials are plotted in random order such that the data shown are representative of the materials hidden due to having similar $Q_{st,CO_2}$ and SPP.

Figure A.17: SPP of the hypothetical zeolites as a function of $Q_{st,CO_2}$ and largest included sphere diameter, for the (a) LFG, (b) APG, and (c) NAG processes, all carried out with PSA. The materials are plotted in random order such that the data shown are representative of the materials hidden due to having similar $Q_{st,CO_2}$ and SPP.
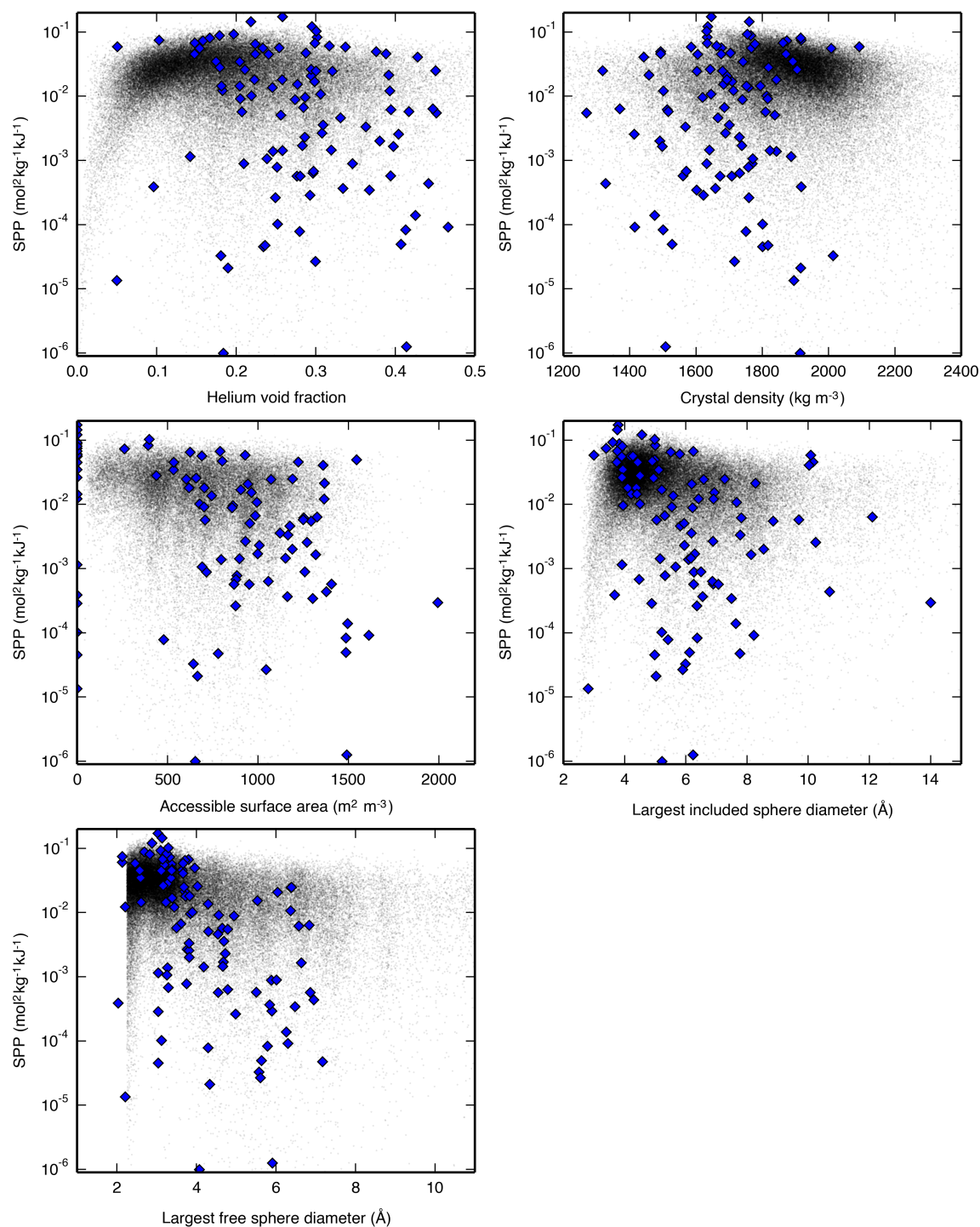
Figure A.18: SPP as a function of geometric parameters for the LFG process carried out with PSA. Hypothetical zeolites are shown as black dots, and IZA zeolites are shown as blue diamonds.
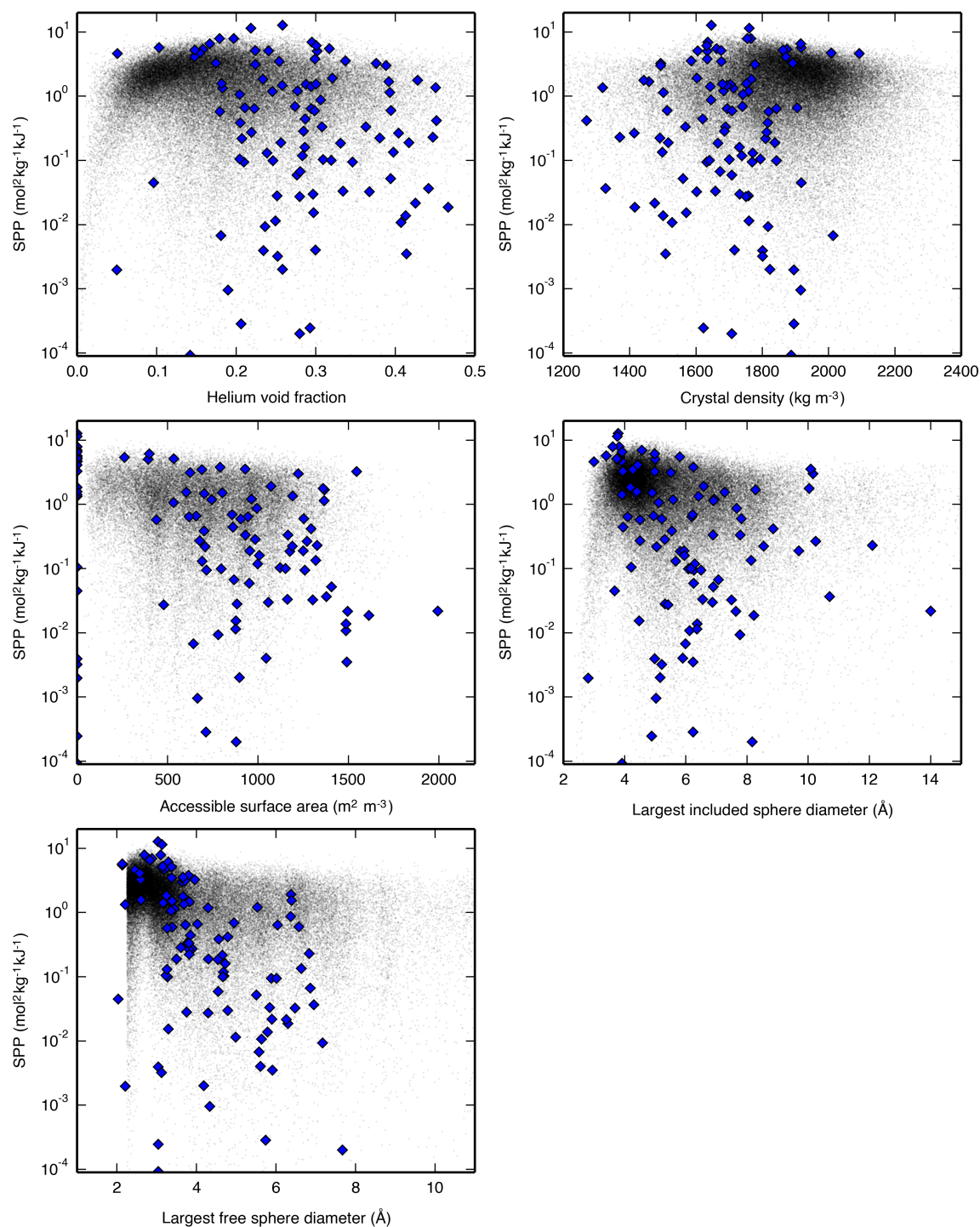
Figure A.19:   SPP as a function of geometric parameters for the APG process carried out with PSA. Hypothetical zeolites are shown as black dots, and IZA zeolites are shown as blue diamonds.
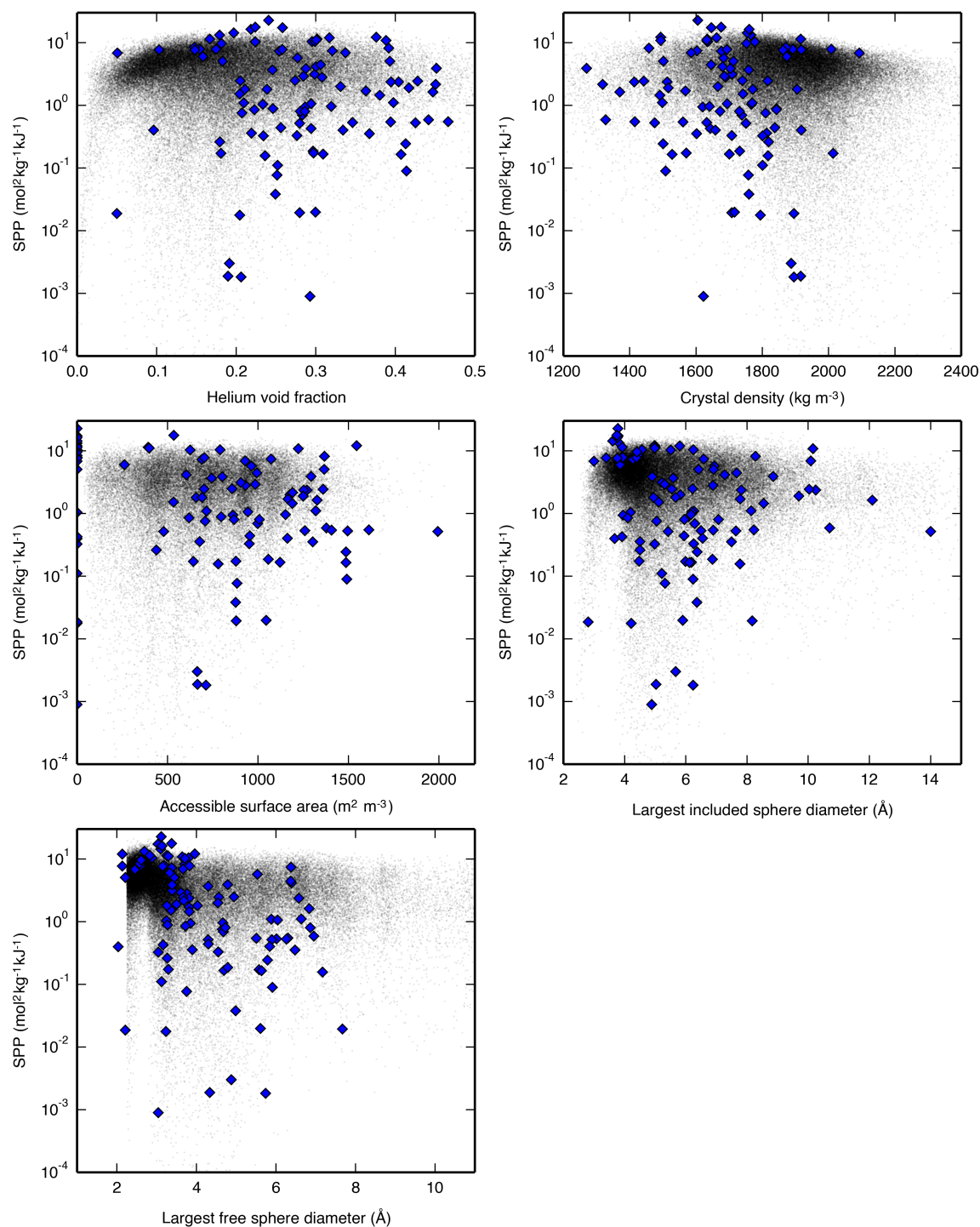
Figure A.20: SPP as a function of geometric parameters for the NAG process carried out with PSA. Hypothetical zeolites are shown as black dots, and IZA zeolites are shown as blue diamonds.
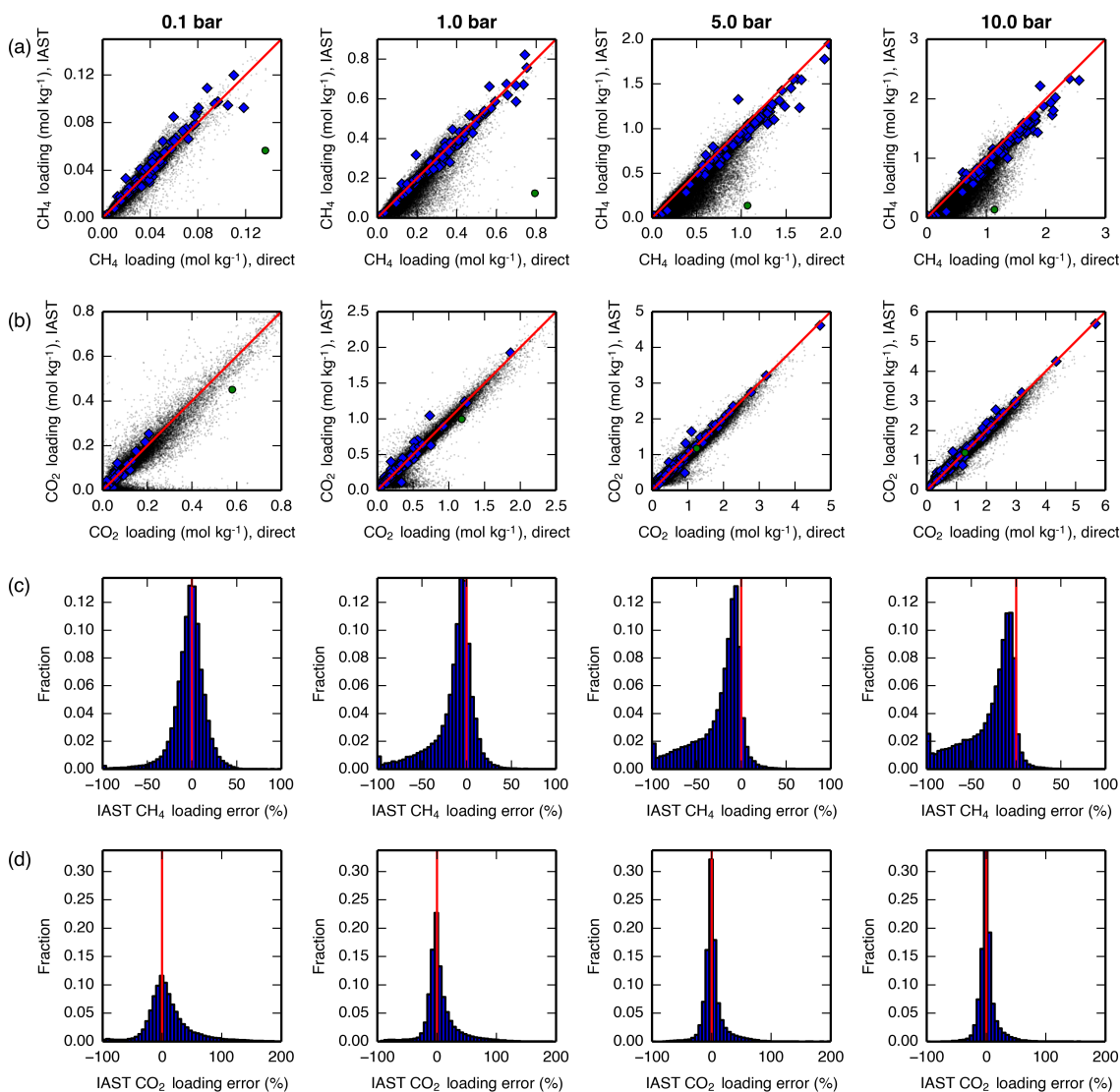
Figure A.21: Loadings of (a) $CH_4$ and (b) $CO_2$ at various total pressures, $300\,K$, and $90\,mol\,\%$ $CH_4$. Data on the x-axis are taken from directly-simulated mixture isotherms, while data on the y-axis are taken from applying IAST to pure-component isotherms. A line is drawn at $y = x$ for reference. Hypothetical zeolites are shown as black dots, IZA zeolites are shown as blue diamonds, and hypothetical zeolite PCOD8205017 is shown as a green circle. Histograms of the fractional IAST error $\left(\frac{\text{IAST loading} - \text{direct loading}}{\text{direct loading}} \times 100\%\right)$ of the hypothetical zeolites only for (c) $CH_4$ and (d) $CO_2$ at the same conditions. A line is drawn at $x = 0$ for reference. All plots in the left-column are at $0.1\,bar$, all plots in the middle-left-column are at $1\,bar$, all plots in the middle-right-column are at $5\,bar$, and all plots in the right-column are at $10\,bar$. A similar plot showing results at $60\,mol\,\%$ $CH_4$ is given in Fig. 4.7.
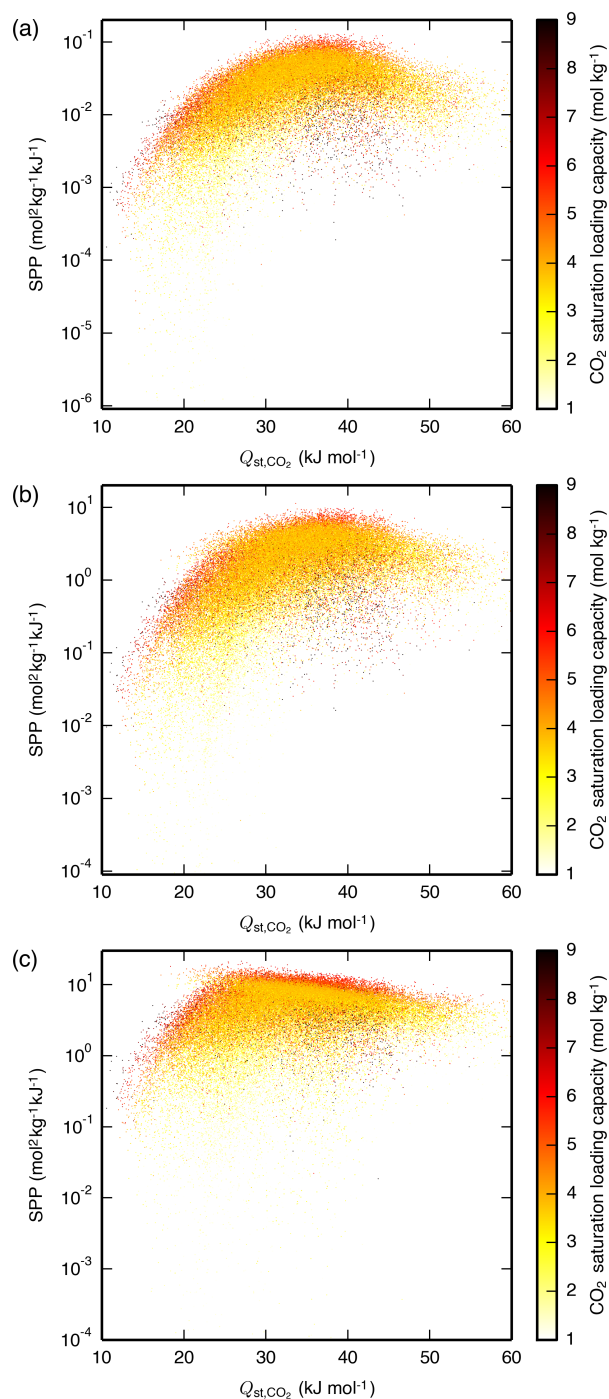
Figure A.22:  SPP of the hypothetical zeolites as a function of $Q_{st,CO_2}$ and the $CO_2$ saturation loading capacity for the (a) LFG, (b) APG, and (c) NAG processes, all carried out with PSA. The materials are plotted in random order such that the data shown are representative of the materials hidden due to having similar $Q_{st,CO_2}$ and SPP. Mixture isotherms were obtained using IAST, whereas Fig. 4.2 used directly-generated mixture isotherms.
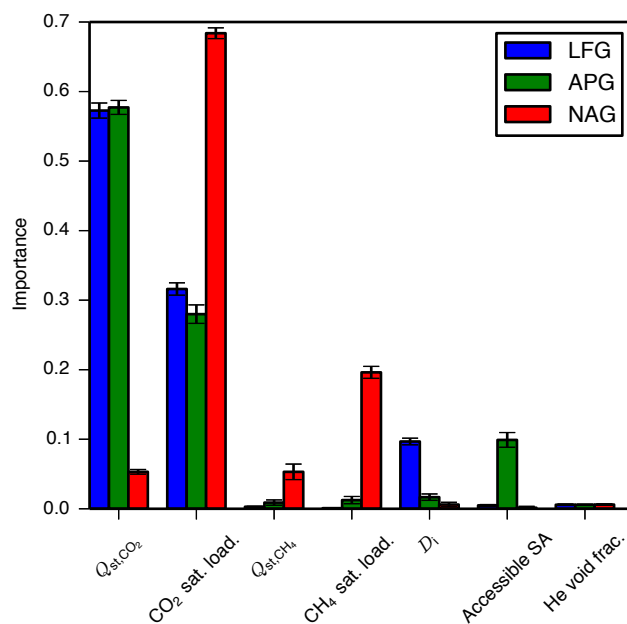
Figure A.23: Importance of geometric and isotherm descriptors to the SPP of the hypothetical zeolites undergoing PSA processes, as determined using a random forest of decision trees. The importance of a descriptor is calculated by summing the reductions in mean squared error brought about at each node where that descriptor splits a decision tree, averaging over all decision trees, and normalizing.[194] Here, sat. load. is the saturation loading capacity, $D_i$ is the largest included sphere diameter, Accessible SA is the accessible surface area, and He void frac. is the helium void fraction. Mixture isotherms were obtained using IAST, whereas Fig. 4.4 used directly-generated mixture isotherms.
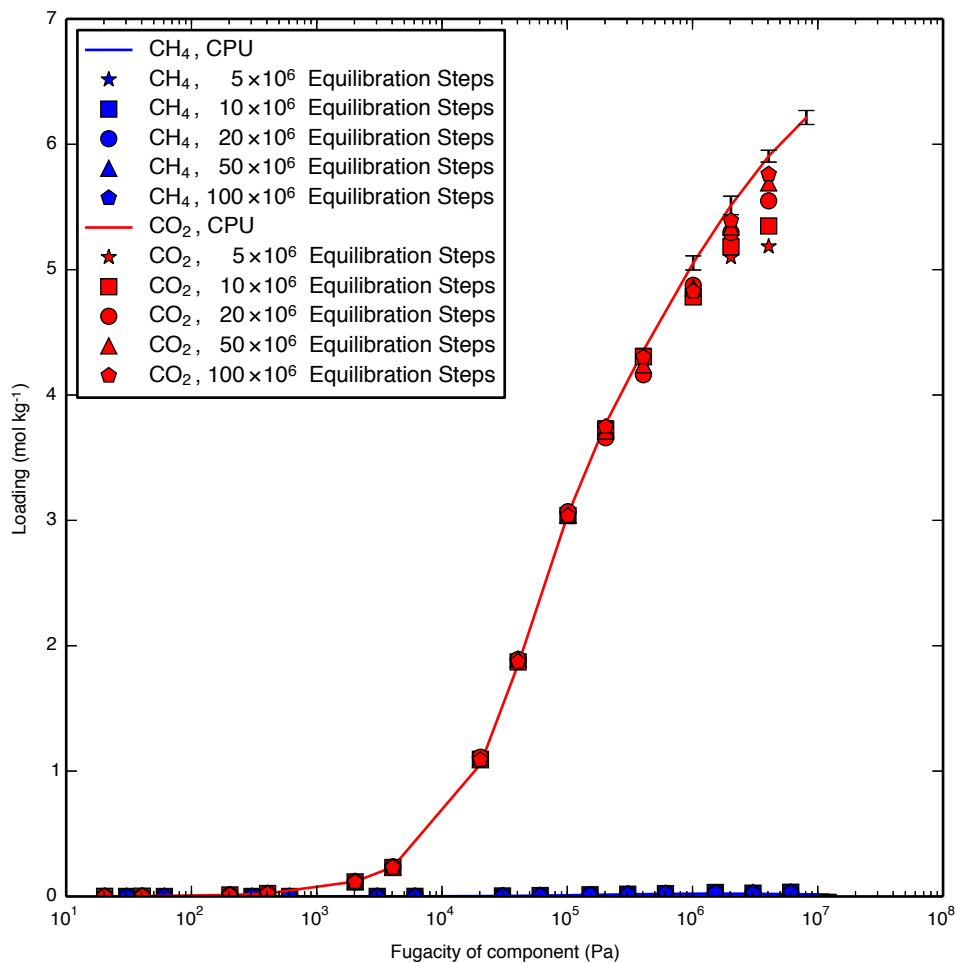
Figure A.24:   Mixture isotherms of $CH_4$ $(60\,\mathrm{mol\,\%})$ and $CO_2$ $(40\,\mathrm{mol\,\%})$ in the IZA zeolite ABW, used to verify the GPU code. Unlike the isotherms used in the screening, these used the ideal gas equation of state and no pocket blocking. The CPU data were computed using 200,000 equilibration cycles and 100,000 production cycles, with error bars presented. The remainder of the data were generated using the GPU code with varying numbers of equilibration steps and 1,000,000 production steps.
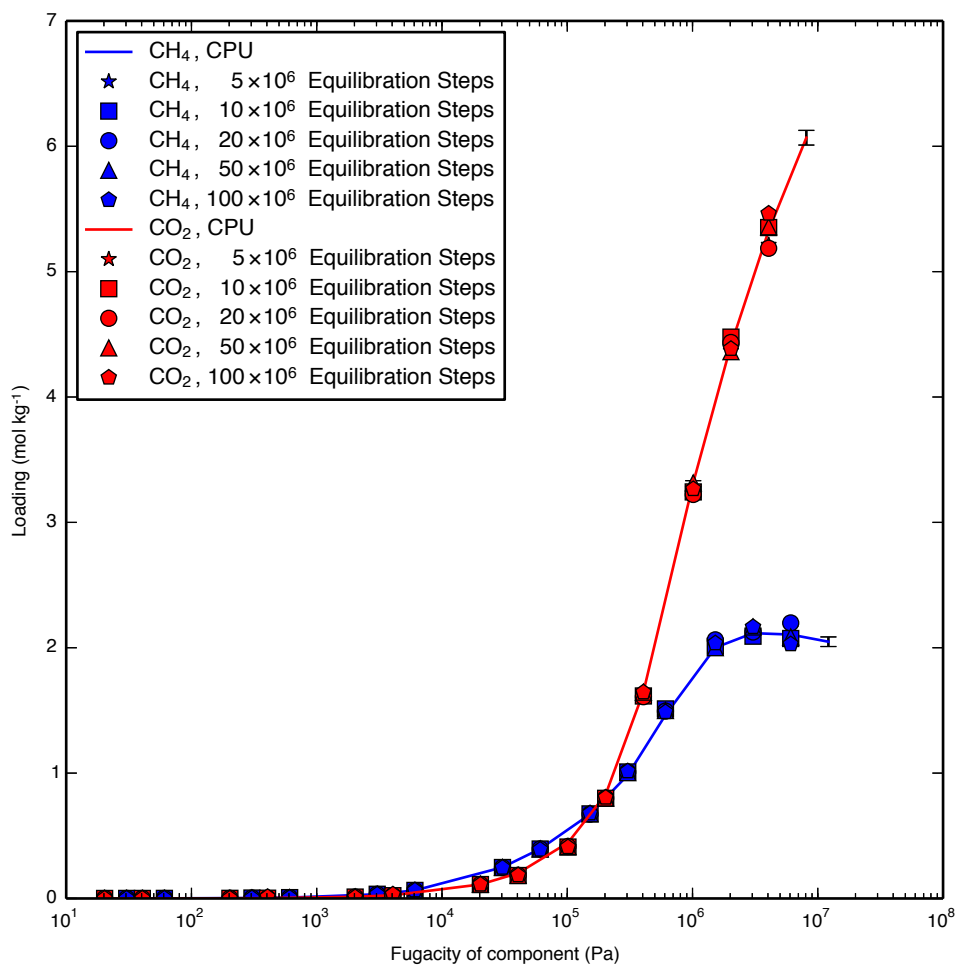
Figure A.25:   Mixture isotherms of $CH_4$ (60 mol %) and $CO_2$ (40 mol %) in the IZA zeolite FAU, used to verify the GPU code. Unlike the isotherms used in the screening, these used the ideal gas equation of state and no pocket blocking. The CPU data were computed using 200,000 equilibration cycles and 100,000 production cycles, with error bars presented. The remainder of the data were generated using the GPU code with varying numbers of equilibration steps and 1,000,000 production steps.
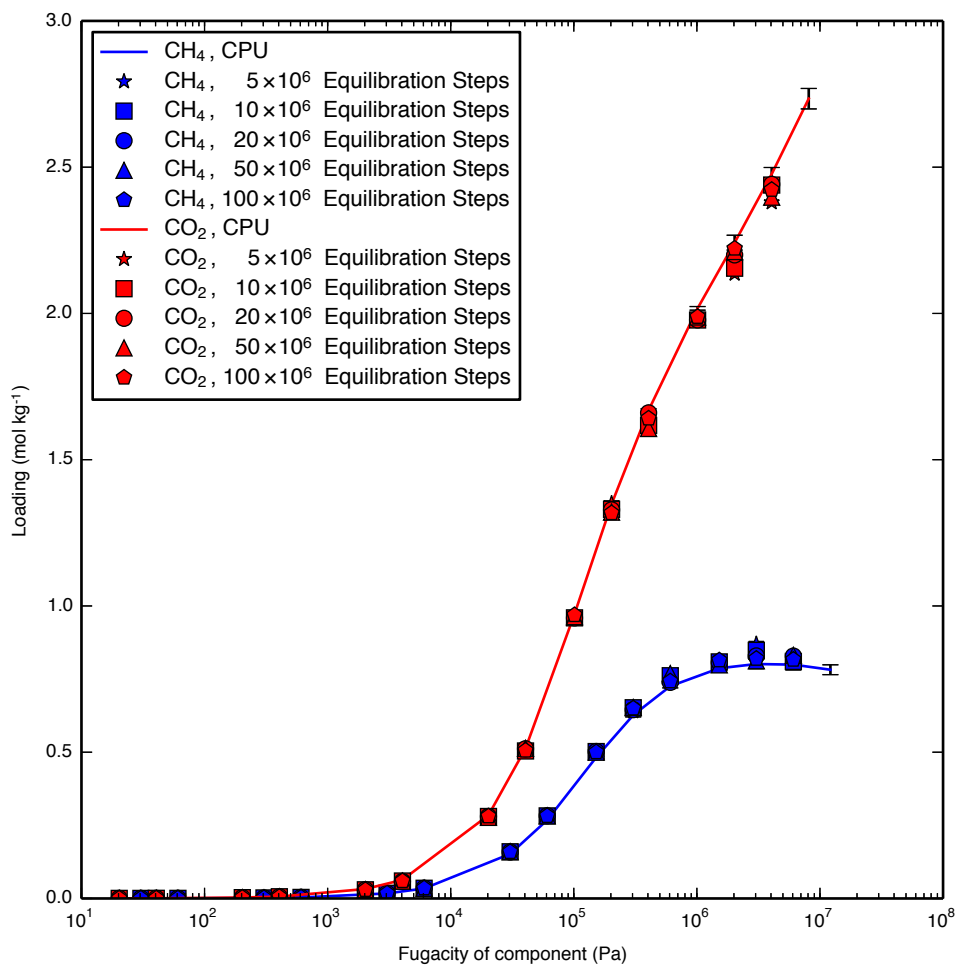
Figure A.26: Mixture isotherms of $CH_4$ ($60\,mol\%$) and $CO_2$ ($40\,mol\%$) in the IZA zeolite MFI, used to verify the GPU code. Unlike the isotherms used in the screening, these used the ideal gas equation of state and no pocket blocking. The CPU data were computed using 200,000 equilibration cycles and 100,000 production cycles, with error bars presented. The remainder of the data were generated using the GPU code with varying numbers of equilibration steps and 1,000,000 production steps.
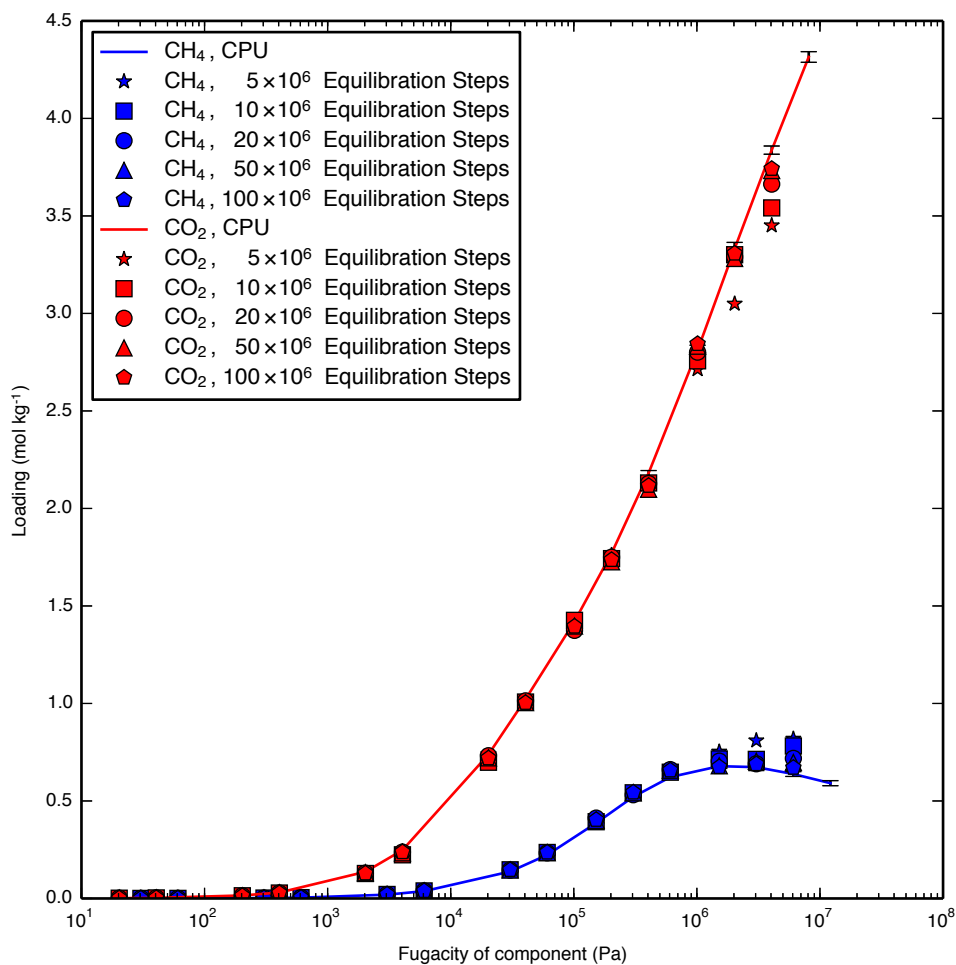
Figure A.27: Mixture isotherms of $CH_4$ (60 mol %) and $CO_2$ (40 mol %) in the IZA zeolite MOR, used to verify the GPU code. Unlike the isotherms used in the screening, these used the ideal gas equation of state and no pocket blocking. The CPU data were computed using 200,000 equilibration cycles and 100,000 production cycles, with error bars presented. The remainder of the data were generated using the GPU code with varying numbers of equilibration steps and 1,000,000 production steps.
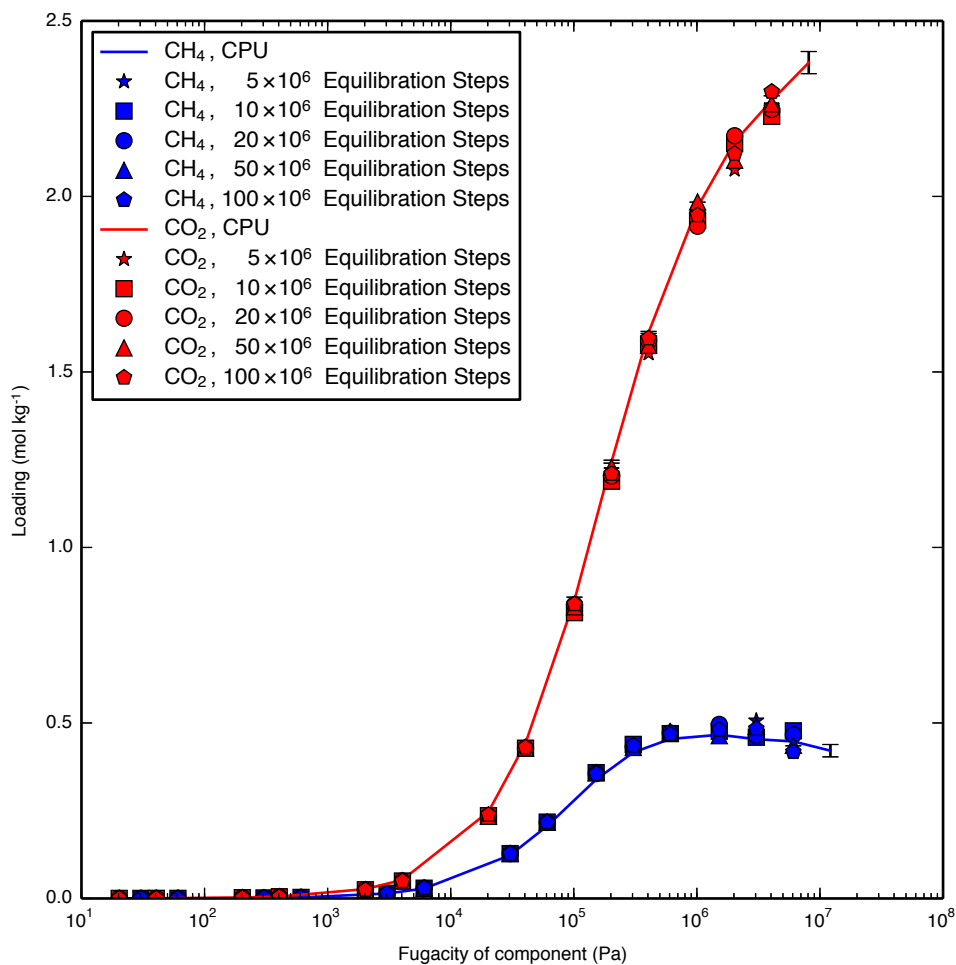
Figure A.28: Mixture isotherms of $CH_4$ (60 mol %) and $CO_2$ (40 mol %) in the IZA zeolite TON, used to verify the GPU code. Unlike the isotherms used in the screening, these used the ideal gas equation of state and no pocket blocking. The CPU data were computed using 200,000 equilibration cycles and 100,000 production cycles, with error bars presented. The remainder of the data were generated using the GPU code with varying numbers of equilibration steps and 1,000,000 production steps.
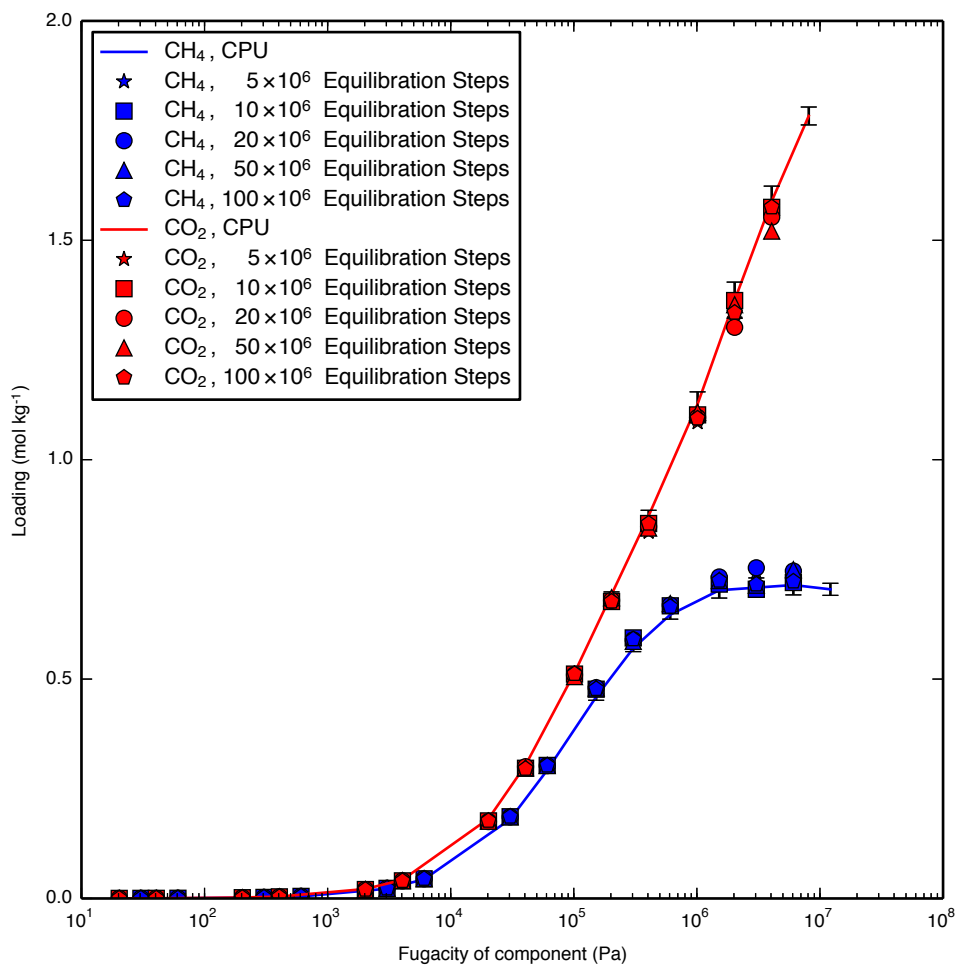
Figure A.29: Mixture isotherms of $CH_4$ ($60 \, mol \%$) and $CO_2$ ($40 \, mol \%$) in the hypothetical zeolite PCOD8170391, used to verify the GPU code. Unlike the isotherms used in the screening, these used the ideal gas equation of state and no pocket blocking. The CPU data were computed using 200,000 equilibration cycles and 100,000 production cycles, with error bars presented. The remainder of the data were generated using the GPU code with varying numbers of equilibration steps and 1,000,000 production steps.
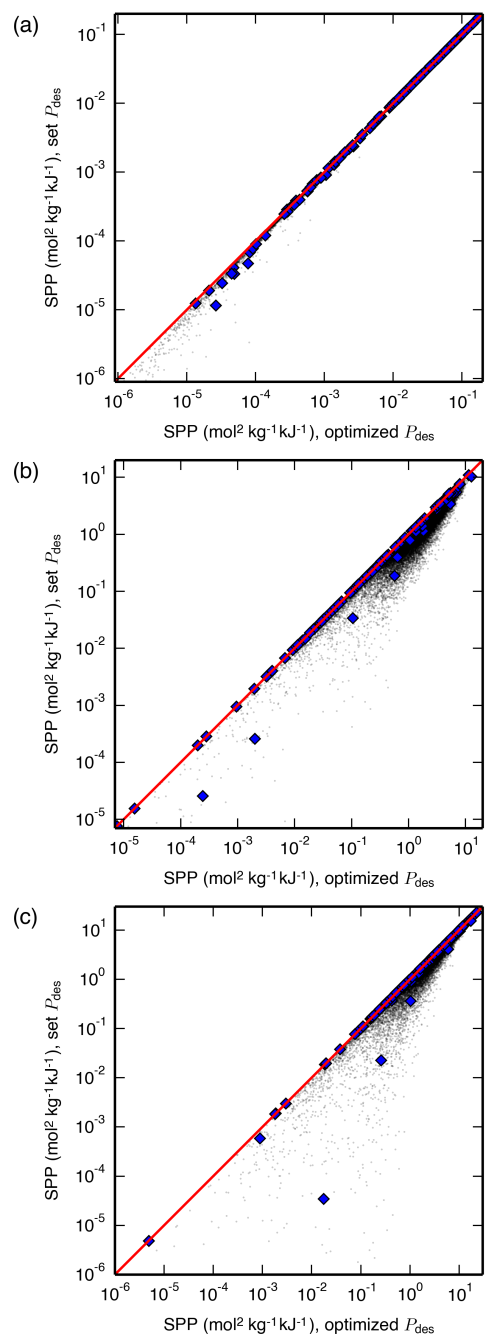
Figure A.30: A correlation between the SPP of materials evaluated at a set desorption pressure and the SPP evaluated at each material's optimal desorption pressure for the (a) LFG, (b) APG, and (c) NAG processes undergoing PSA. The LFG process used a set desorption pressure of 0.1 bar and the APG and NAG processes used a set desorption pressure of 1 bar, which might have been suspected as near-optimal desorption pressures for these processes prior to completing the present study.[165] Hypothetical zeolites are shown as black dots, and IZA zeolites are shown as blue diamonds. A line is drawn at $y = x$ for reference.
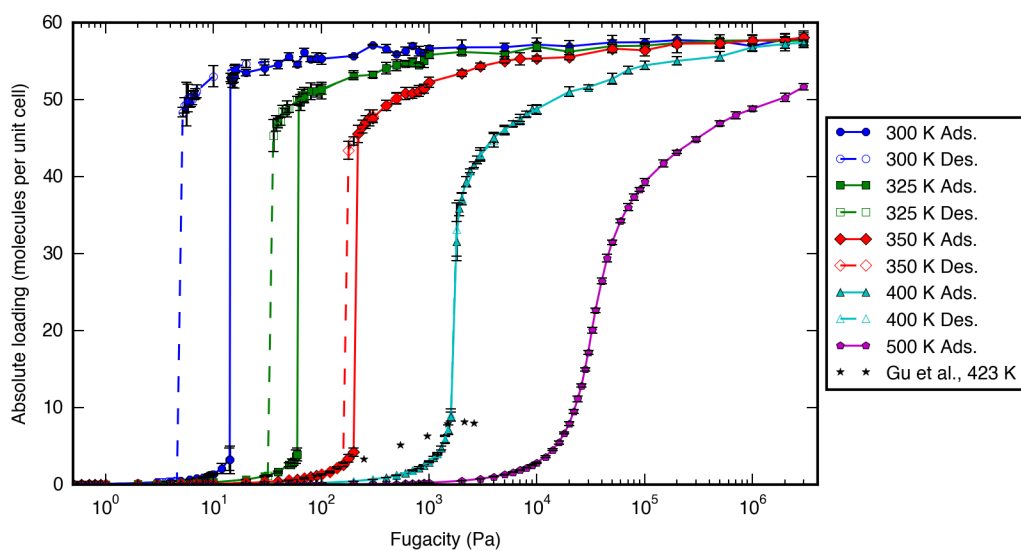
Figure A.31: Adsorption isotherms of p-xylene in IRMOF-1. Solid lines and dashed lines are guides to the eye showing adsorption and desorption branches, respectively. Black symbols are experimental data taken from the literature.[244] Although the experimental data shown are actually excess loadings as a function of pressure, the experimental conditions were such that excess loading is approximately equal to absolute loading and pressure is approximately equal to fugacity.
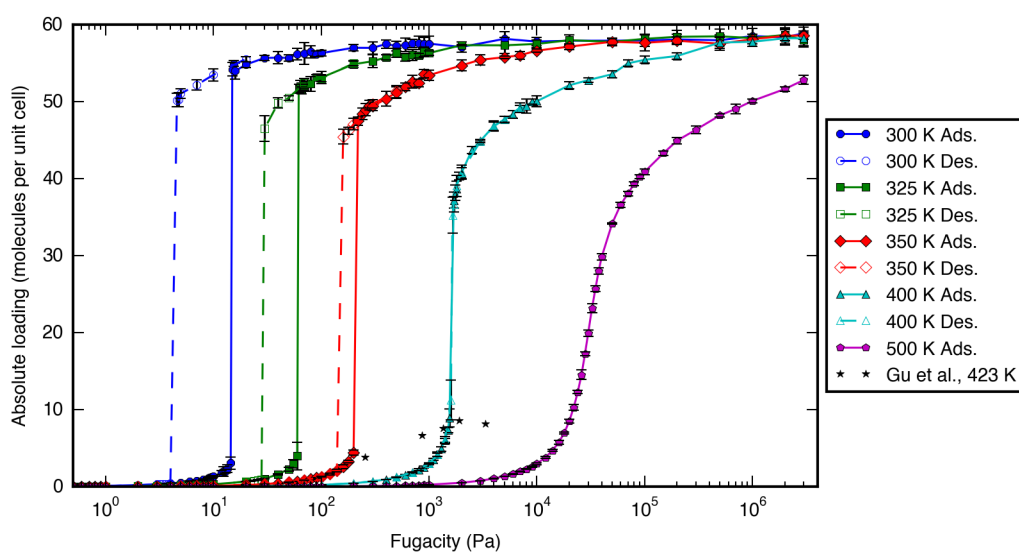
Figure A.32:  Adsorption isotherms of o-xylene in IRMOF-1. Solid lines and dashed lines are guides to the eye showing adsorption and desorption branches, respectively. Black symbols are experimental data taken from the literature.[244] Although the experimental data shown are actually excess loadings as a function of pressure, the experimental conditions were such that excess loading is approximately equal to absolute loading and pressure is approximately equal to fugacity.
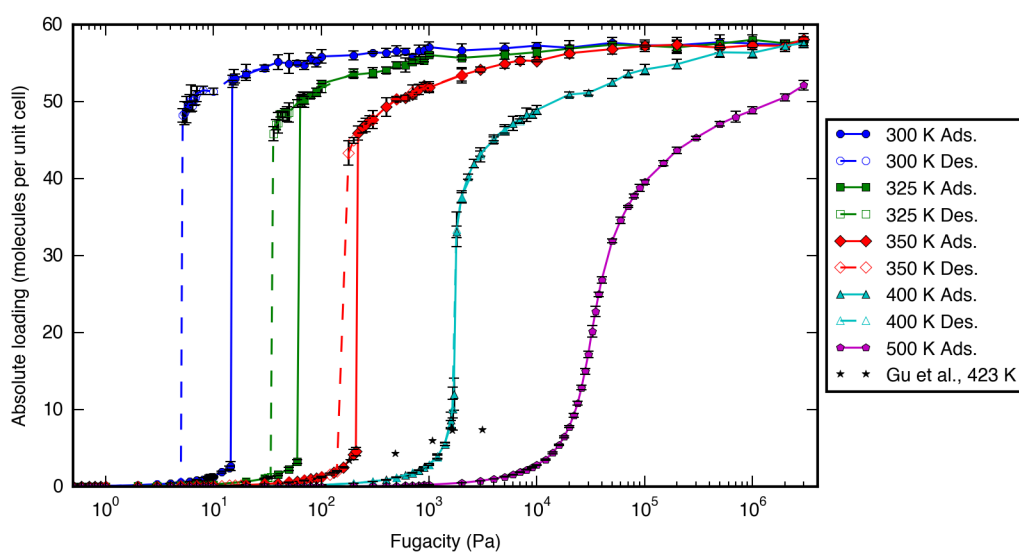
Figure A.33:  Adsorption isotherms of m-xylene in IRMOF-1. Solid lines and dashed lines are guides to the eye showing adsorption and desorption branches, respectively. Black symbols are experimental data taken from the literature.[244] Although the experimental data shown are actually excess loadings as a function of pressure, the experimental conditions were such that excess loading is approximately equal to absolute loading and pressure is approximately equal to fugacity.
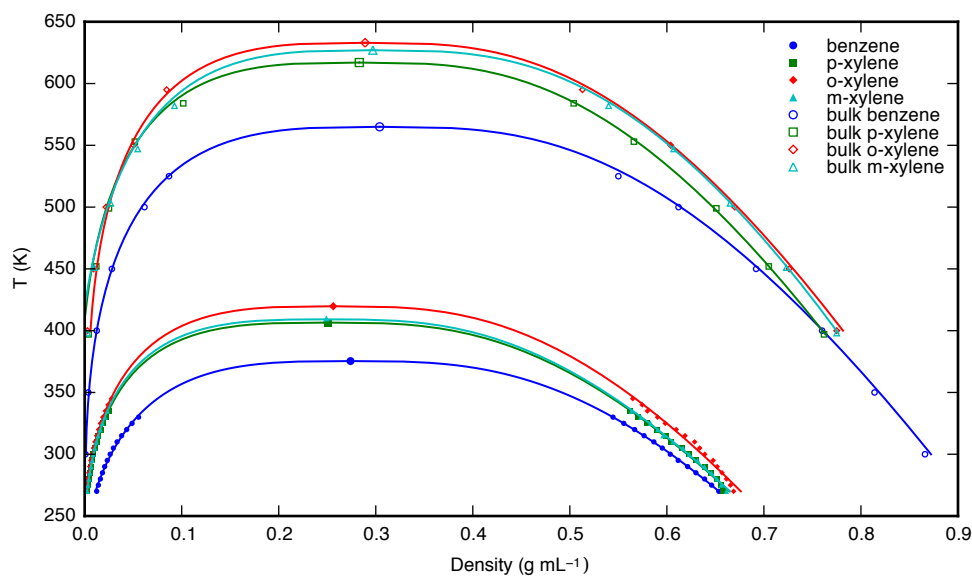
Figure A.34: Vapor-liquid coexistence curves for benzene, p-xylene, o-xylene, and m-xylene in IRMOF-1 and in bulk. The density scaling law (with the three-dimensional Ising critical exponent $\beta = 0.32$) and the law of rectilinear diameters were used to estimate the critical point and to interpolate the data.[4] Note that the error bars of these calculations are smaller than the symbols. The bulk phase diagrams are taken from simulations performed by Wick *et al.*[242] Densities are given in mass per volume void space using a void fraction of 0.832 for IRMOF-1.
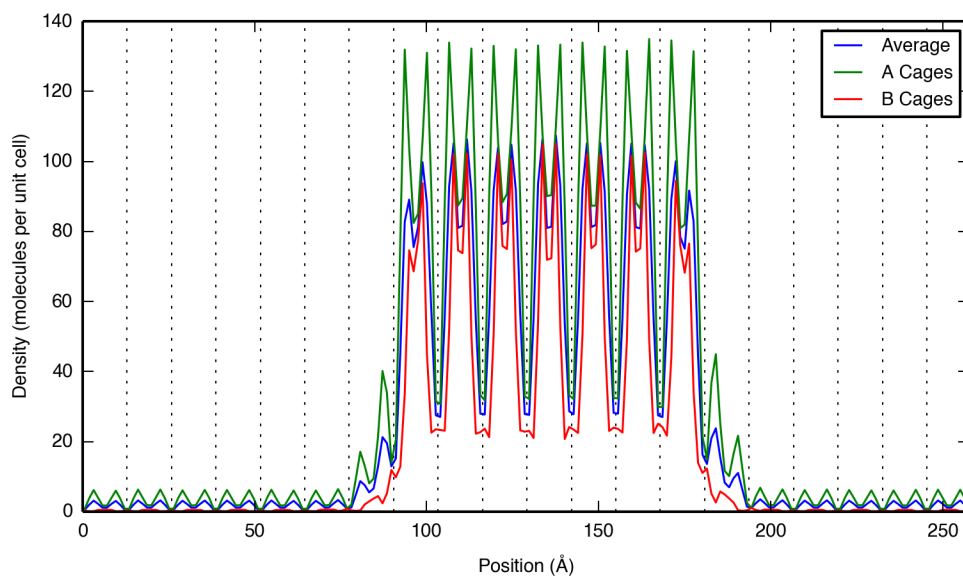
Figure A.35: Cross-sectionally averaged density profile of benzene in IRMOF-1 at 270 K, distributed by type of cage. IRMOF-1 has two different types of cages that make up the unit cell: A cages where the plane of the phenylene ring faces inside and B cages where the side of the phenylene ring faces inside. In the liquid phase, 63 % of benzene molecules reside in an A cage and 37 % reside in a B cage, and density maxima are near the corners of the cages. In the vapor phase these percentages are 88 % and 12 %, respectively, and density maximas are near the corners of the A cages but near the centers of the B cages. Our fractional loadings and positions of density maximas in the vapor phase agree with results from a previous computational study that examined the probability density of benzene in IRMOF-1 at a loading of 10 molecules per unit cell and a temperature of 300 K.[63]
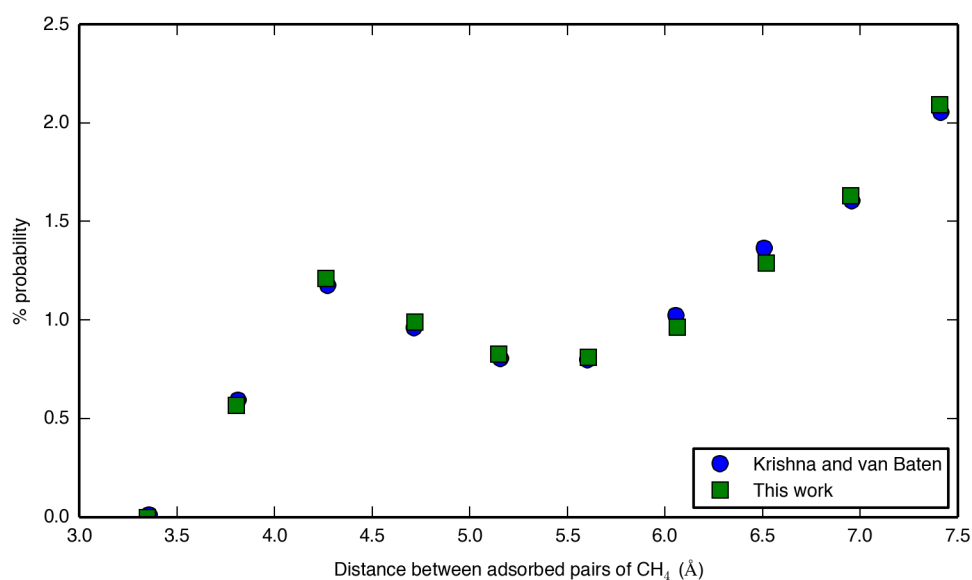
Figure A.36: Probability density function for finding pairs of methane molecules at a given separation in IRMOF-1 at 125 K at a loading of 40 molecules per unit cell, normalized such that integration over the range presented gave the same value as integration over the range of the data in Krishna and van Baten[227]. Note the excellent comparison between our results and those of Krishna and van Baten[227]
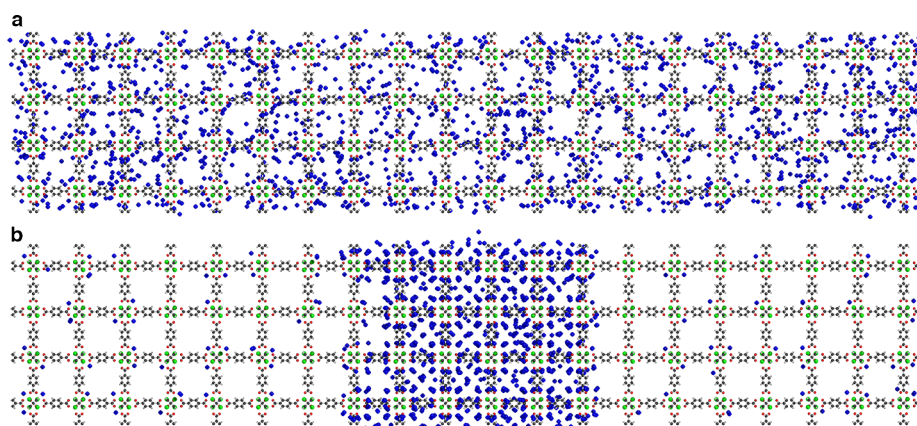
Figure A.37: Snapshots of methane in IRMOF-1 at (a) 125 K and (b) 50 K, both with a loading of 40 molecules per unit cell. Vapor-liquid phase coexistence was only observed at the lower temperature.

Table A.1: Energies of DFT-relaxed ZTCs, $D_{f,3p}$ of their parent zeolites, and descriptions of 1,000 K MD simulations results conducted on the non-relaxed ZTCs.

| zeolite | zeolite $D_{f,3p}$ (Å) | zeolite T atoms in unit cell | ZTC carbon atoms in unit cell | ZTC relative energy per carbon atom (eV/atom) | MD results[a] |
|---|---|---|---|---|---|
| diamond[b] | | 0 | 8 | 0.000 | |
| $C_{60}$[b] | | 0 | 60 | 0.340 | |
| -IRY | 8.61 | 76 | 384 | 0.232 | stable |
| FAU (model 1)[c] | 7.29 | 48 | 186 | 0.263 | stable |
| FAU (model 2)[c] | 7.29 | 48 | 164 | 0.315 | stable |
| FAU (model 3)[c] | 7.29 | 48 | 176 | 0.301 | stable |
| SAO | 6.73 | 56 | 210 | 0.516 | stable |
| EMT | 6.48 | 96 | 326 | 0.335 | stable |
| IWS (model 1)[d] | 6.32 | 136 | 444 | 0.562 | stable |
| IWS (model 2) | 6.32 | 136 | 562 | 0.401 | stable |
| RWY | 6.23 | 48 | 386 | 0.281 | stable |
| ISV | 6.22 | 64 | 208 | 0.548 | stable |
| -IFU | 6.05 | 128 | 610 | 0.378 | not stable |
| BEC | 6.03 | 32 | 106 | 0.566 | stable |
| IRR (model 1)[d] | 5.94 | 52 | 256 | 0.308 | stable |
| IRR (model 2) | 5.94 | 52 | 244 | 0.302 | stable |
| BEA | 5.89 | 64 | 240 | 0.402 | stable |
| BEB | 5.73 | 64 | 218 | 0.491 | stable |
| SBS | 5.65 | 96 | 346 | 0.335 | stable |
| SBT (model 1)[d] | 5.65 | 144 | 530 | 0.349 | stable |
| SBT (model 2) | 5.65 | 144 | 556 | 0.328 | stable |
| POS | 5.62 | 64 | 250 | 0.495 | stable |
| ITT | 5.16 | 46 | 232 | 0.384 | stable |

[a] We classify these results as "stable" if during the relaxation and NVT parts of the simulation no atom moved more than 3.0 Å from its starting position and the average squared displacement of all atoms from their starting positions was less than 1.5 Å, and if during the NPT parts of the simulation none of the six lattice constants changed by more than 10 %. We visually observed that if one of these conditions was violated, some degree of atomic rearrangement was occurring, so we classify these cases as "not stable."

[b] Given as a reference.

[c] The multiple models of this ZTC were plotted together as a single point in Fig. 3 using the average energy.

[d] This ZTC model contains defects and is not shown in Fig. 3.

[e] This zeolite is from Deem's SLC hypothetical zeolite database.[138] The ZTC templated by this zeolite has its associated TPMS given in parentheses following the zeolite's database entry number.

[f] This zeolite is from Treacy and Foster's silver hypothetical zeolite database.[137] The ZTC templated by this zeolite has its associated TPMS given in parentheses following the zeolite's database entry number.

Table A.1: (continued)

| zeolite | zeolite $D_{f,3p}$ (Å) | zeolite T atoms in unit cell | ZTC carbon atoms in unit cell | ZTC relative energy per carbon atom (eV/atom) | MD results[a] |
|---|---|---|---|---|---|
| MEL | 5.13 | 96 | 222 | 0.720 | not stable |
| DFO | 5.02 | 132 | 592 | 0.443 | stable |
| JSR (model 1)[c] | 4.99 | 96 | 608 | 0.409 | stable |
| JSR (model 2)[c] | 4.99 | 96 | 508 | 0.600 | not stable |
| JSR (model 3)[c] | 4.99 | 96 | 624 | 0.413 | stable |
| UOV | 4.97 | 176 | 584 | 0.615 | not stable |
| ITG | 4.94 | 56 | 200 | 0.522 | stable |
| MSE | 4.94 | 112 | 376 | 0.456 | stable |
| TUN (model 1)[c] | 4.94 | 192 | 480 | 0.633 | not stable |
| TUN (model 2)[c] | 4.94 | 192 | 458 | 0.641 | not stable |
| TUN (model 3)[c] | 4.94 | 192 | 508 | 0.624 | not stable |
| UWY | 4.90 | 60 | 184 | 0.652 | stable |
| BOG | 4.86 | 96 | 246 | 0.579 | stable |
| IWW | 4.85 | 112 | 278 | 0.714 | not stable |
| CON | 4.83 | 56 | 160 | 0.521 | stable |
| IWR | 4.80 | 56 | 180 | 0.467 | stable |
| -SVR (model 1)[c] | 4.59 | 92 | 242 | 0.752 | stable |
| -SVR (model 2)[c] | 4.59 | 92 | 198 | 1.069 | not stable |
| -ITN | 4.48 | 54 | 174 | 0.576 | stable |
| MFI | 4.40 | 96 | 210 | 0.793 | stable |
| -EWT | 4.23 | 70 | 250 | 0.588 | not stable |
| LTA | 4.14 | 24 | 76 | 0.470 | stable |
| SOF | 4.11 | 40 | 140 | 0.955 | not stable |
| AFY | 4.02 | 16 | 84 | 0.618 | stable |
| RHO | 4.00 | 48 | 118 | 0.595 | not stable |
| BSV | 3.78 | 96 | 236 | 1.045 | not stable |
| SAV | 3.69 | 48 | 192 | 0.621 | stable |
| AFX | 3.67 | 48 | 136 | 0.747 | not stable |
| CHA | 3.66 | 36 | 108 | 0.717 | not stable |
| AEI | 3.58 | 48 | 172 | 0.668 | stable |
| ITR | 3.52 | 112 | 226 | 0.958 | not stable |
| ITH | 3.47 | 56 | 146 | 0.781 | not stable |
| OFF | 3.37 | 18 | 66 | 0.584 | stable |
| ERI | 3.36 | 36 | 124 | 0.659 | stable |
| OSO | 3.35 | 9 | 64 | 0.711 | not stable |
| GME | 3.35 | 24 | 94 | 0.735 | not stable |
| PUN | 3.35 | 36 | 166 | 0.826 | stable |
| STW | 3.32 | 60 | 204 | 0.833 | stable |
| SZR | 3.26 | 36 | 100 | 0.737 | not stable |
| 8326829 (G-W)[e] | 6.24 | 66 | 280 | 0.325 | |
| 8326836 (H)[e] | 6.06 | 84 | 306 | 0.298 | |
| 8326837 (H)[e] | 6.26 | 44 | 164 | 0.308 | |
| 8326849 (H)[e] | 6.25 | 66 | 182 | 0.334 | |
| 8326896 (H)[e] | 6.25 | 96 | 304 | 0.207 | |
| 8327291 (H)[e] | 6.47 | 60 | 218 | 0.254 | |
| 8331018 (G)[e] | 5.82 | 168 | 320 | 0.431 | |
| 221_2_6 (P)[f] | 6.63 | 72 | 224 | 0.275 | |

# Appendix B

# Acknowledgements of each study