UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title
Computation assisted discovery of nanoporous materials for gas storage and separations

Permalink
https://escholarship.org/uc/item/8083b8c6

Author
Simon, Cory

Publication Date
2016

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Computation assisted discovery of nanoporous materials for gas storage and separations

by

Cory M. Simon

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 Jeffrey R. Long
Professor Carlo Carraro

Fall 2016

Computation assisted discovery of nanoporous materials for gas storage and separations

Copyright 2016
by
Cory M. Simon

Abstract

Computation assisted discovery of nanoporous materials for gas storage and separations

by

Cory M. Simon

Doctor of Philosophy in Chemical Engineering

University of California, Berkeley

Professor Berend Smit, Chair

Nanoporous materials, such as metal-organic frameworks (MOFs), have enormous internal surface areas. Their consequent adsorption properties demonstrate promise towards solving energy-related problems in gas storage and gas separations. Owing to their modular and versatile chemistry, millions of possible nanoporous materials can be synthesized. This vast chemical space allows a material to be tailor-made or fine-tuned to target specific adsorbate molecules and conditions. In this thesis, we utilize molecular models and simulations of gas adsorption in both existing and predicted nanoporous material structures to accelerate the discovery of new materials targeted for gas storage and separations at specific conditions.

In the first part of this work, we approach the problem of identifying an optimal porous material to densify natural gas for storage onboard vehicles as fuel. We developed a series of statistical mechanical models to find the thermodynamic parameters that optimize the deliverable capacity of a material. We conclude that the heat of adsorption, which is a commonly used metric to evaluate materials for natural gas storage, is a misleading metric because the optimal heat of adsorption depends on the pore size. Our models also reveal that adsorbate-adsorbate attractions– in the case where multiple methane molecules can fit into a pore– can enhance the deliverable capacity.

Next, we carried out a high-throughput computational screening of metal-organic frameworks, porous polymer networks, zeolites, and zeolitic imidazolate frameworks for natural gas storage. The data that we collected provide candidate structures for synthesis, reveal relationships between structural characteristics and performance, and suggest that it may be difficult to reach the current Advanced Research Project Agency-Energy (ARPA-E) deliverable capacity target. To assess thermodynamic limits to the methane deliverable capacity, we then built a model of an extreme scenario where an energy field can be created without taking up space with material. This model suggests that, while the failure to reach the ARPA-E storage target is due to material design constraints rather than purely thermodynamic constraints, the ARPA-E storage target is ambitiously close to the thermodynamic limit.

In the second part of this work, we approach the problem of identifying a material that selectively adsorbs xenon over krypton. With over half a million nanoporous material structures to consider as candidate adsorbents, the computational cost of a brute-force computational screening strategy was prohibitive. Instead, we employed a machine learning algorithm, a random forest, to learn the relationship between quickly computed structural descriptors and Xe/Kr selectivity, which is more expensive to compute. The trained random forest allowed us to rule out a large percentage of the materials on the basis of quickly-computed structural descriptors. Our machine learning accelerated screening pinpoints top candidates on which to focus experimental efforts and elucidates structure-property relationships for design guidelines for a Xe-selective material.

As we are now working with mixed gas adsorption, we developed a user-friendly software package in Python, pyIAST, for ideal adsorbed solution theory (IAST) calculations. IAST is a thermodynamic framework to predict mixed gas adsorption from pure-component adsorption isotherms, which are easier to measure. We provide practical guidelines for applying IAST.

Finally, we carry out a high-throughput computational screening of metal-organic frameworks for capturing Xe from air at dilute conditions, a separation encountered in used nuclear fuel reprocessing. Our computational screening, facilitated by a parallelized code on GPUs, predicted a metal-organic framework, SBMOF-1, to be among the most Xe-selective. Our experimental collaborators synthesized and tested SBMOF-1 and found it to exhibit the highest Xe/Kr selectivity and Xe Henry coefficient reported in the literature. Column-breakthrough experiments reveal that SBMOF-1 is a near-term material for capturing xenon from the off-gases of used nuclear fuel reprocessing plants. This is a rare case of a computation-assisted materials discovery.

Contents

Contents **i**

1 Introduction **1**
1.1 Applications of nanoporous materials . . . 3
1.2 Molecular Simulation Methods . . . 10

2 Optimizing nanoporous materials for gas storage **16**
2.1 Introduction . . . 16
2.2 Zeolite screening data . . . 18
2.3 Models for Adsorption: Seeking an optimal heat of adsorption . . . 22
2.4 Analysis of top performing zeolite structures . . . 43
2.5 Discussion . . . 46

3 High-throughput screening for methane storage **49**
3.1 Introduction . . . 50
3.2 Metric for methane storage . . . 50
3.3 The Nanoporous Materials Genome . . . 51
3.4 Screening the materials genome . . . 54
3.5 Results and discussion . . . 56
3.6 Conclusions . . . 59

4 Thermodynamic limits to deliverable capacity **73**
4.1 Introduction . . . 73
4.2 Scenario 1: Lattice model . . . 74
4.3 Scenario 2: Lennard-Jones fluid in a uniform energy field . . . 76
4.4 Methods . . . 77

5 Machine learning-accelerated computational screening **83**
5.1 Introduction . . . 84
5.2 Methods . . . 87
5.3 Results . . . 96
5.4 Conclusions . . . 120

6 Software for Ideal Adsorbed Solution Theory **124**
6.1 Introduction 124
6.2 Ideal Adsorbed Solution Theory (IAST) 126
6.3 Characterizing the pure-component adsorption isotherm $n_i^\circ(P)$ 129
6.4 IAST Package Case Study 138
6.5 Limitations of IAST 148
6.6 Conclusions 150

7 Discovery of MOF for xenon capture **152**
7.1 Introduction 152
7.2 High-throughput screening 154
7.3 Synthesis and equilibrium adsorption measurements 156
7.4 Adsorption kinetics and column breakthrough experiments 158
7.5 Revealing the Xe adsorption sites in SBMOF-1 160
7.6 Discussion 163

8 Conclusions **164**
8.1 Outlook 166

Bibliography **167**

Acknowledgments

My time spent during the PhD has been a period of intellectual growth. This is evident from comparing my first paper to my last one while assembling this thesis. I have several people to thank for this period of learning.

First and foremost, thanks to my advisor, Berend Smit, for guidance and support and for giving me the freedom to explore my ideas and interests. Thanks to Maciej Haranczyk for advising me on several projects during my time at Lawrence Berkeley National Laboratory. Finally, thanks to Carlo Carraro for working with and advising me on a statistical mechanical model of porous crystals with rotating ligands during the past year.

Through osmosis, I've learned much from collaborating with many talented researchers. And, these collaborations were essential to many of the scientific results obtained during my PhD; for this reason, I write using "*we*" in this thesis instead of "*I*". During the beginning of my PhD, I learned a lot from working with Richard Luis Martin and Li-Chiang Lin. And thank you Jihan Kim for collaborating with me overseas. I'd also like to thank Profs. Randall Snurr, Michael W Deem, Hong-Cai Zhou, and David Sholl for collaborations in methane storage, as well as members of their groups, particularly Diego Gomez Gualdron, Yongchul (Greg) Chung, Dawei Feng, and Yi Bao. For collaborations in Xe/Kr separations, I'd like to thank Sondre Schnell and Rocio Mercado. Finally, it has been a pleasure collaborating with Praveen Thallapally at Pacific Northwest National Lab, along with Debasis Banerjee and Sameh Elsaidi. It has also been a pleasure working with Efrem Braun.

Generally, it has been great to be lab-mates with Bess, Daniele, Efrem, Forrest, Greg, Johanna, Josh, Katie, Kyuho, Li-Chiang, Matt, Michelle, Rocio, Shachi, Sondre, and Sudi.

Thanks to my family for support– particularly my mom– and my dad, brother, grandma, and Ed.

Chapter 1

Introduction

> I can't see exactly what would happen, but I can hardly doubt that when we have some control of the arrangement of things on a small scale, we will get an enormously greater range of possible properties that substances can have and of different things that we can do.
>
> Richard Feynman, American Physical Society Meeting, 1959

Nanoporous materials are solid-state materials with nano-sized pores. Several classes of nanoporous materials have been synthesized, including metal-organic frameworks (MOFs) [1], porous polymer networks (PPNs) [2], covalent organic frameworks (COFs) [3], zeolitic imidazolate frameworks (ZIFs) [4], and porous organic cages [5]. For example, metal-organic frameworks (MOFs) are porous, crystalline materials [1] with record-breaking surface areas [6]. The consequent adsorption properties of MOFs have demonstrated promise to help solve paramount energy problems [7] in gas storage [8], e.g., storing natural gas [9] or hydrogen [10] on board passenger vehicles, and gas separations [11], e.g., capturing carbon dioxide from the flue gas of coal-fired power plants [12] or radioactive noble gases from used nuclear fuel reprocessing facilities [13]. In addition, MOFs hold promise for applications in chemical sensing [14], catalysis [15], drug delivery [16, 17], ion exchange [18], water desalination [19], electronics [20], and electro-optics [21].

The unprecedented chemical tunability in the synthesis of MOFs, whereby metal nodes or clusters and organic moieties self-assemble to yield predetermined, ordered structures [22], has enabled chemists to synthesize tens of thousands of MOFs to date [23]. To illustrate the tunability of MOFs, consider pillared square grid MOFs, which have shown unprecedented selectivities for several energy-intensive separations [24, 25, 26, 27]. Pillared square grid MOFs are assembled from metals, organic linker molecules, and pillars. The linkers and metals self-assemble to form two-dimensional square grids, and pillars connect the square grids to construct a three-dimensional, porous framework. Several pillared square grid MOFs have been synthesized by varying the metal, linker, and pillar. See Fig 1.1. First, different metal nodes, Ni^{2+} [28], Zn^{2+} [29], Cu^{2+} [27], and Co^{2+} [28], have been utilized and shown

to affect the affinity for CO_2 [27]. Second, different linkers, all of which possess electron-rich nitrogen atoms at two ends to coordinate to the metal, have been utilized to more significantly modulate the size of the pores [30]. Finally, two different pillars, $(SiF_6)^{2-}$ [28] and $(NbOF_5)^{2-}$[26], have been utilized. Numerous MOF topologies and coordination chemistries have been reported [31, 23] beyond the pillared square grid architecture. In summary, owing to their modular and versatile chemistry, the chemical space [32] of MOFs is vast.

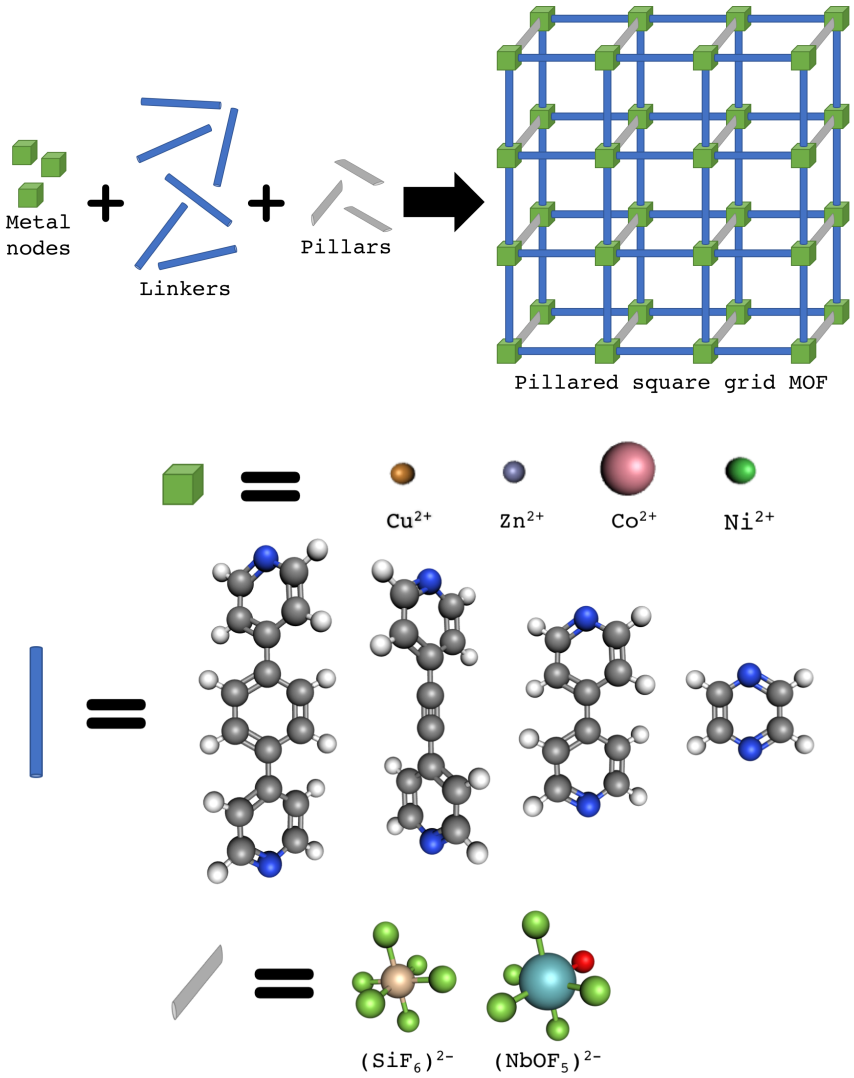


Figure 1.1: Illustrating the tunability of MOFs with the pillared square grid architecture. The molecular building blocks in a pillared square grid MOF are metal nodes, organic linker molecules, and pillars. A variety of pillared square grid MOFs have been synthesized using combinations of the depicted metals, linkers, and pillars.

As an example metal-organic framework, HKUST-1 [33], also known as Cu_3BTC_2 (BTC = benzene-1,3,5-tricarboxylate), is given in Fig 1.2. Shown is a visualization of the crystal structure and how the MOF appears in reality. The coordination chemistry of Cu_3BTC_2 is more complicated than in the pillared square grid architecture in Fig 1.1; Cu_3BTC_2 consists of interconnected dicopper(II) tetraBTC units (see Fig 1.2c) [33]. The secondary building unit of HKUST-1 is an octahedral unit with copper dimers at its six vertices; the assembly of these octahedral secondary building units engenders HKUST-1 with its cubic symmetry [33].

Excitingly, as we can predict the structure that assembles on the basis of the starting molecular building blocks, we have *control* over the structure of modular nanoporous materials. Recent advances in the postsynthetic modification [35, 36] afford another degree of tunability and control. This allows us to fine-tune existing MOF architectures– or tailor-make a completely *new* MOF– to exhibit optimal properties for a given application. For example, we can tune the MOF chemistry to target the adsorption of specific gas molecules from a mixture.

Our understanding of the intricate structure-property relationships for the rational design and tuning of nanoporous materials for a given application, however, is limited; trial and error is heavily involved in the laboratory today. As it can take months to find how to synthesize and activate a single material, then characterize and test it, only a small subset of possible materials can be investigated due to time and resource constraints. In line with the Materials Genome Initiative [37], we use molecular models and simulations to rank nanoporous material structures for gas storage and separations. By conducting high-throughput computational screenings of databases of nanoporous material structures, we can predict performance rankings, pinpointing the top predicted performers on which to focus experimental efforts. Further, the large volume of data from a high-throughput screening elucidates useful structure-property relationships for the rational design of new materials tailored towards a given application. Finally, by considering so many nanoporous material structures, we can obtain performance expectations.

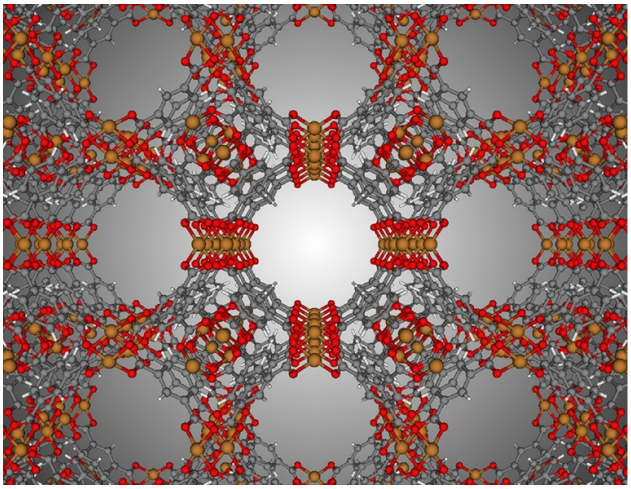
1.1 Applications of nanoporous materials

Here, I outline in detail applications of nanoporous materials on which I focused during my doctoral research, namely (i) natural gas storage, (ii) Xe/Kr separations, and (iii) Xe capture from air at dilute conditions.

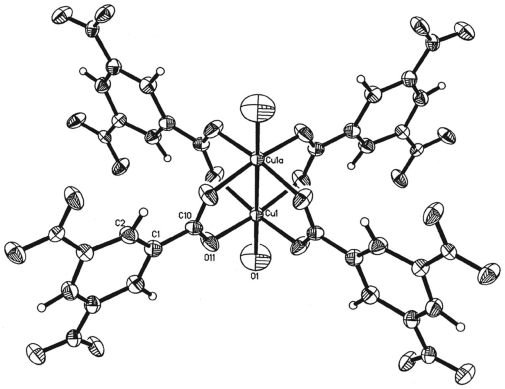
Natural gas storage

Natural gas is often called a 'bridging fuel' from our incumbent petroleum economy to one predominantly based on renewable energy sources. The transportation sector, currently fueled predominantly by gasoline, is an enormous potential market for natural gas. Mostly methane, natural gas emits less CO_2 per energy harvested upon combustion than any other

(a)



(b)



(c)

Figure 1.2: HKUST-1 [33], a metal-organic framework (MOF), as an example nanoporous material. (a) A photo of HKUST-1 in the laboratory. The color of HKUST-1 is blue due to the copper content. The dry erase marker is for scale. Figure from Ref [34], used under Creative Commons Attribution License. (b) A visualization of the crystal structure of HKUST-1, which has chemical formula Cu_3BTC_2, where BTC = benzene-1,3,5-tricarboxylate. (orange: copper, red: oxygen, gray: carbon, white: hydrogen). The walls of the hole-like pores in this material form an interior surface to which gas molecules adsorb. The diameter of these pores is 0.7 nm. Figure from Ref [34], used under Creative Commons Attribution License. (c) Dicopper(II) tetracarboxylate building block for HKUST-1. From Ref. [33]. Reprinted with permission from AAAS.

hydrocarbon [38] (25% less CO_2 than gasoline [39]). Because transportation accounts for 28% of US greenhouse gas emissions [40], the widespread adoption of natural gas as a transportation fuel has the potential to significantly mitigate anthropogenic climate change. Furthermore, natural gas as a transportation fuel emits less volatile organic compounds (VOCs), CO, SO_x, and particulates than gasoline [41]. The reduction of greenhouse gases by adopting natural gas as a fuel is predicated on mitigating releases of methane that may come with increased production, transportation, and use [42, 43]; methane itself is a greenhouse gas 21 times more potent than CO_2 [40, 44], considering its 100-year global warming potential (GWP).

Economic incentives for replacing petroleum energy sources with natural gas are in place. Dry shale gas production in the US has skyrocketed in recent years [47] (Fig 1.3a) due to hydraulic fracturing and horizontal drilling technologies that enable the extraction of shale gas that was once inaccessible [48, 49]. Estimated world-wide natural gas reserves have increased substantially [39]. This has driven natural gas prices to historic lows (Fig 1.3b) [50], rendering methane economically competitive with petroleum fuels. The US in particular has a national security incentive to decrease petroleum imports from unstable regions and instead utilize domestic energy reserves. Further, the barrier for manufacturers to tailor vehicular engines to consume natural gas is low, and gasoline engines can be retrofitted to run on natural gas [51]. In the US, tax incentives for this are in place [52]. Still, the hidden environmental costs of hydraulic fracturing [53], such as drinking water contamination [54], are not yet well understood.

Despite the positive environmental and economical outlook, the comparatively low volumetric energy density of methane at ambient conditions is a major obstacle to adopting natural gas as a vehicular fuel instead of gasoline. On a mass basis, the heat of combustion of natural gas is actually slightly higher than that of gasoline. However, the fact that methane is a gas at ambient conditions, leaving it with a volumetric energy density 1/950 that of gasoline, necessitates an energy-intensive densification strategy to get an acceptable driving range with a reasonably-sized vehicular fuel tank. The incumbent densification processes, compressed natural gas (CNG) and liquified natural gas (LNG), are a major hindrance to the commercial adoption of passenger vehicles fueled by natural gas because of the costly infrastructure required for refilling stations and vehicular fuel tanks [55, 56]. For LNG, the energy-intensive process of cooling methane to a liquid at c.a. -162°C [9] requires an expensive liquifaction process. To store LNG onboard the vehicle, expensive cryogenic fuel tanks are required, and LNG suffers from boil-off losses [57]. For CNG, where natural gas is compressed to 200-300 bar at ambient temperature, costly multi-stage compressors are required at refilling stations to achieve the high pressures. To withstand such high pressures, CNG tanks are thick-walled and therefore heavy and relatively expensive. Further, the tanks must be cylindrical or spherical to evenly distribute stress [58]. This makes for a cumbersome, large, non-conformable tank that must be placed in the trunk of passenger vehicles.

Adsorbed natural gas (ANG) technology is an alternative densification strategy at ambient temperatures where we pack the vehicular fuel tank with a nanoporous adsorbent material. The natural gas would be delivered to the engine from the ANG tank by a pres-

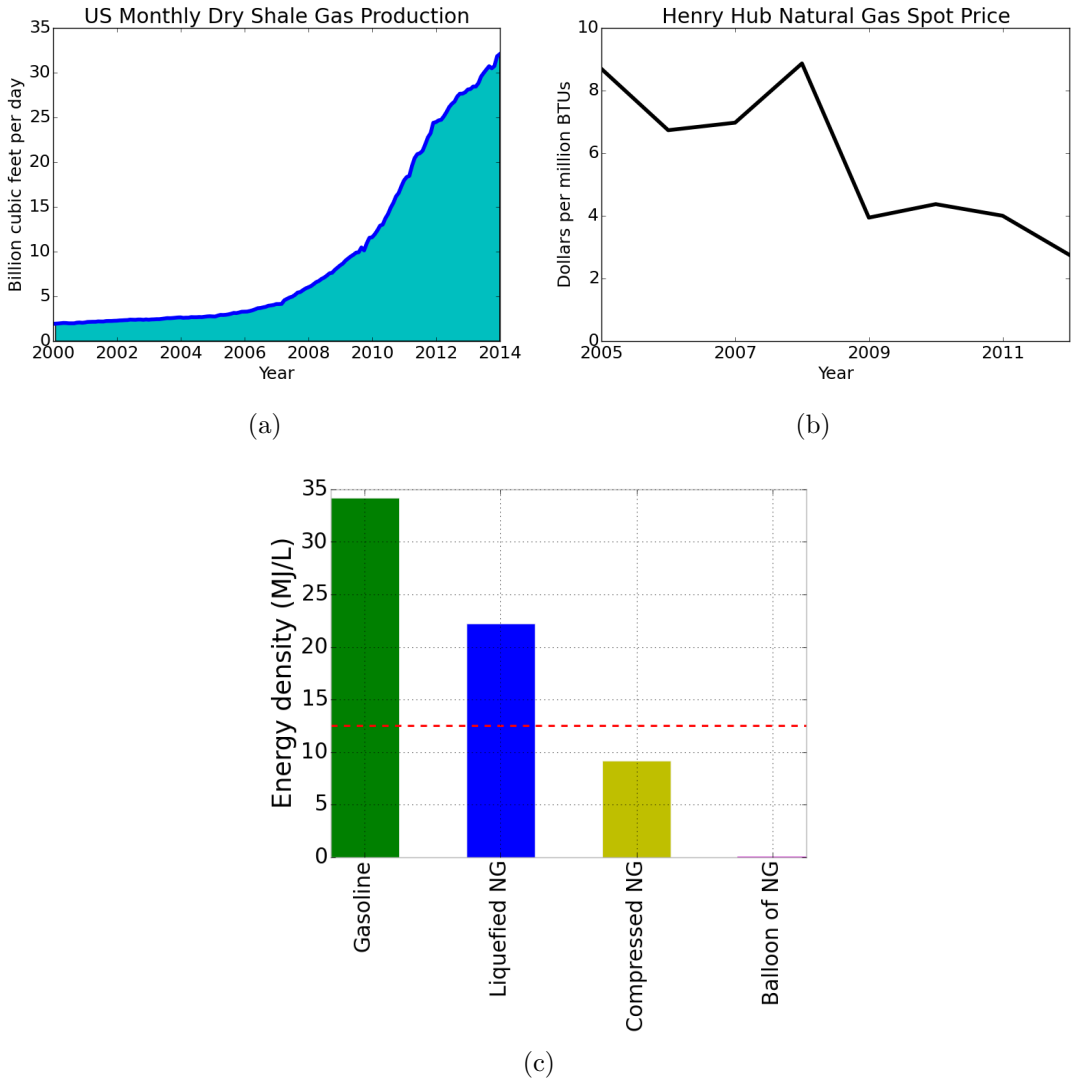


Figure 1.3: Adsorbed natural gas technology motivation. (a) Dry shale gas production in the US has increased dramatically due to the implementation of hydraulic fracturing and horizontal drilling technologies. [45] (b) Natural gas wholesale prices have dropped substantially in recent years, likely as a result of the increased supply. Data from Ref [46]. (c) The volumetric energy density of gasoline, LNG, CNG, and methane at ambient conditions. The energy density of methane gas at ambient conditions is 0.036 MJ/L, highlighting the importance of a densification strategy for adoption as a vehicular fuel. The dashed line indicates the ARPA-E single-crystal deliverable capacity target, set above the energy density of CNG in anticipation of packing losses. (c) from Ref. [34], used under Creative Commons Attribution License.

sure swing. By exploiting attractive interactions between the solid and methane gas, the goal is to recruit methane into the ordered pores of the material to result in storing an amount of methane per volume competitive to CNG, albeit at a more moderate pressure (35-65 bar). This lower storage pressure mitigates refilling infrastructure costs because less expensive, two-stage compressors can achieve 65 bar of pressure in comparison to CNG that stores at 200 bar. As 60 million US households have natural gas delivery [55], homeowners could potentially refill their fuel tank in their own garage with these cheaper compressors. Further, a thinner-walled and thus lighter and less expensive fuel tank is required to withstand 65 bar, possibly enabling conformable fuel tanks that, in contrast to CNG tanks, do not occupy cargo space in the vehicle. ANG technology thus could allow us to realize the economic and environmental benefits of natural gas as a vehicular fuel in passenger vehicles [9].

The Advanced Research Projects Agency - Energy (ARPA-E) of the US Department of Energy uses a target of 12.5 MJ/L deliverable capacity for natural gas sorbents [59] at 298 K using a pressure range of 65 to 5.8 bar, which was recently modified from the old pressure range of 35 to 1 bar. The discharge pressure was raised from 1 bar in recognition that a sufficient pressure differential must be present to drive a sufficient methane flow rate from the adsorbent to the engine. The charging pressure of $P_H = 65$ bar considers the threshold pressure for cheaper infrastructure at refilling stations and lighter-weight, conformable tanks. For comparison with the ARPA-E target, the volumetric energy density of CNG is 8.9 MJ/L (200 bar, 298 K, Peng-Robinson EOS) and of LNG is 22.2 MJ/L (-161.5°C), which is only 27% and 64% of gasoline (34.2 MJ/L), respectively [9]. Using the energy of combustion of methane (-889 kJ/mol), the DOE target for adsorbed natural gas is 315 v STP/v, where v denotes a unit of volume (this is non-dimensional, so the units for volume are arbitrary). Once materials survive the 315 v STP/v deliverable capacity screening process, other properties must be considered, e.g., stability, good dynamic performance under fill-refill cycles [60, 39], resistance to impurities, de/adsorption kinetics, and cost. APRA-E also has a gravimetric deliverable capacity target, but the volumetric deliverable capacity is more important for improving the driving range from a tank fill-up [51].

In Chapter 2, we build a series of simple models of adsorption to gain insights into how to optimize nanoporous materials for methane storage. In parallel, we conducted a high-throughput screening of zeolites for methane storage and interpreted this data in light of our simple models.

In Chapter 3, we conduct a high-throughput screening of several classes of nanoporous materials, including metal-organic frameworks, zeolites, porous polymer networks, and zeolitic imidazolate frameworks, for methane storage. Our data reveals design rules and, although we originally set out to identify materials that exceed methane storage targets, our data strongly suggests that current methane storage targets are too ambitious for rigid materials.

In Chapter 4, we constructed two simple models to assess thermodynamic limitations to the deliverable capacity of methane in a nanoporous material.

Xe/Kr separations

The noble gases xenon (Xe) and krypton (Kr) have several important applications [61]. Xenon is used as an anesthetic [62, 63, 64] and for imaging [65] in the health industry and as a satellite propellant in the space industry [66]. Both xenon and krypton are used in lighting [67], in lasers [68, 69], in double glazing for insulation [70, 71], and as carrier gases in analytical chemistry [72]. As krypton and xenon are present in Earth's atmosphere at concentrations of 1.14 and 0.087 ppm, respectively [73], the conventional method to obtain xenon and krypton is as a byproduct of the separation of air into oxygen and nitrogen by cryogenic distillation [74]. This byproduct stream from air separation consists of 80% krypton and 20% xenon [75]. At Air Liquide, this mixture is compressed to 200 bar and stored in cylinders, then sent to a separate Xe-Kr separation plant to undergo *another* cryogenic distillation to obtain pure xenon or krypton [76]. Cryogenic distillation for the separation of krypton and xenon has a very high energy and capital requirement, reflected by the cost of high-purity xenon, ca. $5,000/kg [77]. An alternative, potentially energy- and cost-saving Xe/Kr separation process is to use a nanoporous material as an adsorbent to selectively adsorb either Kr or Xe at ambient temperature through a temperature- and/or pressure-swing adsorption process [78, 79].

Experiments regarding xenon and krypton adsorption [80, 81, 82, 83, 84, 85, 77, 86, 87, 88, 89, 90, 91, 92, 93] suggest that it may be feasible to use nanoporous materials in an selective adsorption-based process to separate a Xe/Kr mixture.

In a distillation, one exploits the differing boiling points of Xe (119.8 K at 1 atm) and Kr (164.9 K at 1 atm). For an adsorption-based separation of xenon and krypton, we seek to exploit the differences in their size and van der Waals interactions with the framework atoms of the nanoporous materials. Xenon is larger than krypton (van der Waals radii: $r_{Xe} = 1.985$ Å and $r_{Kr} = 1.83$ Å) and has a deeper Lennard-Jones attractive potential well (see Figure 1.4). Thus, we expect the majority of materials to selectively adsorb xenon over krypton; other materials will be reverse selective if their pores are the appropriate size for krypton, but too small for xenon.

In Chapter 5, I outline how we screened over half a million experimentally reported and predicted/hypothetical nanoporous materials for separating a 20/80 mol % Xe/Kr mixture at 1 bar and 298 K. The screening reveals design rules for synthesizing a Xe-selective material and rationalizes the importance of pore size. Because of the large number of structures screened, a brute-force computational strategy would require exorbitant resources. Instead, we employed a machine learning algorithm to learn the relationship between more quickly computed structural descriptors and selectivity, allowing us to quickly rule out a large fraction of materials on the basis of their quickly computed structural descriptors. This introduction is from Ref. [95].

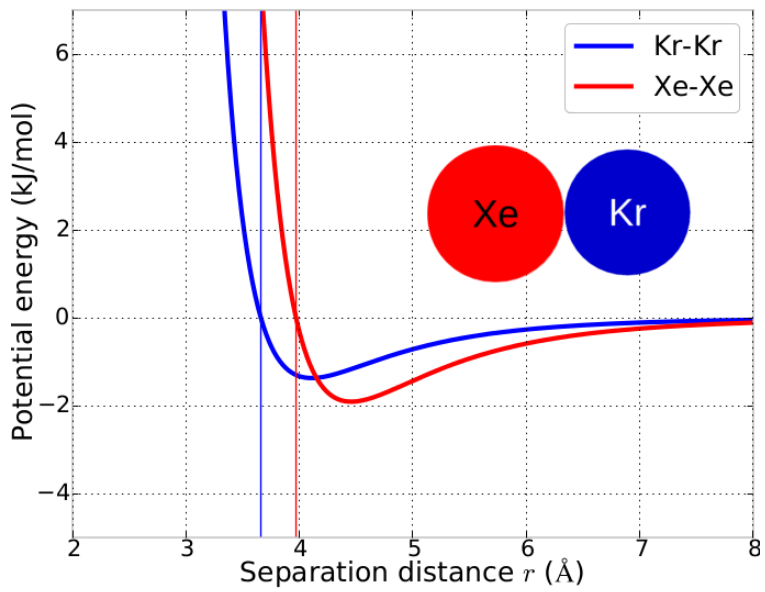


Figure 1.4: Exploiting difference in size and potential well depth of Xe/Kr for separation. The Kr-Kr and Xe-Xe Lennard-Jones potentials from Ref [94]. Vertical lines denote the zero of the potential, yielding van der Waals radii of $r_{Xe} = 1.985$ Å and $r_{Kr} = 1.83$ Å. Xe and Kr atoms are depicted with sizes proportional to their van der Waals radii to visualize the difference in size. Reprinted with permission from Ref. [95]. Copyright 2015 American Chemical Society.

Xe capture from air

One of the grandest challenges of our generation is to meet our rapidly growing energy demand without further increasing the emission of greenhouse gases[96, 55]. Nuclear energy is one of the cheapest alternatives to carbon-based fossil fuels that, because of its high energy density and minimal land use requirements, can be scaled up to help meet global energy demands. Life cycle analyses indicate that greenhouse gas emissions of a nuclear power plant are significantly lower than fossil fuel technologies and comparable to other renewable electricity generation technologies, such as solar photovoltaics[97]. For the mass implementation of nuclear energy as a low-emissions energy source, we must also safely sequester the associated high-level radioactive waste[98]. In this, most attention is given to recovering the heavy, long-lived nuclear elements in used nuclear fuel (UNF), such as uranium and plutonium. Less discussed are the volatile radionuclides (for example, Xe, Kr) that evolve into the off-gas of UNF aqueous reprocessing facilities[98]. In these off-gases, gaseous radioactive ^{85}Kr has a long half-life ($t_{1/2} = 10.8$ years) and therefore must be captured and removed from the off-gas to prevent its uncontrolled release into the atmosphere[98]. In contrast, the radioactive Xe isotopes ($t_{1/2} \sim 36.3$ days for ^{127}Xe) have decayed by the time the fuel is reprocessed. As high purity Xe is used in many applications, including commercial lighting, propulsion, imaging, anesthesia and insulation, the recovered Xe could be sold into

the chemical market to offset operating costs. At present, cryogenic distillation is the most mature technology to separate Xe and Kr from air, but it is energy- and capital-intensive and therefore expensive. Furthermore, the radiolytic formation of ozone poses an explosion hazard during cryogenic distillation[98]. These factors incentivize the development of an alternative technology for a less energy-intensive, more cost-effective and safer process to capture Kr and Xe from UNF reprocessing facility off-gas.

A promising alternative technology for Xe/Kr removal from reprocessing off-gas is an adsorption-based process at room temperature using a selective, solid-state adsorbent [13]. These solid-state adsorbents are found to be almost exclusively Xe-selective, and thus a dual step process whereby, first, the Xe is selectively removed from the off-gas, is a necessary requirement for a practical application[99]. In the subsequent step, the radioactive Kr can be removed from the Xe-free effluent using the same material or a different material. Adsorbents such as activated carbon have been proposed[86, 98], but these fall short compared with high surface area, crystalline metal-organic frameworks (MOFs) and porous organic cage compounds. Among the many novel materials tested thus far, SBMOF-1 has shown to be the most promising for Xe/Kr separations, showing high capacity and good selectivity for Xe over Kr. See Fig 1.5.

In Chapter 7, our high-throughput computational screening of MOFs for Xe capture at dilute conditions leads to the characterization of SBMOF-1 as the most Xe-selective material reported in the literature. Our colleagues at Pacific Northwest National Laboratory carried out the experimental work, which is an integral part of this computer-assisted materials discovery. Part of this introduction is from Ref. [100].

1.2 Molecular Simulation Methods

Here, we outline the methods used throughout this thesis to simulate the adsorption of gas into the pores of nanoporous crystals. Throughout, we treat the crystal structure as a rigid host, and thus the nanoporous crystal is determined by the Bravais unit cell, the identity of the atoms in a unit cell, and the fractional coordinates of these atoms.

Molecular models

For a molecular simulation, we require a model of the energetic landscape of a gas molecule in the pore of a nanoporous crystal. Consider a nonpolar gas molecule such as xenon (Xe) or krypton (Kr). We model the [dominantly dispersive] interactions of such a gas molecule with atom i of the nanoporous material using a Lennard-Jones potential:

$$U_i(r) = 4\epsilon_i \left[\left(\frac{\sigma_i}{r}\right)^{12} - \left(\frac{\sigma_i}{r}\right)^{6} \right], \tag{1.1}$$

where U_i is the energy of the interaction and r is the distance between the gas molecule and atom i of the nanoporous crystal. The meaning of the parameters σ_i and ϵ_i are illustrated

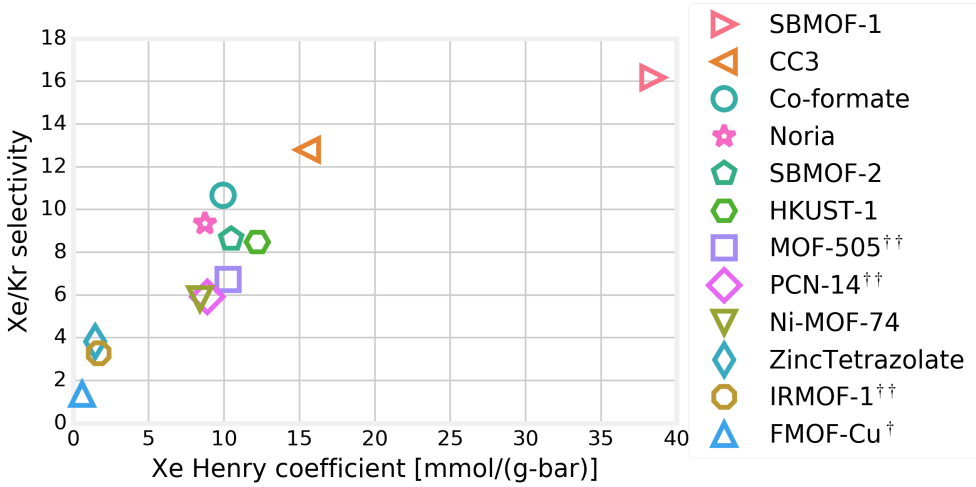


Figure 1.5: Survey of thermodynamic Xe/Kr separation performance in top-performing materials. Henry coefficients are extracted from pure-component Xe and Kr adsorption isotherms reported in the literature (see Refs. [100, 101]). Data at 298 K, exceptions denoted by a dagger (†) for 297 K and a double dagger (††) for 292 K. Survey includes CC3 [77], Co-formate [90, 85], noria [101], SBMOF-2 [102], HKUST-1 [103], MOF-505 [92], PCN-14 [104], Ni-MOF-74 [86], Zinc tetrazolate [105], IRMOF-1 [83], and FMOF-Cu [87].

in Fig 1.6. The parameters ϵ and σ for different interactions (e.g., xenon-xenon) can then be taken from the literature, where these parameters were fit to reproduce experimental data, for example vapor-liquid equilibria [106] or sublimation energies [107]. Commonly used off-the-shelf molecular models used in this thesis include the Dreiding force field [107], the Universal Force Field [108], and Transferable Potentials for Phase Equilibria (TraPPE) [106]. Mixing rules, such as the Lorentz-Berthelot mixing rules, then dictate cross-interactions. Note that we can model methane (CH_4) as a single, united atom [106].

We can then model the interaction of a gas molecule with the crystal as pairwise additive, as a sum over all atoms of the crystal:

$$U(r) = \sum_i U_i(r), \tag{1.2}$$

where $U_i(r)$ is given in eqn 1.1. As the Lennard-Jones potential diminishes to zero with power r^{-6} as r increases, we typically truncate Lennard-Jones potentials and approximate them as zero beyond a critical radius r_c. This allows us to model the interaction of a gas molecule with a crystal of infinite extent by considering all crystal atoms within a distance r_c from the gas molecule.

Gas-gas interactions for nonpolar gases are also modeled with the Lennard-Jones potential in eqn 1.1. In this thesis, periodic boundary conditions are applied to the gas-gas interactions to, again, model adsorption in a perfect crystal of infinite extent.

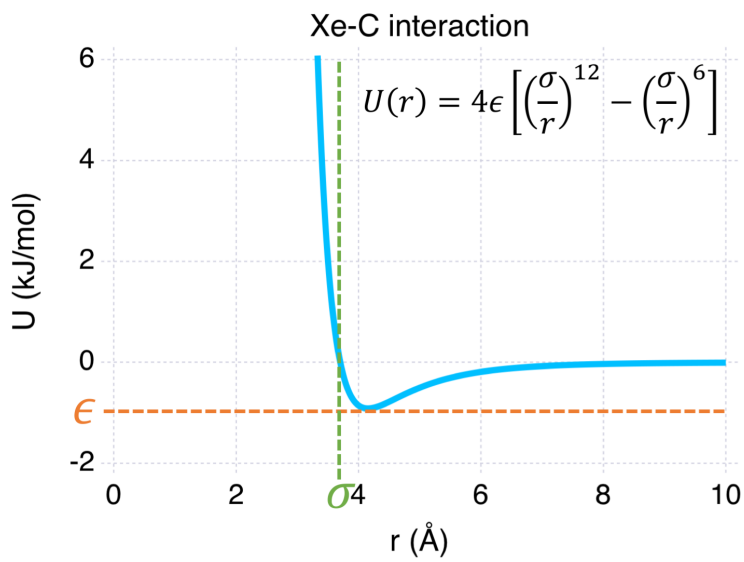


Figure 1.6: The blue curve shows the Lennard-Jones potential modeling the interaction of a xenon and carbon atom at distance r from one another. The ϵ parameter is the minimum value of $U(r)$. The σ parameter is the distance such that $U(\sigma) = 0$.

Given N gas molecules in the pores of a rigid nanoporous material (a Bravais unit cell taken as a simulation box with periodic boundary conditions applied) at positions $\mathbf{x}_1$, $\mathbf{x}_2$, ..., $\mathbf{x}_N$, the total interaction energy of the system is then determined by this molecular model.

Molecular simulations

The grand-canonical statistical mechanical ensemble (μVT ensemble) mimics experimental gas adsorption isotherm measurements, where a sample of a nanporous material is immersed in a bath of gas at pressure P. In the grand-canonical ensemble, the temperature T, volume of nanoporous material V, and chemical potential μ of the bath of gas are held fixed. The energy and number of adsorbed gas molecules in the system fluctuate in the grand-canonical ensemble. The chemical potential μ of a gas is related to pressure via an equation of state, such as the ideal gas law or van der Waals equation of state. For example, for an ideal gas at temperature T and pressure P, the chemical potential μ is [109]:

$$\mu = k_B T \log(\beta P \Lambda^3), \tag{1.3}$$

where k_B is the Boltzmann constant and Λ is the thermal de Broglie wavelength [109].

The partition function of a rigid nanoporous material immersed in a bath of gas in the grand-canonical ensemble is a sum of the statistical weight $e^{\beta\mu N_\nu - \beta E_\nu}$ over all microstates ν:

$$\Xi(\mu, V, T) = \sum_{N=0}^{\infty} \frac{e^{\beta\mu N}}{\Lambda^{3N} N!} \int e^{-\beta U(\mathbf{x}^N)} d\mathbf{x}^N, \tag{1.4}$$

where N is the number of gas molecules, $\mathbf{x}^N$ is a convenient notation for the positions of these gas molecules in the pores of the material, and $\beta := 1/(k_B T)$ [110]. The energy of a particular state of the system, characterized by (N, x^N), $U(\mathbf{x}^N)$ is determined by the molecular model above. The probability density $p(N, \mathrm{x}^N)$ in the μVT ensemble is proportional to the factor:

$$p(N, \mathrm{x}^N) \sim \frac{e^{\beta\mu N}}{\Lambda^{3N} N!} e^{-\beta U(\mathbf{x}^N)}. \tag{1.5}$$

With the goal of finding the expected value of N, we turn to Markov Chain Monte Carlo simulations to simulate the probability distribution in eqn 1.5. We propose three Markov chain transition proposals:

- Particle insertion: we propose to add a particle to the system at a random position $\mathbf{x}$ uniformly distributed in the simulation box of volume V.
- Particle deletion: we propose to remove a uniformly random selected particle.
- Particle translation: we perturb the coordinates of a uniformly random selected particle by placing it in a uniformly distributed position about a cube centered at the current position.

To achieve a stationary distribution, we impose detailed balance:

$$p[(N', \mathrm{x}'^{N'})|(N, \mathrm{x}^N)]p[(N, \mathrm{x}^N)] = p[(N, \mathrm{x}^N)|(N', \mathrm{x}'^{N'})]p[(N', \mathrm{x}'^{N'})]. \tag{1.6}$$

Here, the prime denotes a different state. The conditional probability density $p[(N', \mathrm{x}'^{N'})|(N, \mathrm{x}^N)]$ refers to, in the Markov chain, moving to state $(N', \mathrm{x}'^{N'})$ given that we are currently in state (N, x^N). We now separate the conditional distribution into a *proposal distribution* $g[(N', \mathrm{x}'^{N'})|(N, \mathrm{x}^N)]$ which is the conditional probability density of proposing to move to state $(N', \mathrm{x}'^{N'})$ given that we are currently in state (N, x^N) and an *acceptance distribution* $A[(N', \mathrm{x}'^{N'})|(N, \mathrm{x}^N)]$, which is the conditional probability density to accept the transition:

$$p[(N', \mathrm{x}'^{N'})|(N, \mathrm{x}^N)] = g[(N', \mathrm{x}'^{N'})|(N, \mathrm{x}^N)]A[(N', \mathrm{x}'^{N'})|(N, \mathrm{x}^N)]. \tag{1.7}$$

Then, in terms of the proposal and acceptance distributions, the detailed balance in eqn 1.6 holds if:

$$\frac{A[(N', \mathrm{x}'^{N'})|(N, \mathbf{x}^N)]}{A[(N, \mathbf{x}^N)|(N', \mathrm{x}'^{N'})]} = \frac{p[(N', \mathrm{x}'^{N'})]}{p[(N, \mathrm{x}^N)]} \frac{g[(N, \mathrm{x}^N)|(N', \mathrm{x}'^{N'})]}{g[(N', \mathrm{x}'^{N'})|(N, \mathrm{x}^N)]}. \tag{1.8}$$

The Metropolis-Hastings acceptance distribution A achieves the above condition:

$$A[(N', \mathrm{x}'^{N'})|(N, \mathrm{x}^N)] := \min\left(1, \frac{p[(N', \mathrm{x}'^{N'})]}{p[(N, \mathrm{x}^N)]} \frac{g[(N, \mathrm{x}^N)|(N', \mathrm{x}'^{N'})]}{g[(N', \mathrm{x}'^{N'})|(N, \mathrm{x}^N)]}\right) \tag{1.9}$$

If the particle insertion and deletion moves are proposed with equal probability:

$$g[(N+1, \mathbf{x}^{N+1})|(N, \mathbf{x}^N)] = g[(N, \mathbf{x}^N)|(N+1, \mathbf{x}^{N+1})], \tag{1.10}$$

and by construction of the particle translation move:

$$g[(N, \mathbf{x}^N)|(N, \mathbf{x}'^N)] = g[(N, \mathbf{x}'^N)|(N, \mathbf{x}^N)]. \tag{1.11}$$

Therefore, using eqn 1.5, the acceptance probability of a particle insertion proposal is:

$$A[(N+1, \mathbf{x}^{N+1})|(N, \mathbf{x}^N)] = \min\left(1, \frac{V}{\Lambda^3(N+1)} e^{\beta\mu N} e^{-\beta(U(N+1,\mathbf{x}^{N+1})-U(N,\mathbf{x}^N))}\right); \tag{1.12}$$

of a deletion proposal:

$$A[(N-1, \mathbf{x}^{N-1})|(N, \mathbf{x}^N)] = \min\left(1, \frac{\Lambda^3 N}{V} e^{-\beta\mu N} e^{-\beta(U(N-1,\mathbf{x}^{N-1})-U(N,\mathbf{x}^N))}\right); \tag{1.13}$$

and of a translation proposal:

$$A[(N, \mathbf{x}'^N)|(N, \mathbf{x}^N)] = \min\left(1, e^{-\beta(U(N,\mathbf{x}'^N)-U(N,\mathbf{x}^N))}\right). \tag{1.14}$$

These state Markov chain transition proposals and associated acceptance rules then determine how we conduct a grand-canonical Monte Carlo simulation in practice.

This algorithm is outlined in the textbook in Ref. [110].

Infinite dilution

At dilute conditions ($P \to 0$), we can use a much cheaper algorithm than Markov chain Monte Carlo simulations. Let us use eqn 1.3 to write the grand-canonical partition function in eqn 1.4 in terms of pressure:

$$\Xi(P, V, T) = \sum_{N=0}^{\infty} \frac{(\beta P)^N}{N!} \int e^{-\beta U(\mathbf{x}^N)} d\mathbf{x}^N. \tag{1.15}$$

As $P \to 0$, the leading order terms are:

$$\Xi(P, V, T) \sim 1 + \beta P \int e^{-\beta U(\mathbf{x})} d\mathbf{x}. \tag{1.16}$$

The expectation of N is given by a derivative of the partition function:

$$\langle N \rangle = \frac{\partial \log \Xi}{\partial(\beta\mu)} = P \frac{\partial \log \Xi}{\partial P}. \tag{1.17}$$

Using the expansion in eqn 1.16 and carrying out the above derivative, we find that, as $P \to 0$:

$$\langle N \rangle = \left(\beta \int e^{-\beta U(\mathbf{x})} d\mathbf{x}\right) P + O(P^2). \tag{1.18}$$

The term in the parenthesis multiplying P is the *Henry coefficient*, K_H. This is a derivation of Henry's law $\langle N \rangle = K_H P$ which is often used to model adsorption at dilute conditions. The Henry coefficient K_H can thus be computed as β times the spatial average of the Boltzmann factor of a gas molecule in the pores of the material:

$$K_H = \beta \int e^{-\beta U(\mathbf{x})} d\mathbf{x}. \tag{1.19}$$

Eqn 1.19 is also derived in Ref. [110]. While one could use numerical quadrature to compute the above average, in this thesis, we use Monte Carlo integration to compute the spatial average when we seek to compute the Henry coefficient. As outlined by Kim and Smit [111], the computation of a Henry coefficient is an embarrassingly parallel algorithm. We can assign each thread on a graphics processing units (GPU) or any parallel computing architecture a different position $\mathbf{x}$ of an adsorbed molecule and compute its energy there in parallel, as a sample for the average in eqn 1.19.

Chapter 2

Optimizing nanoporous materials for gas storage

In this work, we address the question of which thermodynamic factors determine the deliverable capacity of methane in nano-porous materials. The deliverable capacity is one of the key factors that determines the performance of a material for methane storage in automotive fuel tanks. To obtain insights in how the molecular characteristics of a material are related to this deliverable capacity, we developed several statistical thermodynamic models. The predictions of these models are compared with the classical thermodynamics approach of Bhatia and Myers [Bhatia and Myers *Langmuir*, 2005, **22**, 1688] and with the results of molecular simulations in which we screen the IZA zeolite structures and a hypothetical zeolite database of over 100,000 structures. Both the simulations and our models do not support the rule of thumb that, for methane storage, one should aim for an optimal heat of adsorption of 18.8 kJ/mol. Instead, our models show that one can identify an optimal heat of adsorption, but that this optimal heat of adsorption depends on the pore of the material. The different models we have developed are aimed to provide guidelines on how this optimal heat of adsorption is related to the molecular structure of the material.

This chapter is based on the following publication [112]:

C. Simon, J. Kim, L.C. Lin, R.L. Martin, M. Haranczyk, B. Smit. Optimizing nanoporous materials for gas storage. *Physical Chemistry Chemical Physics.* (2014) 16 (12): 5499 -5513.

2.1 Introduction

A potential solution to enhance the volumetric energy density of methane for use as a vehicular fuel without an expensive or energy-intensive densification process is adsorbed natural gas technology. See Sec. 1.1.

A large body of research [113, 114, 115, 116, 117] aims to improve nanoporous materials for methane storage by modifying the framework chemistry, thereby tuning the interactions between the framework atoms and the methane [118]. With the goal of achieving an energy that is competitive with gasoline using an adsorbent, we must maximize the quantity of methane stored in the material at the high pressure, P_H, of a filled tank while, at the same time, minimizing the retention of gas at the low pressure, P_L, of a depleted tank, the "cushion gas". The *deliverable capacity* or *working capacity* is defined as the difference in the loading in the adsorbent between the two pressure ranges.

Here, we use a computational approach for a high-throughput screening of over 100,000 zeolites from the hypothetical database generated by Deem and coworkers [119, 120] and the IZA database as adsorbent candidates for methane storage. While our screening study shows that zeolites are not particularly promising for exceeding ARPA-E targets, they serve here as a data set to investigate what properties endow a material with a large deliverable capacity, which is the goal of this work. Wilmer et al. [121] similarly explored geometrical relationships with methane storage capacity in a large number of hypothetical metal-organic frameworks for methane storage.

In addition, we aim to understand how the heat of adsorption affects the deliverable capacity. Bhatia and Myers [122] were among the first to systematically study the relationship between the heat of adsorption and the deliverable capacity for a Langmuir isotherm. They noted that, to yield a high deliverable capacity, the affinity for methane in the adsorbent framework must be strong enough to store a large amount at the charging pressure P_H, yet weak enough to release most of the methane at the discharge pressure P_L [122]. Using experimental data for carbonaceous materials, Bhatia and Myers use classical thermodynamic arguments to demonstrate that the optimal heat of adsorption for methane storage is 18.8 kJ/mol for $P_H = 30$ and $P_L = 1.5$ bar. Their highly cited work provided much needed guidance, so much that the 18.8 kJ/mol target is now frequently used for a far wider class of materials than those that were envisioned in the original work of Bhatia and Myers. This raises the question whether the assumptions underlying the work of Bhatia and Myers hold for such a wide class of materials.

Snurr and coworkers investigated the relation between the heat of adsorption and deliverable capacity by artificially changing Lennard Jones interaction parameters in a few select MOFs to tune the heat of adsorption *in-silico* [123]. They indeed observed an optimal heat of adsorption, but it was different for each material. In other work[124], they derive an upper bound for the optimal heat of adsorption by considering an adsorption site that results in complete loss of entropy of the adsorbate molecule. Garrone et al. [125] pointed out that the entropy change upon adsorption is not the same for all materials and, using a positive correlation found between enthalpy and entropy changes upon adsorption, adjusted Bhatia and Myers' formula. We explain the enthalpy-entropy correlation and its consequences for designing the optimal material.

Our high-throughput, *in-silico* screening facilitates our study of the relationship between material performance and key geometric and thermodynamic parameters. Moreover, to obtain a fundamental understanding of these relationships, we build a sequence of statistical

mechanical models of adsorption.

2.2 Zeolite screening data

Here, we computationally screen 136,328 hypothetical and 187 IZA zeolites for methane storage and investigate what properties endow zeolites with a high deliverable capacity.

Methods

The zeolite crystal structures were obtained from the SLC hypothetical database generated by Deem and coworkers [120]. We analyzed only structures with pore sizes accessible to methane (> 1.625 $\AA$ radius). Geometrical properties of the zeolite structures, such as accessible surface area and largest-included-sphere pore diameters are obtained using a Voronoi decomposition approach in the Zeo++ software [126]. For every structure, we computed the equilibrium number of methane guest molecules per unit crystal cell as a function of temperature and chemical potential (related to pressure) with the Grand-canonical Monte Carlo Algorithm [110] (see Sec. 1.2). A Lennard-Jones potential models the framework-guest and guest-guest interactions using force field parameters developed in Refs. [127, 128]. This force field [127, 128] was specifically tuned to reproduce the adsorption isotherms of methane and other alkanes in zeolites. To mimic an infinite crystal, we implement periodic boundary conditions with a shifted and truncated Lenard Jones potential with a 12.0 $\AA$ cutoff. For identification of top performing materials, we use the Peng-Robinson equation of state to convert fugacity to the experimentally relevant mechanical pressure.

We take into account in our simulations that some zeolites have adsorption pockets that can accommodate a methane molecule, but are not accessible from the gas phase. In our Monte Carlo simulations, we insert molecules at random position and we therefore need to block these sites to ensure that these are not sampled. In most conventional simulations, such pockets are identified by visual inspection. For screening large databases, we have developed a high-performance algorithm [129, 130] for this task. We segment the void space in the material into *pockets* (inaccessible regions of void space) and *channels* by inspecting the three-dimensional potential energy profile of methane in the material, computed prior to Monte Carlo sampling for speed-up of force computations. We assume that the guest molecule is not able to cross an energetic barrier greater than $15k_BT$ and perform a multicore CPU flood fill process to segment the energy grid with respect to this barrier. We then identify how these segments are connected periodically, which affords the use of a convenient definition of a channel - a channel is a region of space from which you can travel to the equivalent region in another unit cell; everything else must be a pocket. Grid points within pockets are reassigned to a very large energy value, such that when Monte Carlo moves land within a pocket, they will experience a very large repulsive energy and be rejected, just as if the move had landed within an atom of the material.

The isosteric heat of adsorption is defined as the negative of the enthalpy change upon adsorption ΔH. The energy of adsorption differs from the enthlapy of adsorption by a work term ($\Delta H = \Delta U - RT$) [131]. We perform Widom insertions [110] to obtain the ensemble average energy of adsorption at infinite dilution (U_0 in following equations) and the Henry coefficient.

Adsorption isotherms, Henry coefficients, heats of adsorption, and geometric data of each structure are available as Supplementary data, including a list of the top 50 performers in both the 65 to 5.8 bar and 35 to 1 bar pressure ranges. In the scatter plots that follow, each zeolite structure is plotted as a point.

We point out that our molecular simulations, which implement periodic boundary conditions, predict the uptake in a perfect single crystal. Ultimately, the adsorbent in the final methane storage tank, which is not a single-crystal, will depend upon the bulk adsorbent properties. For an adsorbent with a high-pressure uptake greater than free space, the deliverable capacity of the final, bulk adsorbent will necessarily be less than that of the single-crystal; hence, our reported single-crystal deliverable capacities of high-performing materials are an upper bound for the deliverable capacity of an adsorbed natural gas tank. However, the DOE target of 315 v STP/v anticipates this and applies to the single crystal deliverable capacity [59].

Results

We obtain adsorption isotherms at 298 K from 1-200 bar from Grand-canonical Monte Carlo simulations. We used these to plot the deliverable capacity against loading at $P_H = 65$ bar in Fig 2.1a. To compare with using a free-space tank, the deliverable capacity and density at 65 bar of an ideal gas is plotted as horizontal and vertical lines, respectively.

Fig 2.1a illustrates that a high loading is necessary, but not sufficient, for a high deliverable capacity. Two materials with the same high-pressure loading can exhibit substantially different deliverable capacities depending on the shape of the isotherm (see Fig 2.1b). As an example, structure PCOD8294531 (red star) has a promising 65 bar loading of 163 v STP/v. However, there is still so much methane left at the discharge pressure $P_L = 5.8$ bar that it exhibits a deliverable capacity of only 33% of this, rendering it useless (a free-space tank is better). This highlights the importance to consider more than the absolute storage capacity at 65 bar.

Next, we investigate relationships between deliverable capacity and geometrical properties. First, we explore if the size of the adsorption pore affects the deliverable capacity in Fig 2.1c. We characterize the largest pore size by assigning radii to the framework atoms, modeling them as hard-spheres, and finding the largest sphere that can be included in the framework without overlapping with a framework atom. Structures with the highest deliverable capacities generally occur at an optimal pore size range around 4.6 $\AA$. A second peak occurs around twice 4.6 $\AA$, likely corresponding to more than one methane molecule fitting inside the pore. Pore diameter is not, however, a good predictor of deliverable capacity and thus other variables must be considered. Second, we corroborate claims in the literature

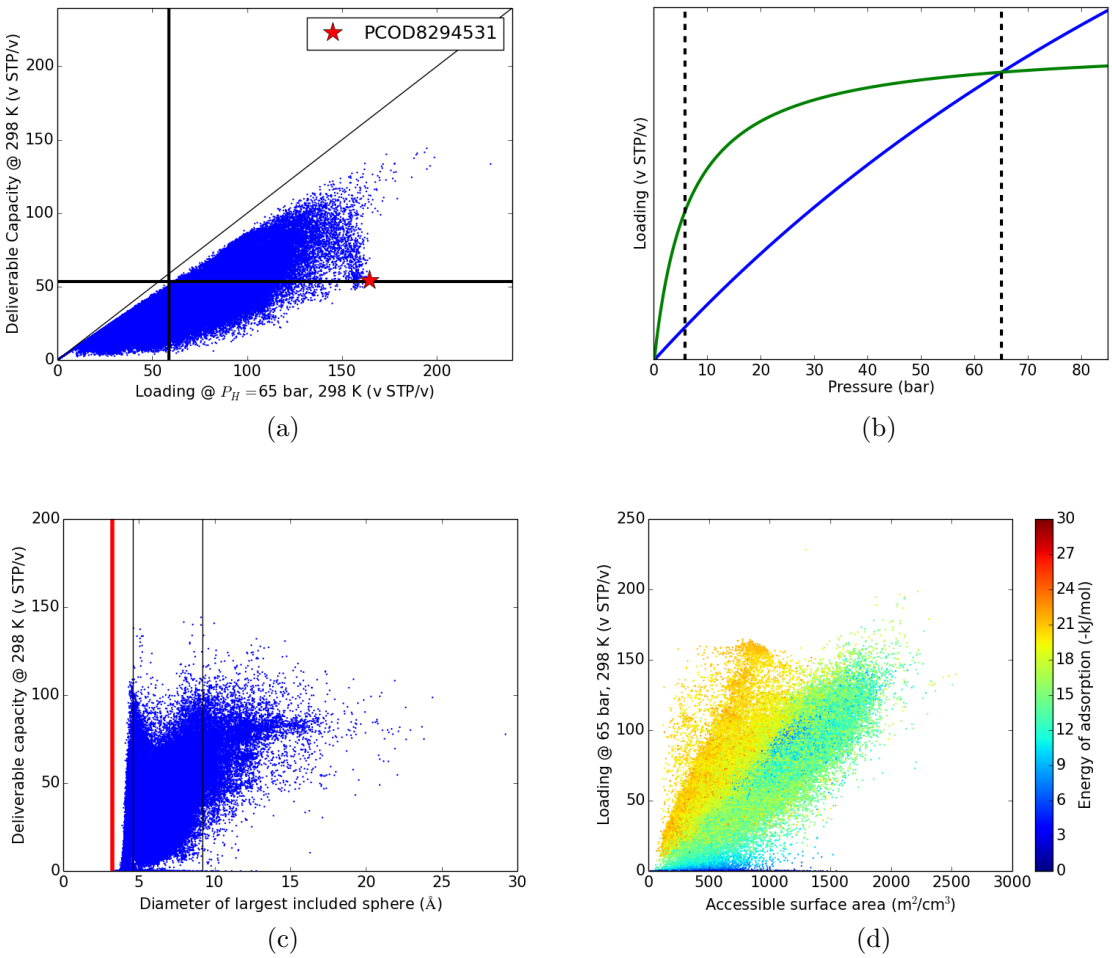
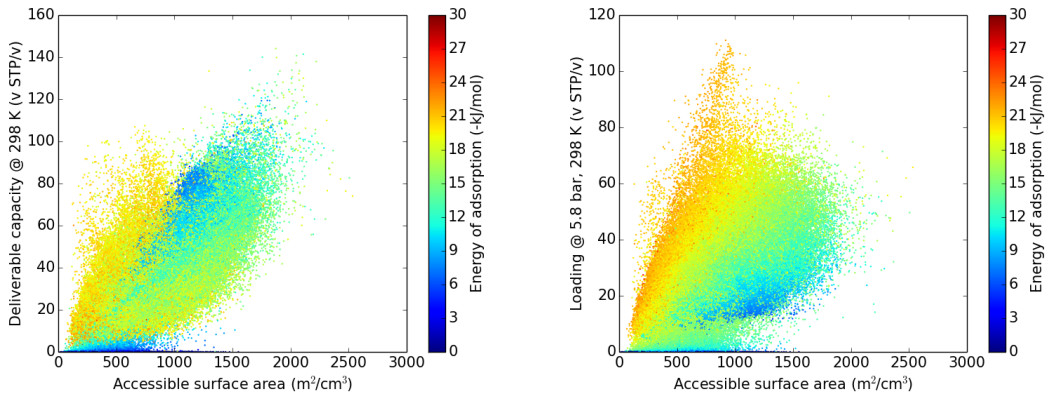


Figure 2.1: (a) Deliverable capacity of structures in the zeolite database against high-pressure loading. The cushion gas (loading at 5.8 bar) can significantly degrade the deliverable capacity from the high-pressure loading. Vertical line: ideal gas density at 65 bar (58.76 v STP/v) Horizontal line: deliverable capacity of free space tank (ideal gas, 53.52 v STP/v) (b) Depending on isotherm shape, two materials with the same high-pressure loading can have very different deliverable capacities. The green isotherm saturates quickly, resulting in a high uptake at the low pressure as well, and thus lower deliverable capacity. (c) Deliverable capacity plotted against the largest included sphere diameter. Red vertical line is hard-sphere diameter of methane, 3.25 $\AA$ [132]. Black vertical lines denote 4.6 $\AA$ and 2*4.6=9.2 $\AA$. (d) Loading at 65 bar vs accessible surface area color-coded according to the energy of adsorption. Figures reproduced from Ref. [112] with permission from the PCCP Owner Societies.

[133] that high surface area polymeric materials result in high uptake with the positive correlation present in Fig 2.1d. Based on the color-coding in Fig 2.1d, a material can somewhat compensate for its lack of accessible surface area by achieving a high heat of adsorption. Because high uptake is a prerequisite for high deliverable capacity, there is also a positive correlation between the deliverable capacity and the surface area, but the relationship with the heat of adsorption is more complicated. See Fig 2.2.



(a) (b)

Figure 2.2: Deliverable capacity (a) and loading at 5.8 bar (b) plotted against accessible surface area. We still see a positive correlation between surface area and deliverable capacity, but the heat of desorption actually decreases the deliverable capacity at high surface areas, unlike what is seen in the loading correlations. Figures reproduced from Ref. [112] with permission from the PCCP Owner Societies.

One of the important predictions of the thermodynamic analysis of Bhatia and Myers is that there is an optimal heat of adsorption for given operating pressures. Using the assumptions in Ref. [122] that the standard entropy change upon adsorption for methane is $\Delta S^\circ = -9.5R$ and our relevant operating pressures $P_H = 65$ bar and $P_L = 5.8$ bar, we expect a heat of adsorption of 16.2 kJ/mol, plotted as vertical line in Fig 2.3, to yield materials with the highest deliverable capacity. We plot the deliverable capacity against the heat of adsorption in Fig 2.3. However, materials endowed with a heat of adsorption of 16.2 kJ/mol exhibit a drastic range of deliverable capacities. Further, there are many high-performance materials with heats of adsorption far from the optimal. To try to better understand this, we color-code the points in Fig 2.3 by the void fraction of the material. To an extent, the void fraction helps distinguish deliverable capacities between materials with the same heat of adsorption; a larger void fraction, implying a more porous material, generally results in a higher deliverable capacity. To understand the discrepancy between Fig 2.3 and the concept of a blanket optimum heat of adsorption, we proceed to develop a series of statistical mechanical models of adsorption.

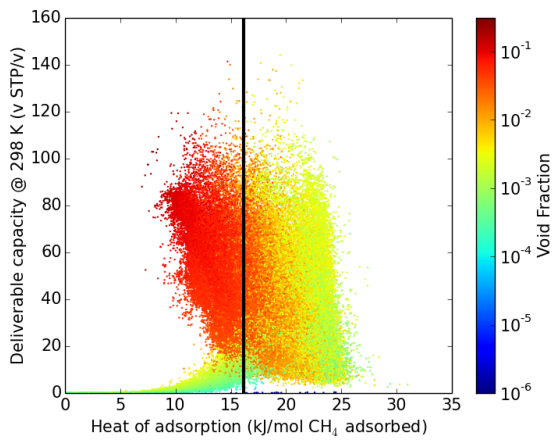


Figure 2.3: Deliverable capacity vs. heat of adsorption color-coded according to the void fraction. Vertical line denotes the optimal heat of adsorption in eqn 2.19 using the assumption $\Delta S^0 = -9.5R$ and a pressure range of 65 and 5.8 bar, 16.2 kJ/mol. Figures reproduced from Ref. [112] with permission from the PCCP Owner Societies.

2.3 Models for Adsorption: Seeking an optimal heat of adsorption

We develop a series of models to quantify and intuit how the heat of adsorption affects the deliverable capacity. We start with a simple model (model 0), which is constructed such that the heat of adsorption is independent of structure of the material. As in this model the entropy can be varied independently, we obtain insight in the role of entropy in determining the optimal adsorption conditions. In model 0, we assume that the gas molecules do not interact; in model 1 we study the effect of gas-gas interactions on the performance of the materials. In model 2, we further study the correlation between heat of adsorption and entropy of adsorption that has been found in experimental studies. Model 3 similarly explores how a heterogeneous energy landscape is linked to the entropy of adsorption.

Model 0: Langmuirian material

In model 0, our idealization of a nanoporous crystalline material is that it consists of M identical, independent adsorption sites in a unit volume. Each crystalline unit cell has a void space Ω_s that is of uniform energy U_0 and large enough for only one gas molecule. The region outside the binding site is inaccessible because it is occupied by framework atoms. See Fig 2.4. The energy landscape of the material is thus:

$$U(x) = \begin{cases} U_0, & x \in \Omega_s \\ \infty, & x \notin \Omega_s. \end{cases} \tag{2.1}$$

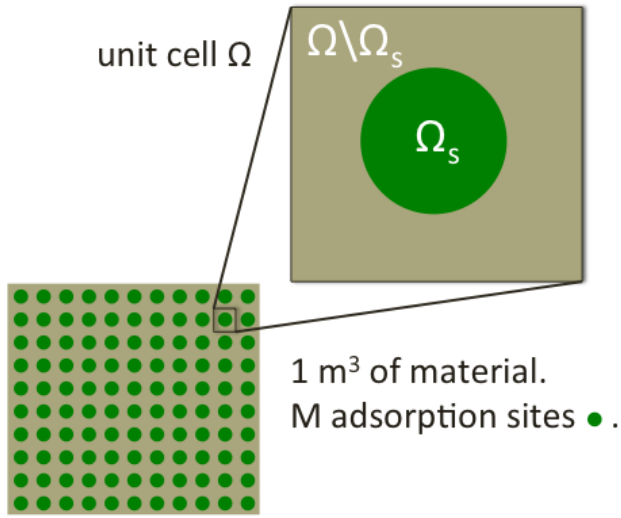


Figure 2.4: Our model material. Crystalline unit cell consists of binding site Ω_s (green) that can be occupied by one guest molecule. The region outside the binding pocket (brown) is unoccupiable due to the presence of framework atoms. The bulk crystal is composed of M unit cells per volume, endowing it with M adsorption sites per unit volume. Figures reproduced from Ref. [112] with permission from the PCCP Owner Societies.

With this energy landscape, we now derive from statistical mechanics the isotherm for this model material under the assumption that nearby guests do not interact. Consider a gaseous phase of methane in equilibrium with an adsorbed phase of methane in a volume V of our model nanoporous material.

Approach 1

The condition for equilibrium between the adsorbed phase (ads) and the ideal gas phase (ig) is the equality of the chemical potential in the two phases:

$$\mu_{ads} = \mu_{ig}. \tag{2.2}$$

We obtain both μ_{ads} and μ_{ig} using the identity that relates the chemical potential to the Canonical partition function Q:

$$\mu = \left(\frac{\partial A}{\partial \sigma}\right)_{T,V} = -\frac{1}{\beta}\left(\frac{\partial \ln Q}{\partial \sigma}\right)_{T,V}. \tag{2.3}$$

When considering the adsorbed phase, M plays the role of volume. Our approach is to find the Canonical partition function for an ideal gas Q_{ig} and for the adsorbed phase Q_{ads}, perform the derivative in eqn 2.3, and equate the two chemical potentials.

The ideal gas. The single-particle partition function is generally:

$$q(V,T) = \frac{1}{\Lambda^3}\int_{\Omega} e^{-\beta U(x)} dx, \tag{2.4}$$

where Ω represents the system volume and $\Lambda = \Lambda(T)$ is the thermal de Broglie wavelength. Because ideal methane molecules do not interact, $U(x) = 0$ and we arrive at $q(V,T) = V/\Lambda^3$. Because ideal gas particles behave independently of one another, the Canonical partition function for an entire system of σ ideal gas particles is:

$$Q_{ig}(\sigma, V, T) = \frac{q(V,T)^\sigma}{\sigma!}, \tag{2.5}$$

where the factorial term takes into account that the particles are indistinguishable. Using Stirling's approximation for the factorial term and the identity in eqn 2.3, we take the derivative of Q_{ig} to arrive at the chemical potential for an ideal gas:

$$\mu_{ig} = \frac{1}{\beta} \ln\left(\Lambda^3 \rho\right), \tag{2.6}$$

where ρ is the density of the ideal gas.

The adsorbed phase. The Canonical partition function q of a single guest particle in a single adsorption site of our model material, considered the unit cell Ω in Fig 2.4, is found by using eqn 2.4 and the energy landscape in eqn 2.1:

$$q = \frac{1}{\Lambda^3} \epsilon \frac{V}{M} e^{-\beta U_0}, \tag{2.7}$$

since the volume of the unit cell is the total macroscopic volume V under consideration divided by the number of sites in that volume, M. The void fraction ϵ is the ratio of the volume of the binding pocket Ω_s to the volume of the entire unit cell Ω. Since the particles do not interact, the Canonical partition function $Q(\sigma, M, T)$ for σ molecules adsorbed onto a solid with M sites is then:

$$Q(\sigma, M, T) = \frac{M!}{\sigma!\,(M-\sigma)!} q^\sigma. \tag{2.8}$$

The binomial term $\binom{M}{\sigma}$ counts the number of arrangements of the σ molecules among the M sites. Using Stirling's formula for the $\binom{M}{\sigma}$ term and taking the derivative as in eqn 2.3, we arrive at the chemical potential for the adsorbed phase:

$$\mu_{ads} = -\frac{1}{\beta} \ln\left(\frac{M-\sigma}{\sigma} \frac{1}{\Lambda^3} \epsilon \frac{V}{M} e^{-\beta U_0}\right). \tag{2.9}$$

Finally, we equate the chemical potentials due to the equilibrium condition in eqn 2.2 and solve for the number of occupied sites σ:

$$\sigma = \frac{\beta \epsilon V e^{-\beta U_0} P}{1 + \frac{\beta \epsilon V e^{-\beta U_0}}{M} P}, \tag{2.10}$$

where we used the ideal gas law to relate the density to the pressure via $\rho = P\beta$. Since $\frac{\beta \epsilon V e^{-\beta U_0}}{M}$ is constant with temperature, we define it as a constant K to get the familiar Langmuir isotherm for fractional occupancy:

$$\frac{\sigma}{M} = \frac{KP}{1+KP}. \tag{2.11}$$

Approach 2

In the second approach, we directly find the average number of particles in the adsorbed phase when the particle number is allowed to fluctuate in the Grand-canonical ensemble. The Grand-canonical partition function under the equilibrium condition $\mu = \mu_{ig} = \mu_{ads}$ is:

$$\Xi(\mu, M, T) = \sum_{\sigma=0}^{M} Q(\sigma, M, T) e^{\beta\mu\sigma}. \tag{2.12}$$

We seek to derive the Langmuir isotherm by finding the average σ:

$$\langle\sigma\rangle = \sum_{n=0}^{M} nP(N=n) = \frac{1}{\Xi}\sum_{n=0}^{M} ne^{-\beta n U_0} e^{\beta\mu n} = \frac{\partial \ln \Xi}{\partial(\beta\mu)}. \tag{2.13}$$

First, we use eqn 2.8 to write the Grand Canonical partition function in eqn 2.12 as:

$$\Xi(\mu, M, T) = \sum_{\sigma=0}^{M} \binom{M}{\sigma} \left[\frac{\epsilon V}{M\Lambda^3} e^{-\beta U_0} e^{\beta\mu}\right]^{\sigma}, \tag{2.14}$$

which is a binomial expansion of the function $(1 + \frac{\epsilon V}{M\Lambda^3} e^{-\beta U_0} e^{\beta\mu})^M$. Taking the logarithm and performing the derivative in eqn 2.13:

$$\langle\sigma\rangle = \frac{\epsilon V e^{-\beta U_0} e^{\beta\mu}}{\Lambda^3 + \frac{\epsilon V e^{-\beta U_0}}{M} e^{\beta\mu}}. \tag{2.15}$$

Using the ideal gas law and the expression for the chemical potential of the bulk gas in equilibrium with the solid in eqn 2.6, we write this as:

$$\langle\sigma\rangle = \frac{\beta\epsilon V e^{-\beta U_0} P}{1 + \frac{\beta\epsilon V e^{-\beta U_0}}{M} P}, \tag{2.16}$$

which is equivalent to eqn 2.10 and therefore eqn. 2.11 in Approach 1, and we arrive at the Langmuir isotherm, where ϵ is the void fraction, or porosity, of the material ($\epsilon = |\Omega_s|/\Omega$) and $\beta = 1/(RT)$ with R as the universal gas constant. Eqn 2.16 is the familiar Langmuir isotherm with Henry coefficient $K_H = \beta\epsilon e^{-\beta U_0}$ and Langmuir constant $K = \frac{\beta\epsilon e^{-\beta U_0}}{M}$.

The isothermal deliverable capacity D of a material is the difference in loading at the high pressure P_H of a filled tank and the low pressure P_L of a depleted tank. Using our model, the methane delivered to the engine per volume of model material is:

$$D(U_0; P_L, P_H) = \frac{\beta\epsilon e^{-\beta U_0} P_H}{1 + \frac{\beta\epsilon e^{-\beta U_0}}{M} P_H} - \frac{\beta\epsilon e^{-\beta U_0} P_L}{1 + \frac{\beta\epsilon e^{-\beta U_0}}{M} P_L}, \tag{2.17}$$

which is lucidly a function of the energy of adsorption U_0. The two terms show the prerequisite to obtain a large deliverable capacity is a high uptake at P_H, but uptake at the depletion pressure P_L reduces the deliverable capacity below the high-pressure loading (see Fig 2.1a).

To search for a U_0 that optimizes the deliverable capacity, we solve the eqn $\frac{\partial D}{\partial U_0} = 0$ and find:

$$U_{0,opt} = U_{0,opt}\left(\frac{\epsilon}{M}\right) = RT \ln\left(\frac{\sqrt{P_L P_H}}{RT}\frac{\epsilon}{M}\right). \tag{2.18}$$

The optimum enthalpy of adsorption is then $\Delta H_{opt} = U_{0,opt} - RT$. Bhatia and Myers' derivation from classical thermodynamics yields

$$\Delta H_{BM} = \frac{RT}{2}\log\left(\frac{P_L P_H}{(P^\circ)^2}\right) + T\Delta S^\circ, \tag{2.19}$$

where ΔS° is the standard entropy change upon adsorption with reference pressure $P_0 = 1$ bar [134].

Bhatia and Myers show that, for activated carbons and carbon nanotubes, the entropic term is roughly the same, and they obtain their optimal heat of adsorption by using $\Delta S^\circ = -9.5R$ in eqn 2.19. If we assume $\Delta S^\circ = -9.5R$ holds for all materials, we obtain an optimal heat of adsorption for methane storage of 16.2 kJ/mol at ambient temperature with P_H =65 bar and P_L =5.8 bar. As Bhatia and Myers assume that the isotherms are described with a Langmuir model, $U_{0,opt}$ for our model 0 is consistent with eqn 2.19. In addition, our model links the entropic term to the structural properties in model 0, manifesting precisely how this term differs among materials.

To show that the eqn derived in Ref [122] for the optimum heat of adsorption:[1],

$$\Delta H_{BM} = \frac{RT}{2}\log\left(\frac{P_L P_H}{(P^\circ)^2}\right) + T\Delta S^\circ, \tag{2.20}$$

for a Langmuirian material is consistent with eqn 2.18 for our model material in model 0 by applying eqn 2.20 to the model material in model 0. The term ΔS° is the standard entropy change upon adsorption with reference pressure $P^\circ = 1$ bar. The details of the reference state are in [134], briefly,"The reference states for the standard entropy of adsorption are a hypothetical perfect gas at one atmosphere for the gas phase and a hypothetical gas in the nanopores which obeys Henry's law exactly". In [134], it is show that ΔS° can be written:

$$\Delta S^\circ = \frac{\Delta H}{T} + R\ln\left(\frac{K_H P^\circ}{M}\right), \tag{2.21}$$

where K_H is Henry's constant and M is the saturation loading. For model 0:

$$K_H = \beta\epsilon e^{-\beta U_0} \tag{2.22}$$

since $K_H = \beta\langle e^{-\beta U}\rangle$ [110]. Thus, the standard entropy change of adsorption for model 0 is:

$$\Delta S^\circ = -R + R\ln\left(\beta P^\circ \frac{\epsilon}{M}\right). \tag{2.23}$$

[1]Note that the units are consistent here because we substituted one unit volume in the Langmuir isotherm expressions for σ

If we plug the expression for the standard entropy change upon adsorption in eqn 2.23 into Bhatia and Myers' formula in eqn 2.20, we arrive at:

$$\Delta H_{opt} = U_{0,opt} - RT = RT \ln\left(\frac{\sqrt{P_L P_H}}{RT}\frac{\epsilon}{M}\right) - RT, \tag{2.24}$$

which is equivalent to eqn 2.18 in the main text. This completes our proof of equivalence.

Plugging $U_{0,opt}$ into eqn 2.17, we get the best possible deliverable capacity of our model material:

$$D_{opt}(P_L, P_H) = M\left(\frac{P_H}{\sqrt{P_L P_H} + P_H} - \frac{P_L}{\sqrt{P_L P_H} + P_L}\right), \tag{2.25}$$

equivalent to the maximum delivery obtained by optimizing the Langmuir constant in Ref. [135]. At extreme operating conditions,

$$\lim_{P_H \to \infty} D_{opt}(P_L = 0, P_H) = M, \tag{2.26}$$

and each adsorption site will be fully utilized and deliver a methane molecule to the engine or fuel cell. The two previous equations bound the deliverable capacity below the number of adsorption sites per unit volume– and tuning the heat of adsorption does not change this. The heat of adsorption optimizes the working capacity for a fixed number of sites M. More precisely, it maximizes the *fraction* of sites ($\frac{D}{M}$) *in a given material* that can deliver a methane molecule [135]. The density of adsorption sites M is thus important in obtaining a high deliverable capacity, and from here one can tune the chemistry to obtain the optimal energetics. The analysis in Ref. [122] was aimed to demonstrate that for a fixed maximum loading, there exists an optimal heat of adsorption. This explains the discrepancy between Fig 2.3, where the heat of adsorption is not a good predictor of deliverable capacity, and the rule of thumb in the literature that 18.8 kJ/mol provides an optimum material for methane storage.

First, we derive the equations for fitting a Langmuir model to adsorption data. We have N pressure-loading data pairs (P_i, L_i) for $i = 1, ..., N$ organized into a vector $\mathbf{P}, \mathbf{L} \in \mathbb{R}^N$. We also have the Henry coefficient from molecular simulation. The model prediction $\mathbf{g} \in \mathbb{R}^N$ is a function of the maximum uptake capacity M since we already have the Henry coefficient K_H from simulation:

$$g_i(M) := \frac{K_H P_i}{1 + \frac{K_H}{M} P_i}. \tag{2.27}$$

To find the M in the Langmuir model that fits the data, we solve the nonlinear minimization problem

$$\min_M \frac{1}{2}||\mathbf{g}(M) - \mathbf{L}||^2. \tag{2.28}$$

The M^* that is the minimizer satisfies

$$A(M^*)^T(\mathbf{g}(M^*) - \mathbf{L}), \tag{2.29}$$

where A is the N by 1 Jacobian matrix of $\mathbf{g}$. We solve the minimization problem using the Gauss-Newton method [136]. We start with a guess M_0 and iterate on M using a search direction p_k:

$$M_{k+1} = M_k + p_k. \tag{2.30}$$

The search direction is found by, at each iteration, finding the solution p_k to the linear least squares problem:

$$\min_p ||A(M_k)p - (\mathbf{L} - \mathbf{g}(M_k))||. \tag{2.31}$$

For this, we use the Scipy package in Python. The convergence criterion is that the relative change in the sum of square errors is less than 0.0000001. The Langmuir constant K is then determined by $K = \frac{K_H}{M}$, where M is found by the above fitting routine.

Next, we selected ∼2500 structures identified as Langmuirian by our isotherm fitting routine to analyze in the context of model 0. As a criterion, we use the residual normalized by the predicted maximum loading, also obtained from the Langmuir fitting routine. We selected only the zeolites that have a good Langmuir model fit because (i) model 0 applies to only Langmuirian materials (ii) we want to ensure confidence in the accuracy of identification of M from our high-throughput data fitting routine. In Fig 2.5, we show that this set of selected materials explores a vast range of geometric parameters and heats of adsorption for it to be considered diverse.

In Fig 2.6 (top, blue points), we plot the *fractional* deliverable capacity as a function of the energy of adsorption for these Langmuirian materials. Also for the fractional deliverable capacity, we do not see that the best structures have the predicted optimal heat of adsorption of 16.2 kJ/mol. Further, several structures are endowed with a heat of adsorption of 16.2 kJ/mol, but do not have the maximum fractional deliverable capacity in eqn 2.25. The reason is that the assumption in Ref. [122] that the entropy of adsorption is constant may not hold, implying that the optimal heat of adsorption varies from material to material. This is evident from eqn 2.18 where the structural characteristic $\frac{\epsilon}{M}$ appears.

The standard entropy loss upon adsorption is related to the volume of the adsorption site, $\frac{\epsilon}{M}$ (see eqn 2.23). For the zeolite structures, we estimate the void fraction using an energetic approach, where points in space that have less potential energy than a threshold are deemed occupiable (see eqn 2.46). Fig 2.7a is an M-ϵ scatter plot for the Langmuirian materials in the zeolite database. If the entropic change upon adsorption were the same for all zeolites, all points would collapse to a single line. The color-coding illustrates that we have many materials with different $\frac{\epsilon}{M}$ and thus entropic changes upon adsorption. To see the relevant range of $\frac{\epsilon}{M}$ experienced in the zeolite database, which will dictate the range of the optimal heat of adsorption via eqn 2.18, we take the range of the inverse slopes in Fig 2.7a. Also note that ϵ and M have a weak correlation; if a material has a large void fraction ϵ, it may be a result of having many adsorption sites.

Model 0 predicts that if a material has a different $\frac{\epsilon}{M}$ (related to ΔS°), it has a different optimal heat of adsorption. In Fig 2.7b, we quantify this effect for the range of $\frac{\epsilon}{M}$ we observed in the zeolite database in Fig 2.7a. Fig 2.7b shows that as the adsorption site decreases in volume, corresponding to a larger entropic penalty upon adsorption, the optimal heat of

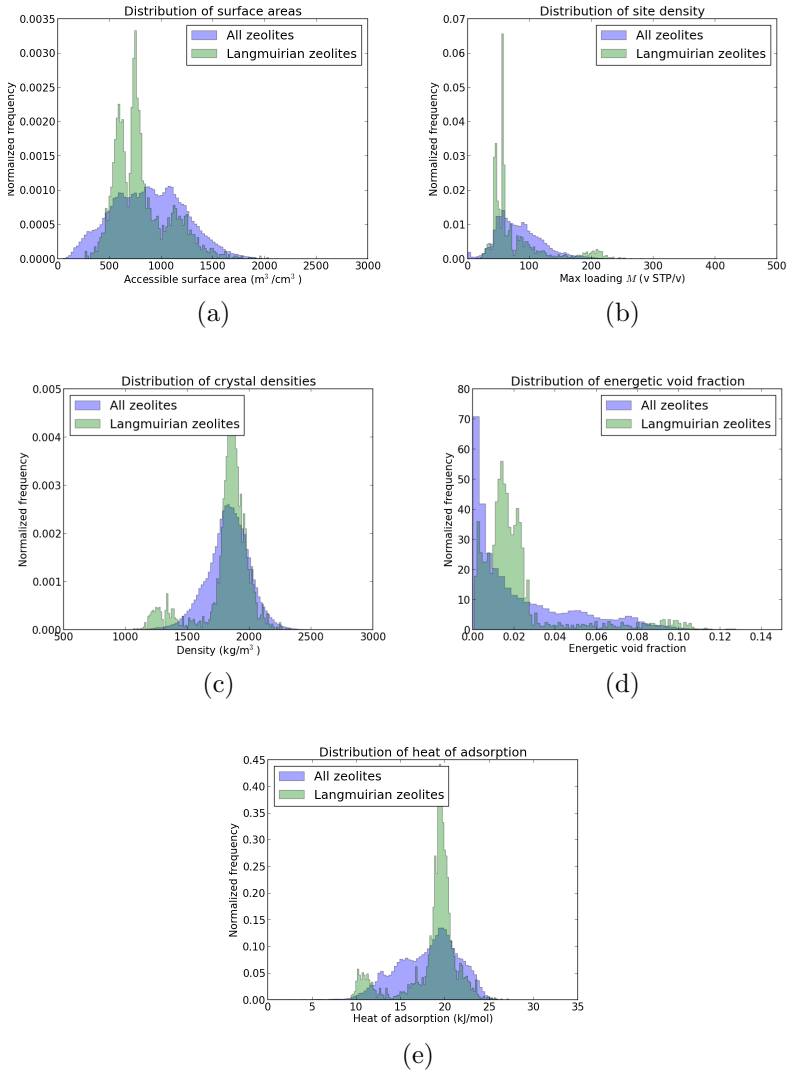


Figure 2.5: Here, we present probability distributions of different properties of the hypothetical zeolite structures in the SLC database [120] analyzed in this work (those with pore size accessible to methane). For comparison, we also show the distribution for the materials that have the best fit to the Langmuir isotherm in our data fitting routine. These materials were analyzed in the context of model 0. We see that these Langmuirian materials explore a range of these properties, and are thus a diverse set for the analysis. Figures reproduced from Ref. [112] with permission from the PCCP Owner Societies.

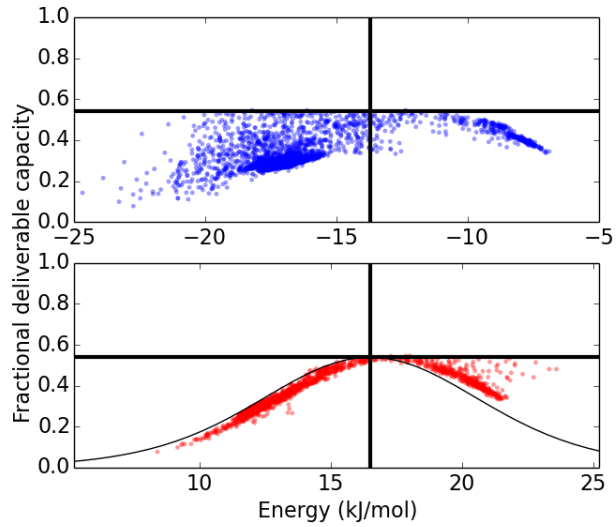


Figure 2.6: Fractional deliverable capacity plots for Langmuirian structures. Horizontal line: maximum theoretical fractional working capacity in eqn 2.25. (Top) The x-axis is the infinite dilution energy of adsorption U_0 computed from simulation. Vertical line: predicted optimal energy of adsorption, -13.7 kJ/mol, using Bhatia and Myers' assumption of $\Delta S^\circ = -9.5R$, corresponding to an enthalpy of adsorption of -16.2 kJ/mol. (Bottom) The x-axis is the effective energy of adsorption U^* in eqn 2.32. Vertical line: $U^*_{opt} = \frac{RT}{2}\log(P_H P_L)$. The black curve is the theoretical universal curve in eqn 2.34, a function of the effective energy of adsorption, on which all materials should fall. Figures reproduced from Ref. [112] with permission from the PCCP Owner Societies.

adsorption increases to compensate for the entropic loss. Depending on the $\frac{\epsilon}{M}$, the optimal heat of adsorption ranges from 11 to 22 kJ/mol.

According to our model, all materials will fall on a universal curve when we subtract off the entropic dependence in $U_{0,opt}$ and define a free energy of adsorption:

$$U^* := U_0 - RT \ln\left(\frac{1}{RT}\frac{\epsilon}{M}\right). \tag{2.32}$$

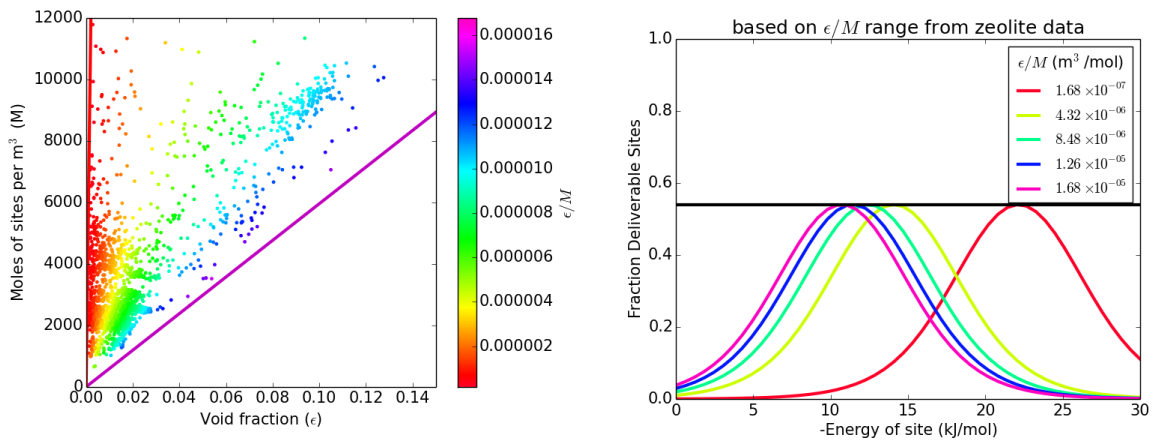
Using the definition of the effective heat of adsorption U^* in eqn 2.32, we substitute:

$$U_0 = U^* + RT \ln\left(\beta\frac{\epsilon}{M}\right), \tag{2.33}$$

into eqn 2.17 to get:

$$\frac{D}{M} = \frac{e^{-\beta U^*}P_2}{1+e^{-\beta U^*}P_2} - \frac{e^{-\beta U^*}P_1}{1+e^{-\beta U^*}P_1}, \tag{2.34}$$

This theoretical curve of $\frac{D}{M}$ depends only upon the effective heat of adsorption, and not the volume of the adsorption pocket in the material. It is plotted in Fig 2.6 as the black line.



(a) (b)

Figure 2.7: (a) The void fraction ϵ and number of sites M are correlated in the Langmuirian materials, but $\frac{\epsilon}{M}$ (inverse slope here) is constrained to a range. Points are color-coded according to their $\frac{\epsilon}{M}$ value. The inverse slopes of the two solid lines (pink, red) at the boundary of the point cloud are the extremum in the set of $\frac{\epsilon}{M}$ in the zeolite database. (b) The optimum energy of adsorption U_0 is a function of $\frac{\epsilon}{M}$. The range of $\frac{\epsilon}{M}$ is taken from the slopes in (a) and the colors of the curves correspond to (a). Figures reproduced from Ref. [112] with permission from the PCCP Owner Societies.

We plot eqn 2.34 as a black curve in Fig 2.6 (bottom). Now, the theoretical optimal effective energy of adsorption is $U^*_{opt} = \frac{RT}{2}\log(P_L P_H)$, constant for every material. To test our model, we plot the fractional deliverable capacity against the effective energy of adsorption in Fig 2.6 (bottom, red points) along with U^*_{opt} as a vertical line. The cloud of blue points collapses onto a near universal curve when we consider the effect of differing $\frac{\epsilon}{M}$ among the materials. The universal curve corresponds to the theoretical predictions. Knowledge of $\frac{\epsilon}{M}$ and the heat of adsorption is thus enough to make a reasonable prediction of the fractional deliverable capacity for Langmuirian materials. The slight shift of the blue points from the theoretical universal curve may be explained from model 3.

Model 1: Incorporating guest-guest interactions

Model 1 builds upon model 0 to include the feature of guest-guest interactions. A reasonable abstraction of adsorbed gas molecules interacting with each other is that each nearby guest enhances the energy of a free adsorption site by energy U_{gg} (subscripts for "guest-guest"). Using a mean-field approximation, the energy landscape is now a function of the number of occupied adsorption sites:

$$U(x) = \begin{cases} U_0 + z\frac{\sigma}{M}U_{gg}, & x \in \Omega_s \\ \infty, & x \notin \Omega_s. \end{cases} \tag{2.35}$$

The energy of each site at infinite dilution, without any guest-guest interactions, is U_0, and z is the coordination number around a site (e.g., $z = 6$ in $\mathbb{R}^3$ for a cubic lattice).

With guest-guest interactions, the total energy U of a system of σ guest molecules adsorbed in the model material is:

$$U = \sigma U_0 + \sum_i \sum_{j>i} z U_{gg} \frac{\sigma}{M} = \sigma U_0 + \frac{1}{2} \sigma z U_{gg} \frac{\sigma}{M}, \tag{2.36}$$

where the first term is the guest-framework interaction energy and the second term is the guest-guest interactions (the factor of $\frac{1}{2}$ avoids over-counting). Now, we write the canonical partition function as:

$$Q(\sigma, M, T) = \frac{M!}{\sigma!\,(M-\sigma)!} \frac{1}{\Lambda^3} \int_\Omega e^{-\beta U} dx_1 dx_2 \cdots dx_\sigma. \tag{2.37}$$

Using eqn 2.36 and the void fraction ϵ, we write evaluate the integral to get:

$$Q(\sigma, M, T) = \frac{M!}{\sigma!\,(M-\sigma)!} \frac{1}{\Lambda^3} \left(\frac{V}{M}\epsilon\right)^\sigma e^{-\beta\sigma U_0 - \beta z U_{gg} \frac{\sigma^2}{2M}}. \tag{2.38}$$

Following Approach 1 above, we apply Stirling's formula for the factorial terms and take the derivative of the logarithm with respect to σ to get an expression for μ_{ads} and then equate this to μ_{ig} to arrive at:

$$\frac{M-\sigma}{\sigma} \frac{\epsilon V}{M\Lambda^3} e^{-\beta U_0} e^{-\beta z g \sigma / M} = (\Lambda^3 \rho)^{-1}. \tag{2.39}$$

In contrast to the Langmuir isotherm derivation, we cannot solve this equation explicitly for σ. However, we gain some intuition by writing this as:

$$\frac{\sigma}{M} = \frac{K e^{-\theta\sigma/M} P}{1 + K e^{-\theta\sigma/M} P}, \tag{2.40}$$

where $\theta := \beta U_{gg} z$ is a guest-guest interaction dependent term and $K := \frac{\beta\epsilon V e^{-\beta U_0}}{M}$ as previously. We can intuit this in the context of a traditional Langmuir isotherm with an effective Langmuir constant $\hat{K} := K e^{-\theta\sigma/M}$, where $\theta := \beta U_{gg} z$ is a guest-guest interaction dependent term and $K := \frac{\beta\epsilon e^{-\beta U_0}}{M}$ as in model 0. That is, now the Langmuir constant is a function of loading. The effect is that, for attractive guest-guest interactions ($U_{gg} < 0$), the effective Langmuir constant increases as loading increases. As the solid loads with adsorbates upon increasing the pressure, the loading lies on a Langmuir isotherm with a larger and larger Langmuir constant. The opposite is true for repulsive guest-guest interactions.

To gain insight into how the guest-guest interactions cause a deviation from a Langmuirian curve, we seek to make an approximation to eqn 2.40 for small θ. We Taylor expand the exponential term in eqn 2.40 as $e^{-\theta\sigma/M} \approx 1 - \theta\sigma/M$ and get a quadratic equation for $\frac{\sigma}{M}$:

$$KP\theta \left(\frac{\sigma}{M}\right)^2 - (1 + KP(1+\theta)) \frac{\sigma}{M} + KP = 0. \tag{2.41}$$

Next, we seek an asymptotic expansion:

$$\frac{\sigma}{M} \sim \sigma_0 + \theta\sigma_1 + ... \tag{2.42}$$

and substitute this into eqn 2.41. Equating terms at order 1, we get the Langmuir isotherm at leading order:

$$\sigma_0 = \frac{KP}{1+KP}, \tag{2.43}$$

and at order θ we get the correction term:

$$\sigma_1 = \sigma_0^2(\sigma_0 - 1). \tag{2.44}$$

We arrive at a correction to a Langmuir isotherm:

$$\frac{\sigma}{M} \sim \frac{KP}{1+KP} + \theta\left(\frac{KP}{1+KP}\right)^2\left(\frac{KP}{1+KP} - 1\right). \tag{2.45}$$

The correction term goes to zero as both $P \to 0$ and $P \to \infty$, consistent with our expectation that guest-guest interactions cause a negligible deviation from a Langmuir isotherm at low and high enough pressures. At low pressures, and hence dilute loading, there are very few nearby guests for interaction. At high enough pressures, there is enough chemical potential driving force such that each site is occupied, regardless of its energy, and the loading is near M. We plot in Fig. 2.9 the numerical solution to eqn 2.40 with $\theta = \pm 0.5$ to represent repulsive and attractive guest-guest interactions, respectively to visualize the deviation from a Langmuir isotherm. The asymptotic approximation to the solution in eqn 2.45, shown with the dashed line, is in excellent agreement with the numerical solution to eqn 2.40.

From the behavior of the correction term in eqn 2.45, guest-guest interactions can benefit the deliverable capacity of a material by substantially increasing loading at P_H, but not the loading at P_L. If the isotherm begins to saturate before P_L, however, guest-guest interactions can degrade the deliverable capacity by increasing uptake at P_L more than P_H. Fig 2.8 shows illustrative sketches.

In Fig 2.10a, we plot the fractional deliverable capacity against the infinite dilution energy of adsorption to show that there still exists a U_0 that optimizes fractional deliverable capacity. For (repulsive) attractive guest guest interactions $\theta < 0$ ($\theta > 0$), the optimum heat of adsorption is below (above) that of a Langmuirian material. Attractive (repulsive) guest-guest interactions yield a larger (smaller) maximum fractional deliverable capacity than for a Langmuirian isotherm in eqn 2.25, which is plotted as a horizontal line in Fig 2.10a, consistent with our interpretation of the correction term in eqn 2.45. Analogous with Fig 2.7b, we show how the optimum U_0 changes from material to material when there are guest-guest interactions in Fig 2.10b by varying the structural characteristic $\frac{\epsilon}{M}$. Note that the maximum theoretical deliverable capacity does not change with $\frac{\epsilon}{M}$. We numerically solve for the maximum theoretical fractional working capacity as a function of guest-guest interaction strength in Fig 2.10c to provide an analog of eqn 2.25. In Fig 2.10d, we show how the optimal

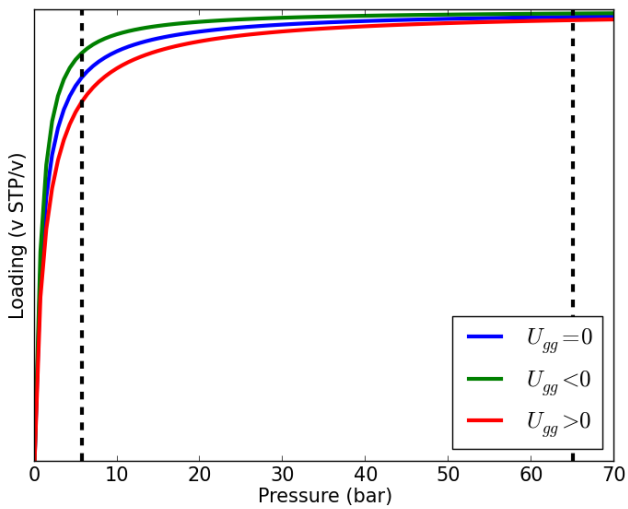


Figure 2.8: An example of how guest-guest interactions that lower the energy of an adsorption pocket can decrease the deliverable capacity by increasing uptake at P_L more than it increases uptake at P_H. Here, repulsive guest-guest interactions would enhance the deliverable capacity. Figures reproduced from Ref. [112] with permission from the PCCP Owner Societies.

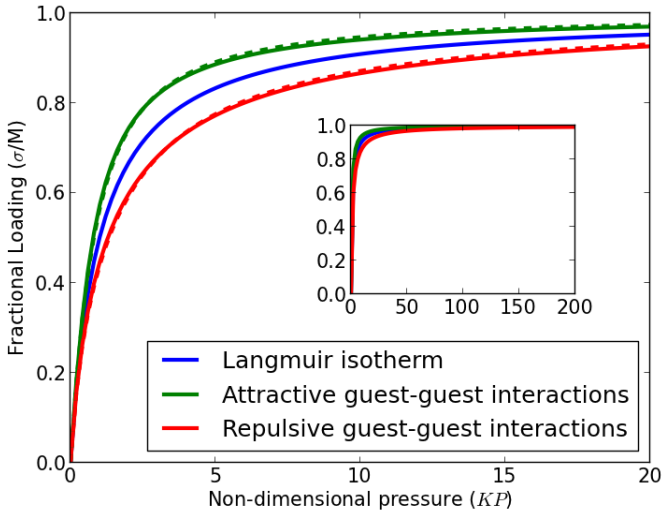


Figure 2.9: Model 1. The influence of guest-guest interactions on the shape of the isotherm in the context of a Langmuirian curve. Depending on where P_H and P_L fall relative to KP, guest-guest attractions or repulsions may increase the deliverable capacity. Dashed lines are asymptotic approximation in eqn 2.45. Figures reproduced from Ref. [112] with permission from the PCCP Owner Societies.

heat of adsorption decreases with attractive guest-guest interactions. Fig 2.10d serves as a reference to, for a given pore volume (ϵ/M) and pocket-to-pocket distances and coordination number ($\rightarrow zU_{gg}$), determine the optimal heat of adsorption. Each vertical line is a multiple of the Lennard-Jones interaction parameter for methane using the TraPPE force field [106] to provide a relevant reference for guest-guest interaction energies. The void fraction values ϵ were obtained from the Widom insertion method. The total number of random methane insertion configurations is set to be $N_{\text{Widom}} = 35.84$ million. Out of N_{Widom}, we count the number of test points $N_{\text{Low energy}}$ where the methane-zeolite framework interaction energy is smaller than $\langle U \rangle + RT$, with $\langle U \rangle$ being the Boltzmann-weighted average energy of the system. The void fraction is then

$$\epsilon = \frac{N_{\text{Low energy}}}{N_{\text{Widom}}}. \tag{2.46}$$

We can understand the reason that the optimal heat of adsorption for attractive guest-guest interactions is below that of a Langmuirian curve. For a Langmuirian curve, a high $-U_0$ ensures that loading is close to M at P_H, but there is a tradeoff because a large $-U_0$ also dictates a high loading at P_L, eroding the deliverable capacity. This is the intuition behind why there is an optimum heat of adsorption. The tradeoff shifts towards having a lower $-U_0$ when guest-guest interactions are considered because these will have the greatest influence at P_H (if P_L is sufficiently low, or if K is sufficiently high). Thus, we can lower the value of $-U_0$ to reduce the uptake at P_L and rely on the guest-guest interactions to provide the extra energy to recruit adsorbate molecules at P_H.

As an example of beneficial guest-guest attractions, consider the zeolite SBN, which has been identified as a potentially attractive material for methane separations [137]. In Fig 2.11a, we see that the energy of adsorption of SBN increases linearly with loading, suggesting that attractive guest-guest interactions are present. Fig 2.11b is a contour plot of the energy landscape of SBN. The binding pockets are ≈ 4 and 4.6 $\AA$ distance from one another [137], close to the optimal distance for maximal methane-methane potential overlap of 4.15 $\AA$ based on the Lennard Jones parameter for methane (taken from [127, 128]). From a visual inspection of Fig 2.11b, we observe 4 binding sites in a unit cell of volume 619.36 $\AA^3$, yielding $M = 240$ v STP/v for SBN. Specified with knowledge of the Henry coefficient and M, we plot a Langmuirian SBN isotherm (Model 1, $U_{gg} = 0$) in Fig 2.11c– a thought experiment for when methane-methane interactions are not present. To corroborate this prediction, we 'turn off' guest-guest interactions by letting the Lennard-Jones parameter ϵ for CH_4-CH_4 interactions go to zero and running GCMC simulations on SBN. We do not set $\epsilon_{CH4-CH4} = 0$ because then the feature of excluded volume would be lost; by letting $\epsilon_{CH4-CH4}$ be close to zero, the potential still goes to infinity with two overlapping methane molecules. We plot the loading at 1 bar and 35 bar as a function of guest-guest interaction strength in Fig 2.11d to show that (i) the loading converges to a particular loading as $\epsilon_{CH4-CH4} \rightarrow 0$ and (ii) guest-guest interactions enhance the loading at 35 bar in comparison to if the SBN isotherm were Langmuirian. Using the expression for $U(\sigma)$ from linear regression in Fig 2.11a, we then

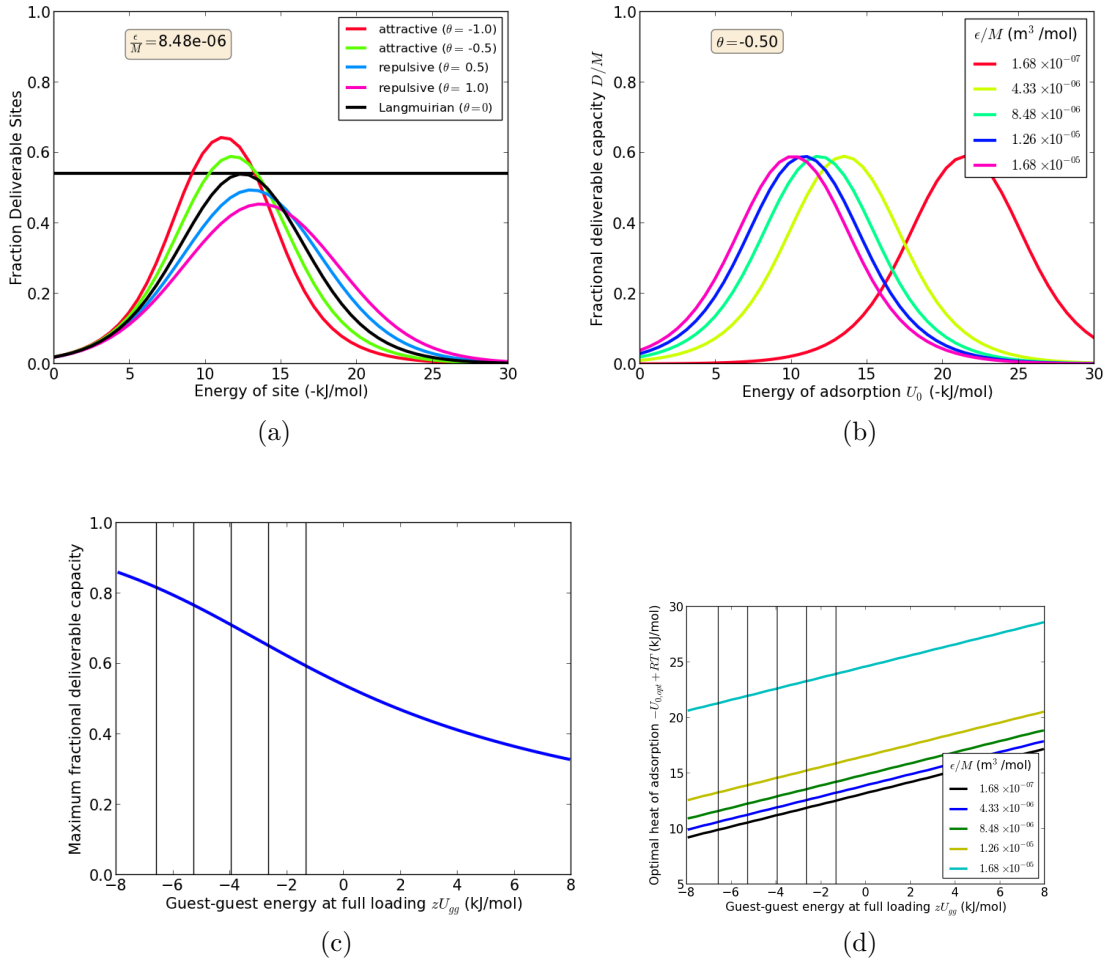


Figure 2.10: Model 1. (a) For a given material (i.e. $\frac{\epsilon}{M}$), attractive guest-guest interactions enable a higher maximum fractional deliverable capacity. The horizontal line is the maximum deliverable capacity in eqn 2.25 for a Langmuirian curve. The optimum energy of the site $-U_0$ decreases with more attractive guest-guest interactions (θ becoming negative). (b) Again, the optimal infinite dilution heat of adsorption depends on the material (through $\frac{\epsilon}{M}$) with fixed guest-guest interactions. The maximum possible fractional deliverable capacity does not depend upon $\frac{\epsilon}{M}$. Colors of curves based on Fig 2.7a. (c) The maximum fractional deliverable capacity as a function of the guest-guest interaction energy (same curve for all $\frac{\epsilon}{M}$). (d) For a fixed $\frac{\epsilon}{M}$, the optimal heat of adsorption decreases with attractive guest-guest interactions. Vertical lines in d) and e) are multiples of the methane-methane Lennard-Jones parameter in Refs. [127, 128], 1.44 kJ/mol. Figures reproduced from Ref. [112] with permission from the PCCP Owner Societies.

predict the actual isotherm of SBN from eqn 2.40 (Model 1), which successfully matches with the simulated loading of SBN in Fig 2.11c.

Model 2 and the entropy-enthalpy correlation

Model 2 presents a different abstraction of a material than in Fig 2.4 that captures the coupling between the pore size and the energy of adsorption in the pore. In model 0, we take the energy of adsorption U_0 to be independent of the structural characteristics ϵ and M of the material. Experimentally, however, a correlation between pore size and heat of adsorption is observed [138, 139]. Ref. [125] adjusted Bhatia and Myers' formula for the optimal heat of adsorption in the context of hydrogen storage by taking into account the positive correlation found between enthalpy and entropy changes upon adsorption. The volume of the adsorption site in model 0, $\frac{\epsilon}{M}$, determines the standard entropy change upon adsorption (eqn 2.23).

For the zeolite database, we calculated the standard entropy change upon adsorption from the computed Henry coefficient and estimated maximum loading and plot it against the computed heat of adsorption in Fig 2.13a. We used equation 2.21 to calculate the standard entropy change upon adsorption. The enthalpy change upon adsorption is $\Delta H = \langle U \rangle - RT$. [To help keep track of the signs: $\langle U \rangle$ will be negative for a good material.] The Henry coefficient is computed by $\langle \beta e^{-\beta U} \rangle$. Finally, we get the maximum loading following the method outlined in [134], where we first compute the Helium pore volume V_p:

$$V_p = \int e^{-\beta U_{He-s}} dV, \tag{2.47}$$

from Monte Carlo simulation. The term U_{He-s} is the solid-helium interaction energy. In the effective potential in eqn 2.49, we use the helium-methane Lennard-Jones parameters to model helium adsorption. We used a surface density of carbon atoms $\alpha = 0.13\ \mathring{A}^{-2}$. From the Helium pore volume, we estimate the maximum loading by:

$$M = V_p \rho_{CH4,l}, \tag{2.48}$$

where $\rho_{CH4,l} = 422.62$ kg/m^3 is the liquid density of methane.

In Fig 2.13a, we indeed observe a positive correlation between $|\Delta H|$ and $|\Delta S^\circ|$ up until $|\Delta S^\circ|$ becomes large; at this point, the correlation quickly drops off, an effect not observed in Ref. [125]. Two properties of the plot in Fig 2.13a underscore our aforementioned conclusions. First, it is clear from the variance in ΔS° that it cannot be assumed to be the same for all materials– even within the class of zeolites. Second, Fig 2.13a is color-coded according to the deliverable capacity; we see that a *range* of heats of adsorption, between 8 and 23 kJ/mol, yields the zeolites with the highest deliverable capacities, depending upon the associated ΔS°. Note that several poor-performing structures are invisible, buried beneath the high performance structures, highlighting that an optimal heat of adsorption does not guarantee a high deliverable capacity.

The work in [125] intuitively pointed out that there is a bound to the change in entropy that a guest molecule can experience in the framework. In accordance, we see a sharp bound

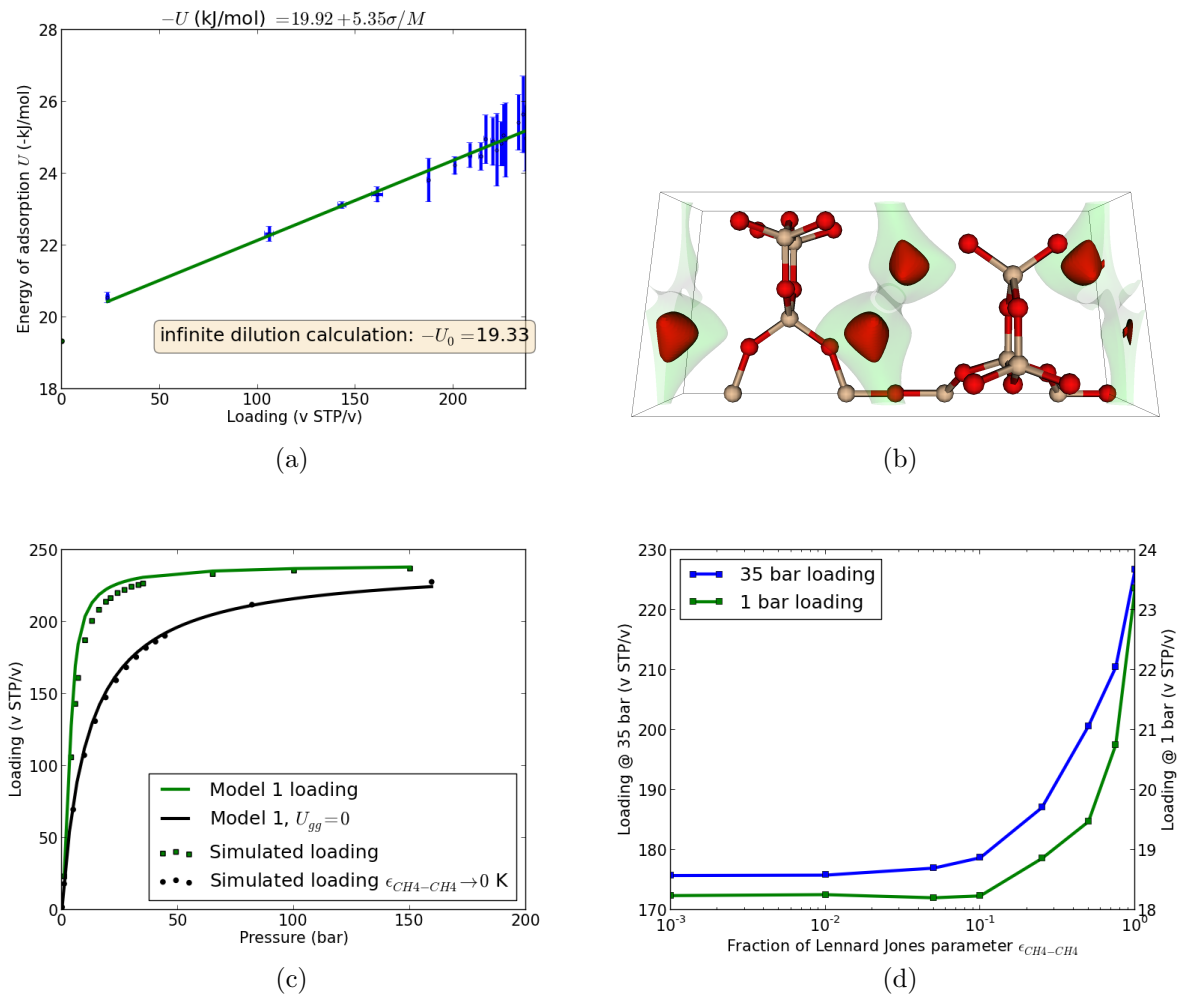
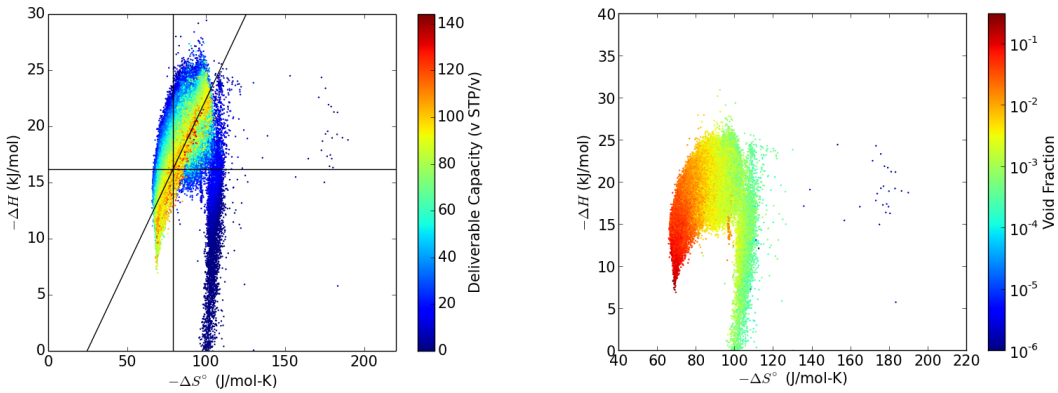


Figure 2.11: SBN has attractive guest-guest interactions. (a) Simulated heat of adsorption increases linearly with loading. We fit to eqn 2.35. (b) Energy landscape of SBN shows binding sites that are 4 and 4.6 $\AA$ [137] distance from each other, close to the optimal distance of 4.15 $\AA$ that maximizes potential overlap for CH_4-CH_4 interactions. Red contour: -2000 K, green contour: 4470 K = $15T$. (c) SBN simulated loading as a function of pressure (green points). Black line is Langmuir isotherm of SBN using $K = \frac{K_H}{M}$, where K_H is from molecular simulation and M is from visually counting adsorption sites per unit cell in (b). Black circles are GCMC simulations with guest-guest interactions turned off via reducing $\epsilon_{CH4-CH4}$ by a factor of 10^{-3}. Green curve is predicted SBN isotherm using K, M, and zU_{gg} from the fit in Fig 2.11a. (d) Loading in SBN decreases and converges to a value as guest-guest interactions are neglected, simulated by reducing the CH_4-CH_4 Lennard-Jones interaction parameter. Figures reproduced from Ref. [112] with permission from the PCCP Owner Societies.

in Fig 2.13a. However, there are some outliers that we omitted, and we show them here in Fig 2.12a. Fig 2.12b shows that these outliers have a very low void fraction and thus experience the greatest change in entropy upon adsorption (via eqn 2.23).



(a) (b)

Figure 2.12: Outliers in the dataset, as they experience entropy changes > 120 J/mol-K. These structures have a very small void fraction, consistent with eqn 2.23. Figures reproduced from Ref. [112] with permission from the PCCP Owner Societies.

We seek to explain the shape of the $|\Delta H|$-$|\Delta S^\circ|$ correlation observed in Fig 2.13a, by posing a simple model that captures the coupling between the size of the adsorption site and the energy of the site, both of which are dictated by the placement of framework atoms. We assume that a gas molecule is adsorbed in a spherical cavity of radius R (Fig 2.13b for the unit cell). Inside this cavity, the gas molecules interact with a shell of uniformly distributed oxygen atoms that defines the adsorbent. Given that the gas molecule interacts with the oxygen atoms through a Lennard-Jones potential, an effective potential $U_{eff}(d;R)$ that is a function of the guest's distance d from the shell wall was derived in Ref. [140] by smearing (performing a surface integration [141]) the framework atoms over the spherical cage:

$$U_{eff}(d;R) = 4\alpha\epsilon\frac{2\pi R}{R-d}\left[\frac{\sigma^{12}}{10}\left(\frac{1}{d^{10}} - \frac{1}{(2R-d)^{10}}\right) - \frac{\sigma^6}{4}\left(\frac{1}{d^4} - \frac{1}{(2R-d)^4}\right)\right]. \tag{2.49}$$

We use the oxygen-methane Lennard Jones parameters [127, 128] for ϵ and σ in eqn 2.49. The parameter α is the surface density of the oxygen atoms in the spherical shell, which we take as constant 0.13 $\mathring{A}^{-2}$ obtained by fitting to the maximum enthalpy change in Fig 2.13a.

By varying the radius of the spherical shell, we simultaneously change the standard entropy loss upon adsorption (the "size" of the site) and the enthlapy upon adsorption (through the energy potential). Fig 2.13c shows how the potential inside the spherical shell

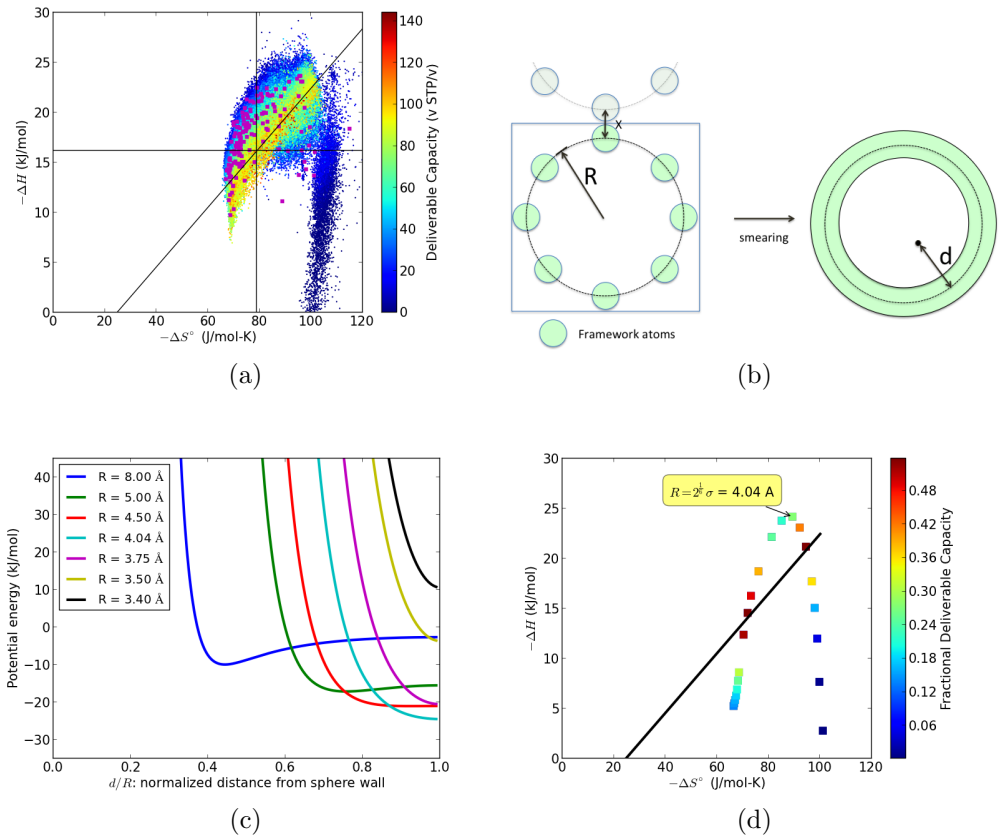


Figure 2.13: Entropy-enthalpy correlation. (a) Correlation between enthalpy and entropy change upon adsorption calculated for the zeolite database. Points, which represent structures, are color-coded according to the deliverable capacity. The black diagonal is the optimal heat of adsorption a function of ΔS° in eqn 2.19; horizontal line is the blanket 16.2 kJ/mol optimal heat of adsorption for all materials from using the assumption $\Delta S^\circ = -9.5R$ (vertical line) in Ref. [122], that the standard entropy change upon adsorption is the same for all materials. The structures with the highest deliverable capacity were plotted last, so there are many poor performing structures buried by the high performance structures. The IZA zeolite structures are plotted as pink points. (b) A 2D representation of our toy model to understand the entropy-enthalpy correlation. A spherical cage of framework atoms with radius R creates the adsorption pocket inside. The square is the unit cell. The potential $U_{eff}(d; R)$ smears the framework atoms over the spherical shell. The variable d is the closest distance from a point to the shell surface. The variable x is the distance between two spherical shells, which we set to be the optimal oxygen-oxygen distance $2^{1/6}\sigma_{O-O}$. (c) The potential $U_{eff}(d; R)$ plotted for a series of sphere radii R. (d) The entropy-enthalpy correlation from our toy model mimics that of the zeolite database in (a). Each point corresponds to a simulation of a toy model of a different radius R. The radius R decreases monotonically with the entropic loss. The deepest potential well in (c) occurs for $R = \sigma 2^{\frac{1}{6}}$ and yields the largest $|\Delta H|$. The points are color coded according to their fractional deliverable capacity. The optimal heat of adsorption as a function of the standard entropy change in eqn 2.19 is plotted as the black line. Figures reproduced from Ref. [112] with permission from the PCCP Owner Societies.

changes as a function of the radius R. For large radii, adsorbed methane molecules will prefer to reside near the sphere surface, where the potential is deepest. The potential gets deeper as the radius decreases, as methane gets the opportunity to interact with a greater fraction of the framework atoms in the cage. Eventually, the potential well is deepest in the center of the sphere, the point where methane maximizes its interactions with all framework atoms in the cage. As the radius gets too small, however, repulsive forces take effect, and the potential becomes positive. This repulsive effect is the source of the dropoff in Fig 2.13a.

We now assume that our material consists of these spherical shells packed in a cubic lattice at a shell-to-shell distance $x = 2^{1/6}\sigma_{O-O}$. We compute for spheres of different radii R the Henry coefficient and ensemble-average energy of adsorption using Widom insertions. To estimate the saturation loading, we compute the Helium pore volume and use the density of liquid methane. See eqns 2.47 and 2.48.

Fig 2.13b shows the coupling between the heat of adsorption and the entropy of adsorption for our model. Our model nicely captures the shape of the corresponding data for the zeolites in Fig 2.13a. We observe a maximum in the heat of adsorption, which corresponds to a pore diameter of $R = R_c := 2^{1/6}\sigma$ in which we have the maximum interactions of methane with the spherical shell. If we increase the pore diameter from R_c, the heat of adsorption decreases until we reach such a large pore that the effect of the pore walls is negligible. If we, on the other hand, decrease the pore diameter from R_c, the interactions are dominated by the repulsive part of the potential, and the magnitude of the enthalpy of adsorption quickly decreases and will eventually become positive. The entropy loss upon adsorption increases monotonically with decreasing pore size. The entropy loss in our model has then reached its maximum as the methane molecule is frozen between the wall of the sphere.

In our model, we assumed a cubic lattice packing of spherical shells with distance $x = 2^{1/6}\sigma_{O-O}$. However, the packing can vary, which shifts the ΔH-ΔS° curve. In addition, the shape of the cavity and surface density of atoms can change, which results in different heats of adsorptions. The net effect of these changes will be a collection of curves that each have a similar shape, explaining the thick band we observe in Fig 2.13a. The color coding shows the fractional deliverable working capacity for this material. Again, the cage radii with the highest fractional deliverable capacity fall on the line dictated by eqn 2.19.

The IZA zeolite structures are plotted in the $(\Delta S^\circ, \Delta H)$ plane along with the hypothetical zeolite structures in Fig 2.13a. The IZA structures explore a similarly sizable range of ΔS°. As supplementary information, we provide an interactive plot of the predicted deliverable capacities of the IZA zeolites in both the 65 - 5.8 bar and 35 - 1 bar pressure ranges.

Model 3: Spatially inhomogeneous adsorption site

In models 0 and 1, the energy of a guest molecule inside the adsorption site is assumed to be spatially homogeneous. As a consequence, the energy of adsorption in models 0 and 1 does not depend on temperature. If we increase the temperature in a system with a real material, methane will explore higher energy configurations and, as a consequence, the heat of adsorption will increase (become less negative). To include this effect, we consider a

spherical adsorption site with a radially symmetric harmonic energy landscape ($r = 0$ is the center of the site with radius r_0 in Fig 2.4):

$$U(r) = \begin{cases} -U_d + \frac{1}{2}kr^2, & r < r_0 \\ \infty, & r \geq r_0. \end{cases} \tag{2.50}$$

As $T \to 0$, entropy becomes irrelevant and the adsorbed particle resides at the center of the adsorption site $r = 0$. As the temperature increases from zero, entropic considerations take effect, and the guest molecule explores a greater region of the adsorption site, and hence a greater region of the bottom of the harmonic potential in eqn 2.50. The ensemble average energy of adsorption with the energy landscape in eqn 2.50 is:

$$-U_d + \frac{3}{2}RT. \tag{2.51}$$

The increase with temperature reflects the greater exploration of the trough in the harmonic potential at higher temperatures in the ensemble. If the harmonic potential is steep enough (i.e., sufficiently strong such that the methane remains close to the minimum), we can derive from statistical mechanics an approximate isotherm of the model material in Fig 2.4 with the energy landscape in eqn 2.50:

$$\frac{\sigma}{M} = \frac{\beta\Psi e^{\beta U_d} P}{1 + \beta\Psi e^{\beta U_d} P}, \tag{2.52}$$

where

$$\Psi := \sqrt{2}\left(\frac{\pi RT}{k}\right)^{\frac{3}{2}}. \tag{2.53}$$

Comparing eqn 2.52 with the Langmuirian isotherm in eqn 2.16, Ψ is the analogy of $\frac{\epsilon}{M}$– the volume of an adsorption site– in model 0. As k increases, i.e., as the harmonic potential well becomes steeper, the effective volume per site Ψ decreases. As the temperature increases, the effective volume per site Ψ increases since an adsorbed guest particle can explore more of the trough for a fixed k. This derivation gives a natural definition for the void fraction of a material with a harmonic potential and shows that the effective volume per adsorption site scales with $T^{\frac{3}{2}}$.

Defining the deliverable capacity with the isotherm in eqn 2.52 and optimizing with respect to U_d results in:

$$U_{d,opt} = U_{d,opt}(\Psi) = RT \ln\left(\frac{\sqrt{P_L P_H}}{(RT)}\psi\right). \tag{2.54}$$

essentially the same as eqn 2.18. In real materials, whose energy landscape is a more complicated function of space than the prescription in eqn 2.1, the void fraction is ill-defined without a threshold energy beyond which the framework is deemed as unoccupiable by a guest molecule. Still, the result in eqn 2.52 suggests that real materials may be analyzed in

the context of models 0 and 1, despite the "binding site" region Ω_s depicted in Fig 2.4 being ambiguous in real materials. i.e., the energy landscape of a real material can be mapped to a void fraction ϵ to be used in eqn 2.18. For a Harmonic potential, the mapping is $\frac{\epsilon}{M} \to \Psi$. We found that an energy threshold of $\langle U \rangle + RT$ to determine which points are occupiable for defining the void fraction for Fig 2.6 provided a reasonable mapping of for the Langmuirian zeolites with an energy landscape defined by a pair-wise Lennard-Jones model.

Here we derive the isotherm in eqn 2.52. The single particle partition function for a single site is:

$$q = \frac{1}{\Lambda^3} \int_\Omega e^{\beta U_d - \frac{1}{2}k\beta r^2} dx = \frac{e^{\beta U_d} 4\pi}{\Lambda^3} \int_0^{r_0} r^2 e^{-\frac{1}{2}k\beta r^2} dr. \tag{2.55}$$

Next, we rescale the r in the integral above by $\hat{r} := \frac{r}{r_0}$ to get a new integral:

$$\Lambda^3 q = \int_0^1 r_0^3 \hat{r}^2 e^{-\frac{1}{2}k\beta r_0^2 \hat{r}^2} d\hat{r}. \tag{2.56}$$

Our approximation is as follows. If $\frac{1}{2}k\beta r_0^2 >> 1$, the above integration in $\hat{r}$ can be approximated by the same integral from 0 to ∞ since $\int_1^\infty (\cdots) d\hat{r} \approx 0$. With this approximation, we get:

$$q = \frac{1}{\Lambda^3} 4\pi \sqrt{\frac{\pi}{2}} \left(\frac{RT}{k}\right)^{\frac{3}{2}} e^{\beta U_d}. \tag{2.57}$$

In words, the approximation is reasonable if the harmonic potential in eqn 2.50 is steep in the scaled coordinate $\hat{r}$. Using this single-particle partition function for constructing the Grand Canonical paritition function, the logic starting from eqn 2.14 follows for deriving $\langle \sigma \rangle$ for the harmonic potential, and we arrive at eqn 2.52.

The ensemble average heat of adsorption in expression 2.51 is calculated by the expression:

$$\frac{\int_\Omega U e^{-\beta U} dx}{\int_\Omega e^{-\beta U} dx}, \tag{2.58}$$

and invoking the same approximation above under the assumption $\frac{1}{2}k\beta r_0^2 >> 1$.

2.4 Analysis of top performing zeolite structures

It is interesting to compare the results of our models with the characteristics of the top performing structures. Here, we analyze the two zeolites with the largest 35 to 1 bar deliverable capacity, PCOD8124791 and PCOD8330975. PCOD8124791 is topologically identical to the IZA zeolite SBN, however with perturbed oxygen atom positions due to the relaxation step in hypothetical zeolite generation [119]. These two structures have outlying deliverable capacities because a combination of factors must arise to, under the constraint of using silica, provide an optimum material. Namely, to resonate with models 0 and 1, the optimal material has a high density of sites, a moderate heat of adsorption, and adsorption pockets positioned for guest-guest attractions.

The two top performing materials PCOD8124791 and PCOD8330975 have predicted 35 to 1 bar deliverable capacities of 200 v STP/v and 172 v STP/v, respectively. Satisfying the first prerequisite, they both display a high density of binding sites, obtained by counting the adsorption pockets per unit cell from Fig 2.15, corresponding to M =245 v STP/v and M =221 v STP/v, respectively. PCOD8330975 has a 10% lower density of sites due to the cubic lattice-like arrangement compared to the more efficient site packing in PCOD8124791, resulting in reduced performance. This corresponds to an 82% and 78%, respectively, utilization of sites to deliver a methane molecule– significantly higher than the 70% theoretical maximum utilization by eqn 2.25 for a Langmuirian material. The pairwise distances between sites in zeolites PCOD8124791 and PCOD8330975 are near the optimal Lennard-Jones distance for maximizing potential overlap (∼4.15 $\AA$, see Fig 2.15), leading to attractive guest-guest interactions and enabling these materials to utilize a greater fraction of their sites to deliver a molecule to the tank/fuel cell. Indeed, the simulated heats of adsorption of PCOD8124791 and PCOD8330975 increase with loading (Fig 2.14). This resonates with the story of Fig 2.10c, where guest-guest attractions were shown to increase the fraction of sites that are utilized to deliver a methane molecule to the engine beyond that possible with a Langmuirian material. We found that zeolites perform better in the 35 to 1 bar pressure range operation than in the 65 to 5.8 bar operation because their isotherms are generally steep at low pressures (1-5.8 bar).

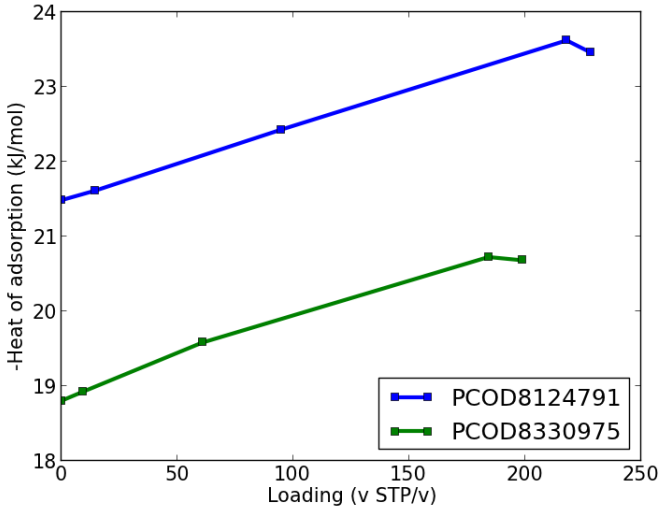
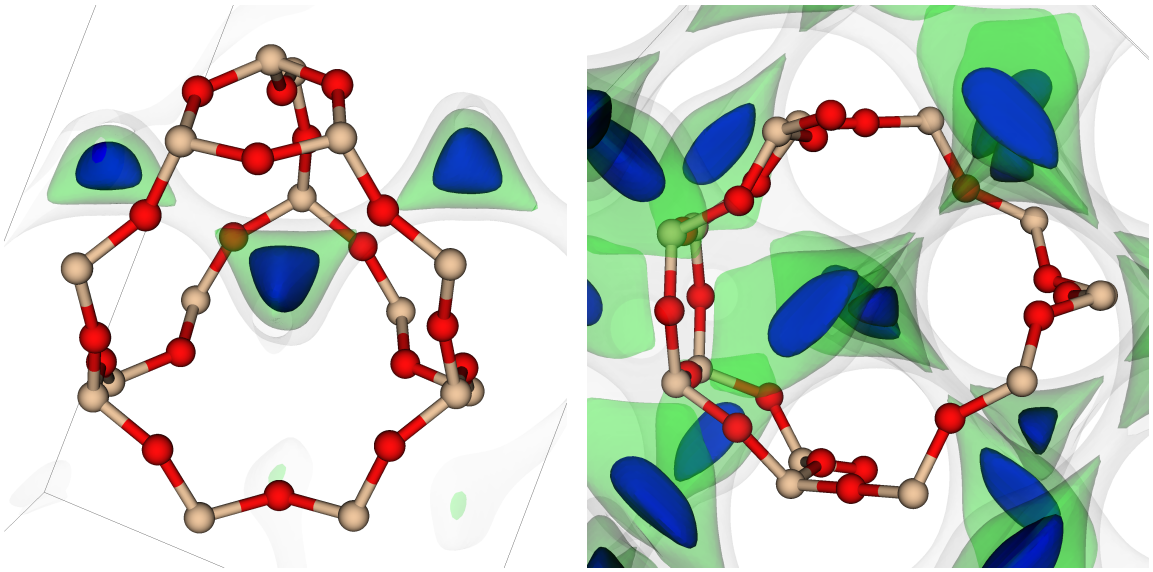


Figure 2.14: The top two performing 35 to 1 bar deliverable capacity zoelites have heats of adsorption that increase with loading, suggesting that attractive guest-guest interactions are present. Figures reproduced from Ref. [112] with permission from the PCCP Owner Societies.

For attractive guest-guest attractions to benefit the deliverable capacity, the infinite dilution heat of adsorption must be at the respective optimum to ensure that the isotherm

saturates at the correct point with respect to P_L and P_H. Many other structures have heats of adsorption that increase with loading and thus attractive guest-guest interactions, but are not outlying performers. That is, P_L must be low enough in the isotherm so that the guest-guest attractions do not enhance uptake at P_L and degrade the deliverable capacity (Fig 2.8).



(a) (b)

Figure 2.15: Visualization of the structure unit cell and energy landscape of the top performing zeolites. Contours in the potential surfaces correspond to energies of -2000 K (blue, the strongest binding regions), 0 K (green), and 4470 K (gray, representing the channel system defined approximately as $15T$). Silicon atoms are shown in tan; oxygen red. a) first place deliverable capacity material, PCOD8124791 b) second place deliverable capacity material, PCOD8330975. Figures reproduced from Ref. [112] with permission from the PCCP Owner Societies.

The methane binding sites in PCOD8124791 and PCOD8330975 arise from similar local atomic arrangments in their structures. Biding sites in PCOD8124791 are comprised of three 8-Si rings (see Fig 2.15a), which form the apertures in the channel system; these are capped by a 3-Si ring (top in figure) and a 9-Si ring (bottom), each of which is too small to form an aperture for methane. In PCOD8330975 (see Fig 2.15b), we see the same 8-Si aperture shape, but the ring systems differ: three 4-Si rings cap the binding site (left), while the apertures comprise 10-Si rings (due to the two additional Si seen at figure right which do not constitute part of the 8-Si shape). Accordingly, the atoms contributing to the creation of the binding site in these two materials are strikingly similar: 15 Si and 18 O, and 14 Si and 19 O, respectively (note these atoms are shared between multiple binding sites). In previous work [142, 143], we identified these 8-Si shapes for achieving favorable binding sites

in zeolites for small hydrocarbon and CO_2 guest molecules; here we show that these features are also yielded optimal methane binding pockets in zeolites.

Some materials do not have defined sites such as in model 0, but rather large, open surface areas of framework atoms that attract guest molecules. Bhatia and Myers showed that the optimal material will not have heterogeneous sites, which is intuitive because the "it is the optimal site that is most productive" [122]. We see this in the top-performing zeolites PCOD8124791 and PCOD8330975, whose adsorption pockets are relatively homogeneous, and they do not have large cavities (which will certainly have regions of suboptimal energies) that provide open surface areas for adsorption.

There are no known synthesis routes to obtain any of the given hypothetical zeolites that we identified as top performers. However, advancements in the design of organic structure-directing agents [144] may change this situation. For example, a recent work by the Deem group [145] demonstrates a prototype approach for such a design.

2.5 Discussion

In this work, we have extended the ideas developed by Bhatia and Myers [122] on an optimum heat of adsorption for methane storage in nanoporous materials using molecular simulation results from the zeolite database and thermodynamic models as guides. In the context of a Langmuir isotherm, there exists a heat of adsorption that maximizes the *fraction* of adsorption sites that will deliver a methane molecule to the engine. An optimal heat of adsorption does not, however, guarantee that a material will yield a large deliverable capacity (Fig 2.3) because this depends on the density of adsorption sites. Moreover, our analysis has shown that this optimal heat of adsorption is material-dependent because of differing entropy changes upon adsorption, and hence there does not exist a blanket optimal heat of adsorption target for all materials.

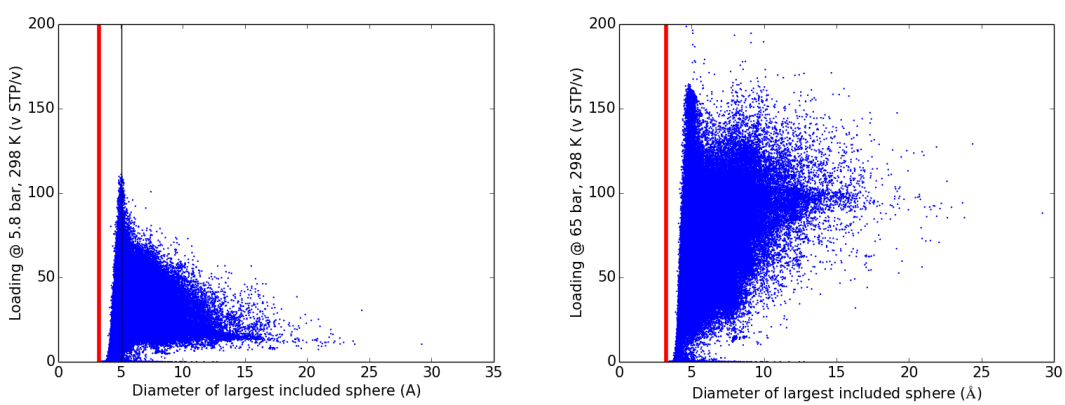
Our sequence of models manifests features that endow a material with a large deliverable capacity. The first prerequisite is that a material must have a high density of sites, as this is an upper-bound for the deliverable capacity. Given a density of sites, designing an optimal material reduces to optimizing the fraction of sites utilized to deliver a guest. If the number of adsorption sites per volume multiplied by the maximum deliverable fraction of sites (eqn. 2.25 or Fig 2.10c)) is less than the target deliverable capacity, the material is hopeless. To maximize the fractional deliverable capacity, one would ideally substitute chemical moieties to effectively tune the infinite dilution heat of adsorption, using our model as a guide (eqn 2.18 or Fig 2.10c (green line for relevant $\frac{\epsilon}{M}$). One must consider that, if differently sized chemical moieties are substituted, this changes the entropic loss upon adsorption (via $\frac{\epsilon}{M}$), and thus the optimal heat of adsorption shifts from what it was at the starting point.

Wilmer et al. [121] found that, in analyzing a large set of hypothetical MOFs, the materials with the highest deliverable capacities had a void fraction of ∼0.8. The zeolites analyzed in this work explore void fractions (calculated from Zeo++ with a methane probe of 1.625 $\mathring{A}$ radius) in the range $\sim [0, 0.5]$. Because materials with more adsorption sites per

volume will generally have higher void fractions (correlation in Fig 2.7a), this suggests that zeolites have a smaller density of sites, rendering them, on average, poorer materials for gas storage. Zeolite are relatively dense nanoporous materials, thus their gravimetric deliverable capacities are generally poorer compared to the MOFs considered in Ref. [121].

The heat of adsorption is generally a function of the loading. Attractive guest-guest interactions cause the energy of adsorption to become more negative with loading. The distance between two methane molecules that maximizes attractive energies is ~ 4.15 $\AA$ (based on both [127, 128] and TraPPE [106]). Model 1 shows that, if the infinite dilution heat of adsorption is at an optimal value (see Fig 2.10a), attractive guest-guest interactions can substantially enhance the fractional deliverable capacity beyond that of a Langmuirian material. Ideally, one would like to have control on the positioning of the adsorption pockets such that they are $\sim$4.15 $\AA$ apart. The deliverable capacity of the most promising zeolites, SBN, PCOD8124791, and PCOD8330975 among them, is enhanced by having their pockets positioned at the optimal distance for attractive guest-guest interactions. Note, however, that if the infinite dilution heat of adsorption is too high, guest-guest *repulsions* will increase the fractional deliverable capacity (Fig 2.10a) by reducing uptake at P_L.

The void fraction can help distinguish the deliverable capacity of materials with the same heat of adsorption in Fig 2.3 precisely because of the ϵ-M correlation in Fig 2.7a: more porous materials will have a higher density of sites. The story of Fig 2.3 is that, to obtain a material with a high deliverable capacity, first maximize the density of sites (related to void fraction), and then optimize the heat of adsorption.



(a) (b)

Figure 2.16: Loading at 5.8 bar (a) and 65 bar (b) plotted against the largest included sphere diameter. Red vertical line is hard-sphere diameter of methane. Black vertical lines denote (a) 5.1 $\AA$ (b) 4.8 $\AA$. A comparison between (a) and (b) reveals that design heuristics are sensitive to the conditions. Figures reproduced from Ref. [112] with permission from the PCCP Owner Societies.

With our model in Fig 2.13b, we intuitively highlight how the entropy and enthalpy

change upon adsorption are coupled. Model 2 explains the enthalpy-entropy correlation reported in Ref. [125]: the largest entropy changes are for small adsorption sites, which will have framework atoms in close proximity to the adsorbed methane. These proximal framework atoms will provide potential overlap to make the energy of adsorption more negative. As the site gets too small, however, repulsion takes effect and the enthlapy change quickly drops off as the site gets smaller. In this regime, structures will have very low guest loading. The reason the dropoff in Fig 2.13a was not observed in Ref. [125] is that they did not consider very poorly performing structures in their dataset.

Düren et al. [114] focus on the problem of obtaining the largest methane uptake in a material at the charging pressure P_H, assuming that the gas tanks are designed in such a way that we can use *all* methane adsorbed for the combustion. With this criteria, the best material must have a large accessible surface area and free volume and strong energetic interactions with adsorbed methane. They note that it is not straightforward to tune these properties by changing the pore size because, "to improve one of these may worsen the others in a complex manner". The model in which we vary the radius of the spherical shell of framework atoms that build an adsorption pocket (model 2) quantitatively elucidates why: changing the pore size affects the energy landscape inside the pore in addition to the the free volume. In the context of maximizing the deliverable capacity, the pore size must be tuned such that the optimal heat of adsorption is obtained. In Fig 2.13d, there are two pore sizes that yield the optimal heat of adsorption, but the one with the larger pore size will have a higher deliverable capacity because it achieves a higher free volume (and thus more sites for methane to adsorb).

The free energy landscape ultimately dictates the deliverable capacity of methane in a material. Correlations with surface areas and pore sizes such as in Figs 2.1c and 2.1d attempt to distill a complicated thermodynamic problem into simple, one-dimensional, guidelines, but should be taken with caution. One cannot always extrapolate geometric correlations with deliverable capacity far beyond the 298 K, 65 to 5.8 bar pressure range and the space of non-silica nanoporous materials. As an example, correlations of uptake with pore size and accessible surface area look different for different pressures (Figs 2.16 and 2.2). Model 2 shows that the best pore size, even under the assumption of a spherical adsorption pocket, depends upon the atoms used to build the adsorption pocket and the surface density of those atoms; zeolites will have a different optimal pore size than, for example, MOFs. Futher, we cannot be sure that the zeolite database, despite its vastness, exhausts the complete "space" of materials for a comprehensive analysis. That is, spurious relationships can arise from the fact that only a subset of possible materials were analyzed.

In summary, the potential for adsorbed natural gas technology is promising, and we hope our conclusions from the analysis of the zeolite dataset and our models help the materials-synthesis community intuit what features endow a material with a high deliverable capacity to guide in designing the optimum material. Finally, it is important to note that our conclusions can be applied to storing other gases in nanoporous materials, such as hydrogen [146]. Subject to the storage condition, one can perform a similar analysis as in this study.

Chapter 3

The Materials Genome in Action: Identifying the Performance Limits to Methane Storage

Analogous to the way the Human Genome Project advanced an array of biological sciences by mapping the human genome, the Materials Genome Initiative aims to enhance our understanding of the fundamentals of materials science by providing the information we need to accelerate the development of new materials. This approach is particularly applicable to recently developed classes of nanoporous materials, such as metalorganic frameworks (MOFs), which are synthesized from a limited set of molecular building blocks that can be combined to generate a very large number of different structures. In this Perspective, we illustrate how a materials genome approach can be used to search for high-performance adsorbent materials to store natural gas in a vehicular fuel tank. Drawing upon recent reports of large databases of existing and predicted nanoporous materials generated in silico, we have collected and compared on a consistent basis the methane uptake in over 650,000 materials based on the results of molecular simulation. The data that we have collected provide candidate structures for synthesis, reveal relationships between structural characteristics and performance, and suggest that it may be difficult to reach the current Advanced Research Project Agency-Energy (ARPA-E) target for natural gas storage.

This chapter is based on the following publication [147]:

C. Simon, J. Kim, D. Gomez-Gualdron, J. Camp, Y. Chung, R.L. Martin, R. Mercado, M.W. Deem, D. Gunter, M. Haranczyk, D. Sholl, R. Snurr, B. Smit. The Materials Genome in Action: Identifying the Performance Limits to Methane Storage. *Energy & Environmental Science*. (2015) 8 (4): 1190-1199

3.1 Introduction

In this chapter, we carry out a high-throughput computational screening of metal-organic frameworks, porous polymer networks, zeolitic imidazolate frameworks, and zeolites for materials that meet the ARPA-E methane deliverable capacity target for vehicular methane storage (see Sec 1.1). At present, the most promising materials have a deliverable capacity well below the ARPA-E target.[117, 148, 149] An important practical question is whether any sorbent material can be designed to deliver 315 v STP/v of methane under these conditions, especially given that the ARPA-E target is based solely on economic considerations for competitiveness with CNG.

One of the exciting aspects of advanced nanoporous materials is their analogy with the human genome, where a small number of building blocks (e.g., amino acids) can be combined to yield the nearly infinite variety of biology. Similarly, by combining different molecular building blocks, researchers have recently obtained unprecedented chemical and geometrical tunability of nanoporous materials, enabling optimization and tailoring for specific applications. For example, for materials such as metalorganic frameworks (MOFs) [117] or porous polymer networks (PPNs) [2], 10 millions of different materials can, in principle, be synthesized by combining different molecular building blocks of varying topologies (see Fig. 3.1). In practice, however, due to the limits on time and resources, only a small subset of the possible materials will ever be synthesized and tested as natural gas adsorbents. Similar challenges exist in the development of Li-ion batteries [150], solar water-splitters [151], piezoelectrics[152], scintillators[153], photovoltaics[154], and thermoelectrics [155]. In recognition of this problem, the Materials Genome Initiative is a 100 million dollar effort from the White House that aims to "discover, develop, and deploy new materials twice as fast" as the current methods. [37, 156] Part of this initiative is to develop computational tools to explore the vast space of materials and identify those that are optimal for a given application. In this Perspective, we analyze recent reports from the literature and demonstrate how this approach can address a very practical question: can we find a material that meets the ARPA-E target for methane storage?

3.2 Metric for methane storage

There are many important requirements for a material to be a successful candidate for a gas tank adsorbent. At present, it is impossible to take all these factors into account in a screening study. For example, it can be difficult to estimate the costs to synthesize a given material. Screening strategies have therefore focused on the metric that primarily determines the driving range from a given tank volume,[148] the deliverable capacity. The deliverable capacity is defined as the methane stored per volume of material in a fully loaded tank at the storage pressure, 65 bar, minus the methane that remains at the depletion pressure. The depletion pressure set by ARPA-E, 5.8 bar, was chosen because a sufficient pressure differential must be present to drive flow from the adsorbent to the engine. The deliverable

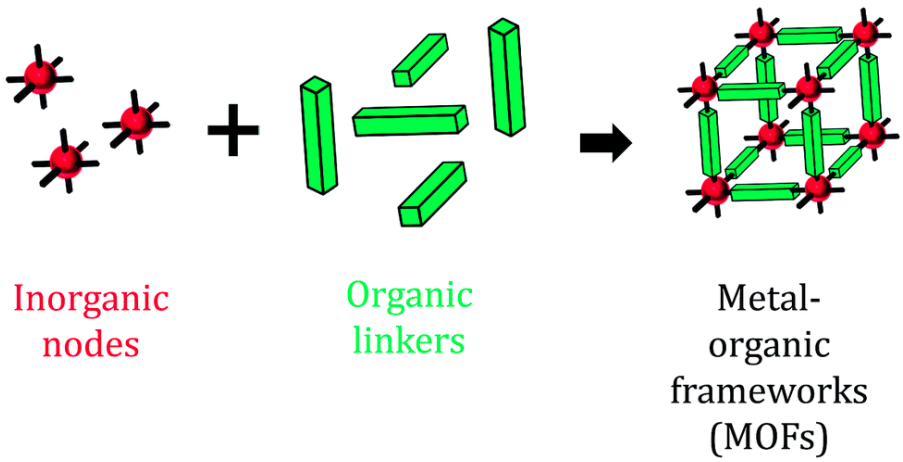


Figure 3.1: By combining different molecular building blocks, one can obtain an unprecedented large number of materials. For example, inorganic nodes (metals) are combined with organic linkers to form a metalorganic framework (MOF). Reproduced from Ref. [147] with permission from The Royal Society of Chemistry.

capacity metric takes into account that a material must not only store a large amount of methane at the fuel station; it must also release the methane from the pores when the valve from the tank is open to fuel the engine.

3.3 The Nanoporous Materials Genome

Over 3,000,000 different materials have been generated in silico by various research groups within the past few years. This set of materials includes libraries of in silico-generated zeolites [119, 120] MOFs [121, 157], zeolitic imidazolate frameworks (ZIFs) [158], and PPNs [159], as well as a library of ∼5000 MOFs that have been synthesized [160].

Materials

Fig. 3.2 summarizes the building blocks and structures of the different classes of porous materials examined in this Perspective. Fig 3.3 illustrates the different strategies used to combine the building blocks to build the structures.

In the materials genome approach, we aim to generate models for a large number of structures that are representative of a class of materials. This is, of course, very different from enumerating all possible structures, which would be an impossible aim given the infinite possible combinations of inorganic and organic building blocks. An intuitive strategy for designing new materials is to take inspiration from known experimental materials. In Fig 3.3 we illustrate different strategies, which apply this concept to generate databases of predicted materials. In the first work of this kind, Wilmer *et al.* [121] compiled a set of

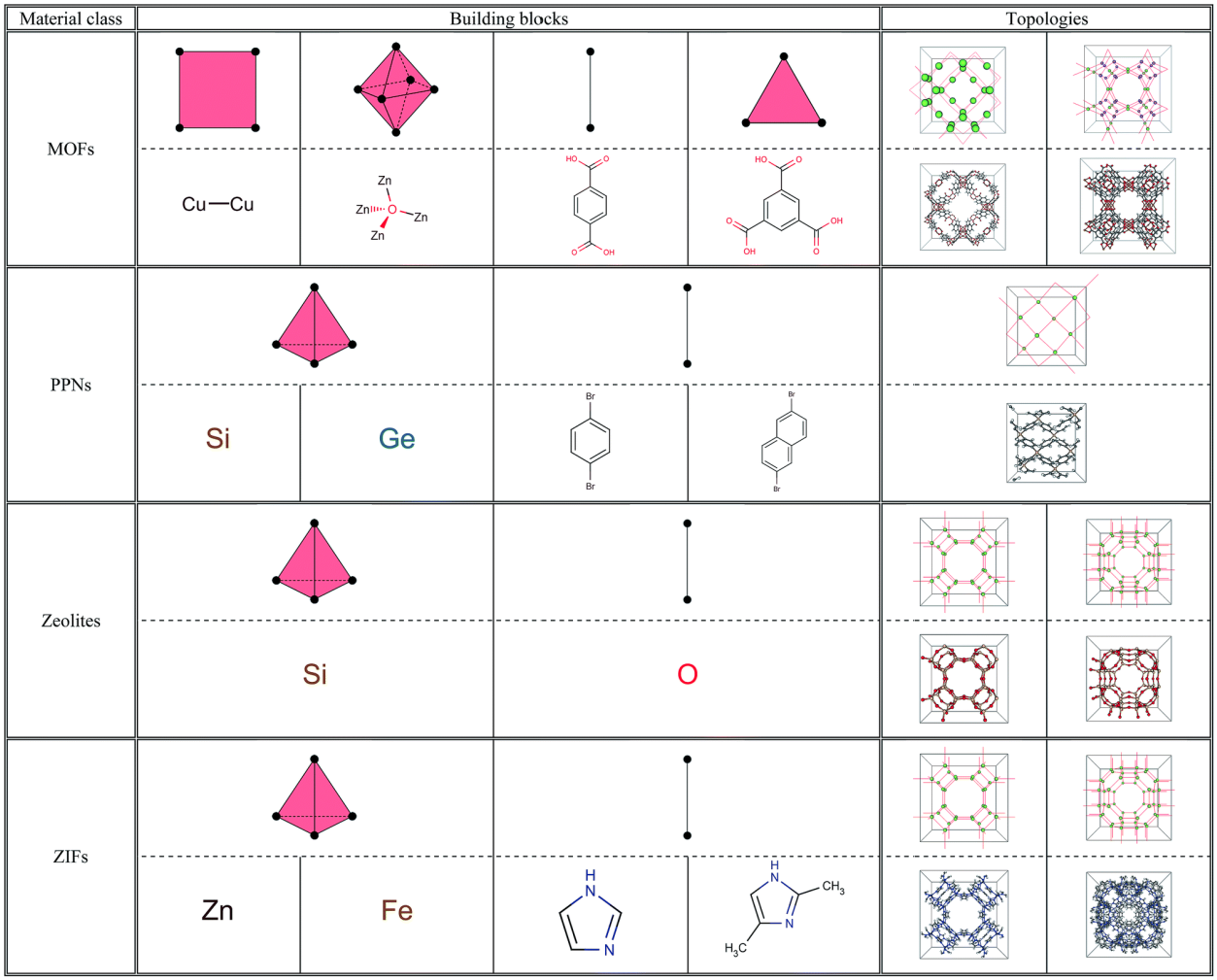


Figure 3.2: Materials studied in this work. MOFs, PPNs, zeolites and ZIFs are described in terms of their constituent building blocks and resulting topologies. The "Building blocks" column illustrates both symbolic representations of the geometry of possible building blocks (above) and chemical examples thereof (below). Similarly, the "Topologies" column provides graph representations (nets) of possible topologies for each material class (above) and examples of chemical structures exhibiting these topologies (below). Solid vertical lines distinguish between types of building blocks and/or variations in topology and geometry; note also that in some cases, many more possibilities exist beyond the few illustrated. Reproduced from Ref. [147] with permission from The Royal Society of Chemistry.

inorganic secondary building units (SBUs) and bridging organic linkers known from experimental MOF structures and combined these building blocks to generate over 130,000 new MOF structures. In such a scheme, the size of the resulting database can be arbitrarily expanded by the addition of linkers (or metal SBUs), but not all of these might be synthe-

sizable. In a complementary strategy, Martin *et al.* [157] used a database of commercially available molecules as a starting point and identified those with the appropriate linker characteristics; this strategy resulted in over 100 analogues of the well-studied MOF-5 system, and more recently, approximately 18,000 new PPN materials [159]. A similar strategy was employed to identify alternative imidazole-like substituents for ZIF materials [158]. Gomez-Gualdron *et al.* [161] used a small library of building blocks compatible with experimentally observed nets based on the metal SBU of the UiO-66 system[162]– which is renowned for its stability –resulting in over 200 MOF structures. These are examples of "building block selection" strategies for advanced porous materials; we duly note that this terminology does not readily apply to other materials such as zeolites.

The next step in computational materials design is developing an algorithm to combine the organic and/or metallic SBUs to form a particular structure. Here we also see different approaches. Wilmer *et al.* [121] defined rules specifying how building blocks may coordinate to one another (much like the limited ways in which Lego bricks may connect) and used a recursive algorithm to exhaustively generate MOF structure models. Martin *et al.* [163] enumerated highly symmetric crystal topologies consistent with particular SBU combinations and constructed models by positioning the building blocks according to the underlying net. These schemes are not limited to MOFs and can also be used to generate PPNs, ZIFs and related materials such as covalent organic frameworks (COFs).

An alternative approach was employed for enumeration of novel zeolite structures, which differ in their pore topologies. Deem *et al.* [119, 120] sampled alternative zeolite topologies by positioning silicon atoms within a unit cell through a Monte Carlo procedure, exploring a wide range of unit cell dimensions and crystal densities for all 230-space groups. Symmetry operations acting on crystallographically unique atoms then generate the full unit cell structure, resulting in 2.6 million topologically distinct zeolite-like structures. The structures were then optimized using a classical force field. Out of the 2.7 million structures, 331,172 were identified to be thermodynamically accessible, i.e., exhibiting energies no more than 30 kJ mol^{-1}-Si above α-quartz. Finally, since ZIFs are a class of MOFs possessing the same pore topologies as zeolites, Lin *et al.* [158] used the zeolite database to generate a corresponding ZIF database through chemical substitution. Most experimentally synthesized MOF structures are deposited in the Cambridge Structural Database [164]. However, many of these structures contain solvent molecules and other impurities, which makes it very difficult to directly compare these structures with those generated computationally. To make such a comparison possible, Chung *et al.* [160] developed computational methods to automatically remove these impurities from the $\sim$5000 deposited structures; the resulting dataset is known as the computation-ready, experimental (CoRE) MOF database [160]. All of these structures are accessible through the Nanoporous Material Website [165].

Properties

To characterize and compare the libraries of different classes of materials, we show in Fig. 3.4 the distributions of their geometric properties. These properties were calculated with the

open-source software suite Zeo++ [166], which uses a Voronoi decomposition to translate structural information (positions and hard-sphere radii of atoms, size and shape of the periodic unit cell) into a periodic graph representation of the material's porosity. This graph is then analyzed to obtain the accessible surface area and the largest included sphere. The largest included sphere is defined as the largest hard-sphere that fits in the pore network of the material without overlapping any framework atoms, also modeled as hard-spheres.

Fig. 3.4 illustrates the large range of geometric properties spanned by these materials. Fig. 2 (top left) shows that zeolites typically have the highest crystal densities while PPNs have the lowest; MOFs have a very broad density distribution. Fig. 3.4 (top right) shows that the crystal density correlates inversely with the geometric void fraction, as the most porous materials are generally the least dense. The distribution of pore sizes (Fig. 3.4, bottom left), measured by the largest included sphere, mimics the distribution of void fractions as expected from intuition. Fig. 3.4 (bottom right) shows that PPNs and MOFs tend to achieve the highest surface areas. Characterizing materials by their geometric properties is important because we can link these simple descriptors to the deliverable capacity and elucidate guidelines for material design and selection. Furthermore, these descriptors serve as a multi-dimensional fingerprint that characterizes each material in a high-dimensional space, opening up the possibility for data mining and machine learning techniques to rapidly search this space and predict material performance from simple, easily computed structural descriptors [167, 168, 95].

3.4 Screening the materials genome

Monte Carlo simulations

Grand-canonical Monte Carlo simulations have been used in the literature to calculate the adsorption of methane in 650,000 structures in the libraries of zeolites [112], MOFs [121, 157, 161, 169], and PPNs [159]. Some of these studies have used slightly different force fields or pressures. To allow for a direct comparison of these studies, we recomputed the deliverable capacity for these materials. In these calculations, we used the following force fields: UFF[108] for PPNs, MOFs, and ZIFs; that reported by Dubbeldam *et al.* [127, 170] for zeolites; and TraPPE [106] for methane. Lorentz-Berthelot mixing rules were used to obtain the solid-adsorbate Lennard-Jones parameters. We used periodic boundary conditions, and the Lennard-Jones potentials were truncated and shifted with a cut-off radius of 12.8 Å. We assumed the crystal structures to be rigid, which may induce some errors, as some materials are known to expand and contract with gas adsorption[171, 172, 173, 174, 175]. These force fields generally provide good descriptions of most materials in our study. Notable exceptions are MOFs containing open-metal sites [118, 176, 104].

To test the reliability of our methodology, we compared the methane adsorption isotherms of experimental structures that were available in the literature to the analogous structures in our dataset. Using Grand-canonical Monte Carlo simulations [110] (see Sec. 1.2), we

simulated the isotherms for HKUST-1 (see Fig. 3.5); Mg-MOF-74, Ni-MOF-74, and Co-MOF-74 (see Fig. 3.6); and MOF-5 and PCN-14 (see Fig. 3.7) from 0 to 80 bar at 298 K.

For PCN-14 and MOF-5, our simulations give a good agreement with the experimental data. For the MOF-74 series, it is known that not all open metal sites are activated [148], giving a lower maximum adsorption compared to the perfect crystal structures used in our simulations. To correct for these blocked adsorption sites, we scaled the isotherms by the ratio of normalized surface areas in Mg-, Co-, and Ni-MOF-74, as reported in Table 2 of Ref. [177], where the surface areas are derived from a Langmuir fit of 77K N_2 adsorption isotherms. In the case of Zn-MOF-74, $\approx$ 24% of the surface area, and thus binding sites, is not accessible; in Ni-MOF-74, $\approx$ 10%; in Mg-MOF-74, $\approx$ 13%; and in Co-MOF-74, $\approx$ 18% of binding sites are not accessible. We see that in frameworks with these particularly strong binding sites, such as Ni-MOF-74 and HKUST-1, the UFF + TraPPE force field underestimates the adsorption isotherms.

In conclusion, our force field tends to slightly underestimate the measured methane uptake in certain materials, by anywhere from 7% (MOF-5) to 20% (Ni-MOF-74) at 65 bar, which explains why in our screening study our top performing structures have a slightly lower deliverable capacity compared to the experimental structures.

Methane isotherms at 298 K were calculated using a high-performance, parallel algorithm developed for GPUs (graphics processing units), allowing us to compute such a large number of adsorption isotherms in a reasonable time.[111] The Peng-Robinson equation of state was used to relate the pressure of methane gas to its chemical potential.

Screening

Our initial screening included 650,000 materials: 139,407 predicted zeolite structures[119, 120] with isotherms from Simon *et al.* [112]; 137,953 predicted metal-organic framework (hMOF) structures [121] of which we made a random selection of 10,000 structures; 17,970 predicted porous polymer network (PPN) structures [159]; 381,178 predicted zeolitic imidazolate framework (hZIF) structures [158], of which we simulated a random selection of 20,000 structures; and a set of 5109 Computation-Ready Experimental (CoRE) metal-organic framework structures [160].

We computed the deliverable capacity for the initial set of 20,000 hZIFs and 10,000 hMOFs. These calculations showed that the materials with the highest deliverable capacity exhibit a largest included sphere in the range of 8.0-14.5 Å and void fractions in the range of 0.25-0.7. See Fig 3.8. We then applied these criteria in void fractions and largest included spheres to pre-screen the remaining hMOFs and hZIFs for top performers and added 34,363 hMOF and 37,437 hZIF structures to our final list for which we performed Grand Canonical Monte Carlo simulations.

3.5 Results and discussion

Large datasets of structures such as this can be used to identify promising candidate materials for different applications, assess thermodynamic limits to performance, and elucidate relationships between geometric properties and performance, as illustrated below for natural gas storage. Fig. 3.10 shows the distribution of the deliverable capacities of different materials. Generally, PPNs and MOFs have significantly higher deliverable capacities than zeolites. While the best structures are those that show a significant improvement over the 62 v STP/v deliverable capacity of a free-space tank, no material examined meets the ARPA-E deliverable capacity target. The highest predicted deliverable capacity in the set of 650,000 materials is 196 v STP/v. The material with the highest predicted deliverable capacity in our study is a hypothetical porous polymer network (PPN), Adamantane 4387 1-net 004 [159], exhibiting a 65 - 5.8 bar deliverable capacity of 196 v STP/v. The structure is shown in Fig. 3.9. We plotted the potential energy contours of a methane molecule in the pores at -12 kJ/mol (orange) and 0 kJ/mol (gray) to highlight the binding regions. This material exhibits a largest included sphere of 11.75 Å; this is larger than a single methane molecule. The strong binding regions in orange indicate that multiple methane molecules can be efficiently packed into the pores. The computed surface area of this material is 1992 m^2/cm^3.

The PPN in Fig 3.9 can in principle be synthesized from known synthesis routes [178] using an adamantane core and 1,2-dibromoethylene linkers: Four of these linkers are appended to each tetrahedral point of the adamantane cage. When two of these tetrahedral monomers then react, we get the C_4H_4 linkage spanning the adamantane cages. This 'two-monomer synthesis route' is discussed in detail in Ref. [159].

The materials with the best experimentally measured deliverable capacities to date are MOF-5 (185 v STP/v)[148], HKUST-1 (185 v STP/v) [117], UTSA-76 (194 v STP/v) [179], and MOF-519 (208 v STP/v) [180]. Given possible errors in simulation force fields, experimental sample quality, and experimental adsorption measurements, it is striking that extensive experimental and computational work all converge on a maximum deliverable capacity of approximately 200 v STP/v. The very large number of structures examined computationally suggests that the top experimentally tested adsorbents are already at the boundary of the thermodynamic and material performance limits.

Analysis of these large datasets in the materials genome approach can also provide fundamental insights that are difficult to draw from the limited experimental data. For example, Fig. 3.11 shows that the materials with the largest deliverable capacities occur in a narrow range of crystal densities. Fig. 3.11 also shows the relationship between the deliverable capacity and the pore size. We included in Fig. 3.11 a selection of experimental structures from each material class that exhibit high deliverable capacities and for which experimental methane adsorption isotherms were available. The experimental data generally show the same trends as the simulated data, but the limited number of data points makes it difficult to spot the trends. Fig. 4 illustrates how larger pores increase the capacity to an extent; if the pores become too large, the methane molecules in the middle of the pores do not feel the van der Waals interactions with the walls, and the deliverable capacity approaches that

of a free-space tank. The optimal pore diameter of ∼11 Å for methane storage is a useful guideline because the size of the molecular building blocks used in the synthesis directly controls the pore size.

Fundamental physical models are an important complement to the "big-data" approach of plots as presented in Fig. 3.11. An empirical formula for methane storage of MOF materials has been developed by Kong *et al.*[188] and He *et al* [189]. For example, Simon *et al.* [112] developed a statistical thermodynamic model in which the deliverable capacity of a material is viewed as the product of the number of effective adsorption sites per volume, M, and the fractional deliverable capacity f. The fractional deliverable capacity is maximal if the energetics of the adsorption sites are optimal; if the adsorption of methane is too strong, too much methane remains in the tank at the discharge pressure, and, if the adsorption is too weak, too little methane will be in the tank at the storage pressure [112, 122].

To estimate the saturation methane loading in each material, which we call the effective density of adsorption sites, we fit the simulated methane adsorption isotherms from 1 – 160 bar fugacity to model 1 in Ref. [112], which builds upon a Langmuir model by including adsorbate-adsorbate interactions. The model is a Langmuir model at first order, with a correction term that is weighted by the strength of the adsorbate-adsorbate interactions (θ) [112]

$$\frac{\sigma}{M} = \frac{\frac{K_H}{M}\phi P}{1+\frac{K_H}{M}\phi P} + \theta\left(\frac{\frac{K_H}{M}\phi P}{1+\frac{K_H}{M}\phi P}\right)^2\left(\frac{\frac{K_H}{M}\phi P}{1+\frac{K_H}{M}\phi P} - 1\right). \tag{3.1}$$

The variable σ is the methane loading per volume of material; M is the saturation loading or effective density of sites; K_H is the Henry coefficient; P is the pressure; ϕ is the fugacity coefficient such that ϕP is the fugacity of methane corresponding to pressure P. The Henry coefficient in eqn 3.1 is independently obtained from the Widom insertion method. We thus fit each methane adsorption isotherm to the model in eqn 3.1 using the parameters M and θ with a nonlinear least squares data fitting routine implemented with the OPTIMIZE function in Scipy, an open-source computing package in Python. If the highest-pressure point in the simulated isotherm (160 bar fugacity) was less than 60% of the identified M, we extended the isotherms to a fugacity of 700 bar to obtain enough curvature in the simulated isotherms for our fitting routine to reliably estimate M. For plots involving M, we only include structures whose residual sum of squares (including all 14 data points on the isotherm) is below 5% of M to help ensure the estimation of M is reliable.

In Fig. 3.12, we plot the fractional deliverable capacity f of each material against the saturation loading, M. A hyperbola $f = \sigma_{del}/M$ is a curve of constant deliverable capacity σ_{del}. If the adsorbent in the fuel tank were filled at infinite pressure and depleted by pulling a vacuum to remove all residual methane, all sites would be fully utilized and the deliverable capacity would be equal to the saturation loading (M). However, the adsorption sites may not be fully occupied at 65 bar, and some methane will remain in the pores at 5.8 bar; therefore even for an optimal material, the deliverable capacity is only some fraction f of the saturation loading. For all material classes, we find similar behavior in Fig. 3.12. For low M (∼350 v STP/v), there are many materials that have adsorption sites with

the optimal methane affinity, endowing the material with a maximal fractional deliverable capacity. However, there is a tradeoff between the density of adsorption sites and the density of the material, such that in the case of high M, we achieve an effective density of sites where none of the materials have sites that have optimal affinities for methane. We can only increase the number of adsorption sites if we decrease the number of framework atoms. Lowering the density of framework atoms causes the additional adsorption sites to have little or no interactions with the framework. Hence, the deliverable capacity of these materials approaches that of an empty tank.

From these data, we can identify the ideal material for methane storage: the material with the largest density of adsorption sites under the constraint that the sites have optimal methane interactions. The distinguishing characteristic between different classes of materials is the number of atoms that are involved in creating an optimal adsorption site. Fig. 3.11 shows that PPNs and MOFs perform better than zeolites and ZIFs, as the former require fewer atoms to create optimal adsorption sites.

An intriguing question is whether a material exists that meets the ARPA-E target. The experimental results to date and computational screening of over 650,000 structures suggest that there may be a physical limitation on the deliverable capacity, around 200 v STP/v, which is quite far from the ARPA-E target of 315 v STP/v. Adopting a materials genome approach allowed this conclusion to be reached rather quickly. Note that the computational studies that provided the data in Fig.3.10 were all published in the past 2 years. Given this tentative conclusion, we can ask whether some other set of operating temperatures and pressures might, then, make adsorbed natural gas tanks more attractive. With the computational infrastructure in place, we can readily recalculate the deliverable capacity for other conditions.

We now outline our methods for calculating the deliverable capacity of our materials under different conditions than set in the ARPA-E target. To avoid having to carry out simulations at many different temperatures and pressures, we characterize each material by the parameters M, K_H, and θ obtained from fitting the simulated isotherms at 298 K to the model in eqn 3.1. To extrapolate the loading at a higher temperature, we assume that the temperature dependence of the Henry coefficient K_H is given by the Van't Hoff equation and the temperature dependence of θ (see Ref. [112]):

$$K_H \sim e^{-\frac{\Delta H}{RT}} \tag{3.2}$$

$$\theta \sim \frac{1}{T}. \tag{3.3}$$

We calculated the enthalpy of adsorption ΔH (= negative of the heat of adsorption) during our simulations. We subsequently use these parameters to estimate the methane adsorption at alternative conditions.

To test that the various predictions made with these fitted parameters gives a sufficiently accurate description of the isotherms that would be obtained by simulations we carried out several tests. In Fig. 3.13 we show that the fitted isotherms describe the simulated deliverable

capacity sufficiently accurate. Fig. 3.14 shows that this approach gives a sufficiently accurate description of the methane adsorption isotherms at these different temperatures.

In the calculations that follow, we include a random sample of 3,701 materials from each class to assign each class an equal prior. We amalgamate all material classes together and plot the distribution of deliverable capacities at the different storage conditions. For each alternative operating condition, we also depict how one class of materials may perform better over another class at different conditions by stacking the probability distributions for each class. This allows us to visualize the contributions of each material class to the probability distributions.

We now assess the impact of changing the operating pressures on the deliverable capacity at 298 K. Methane is stored in the adsorbed natural gas tank at P_H bar at the refilling station, and a tank is considered depleted if it exhibits a pressure of P_L bar. By changing P_H and P_L, we compute the deliverable capacity under two scenarios that will benefit the deliverable capacity: (i) Increasing the storage pressure P_H but keeping the ARPA-E target's discharge pressure of $P_L = 5.8$ bar (ii) Decreasing the discharge pressure P_L but keeping the ARPA-E target's storage pressure of $P_H = 65$ bar.

Fig. 3.15 shows how the distribution of deliverable capacities is changing if we increase the storage pressure $P_H = 65$, 100, 150, and 200 bar. As expected, the higher the storage capacity the larger the deliverable capacity; however, even at 200 bar we do not reach the ARPA-E target.

Fig. 3.16 shows the effects of decreasing the recharging pressure $P_L = 5.8$, 3, 1, and 0 bar, while keeping the charging pressure $P_H = 65$ bar. In these calculations the deliverable capacity with $P_L = 0$ is equal to the loading at $P_H = 65$ bar. As expected, decreasing the decharging pressure increases the deliverable capacity. In particular materials for which methane is strongly bound will have a higher deliverable capacity.

Another interesting option is to re-engineer the tank to use residual heat from the engine to drive off the residual methane at the delivery pressure [169]. Fig. 3.17 shows that by discharging at a higher temperature, one can significantly increase the deliverable capacity. We also see that the shape of the distribution changes, the reason being that a different set of materials is optimal for different storage conditions (See Fig. 3.15 and 3.16). For example, a lower discharge pressure or higher discharge temperature favors materials that have stronger interactions with methane, while a higher storage pressure favors materials with a larger pore volume and weaker interactions. Still, none of these more favorable storage conditions allow materials to reach the ARPA-E target.

3.6 Conclusions

Screening hundreds of thousands of different materials gives us new insights that cannot be obtained otherwise. First of all, the materials genome approach allows us to explore a much larger range of materials than researchers could ever synthesize and is therefore ideal to obtain insights into the limits of an entire class of materials and into their structure/performance

relationships. In the past, if we wanted to know how methane uptake correlates with, for example, crystal density, we might measure or simulate the properties of a dozen materials and look for a simple linear correlation. Now, with hundreds of thousands of structures, we obtain a much more complete insight into the full complexity of this question, as shown in Fig. 3.11. In addition, the huge amount of data generated in high-throughput screening opens opportunities for data mining as a tool to extract additional, possibly unexpected relationships between material characteristics and performance. For example, in Fig. 3.11, we can visualize the relationship between the pore size and the deliverable capacity with a simple scatter plot. As we see drastically different deliverable capacities for a given pore size, it is clear that certain combinations of properties are required and thus including other descriptors will add predictive value. In such a high-dimensional space, machine-learning techniques are useful for extracting complicated relationships and making more accurate performance predictions from structural descriptors.

Although hundreds of thousands of predicted nanoporous materials have been generated within the past few years, the chemical search space is so large that all possible structures cannot be enumerated. Fig. 3.4 shows that the experimental MOF structures have different crystal density and surface area distributions than structures in the predicted MOF database. These differences raise questions about how to ensure that the chemical space is adequately sampled. In addition, only a subspace of all possible materials can ever be synthesized. An interesting application of material genomics would be to identify the feasibility of synthesizing a given material. From the screening studies summarized here, we conclude that the classes of materials currently being investigated are unlikely to meet the ARPA-E target for natural gas storage. This is not surprising as the ARPA-E target was not based on thermodynamic or material arguments but rather based on economic competition with CNG. High-throughput screening allowed an assessment of the feasibility of the target in a very short period of time, allowed alternative operating scenarios to be quickly evaluated, and provided useful structure/performance relationships. This is an illustrative example of how a materials genome approach can drastically reduce experimental time and effort by eliminating unproductive tasks and focusing experimental efforts on the most interesting candidates.

An important practical question is whether an energy density of ca. 200 v STP/v under the currently proposed storage conditions is sufficient to justify the adoption of vehicular adsorbed natural gas fuel tanks. However, natural gas storage is not the only application of nanoporous materials. These materials are also of interest for gas separations [11], including CO_2 capture [12], gas sweetening [137], and separation of hydrocarbons [190], as well as membranes [191], catalysis [15], sensors [14], and drug delivery[16]. For these applications, one can screen the databases discussed in this Perspective to reveal any performance limits, identify specific materials with exceptional properties, and elucidate relationships between material characteristics and performance using data mining techniques.

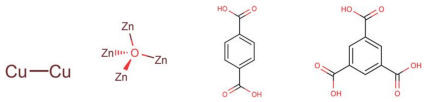
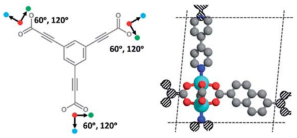
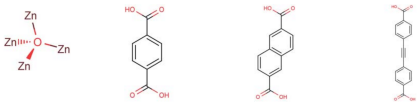
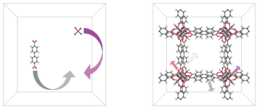
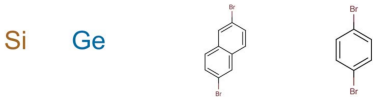
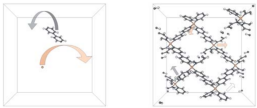
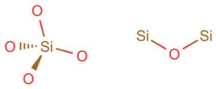
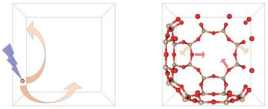
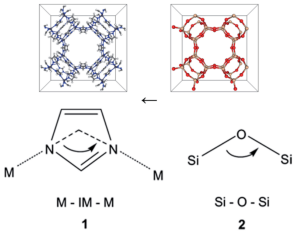
Name of the material dataset	Building block selection strategy	Model construction strategy
MOFs		
	Informed selection of previously utilized metal cations and organic linkers or design of similar, permitting additional functional groups	Modes of coordination between building blocks are specified; recursive algorithm combines building blocks in permissible combinations and infers a unit cell
MOF-5s		
	Zn_4O with commercially available dicarboxylic acids; pruning of linkers for linearity, rigidity, *etc.* prior to construction	Based on the underlying net of MOF-5 (pcu), building blocks are positioned in space and symmetry operations complete the unit cell; structures subsequently relaxed with semi-empirical PM6-DH2
PPNs		
	Si/Ge/C/adamantane with commercially available dibromides; pruning of linkers for linearity, rigidity, *etc.* prior to construction	Based on the underlying net of PPN (dia), building blocks are positioned in space and symmetry operations complete the unit cell; structures subsequently relaxed with semi-empirical PM6-DH2
Zeolites		
	Tetrahedrally coordinated silicon atoms, bridged by oxygen atoms	For each symmetry space group and a variety of unit cell parameters, Si atoms are positioned *via* a Monte Carlo process; symmetry operations complete the unit cell and bridging oxygens are placed; structures are subsequently relaxed with one of two interatomic potentials
ZIFs		
	Tetrahedrally coordinated zinc atoms, bridged by imidazolate linkers	
ZIFs (decorated)		For each zeolite topology, the Si–O–Si motif is replaced by a Zn–Im–Zn motif (Im being an imidazolate-like group), the unit cell is scaled by 1.95× compared to the original zeolite
	Tetrahedrally coordinated zinc atoms, bridged by commercially available, functionalized imidazolate linkers	

Figure 3.3: The complementary strategies employed to navigate the vast chemical space of possible nanoporous materials, exemplified by the specific datasets discussed in this work. Each strategy comprises a "Building block selection" and "Model construction" component, as described. Reproduced from Ref. [147] with permission from The Royal Society of Chemistry.

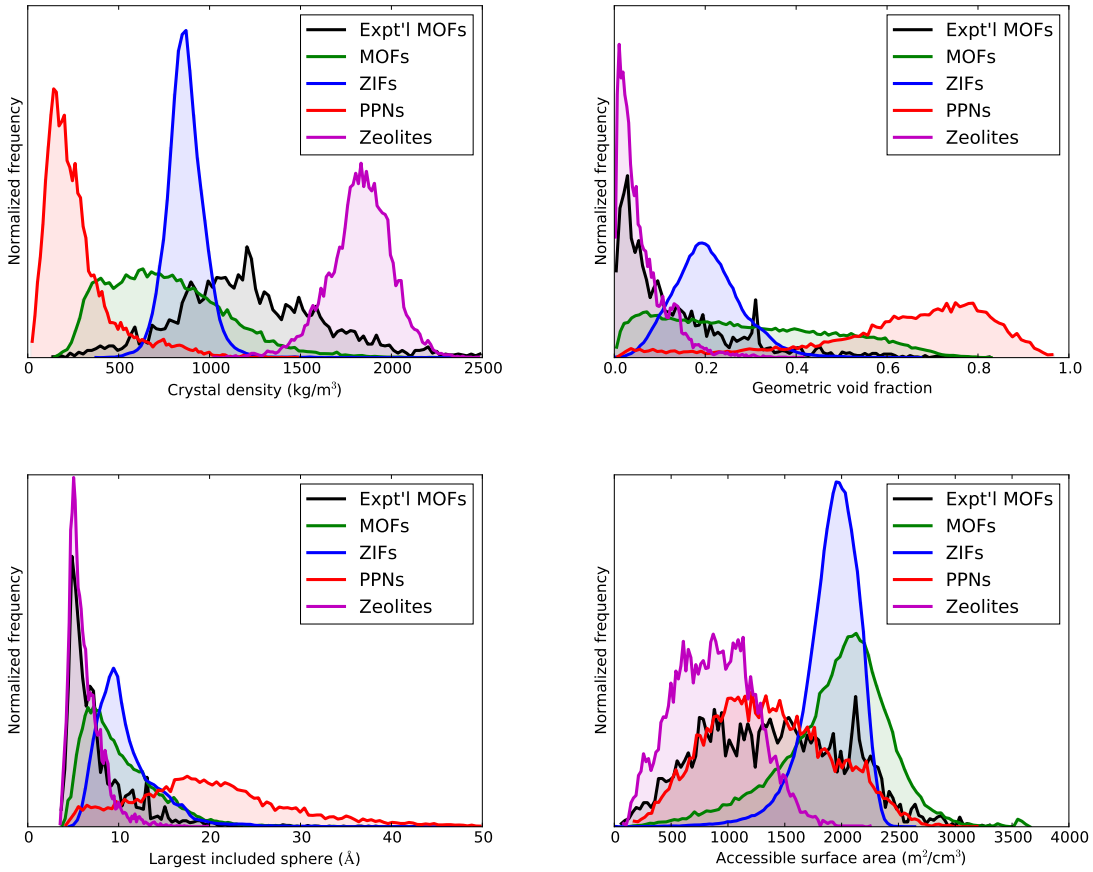


Figure 3.4: Comparison of the geometric properties in the predicted materials datasets (zeolites, PPNs, ZIFs, and MOFs) and the set of experimental MOFs, as computed from the crystal structures: (top left) crystal density, (top right) geometric void fraction, (bottom left) largest included sphere, (bottom right) accessible surface area. In these histograms we include only those structures that are methane-accessible. A structure is declared methane-accessible (using a methane probe of radius 1.625) if the largest free sphere is greater than the size of our methane probe. We also do not include materials that have deliverable capacities of less than 1 v STP/v. Reproduced from Ref. [147] with permission from The Royal Society of Chemistry.

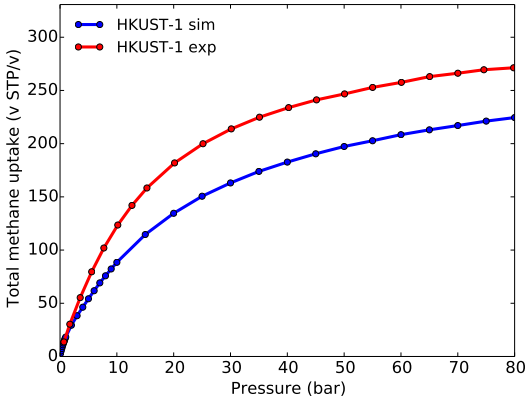


Figure 3.5: Adsorption isotherm of methane in HKUST-1. The experimental data are from Mason *et al.* [148]. Reproduced from Ref. [147] with permission from The Royal Society of Chemistry.

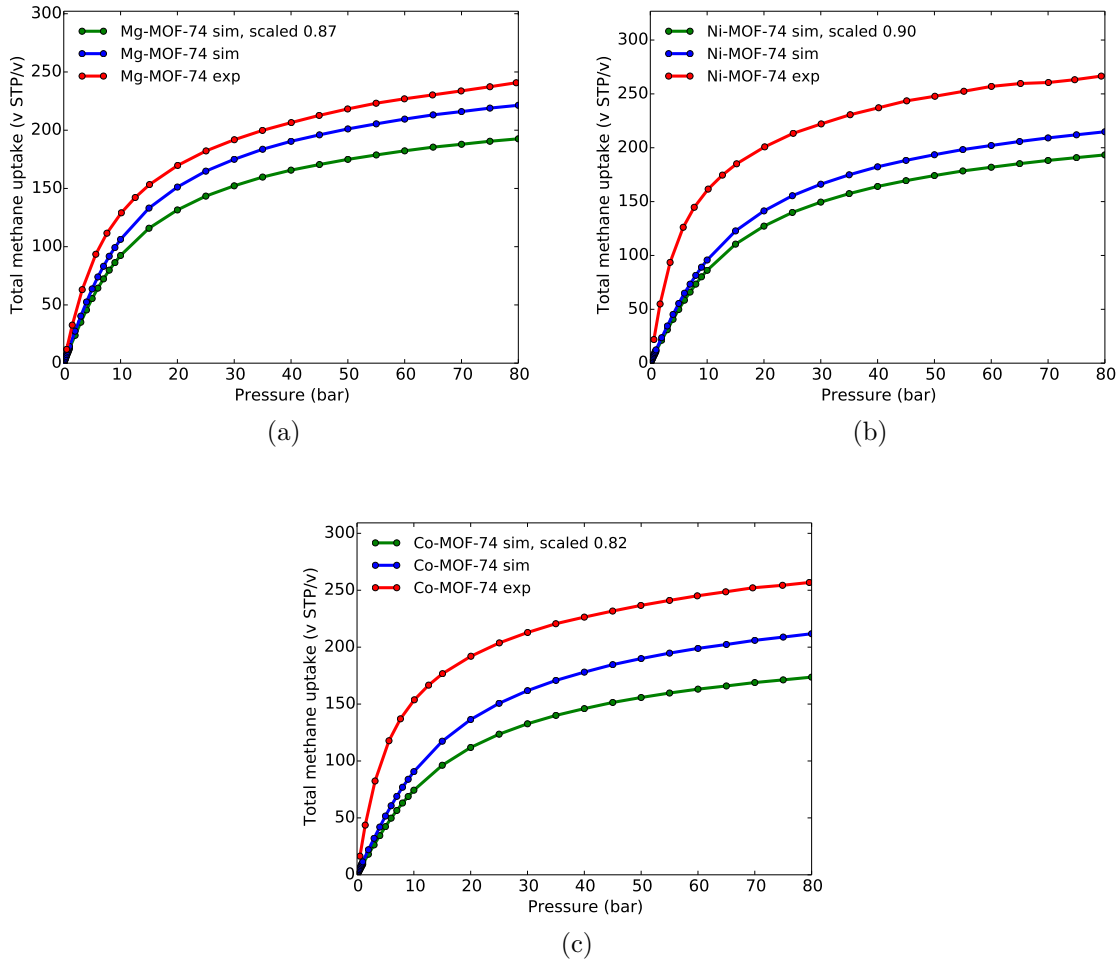
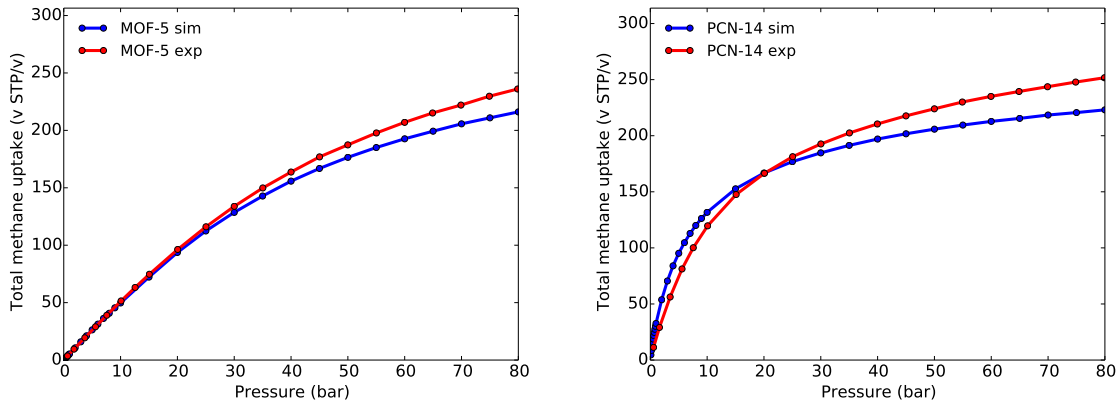


Figure 3.6: Adsorption isotherm of methane in (a) Mg-MOF-74, (b) Ni-MOF-74, and (c) Co-MOF-74. The experimental data are from Mason *et al.* [148]. The simulated isotherms are scaled to take into account that in the experimental structures not all open metal sites are accessible. Reproduced from Ref. [147] with permission from The Royal Society of Chemistry.



(a) (b)

Figure 3.7: Adsorption isotherm of methane in (a) MOF-5 and (b) PCN-14. The experimental data are from Mason *et al.* [148]. Reproduced from Ref. [147] with permission from The Royal Society of Chemistry.

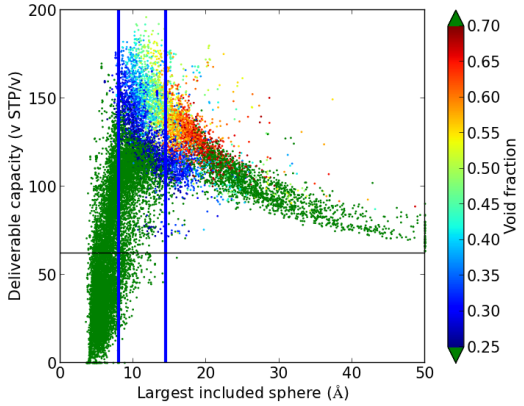


Figure 3.8: Screening strategy with geometric descriptors. Deliverable capacity plotted against the diameter of the largest included sphere. Color indicates geometric void fraction. We computed methane adsorption isotherms of all structures falling between the two vertical blue lines and having a void fractions between 0.25 and 0.70. Reproduced from Ref. [147] with permission from The Royal Society of Chemistry.

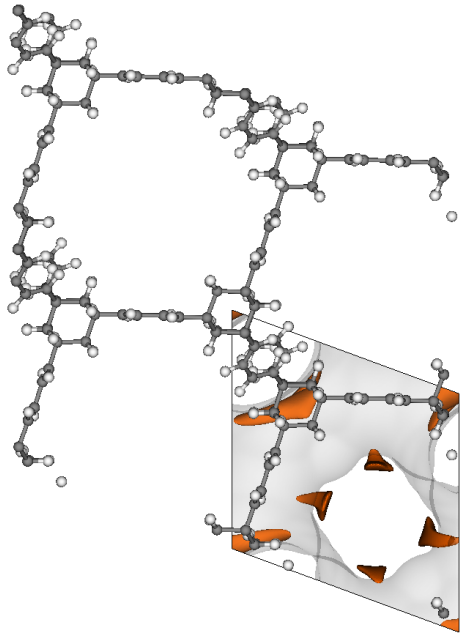


Figure 3.9: Adamantane 4387 1-net 004 exhibits a 65 - 5.8 bar methane deliverable capacity of 196 v STP/v. Contours of the potential energy of a methane molecule in the pores are shown at -12 kJ/mol (orange) and 0 kJ/mol (gray). Reproduced from Ref. [147] with permission from The Royal Society of Chemistry.

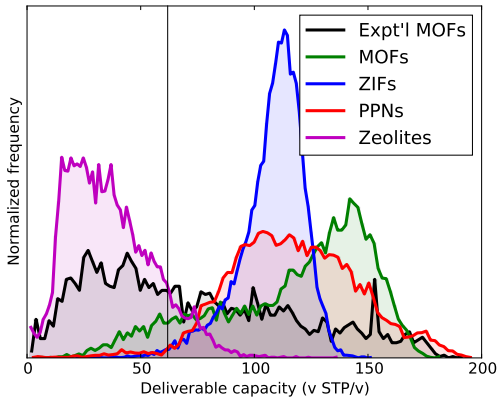


Figure 3.10: Distribution of the deliverable capacity for different materials; the vertical line marks the deliverable capacity of an empty tank (calculated from the density of bulk methane at 65 bar minus that at 5.8 bar using the PengRobinson equation of state). Reproduced from Ref. [147] with permission from The Royal Society of Chemistry.

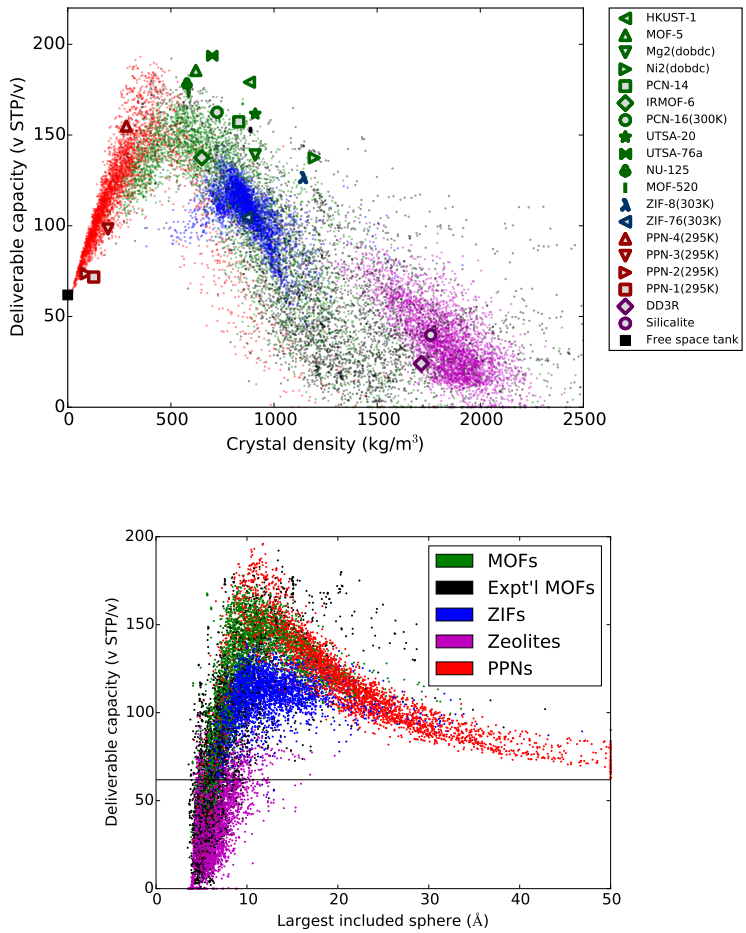


Figure 3.11: Deliverable capacity for different materials: (top) relationship between the deliverable capacity and the crystal density. The symbols are some examples of structures for which we computed the deliverable capacity from the experimental isotherms. The experimental data were taken from the following references: MOF-519 and MOF-520 [180]; HKUST-1, Mg2(dobdc), Ni2(dobdc), MOF-5, PCN-14 [148]; IRMOF-6 [181]; PCN-16, [182]; NU-125, UTSA-20 [117]; ZIF-8, ZIF-76, [183]; PPN-4 [184]; PPN-1, PPN-2, PPN-3 [185]; DD3R [186]; and Silicate [187]. In all figures, the materials are color-coded according to their class. (bottom) Relationship between the deliverable capacity and the largest included sphere diameter; the horizontal line indicates the deliverable capacity of an empty tank. In these scatter plots, we include only a random sample of 3701 structures from each class to avoid cluttering the plots and to assign equal importance to each class of materials. Reproduced from Ref. [147] with permission from The Royal Society of Chemistry.

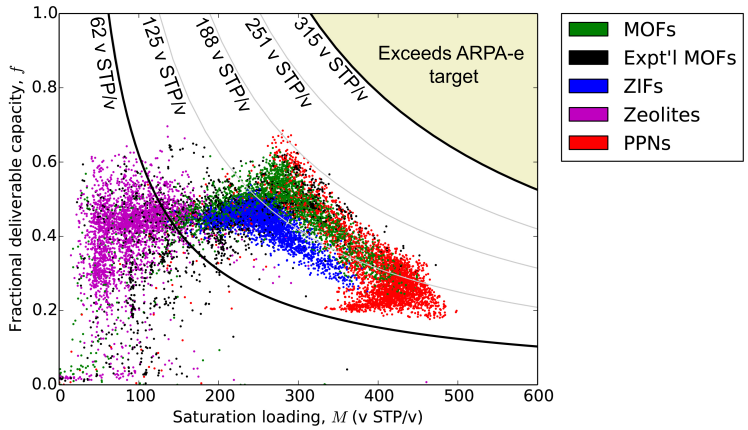
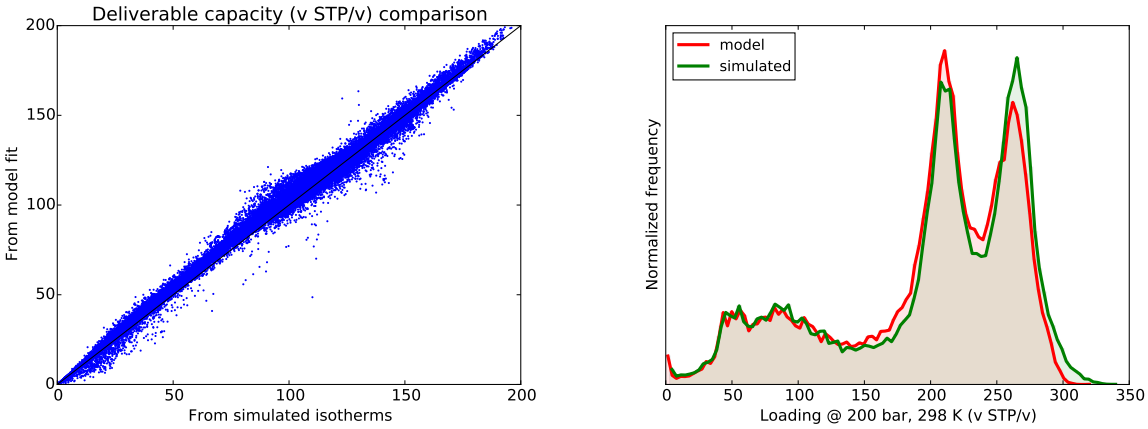


Figure 3.12: Fractional deliverable capacity f plotted against the saturation loading M for each material. Hyperbolas indicate lines of constant deliverable capacity. The bottom and top hyperbolas give the deliverable capacity of a free space tank and the APRA-E target, respectively; the region exceeding the ARPA-E target is highlighted in yellow. Reproduced from Ref. [147] with permission from The Royal Society of Chemistry.



(a) (b)

Figure 3.13: Validating the approximation of isotherms with the model in eqn. 3.1. (a) For each material, we plotted the deliverable capacities at 298 K between 65 and 5.8 bar obtained from the model fit against that obtained from the simulated isotherm. The points follow the diagonal with a modest variance. (b) We compare the distribution of 200 bar loadings using the simulated loading and the fitted model. Reproduced from Ref. [147] with permission from The Royal Society of Chemistry.

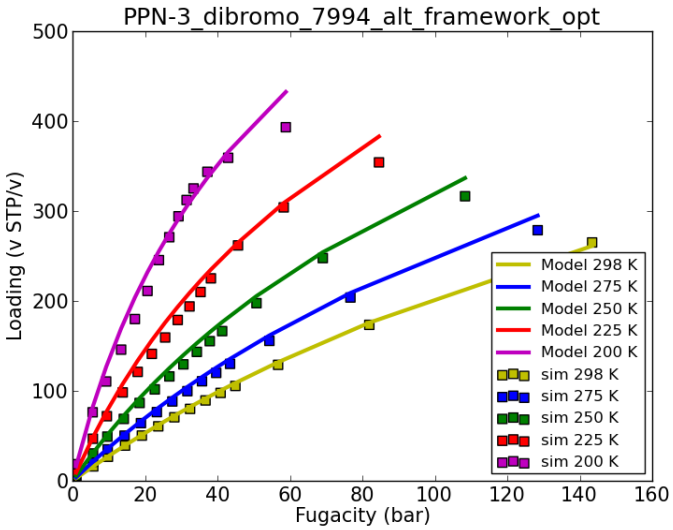


Figure 3.14: Simulated methane adsorption isotherms in a sample structure (name depicted as plot title) with simulations (points) against the extrapolations from the model in eqn 3.1 that was fit to the 298 K isotherm (solid lines) for varying temperatures. Reproduced from Ref. [147] with permission from The Royal Society of Chemistry.

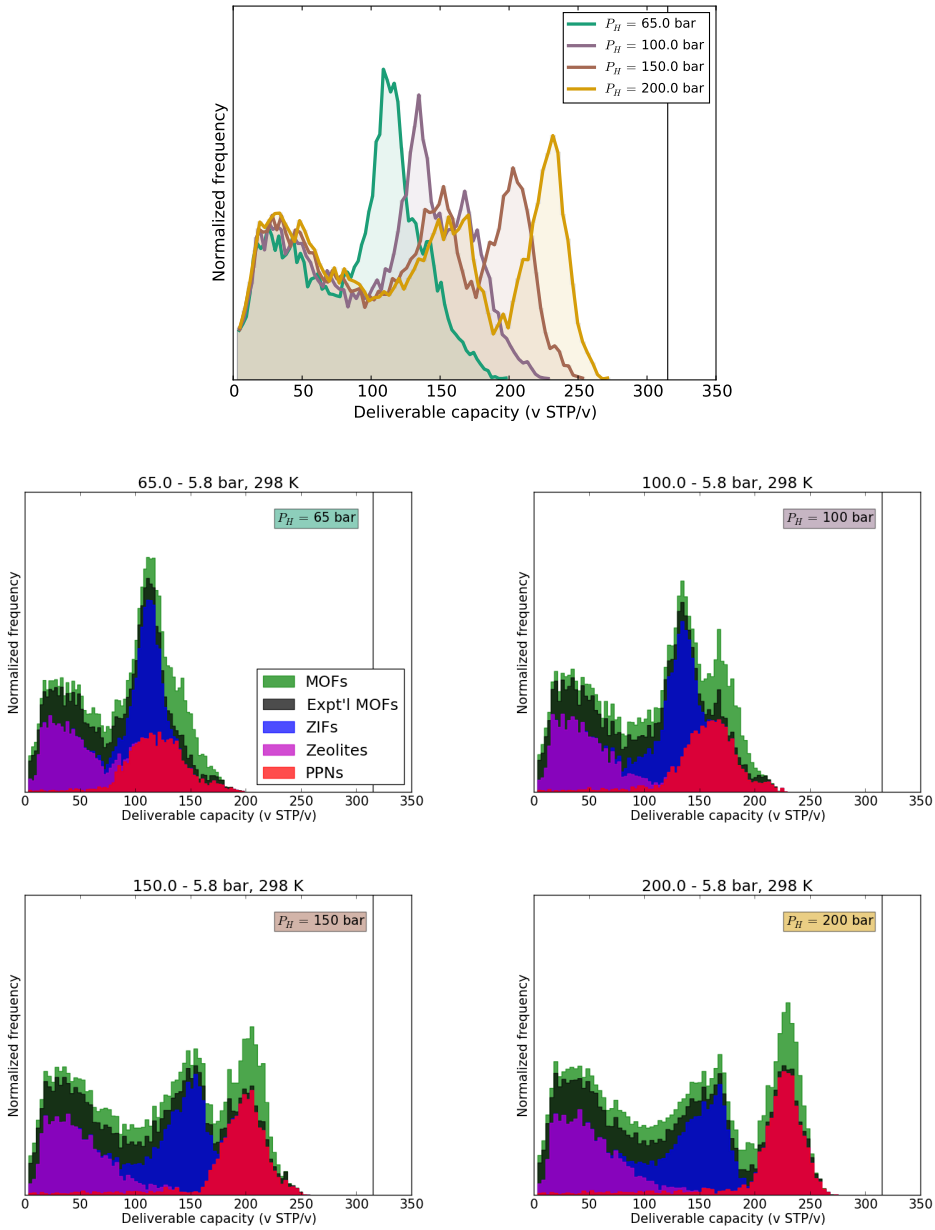


Figure 3.15: Changing the storage pressure: deliverable capacity of methane for materials operating between pressure P_H and 5.8 bar. [Top] Deliverable capacity distributions under differing P_H. [Bottom] For each scenario, the distribution is partitioned into material classes. The vertical line is the ARPA-E target. Reproduced from Ref. [147] with permission from The Royal Society of Chemistry.

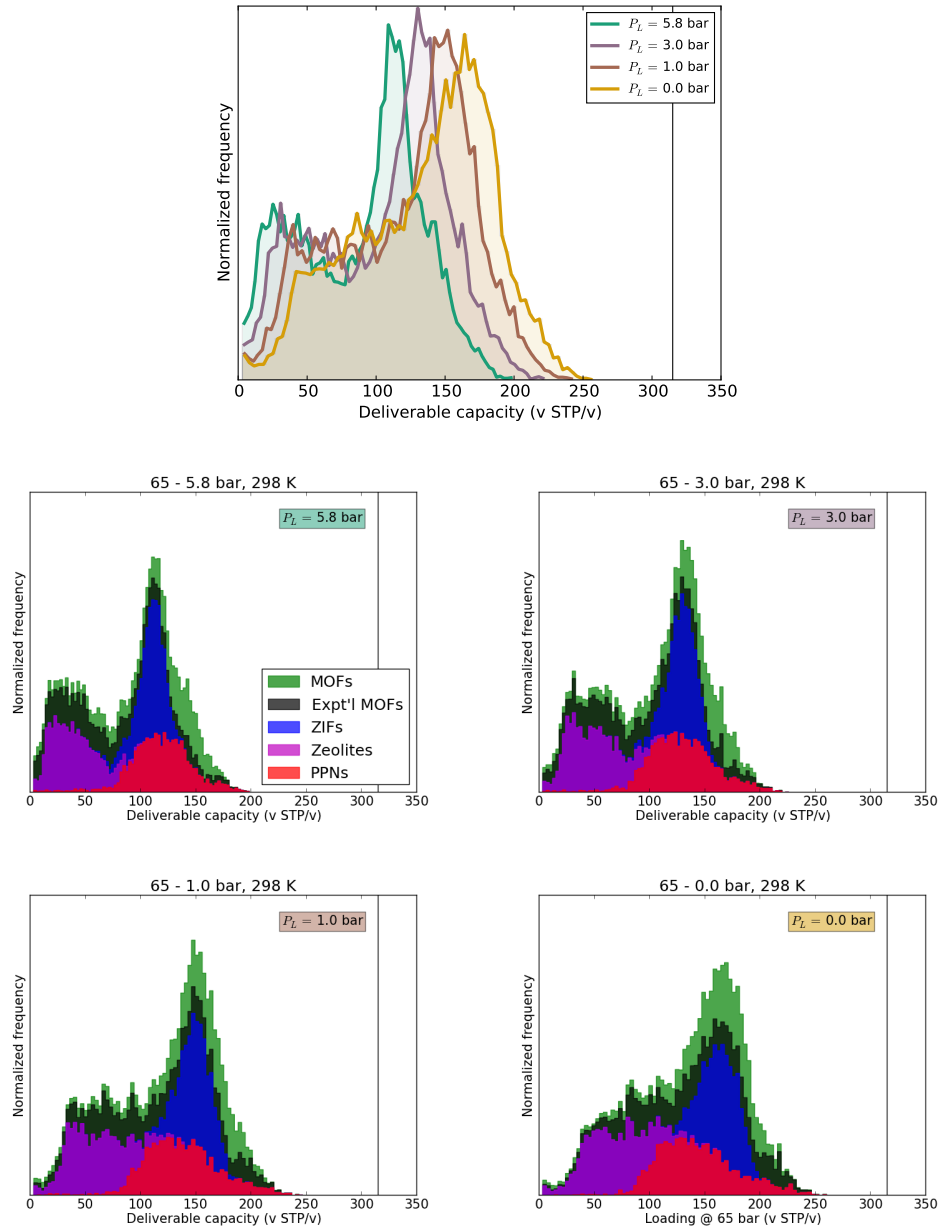


Figure 3.16: Changing the discharge pressure: deliverable capacity of methane for materials operating between pressure 65 and P_L bar. [Top] Deliverable capacity distributions under differing P_L. [Bottom] For each scenario, the distribution is partitioned into material classes. The vertical line is the ARPA-E target. Reproduced from Ref. [147] with permission from The Royal Society of Chemistry.

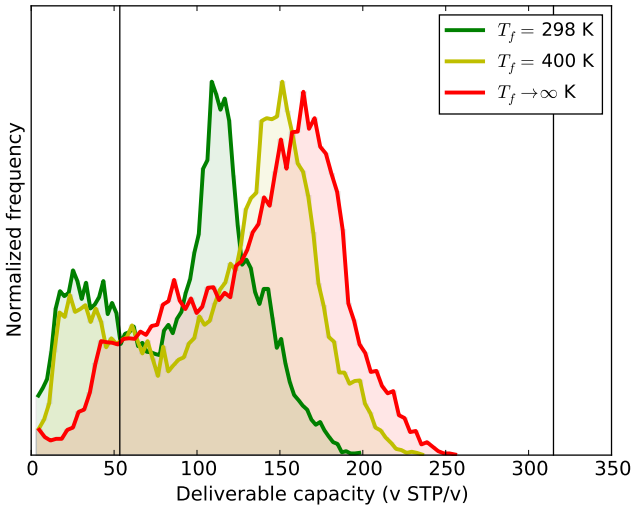


Figure 3.17: Amending the storage conditions and tank design changes the deliverable capacity. The distribution of deliverable capacities between 65 and 5.8 bar for ambient temperature operation (green), when the adsorbent is heated to $T_f = 400$ K when nearing discharge (yellow), and when all residual gas is driven off (red). In the last scenario, the deliverable capacity is equal to the high-pressure loading. Vertical lines denote the deliverable capacity of an empty tank and the ARPA-E target. Reproduced from Ref. [147] with permission from The Royal Society of Chemistry.

Chapter 4

Assessing thermodynamic limits to methane deliverable capacity in nanoporous materials with simple models

The methane deliverable capacity of adsorbent materials is a critical performance metric that will determine the viability of using adsorbed natural gas (ANG) technology in vehicular applications. ARPA-E recently set a target deliverable capacity of 315 cc(STP)/cc that a viable adsorbent material should achieve to yield a driving range competitive with incumbent fuels. However, in Chapter 3, computational screening of hundreds of thousands of materials suggests that the target is unattainable. In this chapter, we aim to determine whether the observed limits in deliverable capacity (200 cc(STP)/cc) are fundamental limits arising from thermodynamic or material design constraints. Our efforts focus on simulating methane adsorption isotherms in a number of systems differing in spatial distributions and energetics of adsorption sites. Our findings suggest that the ARPA-E target is unattainable not due to thermodynamic constraints but due to material design constraints.

This chapter is based on the following publication [192]:

D. Gomez-Gualdron, C. Simon, W. Lassman, D. Chen, R. L. Martin, M. Haranczyk, O. K. Farha, B. Smit, R. Q. Snurr. Impact of the strength and spatial distribution of adsorption sites on methane deliverable capacity in nanoporous materials. *Chemical Engineering Science*. (2016)

4.1 Introduction

For a material to have a high methane (volumetric) deliverable capacity, it needs to have a large number of adsorption sites per volume of material, M [112]. For methane physisorp-

tion, these "sites" are not necessarily small, well defined regions due to the relatively weak adsorbate/adsorbent interactions. Traditionally, adsorption sites are thought of as located on the material (pore) surface, making the density of adsorption sites proportional to the volumetric surface area. Alternatively, it is possible to think of a working definition of an adsorption site as any location that can be occupied by a methane molecule. Thus, if a material has sufficiently large pores, methane molecules can occupy positions far from the pore surface, and thus some "sites" are not related to the volumetric surface area of the material but rather to the void fraction. Under this definition, the density of adsorption sites is equivalent to the methane saturation loading (M) of the material.

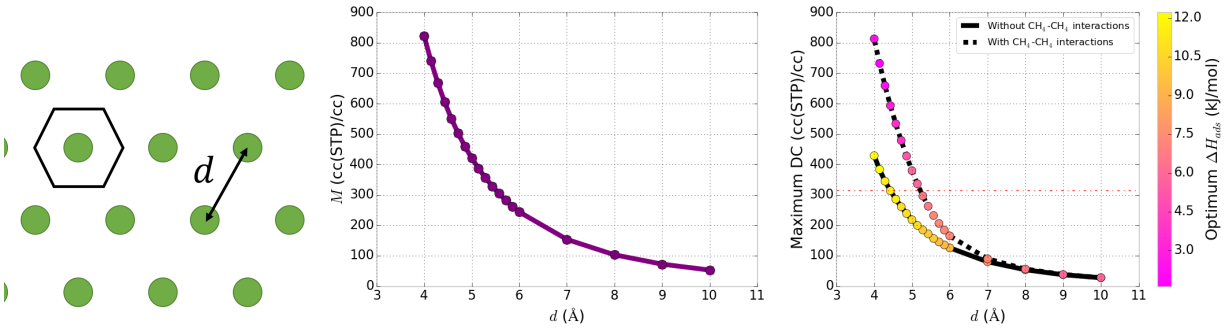
At the storage and discharge pressures (P_H and P_L), different fractions (f_H and f_L) of these sites will be occupied. The deliverable capacity of a material is then $(f_H - f_L)M$. The difference $(f_H - f_L)$ is called the fractional deliverable capacity (FDC), and it depends on the energy of adsorption. Note that sites close to a pore wall provide higher adsorption energies than sites far from any pore wall. Creating walls, and thus sites of higher adsorption energies, requires sorbent atoms, which take up space and limit the saturation loading M. There is, thus, a "material penalty" to create strong adsorption sites. Evidently, the challenge in material design for gas storage applications is to maximize M, i.e. to reduce the material penalty, while simultaneously endowing the adsorption sites with optimal energetics so that FDC is maximized. With this in mind, let us now consider different adsorption scenarios to assess different combinations of M and adsorption energetics.

4.2 Scenario 1: Lattice model

One scenario that can be conceived is a nanoporous material that provides tailored adsorption "pockets" for methane to sit, where each adsorption pocket is capable of hosting one and only one methane molecule. The exact geometrical arrangement and identity of the atoms involved in the creation of these tailored sites determines both the energetics of the methane pockets and the material penalty to create them. The latter, in turn, determines the number of methane pockets per volume of material and thus the saturation loading M.

In a computational simplification of the above scenario, let us consider methane adsorption in a model representing regularly spatially distributed adsorption pockets, where each pocket is denoted by the position of its center and the number of pockets per volume is equivalent to M. In the simplified model, we define the distance between the centers of the pockets as the pocket-pocket distance d, and an increase in d effectively represents an increase in the material penalty. Since we seek to maximize M, we distribute the pockets in space using an FCC (fcu) arrangement, which yields the highest achievable M for a given d. Note, for instance, that the FCC arrangement for d ~4.17 Å (the equilibrium distance of the methane TraPPE model) reproduces the density of solid methane (~690 cc(STP)/cc) [193]. Fig. 4.1a shows a two-dimensional slice for such an arrangement. Fig. 4.1b plots M versus d.

To investigate the effect of energetics, let us endow the pockets with different energies U_0



(a) (b) (c)

Figure 4.1: Methane adsorption in scenario 1. (a) Two-dimensional slice of the FCC (fcu) arrangement of methane pockets (the green dots represent the centers of the pockets), with d being the distance between the centers of adjacent pockets. A two dimensional slice of the contour of a methane pocket is also shown in black. (b) Density of sites or saturation loading, M, versus d. (c) Maximum deliverable capacity (DC) versus pocket-pocket distance d with (top curve) and without (bottom curve) methane-methane interactions. Points are colored according to the optimal heat of adsorption. Reprinted from Impact of the strength and spatial distribution of adsorption sites on methane deliverable capacity in nanoporous materials, D. Gomez-Gualdron, C. Simon, W. Lassman, D. Chen, R. L. Martin, M. Haranczyk, O. K. Farha, B. Smit, R. Q. Snurr, Chemical Engineering Science, (2016), with permission from Elsevier.

and simulate the respective methane deliverable capacity for different d values. See Sec 4.4 for methods. Note that for each d value, we systematically explored different energetics of the pockets until we found the one to maximize the deliverable capacity. Fig. 4.1c shows the maximum deliverable capacity for each d. If methane-methane interactions are neglected, the model is equivalent to and consistent with the Langmuir adsorption model, where the maximum FDC (for storage at 65 bar and delivery at 5.8 bar) is 0.52, obtained with different heats of adsorption depending on d. For the Langmuir model, the maximum deliverable capacity for a given M is thus simply $0.52M$ (indeed the bottom curve in Fig. 4.1c is simply the curve in Fig. 4.1b times 0.52), and the deliverable capacity reaches ∼345 cc(STP)/cc for $d \sim 4.17$ Å (i.e. for M equal to the density of solid methane). On the other hand, if methane-methane interactions between methane molecules in different pockets are considered– a mean interaction energy between two methane molecules in different pockets is considered depending on the distance between the center of the relevant pockets– and the pockets are sufficiently close, the maximum FDC is higher than 0.52, and the maximum deliverable capacity is higher than $0.52M$ as the top curve in Fig. 4.1c shows for low d values. The reason is that methane-methane attractive interactions boost methane loading at the storage pressure while leaving the methane loading at the discharge pressure relatively unaffected.

The top curve in Fig. 4.1c shows that as d becomes smaller (i.e. the pocket centers come

closer) the heat of adsorption needed to maximize the deliverable capacity (i.e. the optimal heat of adsorption) decreases, because the methane-methane attractive interactions compensate for the lower attraction provided by the pockets. From Fig. 4.1c we see that the ARPA-E target is reached for d< 5.3 Å (i.e. $M > 369$ cc(STP)/cc (Fig. 4.1b) if methane-methane interactions are considered. Note that if methane-methane interactions are neglected, the ARPA-E target is reached only for $d < 4.4$ Å (i.e. $M > 609$ cc(STP)/cc, Fig. 4.1b)– recall that 690 cc(STP)/cc is the density of solid methane.

With $d \sim 5.3$ Å, the ARPA-E target is reached with an optimal heat of adsorption of 11.5 kJ/mol (Fig. 4.1c), which yields an FDC of 0.85. Remarkably, as d decreases, FDC approaches one, and thus the deliverable capacity approaches M. So when the pocket-pocket distance d approaches $d \sim 4.17$ Å– and M approaches the density of solid methan–the maximum deliverable capacity approaches ∼690 cc(STP)/cc! The simulations for scenario 1 show a beneficial effect of methane-methane interactions on methane deliverable capacity due to the (unrealistic) confinement of methane molecules to well-defined pockets. It is apparent, however, that such low d or high M is impossible to obtain when actual atoms are involved in the creation of the pockets. Notably, relatively minor increases in d sharply decrease the maximum deliverable capacity. For instance, the maximum deliverable capacity is already at 50% of the ARPA-E target for d values as low as 5.9 Å.

4.3 Scenario 2: Lennard-Jones fluid in a uniform energy field

Contrary to the well-defined adsorption sites that confine methane molecules in scenario 1, the pores of a typical adsorbent material with non-commensurate adsorption provide freedom for methane molecules to spatially organize themselves during adsorption. For instance, in highly porous materials (which also yield the highest M values), adsorbed methane molecules loosely sit in the interaction field created by the pore atoms. Thus, for scenario 2 let us consider methane adsorption onto an artificially created homogeneous interaction field as a simplified limiting case corresponding to no material penalty. Analogous to scenario 1, let us endow the field with tunable adsorption energetics. Here we always consider methanemethane interactions to avoid unrealistic overlap of methane molecules. The overlap is prevented by the repulsive part of the methanemethane interaction potential. See further details on the simulation algorithm used for scenario 2 in Sec. 4.4. Fig 4.2a shows a snapshot from such a simulation.

Fig. 4.2b shows the adsorption isotherms for different background fields corresponding to different heats of adsorption. For comparison, the black dashed isotherm shows the variation of the density of bulk methane gas with pressure, which effectively corresponds to the case with an interaction field featuring a 0 kJ/mol energy of adsorption. Fig. 4.2c shows the deliverable capacities for different heats of adsorption calculated from the isotherms in Fig. 4.2b. In this scenario, deliverable capacities equal to or higher than the ARPA-E target

(315 cc(STP)/cc) were obtained with heats of adsorption in the ∼6.7-10.1 kJ/mol interval. The highest deliverable capacity was 357 cc(STP)/cc for a heat of adsorption of 8.6 kJ/mol.

Remarkably, the optimal heats of adsorption that lead to deliverable capacities surpassing the ARPA-E target are not extraordinarily high and are similar to those observed in highly porous materials synthesized and tested to date [148, 117]. Note, however, that the reported heats of adsorption do not reflect the heterogeneity of adsorption sites in synthesized materials. For instance, locations close to the pore walls have high heats of adsorption, but at other locations such as at the centers of pores, the associated heats of adsorption are much lower. Indeed, the question and associated challenge that derives from this scenario is whether it is possible to synthesize a highly porous material (i.e. low material penalty and thus high M) that can provide a homogeneous interaction field with such an optimal heat of adsorption.

4.4 Methods

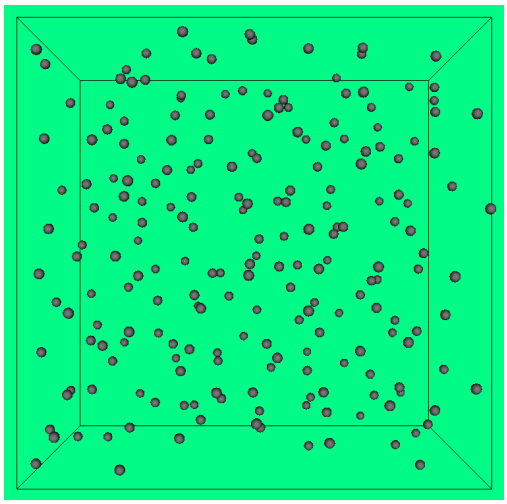
Here, we outline the methods to simulate the adsorption of methane in the model materials in scenarios 1 and 2. In both scenarios, we model methane as a Lennard-Jones sphere with parameters taken from the TraPPE force field [106]. That is, the energy of two methane molecules a distance r apart is modeled with a Lennard-Jones potential, $U_{gg}(r)$. We truncate our Lennard-Jones potential at radius of 12.8 $\mathring{A}$, which allows to mimick a crystal of infinite extent by applying periodic boundary conditions.

$$U_{gg}(r) = \begin{cases} 4\epsilon\left(\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right) & \text{if } r \leq 12.8\ \mathring{A} \\ 0 & \text{otherwise.} \end{cases} \tag{4.1}$$

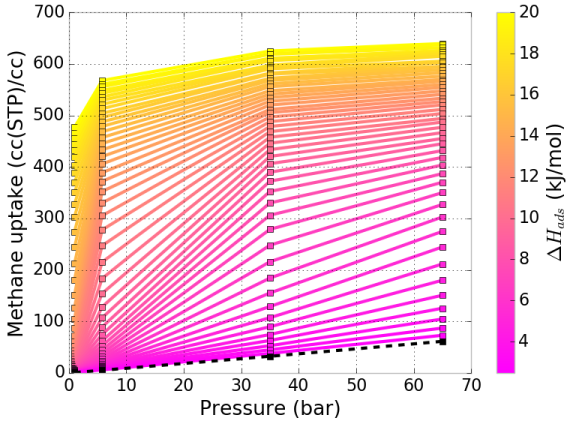
For our grand-canonical Monte Carlo simulations, we convert the pressure of the bulk gas of methane into chemical potential using the Peng-Robinson equation of state.

Scenario 1

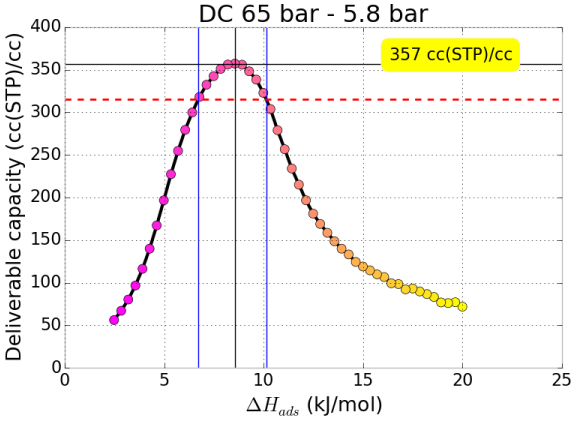
For the model material in Scenario 1, we partition space into a set of voxels. The boundary of each voxel is a rhombic dodecahedral honeycomb; this honeycomb is the Voronoi diagram of an FCC lattice of points. Consequently, we define the *Voronoi centers* of the voxels as this FCC lattice. Each voxel can hold at most one methane molecule. The energy of an adsorbed methane molecule in a voxel is U_0. We simulate the grand-canonical ensemble using a Monte Carlo algorithm under two cases. In the first case, adsorbed methane molecules do not interact with each other at all except through volume exclusion (one methane molecule per site). In the second case, we include interactions between the methane molecules by assigning energies of interaction. We will show that the former case is equivalent to the classic Langmuir adsorption model. The latter case resembles a mean-field approximation of a similar model, which results in the Temkin adsorption isotherm model, a perturbation on the Langmuir model [112].



(a)



(b)



(c)

Figure 4.2: Methane Lennard-Jones fluid in a uniform energy field. Grand-canonical Monte Carlo simulations in a cube of dimension 25 $\AA$. (a) A snapshot of methane positions (represented as spheres) in system at 65 bar with a background energy of U_0 =-12 kJ/mol after equilibration. The methane molecules in this model are free to arrange themselves in continuous space. The green color represents the uniform background energy. (b) Methane adsorption isotherms in a background interaction field. The heat of adsorption associated with the field for each isotherm is denoted by the color of the isotherm according to the vertical color bar. The black dash-line isotherm corresponds to the bulk density of methane at different pressures. (c) Deliverable capacities extracted from panel (b) for different heats of adsorption associated with the background field. Points above the horizontal dashed line exceed the ARPA-E target. The color of the points corresponds with the heat of adsorption in (b). Reprinted from Impact of the strength and spatial distribution of adsorption sites on methane deliverable capacity in nanoporous materials, D. Gomez-Gualdron, C. Simon, W. Lassman, D. Chen, R. L. Martin, M. Haranczyk, O. K. Farha, B. Smit, R. Q. Snurr, Chemical Engineering Science, (2016), with permission from Elsevier.

The model material in Scenario 1 is characterized by the background energy field U_0 and distance d between FCC lattice points, which are the Voronoi centers of the rhobmic dodecahedral voxels.

Generating the lattice

We here outline how to construct an FCC lattice with distance d between points. We start with a primitive unit cell of the FCC lattice, a cube of dimension 1.414 $\mathring{A}$ with four points. In this lattice, the points are a distance 1.0 $\mathring{A}$ apart. To construct a lattice with distance d between points, we simply scale the box length by d. To use the nearest image convention to apply periodic boundary conditions for computing methane-methane interactions, we must replicate this lattice until the cube is at least of length twice of the Lennard-Jones cutoff radius of 12.8 $\mathring{A}$.

For convenience, let $M = M(d)$ be the resulting number of voxels, even though d and U_0 are enough to completely characterize our rhombic dodecahedral honeycomb material.

Case: no methane-methane interactions

We now show that the grand-canonical ensemble in our model material results in a Langmuir isotherm when methane-methane interactions are neglected.

Let V_0 be the volume of each voxel. The single-particle Canonical partition function of an adsorbed methane molecule in a voxel is:

$$q = \Lambda^{-3} \int_{V_0} e^{-\beta U_0} dx = \Lambda^{-3} V_0 e^{-\beta U_0}. \tag{4.2}$$

As we integrate over the entire volume of the voxel, this model assumes that methane molecule is free to explore the entire space inside the voxel, where it experiences an energy U_0.

The Canonical partition of N adsorbed molecules is then:

$$Q(N, M, T) = \binom{M}{N} q^N 1^{M-N}, \tag{4.3}$$

given that each adsorbed molecule is independent, an empty site has a partition function of unity, and there are $\binom{M}{N}$ ways to arrange N indistinguishable particles in the M voxels.

The grand-canonical partition function for our model material in equilibrium with a bath of methane with chemical potential μ is:

$$\Xi(\mu, M, T) = \sum_{n=0}^{M} Q(N, M, T) e^{\beta \mu N} = \sum_{n=0}^{M} \binom{M}{N} [q e^{\beta \mu}]^N. \tag{4.4}$$

Using the binomial theorem:

$$\Xi(\mu, M, T) = (1 + q e^{\beta \mu})^M. \tag{4.5}$$

We obtain the adsorption isotherm in our model material by computing the average particle number in the system of voxels, $\langle N \rangle$:

$$\langle N \rangle = \frac{\partial \log \Xi}{\partial(\beta\mu)} = M \frac{qe^{\beta\mu}}{1 + qe^{\beta\mu}}. \tag{4.6}$$

Using the definition of fugacity f, (units: pressure), this becomes the familiar Langmuir isotherm:

$$\langle N \rangle = M \frac{Kf}{1 + Kf}, \tag{4.7}$$

with the Langmuir constant related to our model material by $K := \beta V_0 e^{-\beta U_0}$.

Case: methane-methane interactions included

Here, the energy of an adsorbed methane molecule in voxel i is:

$$U_i = U_0 + \sum_{j!=i} I(j) U_{gg}(r_{ij}), \tag{4.8}$$

where r_{ij} is the distance between the Voronoi centers of voxels i and j and $U_{gg}(r)$ is the truncated Lennard-Jones potential in eqn 4.1. The function $I(j)$ is an indicator function that is unity if voxel j is occupied and zero otherwise. Thus, in this model, the energy of interaction between two methane molecules in two different voxels is assigned by taking their positions as fixed at the Voronoi centers of the voxels. This is an approximation since eqn 4.2 presumes that each molecule freely explores the space inside a given voxel.

We simulate the grand-canonical ensemble in our model material with the energy prescription in eqn 4.8 since we could not find an analytical solution. Our two Monte Carlo trials for this model are (1) attempt to add a particle (2) attempt to remove a particle. For (1), we choose a random lattice site. If the lattice site is occupied, we do nothing; if the lattice site is empty, we accept this Monte Carlo trial particle insertion with probability p_{add}. For (2), we choose a particle at random and accept the proposal to delete it with probability p_{remove}. Of course, if there are zero molecules in the system, we do nothing. Next, we derive the acceptance probabilities p_{add} and p_{remove} using the Metropolis-Hastings algorithm.

We can characterize a microstate of our system by a vector x of dimension M; entry i of x, x_i, is unity if voxel i is occupied and zero if it is empty. The probability of observing a particular microstate x with N particles ($\sum_i x_i = N$) in the grand-canonical ensemble (see eqn. 4.4) is:

$$p(x) = (V_0 e^{-\beta U(x)} \beta f)^{\sum_i x_i}, \tag{4.9}$$

where $U(x)$ is the energy of the system with microstate x, which can be calculated by the energy prescription in eqn 4.8.

The probability of attempting to insert a particle in a given voxel is $\alpha(N \to N+1) = \frac{1}{2}\frac{1}{M}$ since one of M voxels is chosen at random and we choose an insertion attempt with

probability 1/2. The probability of attempting to delete a given particle is $\alpha(N \to N-1) = \frac{1}{2}\frac{1}{N}$ since we choose one of the N particles at random.

Consider a Monte Carlo trial move to add a particle, changing the current microstate x to a new microstate x'. Our constraint here is thus $\sum_i x_i = N$ and $\sum_i x'_i = N+1$. For detailed balance, the Metropolis-Hastings algorithm enforces:

$$p_{add} = \min\left(1, \frac{\alpha(N+1 \to N)}{\alpha(N \to N+1)} \frac{p(x')}{p(x)}\right). \tag{4.10}$$

From the above discussion,

$$\frac{\alpha(N+1 \to N)}{\alpha(N \to N+1)} = \frac{\frac{1}{2}\frac{1}{N+1}}{\frac{1}{2}\frac{1}{M}} = \frac{M}{N+1}, \tag{4.11}$$

and, using eqn 4.9,

$$\frac{p(x')}{p(x)} = V_0 \beta f e^{-\beta(U(x)-U(x'))}. \tag{4.12}$$

Thus, our acceptance probability for a particle addition is:

$$p_{add} = \min\left(1, \frac{V\beta f e^{-\beta(U(x)-U(x'))}}{N+1}\right), \tag{4.13}$$

where $V = V_0 M$ is the volume of the system.

Using the same procedure for the particle deletion trial, going from microstate x to x':

$$p_{remove} = \min\left(1, \frac{N e^{\beta(U(x)-U(x'))}}{V\beta f}\right). \tag{4.14}$$

For these Monte Carlo trials, we need to calculate $U(x) - U(x')$, where x and x' differ only by one entry. Thus, we only need to calculate the energy of a particle in the particle voxel that we have randomly chosen for the insertion / random particle that we have randomly chosen for the deletion.

We perform $750M$ Monte Carlo trials for each simulation and use $150M$ trials for burn cycles/equilibration (scaling it with the number of voxels, M).

Scenario 2

Here, we simulate the grand-canonical ensemble in a volume V with a uniform, background energy U_0. We simulate in a cubic volume of dimension twice the Lennard-Jones cutoff radius, 12.8 Å, so that the nearest image convention can be applied for periodic boundary conditions.

The energy of particle i in the system is:

$$U_i = U_0 + \sum_{j!=i} U_{gg}(r_{ij}), \tag{4.15}$$

where U_{gg} is the Lennard-Jones potential in eqn 4.1.

We perform grand-canonical Monte Carlo simulations, where the following three Monte Carlo trials are attempted each with probability $\frac{1}{3}$: (1) particle insertion, (2) particle deletion, (3) particle translation. For (1), a particle is inserted at a uniformly random position in the simulation volume. For (2), a particle is chosen at random for removal (if there are no particles, we do nothing). For (3), we perturb each coordinate of the particle by a uniformly distributed number in $[-.05, .05]$ $\mathring{A}$.

The acceptance rules for this standard Grand-canonical Monte Carlo simulation are in Ref. [110].

Chapter 5

Machine learning-accelerated screening of the Nanoporous Materials Genome for Xe/Kr separations

Accelerating progress in the discovery and deployment of advanced nanoporous materials relies on chemical insight and structure property relationships for rational design. Because of the complexity of this problem, trial-and-error is heavily involved in the laboratory today. A cost-effective route to aid experimental materials discovery is to construct structure models of nanoporous materials *in silico* and use molecular simulations to rapidly test them and elucidate data-driven guidelines for rational design. For example, highly-tunable nanoporous materials have shown promise as adsorbents for separating an industrially relevant gaseous mixture of xenon and krypton. In this work, we characterize, screen, and analyze the Nanoporous Materials Genome, a database of ca. 670,000 porous material structures, for candidate adsorbents for xenon/krypton separations. For over half a million structures, the computational resources required for a brute-force screening using grand-canonical Monte Carlo simulations of Xe/Kr adsorption are prohibitive. To overcome the computational cost, we used a hybrid approach combining machine learning algorithms (random forests) with molecular simulations. We compared the results from our large-scale screening with simple pore models to rationalize the strong link between pore size and selectivity. With this insight, we then analyzed the anatomy of the binding sites of the most selective materials. These binding sites can be constructed from tubes, pockets, rings, or cages and are often composed of non-discrete chemical fragments. The complexity of these binding sites emphasizes the importance of high-throughput computational screenings to discover new materials. Interestingly, our screening study predicts that the two most selective materials in the database are an aluminophosphate zeolite analogue and a calcium based coordination network, both of which have already been synthesized but not yet tested for Xe/Kr separations.

This chapter is regarding the following publication [95]:

C. Simon, R. Mercado, S. K. Schnell, B. Smit, and M. Haranczyk. What Are the Best Materials To Separate a Xenon/Krypton Mixture? *Chemistry of Materials*. (2015) 27 (12), 4459-447.

5.1 Introduction

Within the Nanoporous Materials Genome [194], we explore the chemical space of possible nanoporous materials by constructing materials *in silico*; this is achieved by combining different molecular building blocks in various topologies, much like snapping together Lego blocks or K'NEX. In the Nanoporous Materials Genome Center [194], over 670,000 structures have been constructed to date: libraries of predicted MOFs [121], all-silica zeolites [119, 120], PPNs [159], ZIFs [158], and COFs [195] as well as data sets of zeolites [196] and MOFs [160] that have been experimentally synthesized. Henceforth, we will refer to this set of materials as the *Nanoporous Materials Genome*. We can rapidly and cost-effectively test each of these materials for a given application using mathematical models and molecular simulations. In this chapter, we screen the libraries of materials in the Nanoporous Materials Genome for high-performing materials for room temperature xenon/krypton separations relevant to replacing a cryogenic distillation used to produce pure xenon and kyrpton for their myriad of applications (see Sec. 1.1).

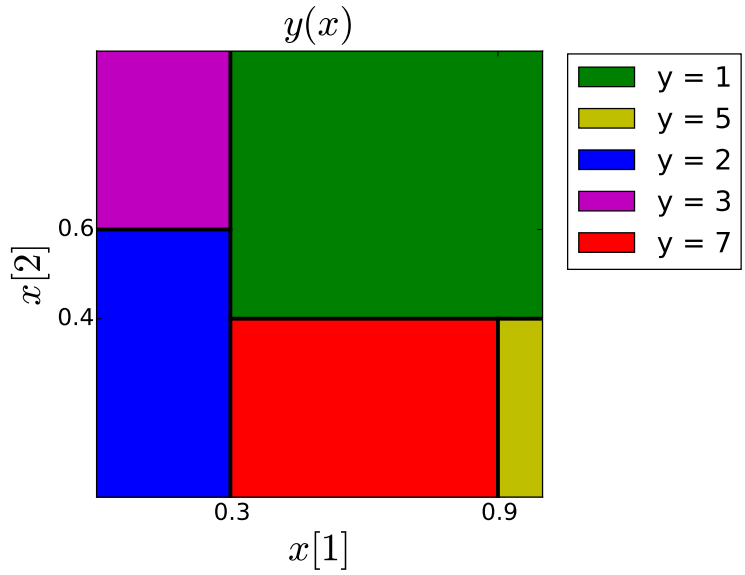
A brute-force high-throughput screening for Xe/Kr separations has been done on the hypothetical MOF dataset [197, 76] and a set of experimental MOFs [198]. We extend these studies to include PPNs, ZIFs, zeolites, and COFs. With such a rapid increase in the number of materials, the required computational resources will become a bottleneck for these brute-force screening techniques. Therefore, it is important to develop algorithms that allow us to screen these databases more efficiently. In this work, we utilize machine learning algorithms to predict performance from structural descriptors. Simple structural descriptors, such as the surface area and void fraction, can be quickly computed. Some structural descriptors, such as crystal density and pore size, are independent of the application and conditions studied and thus can be computed once, stored as a material property, and utilized in several different contexts. These structural descriptors are often correlated to and thus predictive of material performance, underlying the concept behind quantitative structure-property relationship (QSPR) [167] modeling. As such, the array of structural descriptors of a material serves as a representation/fingerprint of the material in a high-dimensional space, a *feature vector*. Machine learning techniques, such as random forests [199], support vector machines [200], and neural networks [201], can then be trained to predict material performance from the feature vector (for a review, see Ref. [167]). With this pre-screening technique, we focus the molecular simulations on only a fraction of the total number of structures, the most promising ones, and avert simulating adsorption in materials that are most likely poor performers.

In the realm of nanoporous materials, Fernandez and coworkers [202] showed that machine learning techniques can predict methane uptake in MOFs from geometric descriptors. In another work, Fernandez and coworkers [203] illustrated the classification of MOFs as high- or low- performing for CO_2 uptake using support vector machines. These works illustrated machine learning as a potential method to be utilized in high-throughput screenings of materials.

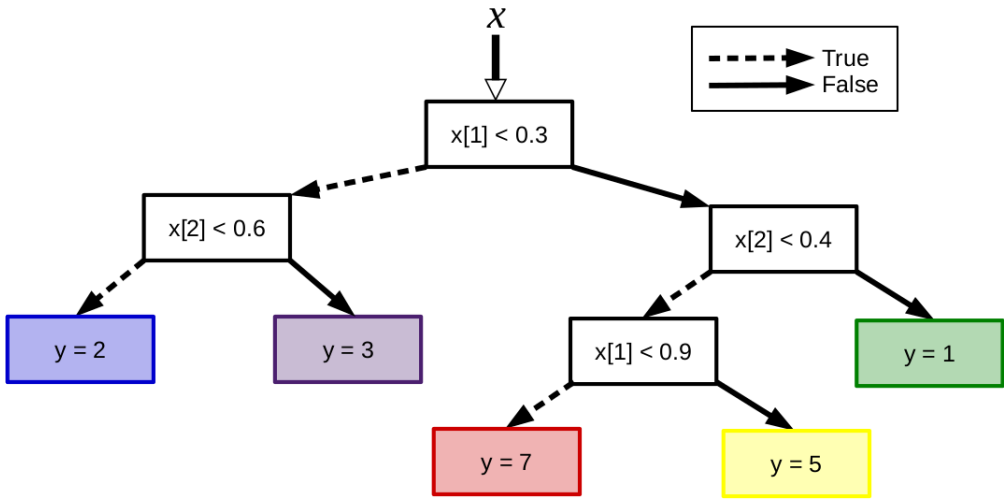
We utilize a random forest [199], an ensemble [204] of decision trees, to predict the Xe/Kr separation performance from structural descriptors. For a given material, we predict the selectivity by running the corresponding vector of descriptors through a set of decision trees. Figure 5.1 illustrates a decision tree regressor. A decision tree is grown using a training set of materials with known selectivities. However, if we train a single decision tree on the entire training set, the prediction is not very accurate, as a single decision tree is highly prone to overfitting [205]. Overfitting leads to an accurate representation of the training set but a high variance prediction for materials that are not in the training set, thus yielding poor generalization error [205]. Much better accuracy on unseen data can be obtained by training many trees – a forest of trees – on randomly selected subsets of the training set, decorrelating the trees from each other. The prediction of the selectivity in a *random forest* is then the average vote among each tree in the forest. The reason that an ensemble of decorrelated trees gives a greater accuracy than a single tree, a Wisdom of the crowds [206] effect, is as follows. Each tree is allowed to overfit its own training data. Thus each tree in the forest accurately describes its own training data, and so we retain a low bias prediction, i.e., the average does not systematically shift away from the true average. While each tree has a high variance prediction on unseen data on its own, by taking the average of each tree in the forest, we reduce the variance, leading to a low-bias, low-variance prediction. Consequently, random forests have been shown to be remarkably accurate [207] in a wide variety of problems. For example, in computer vision, by training a random forest on a set of pictures of humans in different positions, Microsoft successfully implemented random forests to accurately label body parts from video in real-time for the Kinect in the XBox video game console [208]. Similarly, by feeding a random forest training examples of porous material structures and their Xe/Kr separation performance as predicted by molecular simulation, we show that random forests can successfully expedite the discovery of high-performance porous materials.

The success of machine learning in screening materials is predicated on having highly-predictive structural descriptors. In this work, we introduce a new descriptor, the Voronoi energy E_v, that takes into account both geometrical structural information and the energetics of the guest-host interaction. The Voronoi energy is the average energy of a xenon atom at the accessible Voronoi nodes that represent the pore topology; we find this descriptor to be highly predictive of Xe/Kr separation performance.

With the large volume of data from our hybrid machine learning- molecular simulation screening approach, we uncover guidelines for synthesizing a highly selective material. In particular, we rationalize the strong relationship between performance and pore size by considering two simple caricatures of spherical and cylindrical adsorption pockets. With this



(a)



(b)

Figure 5.1: Decision tree regressor illustration for the function $y(x)$, where x is in two-dimensions. (a) A decision tree regressor partitions the feature space into rectangles and models $y(x)$ as a constant in each one. The plane shows the two-dimensional feature space, and the color indicates the value $y(x)$ assigned to a feature vector x that belongs to that region. (b) The decision tree shown takes a feature vector x at the root node (top) and represents $y(x)$ depicted in (a). At each node, a binary decision is made by comparing a component of x and a threshold. The data point is sent to the left (right) if condition in the box is true (false). The terminal/leaf nodes assign a value of y. The colors of the leaf nodes correspond to the regions represented in (a). Reprinted with permission from Ref. [95]. Copyright 2015 American Chemical Society.

insight, we then analyze in detail the top candidates from the Nanoporous Materials Genome for Xe/Kr separations. We identify the characteristic chemistry that makes a binding site selective for these type of separations. This molecular insight provides some guidelines for a possible rational design of these materials.

5.2 Methods

The workflow of our method to screen ca. 670,000 structures in the Nanoporous Materials Genome is illustrated in Figure 5.2 and consists of the following six steps. 1. Characterization: we compute the feature vector of each material, whose components are seven quickly-computed structural descriptors. 2. Selection of the training set: we use a diversity selection algorithm [209] to ensure that our training set of materials adequately covers our seven-dimensional feature space. 3. Label the training set: we perform binary grand-canonical Monte Carlo simulations to compute the Xe/Kr uptake in the training set. 4. Training of the forest: we use these training examples to grow a forest of decision tree regressors to predict Xe/Kr separation performance from the structural descriptors. 5. Pre-screening: we run the structural descriptors of the remaining materials through the trained forest of decision tree regressors. 6. Materials discovery: if the forest predicts the material to be promising for Xe/Kr separations, we run grand-canonical Monte Carlo simulations to refine the prediction.

Molecular simulations

The potential energy of interaction of an adsorbate (Xe or Kr) with the atoms of the material and other adsorbates is modeled with pairwise Lennard-Jones potentials. The Lennard-Jones parameters for Kr-Kr and Xe-Xe interactions are taken from Ref [94]. The parameters for MOF, ZIF, COF, and PPN atoms are taken from the Universal Force Field [108]. For zeolites, the Lennard-Jones parameters for oxygen are back-calculated using Lorenz-Berthelot mixing rules from a force field for krypton developed specially for zeolites by Talu and Myers [210]; in this force field, Kr-Si interactions are not explicitly considered and are instead embedded in the Kr-O parameters. The cross-interaction Lennard-Jones parameters of Xe/Kr with atoms of the framework are then calculated from the Lorenz-Berthelot mixing rules.

We truncate the Lennard-Jones potentials at 12.5 Å and approximate interactions beyond the cutoff distance to be zero. We implement periodic boundary conditions to mimic an infinite crystal and assume the structures to be rigid, which induces some errors since some structures are flexible upon gas adsorption [174, 211].

We simulate the equilibrium Xe and Kr uptake in a material with the grand-canonical Monte Carlo algorithm [110] for a two-component system. We consider a 20/80 molar Xe/Kr mixture at 298 K and 1 atm of total pressure. The Monte Carlo moves are particle exchange (insertion/deletion), particle translation, and particle identity change with probabilities 0.6, 0.2, and 0.2, respectively. We utilize a parallel algorithm written for graphics processing

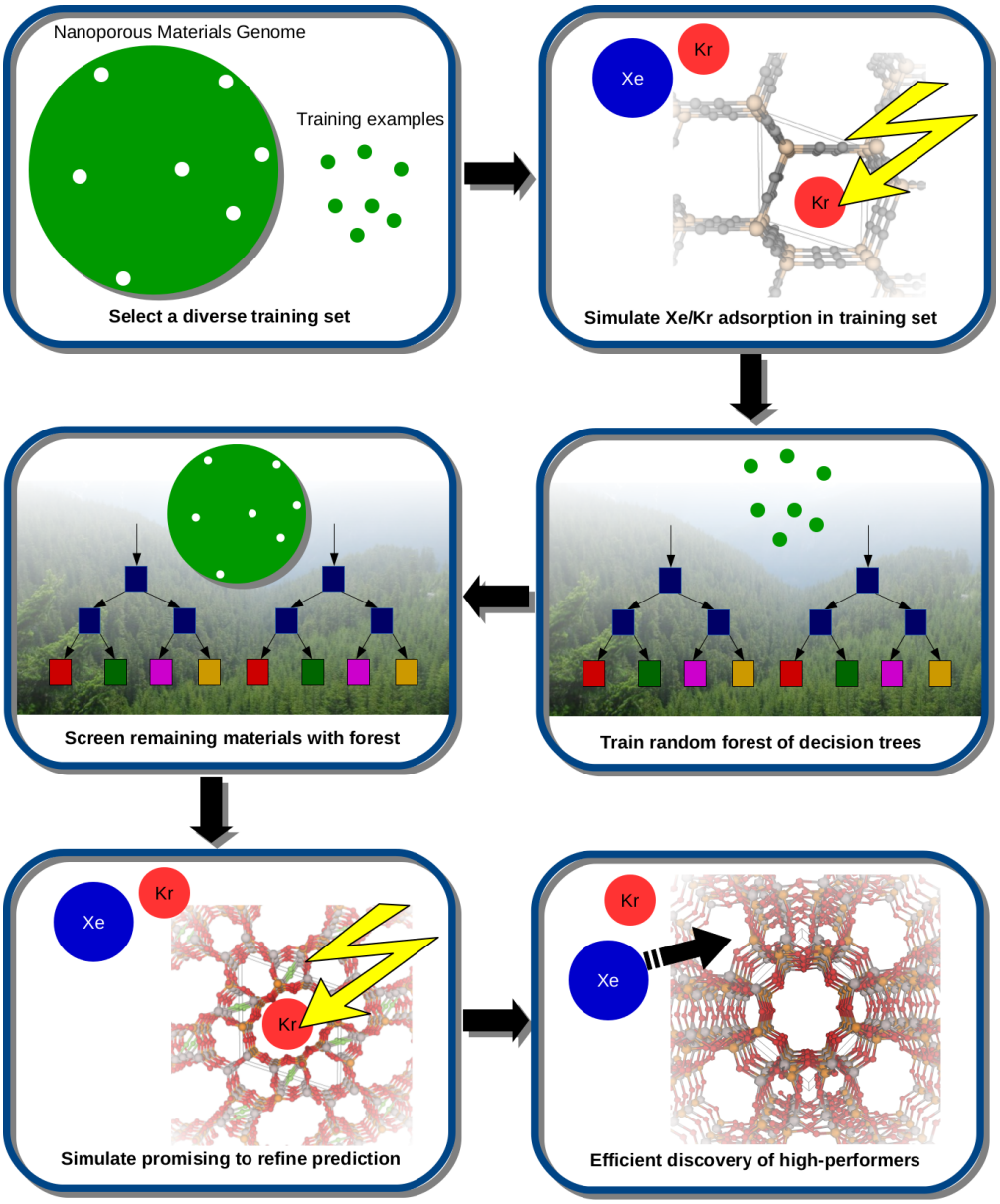


Figure 5.2: Screening procedure. We first select a mathematically diverse training set from the structures in the Nanoporous Materials Genome and simulate the adsorption of Xe/Kr in this training set. Then, we train a random forest of decision tree regressors to predict performance from structural descriptors. We run the structural descriptors of the remaining materials through the trained random forest to identify the most promising materials. We simulate Xe/Kr adsorption in the most promising set to refine the prediction. Note that some high performing materials are discovered during training. Reprinted with permission from Ref. [95]. Copyright 2015 American Chemical Society.

units (GPUs) [111] to pre-compute potential energy interpolation grids (ca. 0.1 Å grid point spacing) to speed-up our simulations.

As evidence that our force field can reasonably predict the experimental Xe/Kr uptake in materials, simulated adsorption isotherms of Xe and Kr match well with the experimental Xe/Kr uptake in metal-organic frameworks IRMOF-1 and IRMOF-2x, for x =Cl, Br, F, I [83], and Silicalite [212]. With a generic force field such as the UFF, it is particularly difficult to predict uptake in MOFs with open metal sites [104]. Still, for the purposes of a high-throughput screening, where we seek to identify the top candidates, the UFF provides a similar *ranking* of materials compared to highly accurate but more computationally expensive *ab initio* force fields [213].

We now provide evidence that the force field used in this work can reasonably predict the Xe/Kr uptake in nanoporous materials. We took experimental Xe and Kr uptake data in IRMOF-1 and the IRMOF-2x series, for $x \in \{\text{Cl, Br, F, I}\}$ from Ref [83] and Kr uptake data in Silicalite from Ref. [212]. Crystal structures for IRMOF-1 and -2 and Silicalite were obtained from Refs [214, 215]; solvent/template atoms were removed; halogens in IRMOF-2x were all assigned the same positions for simplicity. We display the comparison between experimental and simulated noble gas uptake in Figure 5.3. For IRMOF-1 and Silicalite, we obtain excellent agreement[1]– also for Xe adsorption in the IRMOF-2x series. The agreement for Kr adsorption in the IRMOF-2x series is not perfect, but the sorting of uptake depending on the halogens is preserved.

Ref. [104] measured xenon and krypton uptake in ten MOFs with open metal sites, HKUST-1, PCN-14, x-MOF-74, and NOT-10x, and compared the uptake to that predicted from molecular simulations using the Universal Force Field [108]. The agreement is reasonable in some cases, but not others. The Universal Force Field may have difficulty describing the interaction of gas molecules with open-metal sites found in many MOFs [216, 176], likely because the force field was not parameterized using data representative of such a chemical environment. Still, for the purposes of a high-throughput screening, where we seek to identify the top candidates, the UFF provides a similar *ranking* of materials compared to highly accurate but more computationally expensive *ab initio* force fields [213].

Some pores in nanoporous materials are large enough to fit an adsorbate molecule, but are inaccessible in a periodic system due to prohibitively high energy barriers to diffuse into the pore. In our molecular simulations, we identify such pockets automatically in a high-throughput manner and refrain from sampling adsorbate configurations in these regions using an inaccessible pocket-blocking algorithm in Ref. [129].

We first compute the potential energy of a Xe and Kr adsorbate in the structure on a three-dimensional grid of points superimposed on the unit cell. With this grid, we identify regions that are inaccessible to an adsorbate as follows.

We define a pocket of space as inaccessible if the energetic barrier for a Xe/Kr to percolate into this region is greater than $15k_BT$ (k_B is the Boltzmann constant; T = 298 K is the

[1]The agreement for Silicalite is somewhat trivial, given that the force field for zeolites [210] was developed on the basis of fitting to this experimental data

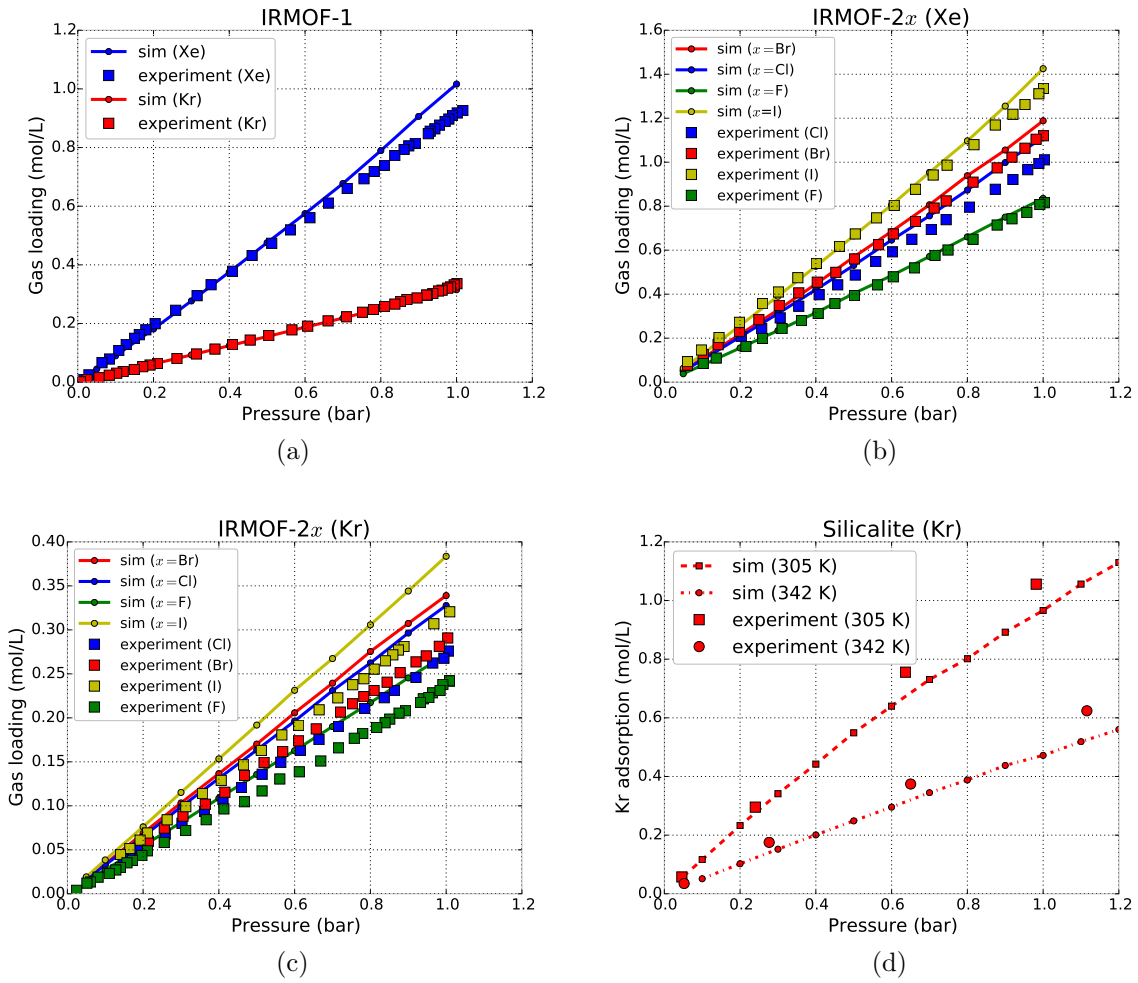


Figure 5.3: Force field validation. Shown is a comparison between experimental equilibrium uptake data and our simulated uptake. (a) Kr and Xe adsorption in IRMOF-1 at 292 K (b) Xe adsorption in IRMOF-2x at 292 K (c) Kr adsorption IRMOF-2x at 292 K (d) Kr adsorption in Silicalite at 305 and 242 K. Reprinted with permission from Ref. [95]. Copyright 2015 American Chemical Society.

Table 5.1: Structural descriptors used to construct a feature vector for each material.

Descriptor	Symbol	Description
Void fraction [unitless]	ϵ_v	fraction of material that is free volume
Crystal density [kg/m^3]	ρ	mass of crystalline material per volume
Largest free sphere diameter [Å]	D_f	largest sphere to percolate through material
Largest included sphere diameter [Å]	D_i	largest sphere to fit inside the material
Accessible surface area [m^2/cm^3]	a	probe-accessible surface area along pore walls
Surface density [kg/m^2]	ρ_s	mass of atoms per acccessible surface area $= \rho/a$
Voronoi energy [kJ/mol]	E_v	average energy of Xe at the accessible Voronoi nodes

temperature). The time scale to diffuse across an energetic barrier of this size is large [129]. To identify such pockets, we begin by defining each point on the potential energy grid as inaccessible if the potential energy of an adsorbate there is greater than $15k_BT$ and accessible otherwise. We then use a flood fill algorithm to identify and label *segments* in the unit cell of the crystal. From here, we build a periodic graph representation [217] of the connections between segments in the periodic system. Walking through this graph starting at each segment, we check that an adsorbate can travel from one unit cell to another, in which case the segment is a *channel*. If not, the segment is an *inaccessible pocket*, and consequently we overwrite the energies in this region of the pore space with large values to deem these inaccessible regions as effectively so during our simulations. [129]

Structural descriptors

Table 5.1 summarizes the structural descriptors used to form the feature vector of the material. The first six descriptors are established geometric structural descriptors. The last descriptor is, to our knowledge, a new descriptor that we invented that incorporates both geometrical structure information and guest-host interaction energy.

While one can imagine many other structural descriptors, those in Table 5.1 were easily computed and, as we will later see, are satisfyingly predictive of Xe/Kr separation performance for our screening.

Established geometric structural descriptors

To compute geometric structure descriptors, we use open-source software Zeo++ [166], which models the accessible void space space inside a porous material with a periodic Voronoi network. The framework atoms as well as Xe/Kr are modeled as hard spheres with radii adopted from the Cambridge Structural Database [218, 219] and the zeros of the Lennard-

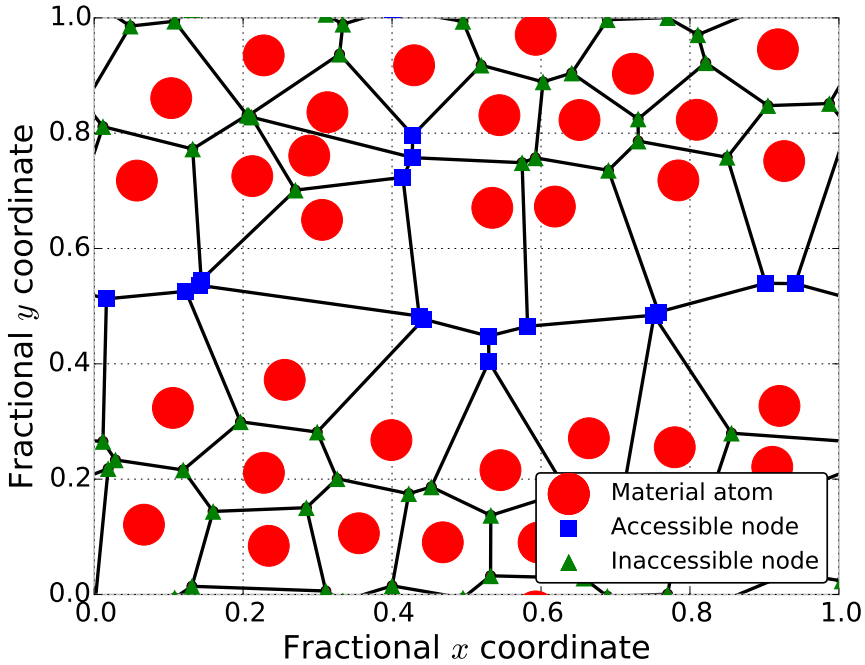


Figure 5.4: Voronoi network model of void space (2D caricature). The unit cell of a toy material is shown. Red circles represent atoms of the material; accessible and inaccessible Voronoi nodes are blue squares and green triangles, respectively. The black lines are the edges in the periodic Voronoi graph that models the void space. The descriptor E_v in eq 5.1 is the average potential energy of a xenon atom adsorbed at the accessible Voronoi nodes. Reprinted with permission from Ref. [95]. Copyright 2015 American Chemical Society.

Jones potentials in Figure 1.4, respectively. The periodic Voronoi network is calculated by Voro++ [220]. Figure 5.4 shows a two-dimensional sketch of the Voronoi network in a toy material. From this periodic graph representation of the pore space, we can identify regions that are accessible to Xe and Kr and calculate the diameter of the largest included and free spheres (D_i and D_f) [221], accessible surface area (a), and void fraction (ϵ_v) [166]. The crystal density (ρ) is easily computed from the crystal structure. Another descriptor we use is the surface density ρ_s, i.e. the mass of material per surface area, $\rho_s = \rho/a$ [kg/m^2].

The Voronoi energy descriptor

Gas uptake at low pressures can be dominated by the strongest adsorption sites due to the exponential contribution of the energy in the Henry coefficient [110]. The established geometric structural descriptors above have difficulty capturing such "sweet spots" [143]. In an attempt to capture such strong binding pockets that are important for Xe/Kr adsorption, we included a descriptor E_v [kJ/mol] that is the average energy of a xenon atom among the

nodes of the xenon-accessible Voronoi network:

$$E_v := \frac{1}{N}\sum_{i=0}^{N} E(v_i), \tag{5.1}$$

where $E(v)$ is the potential energy of a xenon atom at a point v in the pore space of the material as calculated by our force field. The set of points $\{v_i\}$ for $i = 1, 2, ..., N$ are the accessible Voronoi nodes in the Voronoi network used to model the xenon-accessible pore space; the Voronoi network conveniently and automatically identifies points that are located at the 'center' of the accessible pore landscapes (see Figure 5.4). Thus, the descriptor E_v can be viewed as a biased sampling of the potential energy of an adsorbate in the pores. In a material with pores that are near the optimal size for xenon, the energetic minimum will be at the center of the pores, where Xe can maximally interact with all surrounding atoms (see Figure 5.18e later). As the Voronoi energy E_v is a measure of energy at the 'center' of the pore, our intuition is that E_v will capture these strong binding regions in materials with nearly optimal pore diameters. Local structural features, such as a pocket, can also be identified automatically with the Voronoi network, for example, the pocket in Figure 5.4. In contrast, the energetic minimum in materials with large pores will be near the surface instead of the center; however, materials with large pores are poor for Xe/Kr separations (see Figure 5.10 later), and D_i should take responsibility to rule out materials with large pores in the random forest algorithm. Later, we will find that our new descriptor E_v is highly predictive of separation performance.

Selecting a diverse set of materials

To ensure that the training data set adequately represents the chemical space [32] of crystal structures in our dataset and for the decision forest that we train to be accurate, the training data set must include a *diverse* set of materials. To pose a mathematical definition of diversity, we represent each material in a high dimensional space with a *feature vector* x and utilize the Euclidean distance as a notion of similarity. For diversity selection, we define the feature space *a priori* as $x = [\rho, a, D_i, \epsilon_v]$ and normalize each component of the vector to range from zero to unity. To select the diverse set, we use an algorithm in Ref. [209], which we outline here. We selected a diverse subset of ca. 15,000 materials to serve as training data (3,640 experimental MOFs, 2,750 hMOFs, 1,000 hPPNs, 2,319 hZeolites, 194 IZA zeolites, 3,757 hZIFs, and 834 hCOFs ['h' indicates hypothetical or predicted materials that have not been synthesized]).

For diversity selection, we used an algorithm proposed in Ref. [209], which we outline here. The feature vector for material i is x_i. We use the Euclidean distance $d(x_i, x_j) := ||x_i - x_j||^2$ in feature space as a notion of similarity of two materials i and j. This problem is generally known as the maximum diversity problem (MDP) [222, 223] and is NP-hard, but the heuristic search procedure in Ref [209] is suitable for our purposes of selecting a training set of structures.

The algorithm for selecting the diverse set in Ref. [209] is sequential. Let X denote the current, diverse set and Z denote the set of all materials (x_i). The algorithm is[2]:

- Randomly choose an initial point $x_i \in Z$ and set $X = \{x_i\}$.
- While $|X| < n$:
  - Choose a point $x_* \in Z - X$ such that:

$$x_* = \underset{x_j}{\operatorname{argmax}} \left(\min_{j,k} d(x_j, x_k) : x_j \in Z - X, x_k \in X \right) \tag{5.2}$$

  - Set $X \leftarrow X \cup \{x_*\}$

We start by randomly assigning a single material to the diverse set. We sequentially add materials to this set by, for all materals outside of the diverse set, computing the distance to the closest material in the diverse set. We then choose to add the material to the diverse set that has the largest distance to the respective closest material in the diverse set.

A similar algorithm is used to select a diverse set of nanoporous materials in Ref.[225]. We perform the diversity selection within each material class separately to control the number of materials from each family of materials that we use in our training set.

Forest of decision tree regressors

With a random forest of decision tree regressors, we can quickly predict the Xe/Kr separation performance from the seven-dimensional feature vector x_i of a material. We will characterize the performance of a material by the selectivity in eq 5.4, but we will denote the performance score here as y_i for a general discussion. Using the training data set (x_i, y_i) for $i = 1, 2, ..., N$, we show here how to train a forest of decision tree regressors that serves as our regression $y(x)$.

A decision tree [226] is a connected, acyclic graph– a set of nodes and edges– organized in a hierarchical fashion. A node is either a split node, where it has edges leading to two other nodes, or a leaf/terminal node, where it does not have edges emanating from it and is assigned a decision, $c \in \mathbb{R}$. We utilize binary trees, where each split node has two outgoing edges. All nodes have only one incoming edge. [227] We only consider decisions at split nodes of the form $x[j] < \tau$ so that each terminal node speaks for a hyperrectangle in the feature space.

A given decision tree assigns a value $y(x_i)$ to each feature vector x_i by the path that it takes through the tree and thus the terminal node that it reaches. This is determined by a recursive process starting at the root node: each split node makes a decision to send x_i to one of its two child nodes until the data point x_i reaches a terminal node. If x_i reaches terminal node k, which is assigned a decision value c_k, the decision is $y(x_i) = c_k$. To illustrate, we

[2]Notation from Ref. [224]

show the decision tree in Figure 5.1 that takes a feature vector $x \in \mathbb{R}^2$ and corresponding function $y(x)$ in Figure 5.1.

To train/grow a single decision tree regressor starting at the root node, we must determine the splitting criterion made at each split node, criteria for when to stop growing the tree and declare a node as a leaf, and what value c_k to assign to each leaf k. Utilizing the mean square error $\frac{1}{N}\sum_{i=1}^{N}(y(x_i) - y_i)^2$ for the loss function during training is computationally infeasible [201], as the problem of constructing the tree that minimizes the mean-square error is NP-complete [228]. Instead, decision trees are grown using a greedy algorithm to choose the component of the feature vector, j, and threshold, τ, for each split node. Let regions of feature space $R_l = \{x|x[j] < \tau\}$ and $R_r = \{x|x[j] \geq \tau\}$ be those represented by the left l and right r child nodes in the currently proposed split. We choose j and τ by minimizing the sum of square errors at this split node:

$$\min_{j,\tau}\left[\min_{c_l}\sum_{i:x_i\in R_l(j,\tau)}(y_i - c_l)^2 + \min_{c_r}\sum_{i:x_i\in R_r(j,\tau)}(y_i - c_r)^2\right]. \tag{5.3}$$

The inner minimization problem is solved by assigning the child node a decision c that is the average y_i among the data points that fall there [201]. The outer-minimization problem is solved by scanning all possible splits– combinations of components j and thresholds τ between adjacent data points. We can interpret the minimization in eq 5.3 as choosing the split that partitions the data into two subsets that are the most homogeneous within each subset. If a node is declared as a leaf node, the decision that minimizes the mean square error there is the average y_i among the data points that fall there. As a stopping criteria, i.e. when to declare a node as a leaf and stop splitting, one can restrict the depth of the tree, impose a minimum number of data points that a leaf or split node can own during training, impose a maximum number of leaf nodes, and/or stop splitting if the split does not sufficiently improve the mean square error. In a random forest, one grows each tree to the largest extent possible.

A strong advantage of decision tree regressors is interpretability; we can quantify the importance of a particular feature in improving the quality of prediction by keeping track of the reduction of the mean square error at each node and to which feature this reduction in error is attributed [226]. We will later plot the relative importance of each feature in our random forest. Other advantages of decision trees are: (1) normalization or scaling of the data is not required, as decision trees are invariant to monotone transformations of the features, (2) irrelevant features will not severely detriment performance, as the decision nodes can simply ignore these unimportant features and not consider them for splits [201], and (3) decision trees are relatively robust to outliers [205]. Still, a [single] decision tree is often not as accurate as other models, in part due to the greedy construction procedure, and is highly prone to overfitting [205].

A *decision forest* [199] is an ensemble [204] of randomly trained decision trees. In this ensemble method, by taking the average vote among a committee of diverse/de-correlated decision trees [204], we obtain much greater accuracy and generalization than by using a

single decision tree [227], a Wisdom of the Crowds [206] effect. We inject diversity into the trees of the forest by inducing randomness during training using the following techniques: (1) bagging (bootstrap aggregating), where each tree is trained on a random subset of the training data [229], (2) random split selection [230], where at each split node, the split is selected at random from the K best splits, and (3) the random subspace method [231], where a random selection of the features are considered for each tree. The prediction of the forest is the average prediction among each tree in the forest.

In general, the utility of a model for making predictions is determined by its expected error over future data not seen during training time – the generalization error [205]. Typically, one sets aside a validation data set or performs cross validation [201, 205]. As we use bagging in our random forest, despite training on the full data set, we essentially set aside a validation set during training in the process of leaving out a subset of the data for each tree in the forest during the bagging procedure. After training each tree, we take the data left out from the training during bagging, run it through the tree, and store the prediction. In this way, we get the *out of bag error*, which is the error of the forest's prediction on data unseen during training [201] and a quality metric of how well our forest will perform on future data.

5.3 Results

We consider a crystalline nanoporous material in equilibrium with a 20/80 molar Xe/Kr mixture at 298 K and 1 bar. A good material for Xe/Kr separations should have a high selectivity $S_{Xe/Kr}$ for Xe over Kr:

$$S_{Xe/Kr} := \frac{(x_{Xe}/x_{Kr})}{(y_{Xe}/y_{Kr})}, \tag{5.4}$$

where $x_{\{Xe,Kr\}}$ is the mole fraction in the adsorbed phase and $y_{\{Xe,Kr\}}$ is the mole fraction in the bulk gas phase ($y_{Xe} = 0.2$, $y_{Kr} = 0.8$ in our study). A good material could also be reverse selective if its pores are too small for and thus exclude Xe but are large enough to accommodate Kr. In addition to a high Xe selectivity, a good material should also have a high Xe uptake to reduce the amount of material that is needed and the size of the adsorption column to obtain a given amount of Xe.

Training data

Using the diversity selection algorithm and $x = [\rho, a, D_i, \epsilon_v]$ as a feature vector, we selected a mathematically diverse training set of ca. 15,000 materials. The feature space explored by these materials and a scatter plot matrix of these features are depicted in Figure 5.5 and Figure 5.6.

We then performed grand-canonical Monte Carlo simulations to predict the uptake of xenon and krypton in each material in the training set (20/80 molar Xe/Kr mixture at 298 K and 1 atm). The key performance plot for Xe/Kr separations in Figure 5.8 shows the

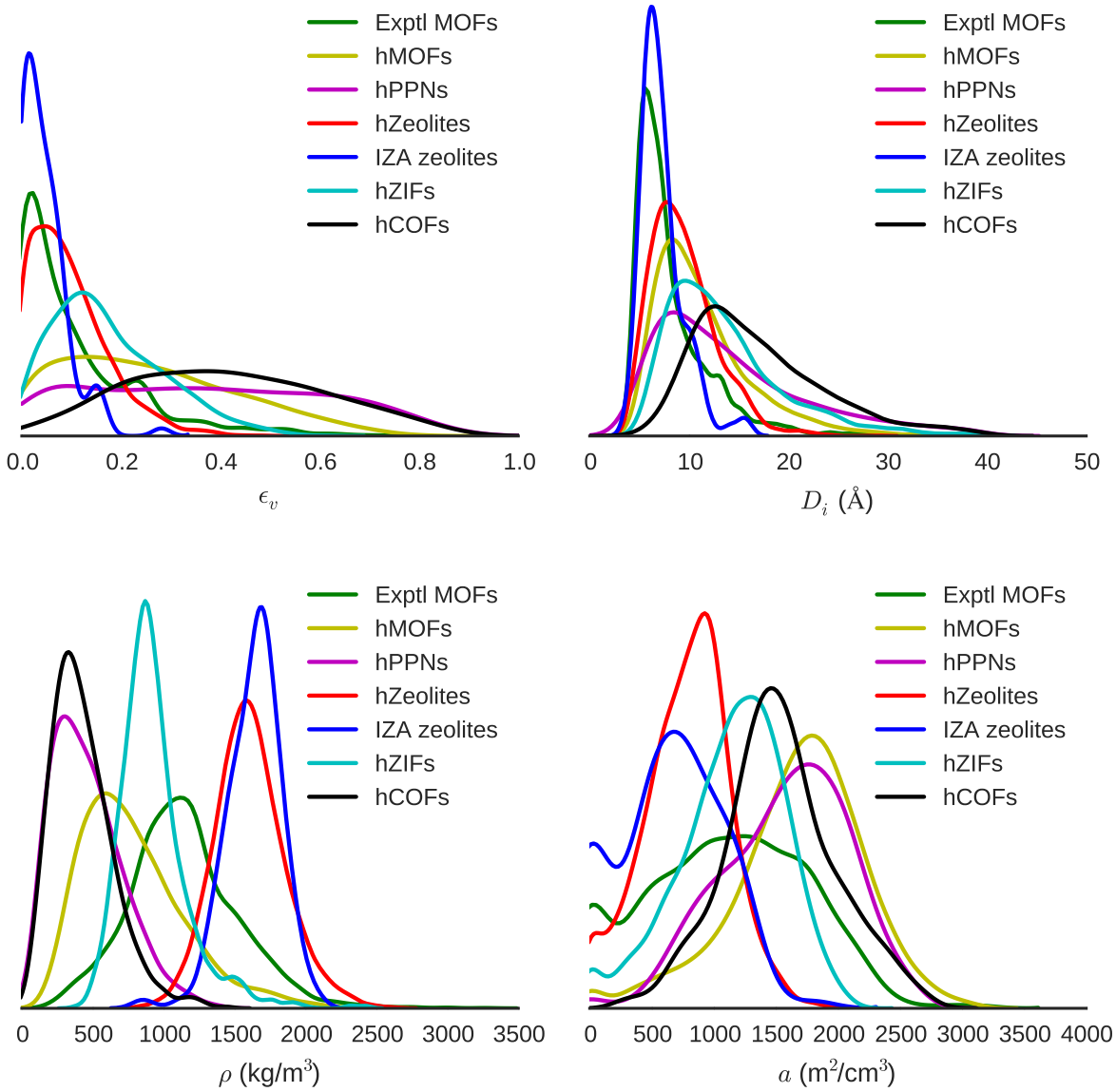


Figure 5.5: Distribution of structural descriptors in training set. Shown is the distribution of structural descriptors in the training set, partitioned by material class. Distribution estimated using kernel density estimation. Plot made in Seaborn. Reprinted with permission from Ref. [95]. Copyright 2015 American Chemical Society.

simulated Xe/Kr selectivity against the Xe uptake in the training set (as in Ref. [76]); the best materials lie in the upper-right region.

For a given purity specification and material with a given selectivity, a high *volumet-*

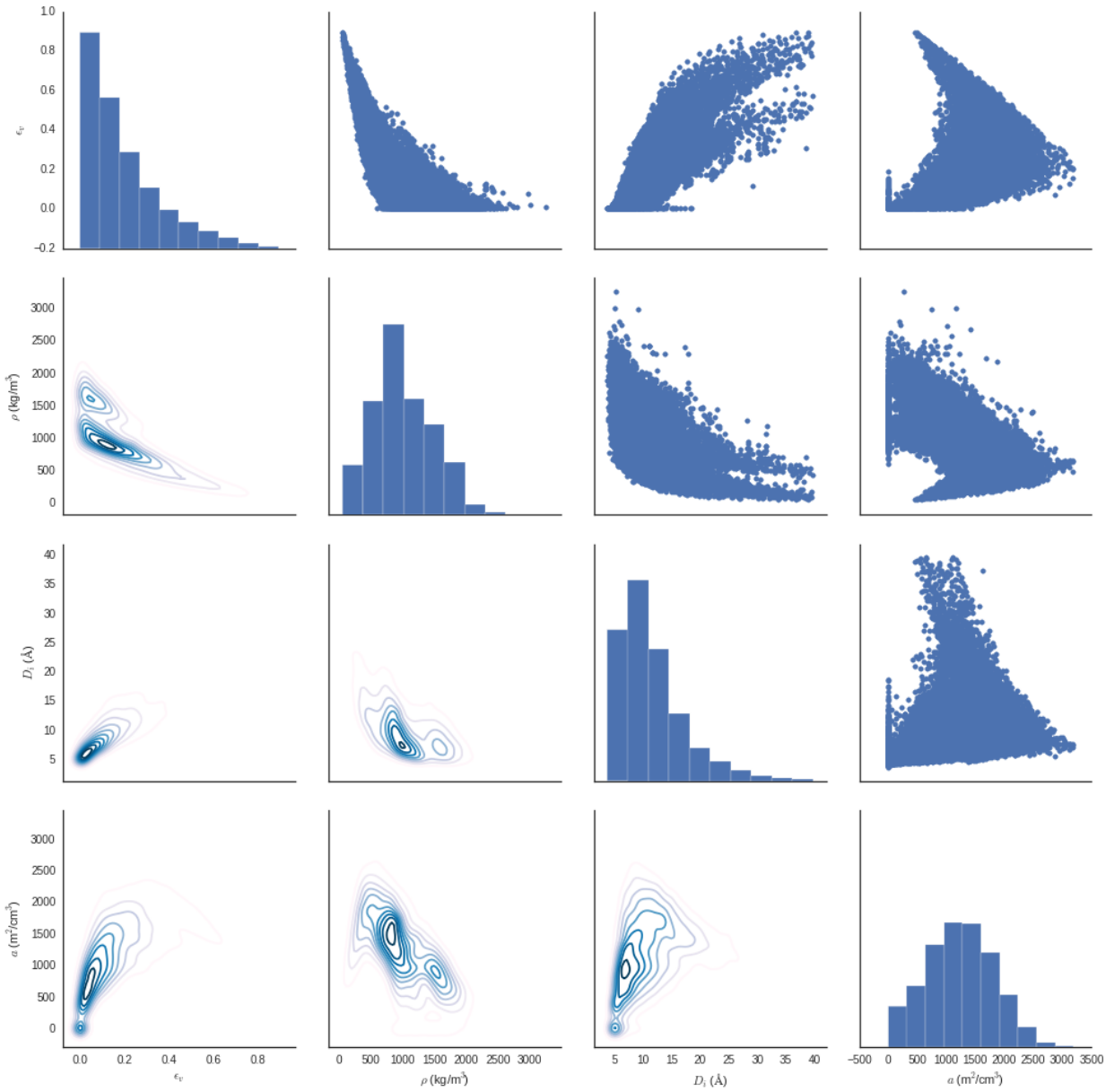
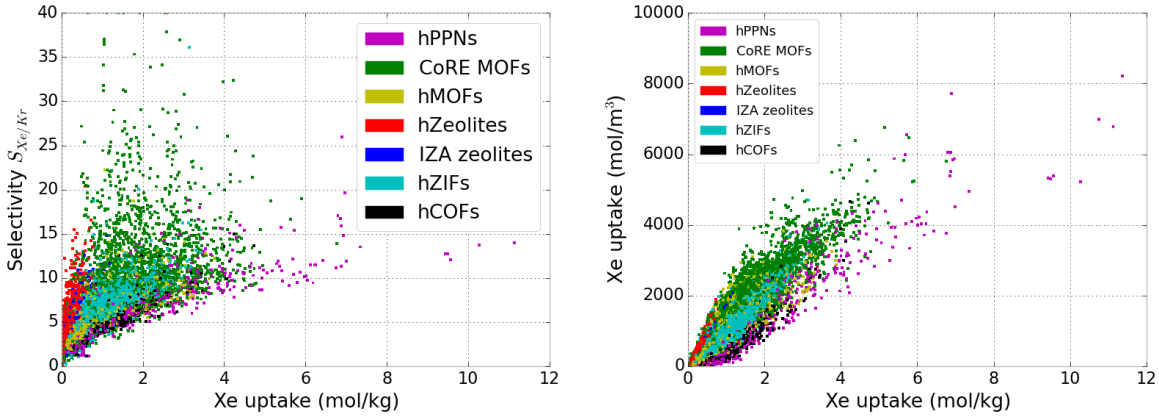


Figure 5.6: Scatter plot matrix of structural descriptors– the void fraction of xenon (ϵ_v), crystal density (ρ), largest included sphere (D_i), and accessible surface area (a)– in the training set. On the diagonal, histograms show the distribution of the respective property. On plot (i,j) of the matrix, we show variable i on the x-axis and variable j on the y-axis; the upper triangle consists of scatter plots and the lower triangle is a contour plot of the joint distribution estimated by kernel density estimation. All material classes are included. Plot made in Seaborn. Reprinted with permission from Ref. [95]. Copyright 2015 American Chemical Society.

ric xenon uptake is important to reduce the capital cost of the adsorption column; a high *gravimetric* uptake is required to reduce the mass and thus cost of material we need. In Figure 5.8, we display the volumetric xenon loading in the performance plot. The performance plot with gravimetric xenon loading on the x-axis instead of volumetric loading is displayed in Figure 5.7a. This performance plot closely resembles the performance plot in Figure 5.8 because volumetric and gravimetric uptake are strongly correlated, as Figure 5.7b shows.



(a) (b)

Figure 5.7: (a) Performance plot with gravimetric xenon loading instead of volumetric loading. (b) Volumetric versus gravimetric xenon uptake. Reprinted with permission from Ref. [95]. Copyright 2015 American Chemical Society.

For machine learning algorithms to be a successful pre-screening tool, the structural descriptors must be correlated to the Xe/Kr separation performance. Scatter plots of selectivity against our structural descriptors indeed show correlations (Figure 5.9). As an example, Figure 5.10 shows a strong relationship between the selectivity and the largest included sphere; the highest selectivities occur when the pore diameter is close to the distance that yields the optimal energy in the Xe-Xe Lennard-Jones potential ($2^{1/6}\sigma_{Xe-Xe}$), plotted as a vertical, dashed line (more on this later). Despite the pore size playing an important role and having some predictive value, other variables are at play; there are several structures with this optimal pore size that have poor selectivities. There is not a simple recipe of a given geometric structural feature that guarantees a material with a high selectivity (Figure 5.9). This motivates the utility of our random forest of decision tree regressors that predict performance by "looking" in higher dimensional feature space and simultaneously considering all structural features. The exceptionally predictive structural descriptor is the Voronoi energy.

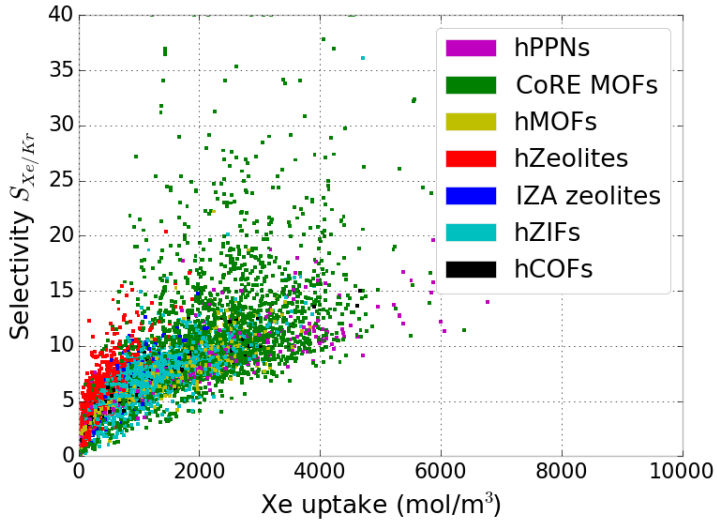


Figure 5.8: Performance plot regarding training data selected from the diversity selection algorithm. Xe/Kr selectivity $S_{Xe/Kr}$ against volumetric Xe uptake. Each structure is a point, and color indicates material class. In this plot as well as others, points are plotted in a random order to avoid seeing only data points plotted last. In Figure 5.7, we show the performance plot for each class of material separately. The performance plot using gravimetric xenon uptake closely resembles this because volumetric and gravimetric xenon uptake are correlated (see Fig 5.7). Reprinted with permission from Ref. [95]. Copyright 2015 American Chemical Society.

Training the random forest

In the absence of a purity specification and an economic model of a Xe/Kr separation plant (describing capital costs of equipment and material and recurring costs of regenerating the material) to construct a sensible performance metric of a material, we use the selectivity for the performance score y of a material for parsimony. This score implicitly takes xenon loading into account since selectivity positively correlates with Xe uptake in Figure 5.8. For the decision forest to predict the selectivity of a material, we use as a feature vector $x = [\rho, \epsilon_v, a, D_i, D_f, E_v, \rho_s]$ (see Table 5.1 for descriptions).

We then grew a random forest of 1000 decision tree regressors using the diverse training set. Figure 5.11 shows that 1000 is an adequate number of trees for our regression. We excluded materials with inaccessible pockets as determined by our pocket blocking algorithm (see Methods) to avoid confusing the trees, as a subset of the geometric structural descriptors cannot possibly predict this phenomena that nonetheless could significantly affect the simulated selectivity. In total, we trained the random forest using 9,376 training examples. We used four features in the random subspace method [231], as this yielded the smallest out of bag error.

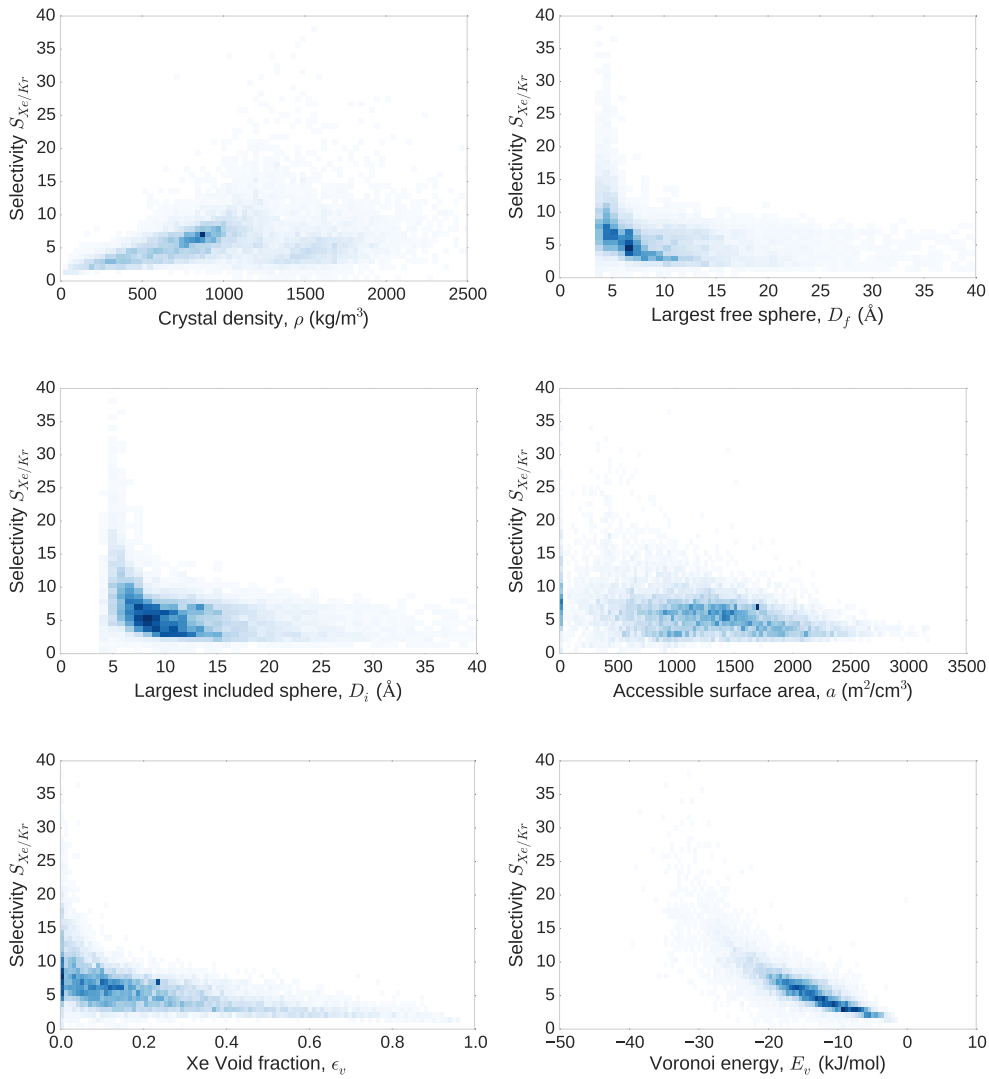


Figure 5.9: Two-dimensional histograms of selectivity vs. various structural descriptors. Strong correlations are present, indicating that there is signal in the descriptors for whether or not a material will have a high selectivity. Reprinted with permission from Ref. [95]. Copyright 2015 American Chemical Society.

Figure 5.17a displays a histogram of the simulated selectivity $S_{Xe/Kr}$ in the training set of materials against the out of bag prediction (see Methods) by the random forest. The red, diagonal line indicates where the random forest's prediction exactly matches the simulated selectivity; the root mean squared error is 1.2. We observe a high variance in the predictions of the decision forest for materials with the highest selectivities, possibly due to the scarcity of data in this performance range, which we will later address, or the inadequacy of our

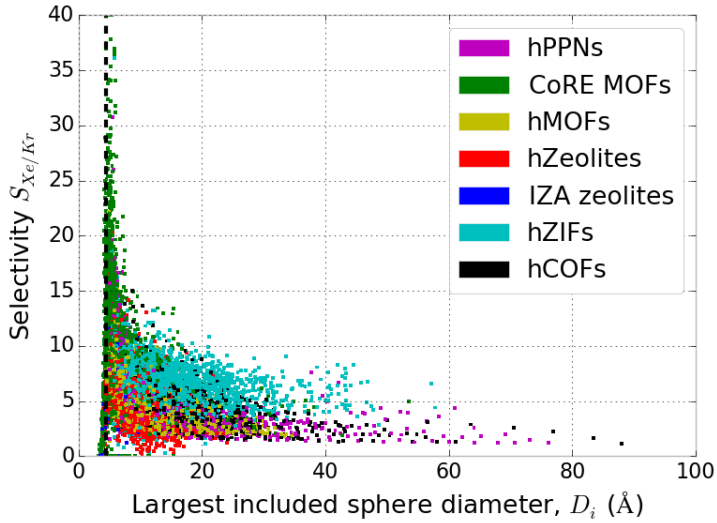


Figure 5.10: Selectivity against pore size. Xe/Kr selectivity $S_{Xe/Kr}$ against diameter of largest included sphere D_i. Vertical, dashed line is $2^{1/6}\sigma_{Xe-Xe}$, the optimal Xe-Xe distance. Selectivities clipped at 40. Reprinted with permission from Ref. [95]. Copyright 2015 American Chemical Society.

descriptors for these materials (irreducible error).

Feature selection

An interesting question is what features are most important in predicting selectivity. We plot the relative importance of each feature (see Methods) during training, averaged over each tree in the forest, in Figure 5.17b; a higher feature importance implies the feature played a larger role in decreasing the mean square error in the decision tree regressor during training. We see that the expected energy of a xenon atom at the accessible Voronoi nodes, E_v, is by far the most important feature. This is not surprising, given that energy is a major thermodynamic driving force for adsorption. The second most important feature is the void fraction. We duly note that correlations between variables can influence this feature importance plot [199, 232]. For reference, the correlations between the features in our training set are depicted in the correlation matrix in Figure 5.12.

We can also get an idea of the *importance* of a feature by removing it from the regression and observing how the mean square error on the out of bag samples changes. We have seven structural descriptors or features of materials. Feature selection involves considering only i features in the regression, for $i = 1, 2, ..., 7$.

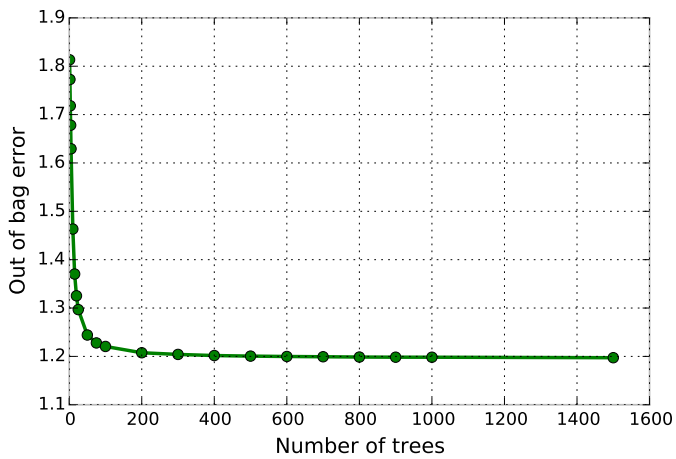


Figure 5.11: The out of bag error in the prediction of the decision forest as a function of the number of trees in the forest. Each point is an average of 200 runs. We trained random forests of differing numbers of trees on the training data and observed the out of bag error. This figure shows that we reach the law of diminishing returns at around 1000 trees. That is, the out of bag error decreases only marginally as we add more trees beyond this point. Thus, 1000 trees is an adequate number of trees for an accurate random forest.The out of bag error in the prediction of the decision forest as a function of the number of trees in the forest. Reprinted with permission from Ref. [95]. Copyright 2015 American Chemical Society.

Forwards stepwise selection

In forward stepwise selection [201], we start with using a single feature ($i = 1$) in the random forest regression. We loop through all seven features, run the regression, and observe which of the single-feature models yields the smallest mean square error on the out of bag examples. This is the most predictive feature.

We then proceed to consider a two-feature ($i = 2$) model, but keep the feature from the last step. Thus, we only loop over the remaining six features and perform a regression with each of these in conjunction with the feature chosen from the $i = 1$ step. We then choose the second most important feature as the addition feature in the two-feature model that yields the lowest error on the out of bag examples.

As we proceed to build a model with $i = 3, 4, 5, 6, 7$, we always keep the feature from the previous step. This is what makes the forward stepwise selection a *greedy* algorithm; it could be the case that, for example, feature 3 and 6 yield the best two-feature model, whereas feature 2 yields the best one-feature model.

Figure 5.13 shows the best root mean square error as a function of the number of features (i) that we include in the forward stepwise selection process. The red point at $i = 0$ is the error when we assign all materials to have the same selectivity, the mean of the training set,

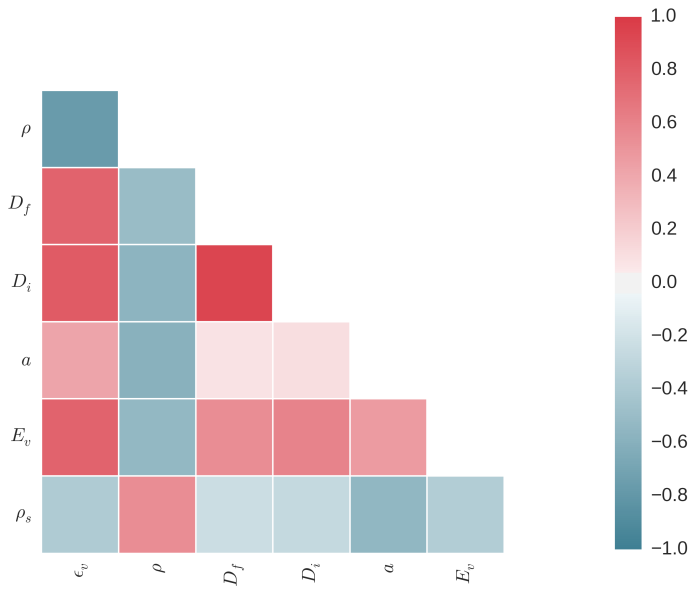


Figure 5.12: Correlation matrix for features in the training set. Each bin depicts the correlation between a pair of features in our diverse training set. Blue represents a negative correlation; red represents a positive correlation. Lighter colors represent weaker correlation, as the colorbar shows. Plot made in Seaborn. Reprinted with permission from Ref. [95]. Copyright 2015 American Chemical Society.

i.e., without any predictions from machine learning. The box on the right shows the rankings of features. The first selected feature is the Voronoi energy E_v; the worst is ρ_s. We see that each feature adds predictive value to the regression, but that the marginal contribution of each descriptor diminishes quickly as we add more features.

Backward stepwise elimination

In backward stepwise elimination [201], we start with the full set of seven features and sequentially remove features that have the smallest impact on decreasing the mean square error on the out of bag examples. In the first step, we consider a six-feature model, loop through each of the seven features, perform the regression with this feature removed, and elect to remove the feature that had the lowest out of bag error when it was removed. Backward stepwise elimination is also a greedy algorithm since, once a feature is removed at a given step, we never consider it again.

Figure 5.14 shows the smallest root mean square out of bag error as a function of the

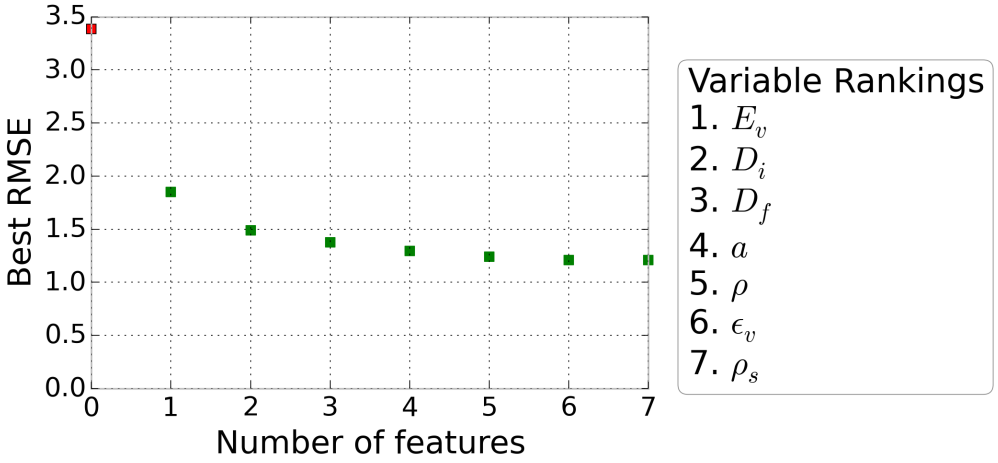


Figure 5.13: Forward stepwise selection. Best root mean square error in the forward stepwise selection process against the number of features that we include in the regression. The box shows the order in which features were selected. Reprinted with permission from Ref. [95]. Copyright 2015 American Chemical Society.

number of features that were eliminated. Again, the red point is when we assign every material the average selectivity in the training set. The order in which features were eliminated is shown in the box beside the figure. Note that this order is not the reverse of forward subset selection, highlighting the greedy nature of these algorithms.

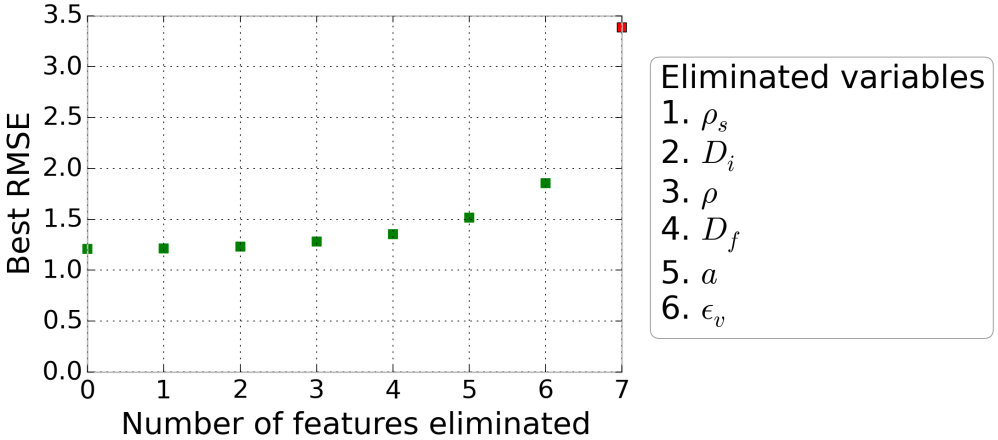


Figure 5.14: Backward stepwise elimination. Best root mean square error in the backward stepwise elimination process against the number of features that we stepwise-eliminated from the regression. The box shows the order in which features were eliminated on the basis of it having a low impact on decreasing the out of bag error. Reprinted with permission from Ref. [95]. Copyright 2015 American Chemical Society.

Exhaustive search

Forward stepwise selection and backward stepwise elimination are greedy algorithms to determine the best subset of features for $i = 1, 2, ..., 7$ feature models. They are often used for computational efficiency as well as to achieve lower variance in the resulting feature subset [201]. As we have only seven features, we can exhaustively explore all

$$\sum_{i=1}^{7} \binom{7}{i} = 127 \tag{5.5}$$

combinations of features.

The best feature combination for a regression with i features is shown in Table 5.2. We see that the $i = 1$ and $i = 2$ models agree with forward stepwise elimination. However, for model $i = 3, 4, 5$, D_i is not included in the best model, highlighting the greedy nature of forward stepwise selection. The $i = 6$ result of course agrees with the backward stepwise elimination; ρ_s is the least important variable.

Table 5.2: Exhaustive search for best combination of features

Number of features (i)	Best feature combination
1	E_v
2	E_v, D_i
3	E_v, ϵ_v, a
4	E_v, ϵ_v, a, D_f
5	E_v, ϵ_v, a, D_f, ρ
6	E_v, ϵ_v, a, D_f, ρ, D_i
7	(all features)

These feature subset selection strategies support that the Voronoi energy E_v is the most important feature in our study. The backward and forward stepwise selection processes show the marginal contribution of each feature in the reduction of the out of bag error.

Screening the remaining materials

The good agreement between the out of bag prediction and simulated selectivity in Figure 5.17a indicates that the trained random forest can reasonably predict the selectivity in the remaining (outside of the training set) ca. 655,000 materials for which we did not yet perform molecular simulations. To screen the remaining materials, we run their feature vectors of structural descriptors through the trained random forest. If the prediction $\hat{S}_{Xe/Kr} > 11$ (to the right of the green vertical line in Figure 5.17), we refine the prediction by a grand-canonical Monte Carlo simulation of the Xe/Kr adsorption. In this manner, we focus the computational resources for molecular simulations on the materials predicted to be the most selective by the random forest. We choose a threshold selectivity of 11 in part because the

random forest exhibits low variance in the region $\hat{S}_{Xe/Kr} \leq 11$, and thus we have greater confidence in its prediction.

The trained random forest predicted that 4,066 remaining structures satisfy the condition $\hat{S}_{Xe/Kr} > 11$, which we refer to as the *promising set*. In Figure 5.17c, we show the simulated selectivities of the promising set against the prediction by the random forest. The root mean squared error is 2.2. This includes materials with blocked pockets that the decision trees were not trained to predict, which account for most of the outliers in Figure 5.17c.

We visualize the efficiency of our screening strategy by comparing the distribution of simulated selectivites in the promising set with the distribution in a *randomly* selected set of ca. 1000 materials from the Nanoporous Materials Genome in Figure 5.17d. The mode of the simulated selectivity distribution for the promising set is significantly greater than that for the randomly selected set, confirming the efficiency of our screening strategy using random forests. The random forest selects the best materials from the database and circumvents wasting simulation time on poorly performing materials, as in a brute-force screening strategy. For comparison, we also show the distribution of selectivities in the diverse training set, which mimics that of the random set.

It is important to ensure that the random forest does not incorrectly label promising materials as low performing. The match between simulation and out of bag prediction for materials with low selectivities in Figure 5.17a is evidence already. To test this further, we simulated Xe/Kr adsorption for the 500 materials in the remaining set that the random forest predicted to have the lowest selectivities. These simulations confirmed (see Figure 5.15) that indeed these materials have low selectivities.

To ensure that the decision forest is not under-predicting the selectivity of some materials because of the paucity of example structures with a high selectivity in the diverse training set (Figure 5.17a), we trained a decision forest using the *union* of the diverse training set in Figure 5.17a and the promising set in Figure 5.17d, providing more training examples of high-performing materials. We then ran the feature vectors of the remaining materials through this re-trained random forest and find 461 additional materials that meet the criteria $\hat{S}_{Xe/Kr} > 11$. We then simulated Xe/Kr adsorption in these materials to refine the prediction. We did not discover any additional materials beyond those found in the first round of training with exceptionally high selectivities (see Figure 5.16). Thus, we are confident that our training set selected by the diversity selection algorithm adequately explores feature space.

By utilizing random forests, we performed grand-canonical Monte Carlo simulations in only 20,000 of the 670,000 materials in our process of screening the Nanoporous Materials Genome for Xe/Kr separations and utilized the quickly-computed feature of structural descriptors to rule out the rest. Before looking at the top materials from our screening, we first gain insight into how the pore size and shape affect the selectivity.

The role of pore size and shape

Snurr and coworkers found that the highest Xe/Kr selectivities in MOFs were obtained with pores around the size of a Xe atom [76], as we see in Figure 5.10. To grasp how the

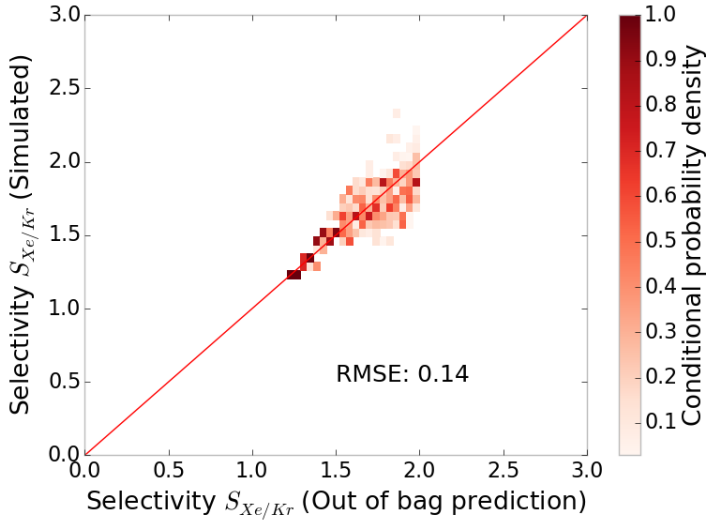
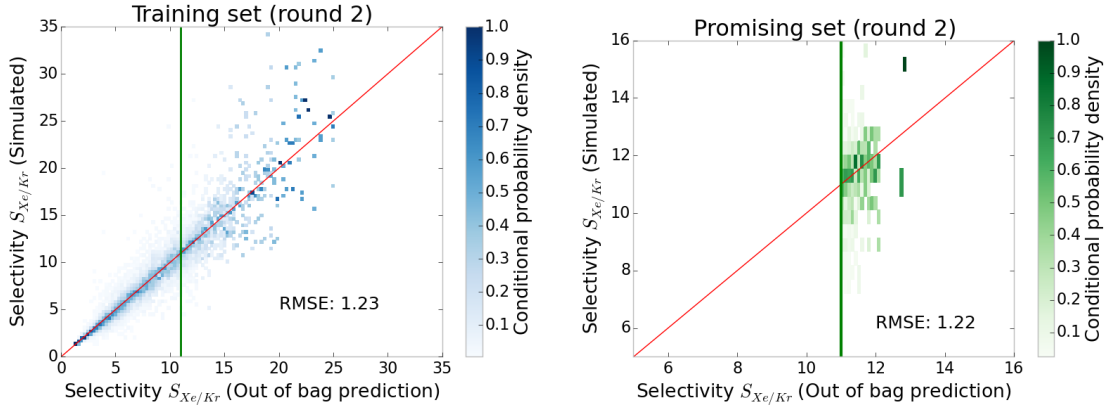


Figure 5.15: False negative analysis. We assess if the materials predicted to have the lowest selectivities by the random forest indeed have low simulated selectivities. This is important to show since we refrain from simulating adsorption in materials that the random forests predicts to be poor, opening up the possibility that we are missing high-performing materials in our screening. We select 500 materials predicted to have the lowest selectivities by the random forest and simulate Xe/Kr adsorption. This two-dimensional histogram shows the simulated against predicted selectivity for 500 of the materials predicted to have the lowest selectivities by the random forest. The red line is a perfect prediction. Reprinted with permission from Ref. [95]. Copyright 2015 American Chemical Society.

selectivity of a material depends on its pore size, we study here two simple model materials which capture the relationship between Xe/Kr selectivity and pore size that we observe in Figure 5.10. Since a given diameter of the largest included sphere D_i can be achieved with differently shaped pores, for example, a spherical and cylindrical pore with the same radius, we also investigate the effect of pore *shape* on the selectivity.

Consider two model adsorption pockets, one constructed from a spherical shell of radius R and the other constructed from an infinite cylindrical shell of radius R. See Figure 5.18a. The framework atoms are uniformly distributed on the surface of the sphere/cylinder. We model the interactions between Xe/Kr and atoms of the shell using Lennard-Jones potentials (energy parameter ϵ, size parameter σ). By integrating the Lennard-Jones potential over the surface of the sphere and cylinder – "smearing" the atoms over the shell surfaces [141] – we obtain effective interaction potentials between an adsorbate and the shells that are functions of r, the radial coordinate of the guest molecule in the spherical[140]/cylindrical[233] shell:



(a) (b)

Figure 5.16: Round 2 of training. We retrained the random forest using the *union* of the diverse training set and the promising set identified from the decision forest. The idea is that including more examples of high selectivities will help the random forest correctly identify the highest-performing materials. (a) Simulated versus out of bag prediction of selectivities for the training set (including selected set from first round of training). (b) Simulated versus predicted selectivities for the set of materials predicted to have selectivities > 11 in the second round of training. Note the change in scale on the x-axis. Reprinted with permission from Ref. [95]. Copyright 2015 American Chemical Society.

$$U_{sphere}(r;R) = \eta 4\epsilon \frac{2\pi R}{r}\left\{\frac{\sigma^{12}}{10}\left[\frac{1}{(R-r)^{10}} - \frac{1}{(R+r)^{10}}\right] - \frac{\sigma^6}{4}\left[\frac{1}{(R-r)^4} - \frac{1}{(R+r)^4}\right]\right\}$$
$$U_{cylinder}(r;R) = \eta 4\epsilon 2\pi\left[\frac{\sigma^{12}}{R^{10}}I_6 - \frac{\sigma^6}{R^4}I_3\right] \quad (5.6)$$
$$I_n := B\left(n-\frac{1}{2},\frac{1}{2}\right){}_2F_1\left(n-\frac{1}{2},n-\frac{1}{2};1;\frac{r^2}{R^2}\right).$$

Here, $B(x,y)$ and ${}_2F_1(a,b;c;z)$ are the beta and hypergeometric funtions, respectively, and η is the surface density of atoms on the shell. These shell models are useful to intuit the relationship between pore size and Xe/Kr separation performance because they tease out the effects of varying pore shapes, topologies, compositions, and surface density of atoms that are present in Figure 5.10 for a given pore size.

For parsimony, we perform all calculations in this section at infinite dilution and assume the adsorption isotherms are in the Henry regime so that loading can be approximated as $K_H P$, where K_H is the Henry coefficient and P is the partial pressure. To compare our model to real material structures, we computed the Kr and Xe Henry coefficients in the dataset of experimental MOFs [160] using Widom insertions [110]. We imposed a uniform,

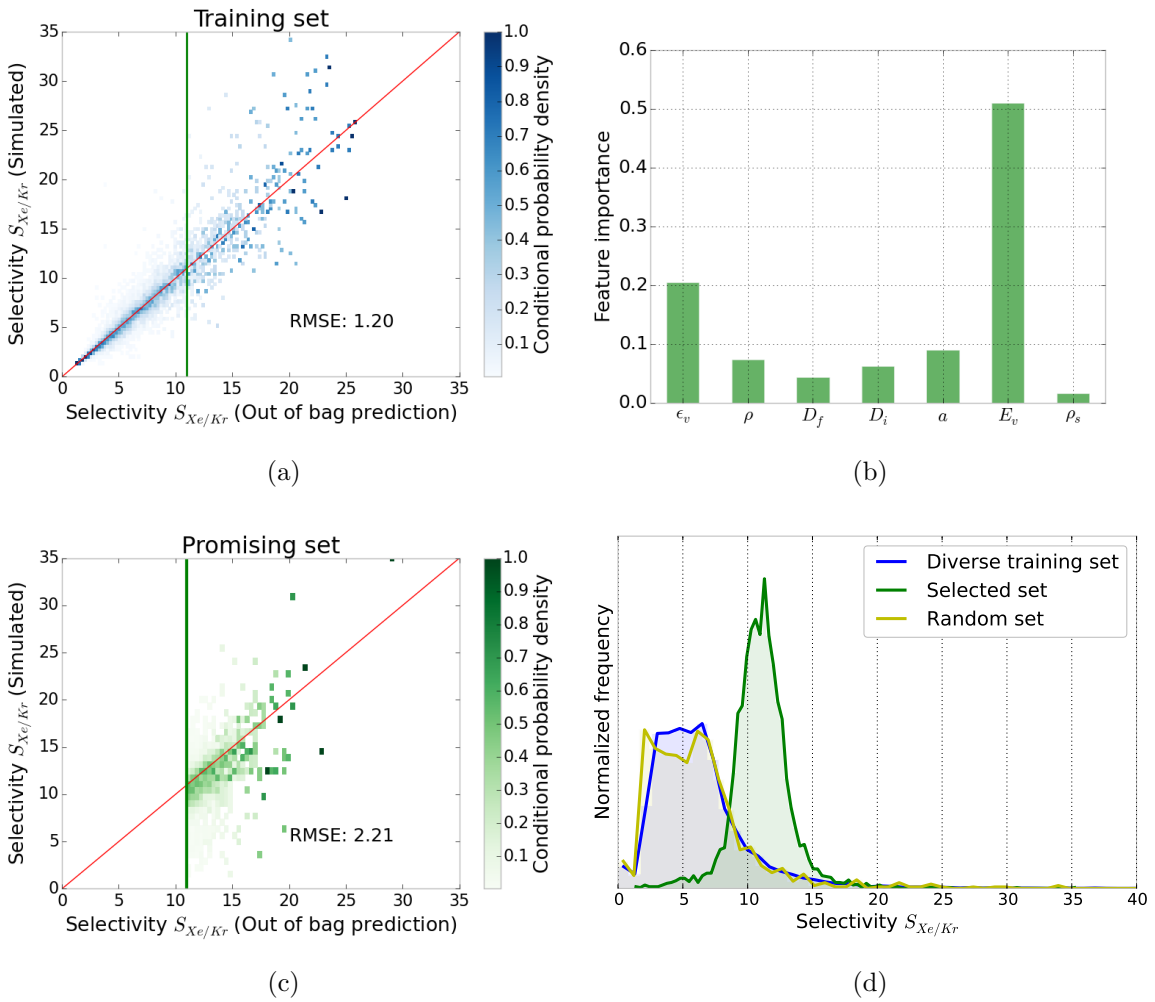


Figure 5.17: Decision forest results. (a) Two dimensional histogram of simulated selectivity against selectivity predicted from feature vector by the random forest. This includes all materials in the training set. The x-axis shows the out of bag prediction; each tree contributing to the prediction for a given structure has not been trained on it. The red, diagonal line is a perfect prediction. For perspective, all materials predicted to lie to the right of the green, vertical line are deemed as promising and assigned to the *promising set*. (b) The average feature importance among the decision tree regressors in the random forest. A higher feature importance implies that the feature is more important in reducing the mean square error. See Table 5.1 for feature descriptions. (c) Two dimensional histogram of simulated against predicted selectivity for the *promising set* of materials (outside the training set). (d) Distribution of simulated selectivities in the diverse training set, the *promising set*, and a random selection of the materials. To ensure a proper comparison, we fixed the percentage of the random set that consists of each material class to be equal to that of the diverse training set. Reprinted with permission from Ref. [95]. Copyright 2015 American Chemical Society.

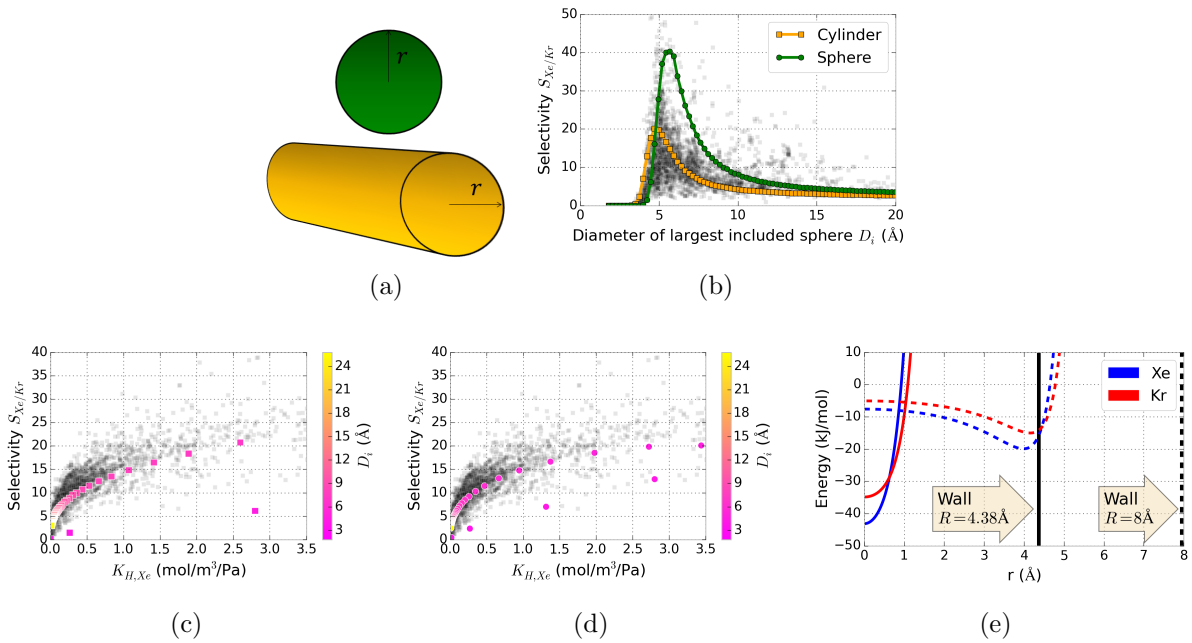


Figure 5.18: Spherical and cylindrical shell models. (a) In the model, framework atoms are uniformly distributed across the shell and 'smeared' over the surface, leaving an effective potential which is a function of the radial coordinate r. (b) The selectivity in the model materials as a function of the pore diameter. (c, d) The performance plot of the spherical (c) and cylindrical (d) shell models. Each point is a sphere/cylinder of a different radius R; colormap indicates D_i. In both (b) and (c), the black, opaque points in the background are the experimental MOFs. (e) The Lennard-Jones potentials in eqn 5.6 for Xe (blue) and Kr (red) for two differently sized spherical shells. Dashed lines are for the larger sphere ($R = 8$ Å); solid lines are for the smaller sphere ($R = 4.38$ Å). Reprinted with permission from Ref. [95]. Copyright 2015 American Chemical Society.

constant surface density of framework atoms, η, and investigated the effect of composition in our model material by using different types of atoms (σ and ϵ in eq 5.6). We found that Universal Force Field [108] Lennard-Jones parameters for silicon and $\eta = 0.1$ Å^{-2} recapitulate the trends of the Henry coefficients of the experimental MOFs.

Using the potentials in eq 5.6, we calculate the Henry coefficient of xenon and krypton in the model materials as a function of pore size R via Widom insertions [110]. We plot the selectivity at infinite dilution for various shell sizes in Figure 5.18b. In our model, we define the largest included sphere diameter $D_i := 2(R - r_{Si})$ to correspond to that calculated by Zeo++, where $r_{Si} = 2.1$ Å is the van der Waals radius of silicon[218]. The models recapitulate the trend in the experimental MOFs (black points), and we observe a well-defined pore diameter that maximizes the selectivity (sphere: 5.7 Å, cylinder: 4.7 Å). The shell models also recapitulate the performance plot of the experimental MOFs in

Figure 5.18c. As these crude model materials recapitulate the trends in the experimental MOFs, it is evident that the pore size plays a central role in Xe/Kr separations.

Let us rationalize the poor performance of materials with large R. Figure 5.18e shows the Xe and Kr potentials in a large sphere ($R = 8$Å) and in a smaller sphere that yields maximal potential well depth for Xe ($R = 4.38$Å). For the larger sphere, the minimum potential energy occurs on a *surface* near the walls; adsorbates at the center ($r = 0$) are too far from the walls to feel strong interactions. As R decreases, eventually the minimum energy becomes a *point* at the center $r = 0$, as we see with the smaller sphere. Thus, one reason large pores are ineffective is that the attractive interactions at the center are weak and resemble an empty space. The size of the shell also determines the degree of potential overlap by multiple framework atoms contributing van der Waals interactions to recruit the adsorbate. In the large pore, an adsorbate near the surface sees a relatively flat surface, whereas in the smaller pore, an adsorbate is proximal to the shell in all radial directions. As a result of these combined interactions, the potentials in Figure 5.18e are deeper in the smaller pore. More importantly, the *difference* in energy of interaction between a Xe and Kr is greater in the smaller pore, yielding a higher selectivity. Of course, if the pores get too small, repulsive forces become dominant and the selectivity and adsorption decrease rapidly with decreasing R. These observations hold for the cylinder as well.

Note that the radius R of shell that yields the minimal energy for Xe (sphere: 4.38 Å) is larger than that for krypton (sphere: 4.2 Å). Thus, *two* factors contribute to the selectivity for xenon over krypton in the model materials: (1) xenon has a deeper Lennard-Jones attractive potential well (larger ϵ) than krypton and (2) the pore diameter that maximizes interactions with xenon is larger than that for krypton, a size effect (see Figure 1.4). We can assess the extent to which the size difference is responsible for the selectivity of the model material for xenon by constructing an artificially enlarged krypton molecule by assigning the Lennard-Jones σ to be equal to that of xenon and but keeping the ϵ equal to that of krypton. By simulating the selectivity of our model material for xenon over this artificially enlarged krypton, the peak selectivity was observed to decrease approximately three-fold (See Figure 5.19). Thus, the difference in the sizes of Xe and Kr accounts for 2/3 of the selectivity of our model material, while the difference in Lennard-Jones potential depths accounts for the other third.

Finally, we remark on the differences between the spherical and cylindrical shell. A spherical shell has a deeper potential well than a cylindrical shell with the same R and surface density of atoms (Figure 5.21). The reason is that a xenon atom in a sphere is encompassed by framework atoms in all directions, whereas inside a cylinder, lateral interactions, from the direction along the cylindrical axis, are lacking. As a result, with the same η, we achieve a higher selectivity with the spherical shell in Figure 5.18b. Also note that the R that yields the maximal selectivity is larger in a sphere than in a cylinder. This is largely because a Xe or Kr atom in the cylindrical shell experiences its minimum energy in the (r, R) parameter space at a smaller R than in the spherical model (Figure 5.20).

The sphere has the minimum energy at $R = 2^{1/6}\sigma$, the distance that yields the minimum in the adsorbate interaction with the atoms of the shell. The cylinder has a smaller optimal

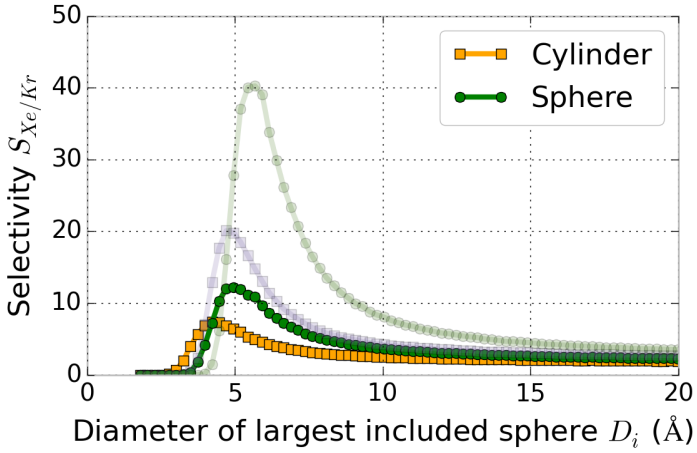


Figure 5.19: Spherical shell model. To analyze the extent to which the difference in potential well depth plays a role in the model material's selectivity, we simulated the selectivity of our model material for xenon over krypton using an artificially enlarged krypton guest. In the enlarged krypton, the Lennard-Jones σ was set to be equal to that of xenon while the ϵ was unchanged. We found that when the two atoms are roughly the same size, the selectivity decreases three-fold. This indicates that the difference in size accounts for 2/3 of the selectivity of our model material for xenon. Shown is selectivity of xenon over an artificially enlarged krypton molecule against pore size. We assign the artificially enlarged krypton the same Lennard-Jones σ as xenon but keep the Lennard-Jones ϵ of krypton. The opaque curve shows Figure 5.18b (unaltered Xe/Kr). Reprinted with permission from Ref. [95]. Copyright 2015 American Chemical Society.

radius, as the suboptimal interactions with the closest atoms are compensated by increasing more interactions in the lateral/axial direction.

While our discussions have centered around energy, we duly note that entropy plays a role in determining the Henry coefficient as well [110].

The scatter about the model curves exhibited by the experimental MOFs in Figure 5.18b is due to real materials having more complex shapes than perfect cylinders and spheres as well as varying compositions and surface densities. We did not observe a clear difference in the pore size-selectivity relationship between MOFs with 1D channels and 2D/3D channels (Figure 5.22) because of these significant and complex perturbations in real materials. While these simple models give us great intuition, the deviation of real materials from spheres and cylinders leads us to perform a detailed analysis of the pores of the top performing materials next.

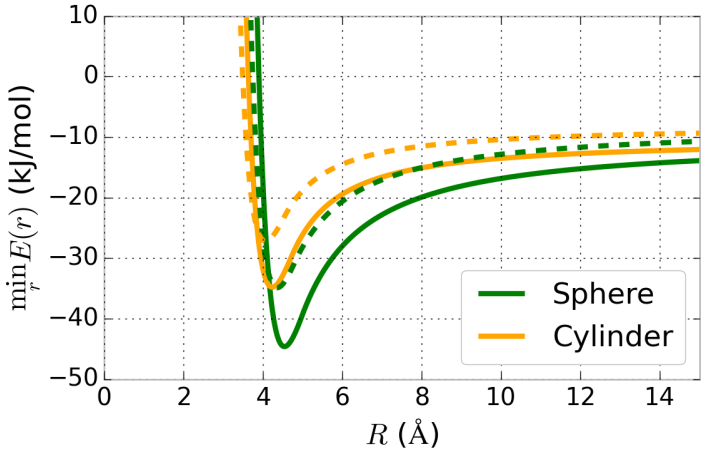
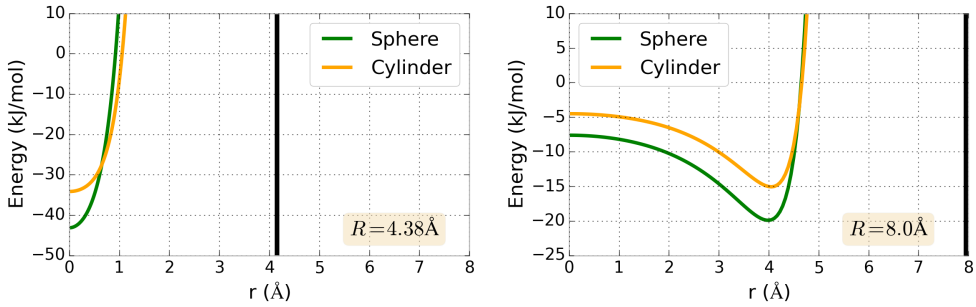


Figure 5.20: The minimum potential energy of xenon (solid lines) and krypton (dashed lines) in the spherical and cylindrical shells as a function of R. The R that yields the minimum $\min_r E(r)$ is smaller for the cylinder than for the sphere. The R that minimizes $\min_r E(r)$ is smaller for Kr than for Xe. Reprinted with permission from Ref. [95]. Copyright 2015 American Chemical Society.

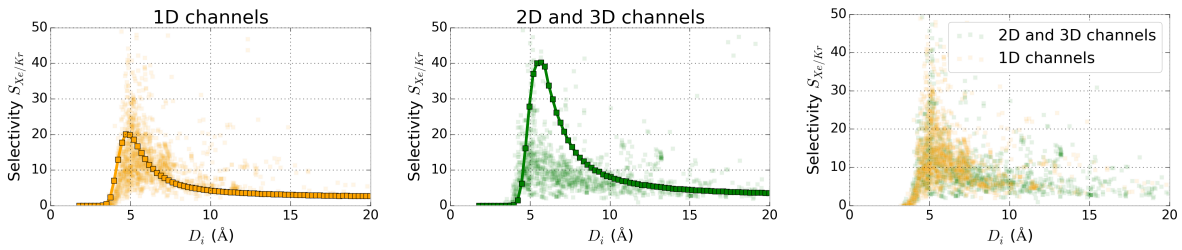


(a) (b)

Figure 5.21: The potentials in eqn 5.6 for xenon in two differently sized cylindrical and spherical shells (a) $R = 4.38$ $\AA$ (b) $R = 8.0$ $\AA$. Solid line indicates location of shell wall. Note that the sphere yields a larger interaction energy than a cylinder with the same surface density of atoms, η. Reprinted with permission from Ref. [95]. Copyright 2015 American Chemical Society.

Top performing materials and binding site analysis

We show the performance plot including the *promising set* in Figure 5.23. Also shown is porous organic cage solid CC3 [5] with groundbreaking measured Xe/Kr separation performance [77], simulated with a force field developed especially for it in Ref. [77]. Our



(a) (b) (c)

Figure 5.22: For these plots, we partition the experimental MOF dataset into structures that exhibit 1D channels (yellow) and structures that exhibit 2D and 3D channels (green). The 1D channels should be cylinder-like structures. (a,b) The cylinder and sphere model simulations on top of the experimental MOF data points. (c) A high degree overlap is exhibited between the MOFs with 1D channels and the MOFs with 2D/3D channels. Reprinted with permission from Ref. [95]. Copyright 2015 American Chemical Society.

models predict many more selective materials for Xe/Kr separations than CC3, which has a selectivity of 13.8 at our conditions.

An interactive dashboard with the performance plot data is openly available at: `http://nanoporousmaterials.org/xekrseparations/`.

Top candidates

We display the structures of the top performing materials in Figure 5.24. The two structures with the highest selectivity in the Nanoporous Materials Genome are JAVTAC, an aluminophosphate zeolite analogue that has been synthesized by Cooper *et al* [235], and KAXQIL, a calcium based coordination network synthesized by Banerjee *et al* [234]. Both structures are found in the experimental MOF database [160]. Figures 5.24a and 5.24b show that KAXQIL and JAVTAC consist of 1-dimensional channels. The structure with the highest volumetric Xe loading (8223 mol/m^3) is predicted porous polymer network hPPN_Si_4080_1-net_001 [159], which has three-dimensional channels (see Figure 5.24c). This structure is a diamond analogue, in which connections between C atoms are extended with -C≡C- groups, therefore the name extended diamond (extDIA). This extDIA structure was previously highlighted for its methane storage capability [236]. The structure with the highest product of selectivity and xenon loading is an aluminophosphate molecular sieve [237] that has been synthesized [238], GOMREG ($S_{Xe/Kr} = 76$, Xe loading =4500 mol/m^3, also found in the experimental MOF database [160]), shown in Figure 5.24d to consist of 1-dimensional channels.

Some materials are reverse selective because their pores are too small for and thus exclude xenon but are large enough to accommodate a krypton atom. The most reverse selective material we found is a predicted all-silica zeolite [119, 120] h8166173, shown in Figure 5.24e,

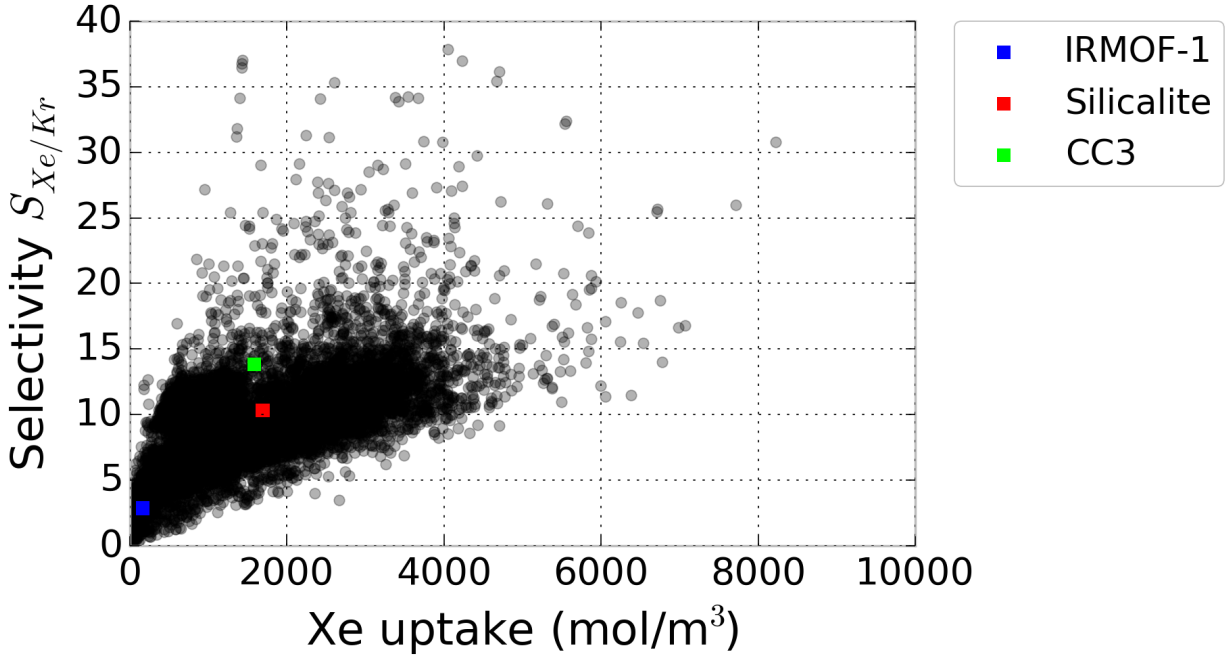


Figure 5.23: Performance plot, including both training and selected set. Opaque, black points are the simulated data for all materials in this study. Colored points show select synthesized structures. IRMOF-1 and Silicalite were simulated using the same force field as in this work. CC3 data point simulated with a force field developed especially for it in Ref. [77] (thank you Linjiang Chen for performing this simulation). Our models predict many better materials for Xe/Kr separations than CC3. Interactive version available at `http://nanoporousmaterials.org/xekrseparations/`. Reprinted with permission from Ref. [95]. Copyright 2015 American Chemical Society.

with a selectivity $S_{Xe/Kr}$ =0.10. The diameter of the largest included sphere in h8166173 is $D_i = 3.73$ Å, between the van der Waals diameter of a krypton and xenon atom ($2r_{Kr} = 3.66$; $2r_{Xe} = 3.97$ Å).

Characteristics of top materials

To further confirm that there is not a simple formula of structural descriptors to yield a material with a high selectivity, we investigate the difference in structural descriptors between highly and poorly selective materials. Figure 5.25 compares the distributions of each structural descriptor among highly selective ($S_{Xe/Kr} \geq 14$) and poorly selective ($S_{Xe/Kr} < 14$) materials separately.

Highly selective materials have low void fractions, very low (favorable) xenon energies at the accessible Voronoi nodes (E_v), and pore sizes around that of a xenon atom (reiterating Figure 5.10). Interestingly, highly selective materials tend to have lower surface areas.

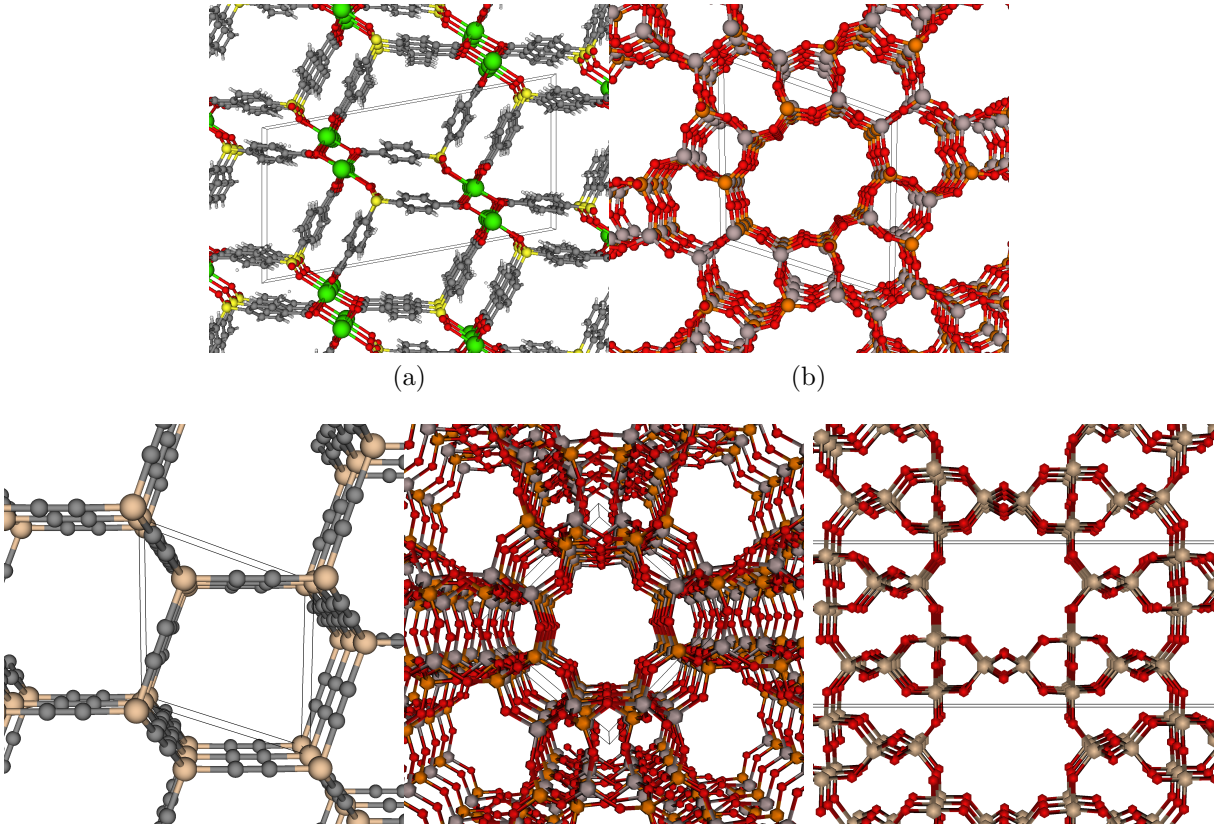


Figure 5.24: Top performing structures. (a, b) Calcium coordination network KAXQIL [234] (a, yellow:S, gray:C, red:O, green:Ca) and aluminophosphate zeolite analogue JAVTAC [235] (b, orange:P, gray:Al, red:O) have the highest simulated selectivities ($S_{Xe/Kr}$ =82). (c) Predicted porous polymer network hPPN_Si_4080_1-net_001 has the highest Xe loading. (tan:Si, gray:C) (d) Aluminophosphate molecular sieve GOMREG has the highest product of selectivity and Xe loading. (orange:P, gray:Al, red:O) (e) A reverse selective predicted zeolite h8166173. (tan:Si, red:O). Reprinted with permission from Ref. [95]. Copyright 2015 American Chemical Society.

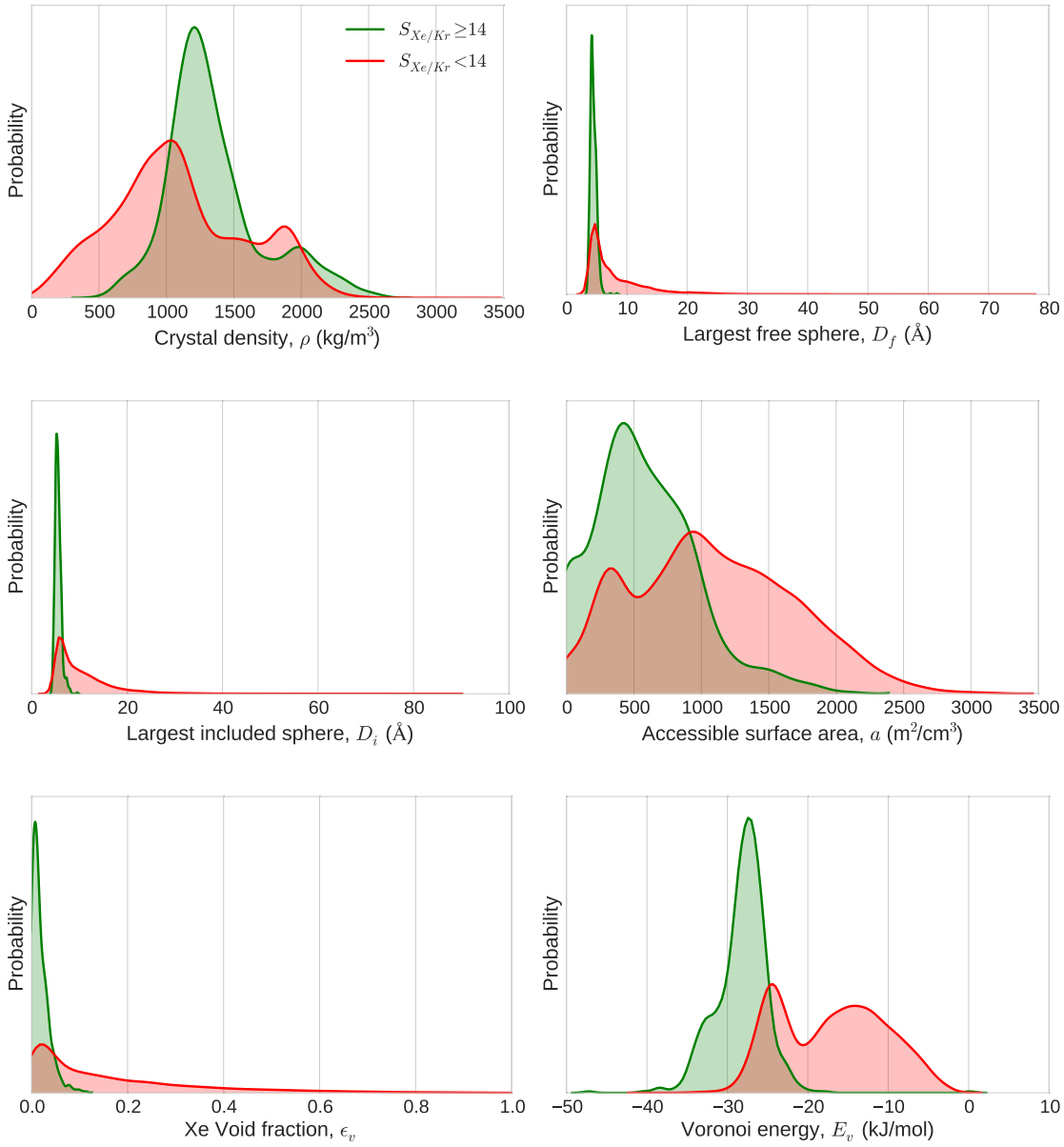


Figure 5.25: Distributions of structural descriptors explored by highly selective ($S_{Xe/Kr} \geq 14$, green) and poorly selective ($S_{Xe/Kr} < 14$, red) materials separately. The distributions are visualized using kernel density estimation with a Gaussian kernel, with the bandwidth chosen by Scott's rule [239]. Reprinted with permission from Ref. [95]. Copyright 2015 American Chemical Society.

Aside from D_i and D_f, the distributions of geometric structural properties among the most selective materials are very broad; if there were a simple recipe of a structural descriptor to yield an optimal material, the distributions would be narrow. Furthermore, the distributions of geometric properties for the poorly and highly selective materials overlap to a large extent. This again motivates using machine learning to look in this high-dimensional feature space to consider all features at once. The exception is the Voronoi energy descriptor, which, while predictive for screening, is not a straightforward quantity to tune in the synthesis of new materials.

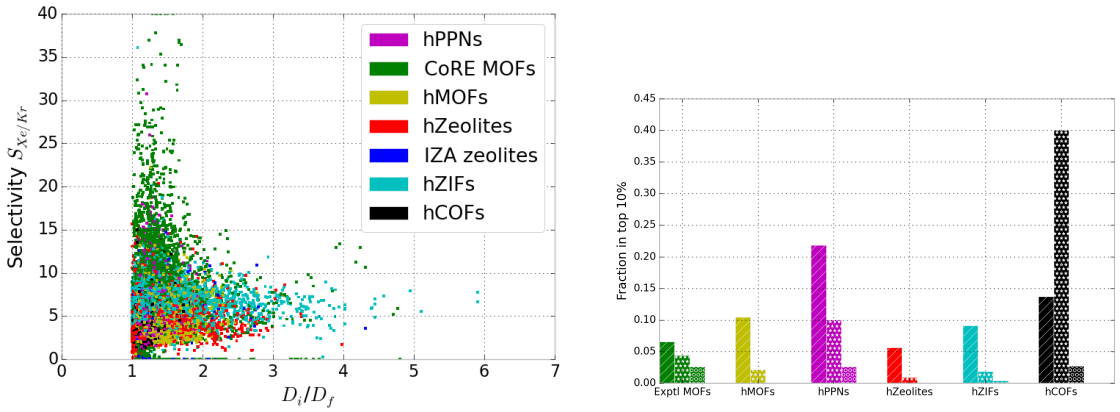
Sikora et al. [76] found that the most selective hMOFs for xenon have cylinder-like pores in contrast to large cavities connected by narrow channels; we found that this generalizes to other classes of materials (Figure 5.26a). We further looked at the *dimensionality* of the channels. We used open source software Zeo++ to calculate the dimension of the channels in the structures in the training and selected sets. Each class of materials consists of structures with a different percentage of 1D, 2D, and 3D channels. To properly compare performance and channel dimensionality, we isolated all materials of class c and channel dimensionality i for $i = 1, 2, 3$. We then looked at the probability $p_{c,i}$ that a structure in class c with channel dimensionality i belongs to the top 10% most selective materials in class c. Fig 5.26b shows $p_{c,i}$, i.e. $P(\text{top 10\% selective } |i, c)$. We found that the probability that a material is within the top 10% most selective materials given that it consists of one-dimensional channels is higher than for two- and three-dimensional channels for all classes of materials except for COFs (See Fig 5.26b). However, 1D channels are not required to form a highly selective binding site.

Our spherical and cylindrical shell models suggested that spherical cages would have higher selectivities than tube-like channels (Figure 5.18b). Inherent in this comparison is that the surface density of atoms η is equivalent in the two shell shapes; however, a spherical shell with a high surface density in all radial directions is not pragmatic, as an opening must exist for the adsorbate to enter the pocket. Also, there is an upper limit to the surface density η that can be obtained in real materials. Due to material design constraints, it may be that this surface density is easier to obtain with 1D tube-like channels.

Binding site analysis

To gain chemical intuition, we detected and analyzed the strongest Xe binding sites in the most selective materials by finding the minimum potential energy position of a xenon atom. Here, we analyze the molecular structure of the most selective binding sites.

First, we create a working definition of which atoms in the structure form the binding site. We expand a sphere centered at the binding site until 90% of the computed potential energy is accounted for (recall we employ a Lennard-Jones cutoff radius of 12.5 Å to define the binding energy); all atoms falling in this sphere are defined to create the binding site. For a selection of materials with the highest selectivities, we display a Xe atom (teal) centered at the binding site with the contributing structure atoms in Figure 5.27. The neighboring bar plot shows the cumulative contribution of each atom type to the binding energy as we



(a) (b)

Figure 5.26: Relationship between pore topology and selectivity. (a) Selectivity for Xe is plotted against the ratio of the largest included to free sphere. Case $D_i/D_f >> 1$: large cavities connected by narrow channels. Case $D_i/D_f \approx 1$: tube-like pores. Consistent with the trend in Ref [76], which motivated this analysis. (b) Probability that a material with a given channel dimensionality will belong to the 10% most selective materials in that class. Diagonal hash, stars, and circles inside the bars represent channel dimension 1, 2, and 3, respectively. For all classes except for COFs, 1D channels are the most likely to belong to the top 10% most selective materials. Reprinted with permission from Ref. [95]. Copyright 2015 American Chemical Society.

expand the sphere. The vertical dashed line is the sphere radius such that 90% of the binding energy is accounted for, and it corresponds to the binding site visualization in the left panel.

Figure 5.27 shows that binding sites can be constructed from a striking diversity of geometries. For example, the binding site in one of the most selective materials, JAVTAC, is constructed inside a 1D channel (a) by forming a cylinder-like *tube* to host Xe atoms. The two structures in (b) form binding sites in a 1D channel using discrete chemical fragments, namely aromatic rings. Both structures suggest that aromatic rings provide a favorable arrangement of carbon atoms to achieve high Xe binding energies; note that C atoms account for most of the binding energy. While in (a) and (b), the structures form a *tube* of optimal diameter to accommodate Xe, the 1D channel in the MOF in (c) exhibits a *pocket* inside the channel with a favorable Xe binding energy. The hPPN in (d) also forms 1D channels, but it creates the binding site in a slightly different manner. With its adamantane core and benzonitrile monomer, it forms a *ring* that provides strong interactions with Xe. The hPPN in (e) exhibits 3D channels. In this hPPN, the binding site is surrounded by a *cage* of structure atoms in all directions.

In summary, binding sites are complex entities, displaying a diversity of often non-discrete chemical fragments and arrangements in space. Based on this analysis of the top materials, we believe it will be difficult to rationally design crystalline materials with such sites embed-

ded in their structures. This underscores the importance of using computational screenings to discover the top materials for Xe/Kr separations.

5.4 Conclusions

We computationally screened a set of 670,000 predicted crystal structures of zeolites, MOFs, ZIFs, PPNs, and COFs as well as already-synthesized zeolites and MOFs for promising materials for Xe/Kr separations at room temperature. The brute-force method of screening a database of materials by performing molecular simulations of Xe/Kr adsorption in each material is prohibitively expensive. To overcome this computational barrier to screening, we used a hybrid molecular simulation- machine learning approach. We characterized each of the structures by more easily-computed geometric properties, namely the largest included and free spheres, void fraction, surface area, and crystal density. We invented a new structural descriptor, the Voronoi energy, that is the average energy of a xenon atom at the Voronoi nodes of the accessible pore space. The Voronoi energy was found to be the most predictive structural descriptor. Using these structural descriptors as a feature vector to serve as a fingerprint of the material in a high dimensional space, we trained a random forest of decision tree regressors to predict selectivity for xenon over krypton by showing the forest a mathematically diverse set of 15,000 training examples. For the remaining 655,000 materials, we predicted the selectivity by running their feature vectors through the trained random forest. We then used molecular simulations to refine the prediction if the random forest predicted the material to be promising. In this manner, we screened the Nanoporous Materials Genome by performing molecular simulations in only 20,000 of the 670,000 structures. Our screening strategy exploits the predictive power of random forests and is a promising high-throughput screening paradigm for expediting the discovery of new materials in the face of rapidly expanding materials datasets. Our screening paradigm can rapidly accelerate high-throughput screenings of the Nanoporous Materials Genome for other applications of porous materials, such as gas storage, gas sensing, drug delivery, and catalysis.

Many materials in our database are predicted to have better Xe/Kr separation performance than CC3 [5], a leading material for Xe/Kr separations [77]. Our models predict that the two most selective materials in the Nanoporous Materials Genome are JAVTAC, an aluminophosphate zeolite analogue [235], and KAXQIL, a calcium coordination network [234]. Both materials have been synthesized but not yet tested for Xe/Kr separations. We hope that our open database of simulated Xe uptake and Xe/Kr selectivities (`http://nanoporousmaterials.org/xekrseparations/`) will inspire the synthesis and characterization of a new material for Xe/Kr separations.

Our analysis of the spherical and cylindrical shell models rationalized the strong link between the selectivity and the pore size. Still, by comparing the structural descriptors of good-performing to poor-performing materials, we found that there is not a simple recipe of geometric descriptors that will guarantee a material to be good for Xe/Kr separations. This motivates using machine learning algorithms to learn the relationship between selectivity

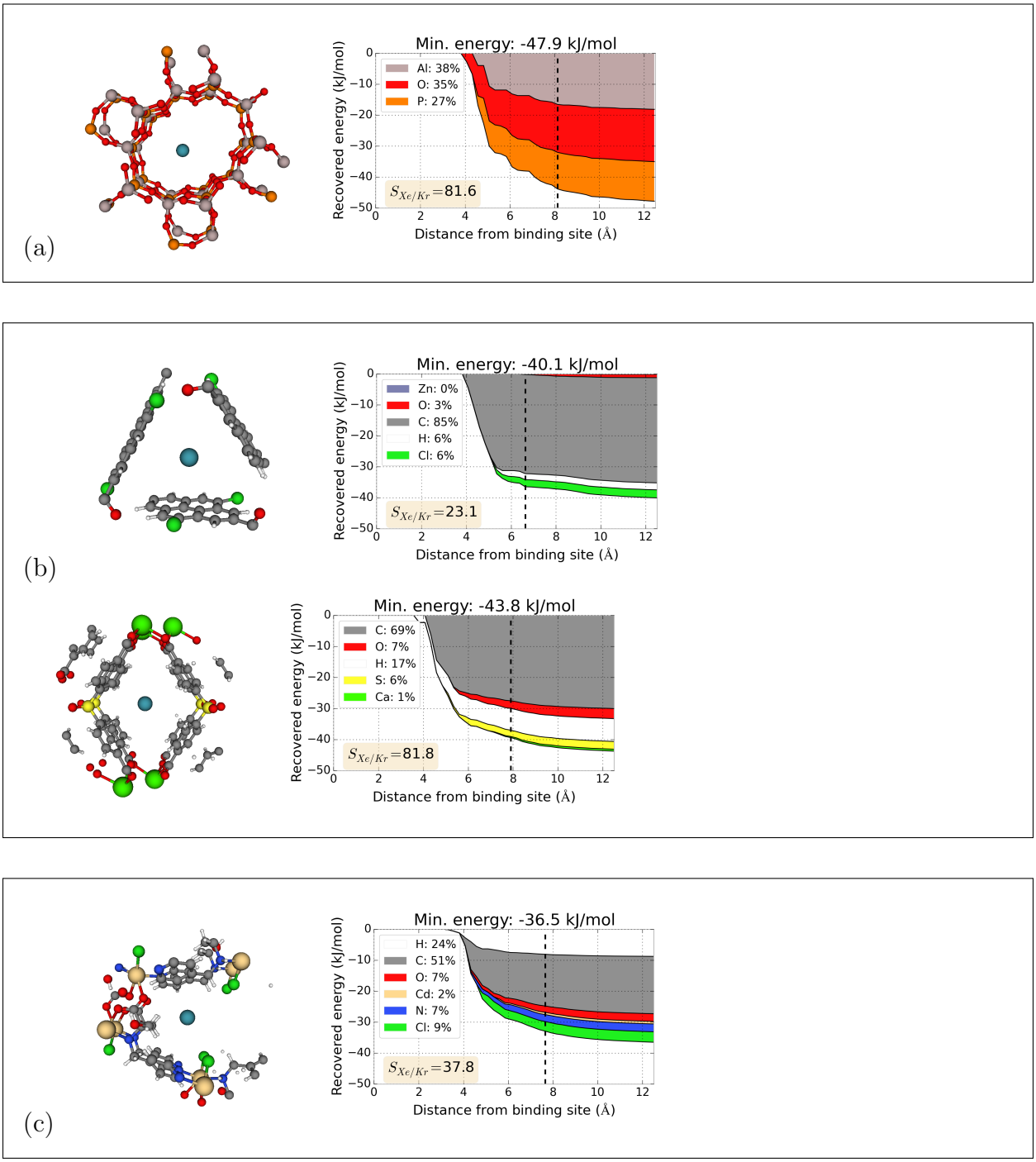

(a)
Min. energy: -47.9 kJ/mol
Al: 38%
O: 35%
P: 27%
$S_{Xe/Kr}$=81.6
Recovered energy (kJ/mol)
Distance from binding site (Å)
(b)
Min. energy: -40.1 kJ/mol
Zn: 0%
O: 3%
C: 85%
H: 6%
Cl: 6%
$S_{Xe/Kr}$=23.1
Recovered energy (kJ/mol)
Distance from binding site (Å)
Min. energy: -43.8 kJ/mol
C: 69%
O: 7%
H: 17%
S: 6%
Ca: 1%
$S_{Xe/Kr}$=81.8
Recovered energy (kJ/mol)
Distance from binding site (Å)
(c)
Min. energy: -36.5 kJ/mol
H: 24%
C: 51%
O: 7%
Cd: 2%
N: 7%
Cl: 9%
$S_{Xe/Kr}$=37.8
Recovered energy (kJ/mol)
Distance from binding site (Å)

Figure 5.27

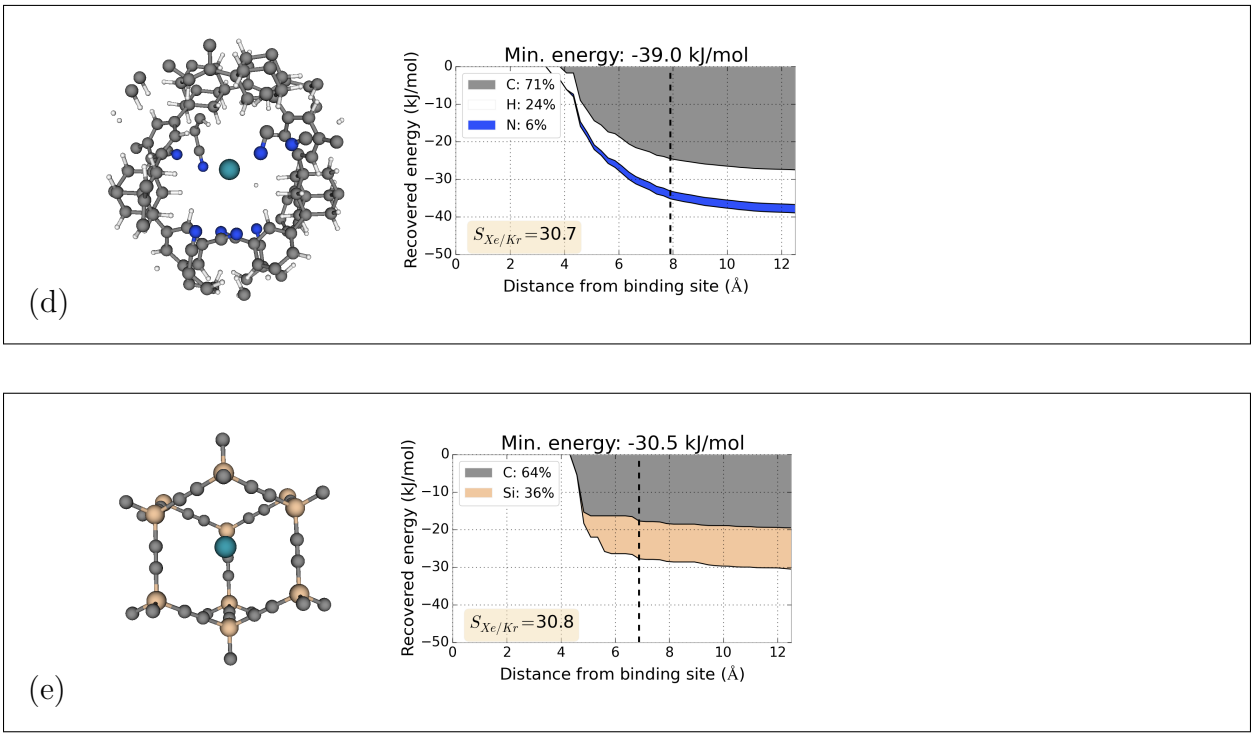


Figure 5.27: Anatomy of binding sites in a few of the most selective materials. (Left column) Xe atom (teal) at the minimum energy position with surrounding atoms contributing to the binding site. (Right column) The binding energy recovered as we include atoms at an increasing distance from the Xe. Colors show the contribution of each atom type to the binding energy; legend shows percentage of binding energy attributed to each atom type. Binding sites can be created by a tube-like arrangement of atoms along 1D channels (a,b). In (b), discrete chemical fragments – aromatic rings – form the binding site. Binding sites can also be *pocket*-like (c), *ring*-like (d), or *shell*-like (e). Structures: CoRE MOF JAVTAC [235], hMOF_15308_i_0_j_17_k_17_m_2_cat_2, CoRE MOF KAXQIL [234], CoRE MOF HEBKEG, hPPN_adamantane_2045_2-net_001, hPPN_Si_4080_1-net_001. Reprinted with permission from Ref. [95]. Copyright 2015 American Chemical Society.

and features in a high dimensional space.

By analyzing the top materials, we revealed chemical insights about their highly selective binding sites. They can be constructed from a diverse array of often non-discrete chemical fragments, arranged into tubes, cages, pockets, or rings. The complexity of these binding sites makes rational design difficult and underscores the importance of high-throughput screening for the discovery of novel and/or optimal materials for a given application.

Chapter 6

pyIAST: Computer program for ideal adsorbed solution theory (IAST)

Ideal adsorbed solution theory (IAST) is a widely-used thermodynamic framework to readily predict mixed-gas adsorption isotherms from a set of pure-component adsorption isotherms. We present an open-source, user-friendly Python package, pyIAST, to perform IAST calculations for an arbitrary number of components. pyIAST supports several common analytical models to characterize the pure-component isotherms from experimental or simulated data. Alternatively, pyIAST can use numerical quadrature to compute the spreading pressure for IAST calculations by interpolating the pure-component isotherm data. pyIAST can also perform reverse IAST calculations, where one seeks the required gas phase composition to yield a desired adsorbed phase composition.

Source code: https://github.com/CorySimon/pyIAST

Documentation: http://pyiast.readthedocs.org/en/latest/

This chapter is based on the following publication [240]:

C. Simon, B. Smit, M. Haranczyk. pyIAST: Ideal Adsorbed Solution Theory (IAST) Python package. *Computer Physics Communications*. (2016) 200: 364-380

6.1 Introduction

For the majority of adsorption-based applications of porous materials, adsorption occurs in the presence of a *mixture* of gases [241]. Even for (pure) gas storage applications, the effect of impurities must typically be considered (i.e. a mixture) [242, 243, 244].

Pure-component adsorption isotherms are routinely measured with high accuracy using commercial instruments [245]. On the other hand, accurate measurements of adsorption isotherms in the presence of a mixture of gases are complicated and time-consuming [241, 246] and require custom-built instruments [245]. To exacerbate this problem, modeling

pressure and temperature swing adsorption processes requires mixed-gas adsorption data at many different gas compositions [247]; the required number of experiments to characterize composition space quickly becomes intractable as the number of components increases.

Ideal adsorbed solution theory (IAST), developed by Myers and Prausnitz [248] in 1965, is a widely used [249] thermodynamic framework for readily predicting multi-component adsorption isotherms from only the pure-component adsorption isotherms at the same temperature. IAST rests upon the assumption that the adsorbed species form an ideal mixture, which is a reasonable approximation in many systems [248, 250, 251, 252, 253, 254, 255, 256, 257, 258]. In these applicable cases, IAST is a powerful tool that can reduce the need for laborious mixed-gas adsorption measurements. Further, for cases where one is interested in several different mixture conditions, taking an array of laborious multi-component adsorption isotherm measurements may be impractical. Even for molecular simulations, simulating pure-component adsorption isotherms and applying IAST can be faster than performing multi-component grand-canonical Monte Carlo simulations [110] depending on the number of conditions of interest.

In this work, we release a Python package, pyIAST, under an MIT license to perform IAST calculations to predict mixed-gas adsorption from pure-component adsorption isotherms. The IAST framework does not enforce any specific adsorption model for the pure-component species. In practice, we have experimental or simulated data to characterize the pure-component adsorption isotherms. pyIAST can fit standard analytical isotherm models (Langmuir, Quadratic, BET [259], approximated Temkin isotherm [260], Henry's law) to this data to characterize the pure-component isotherms. Additionally, pyIAST can interpolate the isotherm data for IAST calculations, circumventing the burden of finding an appropriate analytical model and scrutinizing the quality of its fit to the data [198, 261]. pyIAST can handle an arbitrary number of components in the mixture. In addition, pyIAST performs reverse IAST calculations, where one calculates the required gas phase composition to yield a desired adsorbed phase composition. This feature is useful in catalysis, where one seeks to control the adsorbed phase composition to yield the appropriate stoichiometry in the adsorbed phase [262].

We first review the governing equations of IAST by deriving them from thermodynamics. We then discuss the two approaches to characterize the pure-component adsorption isotherms from data for input to pyIAST. To showcase our pyIAST package, we apply IAST to a gaseous mixture of methane and ethane in equilibrium with an adsorbed phase in a nanoporous material. This case study also serves as a verification of our code by comparing the IAST result to binary grand-canonical Monte Carlo simulations of methane/ethane adsorption. We further verify pyIAST by comparing to the competitive Langmuir isotherm, which is an analytical solution to IAST when the pure-component isotherms follow Langmuir models with the same saturation loading. Finally, we showcase pyIAST in predicting a tertiary mixture adsorption isotherm from pure CO_2, N_2, and H_2O isotherms and compare to a rare experimentally measured mixed-gas adsorption isotherm by Mason et al. [245].

6.2 Ideal Adsorbed Solution Theory (IAST)

Below, we briefly review IAST. This also allows us to define our notation and introduce the governing equations that are solved by pyIAST.

Consider a gaseous mixture of N chemical species at a pressure P_T in equilibrium with an adsorbed phase on a surface. Let y_i and x_i be the mole fractions in the gas phase and adsorbed phase, respectively, for species $i = 1, 2, ..., N$. We assume that we have experimental or simulated data characterizing the pure-component adsorption isotherms, $n_i^\circ(P)$, the moles of gas adsorbed on the surface when in equilibrium with a gas phase of pure species i as a function of its pressure P. In ideal adsorbed solution theory, the goal is to use $n_i^\circ(P)$ for $i = 1, 2, ..., N$ to predict the moles of gas of each species adsorbed on the surface when in equilibrium with the multicomponent gaseous mixture, $n_i = n_i(P_T, \{y_j\})$. IAST assumes throughout that the temperature T is fixed and that the pure-component isotherms are measured at the same temperature as the mixed-gas of interest. In Sec 6.3, we will address how to translate pure-component isotherm data into a representative function $n_i^\circ(P)$ for IAST calculations.

For an isothermal physical adsorption process, we assume (1) the change in each thermodynamic property of the adsorbent when gas molecules adsorb is negligible compared to the change for the adsorbate, (2) each adsorbed species has access to the same area of the adsorbent surface, and (3) a Gibbs dividing surface defines an adsorbed phase. [248]

The first law of thermodynamics in differential form for the Gibbs free energy $G = G(T, \pi, \{n_i\})$ of the adsorbed phase is:

$$dG = -SdT + Ad\pi + \sum_{i=1}^{N} \mu_i dn_i, \tag{6.1}$$

where A is the area of the surface and π is the spreading pressure. The spreading pressure is the analogy of pressure in two dimensions; πdA is therefore the work done on the surroundings in the conceptual process of increasing the area of the adsorbent surface [248]. To further extend this analogy, in a two-dimensional ideal gas, the equation of state is $\pi A = Nk_BT$.

The ideal solution assumption is that the enthalpy and area change upon mixing of the two adsorbate species is zero. In this case, the Gibbs free energy of a multi-component adsorbed phase at spreading pressure π is:

$$G(T, \pi, \{n_j\}) = \sum_{i=1}^{N} G_i(T, \pi, n_i) - T\Delta S_{mix}, \tag{6.2}$$

where $G_i(T, \pi, n_i)$ is the Gibbs free energy of component i in its pure state at the same temperature and spreading pressure as the mixture; thus the Gibbs free energy change upon mixing is determined by the entropy change upon mixing $\Delta S_{mix} = -R\sum_i n_i \log x_i$ [109], where R is the universal gas constant. As the chemical potential of adsorbed species i, μ_i,

is related to the Gibbs free energy by $\mu_i = \left(\frac{\partial G}{\partial n_i}\right)_{T,\pi,n_{j\neq i}}$, it follows from eqn 6.2 that

$$\mu_i(T,\pi,\{x_j\}) = \mu_i^\circ(T,\pi) + RT\log x_i. \tag{6.3}$$

Here, $\mu_i^\circ(T,\pi)$ is the chemical potential of adsorbed pure species i at the same temperature and spreading pressure as the mixture.

At thermodynamic equilibrium, the chemical potentials in the adsorbed and gaseous phase are equal; the gas phase thus can be thought to impose the chemical potential. The chemical potential of species i in an ideal gas (denoted by superscript g) is:

$$\mu_i^g(T,P,y_i) = \mu_i^{g,0}(T) + RT\log p_i, \tag{6.4}$$

where $p_i = y_i P_T$ is the partial pressure of species i in the gas phase and $\mu_i^{g,0}(T)$ is the reference chemical potential. Equating $\mu_i^g(T,P,y_i)$ from eqn 6.4 and $\mu_i(T,\pi,\{x_i\})$ for the adsorbed phase from eqn 6.3, we obtain:

$$p_i = x_i e^{\frac{\mu_i^\circ(T,\pi)-\mu_i^{g,0}(T)}{RT}}. \tag{6.5}$$

Consider the limit $x_i \to 1$, a limit of pure species i. In this limit, p_i becomes the pressure in the gas phase required to yield a spreading pressure π of pure species i in the adsorbed phase. Defining this particular pressure as $x_i \to 1$ to be $p_i^\circ = p_i^\circ(\pi,T)$, we see that the exponential term is equivalent to p_i°. With this, we arrive at the analogy of Raoult's law for an adsorbed phase in equilibrium with a gaseous phase:

$$p_i = x_i p_i^\circ(\pi,T) \text{ for } i = 1,2,...,N. \tag{6.6}$$

By definition of p_i°, it follows that the pure-component spreading pressures π_i at pressure p_i° are all equal to the spreading pressure of the mixture π:

$$\pi = \pi_1(p_1^0) = \pi_2(p_2^0) = \cdots = \pi_N(p_N^0). \tag{6.7}$$

Given expressions for the spreading pressure of a pure-component adsorbed phase, $\pi_i(p_i^\circ)$, eqns 6.6 and 6.7 form $N+N-1$ equations for $2N$ unknowns, $\{p_i^\circ\}$ and $\{x_i\}$. The knowns here are $\{y_i\}$ and P_T, the mole fractions in and the total pressure of the gas phase, respectively. These equations can then be solved for $\{p_i^\circ\}$ and $\{x_i\}$ with the additional constraint $\sum_i x_i = 1$.

We here show that we can indirectly calculate the spreading pressure $\pi_i(p_i^\circ)$ from the pure-component adsorption isotherm $n_i^\circ(P)$. By Euler's theorem for homogeneous functions, the Gibbs free energy for a pure-component system of species i is $G = \mu_i^\circ n_i^\circ$. Taking the differential of $G = \mu_i^\circ n_i^\circ$ and subtracting eqn 6.1 for a pure-component system, we obtain the Gibbs-Duhem equation for a pure-component adsorbed phase at constant temperature:

$$-Ad\pi_i + n_i^\circ d\mu_i^\circ = 0 \text{ at constant } T. \tag{6.8}$$

Since the gas phase imposes the chemical potential, $\mu_i^\circ = \mu_i^g(T, p_i^\circ, 1)$ by eqn 6.4, and $d\mu_i^\circ = RTd\log p_i^\circ$. Thus, we can rewrite the Gibbs-Duhem relation as:

$$Ad\pi_i = RT\frac{n_i(p_i^\circ)}{p_i^\circ}dp_i^\circ. \tag{6.9}$$

Integrating eqn 6.9, we obtain an expression for the spreading pressure of adsorbed pure-component i in equilibrium with a pure-component i gas phase at pressure p_i°:

$$\pi_i(p_i^\circ) = \frac{RT}{A}\int_0^{p_i^\circ}\frac{n_i^\circ(P)}{P}dP. \tag{6.10}$$

We used $\pi_i(0) = 0$ in the integration limit. Eqn 6.10 describes how we use the pure-component adsorption isotherms $n_i^\circ(P)$ to obtain the composition $\{x_i\}$ of the adsorbed phase through eqns 6.6 and 6.7.

We also desire the absolute gas adsorption $n_i(P, \{y_i\})$. Here, we derive an expression for the total moles of gas adsorbed on the surface, n_T. It follows that $n_i = x_i n_T$. The Gibbs-Duhem relation for the adsorbed mixture is:

$$-Ad\pi + \sum_{i=1}^{N} n_i d\mu_i = 0 \text{ at constant } T. \tag{6.11}$$

Dividing by n_T,

$$\frac{A}{n_T} = \sum_i x_i \left(\frac{\partial \mu_i}{\partial \pi}\right)_{T,x_j}. \tag{6.12}$$

The derivative can be simplified using eqn 6.3 [263]:

$$\left(\frac{\partial \mu_i}{\partial \pi}\right)_{T,x_j} = \left(\frac{\partial \mu_i^\circ}{\partial \pi}\right)_T = \frac{A}{n_i^\circ}, \tag{6.13}$$

and the equality on the right follows from the Gibbs-Duhem equation of the pure-component system in eqn 6.8. Eqns 6.12 and 6.13 combined yield an expression to solve for the total moles of gas adsorbed n_T from the pure-component adsorption isotherms $n_i^\circ(P)$, $\{p_i^\circ\}$, and $\{x_i\}$ obtained above:

$$\frac{1}{n_T} = \sum_{i=1}^{N}\frac{x_i}{n_i^\circ(p_i^\circ)}. \tag{6.14}$$

The relationship in eqn 6.14 is a two-dimensional analogy to Amagat's law of additive volumes for a three-dimensional system [263]. The expression A/n_T is the area per adsorbed molecule in the mixture; $A/n_i^\circ(p_i^\circ)$ is the area per adsorbed molecule in the pure-component system at the same temperature and spreading pressure as the mixture. Eqn 6.14 thus expresses that the area per adsorbate in the mixture is the weighted average of the area allocated to each adsorbate in the pure-component systems at the same spreading pressure;

this is consistent with the ideal solution assumption that there is no area change upon mixing [248].

This concludes the derivation of ideal adsorbed solution theory. Given the functions for the pure-component adsorption isotherms $\{n_i^\circ(P)\}$ and a specified bulk gas at pressure P_T and composition $\{y_i\}$, we solve for $\{x_i\}$ and $\{p_i^\circ\}$ using the nonlinear system of equations in eqns 6.6 and 6.7. The pure-component adsorption isotherms allow us to calculate the spreading pressures via eqn 6.10. Finally, we can compute the total number of moles adsorbed using eqn 6.14. The computational strategy we use to solve the nonlinear system of equations closely follows that in Ref. [264] outlined for a binary mixture. Fig 6.1 shows this workflow of an IAST calculation in pyIAST.

Note that we can account for nonideality (i) in the gas phase by replacing the pressure P in the above equations with the fugacity and (ii) in the adsorbed phase by replacing the adsorbed phase mole fractions x_i by $x_i\gamma_i$, where γ_i is the activity coefficient of species i in the adsorbed mixture. [248]

6.3 Characterizing the pure-component adsorption isotherm $n_i^\circ(P)$

For IAST calculations, we need the pure-component adsorption isotherms $n_i^\circ(P)$ to compute the integral in eqn 6.10 to obtain the spreading pressure; in practice, we have a set of experimental or simulated data points to characterize $n_i^\circ(P)$. pyIAST can translate this pure-component isotherm data into a function $n_i^\circ(P)$ in two disparate ways: (1) fit an analytical model to the data and (2) linearly interpolate the data. For the first option, we solve the integral in eqn 6.10 analytically for the given model and evaluate it using the fitted model parameters. For the second option, we compute the integral in eqn 6.10 using numerical quadrature. In Sec 6.3, we discuss the advantages and disadvantages of each of the two approaches.

Analytical models

The first option is to fit the simulated or experimental pure-component adsorption isotherm data to an analytical model $n_i^\circ(P)$ using a least-squares loss function. From this analytical model and identified parameters, we can write an analytical formula for the spreading pressure in eqn 6.10. pyIAST can fit several common [264] analytical models to the data: Langmuir, quadratic, BET, and approximated Temkin adsorption isotherms and Henry's law. See Fig 6.2 for caricatures of these adsorption isotherm models. The list `pyiast._MODELS` lists all models implemented in pyIAST, which we review here. Note that pyIAST supports only thermodynamically consistent isotherm models [265, 266] that obey Henry's law at low coverage, i.e. models such that $\lim_{P\to 0} \frac{dn_i^\circ}{dP}$ is finite. For example, pyIAST does not support the empirical [267] Freundlich isotherm [268]. In each section title, we note the keyword for the respective model in pyIAST.

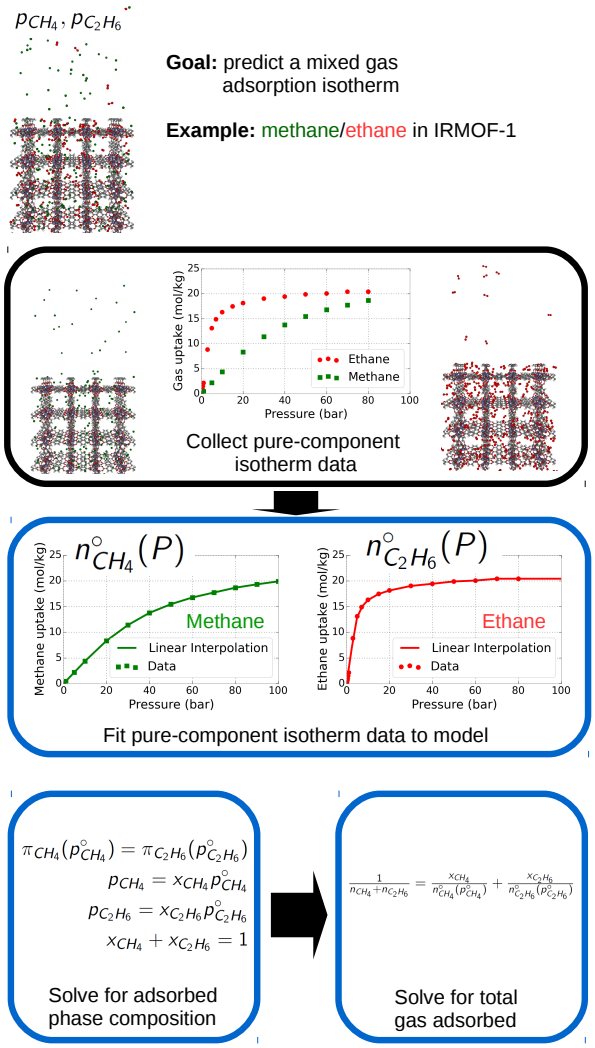


Figure 6.1: Work flow of pyIAST. For this example, the goal is to predict an adsorption isotherm in a nanoporous material, IRMOF-1, immersed in a gaseous mixture of methane and ethane. First, one measures or simulates the pure methane and pure ethane adsorption isotherms. Then, pyIAST will translate the data into a function $n_i^\circ(P)$ for i =CH_4, C_2H_6 by fitting to an analytical model or interpolating the isotherm. For a given gas phase composition and pressure, pyIAST first solves eqns 6.6 and 6.7 for $\{x_i\}$ and $\{p_i^\circ\}$, and then solves for the total gas adsorption by eqn 6.14. Reprinted from Computer Physics Communications, 200, C. Simon, B. Smit, M. Haranczyk, pyIAST: Ideal Adsorbed Solution Theory (IAST) Python package, 364-380, Copyright (2016), with permission from Elsevier.

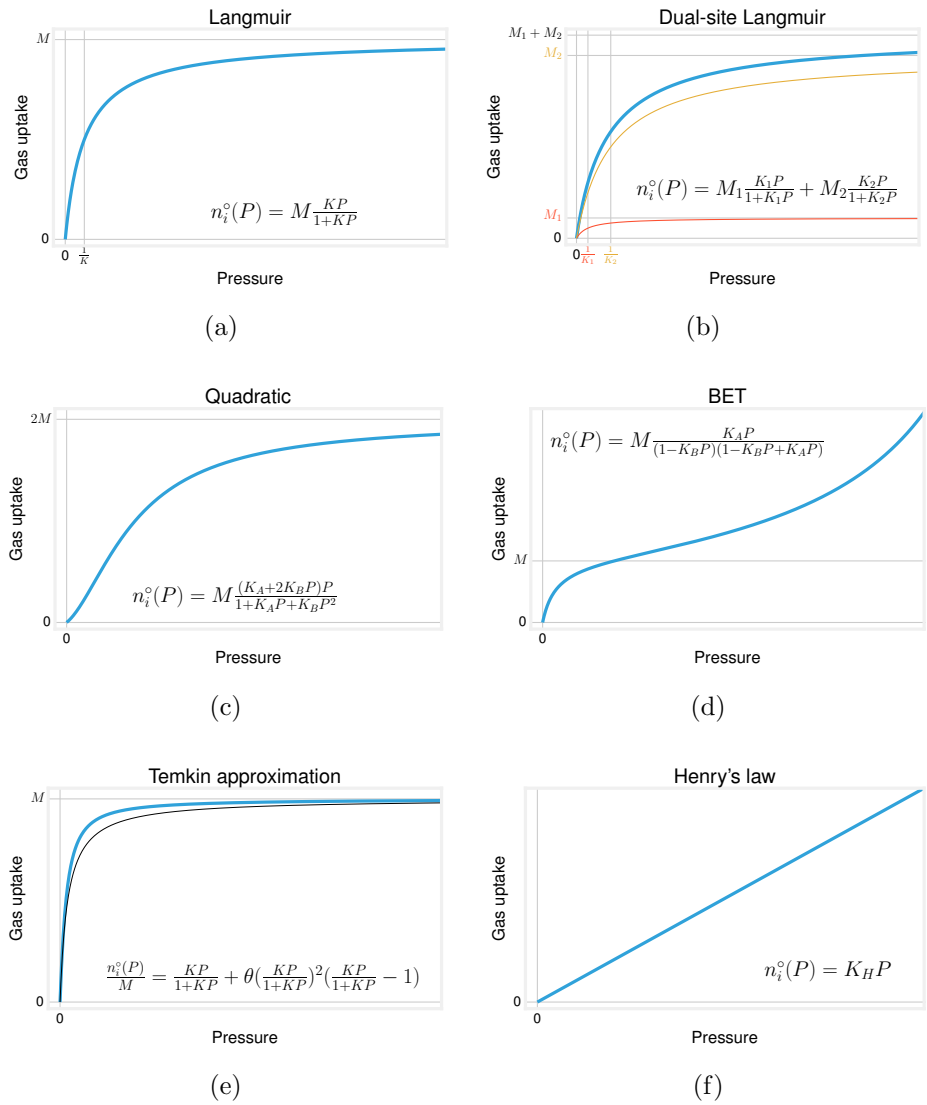


Figure 6.2: Caricatures of adsorption isotherm models implemented in pyIAST. (a) Langmuir, (b) Dual-site Langmuir (c) Quadratic, (d) BET (e) Asymptotic approximation to the Temkin isotherm ($\theta < 0$ here for adsorbate-adsorbate attractions) (f) Henry's law. In each saturating isotherm, M is marked on the y-axis. For (b), we plotted the Langmuir isotherm for sites of type 1 and 2 separately. For (e), we plotted the Langmuir isotherm for comparison, which corresponds to $\theta = 0$, as the solid, black curve. Reprinted from Computer Physics Communications, 200, C. Simon, B. Smit, M. Haranczyk, pyIAST: Ideal Adsorbed Solution Theory (IAST) Python package, 364-380, Copyright (2016), with permission from Elsevier.

Langmuir model, `model="Langmuir"`

The Langmuir adsorption isotherm assumes a single layer of molecules can adsorb onto M distinct, identical, independent adsorption sites; adsorbed molecules do not interact except through volume exclusion. The Langmuir model has an addition parameter K (units: pressure^{-1}) that describes the affinity of the adsorbate with the adsorbent (see Fig 6.2a):

$$n_i^\circ(P) = M\frac{KP}{1+KP}. \tag{6.15}$$

As half of the adsorption sites are occupied at $P = 1/K$, the Langmuir constant K is inversely related to the pressure at which the isotherm saturates. At low pressures, the Langmuir model is approximated by Henry's law $n_i^\circ(P) \sim MKP$.

The spreading pressure for a Langmuir adsorption isotherm via eqn 6.10 is

$$\frac{A}{RT}\pi_i(P) = M\log(1+KP). \tag{6.16}$$

Statistical mechanical derivations of the Langmuir adsorption isotherm can be found in Refs. [260, 269].

Dual-site Langmuir model, `model="DSLangmuir"`

The dual-site Langmuir model is a first attempt to model an adsorbent surface with energetic heterogeneity. In this scenario, the adsorbent surface has two different types of adsorption sites. As in the Langmuir model, we neglect adsorbate-adsorbate interactions and take each adsorption site to be independent, implying that the dual-site Langmuir adsorption isotherm is the sum of the Langmuir isotherm models for the two types of sites:

$$N(P) = M_1\frac{K_1P}{1+K_1P} + M_2\frac{K_2P}{1+K_2P}, \tag{6.17}$$

where M_i is the number of adsorption adsorption sites of type i, which have Langmuir constant K_i, for $i \in \{1, 2\}$.

The spreading pressure for a dual-site Langmuir adsorption isotherm via eqn 6.10 is

$$\frac{A}{RT}\pi_i(P) = M_1\log(1+K_1P) + M_2\log(1+K_2P). \tag{6.18}$$

Quadratic model, `model="Quadratic"`

The quadratic adsorption isotherm exhibits an inflection point; the loading is convex at low pressures but changes concavity as it saturates, yielding an S-shape. See Fig 6.2c. The S-shape can be explained by adsorbate-adsorbate attractive forces [270]; the initial convexity is

due to a cooperative effect [271] of adsorbate-adsorbate attractions aiding in the recruitment of additional adsorbate molecules. The quadratic isotherm model is:

$$n_i^\circ(P) = M\frac{(K_A + 2K_BP)P}{1 + K_AP + K_BP^2}, \tag{6.19}$$

where $2M$ is the saturation loading and K_A (units: pressure^{-1}) and K_B (units: pressure^{-2}) are constants. The spreading pressure for a quadratic isotherm via eqn 6.10 is:

$$\frac{A}{RT}\pi_i(P) = M\log(1 + K_AP + K_BP^2). \tag{6.20}$$

A statistical mechanical derivation of the quadratic adsorption isotherm can be found in [269]. Here, pairs of adsorbate molecules experience attractive forces. The parameter K_A can be interpreted as the Langmuir constant; the strength of the adsorbate-adsorbate attractive forces is embedded in K_B.

Brunauer-Emmett-Teller (BET) model, `model="BET"`

In the Brunauer-Emmett-Teller (BET) adsorption isotherm model [259], we have M independent, identical adsorption sites on the surface – as in the Langmuir model – but now adsorbate molecules are allowed to stack to form an indefinite number of layers. The multilayer BET adsorption isotherm model is:

$$n_i^\circ(P) = M\frac{K_AP}{(1 - K_BP)(1 - K_BP + K_AP)}, \tag{6.21}$$

where K_A is the Langmuir constant for the first layer of adsorbate molecules in direct contact with the surface, and K_B is the constant for the second and higher layers of adsorbate molecules.

The spreading pressure for a BET isotherm via eqn 6.10 is:

$$\frac{A}{RT}\pi_i(P) = M\log\left(\frac{1 - K_BP + K_AP}{1 - K_BP}\right). \tag{6.22}$$

A statistical mechanical derivation of the BET adsorption isotherm can be found in Ref. [269].

Asymptotic Approximation to the Temkin model, `model="TemkinApprox"`

The Temkin adsorption isotherm [272], like the Langmuir model, considers a surface with M identical adsorption sites, but takes into account adsorbate-adsorbate interactions by

assuming that the heat of adsorption is a linear function of the coverage. We independently derived the Temkin isotherm in Ref. [260] using a mean-field argument and used an asymptotic approximation to obtain an explicit equation for the loading:

$$n_i^\circ(P) = M\left(\frac{KP}{1+KP} + \theta\left(\frac{KP}{1+KP}\right)^2\left(\frac{KP}{1+KP} - 1\right)\right). \tag{6.23}$$

Here, M and K have the same physical meaning as in the Langmuir model. The additional parameter θ describes the strength of the adsorbate-adsorbate interactions ($\theta < 0$ for attractions). Fig 6.2e shows how eqn 6.23 is a perturbation on the Langmuir model for $\theta < 0$; the effect of adsorbate-adsorbate interactions are most significant at intermediate coverages.

The spreading pressure of the isotherm model in eqn 6.23 via eqn 6.10 is:

$$\frac{A}{RT}\pi_i(P) = M\left(\log(1+KP) + \theta\frac{2KP+1}{2(KP+1)^2}\right). \tag{6.24}$$

Henry's law, `model="Henry"`

Henry's law is:

$$n_i^\circ(P) = K_H P, \tag{6.25}$$

where K_H is the Henry coefficient (units: loading/pressure). Henry's law describes adsorption only in the low coverage regime because $n_i^\circ(P)$ unrealistically increases indefinitely with P; thus caution must be used when extrapolating experimental data with Henry's law. The spreading pressure for Henry's law is $\frac{A}{RT}\pi_i(P) = K_H P$.

Remark

For some combinations of analytical models for pure-component adsorption isotherms, IAST can be used to obtain an analytical expression for the mixed-gas adsorption isotherm. For example, if all single-component adsorption isotherms obey the Langmuir model in eqn 6.15 with Langmuir constant K_i and exhibit the same saturation loading M – often a strong assumption – IAST recovers the competitive Langmuir isotherm [273]:

$$n_i(P, \{y_j\}) = M\frac{K_i p_i}{1 + \sum_j K_j p_j}, \tag{6.26}$$

where $p_i = y_i P_T$ is the partial pressure of species i in a gas mixture with total pressure P_T. An analytical model for the multi-component adsorption isotherm is more computationally efficient because it circumvents the need to solve the nonlinear system of equations in eqns 6.6 and 6.7. Other derivations for mixture adsorption isotherms can be found in the literature for when the pure-component adsorption isotherm models follow: Langmuir and Freundlich isotherms [273], quadratic isotherms [274], BET and Langmuir isotherms [275], and quadratic and BET and quadratic and Langmuir for the special case when saturation loadings are equal [264].

Numerical quadrature

An alternative to assuming the pure-component adsorption isotherm follows a given functional form is to linearly interpolate the pure-component adsorption isotherm data to approximate $n_i^\circ(P)$ and use numerical quadrature [276] to evaluate the integral in eqn 6.10 to obtain the spreading pressure for IAST calculations.

Consider that we have L data points of the pressure P and uptake n_i° that characterize the adsorption isotherm, $\{(P_j, n_i^\circ(P_j))\}$, $j = 1, 2, ..., L$, where $P_1 < P_2 < \cdots < P_{L-1} < P_L$. Fig 6.3 illustrates the numerical quadrature method to compute $\frac{A}{RT}\pi_i(p_i^\circ)$ from the integral in eqn 6.10, where $p_i^\circ \leq P_L$; the circular points in Fig 6.3 represent $\{(P_j, n_i^\circ(P_j)/P_j)\}$. For interested readers, we outline the details of the numerical quadrature below.

We tackle the integral by splitting the integration into three parts:

$$\frac{A}{RT}\pi_i(p_i^\circ) = \int_0^{P_1} \frac{n_i^\circ(P)}{P} dP + \sum_{j=1}^{k-1} \int_{P_j}^{P_{j+1}} \frac{n_i^\circ(P)}{P} dP + \int_{P_k}^{p_i^\circ} \frac{n_i^\circ(P)}{P} dP, \tag{6.27}$$

where index k is defined such that $P_k \leq p_i^\circ < P_{k+1}$.

The integral $\int_0^{P_1}$ requires extrapolation of the isotherm for $P < P_1$. We assume that Henry's law $n_i^\circ(P) = K_H P$ is obeyed in the region $P \in [0, P_1]$, with the Henry coefficient K_H estimated from the first data point, $K_H = n_i^\circ(P_1)/P_1$. This gives us the limiting behavior:

$$\lim_{P \to 0} \frac{n_i^\circ(P)}{P} = K_H, \tag{6.28}$$

which is added to Fig 6.3 as the star. Thus,

$$\int_0^{P_1} \frac{n_i^\circ(P)}{P} dP \approx \int_0^{P_1} K_H dP = n_i^\circ(P_1), \tag{6.29}$$

equal to the area of the left-most green-shaded rectangle in Fig 6.3.

The integrals $\int_{P_j}^{P_{j+1}}$ are computed by approximating $n_i^\circ(P)$ by linearly interpolating the data points $(P_j, n_i^\circ(P_j))$ and $(P_{j+1}, n_i^\circ(P_{j+1}))$. That is, $n_i^\circ(P) \approx m_j P + b_j$ for $P \in [P_j, P_{j+1}]$, where m_j is the slope and b_j is the intercept of the line passing through the points $(P_j, n_i^\circ(P_j))$ and $(P_{j+1}, n_i^\circ(P_{j+1}))$. Then,

$$\int_{P_j}^{P_{j+1}} \frac{n_i^\circ(P)}{P} \approx m_j(P_{j+1} - P_j) + b_j \log(P_{j+1}/P_j), \tag{6.30}$$

represented by the areas of the next three green-shaded segments in Fig 6.3.

The last portion of the integral, $\int_{P_k}^{p_i^\circ}$, is computed analogously to eqn 6.30, except now m_k and b_k are the slope and intercept of the line passing through the points $(P_k, n_i^\circ(P_k))$ and $(p_i^\circ, \hat{n_i^\circ}(p_i^\circ))$, where $\hat{n_i^\circ}(p_i^\circ)$ is the linearly interpolated value of n_i° at p_i°. This portion of the integral is the right-most green-shaded segment in Fig 6.3.

Depending on the composition and pressure of the gas we are investigating with IAST, we may need knowledge of $\pi_i(p_i^\circ)$ for $p_i^\circ > P_L$, and thus $n_i^\circ(p_i^\circ)$ for $p_i^\circ > P_L$. In this case, the linear interpolation method is inapplicable due to a lack of data for $P > P_L$, unless we assume some behavior for $n_i^\circ(P)$ for $P > P_L$. We will see that an option in pyIAST is to approximate $n_i^\circ(P)$ as a constant for $p_i^\circ > P_L$, which is reasonable if the isotherm data exhibits a plateau at the last few pressure points. Conveniently, the spreading pressure is most sensitive to the low-coverage regime of the pure-component adsorption isotherm [248].

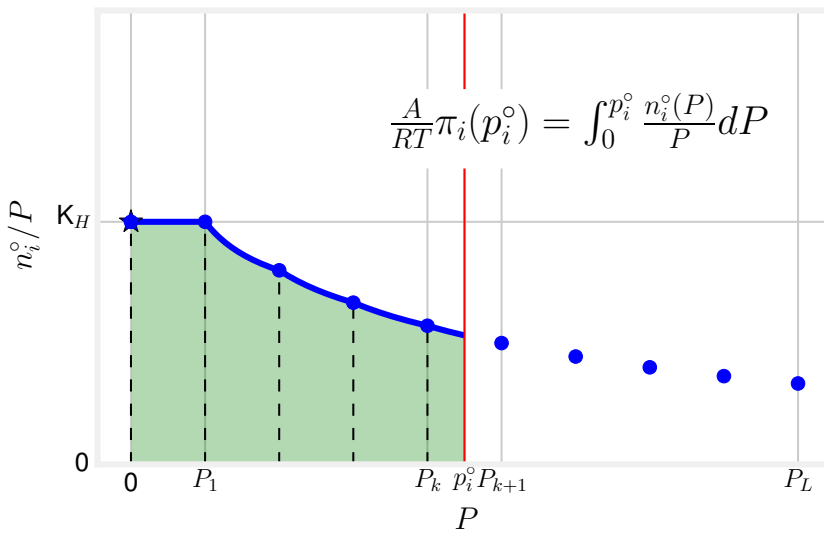


Figure 6.3: Numerical quadrature to compute the spreading pressure $\pi_i(p_i^\circ)$. Using the data $\{(P_j, n_i^\circ(P_j))\}$, $j = 1, 2, ..., L$ that characterize the pure-component adsorption isotherm, the circular points in this visualization are the points $\{(P_j, n_i^\circ(P_j)/P_j)\}$. We model the uptake before the first pressure point with Henry's law and estimate the Henry coefficient K_H from the first data point, $K_H \approx n_i^\circ(P_1)/P_1$. This gives us the limiting behavior of the integrand $\lim_{P\to 0} n_i^\circ(P)/P = K_H$, plotted as the star. The red, vertical line denotes p_i°, which by definition of k falls between P_k and P_{k+1}. The blue curve between P_1 and p_i° is the function $n_i^\circ(P)/P$ approximated by linearly interpolating the data $\{(P_j, n_i^\circ(P_j))\}$. Using numerical quadrature to compute eqn 6.10, $\frac{A}{RT}\pi_i(p_i^\circ)$ is approximately the area shaded in green. Reprinted from Computer Physics Communications, 200, C. Simon, B. Smit, M. Haranczyk, pyIAST: Ideal Adsorbed Solution Theory (IAST) Python package, 364-380, Copyright (2016), with permission from Elsevier.

Fit an analytical model or use numerical quadrature?

Here, we address the question of whether to fit an analytical model to the data or interpolate the data to represent the pure-component adsorption isotherm $n_i^\circ(P)$. The central

issues that can lead to inaccuracy for IAST calculations are (1) imprecisely representing the pure-component adsorption isotherm data and (2) computing the pure-component spreading pressure beyond the range of the data available (extrapolating) [261]. We first discuss issue (1).

Consider fitting an analytical model to the pure-component adsorption isotherm data. Of course, if the user chooses a functional form that does not fit the data, the IAST predictions can be inaccurate. For example, Van Heest et al. [198] show that IAST predictions for a xenon/krypton adsorption in metal-organic framework HKUST-1 lead to incorrect predictions when fitting a dual-site Langmuir model to the pure-component isotherm data, but IAST agrees very well with binary grand-canonical Monte Carlo simulations when using numerical quadrature. Ref. [198] calls for meticulousness when fitting analytical models to the pure-component isotherm data for IAST calculations: the user must test different functional forms to identify one of an appropriate shape and scrutinize the quality of its fit to the data. Numerical quadrature takes this burden off of the user since, by construction, linear interpolation will fit the data perfectly. Further, some isotherms exhibit complex shapes, making it difficult to find an appropriate analytical model.

An appropriately shaped analytical model can better-represent the isotherm than linear interpolation if the data is not highly resolved in pressure space. The numerical quadrature illustrated in Fig 6.3 assumes that $n_i^\circ(P)$ is linear between data points; thus, if the grid of pressure points is coarse, this approximation can introduce errors due to inconsideration of the curvature of $n_i^\circ(P)$ between data points. On the other hand, an appropriate analytical model will capture the curvature between data points. In principle, one could capture the curvature between points in numerical quadrature by using e.g. cubic spline interpolation [276] for $n_i^\circ(P)$, but the high variance in fitting cubic splines often leads to non-physical isotherm shapes, and consequently we restrict pyIAST to linear interpolation for $n_i^\circ(P)$.

In the case that the data is noisy, analytical models may be favorable since they can exhibit low variance in the presence of noisy measurements, whereas linear interpolation will incorporate the noise into the IAST calculations (the bias-variance trade-off [201]).

Depending on the pressure and composition of the mixture investigated with IAST, one may need to compute $\pi_i(P)$ for a pressure P outside the range of the available data (issue 2). This can be an issue for both methods of characterizing $n_i^\circ(P)$: fitting an analytical model and linear interpolation.

By definition of interpolation, pyIAST by default throws an exception when $\pi_i(P)$ is needed for P greater than the largest pressure for which data is available. If the pure-component adsorption isotherm is measured/simulated to a high pressure and exhibits a plateau, it may be reasonable to assume that the loading is a constant value beyond the highest pressure in the data, as implemented as an option in pyIAST to both use interpolation and extrapolate the pure-component adsorption isotherm data when needed.

Fitting to an appropriate analytical model is a more natural way to extrapolate the isotherm beyond the data, as e.g. the Langmuir, Quadratic, and BET adsorption isotherms are derived from physical assumptions. However, if the data does not exhibit much curvature, the fitted parameters in the analytical models exhibit high variance; that is, the optimal

parameters become highly sensitive to noise in the data. Intuitively, it is difficult to estimate the saturation loading of an isotherm in the low-coverage regime where the data does not exhibit much curvature.

Thus, emphatically, even when using analytical models, it may be necessary to collect pure-component isotherm data at higher pressures in order to proceed with IAST if data is needed beyond the largest pressure observed in the available data. To induce caution in users extrapolating their data, pyIAST prints a warning when the IAST result required extrapolation of the pure-component adsorption isotherm data.

Finally, we remark that the convenience of using numerical quadrature to compute the spreading pressure $\pi_i(P)$ comes with a greater computational cost than assuming an analytical model. For the analytical models, we derived explicit expressions for $\pi_i(P)$ (e.g. eqn 6.16), whereas for numerical quadrature we carry out the numerical quadrature procedure illustrated in Fig 6.3. Thus, fitting to an analytical model may significantly reduce computation time in applications where IAST calculations are needed repeatedly, such as in packed-bed a(b)dsorber modeling [277]. In the vast majority of cases we see in the literature, the extra computational cost for numerical quadrature is insignificant.

We note the flexibility of pyIAST to mix different types of adsorption isotherm models for its IAST calculations (e.g., use Langmuir for one component and interpolation for another).

6.4 IAST Package Case Study

As an illustration of how to use pyIAST, we consider a gaseous mixture of methane and ethane at 298 K and 65 bar in equilibrium with an adsorbed phase inside a nanoporous material. This mixture is relevant in the application of porous materials to store natural gas onboard vehicles [242]. Natural gas consists of mostly methane ($> 95\%$), but it also contains higher hydrocarbons, nitrogen, and carbon dioxide [242, 243]; the second largest constituent of natural gas is typically ethane. We consider the adsorption of methane/ethane in metal-organic framework IRMOF-1 [278], a nanoporous material with one of the highest measured [148] methane deliverable capacities to date [147]; see Fig 6.1 for a visualization of the IRMOF-1 crystal structure. IAST is applicable as we abstract the walls of the pores in IRMOF-1 as a ravelled-up surface to which gas molecules adsorb.

Using the grand-canonical Monte Carlo algorithm, we simulated pure-component ethane and methane adsorption isotherms at 298 K to use for an IAST calculation. See Fig 6.4. Our goal is to predict the methane and ethane adsorption in IRMOF-1 at different gas phase mole fractions of ethane $y_{C_2H_6}$ in the methane/ethane mixture at 65 bar and 298 K. As both verification of our code and validation of IAST for this mixture in IRMOF-1, we compare the IAST calculations to binary-component grand-canonical Monte Carlo simulations of methane/ethane adsorption at different $y_{C_2H_6}$. The data and pyIAST code to reproduce this case study are available in the `/test` folder of the Github repository.

For our grand-canonical Monte Carlo simulations, we use the TraPPE united atom model for methane and ethane [279]. To model the energetics of the interaction of the adsorbate

pseudoatoms with the atoms of the framework and with other adsorbate pseudoatoms, we use Lennard-Jones potentials; total energy is assumed to be pairwise additive. We take parameters for the framework atoms from the Universal Force Field [108] and parameters for ethane and methane pseudoatoms from the Ref. [279], applying Lorentz-Berthelot mixing rules to obtain cross-interaction parameters. We take van der Waals interactions beyond a cutoff distance of 12.5 Å to be negligible and thus zero, allowing us to implement periodic boundary conditions to mimic a crystal structure of infinite extent. We replicated the unit cell of IRMOF-1 twice in the a, b, and c crystallographic directions to utilize the nearest image convention in applying periodic boundary conditions. The IRMOF-1 crystal structure that we used for our molecular simulations is in the `/test` directory. The IRMOF-1 structure is taken to be rigid.

With the above energetic model, we use the grand-canonical Monte Carlo (GCMC) algorithm [110] to simulate methane/ ethane adsorption. We treat both methane and ethane as an ideal gas by considering pressure and fugacity to be equal. That is, in all plots, the pressure is actually the fugacity. We first superimposed a three-dimensional grid of points on the unit cell of the crystal and computed the potential energy of each adsorbate psuedoatom (CH_3 and CH_4) at each grid point. For our GCMC simulations, when we calculate the adsorbate-adsorbent interaction, we linearly interpolate this grid to speed-up our simulations. We used a grid resolution of 0.06 $\AA$.

For the pure-component adsorption isotherms, our Monte Carlo moves were particle transitions (30%), particle exchange (60%), and particle regrow moves (10%). We define a Monte Carlo *cycle* as N Monte Carlo moves, where N is the number of adsorbate molecules in our system or 20, whichever is greater. We used 200,000 Monte Carlo cycles, half of which were for equilibration/burn cycles. We found that the average number of particles in a unit cell of IRMOF-1 during the simulation does not change significantly when less cycles are used, indicating convergence. For the binary component GCMC simulations, we utilized 50,000 cycles with 20,000 equilibration/burn cycles. The Monte Carlo moves here are particle translation (30%), particle exchange (50%), particle regrow (10%) and identity change (10%).

Constructing pure-component adsorption isotherm models

We load the pure-component adsorption isotherm data displayed in Fig 6.4, stored as two .csv files, into a Pandas DataFrame. A DataFrame is a data structure for conveniently storing and accessing tabular data.

```
import pandas as pd
df_ch4 = pd.read_csv("IRMOF-1_methane_isotherm_298K.csv")
df_ch3ch3 = pd.read_csv("IRMOF-1_ethane_isotherm_298K.csv")
```

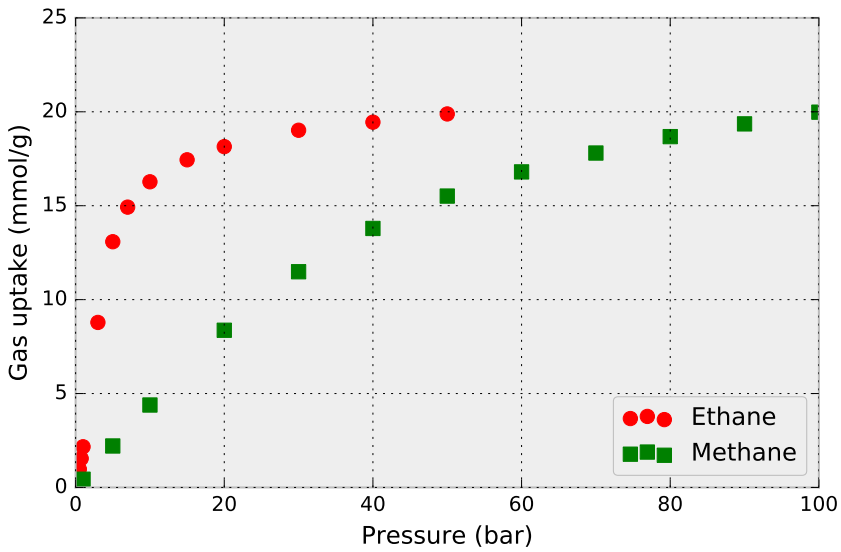


Figure 6.4: Simulated pure-component methane (CH_4) and ethane (C_2H_6) adsorption isotherms in IRMOF-1 at 298 K. Reprinted from Computer Physics Communications, 200, C. Simon, B. Smit, M. Haranczyk, pyIAST: Ideal Adsorbed Solution Theory (IAST) Python package, 364-380, Copyright (2016), with permission from Elsevier.

Next, we use pyIAST to construct an adsorption isotherm object for each pure-component adsorption isotherm. We start with an attempt to fit a Langmuir adsorption isotherm model to characterize $n^{\circ}_{C_2H_6}(P)$. The `ModelIsotherm` class in pyIAST is used for all analytical models; we notify pyIAST which model to fit to the data by passing e.g. `model="Langmuir"` for a model in the list `pyiast._MODELS`. We construct a `ModelIsotherm` object by passing it the Pandas DataFrame containing the pure-component ethane isotherm, `df_ch3ch3`, and the names of the columns (keys) corresponding to the loading and pressure.

```
import pyiast
ch3ch3_isotherm = pyiast.ModelIsotherm(df_ch3ch3,
                     loading_key="Loading(mmol/g)",
                     pressure_key="Pressure(bar)",
                     model="Langmuir")

# Print identified parameters
ch3ch3_isotherm.print_params()
# > Langmuir identified model parameters:
# >   K = 0.229837
# >   M = 21.932328
# >   RMSE =  0.862745372792
```

```
# get dictionary of parameters
ch3ch3_isotherm.params
# > {'K': 0.22983701506470389, 'M': 21.932328016544513}

# predicted loading at 10 bar using fitted model (mmol/g)
ch3ch3_isotherm.loading(10.0)
# > 15.28288389268168
```

In the construction of the `ModelIsotherm` object, a least-squares nonlinear fitting routine finds the parameters of best fit. The above code and resulting output illustrates how to print the identified parameters and root mean square error (RMSE), access a dictionary containing the parameters of best fit, and predict loading using the fitted model from an instance of the `ModelIsotherm` class. We can visualize the fit of the Langmuir model to the adsorption isotherm data with the `plot_isotherm()` function:

```
pyiast.plot_isotherm(ch3ch3_isotherm, xlogscale=True)
```

Fig 6.5 shows that the Langmuir model provides a poor fit to the pure-component ethane adsorption isotherm data and thus will lead to inaccurate IAST calculations. As a better approximation of $n^{\circ}_{C_2H_6}(P)$, we turn to the method of linearly interpolating the ethane adsorption isotherm data in Fig 6.4 since we have data on a sufficiently resolved grid of pressures.

The `InterpolatorIsotherm` object linearly interpolates the pure-component adsorption isotherm data and uses numerical quadrature to compute the spreading pressure (see Sec 6.3). Similar to the `ModelIsotherm` class, we construct an instance of the `InterpolatorIsotherm` class by passing `df_ch3ch3` and the names of the pressure and loading columns. By default, the `InterpolatorIsotherm` will throw an exception when we attempt to extrapolate the pure-component adsorption isotherm data beyond the highest pressure available in Fig 6.4. As we have ethane adsorption data at high enough pressures to see saturation in the isotherm, a reasonable assumption is that the saturation loading is equal to the highest uptake in the data. We enforce this assumption by passing a `fill_value` in the construction of the `InterpolatorIsotherm` object, setting it equal to the maximum loading of ethane observed in Fig 6.4.

```
ch3ch3_isotherm = pyiast.InterpolatorIsotherm(df_ch3ch3,
                    loading_key="Loading(mmol/g)",
                    pressure_key="Pressure(bar)",
                    fill_value=df_ch3ch3["Loading(mmol/g)"].max())
```

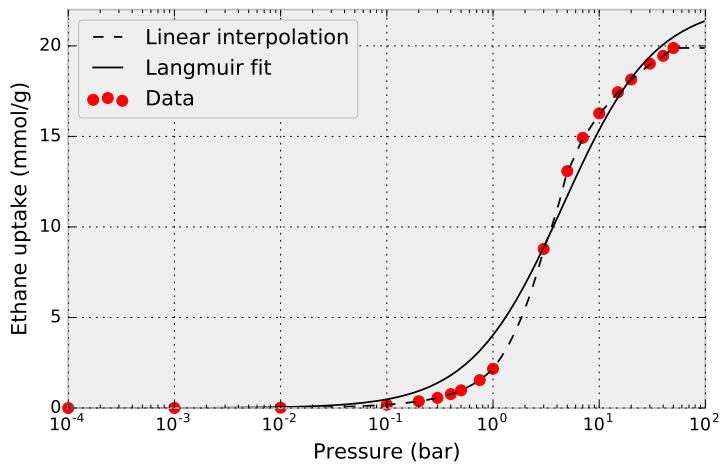


Figure 6.5: Characterizing $n^{\circ}_{C_2H_6}(P)$ from pure-component adsorption isotherm data (red points). Shown is the Langmuir model fit by `ModelIsotherm` (solid line) and the interpolation model fit by `InterpolatorIsotherm` (dashed line). Fits can be visualized by the function `plot_isotherm()`. Reprinted from Computer Physics Communications, 200, C. Simon, B. Smit, M. Haranczyk, pyIAST: Ideal Adsorbed Solution Theory (IAST) Python package, 364-380, Copyright (2016), with permission from Elsevier.

The fit of this `ch3ch3_isotherm` is shown in Fig 6.5.

Analogously, we also construct an `InterpolatorIsotherm` object for methane, `ch4_isotherm`. These adsorption isotherm objects will be passed as inputs to the `iast()` function to perform IAST calculations.

Performing IAST calculations

Given the characterization of the pure-component adsorption isotherms $n^{\circ}_{C_2H_6}(P)$ and $n^{\circ}_{CH_4}(P)$ using `InterpolatorIsotherm` objects, we can now predict binary methane/ethane adsorption isotherms in IRMOF-1 using IAST.

As an example, consider predicting the moles of adsorbed methane and ethane, n_{CH_4} and $n_{C_2H_6}$, in IRMOF-1 in equilibrium with a 95/5 mol % methane/ethane mixture at 298 K and 65 bar. Our function to perform IAST calculations, `iast()`, takes a list or array of partial pressures of the gases and a list of the corresponding instances of `ModelIsotherm`'s and/or `InterpolatorIsotherms`'s. The `verboseflag` prints details of the internal computations, such as p°_i and the spreading pressure of the mixture.

```
# partial pressures of ethane and methane in gas phase (bar)
partial_pressures = [0.05 * 65.0, 0.95 * 65.0]
component_loadings = pyiast.iast(partial_pressures,
```

```
                                [ch3ch3_isotherm, ch4_isotherm],
                                verboseflag=False)
# > array([  4.4612935 ,  13.86364776])
```

The `iast()` function returns an array `component_loadings` of component uptakes in the adsorbed phase, $n_{C_2H_6}$ and n_{CH_4}, computed from IAST. Here, $n_T = n_{CH_4} + n_{C_2H_6}$ and $x_{CH_4,C_2H_6} = n_{CH_4,C_2H_6}/n_T$. pyIAST did not print a warning that it needed to extrapolate the pure-component isotherm data for this IAST calculation, indicating that passing the `fill_value` in the `InterpolatorIsotherm` for ethane was not necessary. That is, we collected pure-component ethane adsorption isotherm data at a sufficiently high pressure for this IAST calculation.

Using the above procedure in a loop, we computed the mixed-gas adsorption isotherms $n_{CH_4}(y_{C_2H_6}, P = 65 \text{ bar})$ and $n_{C_2H_6}(y_{C_2H_6}, P = 65 \text{ bar})$ at varying gas phase compositions $y_{C_2H_6}$. Fig 6.6 shows that the component uptakes predicted by pyIAST match the binary-component grand-canonical Monte Carlo simulations of methane/ethane adsorption in IRMOF-1. This agreement serves as both partial verification of pyIAST and validation of IAST for a methane/ethane mixture in IRMOF-1 at these conditions.

Beyond binary mixtures

While our case study focused on a binary methane/ethane mixture, pyIAST can handle > 2 components. Here, we compare IAST predictions to an experimentally measured tertiary-component gas adsorption isotherm. Mason and coworkers [245] measured pure-component water, carbon dioxide, and nitrogen adsorption isotherms in activated carbon AX-21 at 40 °C. Then, they measured the adsorption of nitrogen and carbon dioxide in the presence of a mixture of 166 mbar CO_2, 679 mbar N_2, and 20 mbar H_2O at 40 °C – a composition relevant to capturing CO_2 from the flue gas of coal-fired power plants. Here, we illustrate how pyIAST is used to predict the result.

Fig 6.7a and 6.7b (points) show the experimentally measured pure-component H_2O, CO_2, and N_2 adsorption isotherms in AX-21 at 40 °C. Using pyIAST, we fit Henry and Langmuir models to the N_2 and CO_2 isotherms, respectively, and use linear interpolation for the H_2O isotherm since it exhibits a relatively complex shape. The lines in Figs 6.7a and 6.7b show the pyIAST fits to the data from the code below.

```
import pandas as pd
import pyiast

df_N2 = pd.read_csv("N2.csv")
N2_isotherm = pyiast.ModelIsotherm(df_N2,
                    loading_key="Loading(mmol/g)",
                    pressure_key="P(bar)",
```

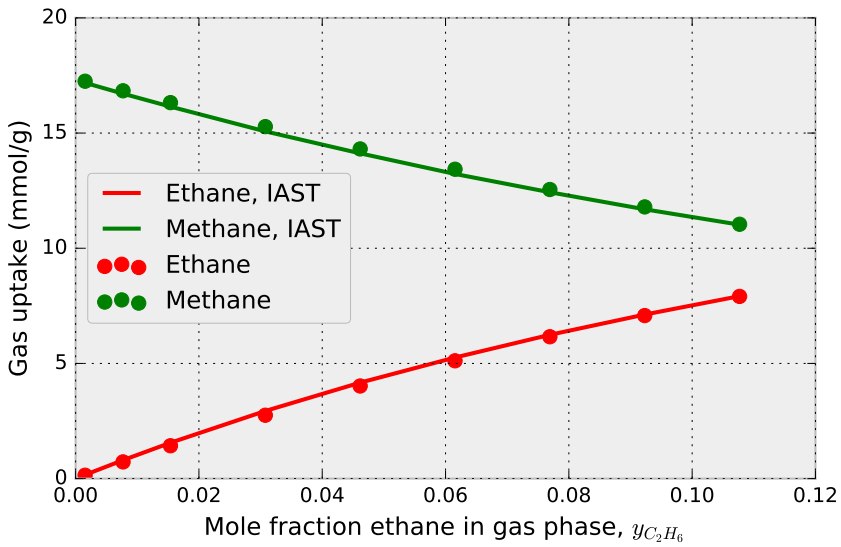


Figure 6.6: Verification and validation of IAST for IRMOF-1 in equilibrium with a methane/ethane mixture at 298 K and 65 bar total pressure. Shown are mixed-gas adsorption isotherms $n_{CH_4}(y_{C_2H_6}, P = 65 \text{ bar})$ (green) and $n_{C_2H_6}(y_{C_2H_6}, P = 65 \text{ bar})$ (red) as a function of the mole fraction of ethane in the gas phase, $y_{C_2H_6}$, computed with IAST (lines) and binary-component grand-canonical Monte Carlo simulations of methane/ethane adsorption (points). Reprinted from Computer Physics Communications, 200, C. Simon, B. Smit, M. Haranczyk, pyIAST: Ideal Adsorbed Solution Theory (IAST) Python package, 364-380, Copyright (2016), with permission from Elsevier.

```
                    model="Henry")

df_CO2 = pd.read_csv("CO2.csv")
CO2_isotherm = pyiast.ModelIsotherm(df_CO2,
                     loading_key="Loading(mmol/g)",
                     pessure_key="P(bar)",
                     model="Langmuir")

df_H2O = pd.read_csv("H2O.csv")
H2O_isotherm = pyiast.InterpolatorIsotherm(df_H2O,
                     loading_key="Loading(mmol/g)",
                     pressure_key="P(bar)",
                     fill_value=df_H2O["Loading(mmol/g)"].max())
```

We calculate the component loadings in AX-21 in the presence of the mixture with pyIAST by passing a list of partial pressures and pure-component isotherms to the `iast()`

function. This returns an array `component_loadings` of component loadings at these gas conditions, predicted by IAST.

```
# list of CO2, N2, and H2O partial pressures (bar)
partial_pressures = [0.166, 0.679, 0.02]
component_loadings = pyiast.iast(p,
          [CO2_isotherm, N2_isotherm, H2O_isotherm])
# > array([ 0.51451255,  0.30433924,  0.61999649])
```

The solid bars in Fig. 6.7c show the experimentally measured CO_2 and N_2 uptakes in the presence of the mixture with their error bars (water uptake not reported). The decorated bars show the corresponding predictions according to pyIAST. Mason et al. [245] used IAST for a *binary* 166 mbar/679 mbar CO_2/N_2 mixture, ignoring the presence of water, and found that the predicted CO_2 uptake matches the experimental measurement very well; pyIAST's binary mixture calculations are consistent with theirs. Interestingly, IAST predicts that the addition of water– considering a 166 mbar/679 mbar/20 mbar $CO_2/N_2/H_2O$ mixture– *enhances* CO_2 and N_2 uptake from the binary case when water was ignored. As a result, the CO_2 uptake in the tertiary mixture according to IAST is greater than experiment. The reason is that IAST imposes that the free energy of mixing is given by the entropy of mixing; this implies that the energetics of the adsorbate-adsorbate interactions are the same between all species in the mixture. The water isotherm in Fig 6.7a is concave up, suggesting strong water-water interactions that yield cooperative adsorption. The ideal solution assumption in IAST imposes that adsorbed water molecules will similarly recruit more CO_2 and N_2, explaining why IAST predicts that water enhances the uptake of CO_2 and N_2 beyond the binary case considered in Mason et al. [245].

Reverse IAST calculations

In reverse IAST, we specify the desired mole fractions in the adsorbed phase, $\{x_i\}$, and the total bulk gas pressure P_T and then calculate the bulk gas composition $\{y_i\}$ that yields these adsorbed mole fractions. This is useful for when one seeks to control the composition of the adsorbed phase, such as in catalysis [262]. As an example, we seek the bulk gas composition – at the same temperature as the pure-component isotherms, 298 K – that will yield a 50/50 mol % methane/ethane mixture in the adsorbed phase at a total bulk gas pressure of 65.0 bar. The following code will return two arrays `gas_mole_fractions` and `component_loadings`, the required mole fractions in the gas phase and the resulting component uptakes in the adsorbed phase, respectively.

```
total_pressure = 65.0  # total bulk gas pressure (bar)
adsorbed_mole_fractions = [0.5, 0.5]  # desired adsorbed mole fractions
```

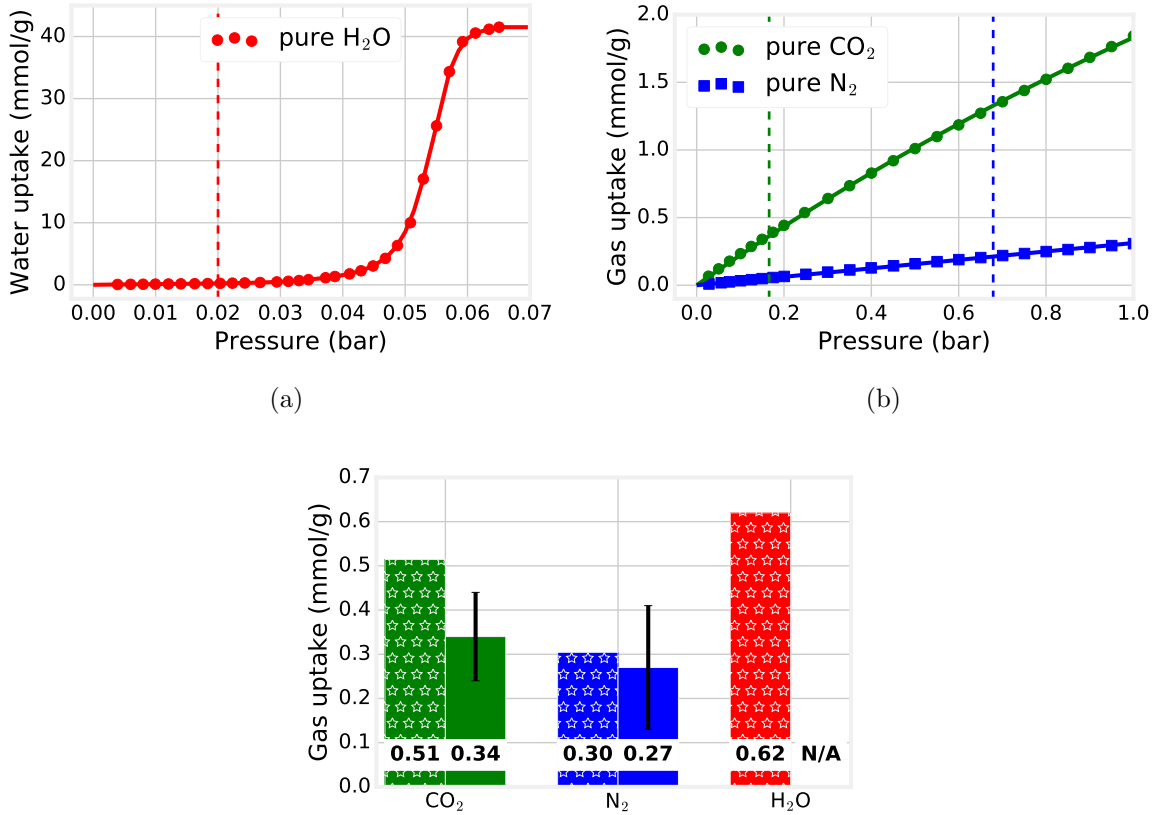


Figure 6.7: Applying IAST to a 3-component mixture of 166 mbar CO_2, 679 mbar N_2, and 20 mbar H_2O at 40 °C in equilibrium with an adsorbed phase in activated carbon AX-21 from Mason et al. [245]. (a, b) Pure-component H_2O (panel a), CO_2 (panel b, green), and N_2 (panel b, blue) isotherms in AX-21 at 40 °C from Ref. [245]. Points are experiment; lines are pyIAST fitted model predictions. Vertical lines show partial pressure of each component in the mixture. (c) Bar chart shows uptakes of CO_2 (green), N_2 (blue), and H_2O (red) in AX-21 in the presence of the mixture predicted by IAST from the pure-component isotherms (decorated) and measured by experiment (solid bars) [245]. Error bars are shown for the two experimental measurements; water uptake was not measured. Reprinted from Computer Physics Communications, 200, C. Simon, B. Smit, M. Haranczyk, pyIAST: Ideal Adsorbed Solution Theory (IAST) Python package, 364-380, Copyright (2016), with permission from Elsevier.

```
gas_mole_fractions, component_loadings = pyiast.reverse_iast(
                                    adsorbed_mole_fractions,
                                    total_pressure,
```

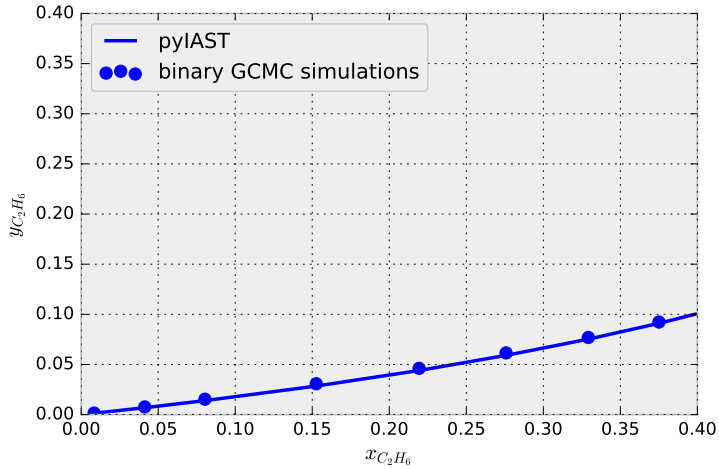


Figure 6.8: Verification and validation of reverse IAST for IRMOF-1 in equilibrium with a methane/ethane mixture at 298 K and 65 bar. In reverse IAST, we specify the desired mole fraction of ethane in the adsorbed phase, $x_{C_2H_6}$, and calculate the gas phase mole fraction, $y_{C_2H_6}$, required to yield this desired adsorbed phase composition. The required $y_{C_2H_6}$ to yield desired $x_{C_2H_6}$, computed with IAST (lines) and found in binary-component grand-canonical Monte Carlo simulations of methane/ethane adsorption (points) are shown. Reprinted from Computer Physics Communications, 200, C. Simon, B. Smit, M. Haranczyk, pyIAST: Ideal Adsorbed Solution Theory (IAST) Python package, 364-380, Copyright (2016), with permission from Elsevier.

```
                                    [ch3ch3_isotherm, ch4_isotherm])
```

As both verification of our code and validation of reverse IAST for this methane/ethane mixture in IRMOF-1, we compare the reverse IAST calculations to binary grand-canonical Monte Carlo simulations of methane/ethane adsorption that yielded different adsorbed phase compositions, $x_{C_2H_6}$ (the same simulations in Fig 6.6). Fig 6.8 shows excellent agreement between reverse IAST calculations and our binary grand-canonical Monte Carlo simulations.

A remark on units

The units of pressure and gas adsorption in each pure-component adsorption isotherm DataFrame should be consistent. The pressures passed to pyIAST functions must be in the same units as the pure-component adsorption isotherm DataFrame corresponding to the `pressure_key` column. The mixed-gas adsorption outputs of pyIAST have the same units as the adsorption in the pure-component adsorption isotherm DataFrame corresponding to the `loading_key` column. Emphatically, gas adsorption must be in a molar unit (e.g. mol/g

or mmol/m^3), and not mass (e.g. not mg/m^3), for the thermodynamic equations of IAST to hold.

Tests for further verification of pyIAST

While the agreement between pyIAST and binary grand-canonical Monte Carlo simulations in Fig 6.6 serves as validation of IAST for this ethane/methane mixture in IRMOF-1 and a partial verification for pyIAST, we outline two more comprehensive test suites here.

Isotherm fitting

To ensure that pyIAST correctly fits analytical models to pure-component isotherm data in the `ModelIsotherm` class, we generated synthetic data following each isotherm model available in pyIAST (a list can be found in `pyiast._MODELS`) for a given set of parameters. We then asserted that pyIAST correctly identified these parameters from the synthetic data. This test is available in the IPython Notebook `test/Isotherm tests.ipynb`.

Comparing to the competitive Langmuir isotherm

In the case that the pure-component adsorption isotherms follow a Langmuir model with the same saturation loading, IAST yields an analytical solution for the mixed-gas adsorption isotherm, the competitive Langmuir model in eqn 6.26. We verify that pyIAST solves the equations of IAST correctly by (1) generating synthetic pure-component isotherm data for fictitious gases A, B, and C that follow a Langmuir isotherm with the same saturation loading M but different Langmuir constants, K_i (see Fig 6.9a), (2) fitting Langmuir models to this synthetic data using pyIAST's `ModelIsotherm`, (3) using pyIAST to predict the mixed-gas adsorption isotherm of A, B, and C at 1 bar total pressure, thoroughly exploring ternary composition space, and (4) asserting that the loadings of A, B, and C predicted by pyIAST match the competitive Langmuir model in eqn 6.26. Fig 6.9 shows the ternary plots of the component loadings of A, B, and C according to the competitive Langmuir model (second from bottom row) and according to pyIAST (bottom row). Across all of ternary composition space, pyIAST matches the competitive Langmuir model. This test is available in the IPython Notebook `test/Test IAST for Langmuir case.ipynb`.

6.5 Limitations of IAST

Widely used [249], ideal adsorbed solution theory (IAST) has shown remarkable success in readily predicting mixed-gas adsorption isotherms from pure-component isotherms with a reasonable accuracy in a variety of systems [249]. Here, we duly remark that mixed-gas isotherms exhibit deviations from IAST when its assumptions are invalid [250, 281, 282, 283, 251, 284, 258]. IAST rests upon the assumptions that:

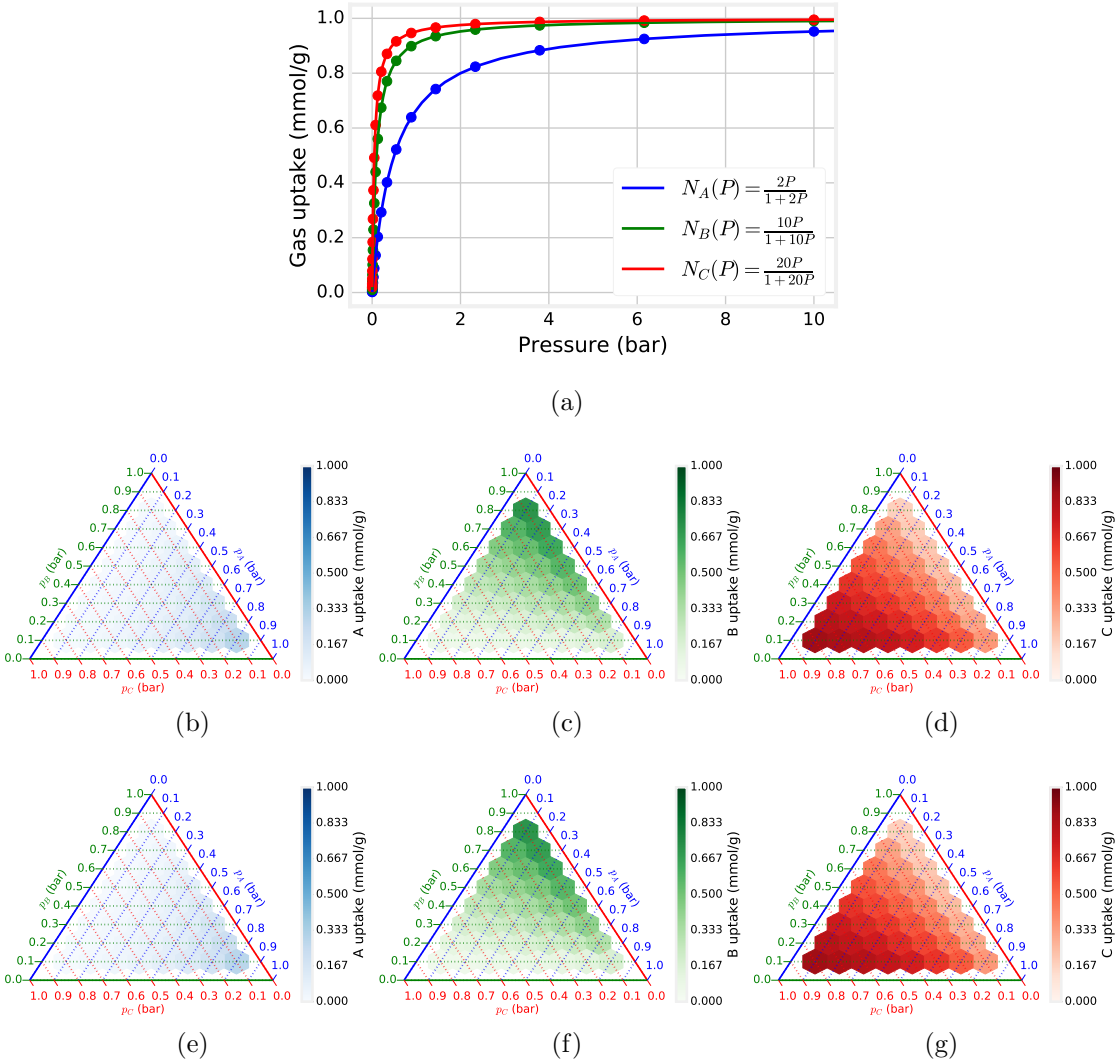


Figure 6.9: Testing pyIAST for a tertiary mixture of fictitious gases A, B, and C at a total pressure of 1 bar. (a) We generated synthetic pure-component A, B, and C isotherm data that follow Langmuir models with saturation loading $M = 1$ and Langmuir constants $K_A = 2$, $K_B = 10$, and $K_C = 20$. (b-g) A point inside the triangle of this ternary plot represents a unique composition of the A/B/C mixture at 1 bar. The triangle is partitioned into hexagons. The color of each hexagon represents the uptake of species i at the center of the hexagon according to the competitive Langmuir isotherm in eqn 6.26 (second from bottom rom) and pyIAST (bottom row). For each row, the three different triangles and colormaps represent the uptakes of species A, B, and C. The colors correspond to (a). Plots created by Ref. [280]. Reprinted from Computer Physics Communications, 200, C. Simon, B. Smit, M. Haranczyk, pyIAST: Ideal Adsorbed Solution Theory (IAST) Python package, 364-380, Copyright (2016), with permission from Elsevier.

.

(i) the adsorbates form an ideal mixture. Deviations from ideal mixing can occur due to appreciable energetic heterogeneity in the adsorbent surface [285, 286, 287] and size and asphericity differences between adsorbates [288, 250]. One can account for non-ideality of adsorbates with an activity coefficient if mixture data is available [248] (real adsorbed solution theory [281, 289]). Energetic heterogeneity can be accounted for using heterogeneous IAST (HIAST [286]), which assumes IAST holds locally at each adsorption site; the overall adsorption isotherm can be integrated over all sites using a distribution of energies [290]. Segregated IAST (SIAST) is another framework where the pore space of a material is partitioned into regions where IAST is applied separately [291].

(ii) adsorbates have access to the same area of surface. The latter assumption cannot hold, for example, in a molecular sieve, where the area accessible to each adsorbent is dependent upon the size of the adsorbate [248].

(iii) changes in the thermodynamic properties of the adsorbent are negligible relative to the same properties for the adsorbate. This assumption is negated in materials that exhibit structural transitions upon gas adsorption [245, 174]. Osmotic Framework Adsorbed Solution Theory (OFAST) [292] has been developed for predicting mixed-gas adsorption isotherms from pure-component adsorption isotherms in frameworks exhibiting structural transitions upon adsorption.

These corrections to IAST must often be implemented on a case-by-case basis and are difficult to include in a transferable code such as pyIAST.

6.6 Conclusions

While measuring pure-component adsorption isotherms is relatively routine, measuring mixed-gas adsorption isotherms is complicated and time-consuming and requires custom instruments [247, 245]. Further, often one is interested in many different gas compositions and pressures; taking mixed-gas adsorption measurements at such an array of conditions is often impractical. Ideal adsorbed solution theory (IAST) [248] is a widely used [249] thermodynamic framework to predict mixed-gas adsorption isotherms from a set of pure-component adsorption isotherms at the same temperature.

We first reviewed the equations of IAST derived from thermodynamics and discussed the two disparate methods of characterizing the pure-component adsorption isotherms: fitting an analytical model and linearly interpolating the data. We then provided a tutorial of how to use our Python package pyIAST to perform IAST calculations using a case study of methane/ethane adsorption in a nanoporous material. pyIAST can handle an arbitrary number of components and is also capable of reverse IAST calculations, where we calculate the gas phase composition to yield a desired adsorbed phase composition.

The source code for pyIAST, registered as an official Python package with the Python Package Index (PyPI) under an MIT license, is openly available on Github `https://github.com/CorySimon/pyIAST`. More extensive online documentation is hosted on Read the Docs,

`http://pyiast.readthedocs.org/en/latest/`. Using the pip Python package manager, pyIAST can be easily installed in a single command in the terminal, `pip install pyiast`.

Chapter 7

Computation-assisted discovery of highly Xe-selective metal-organic framework

Nuclear energy is among the most viable alternatives to our current fossil fuel-based energy economy. The mass deployment of nuclear energy as a low-emissions source requires the reprocessing of used nuclear fuel to recover fissile materials and mitigate radioactive waste. A major concern with reprocessing used nuclear fuel is the release of volatile radionuclides such as xenon and krypton that evolve into reprocessing facility off-gas in parts per million concentrations. The existing technology to remove these radioactive noble gases is a costly cryogenic distillation; alternatively, porous materials such as metalorganic frameworks have demonstrated the ability to selectively adsorb xenon and krypton at ambient conditions. Here we carry out a high-throughput computational screening of large databases of metalorganic frameworks and identify SBMOF-1 as the most selective for xenon. We affirm this prediction and report that SBMOF-1 exhibits by far the highest reported xenon adsorption capacity and a remarkable Xe/Kr selectivity under conditions pertinent to nuclear fuel reprocessing.

This chapter is based on the following publication [100]:

D. Banerjee, C. Simon, A. Plonka, R. Motkuri, J. Liu, X. Chen, B. Smit, J. Parise, M. Haranczyk, P. Thallapally. Metal-organic framework with optimally selective xenon adsorption and separation. *Nature Communications.* (2016) 7, ncomms11831

7.1 Introduction

An important advantage of MOFs is their chemical tunability; by combining different linkers and metal centres that self-assemble to form ordered, pre-determined crystal structures, one can synthesize millions of possible materials [22]. MOFs can thus be tailor-made to

be optimal for applications related to gas storage and separation, chemical sensing [14], drug delivery [16], and catalysis [15]. Our goal here is to identify an optimal MOF for selectively capturing Xe from the off-gas of UNF reprocessing facilities (see Sec. 1.1). In practice, however, constraints in resources allow us to synthesize and test only a small subset of chemical space. Molecular models and simulations of adsorption can rapidly and cost-effectively rank MOFs by their Xe/Kr selectivity with reasonable accuracy. See Fig 7.1.

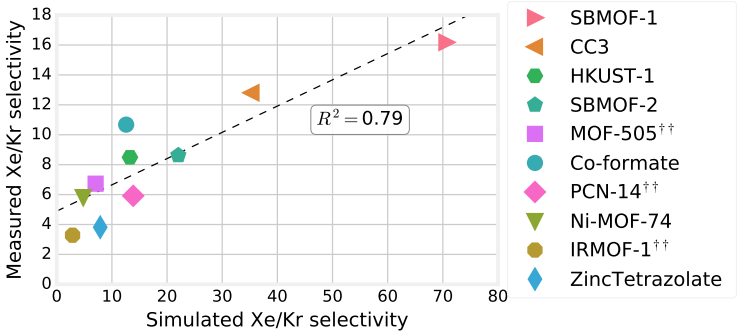


Figure 7.1: Comparison between simulated and experimental Xe/Kr selectivities at dilute conditions. Experimental selectivities calculated from Henry coefficients extracted from pure-component adsorption isotherms. Simulated selectivities calculated from same methods used in the high-throughput screening in this chapter. A correlation of $R^2 = 0.79$ is observed, indicating that there is statistical signal in the ranking of materials that is produced from a high-throughput screening of materials for Xe/Kr separations using the methods in this chapter. Data corresponds to that in Fig 7.5. Reproduced from Ref. [100] under the Creative Commons License Attribution 4.0 International License.

High-throughput computational screenings thus play a valuable role of elucidating design rules, determining performance limits, and predicting performance rankings of materials to focus experimental efforts on the most promising MOFs for Xe/Kr separations [197, 198, 76, 95].

In this work, we use molecular simulations to screen over 125,000 MOF structures [160, 121] for selectively adsorbing Xe over Kr at dilute conditions pertinent to UNF reprocessing. Our computational screening predicts that one of the most Xe-selective MOFs is a calcium-based nanoporous MOF, SBMOF-1 [also known as CaSDB, SDB=4,4-sulfonyldibenzoate], that has not yet been tested for Xe/Kr separations [234]. We affirm this prediction by synthesizing SBMOF-1 and measuring its pure-component Xe and Kr adsorption isotherms. SBMOF-1 exhibits the highest Xe Henry coefficient and thermodynamic Xe/Kr selectivity at dilute conditions among MOFs tested to date. In addition to its high thermal and chemical stability, column breakthrough experiments reveal that SBMOF-1 is a practical, near-term material for capturing Xe from reprocessing facilities.

7.2 High-throughput screening

For capturing Xe from nuclear reprocessing, the Xe/Kr selectivity is the most important thermodynamic property determining the performance of a MOF. We used molecular simulations to predict the Xe/Kr selectivity of 125,000 MOF structures at dilute conditions relevant to UNF reprocessing. We model the energetics of the van der Waals interactions of a Xe and Kr atom with the atoms of the MOF structures as pair-wise additive and with Lennard-Jones potentials. We take Xe-Xe and Kr-Kr Lennard-Jones parameters from Boato and Casanova [94]. We took Lennard-Jones parameters for the atoms in the MOF structure from the Universal Force Field (UFF) [108]. Lennard-Jones parameters for the host-guest interactions then follow from Lorentz-Berthelot mixing rules. The MOF structures are held rigid throughout the simulations. We take interactions beyond a cutoff radius of 12.5 Å to be zero and implement periodic boundary conditions to mimic an infinite crystal. This molecular model is equivalent to that used in the high-throughput screening in Ref. [95]. As Xe and Kr are present at dilute concentrations in the off-gas of used nuclear fuel reprocessing facilities, the adsorption isotherm of Xe and Kr can be modeled with Henry's law as:

$$N_i(p_i) = K_{H,i} p_i \; i = \text{ Xe, Kr,} \tag{7.1}$$

where N_i is the amount of species i adsorbed [moles/mass of material], p_i is the partial pressure of species i in the gas phase [bar], and $K_{H,i}$ is the Henry coefficient of species i in the material [moles/(mass of material - bar)]. The selectivity of Xe over Kr at dilute conditions is then the ratio of the Xe and Kr Henry coefficients.

The Henry coefficient is related to the spatial average of the Boltzmann factor inside the unit cell:

$$K_H = \frac{\langle e^{-U/(RT)} \rangle}{RT}, \tag{7.2}$$

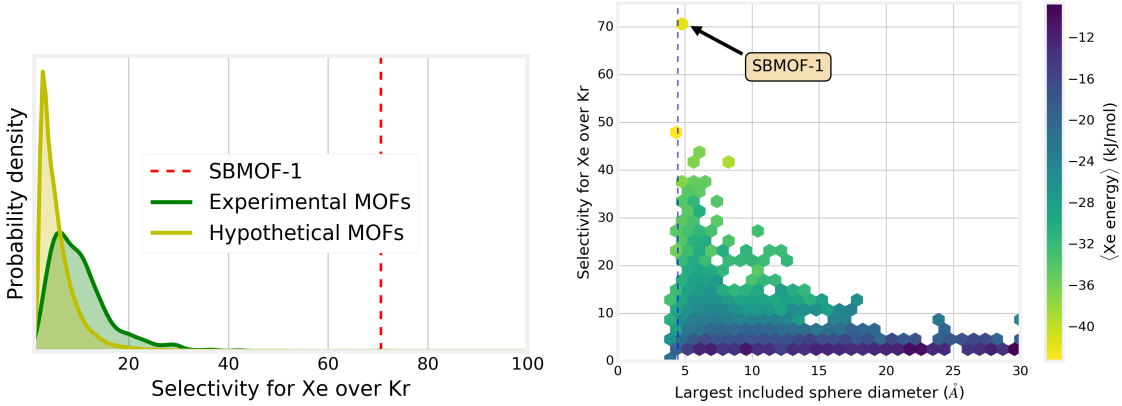
where R is the universal gas constant, T is the temperature (298 K), and U is the energy of the adsorbate in the structure. See Sec. 1.2 for a derivation. We use a Monte Carlo integration algorithm (Widom particle insertions [110], 750 samples per $Å^3$ of material) to compute the above average in each material and predict the Henry coefficient, using the model for the energy U described above. As each Widom particle insertion is independent of the others, this algorithm is highly amenable to parallelization [111]; we utilized a CUDA code for GPUs to speed-up the calculation of the Henry coefficient, which enabled us to simulate Xe/Kr adsorption in large databases of materials in a reasonable time.

We computed the Henry coefficient of Xe and Kr in the following databases of MOF structures: (i) Computation-ready, experimental (CoRE) MOFs [160], a database of over 4,700 computation-ready porous MOFs that have been experimentally synthesized and reported in the Cambridge Structural Database.24 (ii) hMOFs, a database of 120,000 hypothetical/predicted MOFs generated computationally by snapping together molecular building blocks [121], much like Lego blocks or KNEX.

We visualized the distribution of selectivity among the MOFs using kernel density estimation with a Gaussian kernel (bandwidth chosen by Scotts rule [239]) in the Stats models

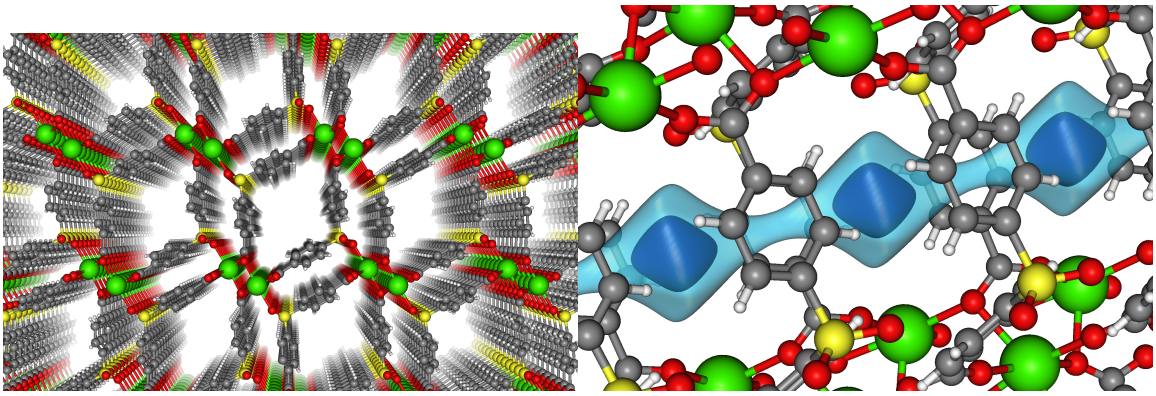
Python module. For the CoRE MOFs, we omitted structures that were identified to have pores that are inaccessible to Xe using a pocket-blocking algorithm. 2,867 CoRE MOFs remained.

The distribution of simulated selectivities in the MOFs is shown in Fig. 7.2a. We partitioned this distribution into a database of existing MOFs (∼5,000 structures) [160] and a database of predicted/hypothetical structures (∼120,000) [121]. These distributions span a large range of selectivities, illustrating the unique tunability of MOF materials. Our simulations predict that the most selective material in the database of existing MOFs is SBMOF-1 (Fig. 7.3a), a three-dimensional, permanently porous MOF (Cambridge Structural Database (CSD) code: KAXQIL) [234]. Furthermore, the Xe/Kr selectivity of SBMOF-1 is ranked in the top 0.01 percentile in the database of 120,000 hypothetical MOFs (Fig. 7.2b). The red line in Fig. 7.2b illustrates the outlying Xe/Kr selectivity of SBMOF-1 predicted by our screening. While SBMOF-1 has been synthesized and considered for CO_2/N_2 separation [234], it has not been tested for Xe/Kr separations.



(a) (b)

Figure 7.2: Computational screening of MOFs for Xe/Kr separations at dilute conditions relevant to UNF reprocessing off-gas. We computed the Henry coefficients of Xe and Kr in ∼125,000 MOF structures; the selectivity at dilute conditions is the ratio of Henry coefficients. (a) Distribution of simulated selectivities for experimentally synthesized (green) and hypothetical (yellow) MOF structures; vertical, dashed line is SBMOF-1 (KAXQIL in the Cambridge Structural Database (CSD)). (b) Histogram showing relationship between selectivity and pore size, with the largest included sphere diameter as a metric; colour shows average energy of Xe adsorption in that bin. SBMOF-1 (KAXQIL in CSD), with simulated selectivity 70.6 and largest included sphere diameter of 5.1 Å, is indicated. Vertical, dashed line is the distance that yields the minimum energy in a Xe-Xe LennardJones potential. Reproduced from Ref. [100] under the Creative Commons License Attribution 4.0 International License.



(a) (b)

Figure 7.3: Crystal structure of SBMOF-1. (c) SBMOF-1 is composed of corner sharing, octahedrally coordinated calcium chains along the crystallographic b direction, which are connected by organic linkers, forming a one-dimensional nanoporous channel. (d) Side view. Shown are the calculated potential energy contours of a Xe atom adsorbed in the pore (blue surface, 32 kJ mol^{-1}; white surface, 15 kJ mol^{-1}).

7.3 Synthesis and equilibrium adsorption measurements

Encouraged by the data from our high-throughput screening, our experimental collaborators at Pacific Northwest National Lab synthesized SBMOF-1 and measured its pure-component Xe and Kr adsorption isotherms at room temperature [234, 100].

SBMOF-1 was originally synthesized using a previously published literature procedure [234]. In a typical synthesis, a mixture of 0.6mmol of $CaCl_2$ (0.074g) and 0.6mmol of 4,4'-SDB (0.198g) were added in 10ml of ethanol and stirred for ∼2h to achieve homogeneity (molar ratio of metal chloride:ligand:solvent=1:1:380). The resultant solution was heated at 180 °C for 3 days. Colorless, needle-shaped crystals were recovered as product and washed with ethanol (yield: 45% based on $CaCl_2$, 0.1g). For scale up, 1.44g of $CaCl_2$ (13mmol) and 3.98g of 4,4'-SDB (13mmol) were added to 120ml of ethanol and stirred for ∼2h to achieve homogeneity (molar ratio: 1:1:156). The well-mixed solution was then transferred to three 100ml Teflon-lined stainless steel Parr autoclaves and heated for 3 days at 180°C. The product was obtained as white powder and washed by ethanol (3 times, 50ml), followed by drying under vacuum (yield: 2.2g, 50% based on $CaCl_2$). The as-synthesized material was then exchanged with methanol (3 times, 50ml) for a total period of 3 days. The product purity was confirmed by powder XRD.

The methanol-exchanged SBMOF-1 was activated at 100°C for 12h under dynamic vacuum. Single-component gas-adsorption isotherms were collected in a Quantachrome Autosorb-1 and dynamic sorption analyzer (ARBC, Hiden Analytical Ltd., Warrington,

UK). The later instrument was also used to collect breakthrough measurement data.

The Xe adsorption isotherm in SBMOF-1 saturates at a low pressure, indicative of a high affinity for Xe (Fig. 7.4). The Kr adsorption isotherm exhibits a smaller slope and does not saturate even at 1 bar, indicative of a much weaker affinity for Kr. This hints that SBMOF-1 is highly discriminatory for Xe over Kr. Indeed, identifying the Xe and Kr Henry coefficients from the pure-component adsorption isotherms, we predict SBMOF-1 to exhibit a thermodynamic Xe/Kr selectivity of 16 at dilute conditions at 298 K.

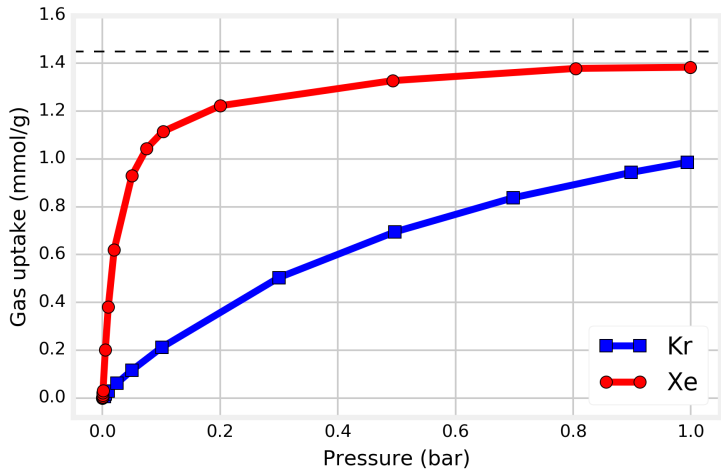


Figure 7.4: Experimental Xe and Kr adsorption isotherms in SBMOF-1 at 298 K. Horizontal line indicates one atom per pore segment. Reproduced from Ref. [100] under the Creative Commons License Attribution 4.0 International License.

It is interesting to compare the equilibrium Xe and Kr uptake of SBMOF-1 with the reported top-performing MOFs. We collected from the literature experimentally measured Xe and Kr adsorption isotherms in CC3 [77], Co-formate [90, 85], noria [101], SBMOF-2 [102], HKUST-1 [103], MOF-505 [92], PCN-14 [104], Ni-MOF-74 [86], Zinc tetrazolate [105], IRMOF-1 [83], and FMOF-Cu [87]. Focusing on the low-pressure regime of the adsorption isotherm that exhibits linear behaviour– the Henry regime where Henrys law in equation 7.1 is valid– we fit a line with zero intercept to this data to identify $K_{H,Xe}$ and $K_{H,Kr}$ in the material. The Henry coefficient of Xe in SBMOF-1 is a factor of two higher than in CC3 [77], the material in our survey with the second highest Xe Henry coefficient; we thus expect SBMOF-1 to have an outstanding Xe uptake under UNF reprocessing off-gas conditions. Figure 7.5 shows that SBMOF-1 exhibits by far the largest Xe Henry coefficient and the highest Xe/Kr selectivity at dilute conditions among all reported Xe and Kr adsorption isotherms in our literature survey.

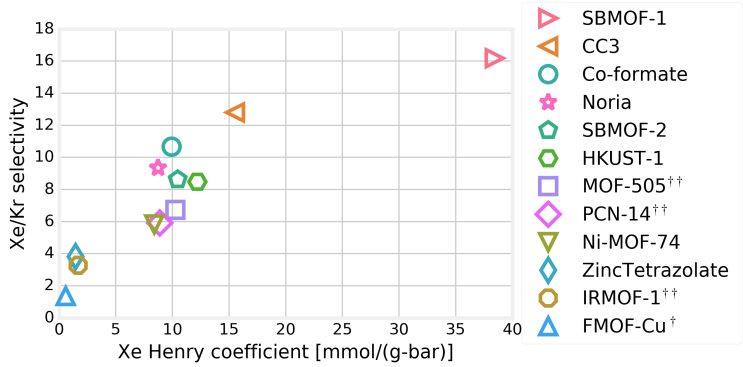
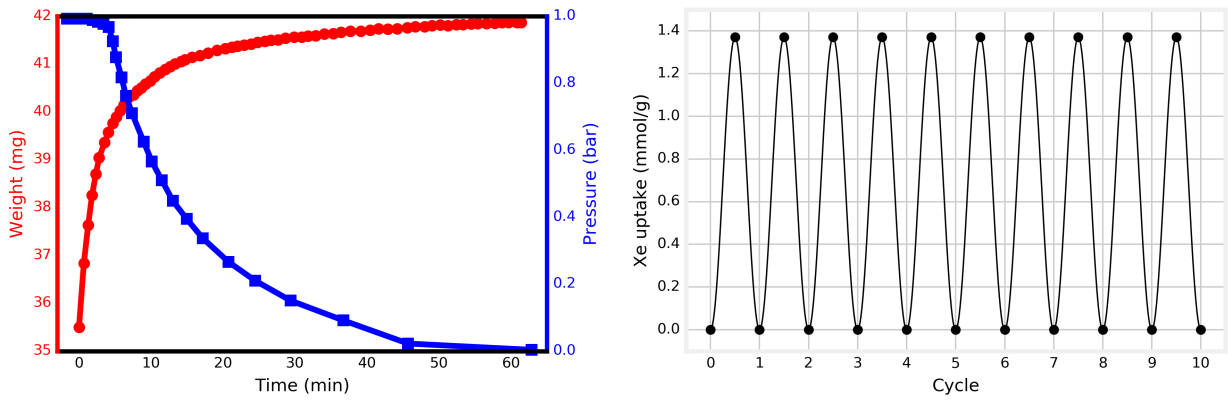


Figure 7.5: Survey of thermodynamic Xe/Kr separation performance in top-performing materials. Henry coefficients are extracted from pure-component Xe and Kr adsorption isotherms reported in the literature. Data at 298 K, exceptions denoted by a dagger (†) for 297 K and a double dagger (††) for 292 K. Data concerns CC3 [77], Co-formate [90, 85], noria [101], SBMOF-2 [102], HKUST-1 [103], MOF-505 [92], PCN-14 [104], Ni-MOF-74 [86], Zinc tetrazolate [105], IRMOF-1 [83], and FMOF-Cu [87].

7.4 Adsorption kinetics and column breakthrough experiments

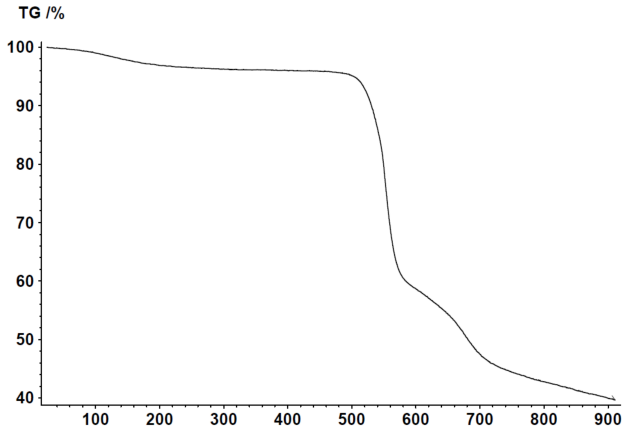
From a practical point of view, it is important that the kinetics of Xe adsorption/desorption are sufficiently fast and the material can undergo multiple ad-/de-sorption cycles without losing capacity. We measured the kinetics of Xe adsorption into an SBMOF-1 sample by connecting a chamber of Xe at 1bar and 298K to an evacuated chamber with the SBMOF-1 sample, then opening a valve to allow flow. Figure 7.6a shows that the rate of Xe uptake is sufficiently fast, reaching 80% of saturation uptake within 10min. Next, we performed 10 ad-/de-sorption cycles to test if SBMOF-1 retains its high Xe adsorption capacity after many cycles. Figure 7.6b shows that SBMOF-1 retains its performance after multiple cycles. In addition, SBMOF-1 shows high thermal stability up to 500K (see Fig. 7.7).

To demonstrate the practical applicability of SBMOF-1 for capturing Xe from UNF reprocessing off-gas, we conducted single-column breakthrough experiments with a representative gas mixture (400ppm Xe, 40ppm Kr, 78.1% N_2, 20.9% O_2, 0.03% CO_2 and 0.9% Ar) [89]. We fed this gas mixture through a column packed with SBMOF-1 and initially purged with He. Figure 7.9a shows that all gases except Xe broke through the column within minutes, whereas Xe was retained in the column for more than an hour). This demonstrates that SBMOF-1 can selectively remove Xe from air at UNF reprocessing conditions. Under these conditions, SBMOF-1 adsorbed 13.2mmol Xe per kg, higher than the reported breakthrough Xe capacities of benchmark materials, Ni-MOF-74 (4.8mmol Xe per kg) [89] and CC3 (11mmol Xe per kg) [77]; see Fig 7.8. Breakthrough measurements were conducted on 20-35 mesh (500-850μm) pellets of SBMOF-1 (1.48g) using a gas mixture composition



(a) (b)

Figure 7.6: Experimental characterization of Xe and Kr adsorption in SBMOF-1. (c) Xe adsorption kinetics experiments. The blue curve shows the pressure drop in a chamber feeding Xe to an initially evacuated chamber with the SBMOF-1 sample; the red curve shows the corresponding weight increase due to Xe adsorption. (d) Xe adsorption/ desorption cycling data; a sinusoidal curve is superimposed on the data. (a,c,d) Data for SBMOF-1 at 298 K. Reproduced from Ref. [100] under the Creative Commons License Attribution 4.0 International License.



Supplementary Figure 20. Thermogravimetric pattern of as-synthesized SBMOF-1.

Figure 7.7: Thermogravimetric pattern of as-synthesized SBMOF-1. Reproduced from Ref. [100] under the Creative Commons License Attribution 4.0 International License.

simulating UNF conditions (400 ppm Xe, 40 ppm Kr, 78.1% N_2, 20.9% O_2, 0.03% CO_2 and 0.9% Ar).

The experimental breakthrough capacity is close to that predicted from the Henry coef-

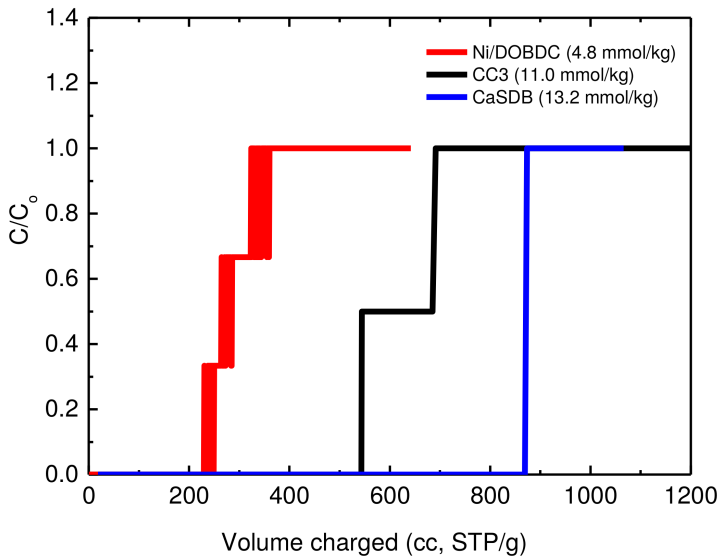


Figure 7.8: Comparison of breakthrough curves for Xe (400 ppm Xe in air) adsorbed on NiMOF-74, CC3 and SBMOF-1 at 298 K. The value (C/C_0) is outlet Xe concentration divided by the inlet Xe concentration (CaSDB = SBMOF-1). Reproduced from Ref. [100] under the Creative Commons License Attribution 4.0 International License.

ficient of the pure-component Xe isotherm (15.4mmolper kg), suggesting minimal diffusion limitations in the SBMOF-1 pellets. Next, we conducted column breakthrough experiments on SBMOF-1 in the presence of 42% relative humidity (Fig. 7.9b). Remarkably, SBMOF-1 retains a high Xe uptake ($\sim$11.5mmolkg^{-1}) even in the presence of water vapor. These results suggest the outstanding stability of SBMOF-1 makes it a practical material for the removal of Xe from UNF reprocessing off-gas. Such stability is a desirable property, as very few metal-organic hybrid materials exhibit such properties [293]. We postulate the absence of open metal sites to be responsible for the stability of SBMOF-1 in the presence of water vapour [294].

7.5 Revealing the Xe adsorption sites in SBMOF-1

To identify the location of adsorbed Xe and Kr, we performed single-crystal X-ray diffraction experiments on activated SBMOF-1. Single-crystal analysis of Xe-loaded SBMOF-1 reveals that Xe adsorbs at a single site, near the midpoint of the channel, interacting with the channel wall composed of aromatic rings by mainly van der Waals interactions. Due to symmetry considerations (space group $P2_1/c$), each Xe atom is positioned at two possible sites (Fig. 7.10a). The distance between each Xe atom along the b axis is 5.56Å, closely matching the b axis length of the unit cell. There are 1.72 atoms of Xe per unit cell based on

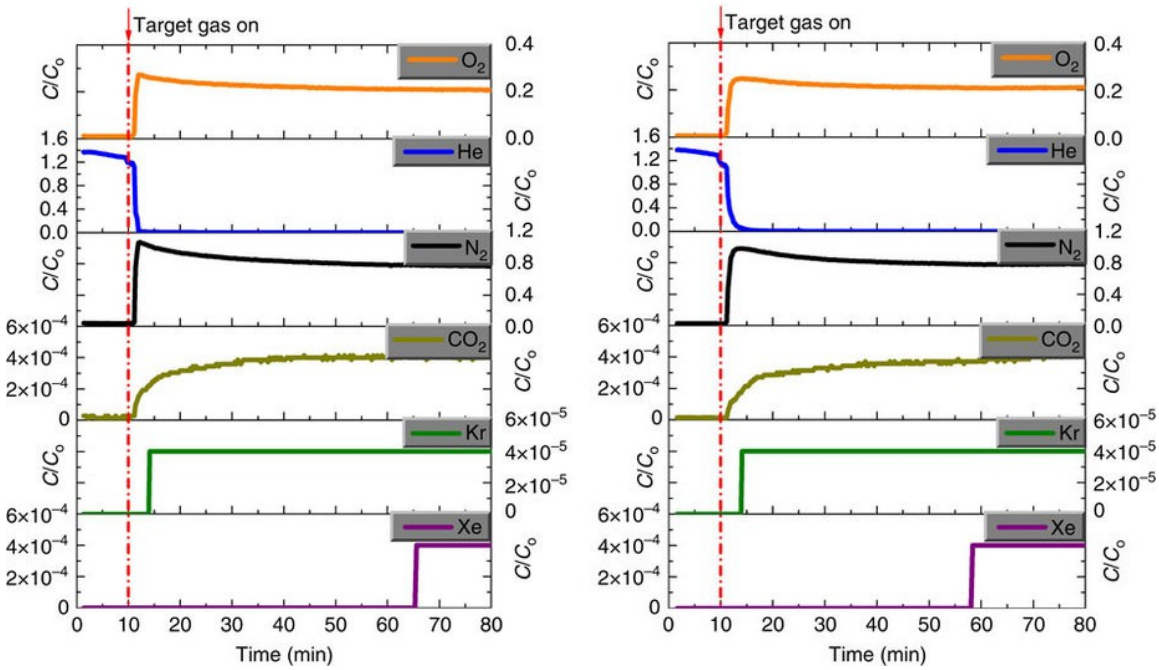
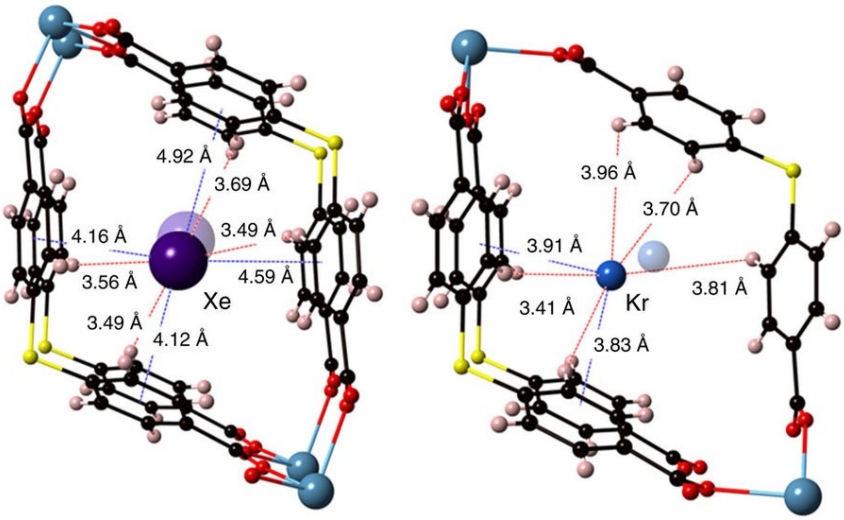


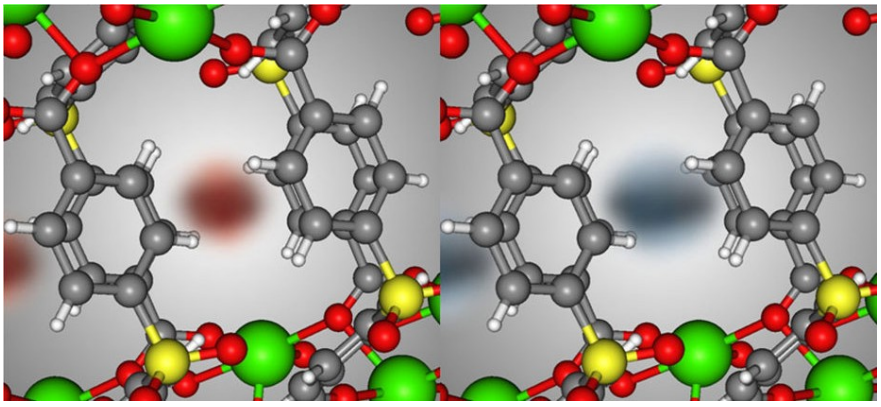
Figure 7.9: Single column breakthrough experiments using SBMOF-1 at room temperature and 1 atm. Column is initially purged with He. (a) Inlet is a dry gas mixture with 400p.p.m. Xe and 40p.p.m. Kr balanced with air. (b) Inlet is the same gas mixture as in (a) with 42% relative humidity. Note that the Xe breakthrough time is only marginally decreased in the presence of water. Reproduced from Ref. [100] under the Creative Commons License Attribution 4.0 International License.

crystallographic analysis (1.25mmol g^{-1}), close to the loading obtained from gas-adsorption data (1.38mmol g^{-1}). The saturation loading of Xe in Fig. 7.4 approaches two atoms per unit cell (see horizontal line), indicating commensurate Xe adsorption, which occurs when the adsorbed amount, location and orientation of an adsorbate are commensurate with the crystallographic symmetry of the adsorbent [295]. Such commensurate adsorption in SBMOF-1 was previously observed for small hydrocarbon molecules (C2-C3)[296, 297]. The observed position of Xe in the pore is consistent with molecular simulations of Xe adsorption (Figs 7.3b and 7.10b).

We can rationalize the high Xe adsorption capacity and selectivity exhibited by SBMOF-1 by its optimal Xe adsorption site. First, the pore size of SBMOF-1 is tailored for Xe [197, 95]. As a metric for pore size, we calculate the diameter of the largest included hard-sphere that can fit inside the pore of SBMOF-1. The pores of the MOF structures were characterized using the open-source Zeo++ [166, 298] code, which calculates the diameter of the largest hard-sphere that can fit within a pore anywhere within a given structure and the surface



(a)



(b)

Figure 7.10: Xe and Kr adsorption sites in SBMOF-1. (a) Xe and Kr positions determined by single-crystal X-ray diffraction. (b) Spatial probability densities from recording adsorbate positions during pure-component grand-canonical Monte Carlo simulations at 1bar and 298K (Xe, red; Kr, blue). Reproduced from Ref. [100] under the Creative Commons License Attribution 4.0 International License.

area accessible to a given probe. We used the high accuracy setting and atomic radii adopted from the Cambridge Crystallographic Data Centre. We calculate the largest included sphere diameter in SBMOF-1 to be 4.2Å, slightly larger than a Xe atom, ∼4.1Å. Simulations of Xe/Kr adsorption in the database of experimental MOFs show that all of the most selective MOFs have pore sizes slightly larger than a Xe atom (Fig. 7.2b). Such a pore diameter is a prerequisite for a highly Xe-selective material, as the pore size controls the proximity and degree of overlap from multiple framework atoms contributing van der Waals interactions

from multiple directions to achieve a highly favorable host-Xe interaction. A pore of optimal size for Xe is suboptimal for Kr because of the size difference, so this forms a pore that is highly discriminatory for Xe over Kr [95]. As shown in Fig. 7.3a, the pore size of SBMOF-1 falls in the optimal pore size window for Xe/Kr separations [76], distinguishing it from other MOFs. Porous organic cage CC3, another outstanding Xe-selective material, also exhibits a pore size tailored for Xe (largest included sphere of 4.8Å), but SBMOF-1 constructs a denser wall of chemical moieties than CC3 to achieve a higher Xe binding energy, enhancing its preference for Xe. This is the second reason why SBMOF-1 is outstanding in Xe adsorption; the color in Fig. 7.2b, representing the energy of adsorption of a Xe atom, shows that the dense wall of SBMOF-1 surrounds a Xe atom to achieve a high energy of Xe adsorption and thus a high Xe selectivity, following the trend in other MOFs.

7.6 Discussion

We demonstrated that a nanoporous MOF, SBMOF-1, identified as an outstanding Xe/Kr selective material from molecular simulations, shows exceptional Xe uptake at low pressure, selectivity for Xe, thermal and water stability, and adsorption kinetics. These attributes make SBMOF-1 potentially useful as a practical, near-term material for removal of Xe and Kr from nuclear reprocessing facilities with a far less energy requirement than cryogenic distillation. The selective adsorption of Xe from relevant gas mixtures even with ∼42% relative humidity demonstrate practicality and offer improvements over current technologies. Our recent economic analysis showed the cost benefits of using Ni-MOF-74 for an adsorption-based separation process at room temperature in comparison with cryogenic distillation [99, 13]. The discovery of the high Xe uptake and selectivity of SBMOF-1 at UNF reprocessing conditions– also in the presence of humidity– will enable an even more cost-effective process. The exceptional selectivity of SBMOF-1 is attributed to its pore size tailored to Xe and its dense wall of atoms that constructs a binding site with a high affinity for Xe, as evident by single-crystal X-ray diffraction and molecular models. As molecular simulations predicted SBMOF-1 to be among the most selective of ∼5,000 experimentally reported MOFs and ∼120,000 hypothetical MOF structures a priori, this work is a rare case of a computationally inspired materials discovery.

Chapter 8

Conclusions

Nanoporous materials have enormous internal surface areas [6] that endow them with adsorption properties useful for gas storage and separations. The modular and versatile chemistry of modern classes of nanoporous materials, such as metal-organic frameworks, allows for millions of possible materials to be synthesized. This is exciting, as we can tune the chemistry of nanoporous materials to target the adsorption of specific molecules for gas storage and separations. But, the chemical space of nanoporous materials is vast, and the rational tuning of MOF chemistry to target a specific molecule is difficult. In this thesis, the theme is that we can use molecular models and simulations to sort through the vast chemical space of nanoporous materials to pinpoint top structures, uncover structure-property relationships, and identify performance expectations. Such efforts help move away from Edisonian approaches to materials discovery and thus accelerate the discovery of materials targeted for a given application. In this thesis, we considered three applications of nanoporous materials applications: (i) vehicular methane densification and storage in Chapters 2, 3, and 4, (ii) Xe/Kr separations in Chapter 5, and (iii) Xe capture from air in Chapter 7. However, our approach of exploring chemical space *in silico* and using molecular models and simulations to rank materials is generalizable to other applications of nanoporous materials in chemical sensing, drug delivery, and catalysis– and generally applicable to other tunable materials as well, for example, batteries [150].

In Chapter 2, we constructed a series of models to gain insights into how to maximize the deliverable capacity of a gas in a material during a pressure-swing adsorption process. For vehicular methane storage, we desire a material with a high affinity for methane to maximize uptake at the fuel station, but the affinity for methane cannot be too high, as then the gas will not desorb to feed the engine. Throughout the literature, the heat of adsorption was commonly used as a metric for evaluating the deliverable capacity of a material. We showed that this can be misguided because the optimal heat of adsorption depends on the pore size, which is related to the entropy of adsorption. Another key conclusion from our models is that adsorbate-adsorbate attractions can enhance the deliverable capacity by recruiting more gas at the charging pressure.

In Chapter 3, we conducted a high-throughput computational screening of several classes

of nanoporous materials– amounting to over half a million structures– to identify materials that meet the deliverable capacity target set by ARPA-E. We concluded that the ARPA-E target is likely too ambitious for rigid, nanoporous materials. Note that we did not consider flexibility. Interestingly, MOF Co(bdp) now holds the record deliverable capacity for vehicular natural gas storage [299]. Co(bdp) is gate-opening, in that it is non-porous and adsorbs negligible gas at low pressures, but then abruptly expands to adsorb gas at higher pressures. The flexing of its unit cell, which induces a step in its methane adsorption isotherm, was a key feature in its achievement of such a high methane deliverable capacity.

In Chapter 4, we constructed a two simple thermodynamic models to explore thermodynamic limits to the methane deliverable capacity in nanoporous materials. Intuitively, we need material atoms to provide a surface for attractions with methane, but these material atoms take up space that methane could otherwise occupy. There is thus a material penalty for creating favorable energetics of adsorption. We considered an extreme in Chapter 4, where a uniform energy field could be created with zero material penalty. Under this scenario, we observed a maximal deliverable capacity of 357 v STP/v at the optimal strength of the energy field. This suggests that the ARPA-E target of 315 v STP/v is not currently met due to material chemistry and design constraints, not purely thermodynamic constraints. However, even under this very optimistic scenario, 357 v STP/v is close enough to 315 v STP/v to question if the ARPA-E deliverable capacity target could ever practically be achieved in a rigid material.

In Chapter 5, we conducted a high-throughput screening of half a million materials for Xe/Kr separations. The novel contribution here is that we moved beyond the brute-force computational screening, where we simulate adsorption in each and every material. Instead, we turned to more quickly computed and generalizable structural descriptors, such as the pore size and surface area. Because performance (i.e., Xe/Kr selectivity) is correlated strongly with such descriptors, we can use them to quickly rule out a large fraction of the materials. By incorporating more descriptors, the prediction becomes more reliable as we add more information, yet also the relationship between performance and these descriptors becomes more complicated as the relationship lies in a manifold in a higher dimensional space. The novel contribution here was the application of a machine learning algorithm, a random forest, to learn the relationship between structural descriptors and selectivity. Employing the random forest greatly accelerated the screening study by ruling out a large fraction of the materials on the basis of their quickly computed structural descriptors; we conducted the more expensive grand-canonical Monte Carlo simulations to predict the selectivity in materials that the random forest predicted would be selective.

Since we started working with gas mixtures, we noticed a demand for an open-sourced code for ideal adsorbed solution theory (IAST) calculations. IAST is a powerful thermodynamic framework, as it allows us to predict mixed gas adsorption isotherms from pure component gas adsorption isotherms, which are much easier to measure in practice. Chapter 6 provides an intuitive derivation of IAST and a thorough, practical understanding of how to apply it. Several researchers are now using the Python package pyIAST.

In Chapter 7, we conducted a high-throughput screening of metal-organic frameworks for

capturing Xe from air. This separation is relevant in used nuclear fuel reprocessing, and the main objective is to maintain a high selectivity for Xe over Kr, which is also in the air stream, since Kr is radioactive in this setting. SBMOF-1 was identified as a Xe-selective material in Chapter 5. In Chapter 7, we found that SBMOF-1 was predicted to be highly selective at dilute conditions as well. Our collaborators at Pacific Northwest National Laboratory synthesized and tested SBMOF-1 and found it to exhibit the highest Xe Henry coefficient and Xe/Kr selectivity reported in the literature. This collaboration and outcome was a remarkable end to my PhD thesis, as it is a rare story of a computer-assisted discovery of a new nanoporous material for gas separations.

8.1 Outlook

The grand vision is to design on the computer materials tailored for niche applications and conditions. While there is statistical signal in a ranking of materials from a high-throughput computational screening with classical models as in this thesis, there exists much room for improvement in the accuracy of current classical models. More accurate property prediction will likely come from a variety of avenues: more work on developing force fields tailored for specific chemical environments, more computational power and efficient algorithms for electronic structure calculations, and software/pipelines for characterizing the energy landscapes of porous materials with electronic structure calculations as input.

In our high-throughput computational screenings in this thesis, we held each structure rigid as an approximation, which is a practical one as (i) intrahost force fields for nanoporous materials are sparse, (ii) the computational resources required to treat flexibility are exorbitant, and (iii) additional parameters/degrees of freedom from intrahost forcefields could induce even more error in the prediction of properties (Occam's razor, overfitting). However, some MOFs are known to flex and significantly change their structure upon gas adsorption [171, 172, 173, 174, 175]. Treating the flexibility of nanoporous materials is another step towards more accurate property prediction and will also be achieved by better models and more computational power.

I am confident that, in the future, materials will be routinely designed and evaluated on a computer before experiments are performed.

My interests now lie in modeling flexible/dynamic, porous materials that respond to guest molecules and external stimuli [171, 172, 173, 174, 175]. Flexible constituents in porous crystals give rise to rich behaviors that can be exploited for engineering processes and may endow materials with enzyme-like specificity [172] for the selective recognition and capture of molecules. Insights from modeling are highly valuable for learning how to harness these flexible materials for separations, chemical sensing [14], catalysis [15], and drug delivery [16, 17]. Already, the flexibility of Co(bdp) has proven useful for achieving a high deliverable capacity for methane storage [299]. The recent Nobel Prize in Chemistry for molecular machines underscores the exciting possibilities enabled by our ability to design at the molecular scale chemical structures with moving and controllable parts.

Bibliography

[1] Hong-Cai Zhou, Jeffrey R. Long, and Omar M. Yaghi. Introduction to metal-organic frameworks. *Chem. Rev.*, 112(2):673–674, 2012.

[2] Weigang Lu, Daqiang Yuan, Dan Zhao, Christine Inge Schilling, Oliver Plietzsch, Thierry Muller, Stefan Bräse, Johannes Guenther, Janet Blümel, Rajamani Krishna, Zhen Li, and Hong-Cai Zhou. Porous polymer networks: Synthesis, porosity, and applications in gas storage/separation. *Chem. Mater.*, 22(21):5964–5972, 2010.

[3] Adrien P. Côté, Annabelle I. Benin, Nathan W. Ockwig, Michael O'Keeffe, Adam J. Matzger, and Omar M. Yaghi. Porous, crystalline, covalent organic frameworks. *Science*, 310(5751):1166–1170, 2005.

[4] Kyo Sung Park, Zheng Ni, Adrien P Côté, Jae Yong Choi, Rudan Huang, Fernando J Uribe-Romo, Hee K Chae, Michael O'Keeffe, and Omar M Yaghi. Exceptional chemical and thermal stability of zeolitic imidazolate frameworks. *Proc. Natl. Acad. Sci.*, 103(27):10186–10191, 2006.

[5] Tomokazu Tozawa, James TA Jones, Shashikala I Swamy, Shan Jiang, Dave J Adams, Stephen Shakespeare, Rob Clowes, Darren Bradshaw, Tom Hasell, Samantha Y Chong, Chiu Tang, Stephen Thompson, Julia Parker, Abbie Trewin, John Bacsa, Alexandra Slawin, Alexander Steiner, and Andrew Cooper. Porous organic cages. 8(12):973–978, 2009.

[6] Omar K Farha, Ibrahim Eryazici, Nak Cheon Jeong, Brad G Hauser, Christopher E Wilmer, Amy A Sarjeant, Randall Q Snurr, SonBinh T Nguyen, A Ozgur Yazaydın, and Joseph T Hupp. Metal-organic framework materials with ultrahigh surface areas: is the sky the limit? *J Am Chem Soc*, 134(36):15016–15021, 2012.

[7] Alexander Schoedel, Zhe Ji, and Omar M Yaghi. The role of metal-organic frameworks in a carbon-neutral energy cycle. *Nat Energy*, 1:16034, 2016.

[8] RussellE. Morris and Paul S. Wheatley. Gas storage in nanoporous materials. *Angew. Chem., Int. Ed.*, 47(27):4966–4981, 2008.

[9] Trevor A. Makal, Jian-Rong Li, Weigang Lu, and Hong-Cai Zhou. Methane storage in advanced porous materials. *Chem. Soc. Rev.*, 41:7761–7779, 2012.

[10] Leslie J Murray, Mircea Dincă, and Jeffrey R Long. Hydrogen storage in metal-organic frameworks. *Chem Soc Rev*, 38(5):1294–1314, 2009.

[11] Jian-Rong Li, Ryan J Kuppler, and Hong-Cai Zhou. Selective gas adsorption and separation in metal-organic frameworks. *Chem Soc Rev*, 38(5):1477–1504, 2009.

[12] Kenji Sumida, David L Rogow, Jarad A Mason, Thomas M McDonald, Eric D Bloch, Zoey R Herm, Tae-Hyun Bae, and Jeffrey R Long. Carbon dioxide capture in metal-organic frameworks. *Chem Rev*, 112(2):724–781, 2011.

[13] Debasis Banerjee, Amy J. Cairns, Jian Liu, Radha K. Motkuri, Satish K. Nune, Carlos A. Fernandez, Rajamani Krishna, Denis M. Strachan, and Praveen K. Thallapally. Potential of metal-organic frameworks for separation of xenon and krypton. *Acc. Chem. Res.*, page DOI: 10.1021/ar5003126, 2014.

[14] Lauren E Kreno, Kirsty Leong, Omar K Farha, Mark Allendorf, Richard P Van Duyne, and Joseph T Hupp. Metal–organic framework materials as chemical sensors. *Chem. Rev.*, 112(2):1105–1125, 2011.

[15] JeongYong Lee, Omar K Farha, John Roberts, Karl A Scheidt, SonBinh T Nguyen, and Joseph T Hupp. Metal–organic framework materials as catalysts. *Chem. Soc. Rev.*, 38(5):1450–1459, 2009.

[16] Patricia Horcajada, Christian Serre, Guillaume Maurin, Naseem A Ramsahye, Francisco Balas, Maria Vallet-Regi, Muriel Sebban, Francis Taulelle, and Gérard Férey. Flexible porous metal-organic frameworks for a controlled drug delivery. *J Am Chem Soc*, 130(21):6774–6780, 2008.

[17] Patricia Horcajada, Tamim Chalati, Christian Serre, Brigitte Gillet, Catherine Sebrie, Tarek Baati, Jarrod F Eubank, Daniela Heurtaux, Pascal Clayette, Christine Kreuz, et al. Porous metal-organic-framework nanoscale carriers as a potential platform for drug delivery and imaging. *Nat Mater*, 9(2):172–178, 2010.

[18] Radu Custelcean and Bruce A Moyer. Anion separation with metal-organic frameworks. *Eur J Inorg Chem*, 2007(10):1321–1340, 2007.

[19] Xinlei Liu, Nilay Keser Demir, Zhentao Wu, and Kang Li. Highly water-stable zirconium metal–organic framework UiO-66 membranes supported on alumina hollow fibers for desalination. *J Am Chem Soc*, 137(22):6999–7002, 2015.

[20] Muhammad Usman, Shruti Mendiratta, and Kuang-Lieh Lu. Metal-organic frameworks: New interlayer dielectric materials. *ChemElectroChem*, 2(6):786–788, 2015.

[21] MD Allendorf, CA Bauer, RK Bhakta, and RJT Houk. Luminescent metal-organic frameworks. *Chem Soc Rev*, 38(5):1330–1352, 2009.

[22] Omar M Yaghi, Michael O'Keeffe, Nathan W Ockwig, Hee K Chae, Mohamed Eddaoudi, and Jaheon Kim. Reticular synthesis and the design of new materials. *Nature*, 423(6941):705–714, 2003.

[23] Hiroyasu Furukawa, Kyle E Cordova, Michael OKeeffe, and Omar M Yaghi. The chemistry and applications of metal-organic frameworks. *Science*, 341(6149):1230444, 2013.

[24] Stephen D Burd, Shengqian Ma, Jason A Perman, Benjamin J Sikora, Randall Q Snurr, Praveen K Thallapally, Jian Tian, Lukasz Wojtas, and Michael J Zaworotko. Highly selective carbon dioxide uptake by [Cu(bpy-n)2 (SiF6)](bpy-1= 4, 4-Bipyridine; bpy-2= 1, 2-bis (4-pyridyl) ethene). *J Am Chem Soc*, 134(8):3663–3666, 2012.

[25] Xili Cui, Kaijie Chen, Huabin Xing, Qiwei Yang, Rajamani Krishna, Zongbi Bao, Hui Wu, Wei Zhou, Xinglong Dong, Yu Han, et al. Pore chemistry and size control in hybrid porous materials for acetylene capture from ethylene. *Science*, page aaf2458, 2016.

[26] A Cadiau, K Adil, PM Bhatt, Y Belmabkhout, and M Eddaoudi. A metal-organic framework-based splitter for separating propylene from propane. *Science*, 353(6295):137–140, 2016.

[27] Osama Shekhah, Youssef Belmabkhout, Zhijie Chen, Vincent Guillerm, Amy Cairns, Karim Adil, and Mohamed Eddaoudi. Made-to-order metal-organic frameworks for trace carbon dioxide removal and air capture. *Nat Commun*, 5, 2014.

[28] Sameh K Elsaidi, Mona H Mohamed, Herbert T Schaef, Amrit Kumar, Matteo Lusi, Tony Pham, Katherine A Forrest, Brian Space, Wenqian Xu, Gregory J Halder, et al. Hydrophobic pillared square grids for selective removal of CO2 from simulated flue gas. *Chem Commun*, 51(85):15530–15533, 2015.

[29] Patrick Nugent, Youssef Belmabkhout, Stephen D Burd, Amy J Cairns, Ryan Luebke, Katherine Forrest, Tony Pham, Shengqian Ma, Brian Space, Lukasz Wojtas, et al. Porous materials with optimal adsorption thermodynamics and kinetics for CO2 separation. *Nature*, 495(7439):80–84, 2013.

[30] Mei-Jin Lin, Abdelaziz Jouaiti, Nathalie Kyritsakas, and Mir Wais Hosseini. Molecular tectonics: modulation of size and shape of cuboid 3-D coordination networks. *CrystEngComm*, 11(1):189–191, 2009.

[31] Mian Li, Dan Li, Michael OKeeffe, and Omar M Yaghi. Topological analysis of metal-organic frameworks with polytopic linkers and/or multiple building units and the minimal transitivity principle. *Chem Rev*, 114(2):1343–1370, 2013.

[32] Christopher M Dobson. Chemical space and biology. *Nature*, 432(7019):824–828, 2004.

[33] Stephen S-Y Chui, Samuel M-F Lo, Jonathan PH Charmant, A Guy Orpen, and Ian D Williams. A chemically functionalizable nanoporous material $[Cu_3(TMA)_2(H_2O)_3]_n$. *Science*, 283(5405):1148–1150, 1999.

[34] Cory Simon, Jihan Kim, Yongchul G Chung, Richard L Martin, Rocio Mercado, Michael W Deem, Dan Gunter, Maciej Haranczyk, Randall Q Snurr, and Berend Smit. Computer-aided search for materials to store natural gas for vehicles. *New Discovery, Front. Young Minds-Understanding the Earth and its Resources*, 2015.

[35] Zhenqiang Wang and Seth M Cohen. Postsynthetic covalent modification of a neutral metal-organic framework. *J Am Chem Soc*, 129(41):12368–12369, 2007.

[36] Kristine K Tanabe and Seth M Cohen. Postsynthetic modification of metal-organic frameworks– a progress report. *Chem Soc Rev*, 40(2):498–519, 2011.

[37] Anubhav Jain, Shyue Ping Ong, Geoffroy Hautier, Wei Chen, William Davidson Richards, Stephen Dacek, Shreyas Cholia, Dan Gunter, David Skinner, Gerbrand Ceder, and Kristin A. Persson. Commentary: The materials project: A materials genome approach to accelerating materials innovation. *APL Mater.*, 1(1):011002, 2013.

[38] Jack H Lunsford. Catalytic conversion of methane to more useful chemicals and fuels: a challenge for the 21st century. *Catalysis Today*, 63(2):165–174, 2000.

[39] Kristina Konstas, Theresa Osl, Yunxia Yang, Michael Batten, Nick Burke, Anita J. Hill, and Matthew R. Hill. Methane storage in metal organic frameworks. *J. Mater. Chem.*, 22:16698–16708, 2012.

[40] Inventory of u.s. greenhouse gas emissions and sinks:1990-2011. pages `http://www.epa.gov/climatechange/Downloads/ghgemissions/US--GHG--Inventory--2013--ES.pdf`, U.S. E.P.A., 2013.

[41] M.Q. Wang and H.S. Huang. A full fuel-cycle analysis of energy and emissions impacts of transportation fuels produced from natural gas. *Center for Transportation Research, Argonne National Laboratory*, 1999.

[42] Ramón A Alvarez, Stephen W Pacala, James J Winebrake, William L Chameides, and Steven P Hamburg. Greater focus needed on methane leakage from natural gas infrastructure. *Proc. Natl. Acad. Sci.*, 109(17):6435–6440, 2012.

[43] Robert W Howarth, Renee Santoro, and Anthony Ingraffea. Methane and the greenhouse-gas footprint of natural gas from shale formations. *Climatic Change*, 106(4):679–690, 2011.

[44] Intergovernmental Panel on Climate Change. `http://www.ipcc.ch/publications_and_data/ar4/wg1/en/ch2s2-10-2.html`.

[45] US Energy Information Administration. Eia official shale gas estimates. 2013. http://www.eia.gov/naturalgas/weekly/#itn-tabs-4.

[46] US Energy Information Administration. http://www.eia.gov/dnav/ng/hist/rngwhhdA.htm,.

[47] U.S. Energy Information Administration. http://www.eia.gov/dnav/ng/ng_prod_shalegas_s1_a.htm.

[48] John N. Armor. Emerging importance of shale gas to both the energy and chemicals landscape. *Journal of Energy Chemistry*, 22(1):21 – 26, 2013.

[49] Adam P. Pacsi, Nawaf S. Alhajeri, Daniel Zavala-Araiza, Mort D. Webster, and David T. Allen. Regional air quality impacts of increased natural gas production and use in texas. *Environ. Sci. Technol.*, 47(7):3521–3527, 2013.

[50] U.S. Energy Information Administration. http://www.eia.gov/dnav/ng/ng_pri_sum_dcu_nus_a.htm.

[51] Jarad A. Mason, Mike Veenstra, and Jeffrey R. Long. Evaluating metal-organic frameworks for natural gas storage. *Chem. Sci.*, 5:32–51, 2014.

[52] NGV America. http://www.ngvc.org/incentives/federalTax.html.

[53] Mark Zoback, Saya Kitasei, and Brad Copithorne. Addressing the environmental risks from shale gas development. *World Watch Institute, Natural Gas and Sustainable Energy Initiative*, 2010. http://www.worldwatch.org/files/pdf/Hydraulic%20Fracturing%20Paper.pdf.

[54] Stephen G Osborn, Avner Vengosh, Nathaniel R Warner, and Robert B Jackson. Methane contamination of drinking water accompanying gas-well drilling and hydraulic fracturing. *Proc. Natl. Acad. Sci.*, 108(20):8172–8176, 2011.

[55] Steven Chu and Arun Majumdar. Opportunities and challenges for a sustainable energy future. *nature*, 488(7411):294–303, 2012.

[56] GA Whyatt. Issues affecting adoption of natural gas fuel in light- and heavy-duty vehicles. 2010. http://s3.amazonaws.com/zanran_storage/www.pnl.gov/ContentPages/184758856.pdf.

[57] MM Faruque Hasan, Alfred Minghan Zheng, and IA Karimi. Minimizing boil-off losses in liquefied natural gas transportation. *Industrial & Engineering Chemistry Research*, 48(21):9571–9580, 2009.

[58] J Wegrzyn and M Gurevich. Adsorbent storage of natural gas. *Applied Energy*, 55(2):71–83, 1996.

[59] Methane opportunities for vehicular energy (move). funding opportunity announcement. pages `http://arpa--e.energy.gov/?q=arpa--e--site--page/view--programs`, Advanced Research Projects Agency - Energy, US Department of Energy, 2012.

[60] J.P.Barbosa Mota, A.E. Rodrigues, E. Saatdjian, and D. Tondeur. Charge dynamics of a methane adsorption storage system: Intraparticle diffusional effects. *Adsorption*, 3(2):117–125, 1996.

[61] Air liquide.

[62] Stuart C. Cullen and Erwin G. Gross. The anesthetic properties of xenon in animals and human beings, with additional observations on krypton. *Science*, 113(2942):580–582, 1951.

[63] N. P. Franks, R. Dickinson, S. L. M. de Sousa, A. C. Hall, and W. R. Lieb. How does xenon produce anaesthesia? *Nature*, 396(6709):324–324, 1998.

[64] Robert D Sanders, Daqing Ma, and Mervyn Maze. Xenon: Elemental anaesthesia in clinical practice. *Brit. Med. Bull.*, 71(1):115–135, 2005.

[65] MS Albert, GD Cates, B Driehuys, W Happer, B Saam, CS Springer, and A Wishnia. Biological magnetic resonance imaging using laser-polarized 129xe. *Nature*, 370:199–201, 1994.

[66] JR Beattie, JN Matossian, RL Poeschel, WP Rogers, and RM Martinelli. Xenon ion propulsion subsystem. *J. Propul. Power*, 5(4):438–444, 1989.

[67] Serdar Yeralan, Douglas Doughty, Rudi Blondia, and Rick Hamburger. Advantages of using high-pressure short-arc xenon lamps for display systems. *Proc. SPIE*, 5740:27–35, 2005.

[68] William B Bridges and Arthur N Chester. Visible and uv laser oscillation at 118 wavelengths in ionized neon, argon, krypton, xenon, oxygen, and other gases. *Appl. Optics*, 4(5):573–580, 1965.

[69] Paul W. Hoff, James C. Swingle, and Charles K. Rhodes. Observations of stimulated emission from high-pressure krypton and argon/xenon mixtures. *Appl. Phys. Lett.*, 23(5):245–246, 1973.

[70] Heinrich Manz. On minimizing heat transport in architectural glazing. *Renew. Energ.*, 33(1):119 – 128, 2008.

[71] G. Weir and T. Muneer. Energy and environmental impact analysis of double-glazed windows. *Energ. Convers. Manage.*, 39(34):243 – 256, 1998.

[72] SR Lipsky and MM Shahin. Use of xenon and krypton as carrier gases for a highly sensitive detection system for gas chromatography. *Nature*, 200:566 – 567, 1963.

[73] Shuen-Cheng Hwang and William R. Weltmer. Helium group gases. In *Kirk-Othmer Encyclopedia of Chemical Technology*. John Wiley & Sons, Inc., 2000.

[74] Air liquide.

[75] Frank G Kerry. *Industrial Gas Handbook: Gas Separation and Purification*. CRC Press, 2010.

[76] Benjamin J Sikora, Christopher E Wilmer, Michael L Greenfield, and Randall Q Snurr. Thermodynamic Analysis of Xe/Kr Selectivity in over 137000 Hypothetical Metal-organic Frameworks. *Chem. Sci.*, 3(7):2217–2223, 2012.

[77] Linjiang Chen, Paul S Reiss, Samantha Y Chong, Daniel Holden, Kim E Jelfs, Tom Hasell, Marc A Little, Adam Kewley, Michael E Briggs, Andrew Stephenson, K. Mark Thomas, Jayne A. Armstrong, Jon Bell, Jose Busto, Raymond Noel, Jian Liu, Denis M. Strachan, Praveen K. Thallapally, and Andrew I. Cooper. Separation of rare gases and chiral molecules by selective binding in porous organic cages. *Nat. Mater.*, 13(10):954–960, 2014.

[78] Berend Smit, Jeffrey R Reimer, Curtis M Oldenburg, and Ian C Bourg. *Introduction to Carbon Capture and Sequestration*. Imperial College Press, London, 2014.

[79] Shivaji Sircar. Pressure swing adsorption. *Ind. Eng. Chem. Res.*, 41(6):1389–1392, 2002.

[80] R.E. Bazan, M. Bastos-Neto, A. Moeller, F. Dreisbach, and R. Staudt. Adsorption equilibria of o_2, ar, kr and xe on activated carbon and zeolites: Single component and mixture data. *Adsorption*, 17(2):371–383, 2011.

[81] Kenzo Munakata, Teruki Fukumatsu, Satoshi Yamatsuki, Kenji Tanaka, and Masabumi Nishikawa. Adsorption equilibria of krypton, xenon, nitrogen, and their mixtures on molecular sieve 5a and activated charcoal. *J. Nucl. Sci. Technol.*, 36(9):818–829, 1999.

[82] Masumeh Foroutan and Amir Taghavi Nasrabadi. Adsorption behavior of ternary mixtures of noble gases inside single-walled carbon nanotube bundles. *Chem. Phys. Lett.*, 497(4):213–217, 2010.

[83] Scott T Meek, Stephanie L Teich-McGoldrick, John J Perry, Jeffery A Greathouse, and Mark D Allendorf. Effects of polarizability on the adsorption of noble gases at low pressures in monohalogenated isoreticular metal–organic frameworks. *J Phys. Chem. C*, 116(37):19765–19772, 2012.

[84] Marie V Parkes, Chad L Staiger, John J Perry IV, Mark D Allendorf, and Jeffery A Greathouse. Screening metal-organic frameworks for selective noble gas adsorption in air: Effect of pore size and framework topology. *Phys. Chem. Chem. Phys.*, 15(23):9093–9106, 2013.

[85] Hao Wang, Kexin Yao, Zhijuan Zhang, Jacek Jagiello, Qihan Gong, Yu Han, and Jing Li. The first example of commensurate adsorption of atomic gas in a mof and effective separation of xenon from other noble gases. *Chem. Sci.*, 5(2):620–624, 2014.

[86] Praveen K Thallapally, Jay W Grate, and Radha Kishan Motkuri. Facile xenon capture and release at room temperature using a metal-organic framework: A comparison with activated charcoal. *Chem. Commun.*, 48(3):347–349, 2012.

[87] Carlos A. Fernandez, Jian Liu, Praveen K. Thallapally, and Denis M. Strachan. Switching kr/xe selectivity with temperature in a metal-organic framework. *J. Am. Chem. Soc.*, 134(22):9046–9049, 2012.

[88] Jian Liu, Denis M Strachan, and Praveen K Thallapally. Enhanced noble gas adsorption in ag@mof-74ni. *Chem. Commun.*, 50(4):466–468, 2014.

[89] Jian Liu, Praveen K. Thallapally, and Denis Strachan. Metal-organic frameworks for removal of xe and kr from nuclear fuel reprocessing plants. *Langmuir*, 28(31):11584–11589, 2012.

[90] Keith V Lawler, Zeric Hulvey, and Paul M Forster. Nanoporous metal formates for krypton/xenon separation. *Chem. Commun.*, 49(93):10959–10961, 2013.

[91] U Mueller, M Schubert, F Teich, H Puetter, K Schierle-Arndt, and J Pastre. Metal-organic frameworks – prospective industrial applications. *J. of Mater. Chem.*, 16(7):626–636, 2006.

[92] Youn-Sang Bae, Brad G. Hauser, Yamil J. Colón, Joseph T. Hupp, Omar K. Farha, and Randall Q. Snurr. High xenon/krypton selectivity in a metal-organic framework with small pores and strong adsorption sites. *Microporous Mesoporous Mater.*, 169(0):176–179, 2013.

[93] OV Magdysyuk, Frank Adams, H-P Liermann, Ioannis Spanopoulos, PN Trikalitis, Michael Hirscher, RE Morris, Morven J Duncan, LJ McCormick, and Robert E Dinnebier. Understanding the adsorption mechanism of noble gases kr and xe in cpo-27-ni, cpo-27-mg, and zif-8. *Phys. Chem. Chem. Phys.*, 16(43):23908–23914, 2014.

[94] G Boato and G Casanova. A self-consistent set of molecular parameters for neon, argon, krypton and xenon. *Physica*, 27(6):571 – 589, 1961.

[95] Cory M Simon, Rocio Mercado, Sondre K Schnell, Berend Smit, and Maciej Haranczyk. What are the best materials to separate a xenon/krypton mixture? *Chem. Mater.*, 27(12):4459–4475, 2015.

[96] Martin I Hoffert, Ken Caldeira, Gregory Benford, David R Criswell, Christopher Green, Howard Herzog, Atul K Jain, Haroon S Kheshgi, Klaus S Lackner, John S Lewis, et al. Advanced technology paths to global climate stability: energy for a greenhouse planet. *science*, 298(5595):981–987, 2002.

[97] Manfred Lenzen. Life cycle energy and greenhouse gas emissions of nuclear energy: A review. *Energy conversion and management*, 49(8):2178–2199, 2008.

[98] Nick R Soelberg, Troy G Garn, Mitchell R Greenhalgh, Jack D Law, Robert Jubin, Denis M Strachan, and Praveen K Thallapally. Radioactive iodine and krypton control for nuclear fuel reprocessing facilities. *Science and Technology of Nuclear Installations*, 2013, 2013.

[99] Jian Liu, Carlos A Fernandez, Paul F Martin, Praveen K Thallapally, and Denis M Strachan. A two-column method for the separation of kr and xe from process off-gases. *Industrial & Engineering Chemistry Research*, 53(32):12893–12899, 2014.

[100] Debasis Banerjee, Cory M Simon, Anna M Plonka, Radha K Motkuri, Jian Liu, Xianyin Chen, Berend Smit, John B Parise, Maciej Haranczyk, and Praveen K Thallapally. Metal-organic framework with optimally selective xenon adsorption and separation. *Nature communications*, 7, 2016.

[101] Rahul S Patil, Debasis Banerjee, Cory M Simon, Jerry L Atwood, and Praveen K Thallapally. Noria: A highly xe-selective nanoporous organic solid. *Chemistry-A European Journal*, 22(36):12618–12623, 2016.

[102] Xianyin Chen, Anna M Plonka, Debasis Banerjee, Rajamani Krishna, Herbert T Schaef, Sanjit Ghose, Praveen K Thallapally, and John B Parise. Direct observation of xe and kr adsorption in a xe-selective microporous metal–organic framework. *Journal of the American Chemical Society*, 137(22):7007–7010, 2015.

[103] Zeric Hulvey, Keith V Lawler, Zhiwei Qiao, Jian Zhou, David Fairen-Jimenez, Randall Q Snurr, Sergey V Ushakov, Alexandra Navrotsky, Craig M Brown, and Paul M Forster. Noble gas adsorption in copper trimesate, HKUST-1: an experimental and computational study. *J Phys Chem C*, 117(39):20116–20126, 2013.

[104] John J Perry IV, Stephanie L Teich-McGoldrick, Scott T Meek, Jeffery A Greathouse, Maciej Haranczyk, and Mark D Allendorf. Noble gas adsorption in metal–organic frameworks containing open metal sites. *J Phys. Chem. C*, 2014.

[105] Shunshun Xiong, Qiang Liu, Qian Wang, Wei Li, Yuanming Tang, Xiaolin Wang, Sheng Hu, and Banglin Chen. A flexible zinc tetrazolate framework exhibiting breathing behaviour on xenon adsorption and selective adsorption of xenon over other noble gases. *Journal of Materials Chemistry A*, 3(20):10747–10752, 2015.

[106] Marcus G. Martin and J. Ilja Siepmann. Transferable potentials for phase equilibria. 1. united-atom description of n-alkanes. *J Phys. Chem. B*, 102(14):2569–2577, 1998.

[107] Stephen L Mayo, Barry D Olafson, and William A Goddard. DREIDING: a generic force field for molecular simulations. *J Phys Chem*, 94(26):8897–8909, 1990.

[108] Anthony K Rappé, Carla J Casewit, KS Colwell, WA Goddard Iii, and WM Skiff. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.*, 114(25):10024–10035, 1992.

[109] M. Scott Shell. *Thermodynamics and Statistical Mechanics: An Integrated Approach.* Cambridge University Press, 2015.

[110] Daan Frenkel and Berend Smit. *Understanding Molecular Simulations: from Algorithms to Applications*, volume 1. Academic Press, San Diego, second edition, 2002.

[111] Jihan Kim and Berend Smit. Efficient Monte Carlo Simulations of Gas Molecules Inside Porous Materials. *J. Chem. Theory Comput.*, 8(7):2336–2343, 2012.

[112] Cory M. Simon, Jihan Kim, Li-Chiang Lin, Richard L. Martin, Maciej Haranczyk, and Berend Smit. Optimizing nanoporous materials for gas storage. *Physical Chemistry Chemical Physics*, 16:5499–5513, 2014.

[113] Christopher E. Wilmer, Omar K. Farha, Taner Yildirim, Ibrahim Eryazici, Vaiva Krungleviciute, Amy A. Sarjeant, Randall Q. Snurr, and Joseph T. Hupp. Gram-scale, high-yield synthesis of a robust metal-organic framework for storing methane and other gases. *Energy Environ. Sci.*, 6:1158–1163, 2013.

[114] Tina Düren, Lev Sarkisov, Omar M. Yaghi, and Randall Q. Snurr. Design of new materials for methane storage. *Langmuir*, 20(7):2683–2689, 2004.

[115] Mohamed Eddaoudi, Jaheon Kim, Nathaniel Rosi, David Vodak, Joseph Wachter, Michael O'Keeffe, and Omar M. Yaghi. Systematic design of pore size and functionality in isoreticular mofs and their application in methane storage. *Science*, 295(5554):469–472, 2002.

[116] Hiroyasu Furukawa and Omar M. Yaghi. Storage of hydrogen, methane, and carbon dioxide in highly porous covalent organic frameworks for clean energy applications. *J. Am. Chem. Soc.*, 131(25):8875–8883, 2009.

[117] Yang Peng, Vaiva Krungleviciute, Ibrahim Eryazici, Joseph T. Hupp, Omar K. Farha, and Taner Yildirim. Methane Storage in MetalOrganic Frameworks: Current Records, Surprise Findings, and Challenges. *J. Am. Chem. Soc.*, 135(32):11887–11894, 2013.

[118] Rachel B Getman, Youn-Sang Bae, Christopher E Wilmer, and Randall Q Snurr. Review and analysis of molecular simulations of methane, hydrogen, and acetylene storage in metal–organic frameworks. *Chem. Rev.*, 112(2):703–723, 2011.

[119] Michael W. Deem, Ramdas Pophale, Phillip A. Cheeseman, and David J. Earl. Computational discovery of new zeolite-like materials. *J. Phys. Chem. C*, 113(51):21353–21360, 2009.

[120] Ramdas Pophale, Phillip A. Cheeseman, and Michael W. Deem. A database of new zeolite-like materials. *Phys. Chem. Chem. Phys.*, 13:12407–12412, 2011.

[121] C. E. Wilmer, M. Leaf, C..Y. Lee, O. K. Farha, Brad G Hauser, Joseph T Hupp, and Randall Q Snurr. Large-scale screening of hypothetical metal-organic frameworks. *Nat. Chem.*, 4(2):83–89, 2011.

[122] Suresh K. Bhatia and Alan L. Myers. Optimum conditions for adsorptive storage. *Langmuir*, 22(4):1688–1700, 2006.

[123] Youn-Sang Bae and Randall Q. Snurr. Optimal isosteric heat of adsorption for hydrogen storage and delivery using metalorganic frameworks. *Microporous Mesoporous Mater.*, 132(12):300 – 303, 2010.

[124] Rachel B. Getman, Jacob H. Miller, Kenneth Wang, and Randall Q. Snurr. Metal alkoxide functionalization in metalorganic frameworks for enhanced ambient-temperature hydrogen storage. *J. Phys. Chem. C*, 115(5):2066–2075, 2011.

[125] E. Garrone, B. Bonelli, and C. Otero Aren. Enthalpyentropy correlation for hydrogen adsorption on zeolites. *Chem. Phys. Lett.*, 456(13):68 – 70, 2008.

[126] Thomas F. Willems, Chris H. Rycroft, Michaeel Kazi, Juan C. Meza, and Maciej Haranczyk. Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials. *Microporous Mesoporous Mater.*, 149(1):134 – 141, 2012.

[127] D Dubbeldam, S Calero, TJH Vlugt, R Krishna, Th LM Maesen, E Beerdsen, and B Smit. Force field parametrization through fitting on inflection points in isotherms. *Physical Review Letters*, 93(8):088302, 2004.

[128] Sofía Calero, David Dubbeldam, Rajamani Krishna, Berend Smit, Thijs JH Vlugt, Joeri FM Denayer, Johan A Martens, and Theo LM Maesen. Understanding the role of sodium during adsorption: A force field for alkanes in sodium-exchanged faujasites. *J. Am. Chem. Soc.*, 126(36):11377–11386, 2004.

[129] Richard L Martin, Mr Prabhat, David D Donofrio, James A Sethian, and Maciej Haranczyk. Accelerating analysis of void space in porous materials on multicore and gpu platforms. *Int. J. High Perform. Comput.*, 26(4):347–357, 2012.

[130] Jihan Kim, Richard L Martin, Oliver Rubel, Maciej Haranczyk, and Berend Smit. High-throughput characterization of porous materials using graphics processing units. *J. Chem. Theory Comput.*, 8(5):1684–1693, 2012.

[131] S. Sircar, R. Mohr, C. Ristic, and M. B. Rao. Isosteric heat of adsorption: theory and experiment. *J. Phys. Chem. B*, 103(31):6539–6546, 1999.

[132] Richard Luis Martin, Berend Smit, and Maciej Haranczyk. Addressing challenges of identifying geometrically diverse sets of crystalline porous materials. *J. Chem. Inf. Model.*, 52(2):308–318, 2012.

[133] V.C. Menon and S. Komarneni. Porous adsorbents for vehicular natural gas storage: A review. *J. Porous Mater.*, 5(1):43–58, 1998.

[134] Alan L. Myers. Characterization of nanopores by standard enthalpy and entropy of adsorption of probe molecules. *Colloids Surf., A*, 241(13):9 – 14, 2004. Proceedings of the Third International TRI/Princeton Workshop, Characterization of Porous Materials: from Angstroms to Millimeters.

[135] Kimberly R. Matranga, Alan L. Myers, and Eduardo D. Glandt. Storage of natural gas by adsorption on activated carbon. *Chem. Eng. Sci.*, 47(7):1569 – 1579, 1992.

[136] U. Ascher and C. Greif. A first course in numerical methods. 2011.

[137] Jihan Kim, Amitesh Maiti, Li-Chiang Lin, Joshuah K Stolaroff, Berend Smit, and Roger D Aines. New materials for methane capture from dilute and medium-concentration sources. *Nat. Commun.*, 4:1694, 2013.

[138] Barbara Schmitz, Ulrich Müller, Natalia Trukhan, Markus Schubert, Gérard Férey, and Michael Hirscher. Heat of adsorption for hydrogen in microporous high-surface-area materials. *ChemPhysChem*, 9(15):2181–4, October 2008.

[139] Steven Tedds, Allan Walton, Darren P. Broom, and David Book. Characterisation of porous hydrogen storage materials: carbons, zeolites, MOFs and PIMs. *Faraday Discuss.*, 151:75, 2011.

[140] J.A. Ripmeester and CI Ratcliffe. Application of xenon-129 nmr to the study of microporous solids. *J. Phys. Chem.*, 94(19):7652–7656, 1990.

[141] B.J. Cox, N. Thamwattana, and J.M. Hill. Mechanics of atoms and fullerenes in single-walled carbon nanotubes. ii. oscillatory behaviour. *Proc. R. Soc. London, Ser. A*, 463(2078):477–494, 2007.

[142] Jihan Kim, Li-Chiang Lin, Richard L Martin, Joseph A Swisher, Maciej Haranczyk, and Berend Smit. Large-scale computational screening of zeolites for ethane/ethene separation. *Langmuir*, 28(32):11914–11919, 2012.

[143] R.L. Martin, T.F. Willems, L.C. Lin, J. Kim, J.A. Swisher, B. Smit, and M. Haranczyk. Similarity-driven discovery of zeolite materials for adsorption-based separations. *ChemPhysChem*, 13(16):3595–3597, 2012.

[144] Raul F Lobo, Stacey I Zones, and Mark E Davis. Structure-direction in zeolite synthesis. *Journal of inclusion phenomena and molecular recognition in chemistry*, 21(1-4):47–78, 1995.

[145] Ramdas Pophale, Frits Daeyaert, and Michael W Deem. Computational prediction of chemically synthesizable organic structure directing agents for zeolites. *J. Mater. Chem. A*, 1(23):6750–6760, 2013.

[146] Leslie J. Murray, Mircea Dinca, and Jeffrey R. Long. Hydrogen storage in metal-organic frameworks. *Chem. Soc. Rev.*, 38:1294–1314, 2009.

[147] Cory M Simon, Jihan Kim, Diego A Gomez-Gualdron, Jeffrey S Camp, Yongchul G Chung, Richard L Martin, Rocio Mercado, Michael W Deem, Dan Gunter, Maciej Haranczyk, et al. The materials genome in action: identifying the performance limits for methane storage. *Energy Environ. Sci.*, 8(4):1190–1199, 2015.

[148] Jarad A Mason, Mike Veenstra, and Jeffrey R Long. Evaluating metal–organic frameworks for natural gas storage. *Chem. Sci.*, 5(1):32–51, 2014.

[149] Felipe Gándara, Hiroyasu Furukawa, Seungkyu Lee, and Omar M Yaghi. High methane storage capacity in aluminum metal–organic frameworks. *Journal of the American Chemical Society*, 136(14):5271–5274, 2014.

[150] Hailong Chen, Geoffroy Hautier, Anubhav Jain, Charles Moore, Byoungwoo Kang, Robert Doe, Lijun Wu, Yimei Zhu, Yuanzhi Tang, and Gerbrand Ceder. Carbonophosphates: a new family of cathode materials for li-ion batteries identified computationally. *Chemistry of Materials*, 24(11):2009–2016, 2012.

[151] Ivano E Castelli, David D Landis, Kristian S Thygesen, Søren Dahl, Ib Chorkendorff, Thomas F Jaramillo, and Karsten W Jacobsen. New cubic perovskites for one-and two-photon water splitting using the computational materials repository. *Energy & Environmental Science*, 5(10):9034–9043, 2012.

[152] Rickard Armiento, Boris Kozinsky, Marco Fornari, and Gerbrand Ceder. Screening for high-performance piezoelectrics using high-throughput density functional theory. *Physical Review B*, 84(1):014103, 2011.

[153] Carlos Ortiz, Olle Eriksson, and Mattias Klintenberg. Data mining and accelerated electronic structure theory as a tool in the search for new functional materials. *Computational Materials Science*, 44(4):1042–1049, 2009.

[154] Liping Yu and Alex Zunger. Identification of potential photovoltaic absorbers based on first-principles spectroscopic screening of materials. *Physical review letters*, 108(6):068701, 2012.

[155] Shidong Wang, Zhao Wang, Wahyu Setyawan, Natalio Mingo, and Stefano Curtarolo. Assessing the thermoelectric properties of sintered compounds via high-throughput ab-initio calculations. *Physical Review X*, 1(2):021012, 2011.

[156] Tom Kalil and Cyrus Wadia. Materials genome initiative: A renaissance of american manufacturing.

[157] Richard L Martin, Li-Chiang Lin, Kuldeep Jariwala, Berend Smit, and Maciej Haranczyk. Mail-order metal–organic frameworks (mofs): designing isoreticular mof-5 analogues comprising commercially available organic molecules. *The Journal of Physical Chemistry C*, 117(23):12159–12167, 2013.

[158] Li-Chiang Lin, Adam H. Berger, Richard L. Martin, Jihan Kim, Joseph A. Swisher, Kuldeep Jariwala, Chris H. Rycroft, Abhoyjit S. Bhown, Michael W. Deem, Maciej Haranczyk, and Berend Smit. In silico screening of carbon-capture materials. *Nat. Mater.*, 11(7):633–641, 2012.

[159] Richard Luis Martin, Cory M Simon, Berend Smit, and Maciej Haranczyk. In silico design of porous polymer networks: High-throughput screening for methane storage materials. *J. Am. Chem. Soc.*, 136(13):5006–5022, 2014.

[160] Yongchul G. Chung, Jeffrey Camp, Maciej Haranczyk, Benjamin J. Sikora, Wojciech Bury, Vaiva Krungleviciute, Taner Yildirim, Omar K. Farha, David S. Sholl, and Randall Q. Snurr. Computation-ready, experimental metal-organic frameworks: A tool to enable high-throughput screening of nanoporous crystals. *Chem. Mater.*, 26(21):6185–6192, 2014.

[161] Diego A Gomez-Gualdron, Oleksii V Gutov, Vaiva Krungleviciute, Bhaskarjyoti Borah, Joseph E Mondloch, Joseph T Hupp, Taner Yildirim, Omar K Farha, and Randall Q Snurr. Computational design of metal-organic frameworks based on stable zirconium building units for storage and delivery of methane. *Chem Mater*, 26(19):5632–5639, 2014.

[162] Jasmina Hafizovic Cavka, Søren Jakobsen, Unni Olsbye, Nathalie Guillou, Carlo Lamberti, Silvia Bordiga, and Karl Petter Lillerud. A new zirconium inorganic building brick forming metal organic frameworks with exceptional stability. *Journal of the American Chemical Society*, 130(42):13850–13851, 2008.

[163] Richard Luis Martin and Maciej Haranczyk. Construction and characterization of structure models of crystalline porous polymers. *Crystal Growth & Design*, 14(5):2431–2440, 2014.

[164] Frank H Allen. The cambridge structural database: a quarter of a million crystal structures and rising. *Acta Crystallographica Section B: Structural Science*, 58(3):380–388, 2002.

[165] Maciej Harancyzk. Nanoporous materials database.

[166] Thomas F Willems, Chris H Rycroft, Michaeel Kazi, Juan C Meza, and Maciej Haranczyk. Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials. *Microporous Mesoporous Mater.*, 149(1):134–141, 2012.

[167] Tu Le, V. Chandana Epa, Frank R. Burden, and David A. Winkler. Quantitative structure-property relationship modeling of diverse materials properties. *Chem. Rev.*, 112(5):2889–2919, 2012.

[168] Michael Fernandez, Tom K Woo, Christopher E Wilmer, and Randall Q Snurr. Large-scale quantitative structure–property relationship (qspr) analysis of methane storage in metal–organic frameworks. *The Journal of Physical Chemistry C*, 117(15):7681–7689, 2013.

[169] Diego A Gomez-Gualdron, Christopher E Wilmer, Omar K Farha, Joseph T Hupp, and Randall Q Snurr. Exploring the limits of methane storage and delivery in nanoporous materials. *The Journal of Physical Chemistry C*, 118(13):6941–6951, 2014.

[170] D Dubbeldam, S Calero, TJH Vlugt, R Krishna, Th LM Maesen, and B Smit. United atom force field for alkanes in nanoporous materials. *The Journal of Physical Chemistry B*, 108(33):12301–12313, 2004.

[171] Ashleigh J Fletcher, K Mark Thomas, and Matthew J Rosseinsky. Flexibility in metal-organic framework materials: Impact on sorption properties. *J Solid State Chem*, 178(8):2491–2510, 2005.

[172] Satoshi Horike, Satoru Shimomura, and Susumu Kitagawa. Soft porous crystals. *Nat Chem*, 1(9):695–704, 2009.

[173] Andreas Schneemann, Volodymyr Bon, Inke Schwedler, Irena Senkovska, Stefan Kaskel, and Roland A Fischer. Flexible metal-organic frameworks. *Chem Soc Rev*, 43(16):6062–6096, 2014.

[174] Francois-Xavier Coudert. Responsive metal–organic frameworks and framework materials: Under pressure, taking the heat, in the spotlight, with friends. *Chem. Mater.*, 27(6):1905–1916, 2015.

[175] Gérard Férey and Christian Serre. Large breathing effects in three-dimensional porous hybrid matter: facts, analyses, rules and consequences. *Chem Soc Rev*, 38(5):1380–1399, 2009.

[176] Allison L Dzubak, Li-Chiang Lin, Jihan Kim, Joseph A Swisher, Roberta Poloni, Sergey N Maximoff, Berend Smit, and Laura Gagliardi. Ab initio carbon capture in open-site metal–organic frameworks. *Nature chemistry*, 4(10):810–816, 2012.

[177] Wendy Queen, Matthew Hudson, Eric Bloch, Jarad A. Mason, Miguel Gonzalez, Jason Lee, David Gygi, Joshua Howe, Kyuho Lee, Tamim Darwish, Michael James, Vanessa K. Peterson, Simon Teat, Berend Smit, Jeffrey Neaton, Jeffrey R. Long, and Craig M. Brown. Comprehensive study of carbon dioxide adsorption in the metal-organic frameworks m2(dobdc) (m = mg, mn, fe, co, ni, cu, zn). *Chem. Sci.*, 2014.

[178] Marcus Rose, Winfried Böhlmann, Michal Sabo, and Stefan Kaskel. Element–organic frameworks with high permanent porosity. *Chemical Communications*, (21):2462–2464, 2008.

[179] Bin Li, Hui-Min Wen, Hailong Wang, Hui Wu, Madhusudan Tyagi, Taner Yildirim, Wei Zhou, and Banglin Chen. A porous metal–organic framework with dynamic pyrimidine groups exhibiting record high methane storage working capacity. *Journal of the American Chemical Society*, 136(17):6207–6210, 2014.

[180] Felipe Gndara, Hiroyasu Furukawa, Seungkyu Lee, and Omar M. Yaghi. High methane storage capacity in aluminum metal-organic frameworks. *Journal of the American Chemical Society*, 136(14):5271–5274, 2014.

[181] Mohamed Eddaoudi, Jaheon Kim, Nathaniel Rosi, David Vodak, Joseph Wachter, Michael O'Keeffe, and Omar M Yaghi. Systematic design of pore size and functionality in isoreticular MOFs and their application in methane storage. *Science*, 295(5554):469–472, 2002.

[182] Daofeng Sun, Shengqian Ma, Jason M Simmons, Jian-Rong Li, Daqiang Yuan, and Hong-Cai Zhou. An unusual case of symmetry-preserving isomerism. *Chem. Commun.*, 46(8):1329–1331, 2010.

[183] Javier Pérez-Pellitero, Hedi Amrouche, Flor R Siperstein, Gerhard Pirngruber, Carlos Nieto-Draghi, Gérald Chaplais, Angélique Simon-Masseron, Delphine Bazer-Bachi, David Peralta, and Nicolas Bats. Adsorption of CO2, CH4, and N2 on zeolitic imidazolate frameworks: experiments and simulations. *Chemistry-A European Journal*, 16(5):1560–1571, 2010.

[184] Daqiang Yuan, Weigang Lu, Dan Zhao, and Hong-Cai Zhou. Highly stable porous polymer networks with exceptionally high gas-uptake capacities. *Advanced Materials*, 23(32):3723–3725, 2011.

[185] Weigang Lu, Daqiang Yuan, Dan Zhao, Christine Inge Schilling, Oliver Plietzsch, Thierry Muller, Stefan Brase, Johannes Guenther, Janet Blumel, Rajamani Krishna, et al. Porous polymer networks: synthesis, porosity, and applications in gas storage/separation. *Chemistry of Materials*, 22(21):5964–5972, 2010.

[186] Shuji Himeno, Toshihiro Tomita, Kenji Suzuki, and Shuichi Yoshida. Characterization and selectivity for methane and carbon dioxide adsorption on the all-silica DD3R zeolite. *Microporous Mesoporous Mater.*, 98(1):62–69, 2007.

[187] EM Flanigen, JM Bennett, RW Grose, JP Cohen, RL Patton, and RM Kirchner. Silicalite, a new hydrophobic crystalline silica molecular sieve. *Nature*, 271:512–516, 1978.

[188] Guo-Qiang Kong, Zhi-Da Han, Yabing He, Sha Ou, Wei Zhou, Taner Yildirim, Rajamani Krishna, Chao Zou, Banglin Chen, and Chuan-De Wu. Expanded organic building units for the construction of highly porous metal–organic frameworks. *Chemistry–A European Journal*, 19(44):14886–14894, 2013.

[189] Yabing He, Wei Zhou, Taner Yildirim, and Banglin Chen. A series of metal–organic frameworks with high methane uptake and an empirical equation for predicting methane storage capacity. *Energy & Environmental Science*, 6(9):2735–2744, 2013.

[190] Eric D Bloch, Wendy L Queen, Rajamani Krishna, Joseph M Zadrozny, Craig M Brown, and Jeffrey R Long. Hydrocarbon separations in a metal-organic framework with open iron(II) coordination sites. *Science*, 335(6076):1606–1610, 2012.

[191] Ying Dai, JR Johnson, Oğuz Karvan, David S Sholl, and WJ Koros. Ultem®/zif-8 mixed matrix hollow fiber membranes for co_2/n_2 separations. *Journal of Membrane Science*, 401:76–82, 2012.

[192] Diego A Gómez-Gualdrón, Cory M Simon, William Lassman, David Chen, Richard L Martin, Maciej Haranczyk, Omar K Farha, Berend Smit, and Randall Q Snurr. Impact of the strength and spatial distribution of adsorption sites on methane deliverable capacity in nanoporous materials. *Chemical Engineering Science*, 2016.

[193] WH Ramsey. On the densities of methane, metallic ammonium, water and neon at planetary pressures. *Monthly Notices of the Royal Astronomical Society*, 125(5):469–485, 1962.

[194] Nanoporous materials genome center.

[195] Richard Martin, Cory Simon, Bharat Medasani, David Britt, Berend Smit, and Maciej Haranczyk. In silico design of three-dimensional porous covalent organic frameworks via known synthesis routes and commercially available species. *J. Phys. Chem. C*, 19(1):186–195, 2014.

[196] *International Zeolite Association*. `http://www.iza-structure.org/databases/`.

[197] Patrick Ryan, Omar K Farha, Linda J Broadbelt, and Randall Q Snurr. Computational screening of metal-organic frameworks for xenon/krypton separation. *AIChE J.*, 57(7):1759–1766, 2011.

[198] Timothy Van Heest, Stephanie L Teich-McGoldrick, Jeffery A Greathouse, Mark D Allendorf, and David S Sholl. Identification of metal–organic framework materials for adsorption separation of rare gases: Applicability of ideal adsorbed solution theory (IAST) and effects of inaccessible framework regions. *J. Phys. Chem. C*, 116(24):13183–13195, 2012.

[199] Leo Breiman. Random forests. *Mach. Learn.*, 45(1):5–32, 2001.

[200] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Mach. Learn.*, 20(3):273–297, 1995.

[201] Trevor Hastie, Robert Tibshirani, Jerome Friedman, T Hastie, J Friedman, and R Tibshirani. *The elements of statistical learning*, volume 2. Springer, 2009.

[202] Michael Fernandez, Tom K. Woo, Christopher E. Wilmer, and Randall Q. Snurr. Large-scale quantitative structureproperty relationship (qspr) analysis of methane storage in metal-organic frameworks. *J. Phys. Chem. C*, 117(15):7681–7689, 2013.

[203] Michael Fernandez, Peter G Boyd, Thomas D Daff, Mohammad Zein Aghaji, and Tom K Woo. Rapid and accurate machine learning recognition of high performing metal organic frameworks for co_2 capture. *J. Phys. Chem. Lett.*, 5:3056–3060, 2014.

[204] Thomas G. Dietterich. *Ensemble Methods in Machine Learning*, volume 1857 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2000.

[205] Kevin P Murphy. *Machine Learning: A Probabilistic Perspective*. MIT Press, 2012.

[206] James Surowiecki. *The Wisdom of Crowds: Why the Many are Smarter Than the Few and How Collective Wisdom Shapes Business*. Doubleday; Anchor, 2004.

[207] Rich Caruana and Alexandru Niculescu-Mizil. An empirical comparison of supervised learning algorithms. In *Proc. Int. Conf. Machine Learn., 23rd, Pittsburgh, PA.*, pages 161–168. ACM, 2006.

[208] Jamie Shotton, Toby Sharp, Alex Kipman, Andrew Fitzgibbon, Mark Finocchio, Andrew Blake, Mat Cook, and Richard Moore. Real-time human pose recognition in parts from single depth images. *Commun. ACM*, 56(1):116–124, 2013.

[209] Ronald W Kennard and Larry A Stone. Computer aided design of experiments. *Technometrics*, 11(1):137–148, 1969.

[210] Orhan Talu and Alan L Myers. Reference potentials for adsorption of helium, argon, methane, and krypton in high-silica zeolites. *Colloids Surf., A*, 187-188(0):83–93, 2001.

[211] Lev Sarkisov, Richard L Martin, Maciej Haranczyk, and Berend Smit. On the flexibility of metal-organic frameworks. *J. Am. Chem. Soc.*, 136(6):2228–2231, 2014.

[212] TC Golden and S Sircar. Gas adsorption on silicalite. *J. Colloid Interface Sci.*, 162(1):182–188, 1994.

[213] Jesse G. McDaniel, Song Li, Emmanouil Tylianakis, Randall Q. Snurr, and J. R. Schmidt. Evaluation of force field performance for high-throughput screening of gas uptake in metal-organic frameworks. *J. Phys. Chem. C*, page null, 2015.

[214] Mohamed Eddaoudi, Jaheon Kim, Nathaniel Rosi, David Vodak, Joseph Wachter, Michael O'Keeffe, and Omar M. Yaghi. Systematic design of pore size and functionality in isoreticular mofs and their application in methane storage. *Science*, 295(5554):469–472, 2002.

[215] Wolfgang Schmidt, Ursula Wilczok, Claudia Weidenthaler, Olaf Medenbach, Richard Goddard, Gernot Buth, and Alexander Cepak. Preparation and morphology of pyramidal mfi single-crystal segments. *J Phys. Chem. B*, 111(48):13538–13543, 2007.

[216] Li-Chiang Lin, Kyuho Lee, Laura Gagliardi, Jeffrey B Neaton, and Berend Smit. Force-field development from electronic structure calculations with periodic boundary conditions: Applications to gaseous adsorption and transport in metal–organic frameworks. *J. Chem. Theory Comput.*, 10(4):1477–1488, 2014.

[217] Olaf Delgado-Friedrichs and Michael O'Keeffe. Crystal nets as graphs: Terminology and definitions. *Journal of Solid State Chemistry*, 178(8):2480 – 2485, 2005. Reticular Chemistry: Design, Synthesis, Properties and Applications of Metal-organic Polyhedra and Frameworks.

[218] A. Bondi. Van der waals volumes and radii. *J. Phys. Chem.*, 68(3):441–451, 1964.

[219] R. Scott Rowland and Robin Taylor. Intermolecular nonbonded contact distances in organic crystal structures: Comparison with distances expected from van der waals radii. *J. Phys. Chem.*, 100(18):7384–7391, 1996.

[220] Chris H. Rycroft. Voro++: A three-dimensional voronoi cell library in c++. *Chaos*, 19(4):041111, 2009.

[221] Hailian Li, Aaron Laine, M. O'Keeffe, and O. M. Yaghi. Supertetrahedral sulfide crystals with giant cavities and channels. *Science*, 283(5405):1145–1147, 1999.

[222] Ching-Chung Kuo, Fred Glover, and Krishna S Dhir. Analyzing and modeling the maximum diversity problem by zero-one programming*. *Decision Sciences*, 24(6):1171–1185, 1993.

[223] Abraham Duarte and Rafael Martí. Tabu search and grasp for the maximum diversity problem. *European Journal of Operational Research*, 178(1):71–84, 2007.

[224] Daniel Cosmin Porumbel, Jin-Kao Hao, and Fred Glover. A simple and effective algorithm for the maxmin diversity problem. *Annals of Operations Research*, 186(1):275–293, 2011.

[225] Richard Luis Martin, Berend Smit, and Maciej Haranczyk. Addressing challenges of identifying geometrically diverse sets of crystalline porous materials. *Journal of chemical information and modeling*, 52(2):308–318, 2011.

[226] Leo Breiman, Jerome Friedman, Charles J Stone, and Richard A Olshen. *Classification and Regression Trees*. CRC Press, 1984.

[227] Antonio Criminisi, Jamie Shotton, and Ender Konukoglu. Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. *Foundations and Trends® in Computer Graphics and Vision*, 7(2–3):81–227, 2012.

[228] Laurent Hyafil and Ronald L. Rivest. Constructing optimal binary decision trees is np-complete. *Information Processing Letters*, 5(1):15 – 17, 1976.

[229] Leo Breiman. Bagging predictors. *Mach. Learn.*, 24(2):123–140, 1996.

[230] Thomas G Dietterich. An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Mach. Learn.*, 40(2):139–157, 2000.

[231] Tin Kam Ho. The random subspace method for constructing decision forests. *IEEE Trans. Pattern Anal. Mach. Intell*, 20(8):832–844, 1998.

[232] Carolin Strobl, Anne-Laure Boulesteix, Thomas Kneib, Thomas Augustin, and Achim Zeileis. Conditional variable importance for random forests. *BMC Bioinform.*, 9(1):307, 2008.

[233] Ngamta Thamwattana, Duangkamon Baowan, and Barry J Cox. Modelling bovine serum albumin inside carbon nanotubes. *RSC Advances*, 3(45):23482–23488, 2013.

[234] Debasis Banerjee, Zhijuan Zhang, Anna M Plonka, Jing Li, and John B Parise. A calcium coordination framework having permanent porosity and high co_2/n_2 selectivity. *Cryst. Growth Des.*, 12(5):2162–2165, 2012.

[235] Emily R Cooper, Christopher D Andrews, Paul S Wheatley, Paul B Webb, Philip Wormald, and Russell E Morris. Ionic liquids and eutectic mixtures as solvent and template in synthesis of zeolite analogues. *Nature*, 430(7003):1012–1016, 2004.

[236] Maciej Haranczyk, Li-Chiang Lin, Kyuho Lee, Richard L Martin, Jeffrey B Neaton, and Berend Smit. Methane storage capabilities of diamond analogues. *Phys. Chem. Chem. Phys.*, 15(48):20937–20942, 2013.

[237] Stephen T Wilson, Brent M Lok, Celeste A Messina, Thomas R Cannan, and Edith M Flanigen. Aluminophosphate molecular sieves: A new class of microporous crystalline inorganic solids. *J. Am. Chem. Soc.*, 104(4):1146–1147, 1982.

[238] Xiaowei Song, Jiyang Li, Yanan Guo, Qinhe Pan, Lin Gan, Jihong Yu, and Ruren Xu. Syntheses and characterizations of transition-metal-substituted aluminophosphate molecular sieves —$(c_3n_2h_5)_8$—$[m_8al_{16}p_{24}o_{96}]$ (m = co, mn, zn) with zeotype lau topology. *Inorg. Chem.*, 48(1):198–203, 2009.

[239] D.W. Scott. *Multivariate Density Estimation: Theory, Practice, and Visualization.* John Wiley & Sons, New York, Chichester, 1992.

[240] Cory M Simon, Berend Smit, and Maciej Haranczyk. pyiast: Ideal adsorbed solution theory (iast) python package. *Computer Physics Communications*, 200:364–380, 2016.

[241] Orhan Talu. Needs, status, techniques and problems with binary gas adsorption experiments. *Adv. Colloid Interface Sci.*, 76:227–269, 1998.

[242] Trevor A Makal, Jian-Rong Li, Weigang Lu, and Hong-Cai Zhou. Methane storage in advanced porous materials. *Chemical Society Reviews*, 41(23):7761–7779, 2012.

[243] Hongda Zhang, Pravas Deria, Omar K Farha, Joseph T Hupp, and Randall Q Snurr. A thermodynamic tank model for studying the effect of higher hydrocarbons on natural gas storage in metal–organic frameworks. *Energy Environ. Sci.*, 8(5):1501–1510, 2015.

[244] Nathaniel L Rosi, Juergen Eckert, Mohamed Eddaoudi, David T Vodak, Jaheon Kim, Michael O'Keeffe, and Omar M Yaghi. Hydrogen storage in microporous metal-organic frameworks. *Science*, 300(5622):1127–1129, 2003.

[245] Jarad A Mason, Thomas M McDonald, Tae-Hyun Bae, Jonathan E Bachman, Kenji Sumida, Justin J Dutton, Steven S Kaye, and Jeffrey R Long. Application of a high-throughput analyzer in evaluating solid adsorbents for post-combustion carbon capture via multicomponent adsorption of CO_2, N_2, and H_2O. *J. Am. Chem. Soc.*, 137(14):4787–4803, 2015.

[246] Orhan Talu. Measurement and analysis of mixture adsorption equilibrium in porous solids. *Chem. Ing. Tech.*, 83(1-2):67–82, 2011.

[247] Shivaji Sircar. Basic research needs for design of adsorptive gas separation processes. *Ind. Eng. Chem. Res.*, 45(16):5435–5448, 2006.

[248] AL Myers and John M Prausnitz. Thermodynamics of mixed-gas adsorption. *AIChE J.*, 11(1):121–127, 1965.

[249] Krista S Walton and David S Sholl. Predicting multicomponent adsorption: 50 years of the ideal adsorbed solution theory. *AIChE J.*, 61(9):2757–2762, 2015.

[250] Naomi F Cessford, Nigel A Seaton, and Tina Duren. Evaluation of ideal adsorbed solution theory as a tool for the design of metal–organic framework materials. *Ind. Eng. Chem. Res.*, 51(13):4911–4921, 2012.

[251] Qingyuan Yang and Chongli Zhong. Molecular simulation of carbon dioxide/methane/hydrogen mixture adsorption in metal-organic frameworks. *J. Phys. Chem. B*, 110(36):17776–17783, 2006.

[252] Ravichandar Babarao, Zhongqiao Hu, Jianwen Jiang, Shaji Chempath, and Stanley I Sandler. Storage and separation of CO_2 and CH_4 in silicalite, C168 schwarzite, and IRMOF-1: a comparative study from Monte Carlo simulation. *Langmuir*, 23(2):659–666, 2007.

[253] Seda Keskin, Jinchen Liu, J Karl Johnson, and David S Sholl. Testing the accuracy of correlations for multicomponent mass transport of adsorbed gases in metal- organic frameworks: Diffusion of H_2/CH_4 mixtures in CuBTC. *Langmuir*, 24(15):8254–8261, 2008.

[254] Youn-Sang Bae, Karen L Mulfort, Houston Frost, Patrick Ryan, Sudeep Punnathanam, Linda J Broadbelt, Joseph T Hupp, and Randall Q Snurr. Separation of CO_2 from CH_4 using mixed-ligand metal- organic frameworks. *Langmuir*, 24(16):8592–8598, 2008.

[255] Bei Liu, Qingyuan Yang, Chunyu Xue, Chongli Zhong, Biaohua Chen, and Berend Smit. Enhanced adsorption selectivity of hydrogen/methane mixtures in metal- organic frameworks with interpenetration: A molecular simulation study. *J. Phys. Chem. C*, 112(26):9854–9860, 2008.

[256] David W Hand, Scott Loper, Metin Ari, and John C Crittenden. Prediction of multicomponent adsorption equilibria using ideal adsorbed solution theory. *Environ. Sci. Technol.*, 19(11):1037–1043, 1985.

[257] R Krishna, S Calero, and B Smit. Investigation of entropy effects during sorption of mixtures of alkanes in MFI zeolite. *Chem. Eng. J.*, 88(1):81–94, 2002.

[258] J. Rother and T. Fieback. Multicomponent adsorption measurements on activated carbon, zeolite molecular sieve and metal–organic framework. *Adsorption*, 19(5):1065–1074, 2013.

[259] Stephen Brunauer, Paul Hugh Emmett, and Edward Teller. Adsorption of gases in multimolecular layers. *J. Am. Chem. Soc.*, 60(2):309–319, 1938.

[260] Cory M Simon, Jihan Kim, Li-Chiang Lin, Richard L Martin, Maciej Haranczyk, and Berend Smit. Optimizing nanoporous materials for gas storage. *PCCP*, 16(12):5499–5513, 2014.

[261] Haibin Chen and David S Sholl. Examining the accuracy of ideal adsorbed solution theory without curve-fitting using transition matrix Monte Carlo simulations. *Langmuir*, 23(11):6431–6437, 2007.

[262] Aaron W Thornton, David Alan Winkler, Ming S Liu, Maciej Haranczyk, and Danielle F Kennedy. Towards computational design of zeolite catalysts for CO_2 reduction. *R. Soc. Chem. Adv.*, 5(55):44361–44370, 2015.

[263] CJ Radke and JM Prausnitz. Thermodynamics of multi-solute adsorption from dilute liquid solutions. *AIChE J.*, 18(4):761–768, 1972.

[264] Abhijit Tarafder and Marco Mazzotti. A method for deriving explicit binary isotherms obeying the ideal adsorbed solution theory. *Chem. Eng. Technol.*, 35(1):102–108, 2012.

[265] O. Talu and A. L. Myers. Rigorous thermodynamic treatment of gas adsorption. *AIChE J.*, 34(11):1887–1893, 1988.

[266] József Tóth. Some consequences of the application of incorrect gas/solid adsorption isotherm equations. *J. Colloid Interface Sci.*, 185(1):228–235, 1997.

[267] K. Y. Foo and B. H. Hameed. Insights into the modeling of adsorption isotherm systems. *Chem. Eng. J.*, 156(1):2–10, 2010.

[268] H. M. F. Freundlich. Over the adsorption in solution. *J. Phys. Chem.*, pages 385–471, 1906.

[269] Terrell L Hill. *An introduction to statistical thermodynamics*. Dover Publications, 1986.

[270] Christoph Hinz. Description of sorption data with isotherm equations. *Geoderma*, 99(3):225–243, 2001.

[271] Charles H Giles, David Smith, and Alan Huitson. A general treatment and classification of the solute adsorption isotherm. I. Theoretical. *J. Colloid Interface Sci.*, 47(3):755–765, 1974.

[272] V. Pyzhev M.I. Tempkin. Kinetics of ammonia synthesis on promoted iron catalyst. *Acta Phys. Chim. USSR*, 12:327–356, 1940.

[273] M Douglas LeVan and Theodore Vermeulen. Binary Langmuir and Freundlich isotherms for ideal adsorbed solutions. *J. Phys. Chem.*, 85(22):3247–3250, 1981.

[274] Milica Ilić, Dietrich Flockerzi, and Andreas Seidel-Morgenstern. A thermodynamically consistent explicit competitive adsorption isotherm model based on second-order single component behaviour. *J. Chromatogr., A*, 1217(14):2132–2137, 2010.

[275] Fabrice Gritti and Georges Guiochon. Band splitting in overloaded isocratic elution chromatography: II. New competitive adsorption isotherms. *J. Chromatogr., A*, 1008(1):23–41, 2003.

[276] Uri M Ascher and Chen Greif. *A First Course on Numerical Methods*, volume 7. SIAM, 2011.

[277] Héctor Octavio Rubiera Landa, Dietrich Flockerzi, and Andreas Seidel-Morgenstern. A method for efficiently solving the IAST equations with an application to adsorber dynamics. *AIChE J.*, 59(4):1263–1277, 2013.

[278] Hailian Li, Mohamed Eddaoudi, Michael O'Keeffe, and Omar M Yaghi. Design and synthesis of an exceptionally stable and highly porous metal-organic framework. *Nature*, 402(6759):276–279, 1999.

[279] Marcus G Martin and J Ilja Siepmann. Transferable potentials for phase equilibria. 1. united-atom description of n-alkanes. *J. Phys. Chem. B*, 102(14):2569–2577, 1998.

[280] M. Harper. Python-ternary: A python library for ternary plots. 2011-2015. `https://github.com/marcharper/python-ternary`.

[281] E Costa, JL Sotelo, G Calleja, and C Marron. Adsorption of binary and ternary hydrocarbon gas mixtures on activated carbon: experimental determination and theoretical prediction of the ternary equilibrium data. *AIChE J.*, 27(1):5–12, 1981.

[282] R Krishna and JM Van Baten. Segregation effects in adsorption of CO_2-containing mixtures and their consequences for separation selectivities in cage-type zeolites. *Sep. Purif. Technol.*, 61(3):414–423, 2008.

[283] AL Myers. Activity coefficients of mixtures adsorbed on heterogeneous surfaces. *AIChE J.*, 29(4):691–693, 1983.

[284] Y. Wang and M. D. LeVan. Adsorption equilibrium of binary mixtures of carbon dioxide and water vapor on zeolites 5a and 13x. *J. Chem. Eng. Data*, 55(9):3189–3195, 2010.

[285] Shivaji Sircar. Influence of adsorbate size and adsorbent heterogeneity of IAST. *AIChE J.*, 41(5):1135–1145, 1995.

[286] DP Valenzuela, Alan L Myers, Orhan Talu, and I Zwiebel. Adsorption of gas mixtures: effect of energetic heterogeneity. *AIChE J.*, 34(3):397–402, 1988.

[287] Manohar Murthi and Randall Q Snurr. Effects of molecular siting and adsorbent heterogeneity on the ideality of adsorption equilibria. *Langmuir*, 20(6):2489–2497, 2004.

[288] Jude Dunne and Alan L Myers. Adsorption of gas mixtures in micropores: effect of difference in size of adsorbate molecules. *Chem. Eng. Sci.*, 49(17):2941–2951, 1994.

[289] Orhan Talu and Imre Zwiebel. Multicomponent adsorption equilibria of nonideal mixtures. *AIChE J.*, 32(8):1263–1276, 1986.

[290] Xijun Hu and Duong D Do. Comparing various multicomponent adsorption equilibrium models. *AIChE J.*, 41(6):1585–1592, 1995.

[291] Joseph A Swisher, Li-Chiang Lin, Jihan Kim, and Berend Smit. Evaluating mixture adsorption models using molecular simulation. *AIChE J.*, 59(8):3054–3064, 2013.

[292] François-Xavier Coudert, Caroline Mellot-Draznieks, Alain H Fuchs, and Anne Boutin. Prediction of breathing and gate-opening transitions upon binary mixture adsorption in metal- organic frameworks. *J. Am. Chem. Soc.*, 131(32):11329–11331, 2009.

[293] Ashlee J Howarth, Yangyang Liu, Peng Li, Zhanyong Li, Timothy C Wang, Joseph T Hupp, and Omar K Farha. Chemical, thermal and mechanical stabilities of metal–organic frameworks. *Nat Rev Mater*, 1:15018, 2016.

[294] Austin C Kizzie, Antek G Wong-Foy, and Adam J Matzger. Effect of humidity on the performance of microporous coordination polymers as adsorbents for co2 capture. *Langmuir*, 27(10):6368–6373, 2011.

[295] Haohan Wu, Qihan Gong, David H Olson, and Jing Li. Commensurate adsorption of hydrocarbons and alcohols in microporous metal organic frameworks. *Chemical reviews*, 112(2):836–868, 2012.

[296] Debasis Banerjee, Hao Wang, Qihan Gong, Anna M Plonka, Jacek Jagiello, Haohan Wu, William R Woerner, Thomas J Emge, David H Olson, John B Parise, et al. Direct structural evidence of commensurate-to-incommensurate transition of hydrocarbon adsorption in a microporous metal organic framework. *Chemical Science*, 7(1):759–765, 2016.

[297] Anna M Plonka, Xianyin Chen, Hao Wang, Rajamani Krishna, Xinglong Dong, Debasis Banerjee, William R Woerner, Yu Han, Jing Li, and John B Parise. Light hydrocarbon adsorption mechanisms in two calcium-based microporous metal organic frameworks. *Chemistry of Materials*, 28(6):1636–1646, 2016.

[298] Marielle Pinheiro, Richard L Martin, Chris H Rycroft, and Maciej Haranczyk. High accuracy geometric analysis of crystalline porous materials. *CrystEngComm*, 15(37):7531–7538, 2013.

[299] Jarad A Mason, Julia Oktawiec, Mercedes K Taylor, Matthew R Hudson, Julien Rodriguez, Jonathan E Bachman, Miguel I Gonzalez, Antonio Cervellino, Antonietta Guagliardi, Craig M Brown, et al. Methane storage in flexible metal-organic frameworks with intrinsic thermal management. *Nature*, 527(7578):357–361, 2015.

Appendix

Funding acknowledgements

Here, I would like to acknowledge my funding.

My work in Chapters 2, 3, and 4 was supported by the U.S. Department of Energy, Office of Basic Energy Sciences, Division of Chemical Sciences, Geosciences and Biosciences under Award DE-FG02-12ER16362. This research used resources of the National Energy Research Scientific Computing Center, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

My work in Chapters 5, 6, and 7 was supported by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists, Office of Science Graduate Student Research (SCGSR) program. The SCGSR program is administered by the Oak Ridge Institute for Science and Education for the DOE under contract number DE-AC05-06OR23100. M.H. was supported by the U.S. Department of Energy, Office of Basic Energy Sciences, Division of Chemical Sciences, Geosciences and Biosciences, under Award DE-FG02-12ER16362.