Computational Design and Analysis of MOF-based Electronic Noses for Disease Detection by Breath

by

Brian A. Day

Bachelor of Science, Chemical Engineering, Lehigh University, 2016

Submitted to the Graduate Faculty of the

Swanson School of Engineering in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2022

UNIVERSITY OF PITTSBURGH

SWANSON SCHOOL OF ENGINEERING

This dissertation was presented

by

Brian A. Day

It was defended on

July 26, 2022

and approved by

Christopher E. Wilmer, Ph.D., Associate Professor,
Department of Chemical and Petroleum Engineering

Eric Beckman, Ph.D., Distinguished Service Professor,
Department of Chemical and Petroleum Engineering

Susan Fullerton, Ph.D., Associate Professor,
Department of Chemical and Petroleum Engineering

Alexander Star, Ph.D., Associate Professor,
Department of Chemistry

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

Computational Design and Analysis of MOF-based Electronic Noses for Disease Detection by Breath

Brian A. Day, PhD

University of Pittsburgh, 2022

Despite the existence of sophisticated analytical gas sensing technologies like gas chromatography – mass spectrometry (GCMS), there are many applications for which sufficient gas sensors are lacking, such as environmental monitoring and disease detection by breath, where there is a need for low-cost, portable devices with high sensitivities and fast response times. A promising strategy for achieving these features is the development of gas sensor arrays, better known as electronic noses, in which multiple sensing elements are used cooperatively to improve detection capabilities.

This dissertation describes my research on the use of metal-organic frameworks (MOFs) as the sensing materials for electronic noses. MOFs are a novel class of nanoporous crystalline materials with high internal surface areas and a large degree of chemical and structural diversity, resulting in similarly impressive and diverse gas adsorption properties. Prior to the work of this lab, there had been few investigations into MOF-based sensor arrays, and they were limited to experimental trial-and-error approaches. In response, our lab pioneered the use computational approaches for high-throughput screening of MOFs and rational design of sensor arrays, resulting in significant improvements to sensing performance.

The focus of this dissertation is on strategies for further improving the design and analysis of MOF-based electronic noses, specifically for the detection of trace gas species in complex gas mixtures for disease detection by breath. We first examined the ternary gas mixtures of carbon

dioxide, nitrogen, and oxygen, the majority species of breath, which provided a valuable starting point for breath analysis and highlighted limitations of our existing method. In order to address the scalability challenges related to the combinatorics of multicomponent mixtures, we developed a novel coefficient-based method for evaluating the adsorption of trace gas species in complex mixtures, as well as a corresponding algorithm for signal analysis, which we used to study five-component gas mixtures relating to the detection of chronic kidney disease by breath. Finally, we developed a strategy for improving the sensitivity and selectivity of arrays, and increasing overall information content, by sampling the gas mixture at various pressures, with low pressures enabling the desaturation of sensors by strongly adsorbing gases, and high pressures enabling us to increase mass uptake for weakly adsorbing gases.

Table of Contents

List of Tables

List of Figures

Preface

There are so many people whom I would like to thank for helping me to where I am today. First and foremost, my advisor, Dr. Chris Wilmer. Chris - you have been patient and supportive, offering me the freedom to explore many avenues of research. You could put things into perspective when I lost my sense of direction, or everything felt overwhelming. And you supported my entrepreneurship endeavors, which I never would have begun without your encouragement. There is so much to thank you for, but suffice to say, I am grateful for your guidance.

Thank you also to my committee members for your time and support. You have all truly been mentors to me at one time or another, and I know I am a better scientist because of you. I would also like to thank my fellow Wilmer lab members: Jenna, Kutay, Alec, Paul, Meiirbek, and Becca. It has been a pleasure to work alongside you all and I appreciate your help and kindness.

Thank you to my parents and the Housers for your endless support and guidance. Thank you to Andy, and our dog Charlie, who have been with me through almost this entire journey and who have both listened to countless ramblings about practically every thought to enter my head. And, lastly, thank you to all of my friends in Pittsburgh who have made this city a great place to call home.

The goal of this work is to develop a universal gas sensing array, called an electronic nose, specifically targeting applications in breath-based disease detection. It was started by Jenna Gustafson and Chris Wilmer, and I am very lucky to have been handed the torch. Building on their work, I developed methods of addressing the scalability challenges associated with sensing highly complex, multicomponent gas mixtures, including a change in the approach to molecular simulations and a corresponding change to the algorithm we use to predict compositions. I also

developed a new strategy for increasing the information content of arrays and improving the sensitivity and selectivity of the device by sampling a gas mixture at various pressures.

Furthermore, I have successfully presented this research in numerous business plan competitions, winning the Big Idea Blast, the Big Idea Blitz, the Randall Family Big Idea Competition, the Product Data and Management Alliance competition, the TiE Pittsburgh Pitch Fest, and was a finalist in the TiE Global pitch competition. Throughout these competitions, I have developed and refined a business model for bringing breath-based disease detection into physicians' offices.

This work includes two peer-reviewed articles, published in *Sensors* and *ACS Sensors*, and another manuscript to be submitted. Additionally, I have worked on several collaborations not mentioned in this dissertation. These include peer-reviewed publications in the *Journal of Physical Chemistry C*, *ACS Sensors*, and another manuscript to be submitted to a peer-reviewed journal in the near future. Finally, this work required a fair amount of software development, most of which is available on GitHub with open-source licenses.

1.0 A Brief Introduction to Gas Sensing Technologies

1.1 Existing Technologies & Limitations

There are numerous gas sensing technologies in existence today including sophisticated analytical instruments like gas chromatography - mass spectrometry (GCMS) and flame ionization detectors; simple mass-based sensors, like surface-acoustic wave (SAW) devices and quartz-crystal microbalances (QCM); and chemiresistive sensors such as metal-oxide semiconducting field-effect transistors (MOSFET), to name just a few.[1,2] However, despite this breadth of technologies, there are many applications for which all of these sensors fall short, as no single technology yet encompasses portability, sensitivity, selectivity, and rapid response times. Furthermore, many of these technologies require special operating conditions, such as elevated temperatures, vacuum chambers, complex calibration procedures, or long recovery periods.[1,3]

One such application is the non-invasive screening and detection of diseases by breath or body odor.[4–8] There are numerous diseases for which the breath biomarkers are well documented (e.g., kidney disease, liver disease, lung cancer), however, because of the limitations of the existing technologies, in particular cost and portability, health systems still rely on other sample types (e.g., blood and urine) and diagnostic techniques.[4,9–17] Hence, there is a need for next generation gas sensors for breath-based diagnostics.

There are also numerous safety applications where a next generation gas sensor would be useful, such as the detection of methane leaks in the natural gas industry, the detection of toxic chemicals in industrial applications, and the detection of chemical warfare agents for defense applications.[1,18,19] Portability is again one the primary barriers to employing existing

technologies, however rapid sensor response times are also critical in the above applications to alert those endangered by the conditions as quickly as possible.

Additional application areas include the food and beverage industry, the pharmaceutical industry, and environmental monitoring. Across all these industries, there is a common desire for the next generation of gas sensors to encompass the following features: low cost, high portability, high sensitivity/selectivity, rapid response, and simple operating conditions. Presently, the most promising strategy for achieving these features in a single device is to build a gas sensor array, commonly referred to as an electronic nose.

1.2 Exhaled Breath Analysis

One of the most exciting application areas for new gas sensing technologies is exhaled breath analysis.[4,20,21] Human breath is a complex and information rich gas mixture. Although breath is composed mostly of nitrogen, oxygen, carbon dioxide, and water, over the past few decades, more than 1,000 different volatile organic compounds (VOCs) have been identified in human breath, ranging in concentration from several parts per million (ppm) down to a few parts per trillion (ppt). Some of these gases originate from processes in the body (endogenic), whereas others originate from surrounding air (exogenic).[21] Notably, the composition of VOCs in breath can vary widely from person to person in both which gases are present and their concentration. Only a subset of VOCs is actually common to all humans, including gases like isoprene, acetone, ethane, and methanol, which are all products of core metabolic processes.

The detection of VOCs in breath for disease detection actually has a long history, stretching back to the times of Hippocrates. Doctors in ancient Greece knew to be on alert for the sweet and

fruity odor of acetone for patients with diabetes, fishy scents for liver disease, urine odors for failing kidneys, and a putrid stench for a lung abscess (though perhaps without these modern diagnoses).

Modern breath analysis started in the 1970s when researchers using gas chromatography identified more that 200 compounds in human breath. Then, along with advances in analytical techniques for gas sensing, researchers began to focus on the correlation between compounds in exhaled breath and human health, resulting in an extensive body of literature of exhaled breath analysis for clinical applications. Examples of known biomarkers include alkanes for oxidative stress; nitric oxide and hydrogen peroxide for lung diseases such as asthma, chronic obstructive pulmonary disease, and cystic fibrosis; isoprene for lung cancer; acetone for diabetes; and hydrogen, methane, and hydrogen sulfide for gastrointestinal diseases.[10,22,11,12]

But despite the ancient and modern history of breath analysis, breath is seldom used in medicine, largely due to a technology gap. There are only a few breath tests on the market today, including a three-compound test for small intestinal bacterial overload (SIBO), a fractional exhaled nitric oxide (FeNO) test for airway inflammation, and most famously, breathalyzer tests for the indirect assessment of blood alcohol content (BAC).[11] For many other diseases, multiple compounds need to be measured simultaneously, and while GCMS is impressive technology capable of such an analysis, it is normally too expensive and low throughput for large patient populations, and no other gas sensors yet exist which can cheaply and quickly analyze gas mixtures as complex as breath.

Instead, most clinical tests use biological samples, such as tissues, fluids (blood, urine, sputum), feces, or alternatively, imaging tests such as x-rays and CT scans. Even so, because breath

testing is completely non-invasive, there is a strong desire to develop breath tests for disease screening, detection, and monitoring applications.

One of the most promising uses for exhaled breath analysis is lung cancer screening. Lung cancer is the second most common and single deadliest cancer in the United States, responsible for around 150,000 deaths per year, more deaths than breast, colon, and prostate cancer combined. This is due, in part, to the fact that lung cancer is most often caught in its late stages, when survival rates are at their lowest.[9,23] Currently, the United States Preventative Services Task Force (USPSTF) recommends lung cancer screening for those over the age of 50 with 20 pack-year smoking histories, which amounts to about 14.5 million people in the United States. However, screening rates for lung cancer are exceptionally low, at around only 5%, contributing to the lack of early diagnosis.[24] Furthermore, the only recommended screening test for lung cancer at present in low dose computed Tomography (LDCT), but this test has an incredibly high false positive rate of 94.5%, causing undo strain on downstream resources. A simple point-of-care breath test for lung cancer could dramatically increase screening rates and consequently, increase the likelihood that lung cancer is caught early, when survival rates are better, and treatment is cheaper.

1.3 Electronic Noses

It is widely known that dogs have an incredible sense of smell. Their noses contain around 1,300 unique olfactory receptor types, each of which is tuned to the detection of different types of molecules.[25] In total, their noses contain well over 100 million olfactory receptor cells, enabling them to detect molecules at parts per trillion concentrations.[26–28] By comparison, a human nose

contains around 900 unique olfactory receptor types, and only 5 million olfactory receptor cells.[25,29] Because of their superior sense of smell, dogs have been employed for numerous tasks, first as hunting dogs, and more recently, for bomb-sniffing, drug-sniffing, disease detection, and disease management.[26,28,30] While dogs have excelled in these areas, the training and management of a team of dogs is difficult and expensive, leading to both practical and economic barriers to their use. For example, a diabetic alert dog costs around $20,000 dollars.[28] And often, when teams of dogs are needed, such as in threat management or disaster response, there are simply not enough trained dogs for the task.[26,27,31] Instead, what if we had the ability to fabricate a device mimicking a dog's nose? Such a device, often called an electronic nose, would be an incredible technological advancement.

Just like a biological nose, an electronic nose can be thought of as an array of inorganic 'olfactory receptors' (i.e., sensing elements), each of which complements the others. In principle, one single sensing element could be used to detect several different gases simultaneously, so long as it has a statistically unique response to all possible combinations of those gases. However, in practice, this is nearly impossible, hence the motivation for building a sensing array. Moreover, for many applications, one may only need to detect a small number of gases, but an electronic nose must be able to discriminate the gases of interest from other interfering compounds in the sample. Consider again exhaled breath analysis for disease detection; a single disease may have anywhere from one to a dozen known biomarkers, but a single breath sample can contain over 700 different VOCs. Detecting even a small set of biomarkers becomes very complicated, as the device must avoid mistaking any of the hundreds of other gases for the clinically relevant compounds. Thus, an electronic nose for such an application may actually need many more sensors than the number of biomarkers in order to minimize the influence of the other compounds.

This leads to into the challenge of materials selection. There are many different types of materials which could be employed in an electronic nose. Metal-organic frameworks (MOFs), for example, which are the focus of this dissertation, are an incredibly large class of materials. To date, there have been over 90,000 MOFs synthesized and over 500,000 which have been proposed.[32] Clearly, selecting the optimal combination of sensing materials for a 700-component gas mixture from hundreds of thousands of materials in non-trivial.[33–37] Going a step further, if one desires to build a truly universal gas sensor, it is likely that hundreds of different sensing materials will be necessary, underscoring the fact that the design of an electronic nose is as much a big data problem benefiting from computational approaches as it is an engineering challenge.

Nevertheless, multi-sensor arrays have existed for decades. The first known multi-sensor array was developed in 1982, with the term 'electronic nose' later being coined by Gardner and Bartlett in 1988, defined as "an instrument which comprises an array of electronic chemical sensors with partial specificity and appropriate pattern recognition system, capable of recognizing simple or complex odors".[19,38–41] Since then, several more electronic noses have been developed, with some having even been commercialized.[19] Even so, these electronic noses have not been widely employed in the previously named application areas, raising the question, why? Generally speaking, the materials and their response to different gases were too similar, and thus there was a diminishing benefit to additional sensing elements such that the desired sensitivity and selectivity could not be achieved. In order to alleviate this problem of diminishing returns, electronic noses must employ sensing mechanisms and materials with a large diversity of responses.[19,42]

Accordingly, electronic nose research can be sectioned as follows: sensing mechanisms, sensing materials, array design, and data analysis.

1.3.1 Sensing Mechanisms

As previously mentioned, for many applications for which electronic noses are being developed, the resulting device must be cheap and portable with high sensitivity/selectivity, and rapid response times. These features, however, limit the sensing mechanisms which one can use. In our opinion, the most promising sensing mechanism is a simple mass-based sensor. Mass-based sensing devices can use a wide variety of sensing materials and because of the simplicity of the transduction mechanism, are sensitive to most gases. Furthermore, they are cheap and portable, require low power input, can be operated at room temperature, and have good mass sensitivity, but can be limited by temperature sensitivity and a poor signal-to-noise ratio.[43–45] Fortunately, these limitations can be mitigated by a well-designed array. In our work, we envision an electronic nose using technologies such as QCMs and SAW devices, as is shown in Figure 1.

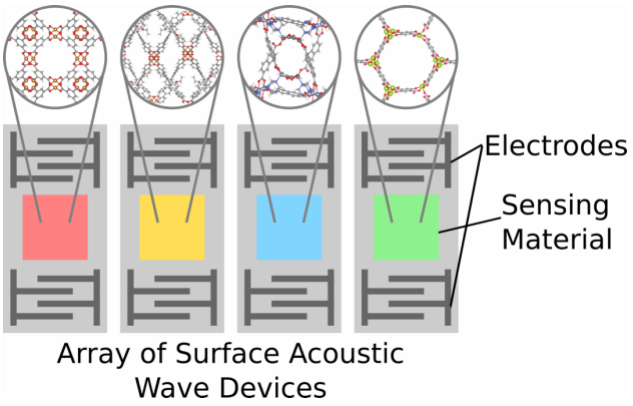

Figure 1. Sample four-element array of surface acoustic wave devices with MOFs the sensing materials.

The transduction mechanism for these two sensors are frequency shifts in a mechanical wave which results from a change in mass of the sensing layer. This frequency shift is then converted into an electrical signal, enabling digital reading of the sensors. The relationship

between the change in mass, frequency shift, and electrical signal is governed by the Sauerbrey Equation (Equation 1), where $\frac{\Delta m}{A}$ is the mass change per area, c_m is a mechanical coefficient, f_o is a resonant frequency used as a reference state, and f is the output frequency.[46]

$$\frac{\Delta m}{A} = \frac{c_m(f - f_o)}{2f^2} \tag{1-1}$$

Thus, for initial computational exploration of these device, one can consider the signal from these devices as effectively a mass measurement.[43–45]

The change in mass is which results in the above signal change is due to the adsorption of gas molecules onto and into the sensing layer upon exposure to a gas mixture. This change in mass is a function of both the sensing material and the composition of the surrounding gas. These devices exhibit very impressive sensitivity, in general yielding nanogram precision, however the sensitivity can be affected by various features of the specific device, as well as the sensing material.

Although we will be focusing only on mass-based sensors, it should be noted that electronic noses can, in principle, use multiple sensing mechanisms concurrently, and there may even be advantages to doing so.[19] However, it is often convenient to limit oneself to a single sensing mechanism, mainly to limit the complexity of screening materials and data analysis.

1.3.2 Sensing Materials

Much research has been done in trying to develop materials with the correct properties for the detection of specific molecules, whether it be through the functionalization of carbon nanotubes, polymer engineering, etc.[47–50] However, as highlighted by the limited use of previous electronic nose devices, precisely engineered materials provide diminishing returns, and so our belief is that improved detection is better addressed by intelligently selecting from a set of

diverse materials, especially when the goal is to detect many different types of molecules with a single device. This belief assumes two things, however; firstly, that there exists a large enough set of materials to choose from, and secondly, that the properties of that materials set are diverse enough to complement for each other. Fortunately, there are several classes of materials to choose from for mass-based sensing, including polymers, zeolites, and metal-organic frameworks (MOFs).[51]

Polymers have been an attractive material for electronic noses for some time, specifically conductive polymers for chemiresistive gas sensors.[52] The exhibit high sensitives, short response times, good mechanical properties, and are able to operate at room temperature. They also have the added bonus of being easily synthesizable. Unfortunately, they have a serious drawback in that they are highly susceptible to humidity, and that humidity can dramatically alter their gas adsorption properties. Furthermore, they are typically less sensitive than metal-oxide films, the other leading chemiresistive sensing material.[52,53]

As for mass-based sensing, molecularly imprinted polymers have been gaining attention for use in aqueous environments, and polymers-carbon nanotube composites have been gaining attention for use in gas sensing. However, the lack of diversity of polymer-CNT composites is potentially limiting for larger arrays, and the non-crystalline nature of these materials makes them hard to study computationally.

Zeolites are a class of microporous crystalline materials made up of interconnected aluminosilicate units, notable for their uniform pores/channels and high internal surface areas.[54–56] The size of the pores/channels of zeolites are easily tuned by changing the number of bonded units.[57] This feature of zeolites makes them useful as molecular sieves, materials which filter molecules according to their size. Their chemical properties can also be tuned to some degree using

various post-synthetic modification techniques.[58] Collectively, these properties of zeolites lead to impressive gas adsorption properties, hence their appeal for gas sensing applications.[56,59] However, even with these tunable parameters and a large number of known hypothetical zeolites (>250,000), they tend to be relatively similar, and it is uncertain whether or not they are diverse enough for use in an electronic nose.[57]

MOFs, on the other hand, combine the best features of polymers and zeolites. Like polymers, MOFs are made by combining structural building units (SBUs), but instead of monomers, the SBUs for MOFs are metal nodes and organic linkers. The resulting structures, like zeolites, are nanoporous crystalline materials with high internal surface areas and again, impressive gas adsorption properties. MOFs also exhibit incredible chemical diversity, owing the sheer variety of metal nodes and organic linkers which can be used. In fact, over 90,000 different MOFs have been synthesized, and there are more than 500,000 known hypothetical MOFs.[32,60] MOFs (and zeolites) also have another convenient feature in that, because of their crystalline nature, their gas adsorption properties are easily examined with well-established computational techniques, enabling large-scale screening. We have chosen to focus on MOFs as the primary sensing material for the electronic nose, believing them to be sufficient for most, if not all, applications.

1.3.3 Array Design & Data Analysis

Finally, there is array design and data analysis. As previously mentioned, a good electronic nose requires that the sensing materials complement each other, and in order to choose complementary materials, one needs to know the behavior of all possible sensing elements. To evaluate the behavior of these materials, one could run experiments, but considering the seemingly

infinite combinations of gases and the sheer number of MOFs to choose from, using experiments alone is combinatorically prohibitive. Likewise, data analysis requires a working knowledge of the behavior of the sensing elements, resulting in the same combinatoric issues. These problems instead require a computational pipeline, which is where the previously mentioned computational techniques become relevant.

The research presented herein assumes arrays of mass-based sensors using MOFs as the sensing materials, and thus the work focuses on the array design and data analysis of an electronic nose, specifically for complex gas mixtures.

2.0 Genetic Algorithm Design of MOF-based Gas Sensor Arrays for CO_2-in-Air Sensing

As a starting point for my research on the electronic nose project, I completed a variation on a study by Gustafson and Wilmer titled “Intelligent Selection of Metal-Organic Framework Arrays for Methane Sensing via Genetic Algorithms”, for which they developed a novel genetic algorithm used to design gas sensing arrays from a set of 50 MOFs for the analysis of ternary gas mixtures of methane, nitrogen, and oxygen. [34]

In my variation of their work, I reimplemented and modified the genetic algorithm to design arrays using the same set of 50 MOFs, but for ternary gas mixtures containing carbon dioxide, nitrogen, and oxygen. The goals of this work were three-fold; firstly, to examine a ternary gas mixture relevant to breath-analysis applications; secondly, to improve the genetic algorithm; and finally, to examine the physical properties of the MOFs in the best and worst arrays and see if there were any features, or combinations of features, which should be targeted or avoided during array design, with the potential to inform array design in future works before running any adsorption simulations. The 50 MOFs used in this work, and the study it is based on, were selected from the CoRE MOF database with the aim of having a diverse representation of surface areas and void fractions.[61] A list of these MOFs and their relevant properties can be found in Appendix A.

2.1 Introduction

There are numerous applications for sensing carbon dioxide in air, prominently environmental monitoring, with carbon dioxide and methane being the two most common greenhouse gases driving climate change. Additionally, with geologic carbon sequestration gaining traction as a carbon capture and storage (CCS) technology, there is a growing need to detect leaks from storage sites grows. [62–64]

Other applications include indoor air quality monitoring, with CO_2 acting as an indoor pollutant, and safety applications for the detection of rapid accumulation in confined spaces. [65] Exposure to elevated levels of CO_2 in air poses a two-fold threat, acting as both an asphyxiant by displacing oxygen and as a toxicant, both with potentially deadly consequences. Adverse health effects have been documented at exposure to concentrations above 1000 ppm, such as headaches, drowsiness, and slight nausea. Exposure to concentrations above 5% can result in the development of hypercapnia, a build-up of CO_2 in the bloodstream, and respiratory acidosis, an inability to clear excess CO_2 the lungs.[66–68] Exposure to concentrations above 10% can result in convulsions, coma, and even death. And lastly, exposure to concentrations above 30% can lead to a loss of consciousness in a matter of seconds.[66,67] Clearly, rapid, sensitive, and portable (ideally wearable) CO_2 sensors would benefit many people.

Finally, along with nitrogen and oxygen, carbon dioxide is one of the primary components of exhaled breath, and this ternary gas mixture provides a useful starting point for the exploration of electronic noses for breath analysis.

2.2 Methods

2.2.1 Molecular Simulations of Gas Adsorption

In order to examine the adsorption of gases into porous crystalline materials, one can use a type of simulation known as Grand Canonical Monte Carlo (GCMC) simulations. In essence, GCMC simulations randomly place gas molecules into the framework in what is known as an insertion move. Then, depending on the energetics of the inserted molecule, the move is either accepted or rejected. GCMC moves also include deletion, translation, regrow (effectively a type of rotation) and identity change, which are either accepted or rejected based on the same energetic criteria. By doing enough of these moves, one can eventually approximate the equilibrium adsorption of the gases in the framework. The specific program we use is called RAPSA.[69] For a detailed description of GCMC simulations, please refer to Appendix A.

For this work, we ran various GCMC simulations of ternary gas mixtures containing CO_2, O_2, and N_2. The compositions of CO_2 and O_2 ranged from 0% to 30%, and the composition of N_2 ranged from 40% to 100%, each in increments of 1%. This resulted in 961 unique gas mixtures. The temperature and pressure of these simulations were 298 K and 1 bar, reflecting atmospheric conditions. To model electrostatic interactions, which are important for accurately predicting CO_2 and, to a lesser extent N_2 adsorption, we assigned partial charges to the atoms of the MOF frameworks via the EQeq method.[70] Similarly, the molecule parameters of the gases also included partial charges, and the forcefield which we used, TrAPPE, has been shown to accurately simulate these effects.[71] The molecule parameters can be found in the Appendix A. The simulations resulted in a library of total and component adsorbed mass data as a function of MOF and gas composition.

2.2.2 Array Design via Genetic Algorithm

The next step is the design of gas sensing arrays. Although it is intuitive that adding more sensing elements to an array should improve performance, to what degree the performance can be improved has not been widely explored. Similarly, there has been little work on systematically finding the best combinations of MOFs for an array, and given that with as few 50 MOFs, one could construct over 1.125×10^{15} unique arrays, determining the top performing arrays is highly non-trivial.

We have developed two strategies for the design of sensing arrays: a brute force approach, in which we evaluate all possible arrays of a given size, and a genetic algorithm approach, in which we continually update a set of arrays with a mutation strategy in order to seek the best (or worst) performing arrays. The brute force strategy is useful only when designing small arrays from a relatively small set of MOFs, for which the total number of arrays is manageable. Here it is used primarily to evaluate the effectiveness of the genetic algorithm approach. In the interest of simplicity, a detailed description of the genetic algorithm approach is left to Appendix B.1. Note, however, that in addition to the mutation strategy used, the genetic algorithm requires evaluating some sample output for each array, and ranking its performance with a quality metric, each of which are described in the subsequent sections.

As a test of the genetic algorithm, we calculated by brute force all 1-, 2-, 3-, 4-, and 5-element arrays, and then screened arrays of the same size with a genetic algorithm (see Figure 2). We ran the algorithm three times seeking the best arrays, and another three times seeking the worst arrays, for a total of 6 runs per array size. Each run of the genetic algorithm included 20 arrays per generation and 200 generations per run. As shown in Figure 2, the genetic algorithm is clearly successful in finding the best and worst performing arrays of each size.

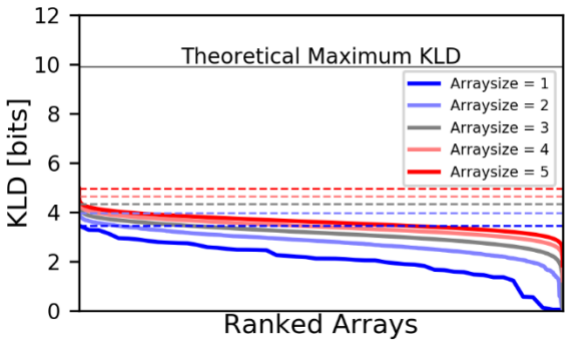

Figure 2. Comparison of the genetic algorithm to brute force screening for 1- to 5-element arrays. The dashed lines highlight only the KLD value of the best array designed by the genetic algorithm, and the solid lines are the KLD values of all arrays designed by brute force screening. Note, there are more arrays as array size increases; the results are simply stretched so that the best/median/worst arrays are vertically aligned.

2.2.3 Analyzing Gas Mixtures Using Arrays

As previously mentioned, the type of sensors which we envision for the array are simple mass-based sensors, specifically SAW devices which measure changes in mass due to gas adsorption. In order to predict the composition of the gas mixture which the sensor is exposed to, we compare the sensor output from each sensing element to our library of simulated adsorption data. Then, a probability is assigned to each of the possible compositions in based on how close the measured and simulated masses are. Effectively, the predicted composition(s) are those with the highest probabilities.

Since we are not yet using real devices, we must create a placeholder for the sensor output, here just using values from the simulated adsorption data. We chose to use values from the simulations of 5% CO_2, 20% O_2, and 75% N_2, as this is approximately the concentration found in human breath and it also corresponds to levels of CO_2 in air which start to become especially concerning.[66] Note that since the sensor outputs are generated directly from the simulated data,

there is a composition whose predicted adsorption values perfectly match the sensor outputs. However, with real sensor device there is always some measurement error. We account for this by artificially imposing measurement error on the sensor output; this ensures that no gas mixture can be predicted with 100% certainty. For a complete description about this prediction methodology, please refer to Appendix B, as well as to the previous work which this is building upon .[34,72,73]

In this work, each MOF showed selective binding for CO_2 (over either O_2 or N_2), and thus one might guess that finding complementary combinations of the 50 MOFs screened to detect O_2 and N_2 would be impossible. Fortunately, even small differences in the adsorption behavior of the individual MOFs have a way of improving predictions once the arrays are sufficiently large and given that the elements of the array are chosen carefully. Figure 3a-c shows the best, median, and worst performing 1-element arrays, and Figure 3d-f shows the best performing 5-, 15-, and 25-element arrays which illustrate this effect.

In general, so long as the simulated adsorption data is accurate and trace gas species which we have not yet accounted for have a negligible effect, this method would work well for predicting the composition of real gas mixtures (assuming that the array used is also sufficient). What is most important to note is that for any composition one wishes to predict, one must run a distinct set of adsorption simulations, one per composition per sensing material. This aspect of the method is severely limiting, creating a computational bottleneck, and thus was changed dramatically for complex gas mixtures, as will be explained in later sections.

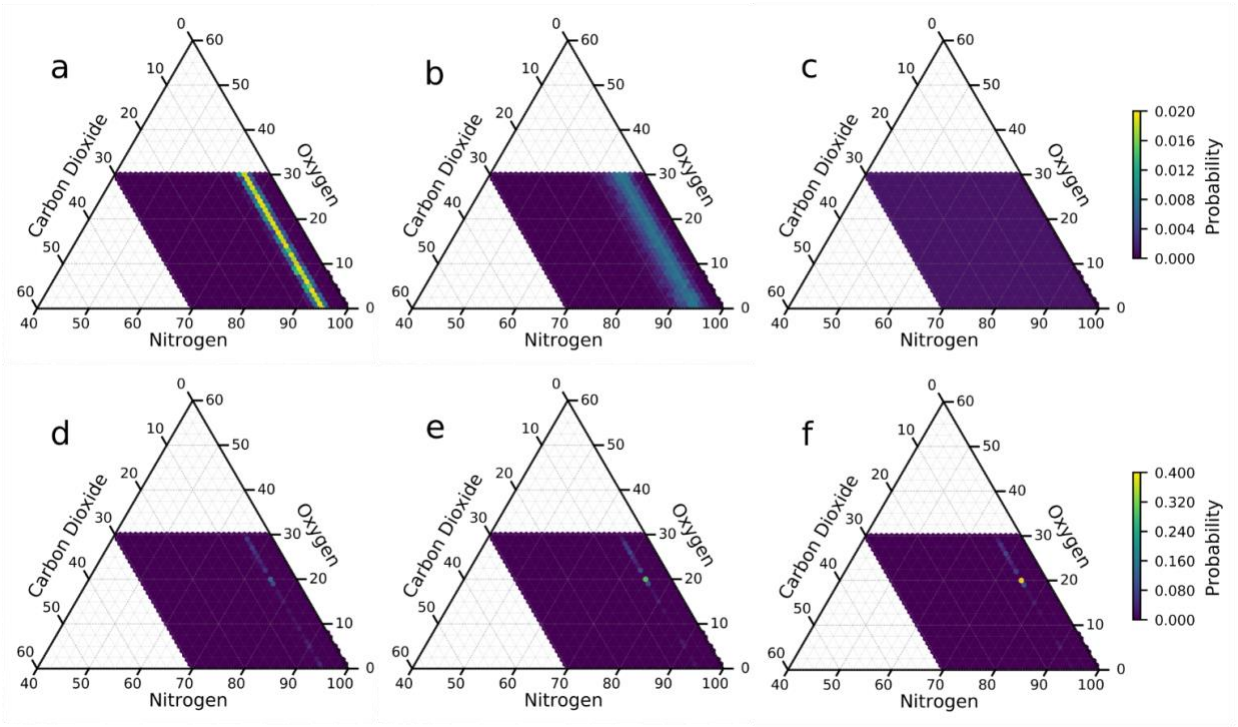

Figure 3. (a-c) Ternary discrete probability plots of (a) the best (Mg-MOF-74)[74], (b) median (Cu4I4(DABCO)2 MOF)[75], and (c) worst (La(PODC)1.5(H2O) MOF)[76] 1-element arrays. (d-f) Ternary discrete probability plots of the best (d) 5-, (e) 15-, and (f) 25-element arrays. Please take note the change in scale of the color bar for the plots in (a-c) versus the plots in (d-f).

2.2.4 Ranking Arrays

In order to quantify the array performance beyond simply looking at the probability of the gas mixture we are testing for, we used a metric known as the Kullbeck-Liebler Divergence (KLD), which in essence, tells us how much better we predict a composition over random chance.[77] The KLD is calculated (in units of bits) as follows:

$$KLD = \sum_{i=1}^{N} P_i \cdot \log_2(P_i \cdot N) \tag{2-1}$$

The range of possible KLD values is limited only by the number of possible outcomes, so in the case of gas sensing, the number of possible compositions. As such, it can be used to compare arrays independent of size, the individual sensing elements, or even the type of sensing mechanism, so long as the number of possible compositions is the same. Consequently, we can use the KLD value to show the advantage of adding more elements to a sensing array, or conversely, the diminishing returns of adding too many elements. Figure 4 shows the KLD values of the best and worst arrays used in this work as a function of array size. Arrays of size 1-5 were examined using the brute force method, whereas all other array sizes were screened via the genetic algorithm.

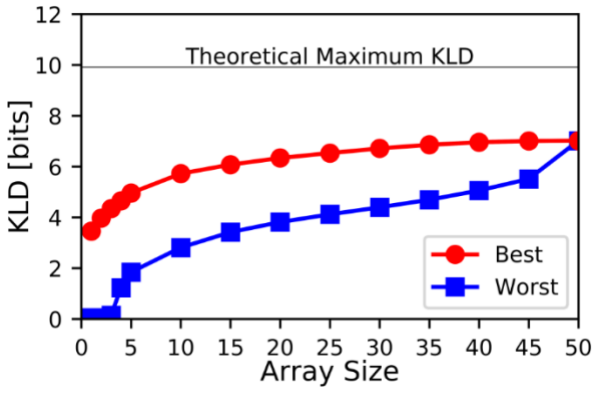

Figure 4. KLD (in bits) of best-found and worst-found arrays as a function of the number of sensing elements.

It is clear that for smaller arrays, the overall performance is worse. The maximum achieved KLD for a 1-element sensor is only 3.46 bits, while for a two-element sensor it is 3.97 bits, and for a three-element sensor it is 4.35 bits. By 45 elements, the maximum achieved KLD is 7.01 bits. It also becomes apparent from Figure 4 that the range of possible KLD values is much greater for small arrays than for large arrays. Although the decrease in variability of the KLD for large arrays moderates the need for advanced screening approaches like the genetic algorithm, it underscores the advantages of employing many-element sensors over 1-element sensors.

2.3 Results & Discussion

Combinations of physical properties that lead to high performance sensor arrays could provide insights into building better sensing arrays that use MOFs beyond the 50 considered in this study. Thus, here we examine the combinations of physical properties of the MOFs in the best and worst performing arrays. Intuitively, one expects that an array made up of a diverse set of materials would perform better than an array of nearly identical materials, as diverse properties should give rise to diverse adsorption behaviors

Although there are many different physical properties one could choose to examine, there are a few that are particularly likely to be important based on known trends in gas adsorption. At low pressures, adsorption behavior is typically governed by the heat of adsorption, a thermodynamic property. However, at moderate pressures, adsorption behavior has been shown to trend with the surface area of the material, and at high pressures, with the free volume of the material, both physical properties.[78,79] Furthermore, at all pressures, pore diameter limits which molecules can diffuse into the bulk of the framework. Thus, we examined the influence of the following properties on array performance: volumetric surface area, void fraction, and pore diameter.

In Figure 5, the 300 best and 300 worst 5-element arrays, ranked from best to worst, are plotted against the physical properties of the elements in the array. We choose to examine 5-element arrays, rather than larger arrays, since with smaller arrays it is harder to compensate for bad elements. Furthermore, with larger arrays, we observed a narrowing of the minimum and maximum KLD values. For 5-element arrays, the KLD of the 300 best-found arrays ranges from 4.58-4.94 bits, and for the 300 worst-found arrays the KLD ranges from 1.87-2.51 bits.

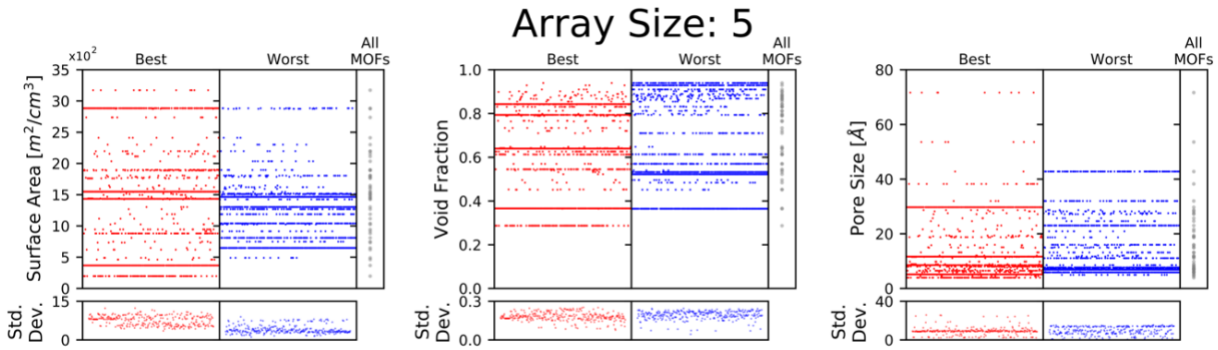

Figure 5. Physical property vs. rank (top/bottom 300 arrays). The bottom plot shows the standard deviation of the corresponding physical property for each array (e.g., one array with the following five void fractions [0.79, 0.84, 0.54, 0.37, 0.64] would have a standard deviation in its void fraction of 0.17). The appearance of horizontal 'lines' corresponds to MOFs which are frequently featured in the best/worst arrays and gives us insight into both the (un)desirable properties as well as the (un)desirable spread of those properties.

For all three properties, the elements of the best arrays span a wide range of the available features, whereas the worst arrays feature elements with a narrow range of those features. Here, however, both the best and worst arrays feature similar ranges of void fractions, and consequently the standard deviations are comparable. For some properties, there seems to be a limited range of useful features. Consider pore sizes, where the best arrays exhibit diversity over a limited range (ca. 4-30Å) when compared to the worst arrays (ca. 5-45 Å). This suggests that, while a diverse set of elements is still beneficial, there is a limit to the usefulness of excessively small (< ~4 Å) or large pores (> ~30 Å), as beyond those limits, they either exclude all gases (small pores) or that the volumetric density of chemically selective adsorption sites is too low for appreciably selective binding (large pores).

In addition to examining the complementarity within an individual physical property, we wanted to examine whether there was complementarity between different physical properties. For

example, it might be desirable to have an element with a high surface area, but a narrow pore diameter so that it selectively adsorbs only smaller gases, but in large quantities, providing a uniquely good signal for detecting those components. This complementarity is examined in Figure 6.

No complementary effect between any two properties becomes readily apparent, so it might be safe to conclude that for this gas mixture at these conditions, it is enough to consider only the individual properties of the elements. Nevertheless, there is still a noticeable clustering of the blue dots (worst array elements) compared to the red dots (best array elements), underscoring the importance of a diverse property set.

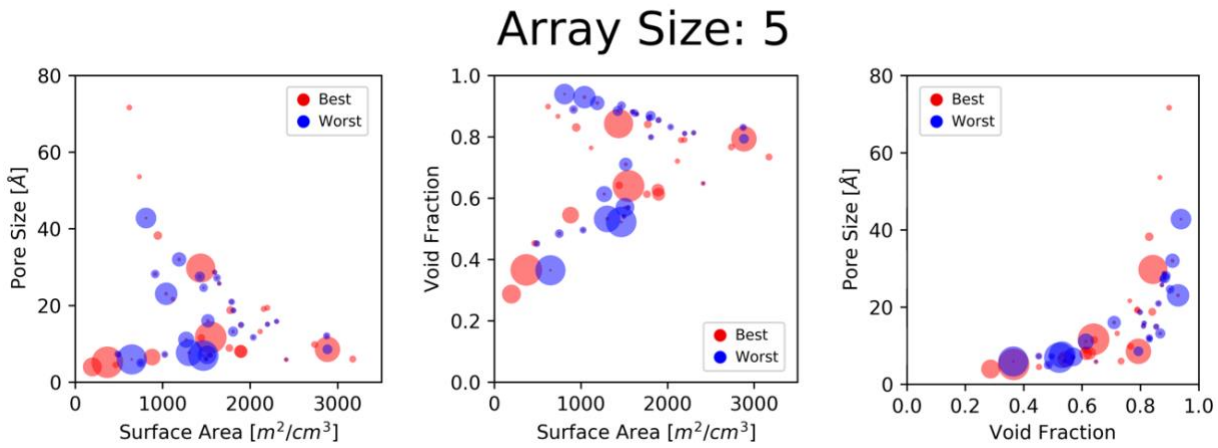

Figure 6. Physical property coupling (top/bottom 300 arrays). Dot color corresponds to elements of the best/worst arrays (red is for best arrays, blue for worst arrays) and dot size corresponds to the number of times the element is featured in the best/worst arrays (a larger dot corresponds to being present in more arrays). This is the same set of arrays used in Figure 5.

When considering the data shown in Figures 5 and 6, it is important to remember that there are only 50 MOFs to choose from, and the gas mixtures feature only three gases of similar sizes (kinetic diameter of CO_2 is 3.30 Å, of O_2 is 3.46 Å, and of N_2 is 3.64 Å).[80] As such, the selected

MOFs may not be representative of the whole material class (though, as a reminder, they were chosen to be relatively diverse) and the gases are clearly not representative of gases species at large. Thus, the conclusions from the above figures, may not be applicable for all gas sensing application, but rather specific to detecting CO_2 in air. Moreover, with CO_2 being strongly polar (relative to N_2 and O_2) and more likely to exhibit strong binding relative to these gases, it is plausible that other features, such as open metal sites, have a greater influence on the adsorption behavior, and thus are better predictors of and design criteria for the best/worst arrays than the features examined here.[81] Nevertheless, our study and the methodology described can help guide intuition with regards to the importance of these features for gas sensing.

2.4 Conclusions

Herein, we have described a methodology for screening gas sensing arrays of MOFs, specifically for the detection of CO_2 in air. Additionally, we examined the physical properties of the MOF elements in the array to improve intuition for the design of sensing arrays.

An interesting feature of the gas mixture studied was the high sensitivity of all MOFs towards CO_2. Although this result gave the impression that reliably detecting the O_2 and N_2 would be difficult, we found that with even relatively small arrays, we were able to accurately resolve their composition. Furthermore, our analysis of the physical properties of the MOFs seems to confirm our intuition that diverse arrays lead to improved sensing. A noteworthy caveat to this is that there seems to be a practical range for certain features, namely pore size, as with excessively small or large pores, selectivity is diminished by the exclusion or inclusion of all the gases present

in the mixture, respectively. Although this limit is likely to change depending on the gases of interest, it seems reasonable that this concept should hold.

In the next iteration of the work, we focus on studying complex gas mixtures (i.e., greater than 3 components). However, this required several changes to the overall procedure in order to keep computational times and costs within reason.

3.0 Computational Design of MOF-based Electronic Noses for Dilute Gas Species Detection: Application to Kidney Disease Detection

3.1 Introduction

The primary limitation of the previous work was the lack of complexity of the gas mixtures which could be considered due to computational expense, as the prior method required enumerating every possible gas composition and then, for each one, simulating its adsorption in every MOF that might be used in an array. Thus, as the library of MOF candidates grows, and as more gases are included in the set of possible compositions, the prior method becomes combinatorically prohibitive. Although there are numerous applications for the sensing of complex gas mixtures, we have chosen to focus on breath analysis, specifically for disease detection applications.

The disease which we chose to target in this work is chronic kidney disease (CKD), for which ammonia is a well-establish biomarker with known healthy and diseased concentration ranges.[13,22] There are several other diseases with known biomarkers in breath, but often the applicable concentration ranges are not well-known, or the concentration of the relevant species in the breath varies dramatically due to external factors. Despite these challenges, disease detection by breath remains a promising area of research owing to its unique and largely untapped clinical potential.

Since breath contains many possible gas species at a wide range of concentrations, we needed to develop a method that was less computationally demanding while still maintaining sufficient accuracy.[82] Fortunately, the majority of the gases of interest are present only in trace

quantities, and so their adsorption in MOFs obeys Henry's law (a linear function of the concentration in the bulk mixture outside of the MOF). Thus, it is sufficient to determine only the corresponding Henry's coefficients to be able to predict the amount of each trace gas absorbed by the MOFs, drastically reducing computational demand. In this next phase of the work, we evaluated a modified form of Henry's coefficients, which we call Combined Linear Adsorption Coefficients (CLACs), where these coefficients quantify not only the amount of trace gas species adsorbed, but also quantify changes in the adsorption of the non-trace gases, all as a function of the trace gas concentration.

CLACs were evaluated for three trace gases (carbon dioxide, argon, and ammonia) in 50 air-filled (nitrogen and oxygen) MOFs, the same 50 previously used. We then designed various a gas sensing arrays which were used to classify a set of artificial breath samples as either healthy or diseased for CKD via a newly developed predication algorithm.

A similar strategy was first developed by Sturluson et al.[36] In their work, they computationally evaluated Henry's coefficients for a set of gases (carbon dioxide and sulfur dioxide) and designed sensing arrays of MOFs. However, the key difference between their work and ours is our simultaneous consideration of both trace and non-trace gas species (i.e., gases for which Henry's law would not apply). Hence where Sturluson et al. used traditional Henry's coefficients, we employ our CLACs. Consequently, the resulting arrays are geared towards different applications, and the subsequent analysis procedures also differ. Their work has nevertheless been influential on ours and we have adapted their methods liberally, as will be discussed later.

The methods employed in this work can be broken into four distinct parts: combined linear adsorption coefficient evaluation, array design, breath sample generation, and breath sample analysis.

3.2 Methods

3.2.1 Combined Linear Adsorption Coefficients (CLACs)

CLACs are most simply described as a modification of a traditional Henry's coefficient. Consider exposing a MOF to a simple gas mixture which contains trace quantities of CO_2 dispersed in non-trace quantities of N_2 and O_2. Henry's Law tells us that the amount of CO_2 adsorbed by the MOF is proportional to its partial pressure, assuming low partial pressures of CO_2. However, what the Henry's coefficient does not quantify is how the adsorption of CO_2 impacts the adsorption of N_2 and O_2. CLACs address exactly this, quantifying both how much CO_2 is adsorbed, and how much N_2 and O_2 is (presumably) desorbed, all as a function of the partial pressure of CO_2. Thus, a CLAC can be thought of as the sum of a traditional Henry's coefficient for the trace gas and a desorption correction for the non-trace gases. This also requires introducing a constant into the model for total adsorbed mass, that constant being the total adsorbed mass of the non-trace gases in the absence of any trace gases and at the total pressure and temperature of interest. The resulting model looks as follows:

$$m_{total} = m_{non-trace\ gases} + \sum_{i=1}^{N} K_i^* \cdot y_i \tag{3-1}$$

where m_{total} is the total adsorbed mass, $m_{non-trace\ gases}$ is the constant mentioned above, and K_i^* and y_i are the CLAC and mole fraction of trace gas species, i, respectively. Note that unlike a Henry's coefficient, which can never have a value less than 0, a CLAC can in principle take of any real value since it is possible that a MOF desorbs more of the non-trace gases that it adsorbs of the trace-gas, resulting in a net decrease in mass. For further discussion of this value, please refer to Appendix D.

Changes in the concentration of the trace gases within the mixture do not impact their CLACs, however, changes in the concentrations of the majority gas species could strongly impact them. In general, one needs a separate set of CLACs for every gas mixture whose majority species' concentrations are different. To address this, we ran a set of GCMC simulations in RASPA for each trace gas species in which the mole fraction was varied from 0-0.05. The remaining non-trace gases were limited to N_2 and O_2 in either a 3:1, 4:1, or 5:1 ratio. A pressure of 1 bar and a temperature at 298 K were used to replicate ambient conditions. The TraPPE forcefield was used for all gases.[71,83] For further details, please refer to Appendix A.

Conveniently, for almost all of the gases and MOFs, the adsorbed mass of the trace gas species was independent of the composition of the background gases as evidenced by the R^2 cutoff used when fitting (described below). The evaluated CLACs could thus be used for a wide variety of air mixture compositions. Moreover, the total mass of air adsorbed seemed to be independent of the composition of air, suggesting that for the remainder of the work, air could be treated as a single gas component, rather than separately as N_2 and O_2. Note that when referring to air in this context, we mean only the nitrogen/oxygen mixture, and not any of the other compounds. This greatly simplified the problem at hand, as these were the only two gases which were present in

greater than trace amounts for the current application, and thus we only needed to determine the constant, $m_{non-trace\ gases}$, once for each MOF.

In order to extract the CLACs from the simulation data, we fit a line to the adsorbed masses of the trace gas species for all ratios of air with the intercept forced to 0. We defined the linear region as the largest portion of the data where the R-squared value of the fit was greater than 0.95. Then, using the same set of compositions, a line was fit to the adsorbed mass of air for all ratios of air, except now with no R-squared cutoff employed, and with the intercept no longer forced to 0 since there is still adsorbed air in the absence of any trace gas species. By adding the slopes of these fits together, we get the CLACs for the system. Although a similar result could have been obtained by fitting to the total adsorbed mass directly, it was more convenient to use the method outlined above. A detailed description of this method, and the reason for using it over the total adsorbed mass data or other methods such as Widom insertion, is given in the Appendix 3.1. An example of the resulting fits is shown in Figure 7, along with the raw adsorption data.

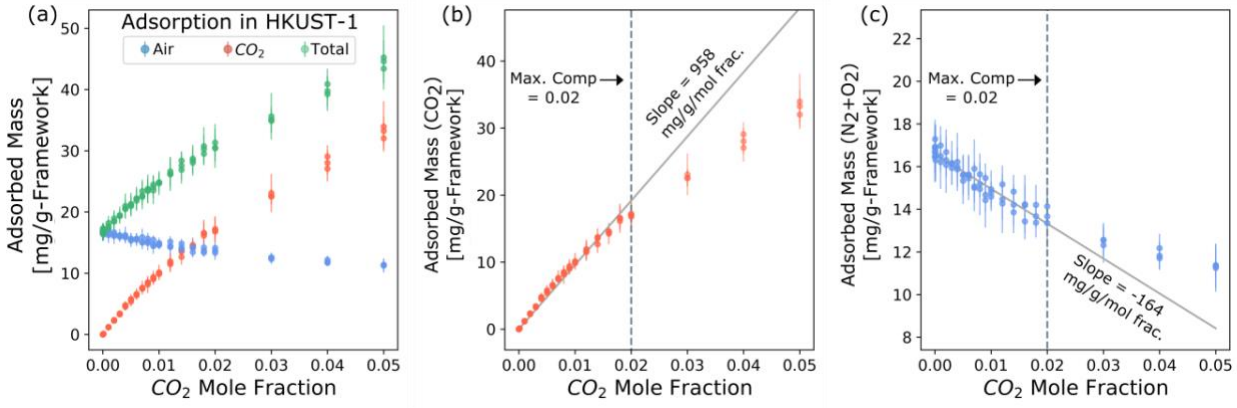

Figure 7. Evaluation of adsorption coefficients from adsorption data (a) Adsorption data in HKUST-1 with a background gas of 3:1, 4:1, and 5:1 N_2:O_2 (i.e., three points for each mole fraction). (b) Linear fit for CO_2 adsorption. (c) Linear fit for air displacement. All fits have an R^2 of at least 0.95. The CLAC which would be used for array design and prediction is 958 - 164 = 794 mg/g/mole fraction.

Figure 8 shows the resulting CLACs for all gas/MOF combinations as a function of the width of their Henry's regime. Please note that a Henry's regime ending at 0.05 mole fraction may extend beyond that point, but additional simulations would need to be done to determine this.

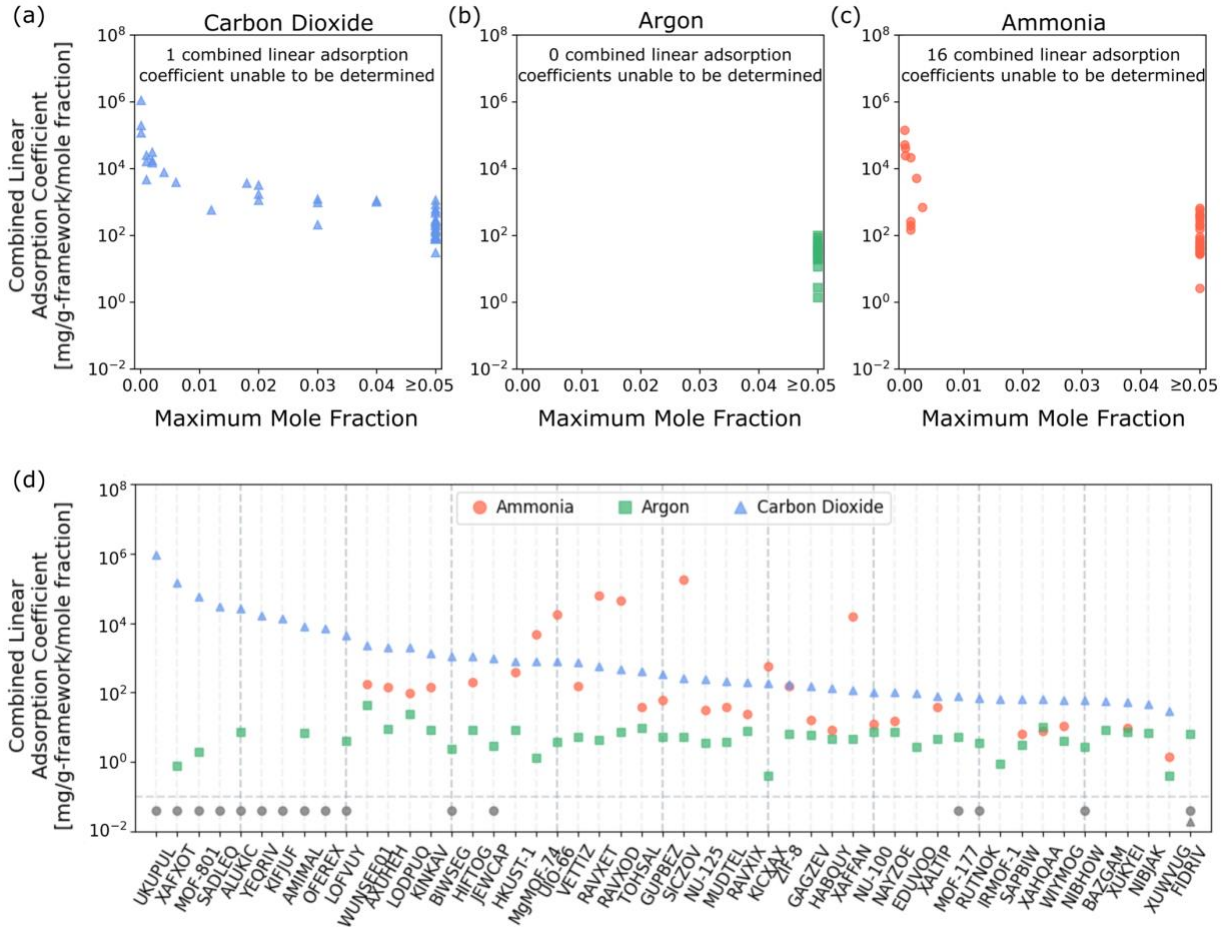

Figure 8. (a-c) CLAC versus maximum concentration for which adsorption of the trace gas species is linear (i.e., end of Henry's regime) for (a) carbon dioxide, (b) argon, and (c) ammonia. A fit was not obtained for all gases/MOFs either because adsorption was highly non-linear or because there was uncertainty in the adsorbed masses from simulations which prevented fitting with the desired R2 cutoff. The number of non-fit MOFs is printed on each plot. (d) CLAC for each gas/MOF combination, sorted by decreasing carbon dioxide CLAC. If a given gas/MOF combination does not have a CLAC, a grey symbol of the same shape is plotted below the horizontal line.

3.2.2 Array Design via Singular Value Decomposition

After calculating CLACs for each gas-MOF pair, the next step in the process is array design. However, we first needed to eliminate MOFs which did not have appropriately wide Henry's regimes for our application, in this case corresponding to a minimum mole fraction of 0.05 for all gases. The concept behind this cutoff is that, if all gases are simultaneously in the Henry's regime, then they will not interact with themselves or with the other trace gases, and thus the linear adsorption assumption should still be valid. Although this value of 0.05 is a restrictive cutoff, it was necessary since we treat carbon dioxide as a trace gas. Fortunately, of the 50 MOFs screened, 23 MOFs met this requirement. With this subset of MOFs, we could start designing arrays. Our approach for this was borrowed from Sturluson et al., who showed that the array with the best sensitivity could be determined by performing a singular value decomposition on the matrix of Henry's coefficients for each array, with the best array having the largest smallest singular value.[36] In our work, we used CLACs rather than Henry's coefficients, but otherwise this method is identical. It should be noted that using CLACs here is preferable to Henry's coefficients, since a MOF could in theory exhibit large Henry's coefficients for certain gases but have very low corresponding CLACs if they desorb non-trace gases in similar amounts. The net result would thus be very little change in the total adsorbed mass as a function of trace gas species concentration, meaning little sensitivity towards that gas.

We opted to design arrays of multiple different sizes (best and worst of each 1-, 2-, 3, 4-, 5-, 10-, and 23-element arrays) so that we could examine how prediction quality changed with size and similarly see how the best and worst arrays of a given size compare. Throughout the paper we refer to the best 5-element array as our baseline array, for which results are presented. A brief

overview of the array design method is given in Appendix C, however for a detailed description, we refer the reader to the original paper by Sturluson et al.[36]

3.2.3 Breath Sample Generation

Breath is primarily composed of nitrogen, oxygen, carbon dioxide, water, and other inert gases, such as argon. However, it has been shown that thousands of other trace compounds (ppm- and ppb-levels) are frequently found in breath, some of which are the byproduct of metabolic processes, while others are simply found in the air we breathe.[84–89] In this work, however, we computationally generated our own simplified breath samples, 50 healthy and 50 diseased, to avoid the complications of handling thousands of compounds. The simplified breath samples include only the gases for which we determined CLACs, and notably exclude water vapor, as it is not only a common interferant, but also destabilizes many MOFs and is notoriously difficult to simulate.[90–96] Fortunately, several breath collection methods involve some form of a dehumidification step.[87,97,98]

We identified relevant concentrations ranges of ammonia for CKD based on the report of Bevc et al., who showed that concentrations of 0.49 ± 0.08 ppm and 3.32 ± 2.19 ppm corresponded to healthy and diseased individuals, respectively.[13] We chose the other gas components in the simplified breath samples to be as follows: carbon dioxide mole fraction between 2-5% (uniform distribution), argon mole fraction between 0.6-1.2% (uniform distribution), and the remainder of the mixture being nitrogen and oxygen in a random ratio between 3:1 and 5:1 (uniform distribution) respectively.[13] The exact compositions of all the healthy and diseased breath samples are given in Appendix D in Tables 10 and 11, respectively.

Next, we created a set of corresponding sensor outputs for each breath sample, again envisioning the sensor array to be comprised of SAW devices, each using a different MOF as the sensing material, such the corresponding sensor output is a change in mass. The detected masses for each MOF for each breath sample were calculated using CLACs in the same fashion as in the prediction algorithm. With no instrument error yet introduced, the mass 'detected' by the sensor is exactly on what would be calculated by the algorithm for that composition. But note that this does not necessarily translate into perfect prediction of the breath sample by the algorithm. It does, however, guarantee that there is a composition within the bounds of the initial composition space which uniquely has the highest probability, thus if the algorithm fails to predict close to this composition, it is poorly behaved.

3.2.4 Breath Sample Analysis

The final step in the process is to predict the composition of a breath sample given a set of sensor outputs. In our case, each MOF sensing element in the array outputs a measured change in mass, within a certainty governed by a fixed instrument error (i.e., independent of the gases being measured) normally distributed about the detected mass. Under certain assumptions the composition can be determined analytically, as outlined in the paper by Sturluson et al.[36] However, we decided to develop and employ a more general numerical approach that could be re-used in future work where analytical solutions are no longer possible, such as when the CLACs and baseline adsorption values vary more significantly with the composition of the majority gas species. An outline of this numerical approach is depicted schematically in Figure 9.

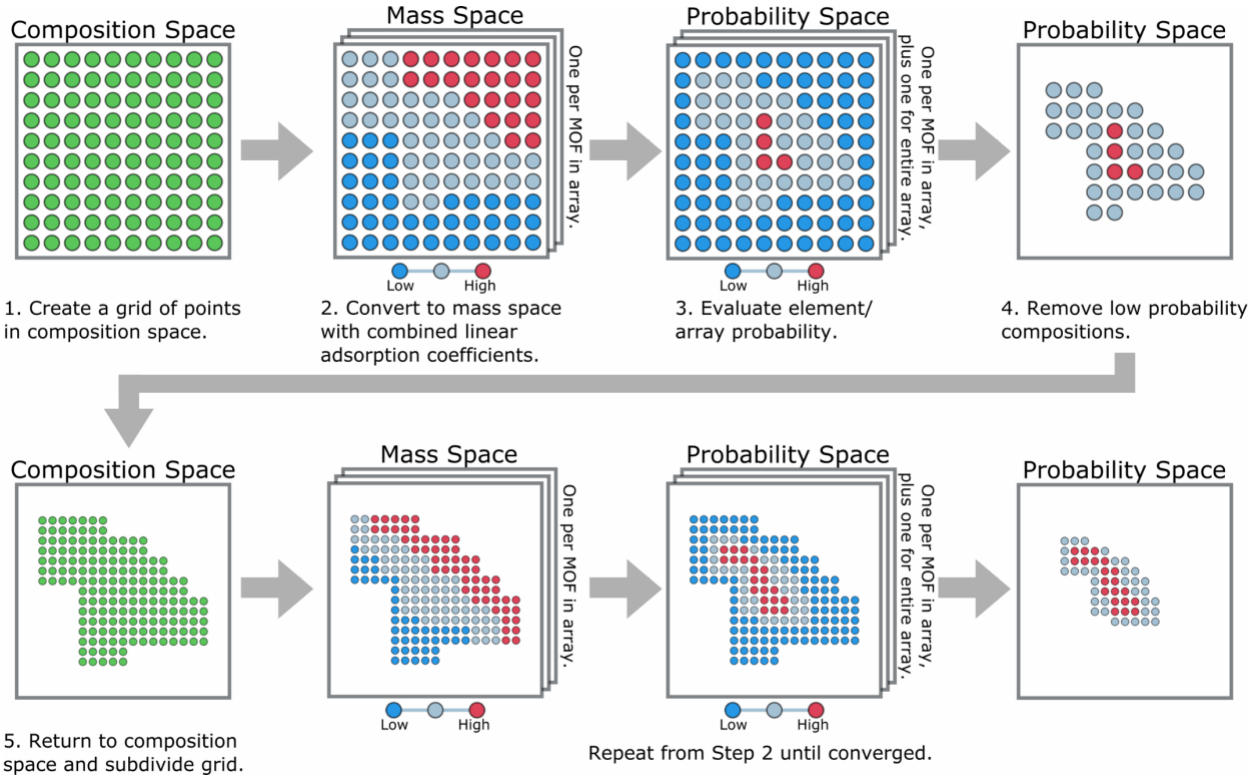

Figure 9. Simplified overview of the algorithm used to predict compositions from a set of sensor data and CLACs. Detailed descriptions of the prediction algorithm are given both below and in Appendix D.

3.2.4.1 Create Compositions

The first step in the numerical prediction method is to generate the initial set of gas mixture compositions; there are three aspects to this step. The first is choosing which gas species should be present in the composition space. Real breath samples can contain hundreds of different compounds, so knowing which compounds must be included, which compounds can be safely excluded, and which compounds can be grouped together is non-trivial. Fortunately, since we consider simplified breath samples, we know exactly which gases are present. The second aspect is deciding what the minimum and maximum value should be for each of the gas species. This is relatively easy since, for most applications, reasonable concentration ranges are already known, or at the very least would need to be determined with more extensive methods prior to designing an

array for that application. Finally, the third aspect is to determine the spacing between points in the initial composition space. This choice can significantly impact the speed of the algorithm and quality of the predicted composition. Generally, the more finely-grained the initial set of compositions is, the better the prediction will be. This point is discussed in more detail in Appendix 3.5.2, but for now all three aspects of creating the initial composition space (i.e., gas species, concentration limits, and spacing) are optimized and standardized for a given array and application prior to use. The specific gases, gas ranges, and gas spacing used in our work are specified in Table 1A in the Results & Discussion section.

3.2.4.2 Evaluate Predicted Masses

The next step in the prediction algorithm is to assign masses to all compositions for all MOF sensing elements. In previous iterations of this work, this step was a bottleneck, as for any combination of MOF and composition, a distinct GCMC simulation was required.[34,35,72,73] As a result, considering a finely spaced multi-component gas mixture for several MOFs would have required significant computing time. Now, masses are determined from equation 2.

Calculating masses in this way enables one to consider all gases individually when using GCMC simulations, dramatically reducing computational time. Consequently, not only can one evaluate more gases, but one can assign a mass to any composition for any MOF so long as the CLACs are known, and the total mole fraction of all trace gas species is within the Henry's regime (i.e., no competitive adsorption).

To illustrate the computational time saved by this change, consider the first phase of this work; we examined a ternary gas mixture of carbon dioxide, oxygen, and nitrogen for a set of 50 MOFs, with the concentration of carbon dioxide and oxygen ranging from 0-30%, and the concentration of nitrogen ranging from 40-100%, all in 1% increments.[35] The result was 48,050

distinct combinations of MOFs and compositions, which of course required 48,050 distinct simulations. In this study, by considering the trace gas species separately, we evaluated the CLACs for 3 gases and 50 different MOFs using only 9,450 distinct simulations. Subsequently, we can examine millions of different five-component compositions in a matter of minutes, all while using less than 20% of the number of simulations. Because of this advantage, designing an array which can handle the thousands of gas species in breath, while still enormously difficult, begins to actually seem plausible.

3.2.4.3 Evaluate Probabilities

Once masses have been assigned, the next step is to compare the calculated masses to the masses detected by the sensor and subsequently assign a probability to each composition for each MOF. This is done by creating a truncated Gaussian normal distribution centered about the detected mass with some known standard deviation, typically chosen to emulate the measurement error for the device (standard deviation = 10mg/g-framework). Array probabilities are calculated by multiplying all of the element probabilities for each composition which are then normalized themselves.

3.2.4.4 Filtering & Convergence Criteria

Finally, the last step is to filter composition space down to the points which have the highest array probability and check for convergence. Assuming the algorithm has not converged, we take the remaining highest probability compositions and subdivide the grid in composition space around those points. This requires choosing both how many points are to be retained and how finely spaced the next grid is. For choosing how many points to retain, we kept only a small fraction such that the number of points in the next cycle was less than or equal to the number of points in

the previous cycle. This guarantees that the number of points does not grow and cause unintended memory or time issues. We set the composition spacing equal to half that which was used in the previous iteration. The net result is an increasingly narrow and fine-grained set of compositions.

As for convergence criteria, after each cycle we check whether the range of concentrations for each gas is within its respective limit. For example, when analyzing the breath samples in this work, we considered the prediction for carbon dioxide and argon converged when the difference between the minimum and maximum concentration values for each gas was less than 1000 ppm, and ammonia converged once the difference was less than 0.1 ppm. Note that, depending on which gas species converged last, the final predicted ranges for the other gases may be far narrower than their specified convergence criteria. Additionally, we do set a limit on the maximum number of iterations so that the algorithm will stop if struggling to converge. For now, each of these parameters (fraction of points retained, convergence criteria, and maximum number of iterations) is optimized for a given array and application prior to use. Figure 10 shows the evolution of the predicted concentration range for a typical breath sample in this work.

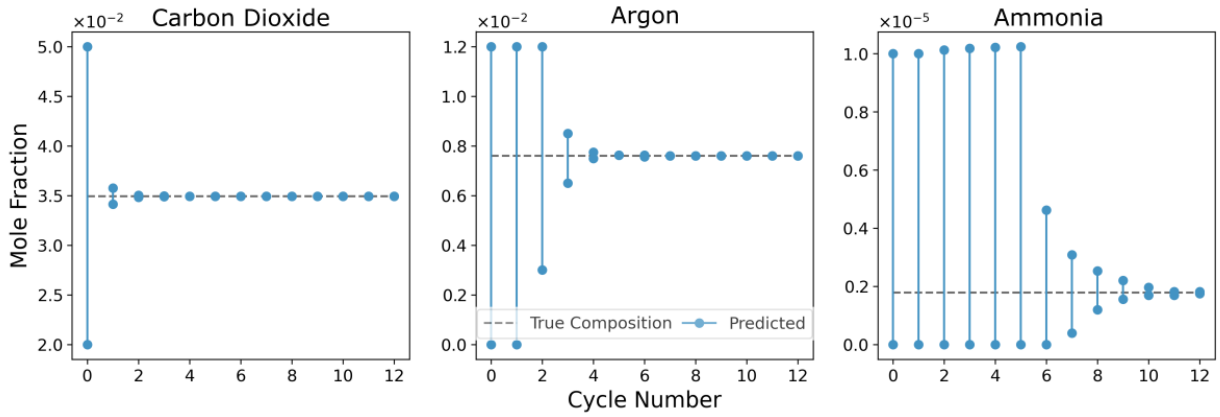

Figure 10. Predicted concentration range for a single breath sample as a function of cycle number for (a) carbon dioxide, (b) argon, and (c) ammonia. The dashed line represents the true concentration of the gas in the breath sample.

Table 1. Initial compostion space for prediction algorithm.

	Initial Range	Initial Spacing	Convergence Limits
Carbon Dioxide	20000-50000 ppm	12.5 ppm	1000 ppm
Argon	0-12000 ppm	2000 ppm	1000 ppm
Ammonia	0-10 ppm	0.25 ppm	0.1 ppm

Table 2. Parameter set for prediction algorithm

Parameter	Value
Maximum Number of Iterations	20 cycles
Fraction to Keep	0.04
Standard Deviation	0.10mg/g-framework

3.3 Results & Discussion

The best 5-element array, shown in Figure 11, includes the following MOFs: ZIF-8[99], XUKYEI, two-fold interpenetrated MOF-5 (HIFTOG)[100], CMOF-4b (XAHQAA)[101], and MOF-399 (BAZGAM)[101].

The prediction results are given in Figure 12, with samples numbered from lowest to greatest ammonia concentration. Clearly, we can reliably predict the composition of the gases, of particular importance being the concentration of ammonia, which is the biomarker for CKD.

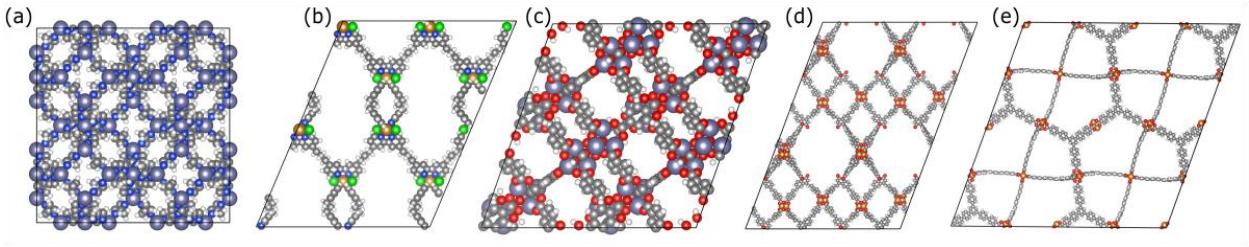

Figure 11. All MOFs which make up the best 5-element sensing array, each displaced as a 2x2x2 unit cell down the crystallographic a-axis. The MOFs are: (a) ZIF-8[99], (b) XUKYEI, (c) 2-fold interpenetrated MOF-5 (HIFTOG)[100], (d) CMOF-4b (XAHQAA)[101], and (e) MOF-399 (BAZGAM)[101]. Note that the common names and first report of MOF given when known, followed by CoRe MOF reference code in parenthesis.

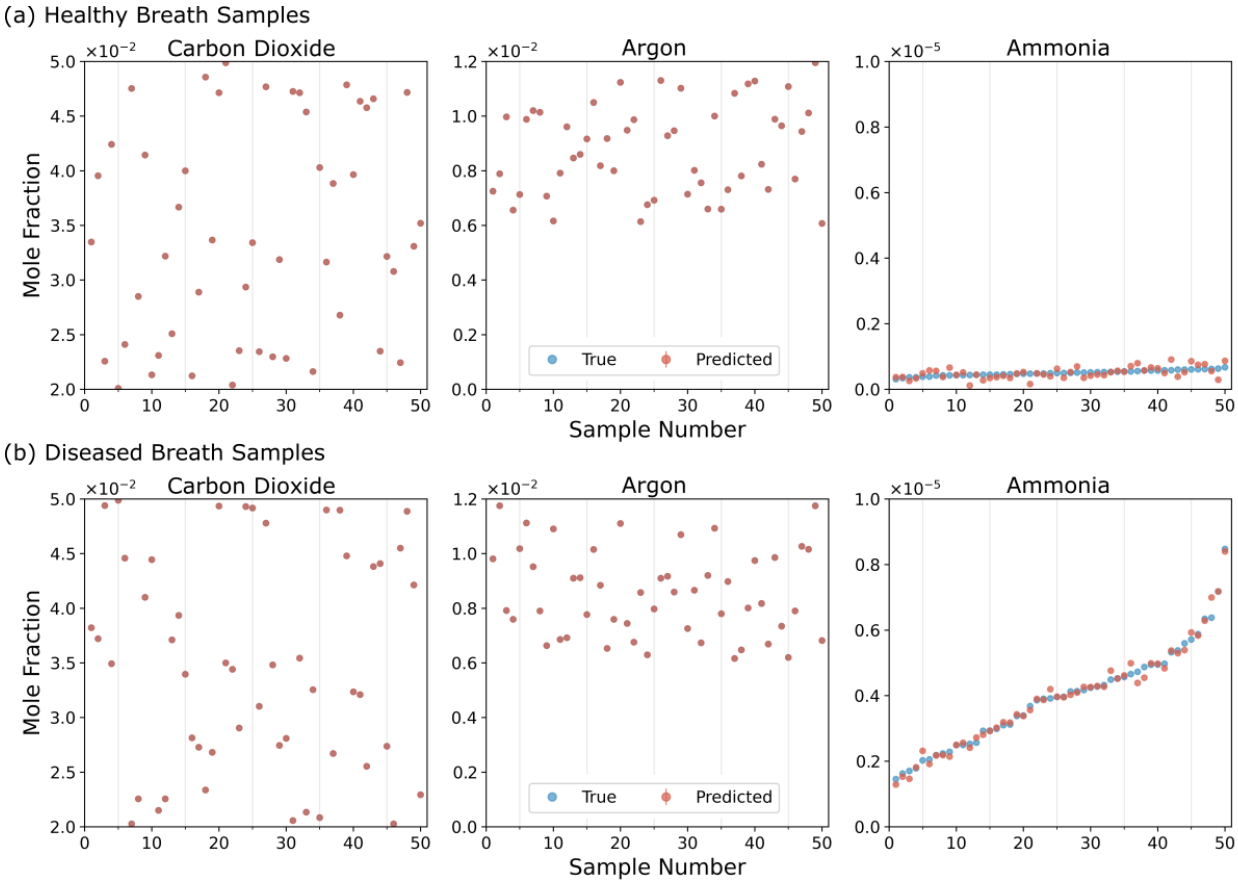

Figure 12. Final predicted concentration ranges for the best 5-element array for (a) 50 healthy and (b) 50 diseased samples, ordered by increasing ammonia concentration. The upper and lower bounds on each plot correspond to the initial concentration range used for prediction.

Although these results were exciting, we did also want to re-examine how both array size and array quality impacted prediction. To this end, we determined the best 1-, 2-, 3-, and 4-element arrays, as well as the worst, median, and best 5- and 10-element arrays, and again predicted the compositions of all 100 samples using the same algorithm parameters, as shown in Figure 13.

It should also be noted that, although there are 5 gases in each breath sample, only 3-elements are required to have a determined system, as the desorption of the two-component air is baked into the CLACs of the other gases. A 5-element array is thus considered over-determined, nevertheless, we use the array shown in Fig. 5 as a baseline for various comparisons, as it is able to reliably differentiate between healthy and diseased samples.

Unsurprisingly, 1- and 2- element arrays struggle to reliably predict the concentration of ammonia, with all the 1-element arrays stopping due to reaching the maximum number of cycles. The 2-element arrays do converge on a concentration, but with a very poor prediction. The best 3-element array substantially improves the overall quality of prediction, but even still, with the chosen parameter set, there would be several false positive/negatives. The 4-, 5-, 10- and 23-element arrays resolve this problem. The 23-element array is a particularly interesting result, as the prediction quality is noticeably poorer than the best 5- or 10-element arrays, which are both subsets of the best 23-element array. Since this analysis uses computer-generated sensor outputs for the breath samples, the algorithm should be able to perfectly predict the composition. An additional consequence of using computer-generated sensor outputs is that we can confidently say that none of the sensors offer contradictory information. Consequently, the reduction in prediction quality of the 23-element array must be an artifact of the algorithm, and further motivates determining a unique set of parameters for each array/application, not just for each application.

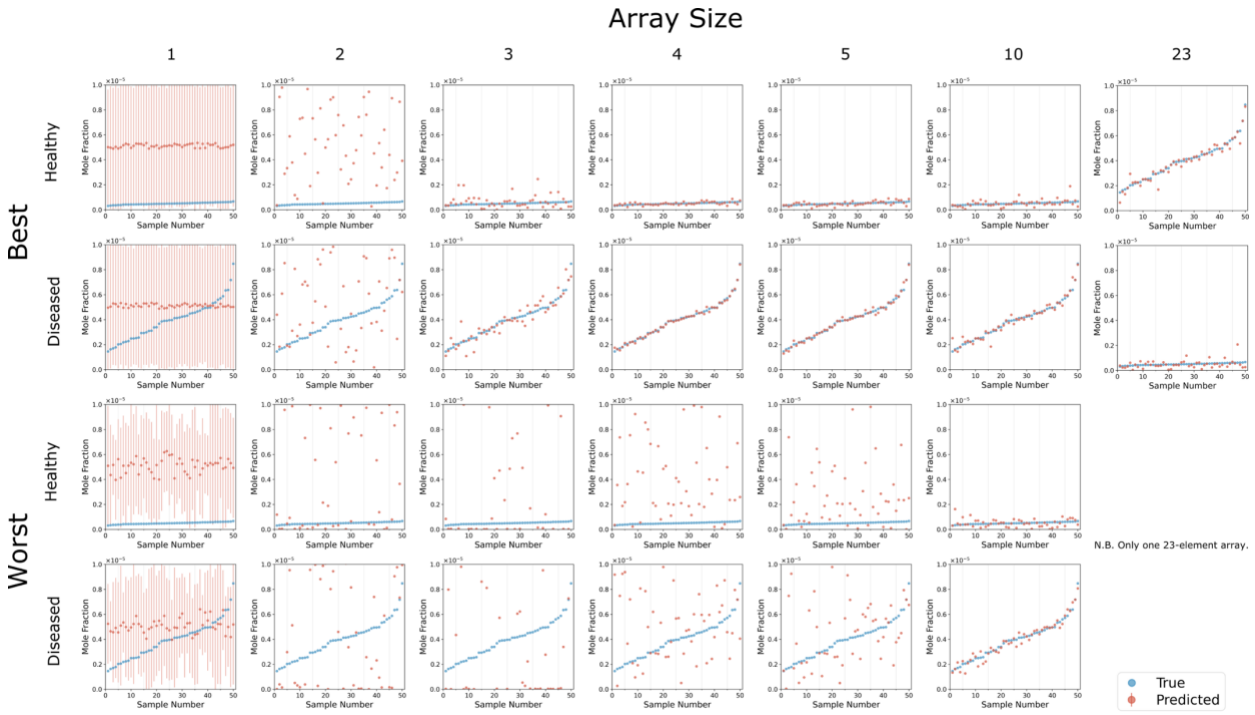

Figure 13. Comparison of the best and worst arrays of various sizes. Note that only the prediction for the concentration of ammonia is shown, as the arrays can reliably predict the concentration of carbon dioxide and argon. Also note that there is only one 23-element array.

Regardless, the overall difference in prediction quality between the best and worst arrays of various sizes underscores that intelligent selection of sensing elements for arrays is still a critical aspect of this electronic nose work, especially considering almost all real arrays will be underdetermined to some degree given that breath can contain thousands of different components.

Finally, as hinted at by the performance of the 23-element array, the algorithm parameters impact the prediction quality at least as much as the array. Thus, even if an array which is capable of making accurate predictions is used, in the absence of a sufficient parameter set for the algorithm, predictions will be unreliable. The effect of specific parameters (i.e., initial composition set and fraction to keep) are examined and discussed in Appendix D. The parameters used within

this work were determined in a guess-and-check fashion, but as the project evolves, a more systematic way of determining algorithm parameters will likely be necessary.

3.4 Conclusions

In conclusion, we determined the CLACs for a set of gases commonly found in breath for 50 MOFs, with the goal of designing a sensor array which can be used to detect kidney disease by breath, for which ammonia is a well-established biomarker. Using the method presented by Sturluson et al., we screened and ranked all arrays of various sizes, and selected the best 5-element array for testing.[36] Then, using our newly designed prediction algorithm, we analyzed a series of 50 healthy and 50 diseased breath samples, successfully quantifying the amount of ammonia for all samples in each set. Although there is still much work which needs to be done, primarily in increasing the number of gases and MOFs considered, accounting for the effects of humidity, and accounting for measurement error, the methods presented here are a marked improvement on our previous work.

4.0 Multi-Pressure Sampling for Improving the Performance of MOF-based Electronic Noses

4.1 Introduction

Throughout most of this project, we have explored a set of only 50 MOFs, far from the over 90,00 synthesized MOFS and further still from the over 500,000 known hypothetical MOFs. Although the 50 MOFs were chosen with the intention of having a diverse set of features, they cannot capture the full extent of the diversity exemplified by a much larger materials set. Even so, from an experimental perspective, synthesizing 50 unique MOFs is quite a challenge, and while the intention of the computational aspect of this project is to inform which MOFs experimentalists should synthesize, there is a strong desire to maximize the amount of information that can be extracted from a single MOF. In a related challenge, for many weakly adsorbing gases such as hydrogen or methane, the total adsorbed mass at low concentrations is often so small that it is hard to extract any information about their concentrations from adsorbed masses alone. Similarly, for strongly adsorbing gases such as acetone, many MOFs can quickly become saturated, again such that it is hard to extract information about their concentrations from adsorbed masses.

Inspired by these challenges, we wanted to explore if sampling gas mixtures at multiple pressures could improve the performance of the arrays. We know from pure adsorption isotherms that adsorption behavior changes as a function of pressure, and thus our hope was that the adsorption of complex gas mixtures would also change sufficiently as a function of pressure, such that sampling a gas mixture with the same MOF at different pressures would be comparable to adding an entirely new MOF to the array. This would help to limit the number of MOFs which

need to be synthesized for device fabrication and could help to find MOFs for detecting particularly strong or weakly adsorbing gases, better than MOFs with atmospheric pressure sampling.

4.2 Methods

In this work, we examine a set of 9 MOFs (HKUST-1, IRMOF-1, MgMOF-74, MOF-177, MOF-801, NU-100, NU-125, UiO-66, ZIF-8), which we had used in previous studies on the design of electronic noses. The molecular modeling software RASPA was used to run GCMC simulations of gas adsorption in these MOFs at a temperature of 298K and various pressures (0.1, 0.5, 1.0, 5.0, and 10 bar).[69] We simulated 4 different sets of 4-component gas mixtures: benzene in $N_2/O_2/CO_2$, methane in $N_2/O_2/CO_2$, hydrogen in $N_2/O_2/CO_2$, and hydrogen sulfide in $N_2/O_2/CO_2$. These 4 gases, along with CO_2, were varied from 0-1% in 0.05% increments. No mixtures contained more than one of benzene, methane, hydrogen, or hydrogen sulfide. N_2/O_2 made up the remainder of the gas mixture, always in a 4:1 ratio (to represent ambient air), for a total of 441 unique compositions per gas mixture. The MOFs are simulated using a combination of universal force field (UFF) and Dreiding parameters, along with partial charges generated by the EQeq method.[70,102,103] Adsorbates are simulated using TraPPE forcefield parameters, which also includes partial charges.[71,104] A detailed description of the simulations, including full parameter sets, is given in the Appendix A.

After generating a complete set of adsorption data for all gas mixtures, MOFs, and pressures, we designed and evaluated various arrays. We created two different types of arrays in this work: single-pressure arrays, in which we sample the gas mixture at only one of the simulated pressures, and all-pressure arrays, in which we sample the gas mixture at all simulated pressures.

With 9 MOFs, there are only 511 possible unique arrays, so both single- and all-pressure were designed via brute force.

The method of predicting gas compositions and quantifying array performance used in this work has been outlined in section 2. A brief overview is given below, and more detailed information can be found in the Supporting Information and relevant papers. To begin analysis, we must first generate a set of detected masses for each of the arrays, here simply using the simulated adsorption data for each sensing element at a known composition, in this case 79% N_2, 19.75% O_2, 0.5% CO_2, and 0.75% of the other gas of interest. Since our goal is to predict the composition given the set of detected masses, we then 'forget' the composition which was used to generate the detected masses. Next, using a truncated normal distribution with a standard deviation of 1 mg/g-framework for hydrogen, methane, and hydrogen sulfide, and 10 mg/g-framework for benzene (because of higher simulation error), we generate a set of probabilities for all compositions based on how close the detected and predicted masses are to each other. Lastly, the array performance is quantified using the KLD (Equation 2) which effectively scores the quality of the predicted composition when compared to random chance, with a higher score corresponding to a more certain prediction.[105,106] The maximum possible KLD is limited by the total number of possible compositions. For 441 unique compositions, the maximum possible value is 6.089 bits.

4.3 Results & Discussion

4.3.1 Benefit of High Pressures (Hydrogen/Methane)

Methane and hydrogen are both small non-polar molecules and, as a result, are normally weakly adsorbing gases. In many MOFs they will make up a very small fraction of the total adsorbed mass, and this effect is exaggerated when they are present in low concentrations. Generally, in order to reliably detect and quantify these gases with mass-based sensing, it helps to increase the amount in which they adsorb relative to the other gas species. We know from many studies that high pressures typically increase the total adsorption in MOFs, especially for small molecules which pack more efficiently than large molecules. With this in mind, we hoped that by increasing the system pressure, we could improve our ability to detect gases like methane and hydrogen. That said, the size difference between nitrogen (3.64 Å) and oxygen (3.46 Å) versus hydrogen (2.89 Å) and methane (3.80 Å) is not very significant, compared to a molecule like benzene (5.85 Å), so it is not obvious what impact pressure would have on the selectivity of adsorbed gases, especially in the presence of more strongly adsorbing gases like CO_2. Fortunately, even if the selectivity does not change, increasing the total adsorbed mass generally improves the performance of mass-based sensors.

One of the MOFs which demonstrates the benefit of high-pressure sensing well is MgMOF-74, shown in **FIGURE 14.**[74] As we increase the pressure of the system, the total adsorbed mass of hydrogen increases, with the amount of hydrogen adsorbed increasing sixty seven-fold from 0.1 bar (0.0008 mg/g-framework) to 10.0 bar (0.0534 mg/g-framework)**.** Methane, as shown in **FIGURE 15**, benefits even more with a seventy-seven-fold increase in mass from 0.1 bar (0.02 mg/g-framework) to 10.0 bar (1.54 mg/g-framework)**.** Please note that, because these

gases are still a small fraction of the total adsorbed mass, we are opting to show only the contribution of the adsorbed hydrogen/methane to demonstrate this effect, rather than the total adsorbed masses.

With such low total adsorbed masses, the mass detection limits of the device become significant, so the significant increases in the adsorbed mass from 0.1 to 10 bar is greatly appreciated for sensing applications. Additionally, high pressures have the added benefit of decreasing the error of our simulations, especially for weakly adsorbing gases, as evidenced by the smoothness of the high-pressure plots. Although this problem can be addressed, to an extent, by increasing the number of cycles or replicating the unit cell during simulations, it is certainly a welcome bonus. In fact, it could be argued that for gases like hydrogen, the challenge in sensing is not in detecting very small changes in mass, as SAW devices and QCMs have impressive mass sensitivities, but rather mapping these changes in mass to a corresponding change in composition with a high degree of certainty.

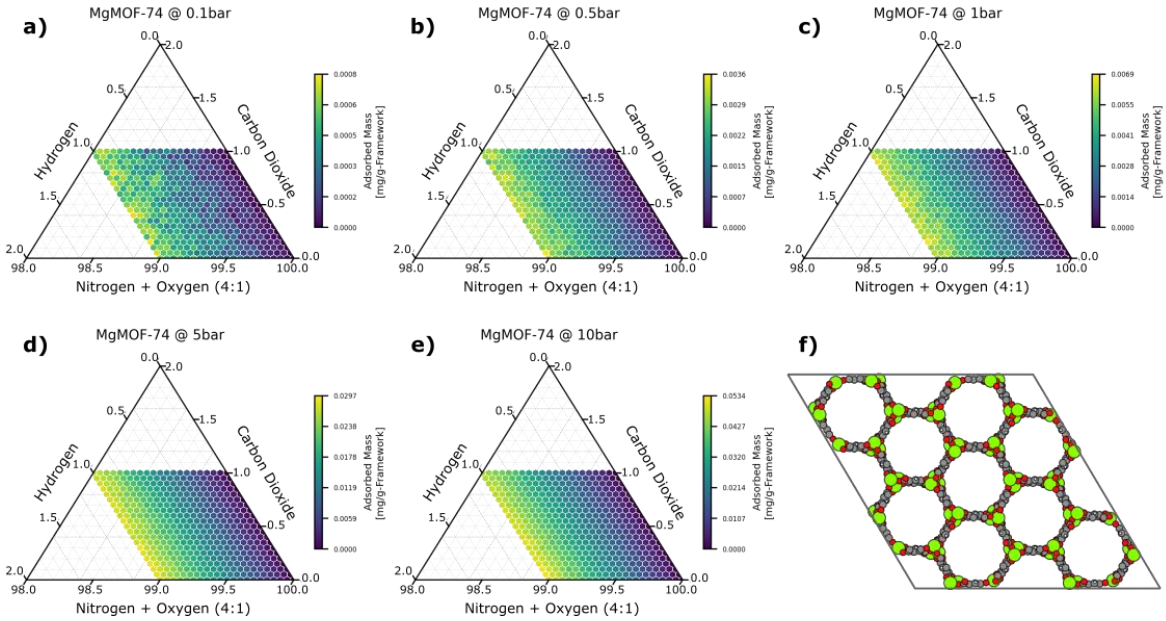

Figure 14. Ternary plots of the adsorbed mass of hydrogen in MgMOF-74 as a function of composition and at the following pressures: a) 0.1 bar, b) 0.5 bar, c) 1 bar, d) 5 bar, and e) 10 bar. f) shows a 2x2x2 unit cell of the MOF projected down the c-axis.

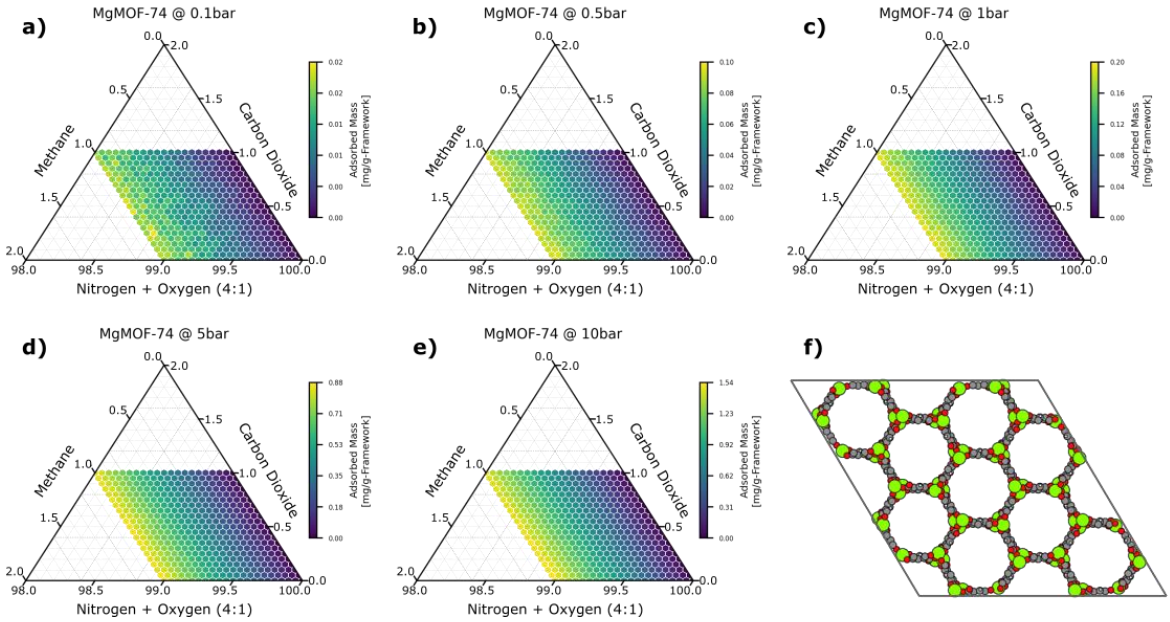

Figure 15. Ternary plots of the adsorbed mass of methane in MgMOF-74 as a function of composition and at the following pressures: a) 0.1 bar, b) 0.5 bar, c) 1 bar, d) 5 bar, and e) 10 bar. f) shows a 2x2x2 unit cell of the MOF projected down the c-axis.

4.3.2 Benefit of Low Pressures (Benzene)

Although it is non-polar, benzene is generally a strongly adsorbing gas due to its large size. Even at low concentrations, it makes up a significant fraction of the total adsorbed mass in most MOFs and can rapidly saturate the sensor response (i.e., a change in the concentration does not result in a change in adsorbed mass). This not only makes benzene difficult to detect, but also the detection of non-benzene gases, since the MOFs lose sensitivity towards those gases in the presence of benzene. Unlike with hydrogen or methane, high pressure sensing largely exaggerates this effect. Instead, in order to improve the detection of benzene, it is actually beneficial to decrease the amount of gas adsorbed, so that the sensor is no longer saturated, and changes in concentration would again result in a change of mass. This is demonstrated well by MOF-177 in **FIGURE 16**.[107] Note that, since benzene is a significant portion of the total adsorbed mass, we are showing that here, rather than just the benzene contribution to adsorption.

Note that the maximum observed total adsorbed mass at 0.1 bar is significantly lower than that observed at 0.5 bar and above. Even so, this decrease in total adsorbed mass is coupled with the necessary desaturation of the sensor, enabling us to reliably predict benzene over this concentration range. Improving benzene sensing by shifting to low pressures highlights an important concept of the sensing elements of electronic noses; the best elements are those in which the *change* in total adsorbed mass from one composition to another is greatest. It is easy to think that highly selective, highly adsorbing MOFs are best, and subsequently, that high mass loadings are universally desired. But as benzene demonstrates, this is not inherently true. At lower pressures, both the selectivity towards benzene and total adsorbed mass decreases, but sensing is still improved because the change in mass as a function of change in composition is improved.

It should, however, be mentioned that for applications where benzene is present in extremely low concentrations (ppm and below), high pressures may not result in saturation of the sensor and may actually benefit from high pressures due to a concentrating effect similar to hydrogen and methane. In fact, one of the MOFs we screened, NU-100, is more useful at high pressures for this reason, as shown in Appendix E.[108] Nevertheless, all other MOFs screened perform best at low pressures, and Figure 16 demonstrates the potential benefits of low-pressure sensing.

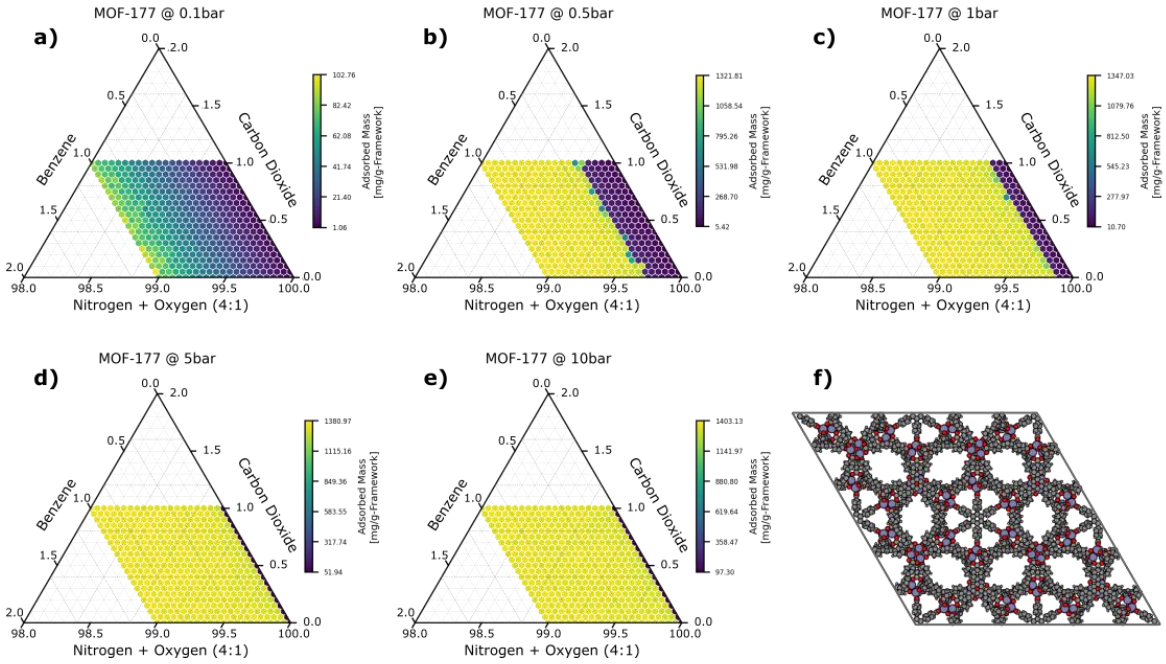

Figure 16. Ternary plots of the adsorbed mass of benzene in MOF-177 as a function of composition and at the following pressures: a) 0.1 bar, b) 0.5 bar, c) 1 bar, d) 5 bar, and e) 10 bar. f) shows a 2x2x2 unit cell of the MOF projected down the c-axis.

4.3.3 Benefit of Multiple Pressure (Hydrogen Sulfide)

Hydrogen Sulfide, like methane and hydrogen, is a fairly small molecule, but it also contains a strong dipole moment, making it a somewhat strongly adsorbing gas. Although there are many MOFs which adsorb hydrogen sulfide appropriately for sensing at ambient pressure, the adsorption behavior can still be beneficially modified by changing system pressure. This is demonstrated well by UiO-66, as shown in Figure 17.[109]

For all pressures, there is an appreciable change in mass in response to a change in composition. That said, at high pressures and high concentrations of hydrogen sulfide, the response appears to flatten out, that is to say, the percent change in mass relative to the total adsorbed mass decreases. However, in this MOF, the apparent flattening of the response is sufficiently compensated by an overall increase in the total adsorbed mass such that the measured change in mass is greater at high pressures, thus offering better performance. Nevertheless, it is highly likely that other MOFs would benefit more from low pressure sampling when the change in mass as a function of concentration is less at high pressure. At low pressures and low concentrations of hydrogen sulfide, the change in mass in response to a change in concentration is smaller, such that arrays using this MOF would benefit more from high pressures. Essentially, for the 9 MOFs screened in this work, hydrogen sulfide, like hydrogen and methane, benefits specifically from high pressures. But unlike hydrogen or methane, it is easy to envision a MOF for which the optimum pressure is actually lower, especially if higher concentrations are expected in the application. Ultimately, the impact of pressure is not as significant for hydrogen sulfide and for gases such as this, it is preferable to focus on screening more MOFs at a single pressure, rather than fewer MOFs at more pressures. Even so, UiO-66 demonstrates that there is some utility to exploring pressures beyond ambient conditions.

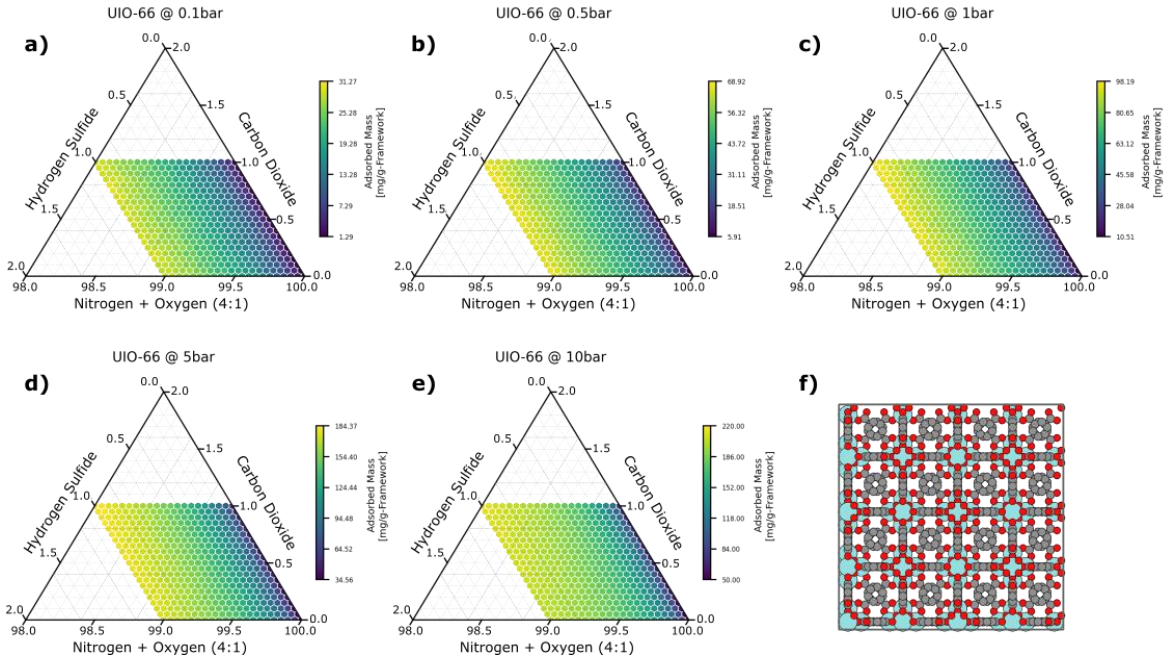

Figure 17. Ternary plots showing the adsorbed mass of hydrogen sulfide in UiO-66 as a function of composition and at the following pressures: a) 0.1 bar, b) 0.5 bar, c) 1 bar, d) 5 bar, and e) 10 bar. f) shows a 2x2x2 unit cell of the MOF projected down the c-axis.

4.3.4 Effect of Pressure & Array Size

Every possible array of each size and pressure was analyzed and its performance quantified with a KLD score. Figure 18 plots the best and worst array performance as a function of array size and operating pressure, including all-pressure arrays (black line).

In general, array performance always improves with array size, however the change in the performance of the best arrays as a function of array size is minimal, suggesting that these arrays rely on only a few high-performing MOFs to make their predictions. Instead, pressure has a much more significant impact. For hydrogen, methane, and hydrogen sulfide, array performance improves specifically with high pressure operation. In fact, of these three gases, hydrogen sulfide

is the only one which exhibits a significant increase in performance beyond size 1 arrays. The jump in performance from size 1 to size 2 arrays for hydrogen sulfide, especially at high pressures, suggests that increasing pressure results in improvement to not just the individual adsorption behaviors, but also the cross-sensitivity of the elements. Even then, beyond size 2 arrays, the improvement in performance is again minimal, consistent with the idea that the best arrays rely on only a few elements.

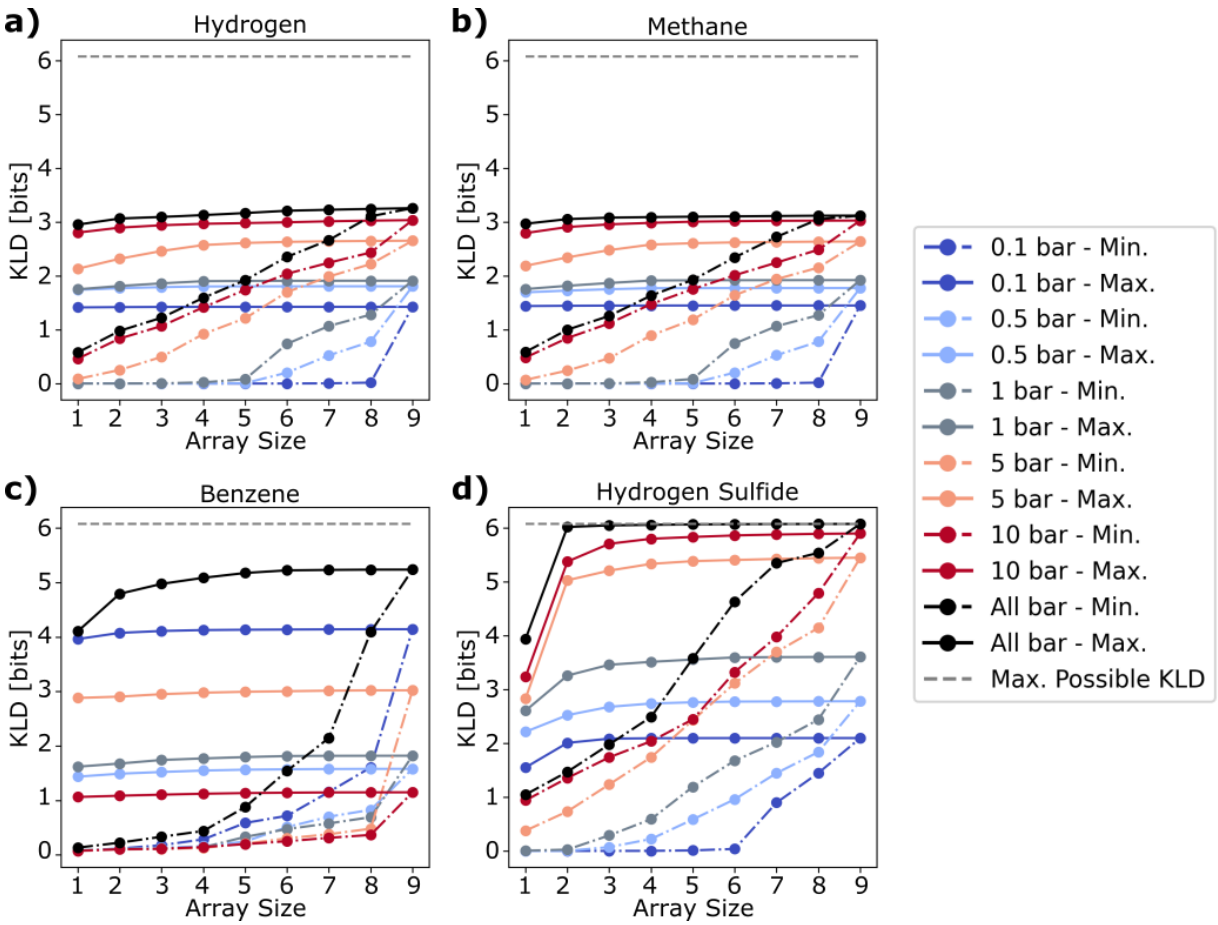

Figure 18. KLD vs. pressure and array size for a) hydrogen arrays, b) methane arrays, c) benzene arrays, and d) hydrogen sulfide arrays. Solid lines are the best performing arrays, and dashed lines are the worst performing arrays.

Conversely, the worst arrays generally improve with both pressure and array size. For example, the KLD scores of the worst arrays at 5 and 10 bar for hydrogen, methane, and hydrogen sulfide all increase steadily with array size. This is because more of the MOFs exhibit useful adsorption behavior at these pressures, even if they are noticeably outperformed by the best MOFs. However, at low pressures, the increase in performance as a function of array size is again limited. For example, with hydrogen sensing at 0.1 bar, the KLD score for the worst arrays is almost 0.0 until all 9 MOFs are used, suggesting that only one of the MOFs has any useful adsorption characteristics for detecting hydrogen. We observe similar behavior for this and other gases at low pressures, suggesting that, under certain conditions, some the screened MOFs are effectively useless as sensors.

Together, these results highlight how varying pressure can change our approach to both the search problem (i.e., screening MOFs) and array design problem (i.e., choosing the correct combination of MOFs and pressures) central to the building an electronic nose. In terms of the search problem, for some gases it will be easier to find materials with useful adsorption behaviors by examining fewer MOFs at more pressures, rather than more MOFs at a single pressure. Similarly, in terms of array performance, using small arrays at an optimized pressure or set of pressures is more beneficial than using large arrays at a single unoptimized pressure. On this note, since methane, hydrogen, and hydrogen sulfide benefit specifically from high pressures, there is only a marginal improvement in the performance of the all-pressure arrays when compared to the single-pressure arrays operating at high pressures. However, this does not mean that there is never any benefit to operating at multiple pressures.

With benzene sensing, most of the MOFs at atmospheric and high pressures saturate at very low concentrations, making detection of benzene beyond these concentrations practically

impossible. By shifting to lower pressures, however, saturation occurs at higher benzene concentrations thus enabling detection. Given this, one might expect benzene to benefit specifically from low pressure sensing, just as hydrogen and methane benefited specifically from high pressures, but NU-100 exhibits unique behavior (**Figure S#)**. It does not saturate at low benzene concentrations until operating at a pressure of 10 bar. At 5 bar, the change in mass as a function of benzene concentration is sharpest, making sensing at this pressure better than either low or atmospheric pressures. In fact, the only single element that outperforms NU-100 at 5 bar is IRMOF-1 at 0.1 bar.[110] As a result, there is a noticeable improvement in the performance of all-pressure arrays, with the all of the best all-pressure arrays of size 2 or more contain NU-100 and IRMOF-1. Figure 19a and Figure 20a-c shows the ternary probability plot and component probability plots of the best 3-element array at ambient pressure (1 bar) sensing conditions, respectively. Figure 19b and Figure 20d-f shows the ternary probability plot and component probability plots of the best 3-element array using multiplex sensing conditions, respectively.

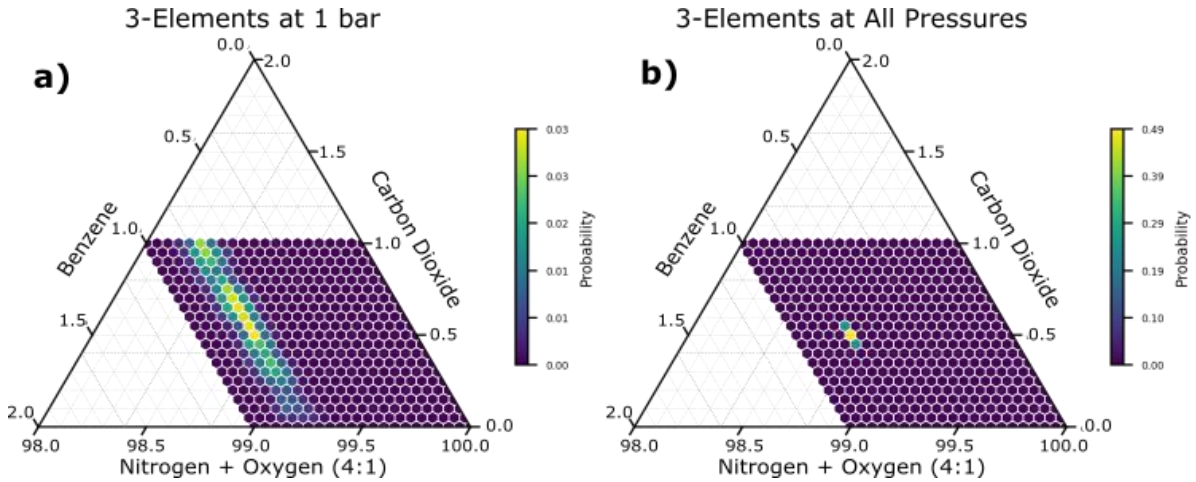

Figure 19. Probability vs. composition for a) the best 3-element array at 1 bar (NU-100, MOF-177, HKUST-1) and b) the best 3-element array at all pressures (NU-100, IRMOF-1, HKUST-1).

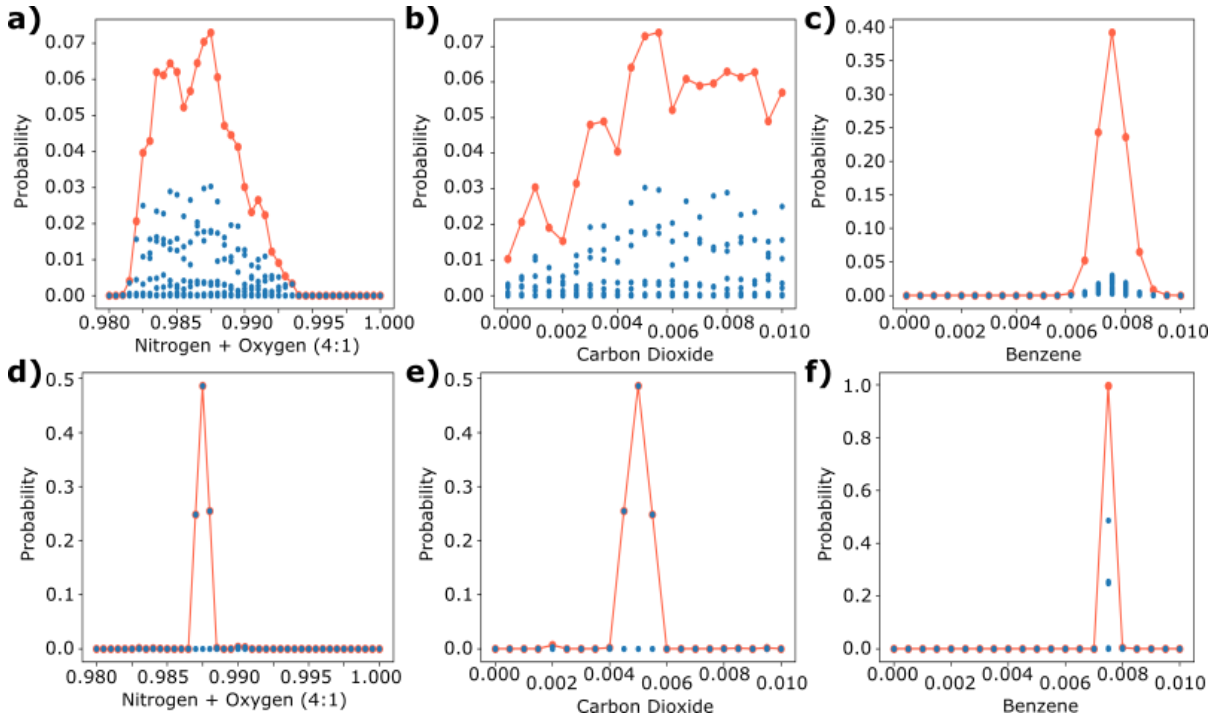

Figure 20. Probability vs. component mole fraction for the best 3-element array at 1 bar (NU-100, MOF-177, HKUST-1) for a) nitrogen/oxygen, b) carbon dioxide, and c) benzene and for the best 3-element array at all pressures (NU-100, MOF-177, HKUST-1) for d) nitrogen/oxygen, e) carbon dioxide, and f) benzene. Blue dots are the probability of individual compositions, and the red line is the total probability for a component concentration.

It is clear from these figures that, just by sampling a few additional pressures, we can dramatically improve the ability to detect gases. Although the 1 bar array does a decent job of detecting benzene, there is still a wide margin of error and the prediction for air and carbon dioxide is very poor. With multiplex sensing, we can narrow down the prediction to almost a single composition.

While it is interesting that benzene had MOFs which performed best at both low and high pressures, and consequently benefits most noticeably from multiplex sensing, this is not a necessary feature of an *individual* component for multiplex sensing to be beneficial; It is simply

important that the *mixture* benefits from multiple pressures. With most real gas mixtures being more complex than these, there will certainly be applications where gases benefiting from low pressures and gases benefiting from high pressures are present simultaneously, such as a system containing benzene and methane in air, which is relevant in natural gas processing.

4.4 Conclusions

For all gas mixtures, designing and operating electronic noses at non-atmospheric pressures modified the adsorption behaviors of the MOFs and resulted in better overall performance. Hydrogen, methane, and hydrogen sulfide benefitted specifically from high pressures for the 9 MOFs screened, whereas benzene benefitted mostly from low pressures. However, NU-100, performed best at 5 bar for benzene sensing, and the corresponding arrays performed best when sampling multiple pressures. Most real gas mixtures will contain gases benefiting from both low and high pressures, and it is not necessary that an individual component benefits from multiple pressures for multiplex sensing to be useful. Rather, it is just important that the mixture benefits.

In general, low-pressure operation seems to benefit strongly adsorbing gases which easily saturate sensors, and high-pressure operation seems to benefit dilute or weakly adsorbing gases. By exploring and operating at multiple pressures, it is easier to find useful MOFs candidates and to design cross-sensitive arrays. We have demonstrated an improvement in the sensing capabilities of the electronic noses while limiting the number materials, making device fabrication cheaper and easier. Looking forward, we plan to combine our coefficient-based method for dilute gas adsorption with the multiple pressure sampling introduced here.

5.0 Updated Approaches to Dilute Gas Species Detection for Breath Analysis

5.1 Introduction

Although the coefficient-based approach to evaluating the adsorption of complex gas mixtures was a notable improvement in our ability to analyze such systems, there were two critical limitations to the approach. The first limitation was the inclusion of CO_2 as a trace gas species, despite its presence in breath of concentrations close to 5%. The second limitation was in what was considered a dilute or non-competitive adsorption system. We required that the total concentration of all trace gas species was below a certain mole fraction, and that all MOFs were in the linear non-competitive adsorption regime for all trace gases on an individual basis up to that same mole fraction. As an example, by our method, we could not use a MOF which saturated with ammonia at concentrations of 3% since CO2 was present in concentrations greater than 3%, despite the fact that in that previous work, ammonia was only present in ppm quantities. Although this was a good assumption in so far as ensuring that all the trace gas species in the mixture were in the dilute adsorption regime, it was unnecessarily restrictive, especially for strongly adsorbing but highly dilute gases.

In order to evaluate more complex gas mixtures such as breath, we wanted to redesign this adsorption coefficient approach and see if we could generate a model with less restrictive assumptions. Again, we decided to split gas mixtures into majority gas species and trace gas species, but this time included CO_2 in the majority gas species group so that the total concentration of the new trace gas species group would almost always be at or below 1%. Already this would help in finding useful MOFs since they would now only need to be in the dilute adsorption regime

for all gases up to concentrations of 1%. However, extra simulations would be required to examine the impact that CO_2 concentrations have on the adsorption coefficients.

Despite the potential improvement of shifting CO_2 to the majority gas species group, we wanted to modify how we determined if a gas mixture was acceptably dilute for a given MOF. Rather than use a universal concentration cutoff, we thought to evaluate mixtures based on the amount of volume the gases would occupy in a MOF after adsorption with respect on their bulk gas concentrations. For example, it is neither practical nor realistic to use the same cutoff for acetone and argon, as they are not only present in very different concentrations in breath, but also exhibit very different adsorption behaviors. A detailed explanation will be given in the methods section below, but to put it generally, we can evaluate how much space there must be per molecule for it to be considered dilute for each molecule type, and then but summing the occupied space for all molecules of a given gas mixture, determine if a MOF has enough volume satisfy the dilute assumption.

5.2 Methods

As an early demonstration of the viability of this method, we examined the adsorption of six trace gas species (acetone, ammonia, argon, hydrogen, isoprene, and methane) in a majority gas species composed of nitrogen and oxygen in a ratio of 4:1, and CO2 at 0%, 1%, 2%, 3%, 4%, and 5% (six unique majority gas mixtures) in HKUST-1 and ZIF-8. We tested this method at pressures of 1 bar, 5 bar, and 10 bar.

As before, in order to evaluate the combined linear adsorption coefficients (CLACs) for the trace gas species, we ran individual sets of GCMC simulations for the trace gas species in each

of the majority gas mixtures, with the trace gas species ranging in concentration for 0 to 1%. We then fit a line to the adsorbed mass of the trace gas only to determine the change in adsorbed mass of that gas as a function of its concentration and the maximum concentration for which that species is considered to be dilute. This same process was done for the adsorbed mass of all the majority gas species to similarly determine the change in adsorbed mass of those species as a function of the trace gas concentration, as well as the baseline adsorption value.

We then ran simulations in which the 6 trace gas species and the 3 majority gas species were present simultaneously. Based on the idea of space filling, we can determine if a mixture satisfies the dilute assumption if the sum of the trace gas mole fraction divided by its maximum allowable mole fraction for all species if less than 1.

$$\sum_{i=1}^{N} \frac{y_i}{y_{i,max}} \leq 1 \tag{4-1}$$

At the maximum allowed mole fraction, all molecules of a given trace gas species are considered to have exactly enough space to non-competitively adsorb within the framework, or to put it from a different perspective, the framework has exactly enough volume to non-competitively accommodate the trace gas. By extension, at concentrations less than the maximum allowable mole fraction, the framework has additional volume to accommodate other trace gases. Specifically, since adsorption the non-competitive regime is linear, we can thus view the ratio between the concentration of a given trace gas species and its maximum allowable mole fraction as a measure of the percent of the volume it occupies in the framework. Now, as long as all the trace gas species occupy less than 100% of the available volume, we can consider the gas mixture to satisfy the dilute assumption. For strongly adsorbing gases which saturate MOFs at relatively low concentrations, as long as the concentration of that species in the bulk gas is sufficiently low, which

for compounds in breath it often is, the MOF will still be considered useable whereas before it would be excluded from further analysis.

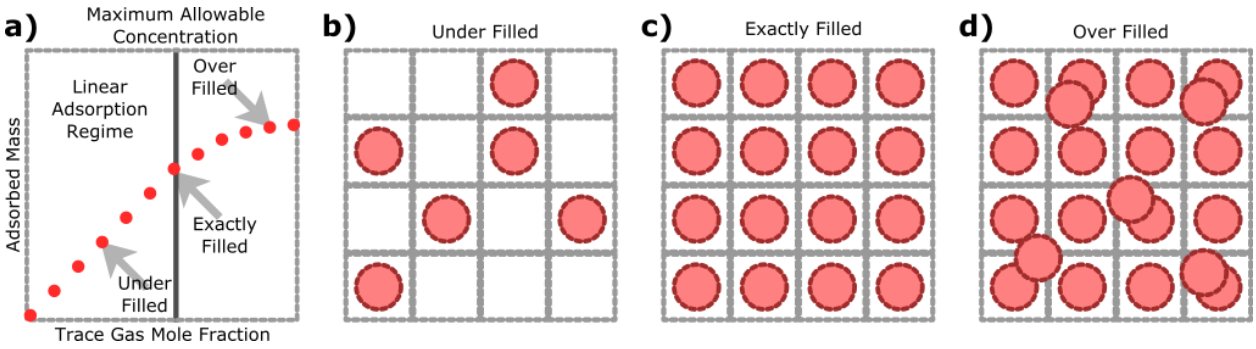

Figure 21. Demonstration of the concept of space filling by trace gases in a framework as a function of (a) trace gas concentration. In (b) the MOF is underfilled and the trace gases adsorb non-competitively, in (c) the MOF is exactly filled, however the gases still adsorb non-competitively, and in (d) the MOF is over filled, and the gases adsorb competitively and non-linearly. The smaller grey boxes represent equal volumes of framework containing majority gas species.

Figure 21 demonstrates this concept for a single gas species. Assuming the grey boxes represent equal volumes of the framework, at concentrations below the maximum allowable concentration, a MOF is considered under filled, at the maximum allowable concentration it is exactly filled, and beyond that over filled and no longer adsorbing linearly. This concept is extended to gas mixtures, with the individual molecules occupying as much volume as needed. Note that since the maximum allowable concentration is defined as the bulk gas concentration which we are trying to measure, this method implicitly accounts for the selectivity of the MOFs.

5.3 Results & Discussion

Figure 22 shows the simulated versus predicted adsorption values for a multicomponent gas mixture in ZIF-8, ordered from lowest to highest CO2 mole fraction, and then secondly by the concentration of the trace gas species for that subplot (thus, the ordering for each subplot is different). The vertical grey lines highlight gas mixtures in which the dilute assumption is not satisfied. For gases such as acetone or isoprene, note that the predicted and simulated values only start to diverge at high concentrations, when the dilute assumption is broken. For weakly adsorbing gases such as argon, hydrogen, or methane, the model seems to predict well, but interestingly, simulations fail to capture dilute adsorption at very low concentration, at least within the number of cycles used in this work, due to a very low likelihood of trying an insertion move for these gases, evidenced by the lack of simulated data. On the other hand, the coefficient-based method is able to readily predict the adsorbed mass of those species, without suffering from a sampling issue. Longer GCMC simulation should be run to validate the predicted adsorption at these very dilute concentrations, but assuming they are accurate, these gases represent another advantage of coefficient-based approaches overcoming sampling issues which can otherwise only be addresses by running much longer simulations. Finally, the large error bars present in the simulated data for some of these gases, in particular ammonia, highlights a similar advantage. Coefficient-based approaches enable prediction of the adsorption of complex gas mixtures with less error, again a problem that is normally overcome by running longer simulations.

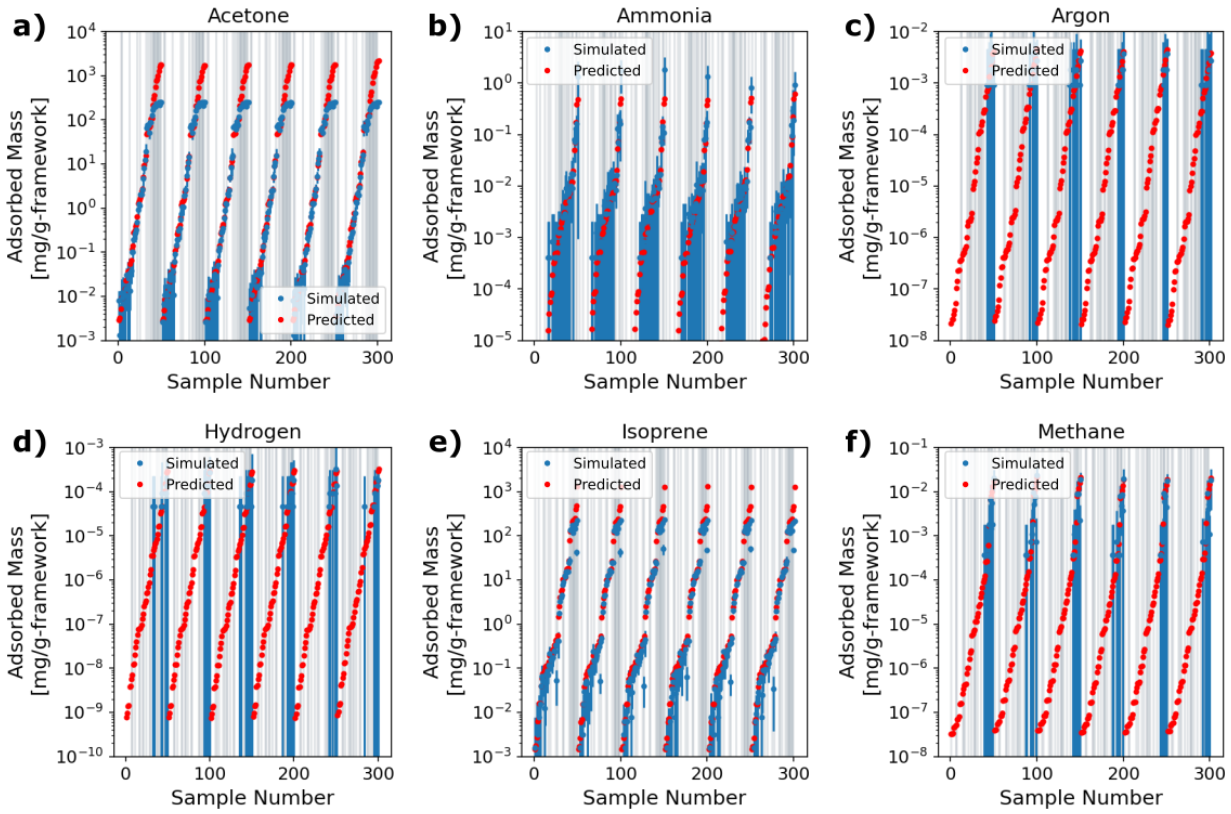

Figure 22. Predicted (red) versus simulated (blue with errorbars) component adsorbed masses for multicomponent gas mixtures in ZIF-8 at 1 bar. Vertical lines are gas mixtures which do not satisfy the dilute assumption. The samples are ordered firstly by the concentration of CO2 in the sample (0%, 1%, 2%, 3%, 4%, and 5%), and secondly by the respective trace gas, that is by (a) acetone, (b) ammonia, (c) argon, (d) hydrogen, (e) isoprene, and (f) methane. Note the logarithmic scale on the y-axis.

Figures 23 and 24 show the total adsorbed mass of the same gas mixtures in both ZIF-8 and HKUST-1 at 1 bar, 5 bar, and 10 bar. Again, the coefficient-based model only seems to deviate from the simulated values for mixtures in which the dilute assumption is not satisfied. Furthermore, these plots suggest that the method holds for different MOFs and at various pressures, so long as specific adsorption coefficients have been calculated.

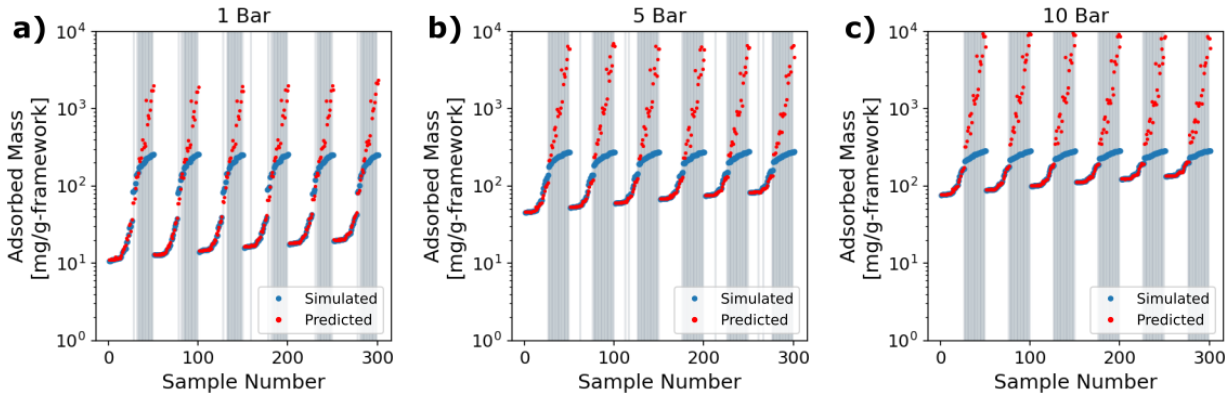

Figure 23. Predicted (red) versus simulated (blue with errorbars) total adsorbed masses for multicomponent gas mixtures in ZIF-8 at (a) 1 bar, (b) 5 bar, and (c) 10 bar. Vertical lines are gas mixtures which do not satisfy the dilute assumption.

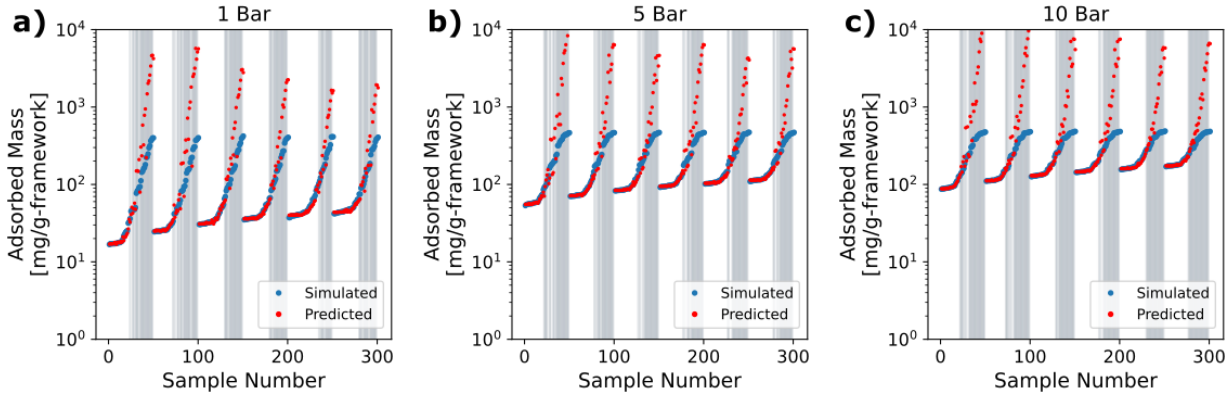

Figure 24. Predicted (red) versus simulated (blue with errorbars) total adsorbed masses for multicomponent gas mixtures in HKUST-1 at (a) 1 bar, (b) 5 bar, and (c) 10 bar. Vertical lines are gas mixtures which do not satisfy the dilute assumption.

5.4 Conclusions

Although this is only preliminary data for two MOFs, these early results suggest a remarkable improvement in our ability to predict the adsorption of multicomponent dilute gas

mixture and address related combinatoric challenges. While it is possible that some gas mixtures of interest will exhibit competitive adsorption in certain MOFs, the space-filling approach to evaluating dilute conditions enables us to consider more MOFs for arrays than we would have previously. Next steps using this approach could include designing arrays in which certain MOFs in arrays are 'turned off' (the output is ignored) for gas mixtures in which they do not satisfy the dilute assumption.

Additionally, since we now fit CLACs that are a function of the majority gas species composition, specifically CO_2 concentration, we cannot solve for the composition analytically, and the algorithm outlined in section 3 will be necessary to evaluate gas mixtures. Array design using the singular value decomposition may also need to change, but if the coefficients do not change significantly as a function of the majority gas species, we could use the same approach and simply use a subset of the adsorption coefficients to design array. Regardless of the change to methods and analyses stemming from this modification to the adsorption coefficient approach, this updated strategy brings us another step closer to building electronic noses for very complex gas mixtures.

6.0 Future Development of MOF-Based Electronic Noses

6.1 Strategies for Accelerating Device Design & Signal Analysis

6.1.1 Machine Learning for High Throughput Screening of MOFs

Despite advances in computing, screening thousands of MOF materials for thousands of gases across many pressures using GCMC is still prohibitively time consuming. However, already throughout the course of this work, we have generated large volumes of simulated adsorption data. As an accelerated approach for high-throughput screening of MOFs, we envision leveraging this data to train a machine learning (ML) model which can approximate the adsorption behavior of MOFs for the purpose selecting materials for further analysis using molecular simulations.

Since the primary goal of this ML model is to act as an initial filter for large libraries of MOFs, it is necessary to use chemical and structural features of the MOFs for analysis. Features which have been used in other ML studies of MOFs include density, void fraction, pore size (pore limiting diameter, largest cavity diameter, and pore size distribution), surface area, topology, structural building units, and bag-of-atoms approaches, to name a few. Practically all of these features can be rapidly calculated from a crystallographic information file of the MOF, and while all of these features may not be necessary, and some may even be redundant, there is clearly a large number of qualitative and quantitative features to use for predicting the adsorption behavior with ML models. Similar feature sets can be generated for the adsorbed gases if necessary.

How to use the output of such a ML model to actually select materials for further analysis is a different question; perhaps we look for MOFs that are especially selective towards only one

gas, or maybe we just select MOFs with the highest predicted adsorption values for a given gas, independent of its response to other gases. Alternatively, consistent with the idea that the best electronic noses should have sensing elements which complement each other, one could also envision using such an ML model not to select single MOFs, but rather small sets of MOFs for detailed analysis based on predicted complementarity.

Regardless of the strategy used, this machine learning approach has the advantage that the adsorption data generated by subsequent analysis can be used to further train the ML model, resulting in a steady improvement over time.

6.1.2 Machine Learning to Predict Adsorbed Masses

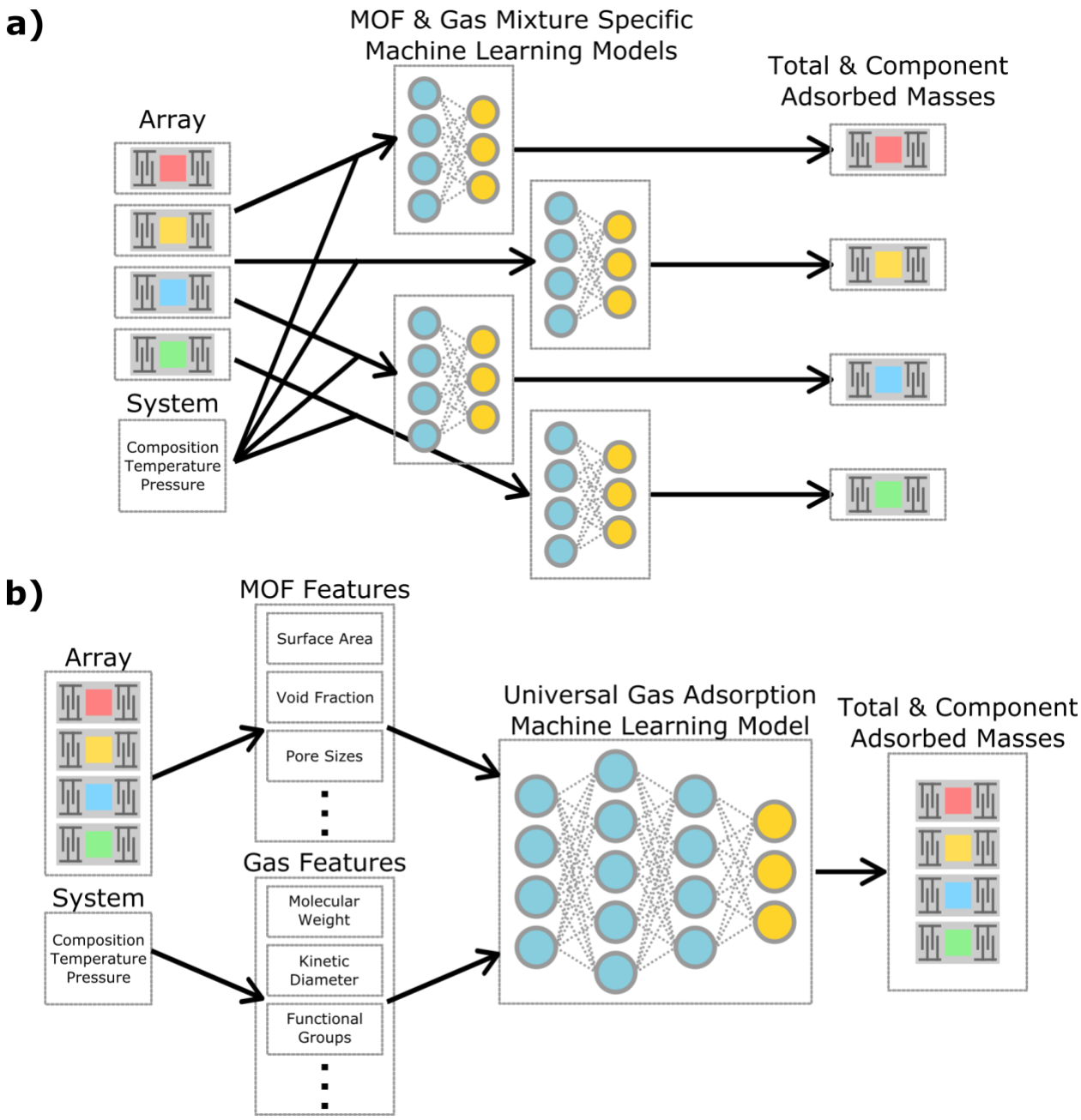

Figure 25. (a) MOF specific Machine learning model used to predict total and component adsorbed mass values from a composition input. (b) Universal gas adsorption machine learning model which uses MOF features, gas features, and composition information to determine total adsorbed mass values.

Another promising area for the employment of ML models is in the analysis of gas mixtures using electronic noses. Although this dissertation included research on simplifying the prediction of the adsorption of complex gas mixtures, this is still an incredibly difficult task requiring a lot of computational time. Furthermore, the method outlined in this dissertation only holds in the absence of competitive adsorption, but this assumption is difficult to guarantee, especially as more MOFs are added to the arrays. Instead, we could keep the concept of dividing complex gas mixtures into majority and trace gas species and generate distinct sets of adsorption data to train ML models for predicting adsorbed mass given only the bulk gas composition, even in instances of competitive adsorption. The model could be further improved with a limited number of simulations containing several trace gases to improve the understanding of competitive adsorption. We may even be able to pull concepts from ideal adsorbed solution theory and use isotherm data and heats of adsorption data to further train the model.

A key advantage of this approach is that the ML models can be built on a per MOF basis, eliminating the need to provide detailed information about the structure. This may also result in a reduction in the amount of training data needed, as well as improved results. Similarly, it would be easy to use just a limited amount of feature data, such as void fraction and pore size, to impose physical constraints to improve the model. Of course, there is also a strong motivation to build a very generalized ML model in which MOF and gas features, bulk gas concentrations, and system parameters like temperature and pressure, can be used as inputs to generate adsorption data, even for gases and MOFs which the model has not yet been trained on, acting as a sort of ‘magic box’ for predicting gas adsorption. Building and training this would likely be a monumental undertaking, but it is certainly a worthy long-term goal with untold benefits. However, regardless of the model type used, what is most important is that by using gas compositions to predict

adsorbed mass values, the generated models are neither array nor application specific, but instead transferable to practically any array for gas sensing application.

Lastly, one could envision developing ML models which tackle the inverse problem: predicting gas compositions from a set of adsorbed masses. However, even though computational data could be used to train such a model, the resulting ML model would almost certainly need to be array specific, meaning that any change in the underlying array would require both new training data and retraining the model. Finally, in a similar train of thought, for applications like disease detection by breath, one could imagine building a model which uses adsorbed mass data to predict disease states, bypassing compositions entirely and effectively fingerprinting the disease state. However, this model would again need to be array specific, and moreover, the training data would need to be clinical in nature. As a result, it would be very expensive and time-consuming, if not outright impossible, to generate enough data to actually train. Furthermore, physicians tend to resist black box diagnostic techniques. For these reasons, I would strongly encourage building some variation of a model which predicts adsorbed masses from composition, and solving the inverse problem with intelligent algorithms, comparable to the one outlined in this dissertation. The resulting compositions could then be used to predict disease states and give a more comprehensive understanding of patient health.

6.2 Strategies for Improving Sensitivity & Selectivity

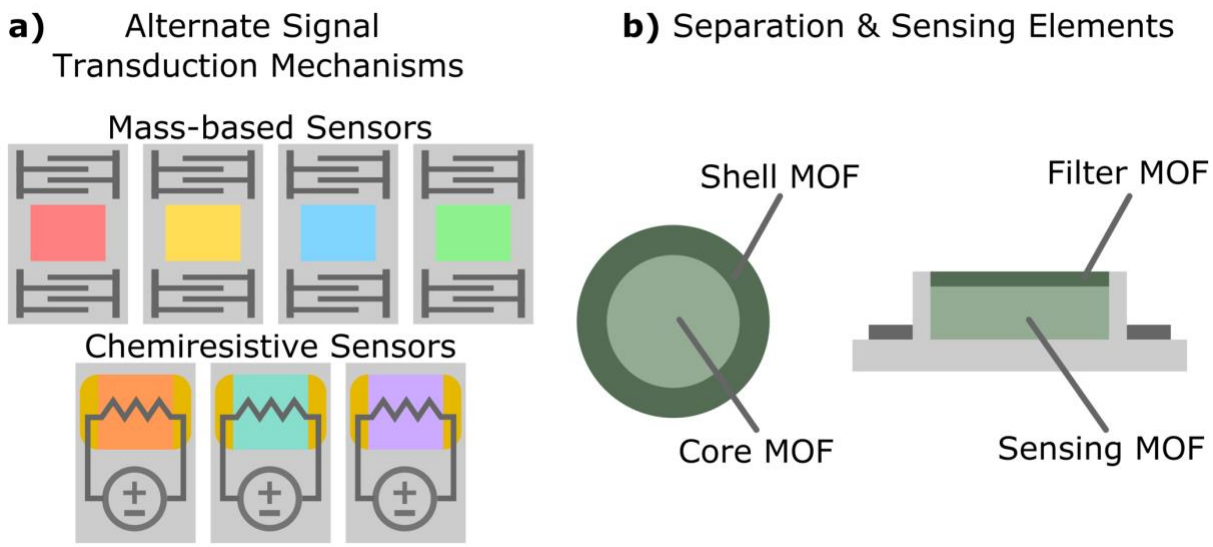

Figure 26. (a) Visual abstract for multiple signal transduction mechanisms in a single electronic nose. (b) Visual abstrat for the sensing elements in an electronic nose, inspired by core-shell MOFs.

While mass-based sensors such as SAW devices and QCMs have impressive sensitivities of nanogram detection limits, there are a number of impressive portable sensing technologies such as chemiresistive and optical sensors which can be used to detect trace gases. Although it is generally harder to predict the response in silico for these types of sensors, they present a great opportunity to improve electronic noses. One can imagine adding these high-sensitivity sensors to arrays for difficult to detect trace gases, or similarly, using them as redundant sensors for gases such as CO_2 to help calibrate the device during operation.

There exist a select few known MOFs with electrical conductivity, but it is certainly a rare phenomenon. Nevertheless, the opportunity to take advantage of the porous nature of MOFs and sensitivity of conductive sensors is a huge opportunity.

6.2.1 Separation & Sensing

Inspired by the use of MOFs in gas separations, in particular core-shell MOFs, a promising approach to improving the selectivity of individual sensing elements would be using MOFs to both separate and sense gases. One could envision a layered structure, in which a thin top layer filters out large gas molecules, and a thicker bottom layer is used to measure the adsorption of the remaining smaller gases, acting as a size-selective filter. Similarly, one could envision a chemical filtering in which the top layer filters out molecules with certain chemistries, eliminating common interferents such as water. While this approach could introduce new challenges, such as accounting for the adsorbed mass of the filter layer, as well as possible time-dependent concerns relating to the diffusion of molecules, it is certainly an interesting approach that could provide useful in detecting particularly hard to detect gases.

6.3 Strategies for Improving Breath Analysis

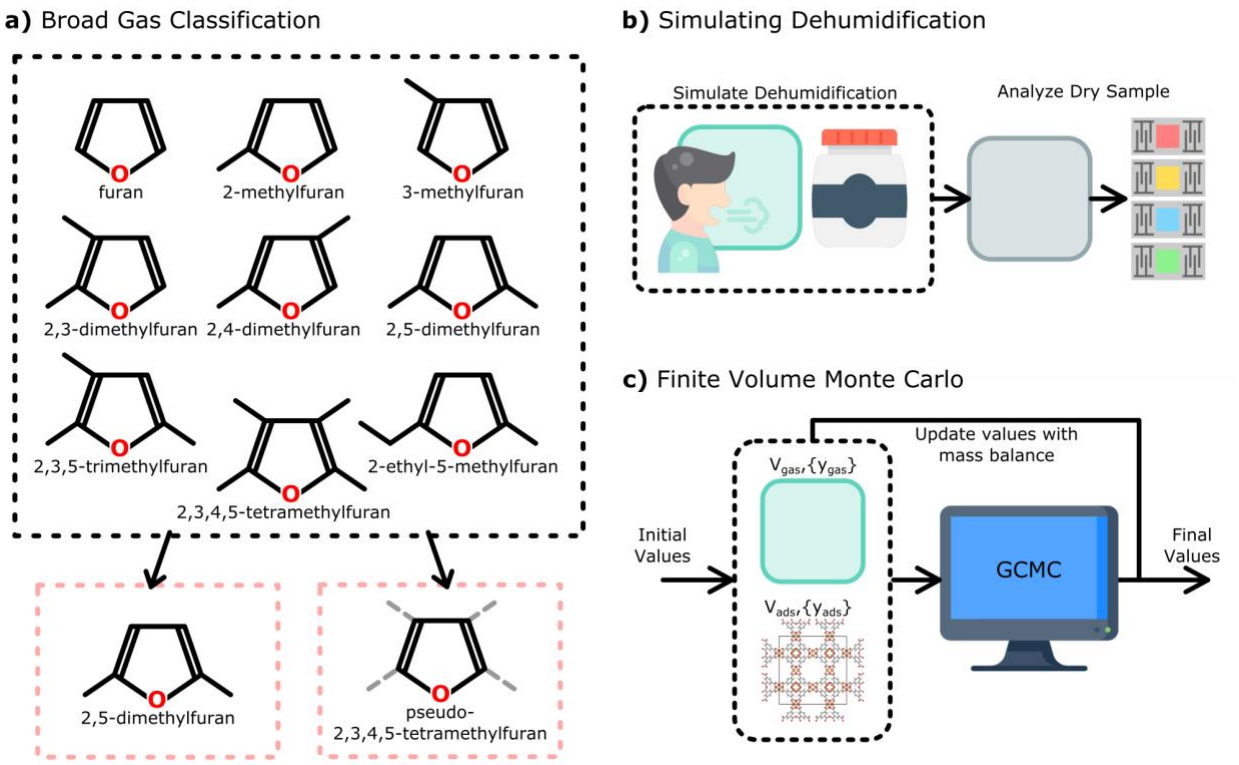

Figure 27. (a) Example approaches to broad gas classification of a furan-type molecules with either a single real molecule representing the set, as shown by 2,5-dimethylfuran on the lower left, or a pseudomolecule with adjusted parameters, as shown by the damped pseudo-2,3,4,5-tetramethylfuran on the lower right. (b) Visual abstract for simulating a dehumidification step for breath samples to limit the impact of humidity on sensing. (c) Visual abstract for an iterative approach to finite volume grand Ccnonical Monte Carlo simulations of gas adsorption in MOFs.

6.3.1 Broad Classification of Gases

As previously mentioned, a single breath sample can contain well over 700 different gases, many of which are exogenous to the body, and thus not biologically relevant. For disease detection applications, it is not imperative to precisely know these compounds or their concentrations,

however it is important that we can minimize their impact on our quantification of biomarkers. Rather than painstakingly enumerating all possible compounds, one strategy for minimizing their influence is broadly classifying similar compounds on the basis of size (e.g., chain length) or chemical features (e.g., functional groups). While the intent of such an approach is to limit the minimum viable array size and streamline analysis, developing such an approach may be difficult. Some MOFs with large pores, for example, may be highly sensitive to functional groups but adsorb gases independently of size. Conversely, one could imagine a MOF with small pores that is very size selective, but minimally effected by chemical features. How to group gas molecules may thus need to be approached on a per MOF basis, complicating analysis.

The best approach may be to go a step beyond grouping similar compounds in the analysis procedure, but actually using a single molecule or pseudomolecule as a placeholder for a small set of molecules when predicting adsorption and analyzing gas mixtures. Obviously, different molecules can and will have different adsorption behaviors, but if the grouping and choice of representative molecule is done carefully, it is possible to dramatically reduce the number of molecules which need to be considered, reining in the combinatorics of the electronic nose problem.

6.3.2 Simulating Dehumidification

In high humidity applications, such as breath analysis, a dehumidification step is likely necessary as many MOFs are susceptible to humidity. However, many dehumidification techniques, such as exposure to silica gel, will also alter the concentrations of the biologically relevant compounds in breath. Thus, it is imperative to understand quantitatively how dehumidification affects the composition such that the initial untreated gas composition can be

correctly calculated, limiting missed or false diagnoses. Fortunately, silica gel is a porous solid material, just like a MOF, and can be simulated using the same techniques, such as GCMC, but special consideration will need to be given since it is non-crystalline. Though complicating the analysis procedure, computational work focused on the pretreatment of a breath sample has enormous potential to improve device performance.

6.3.3 Finite Volume Monte Carlo

When simulating gas adsorption in the grand canonical ensemble, we assume that there is an infinite reservoir of the bulk gas, such that its composition remains constant even when exposed to an adsorbing material (i.e., MOF or silica gel). Given that SAW and QCM devices require very little active material, this is likely a good assumption if a large breath sample is used for direct sampling. However, problems arise if the breath sample is small (Can we guarantee that appreciable volumes of trace gases are not removed from the bulk even small masses of adsorbate?), if the breath samples is stored and then removed from another adsorbent (Do we fully recover the sample without altering the composition?), or if the breath sample if first dehumidified (By definition, we want to alter the composition of the sample to remove water).

There are ensembles for Monte Carlo simulations in which the molecules of the bulk gas are explicitly simulated (i.e., NPT or NVT), however, to accurately simulate trace gas mixtures, one must simulate many more molecules than is computationally practical. That is, at least one million molecules for gases in ppm concentrations, and as many as one trillion molecules for ppt concentrations. GCMC simulations avoid this problem but using equations of state for the bulk gas, returning us to the infinite bulk gas reservoir. To approximate the behavior of finite volumes of bulk gas and adsorbate, one can employ GCMC simulations in an iterative fashion, updating

the composition of the bulk gas species based on the predicted adsorption. The only additional parameters which would need to be specified to do this are the volume of bulk gas and mass of the adsorbent. Then we can perform a simple mass balance to determine the starting concentrations of the bulk gas for the next GCMC simulation. This process can be repeated until convergence is achieved. With this strategy, we can perform accurate simulations for the above limiting cases and improve our analysis using electronic noses.

Appendix A RASPA Simulation Details

All of the adsorption data used throughout this dissertation was generated using RASPA, a grand canonical Monte Carlo simulation software designed by Duddledam et al.[69]

For ternary CO_2-in-air mixtures, simulations were conducted using 1000 initialization cycles and 2000 production cycles. For CLAC gas mixtures and the multiplex sensing work, simulations were conducted using 2000 initialization cycles and 8000 production cycles to improve sampling of low concentration gases. A single cycle consists of n Monte Carlo steps, where n is equivalent to the number of molecules in the simulation. Note that this value fluctuates during a GCMC simulation. The simulations include the following moves: insertion, deletion, translation, regrowth (configuration is changed), and swapping.

To model electrostatic interactions, we assigned partial charges to the atoms of the MOF frameworks via the EQeq method.[70] Similarly, the molecule parameters of the gases also included partial charges, and the forcefield which we used, TrAPPE[71], has been shown to accurately simulate these effects.

Rigid MOF structures, as well as rigid molecule structures, were assumed, and Lennard-Jones (LJ) potentials with a cutoff of 12 Å were used along with Ewald charge interactions to determine the overall energy of the structure and adsorbed gases. The equations for LJ potential are given below, where ε is potential well-depth and σ is radius of interaction.

$$V_{ij} = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \tag{A-1}$$

The equation for Ewald coulombic potential in a periodic system is given as:

$$U^{sys} = U^{real} + U^{rec} \tag{A-2}$$

$$U^{real} = \sum_{i<j} q_i q_j \frac{erfc(\alpha r_{ij})}{r_{ij}} \tag{A-3}$$

$$U^{rec} = \frac{2\pi}{V} \sum_{k} \frac{1}{k^2} e^{-\frac{k^2}{4\alpha^2}} \left(\left| \sum_{i=1}^{N} q_i \cos(\boldsymbol{k} \cdot \boldsymbol{r}_i) \right|^2 + \left| \sum_{i=1}^{N} q_i \sin(\boldsymbol{k} \cdot \boldsymbol{r}_i) \right|^2 \right) - \sum_{i} \frac{\alpha}{\sqrt{\pi}} q_i^2 \tag{A-4}$$

where q_i and q_j are the charges of particle i and j, respectively, $\boldsymbol{r_i}$ is the position of atom i, V is the volume of the cell, α is a damping factor, k is the wavelength, and erfc is the error function complement.

The information about each framework, including minimum number of unit cells, density, volumetric surface area, void fraction, and pore size (largest cavity diameter) are listed below in Table 3. Forcefield parameters (excluding partial charges, which are framework specific and can be found in the CIF) for each framework atom are given in Table 4.

Table 3. Physical properties of MOF structures.

MOF	Min. # of Unit Cells for RASPA [a, b, c]	Density [g/cm^3]	Surface Area [m^2/cm^3]	Void Fraction [---]	Pore Size [Å]
IRMOF-1	1, 1, 1	0.590375	2198.21	0.8108	15.08377
HKUST-1	1, 1, 1	0.879099	2114.54	0.7206	13.18983
NU-125	1, 1, 1	0.57834	2196.18	0.79	19.37323
UIO-66	2, 2, 2	1.22494	1762.62	0.6128	8.88
ZIF-8	2, 2, 2	0.924676	1442.14	0.6416	11.51766
MgMOF-74	1, 1, 4	0.91487	1549.21	0.6396	11.63962
MOF-177	1, 1, 1	0.426775	2035.73	0.8318	11.67849
NU-100	2, 2, 2	0.2843005	1620.675	0.8777	27.190265
MOF-801	1, 1. 1	1.74184	1303.21	0.5322	7.65165
ALUKIC	2, 2, 1	0.56692	2883.21	0.7934	8.54387
AMIMAL	2, 1, 1	0.988926	1269.07	0.6132	11.07211
AXUHEH	2, 2, 1	1.06453	1024.66	0.4958	7.21454
BAZGAM	1, 1, 1	0.126526	810.47	0.9392	42.79818
BIWSEG	1, 1, 1	0.466941	1434.98	0.843	29.73511
EDUVOO	2, 2, 2	0.373403	1788.47	0.862	20.93415
FIDRIV	2, 2, 2	0.698397	1517.46	0.7102	15.99327
GAGZEV	1, 1, 1	0.279149	1594.77	0.8816	28.66522
GUPBEZ	2, 2, 2	2.5399	489.614	0.4518	7.29165
HABQUY	1, 1, 1	0.289452	1646.58	0.8738	25.71531
HIFTOG	2, 2, 2	1.16582	1897.92	0.6126	7.95891
JEWCAP	2, 2. 1	1.11472	880.16	0.5446	6.52252
KICXAX	2, 2, 2	3.58491	368.423	0.3658	5.16268
KIFJUF	3, 2. 2	0.821526	2412.21	0.648	5.86033
KINKAV	3, 2, 2	1.21016	462.625	0.4526	4.49986
LODPUQ	2, 2, 2	1.07351	1496.3	0.5402	6.04771
LOFVUY	2, 1, 1	1.07811	1889.15	0.626	8.04365
MUDTEL	1, 1, 1	0.559282	2154.91	0.789	19.09514
NAYZOE	2, 2, 2	0.498918	2302.42	0.813	15.82314
NIBHOW	1, 1, 1	0.279595	1425.71	0.8844	27.51057
NIBJAK	1, 1, 1	0.223433	1188.94	0.9102	32.00355
OFEREX	3, 3, 2	1.56791	1544.85	0.5694	6.99117
RAVXET	4, 1, 1	0.326773	945.437	0.8304	38.22812
RAVXIX	4, 1, 1	0.23463	734.576	0.8668	53.57674
RAVXOD	4, 1, 1	0.179103	619.991	0.8986	71.64119
RUTNOK	1, 1, 1	0.240823	1468.58	0.9018	24.61263
SADLEQ	3, 3, 2	1.50504	1508.7	0.5702	7.11187
SAPBIW	1, 1, 1	0.305675	915.118	0.889	28.19349
SICZOV	2, 2, 2	0.419881	1773.06	0.8408	18.76086
TOHSAL	2, 2, 1	0.576207	2737.54	0.7668	9.79069
UKUPUL	2, 2, 2	1.43379	1465.65	0.5222	6.90717
VETTIZ	1, 1, 1	0.537597	1117.38	0.7638	21.62456

Table 3 (Continued)

MOF	Min. # of Unit Cells for RASPA [a, b, c]	Density [g/cm^3]	Surface Area [m^2/cm^3]	Void Fraction [---]	Pore Size [Å]
WIYMOG	2, 2, 1	0.408102	2874.44	0.8306	12.0545
WUNSEE01	2, 2, 2	1.20903	749.985	0.4842	5.00556
XAFFAN	2, 2, 2	0.365184	1896.05	0.8544	14.91316
XAFXOT	6, 3, 2	1.88819	647.785	0.3646	5.91185
XAHQAA	1, 1, 1	0.170429	1040.62	0.9292	23.03533
XALTIP	2, 2, 2	0.551216	1809.24	0.7988	18.68299
XUKYEI	2, 2, 2	0.287208	1805.38	0.8682	13.17229
XUWVUG	7, 2, 2	3.19434	197.336	0.287	3.95338
YEQRIV	3, 2, 2	0.74227	3172.25	0.734	5.99633

Table 4. Parameters of framework atoms.

Atom Type	ε/k_B [K]	σ [Å]
H	22.1417	2.886
Be	42.7736	2.44552
B	47.8058	3.58141
C	52.8381	3.851
N	34.7222	3.66
O	30.1932	3.5
F	36.4834	3.092
Na	15.09	2.66
Mg	55.8574	2.69141
Al	155.998	3.91105
Si	155.998	3.80414
P	161.03	3.69723
S	173.107	3.59032
Cl	142.562	3.51932
K	17.61	3.4
Sc	9.56117	2.93551
Ti	8.55473	2.8286
V	8.05151	2.80099
Cr	7.54829	2.69319
Mn	6.54185	2.63795
Fe	6.54185	2.5943

Atom Type	ε/k_B [K]	σ [Å]
Co	7.04507	2.55866
Ni	7.54829	2.52481
Cu	2.5161	3.495
Zn	62.3992	2.46155
Ga	208.836	3.90481
As	155.47	3.77
Br	186.191	3.51905
Zr	34.7221	3.124
Ag	18.1159	2.80455
Cd	114.734	2.53728
In	301.428	3.97608
Sb	225.946	3.93777
Te	200.281	3.98232
I	170.57	4.01
La	8.55	3.14
Ce	6.54	3.17
Nd	5.03	3.18
Eu	4.03	3.11
Tb	3.52	3.07
Dy	3.52	3.05
W	33.71	2.73

Table 5. Simultation parameters of molecular structures.

	Molecule / Atom Type	X [Å]	Y [Å]	Z [Å]	ε/k_B [K]	σ [Å]	Charge [e]
	Acetone						
(0)	CH3_sp3	*Positions defined by flexible bonds			98.0	3.75	0.0
(1)	CH3_sp3				98.0	3.75	0.0
(2)	C_ketone				40.0	3.82	0.424
(3)	O_ketone				79.0	3.05	-0.424
	Ammonia						
(0)	N_NH3	0.0	0.0	0.0757	182.9	3.376	-0.9993
(1)	H_NH3	0.9347	0.0	-0.3164	---	---	0.3331
(2)	H_NH3	-0.4763	0.8095	-0.3164	---	---	0.3331
(3)	H_NH3	-0.4673	-0.8095	-0.3164	---	---	0.3331
	Benzene						
(0)	C_benz	0.0	0.0	0.0	30.70	3.60	-0.095
(1)	C_benz	1.392	0.0	0.0	30.70	3.60	-0.095
(2)	C_benz	2.088	1.2055	0.0	30.70	3.60	-0.095
(3)	C_benz	1.392	2.411	-0.0012	30.70	3.60	-0.095
(4)	C_benz	0.0	2.411	-0.0017	30.70	3.60	-0.095
(5)	C_benz	-0.696	1.2055	-0.0007	30.70	3.60	-0.095
(6)	H_benz	-0.54	-0.9353	0.0004	25.45	2.36	0.095
(7)	H_benz	1.932	0.9353	0.0013	25.45	2.36	0.095
(8)	H_benz	3.1680	1.2055	0.0006	25.45	2.36	0.095
(9)	H_benz	1.932	3.3463	-0.0013	25.45	2.36	0.095
(10)	H_benz	-0.540	3.3463	-0.0026	25.45	2.36	0.095
(11)	H_benz	-1.776	1.2055	-0.0009	25.45	2.36	0.095
	Carbon Dioxide						
(0)	O_CO_2	0.0	0.0	1.16	79.0	3.05	-0.35
(1)	C_CO_2	0.0	0.0	0.0	27.0	2.80	0.70
(2)	O_CO_2	0.0	0.0	-1.16	79.0	3.05	-0.35
	Hydrogen						
(0)	H_H2	0.0	0.0	0.37	36.7	2.958	0.468
(1)	H_e*	0.0	0.0	0.0	---	---	-0.936
(2)	H_H2	0.0	0.0	-0.37	36.7	2.958	0.468
	Hydrogen Sulfide						
(0)	S_H2S	-0.3541	0.2743	-0.4189	122.0	3.60	0.0
(1)	H_H2S	0.9737	0.1714	-0.2619	50.0	2.5	0.21
(2)	H_H2S	-0.6197	-0.4457	0.6808	50.0	2.5	0.21
(3)	H2S_e*	-0.2465	0.1909	-0.2917	---	---	-0.42

Table 5 (Continued)

	Molecule / Atom Type	X [Å]	Y [Å]	Z [Å]	ε/k_B [K]	σ [Å]	Charge [e]
	Isoprene						
(0)	CH2_sp2_isop	*Positions defined by flexible bonds			85.0	3.675	0.0
(1)	C_sp2_isop				22.0	3.85	0.0
(2)	CH_sp2_isop				52.0	3.71	0.0
(3)	CH2_sp2_isop				85.0	3.675	0.0
(4)	CH3_sp3_isop				98.0	3.75	0.0
	Methane						
(0)	CH4_sp3	0.0	0.0	0.0	148.0	3.73	0.0
	Nitrogen						
(0)	N_N_2	0.0	0.0	0.55	36.0	3.31	-0.482
(1)	N_e^*	0.0	0.0	0.0	---	---	0.964
(2)	N_N_2	0.0	0.0	-0.55	36.0	3.31	-0.482
	Oxygen						
(0)	O_O_2	0.0	0.0	0.605	49.000	3.02	-0.113
(1)	O_e^*	0.0	0.0	0.0	---	---	0.226
(2)	O_O_2	0.0	0.0	-0.605	49.000	3.02	-0.113

*X_e = mass-free region of charge

N.B. Lennard-Jones parameters for ammonia are from a paper by Eckl et al.[111]

The Peng-Robinson equation of state, shown below, was used to calculate the fugacities necessary to run the GCMC simulation. The critical parameters for each molecule type are listed below in Table S4.

$$p = \frac{RT}{V_m - b} - \frac{a\alpha}{V_m^2 + 2bV_m - b^2} \tag{A-5}$$

where $a = \frac{0.457235R^2T_c^2}{p_c}$ & $b = \frac{0.077796RT_c}{p_c}$

$$\alpha = \left(1 + k(1 - T_r^{0.5})\right) \tag{A-6}$$

where $k = 0.37464 + 1.54226\omega - 0.26992\omega^2$ & $T_r = T/T_c$

Table 6. Critical parameters of gas molecules.

Molecule Type	T_C [K]	P_C [MPa]	ω	Bond Stretch
Acetone	508.000	5.5300	0.304	Flexible
Ammonia	402.210	10.5200	0.253	Rigid
Argon	154.580	5.0430	0.0	---
Benzene	562.050	4.8940	0.2092	Rigid
Carbon Dioxide	304.128	7.3773	0.22394	Rigid
Hydrogen	33.190	1.3150	-0.214	Rigid
Hydrogen Sulfide	373.400	8.9630	0.09000	Rigid
Isoprene	484.300	3.7896	0.1586	Flexible
Methane	190.564	4.5992	0.01142	---
Nitrogen	126.192	3.3958	0.0372	Rigid
Oxygen	154.581	5.0430	0.0222	Rigid

N.B. The critical constants and the acentric factor for ammonia are from Perry's Chemical Engineering Handbook.

Appendix B Simulating Sensor Measurements & Gas Mixture Analysis

Appendix B.1 Evaluate Element Probabilities

The following section is intended to give an overview of the of the calculations involved in designing arrays and predicting compositions. Specific information about formatting the results and controlling certain parameters of the code is available on the GitHub which hosts this project (https://github.com/WilmerLab/sensor_array_mof_adsorption).

Once these calculations are complete, we create two distinct sets of data; the library of simulated masses, which included the adsorbed mass values for all MOFs and all compositions, and the experimental mass values, which is just a subset of the simulated mass for only a sinlge composition. For work using adsorption coefficients, the simulated mass can actually be calculated as needed. This is outlined further in Appendix D.

The first step in the analysis procedure is to calculate the probability of each composition for each MOF. One MOF at a time, we take the experimental value associated with that MOF and create a truncated normal probability curve centered about the experimental mass, with a standard deviation as either a fixed value or some percent of the experimental mass. The intention of using a truncated probability distribution rather than a true normal distribution is to account for the fact that adsorption will always result in an increase in mass. Consequently, the lower bound is set at 0, and the upper bound is set far beyond the highest simulated mass present in the data set.

The equations which govern the truncated normal distribution are as follows:

$$\psi(\bar{\mu},\bar{\sigma},a,b;x)=\begin{cases}0 & if\ x\ \le a\\ \dfrac{\phi(\bar{\mu},\bar{\sigma}^2;x)}{\Phi(\bar{\mu},\bar{\sigma}^2;b)-\Phi(\bar{\mu},\bar{\sigma}^2;a)} & if\ a<x<b\\ 0 & if\ b\le x\end{cases} \tag{B-1}$$

$$\phi(\bar{\mu},\bar{\sigma}^2;x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{B-2}$$

$$\Phi(\bar{\mu},\bar{\sigma}^2;x)=\int_{-\infty}^{x}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(t-\mu)^2}{2\sigma^2}}dt \tag{B-3}$$

where $\phi(\bar{\mu},\bar{\sigma}^2)$ is the standard normal distribution over the interval $(-\infty,+\infty)$, and $\Phi(\bar{\mu},\bar{\sigma}^2)$ is the cumulative distribution function over the interval $(-\infty,+\infty)$. The variables $\bar{\mu}$ and $\bar{\sigma}$ is the mean and variance of the parent normal distribution, and the variables a and b are the truncation interval.[113]

For each composition, we assign a probability based on where simulated mass sits on the truncated probability curve, as given by:

$$P_{sim,i}=\ \psi\left(\bar{\mu},\bar{\sigma},a,b;m_{sim,i}\right) \tag{B-4}$$

where $\bar{\mu}=m_{exp}$, $\bar{\sigma}=0.05\cdot m_{exp}$, $a=0$, and $b=1.05\cdot m_{sim,max}$. The values used for the standard deviation and upper truncation can be adjusted as needed. Since each mass is assigned a probability independently of each other, the sum of all probabilities does not necessarily equal 1. However, since the intention of this process is to determine which of the simulated compositions the array has been exposed to, we normalize the assigned probabilities so that now their sum equals 1.

$$F=\sum_{i=1}^{N}P_{sim,i} \tag{B-5}$$

$$P_{sim,i}^{norm} = \frac{1}{F} \cdot P_{sim,i} \tag{B-6}$$

This process is repeated for each MOF until we have one normalized probability value for each composition for each MOF.

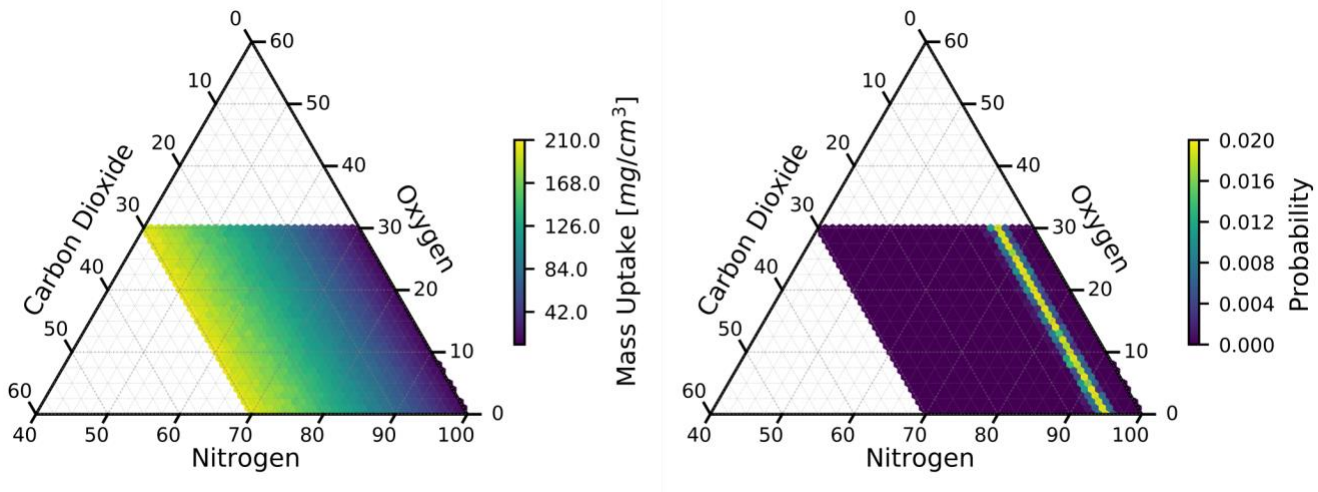

Figure 28. Mapping of Mass uptake to probability as a 1-element sensor of Mg-MOF-74, the top performing 1-element sensor.

Appendix B.2 Evaluate Array Probabilities

Now that we have the probabilities for each MOF, we need to determine the probabilities for arrays. Fortunately, this process is very straightforward. For each composition, we simply multiply all of the normalized probabilities for each MOF with each other, resulting in a non-normalized array probability for each composition. As before, we normalize these probabilities so that they sum to 1. With this information, we can now say which of the simulates gases the array is most likely exposed to.

Appendix B.3 Evaluate Component Probabilities

In addition to calculating the probability for each of the simulated compositions, it is often convenient to be able to predict the mole fraction of each component individually. To this end, we developed a simple approach leveraging the previously calculated probabilities. For whichever component we are trying to predict, we establish a set of bins, typically with the same spacing as the simulated compositions, though this stipulation is not required. Then, for each of the simulated compositions, we assign it to its corresponding component bin. For example, if our bin boundaries for a given CO_2 bin were 29.5 and 30.5, all compositions where the mole fraction of CO_2 was 30% would be placed in that bin. Next, we sum all of the probabilities in that bin to determine the total probability for that bin. Note that since this method uses the already normalized probabilities, the sum of the probabilities for each bin already equals one, and thus no additional normalization is needed.

We can repeat this process for each component in the mixture, until the mole fraction for each component has been predicted individually, though this is not necessary, and may often be undesirable. Nevertheless, it is important to note that if we were to use this approach to individually predict the mole fraction of all components, we are not guaranteed to predict the same composition as we had predicted when considering the mixture as a whole. Furthermore, the sum of each of the mole fractions is not guaranteed to equal one, however both of these scenarios become less likely as the quality of the array improves.

It seems worthwhile to mention the advantages of this approach, as the previously mentioned scenarios would seem to demotivate using it. Notably, one could conceive developing an array which is uniquely sensitive to one primary component (or a set of primary components), and less sensitive to the remaining gases of a typical mixture. It is then conceivable that the array would

continue to predict the primary (set of) component(s) reliably, regardless of how the mole fractions of the remaining components fluctuate. It is additionally possible that the array would continue predicting reliably in the presence of other gases which were not accounted for in the simulations. Conversely, by trying to predict the mixture as a whole, it is foreseeable that in either of these cases, the prediction of the component(s) of interested in negatively impacted by the less important components. Although none of the above situations are guaranteed to hold for all mixtures or arrays, hopefully they demonstrate at least the advantage of having this method available.

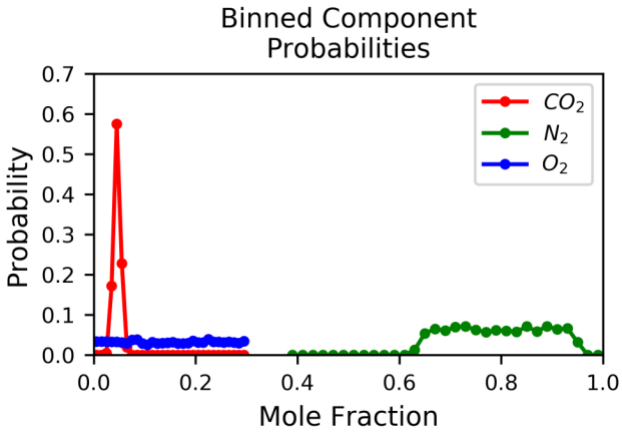

Figure 29. Component-wise probability for MgMOF-74, the top-performing single-element sensor.

Appendix B.4 Ranking Array Performance

Although the set of probabilities for an array enables us to predict the composition of the mixture, it does not lend itself conveniently to comparing the quality of different arrays. Consequently, we wanted a way of quantifying the prediction capabilities of an array, and for this purpose we introduced the Kullbeck-Liebler divergence (KLD).

Rigorously, the KLD quantifies the difference between two probabilities of any form and can be represented mathematically as follows:

$$KLD(P||Q) = \sum_{i=1}^{N} P_i \cdot \log_2\left(\frac{P_i}{Q_i}\right) \tag{B-7}$$

where P and Q are the system and reference probability respectively. When the reference probability is simply a uniform distribution (i.e., random chance), $Q_i = \frac{1}{N}$ for all I, so this simplifies to:

$$KLD = \sum_{i=1}^{N} P_i \cdot \log_2(P_i \cdot N) \tag{B-8}$$

Note that we also drop the $(P||Q)$ notation, since our reference probability is never anything other than a uniform distribution. This form of the equation can be used both when trying to predict the mixture as a whole and when trying to predict the mixture component-wise, the only difference is that the number of points, N, changes. We have taken to calling the these the absolute KLD and component KLD respectively. We also calculate what we have been calling the joint KLD, which is simply the product of all of the component KLDs, though there does not seem to be any advantage to using this in place of the absolute KLD.

Appendix C Array Design

Appendix C.1 Brute Force

In order to determine all possible arrays of a given size, we simply iterate overall available MOFs, repeating this up to the number of elements in the array, and add a MOF only when it has not previously appeared in the array. Once all arrays have been determined, the compound probability, and subsequently the KLD, is evaluated as described above. We can then rank all arrays on the basis of KLD (or component KLD, or any other numerical property of interest), to find the best and/or worst arrays.

Appendix C.2 Genetic Algorithm

With 50 MOFs to choose from, there are over 2.1×10^6 possible 5-element arrays. With an array size of 25 elements, there are over 1.25×10^{14} possible arrays, thus motivating the need for an intelligent screening approach to study these larger arrays. To this end, we developed a genetic algorithm which works in the following way.

Before explaining the details of our genetic algorithm approach, let us first cover some basic terminology. The ‘genetic’ in genetic algorithm refers to the fact that we are using distinct pieces of information about an array to modify it. Here the ‘genes’ correspond to the individual elements in the array. A generation refers to a distinct set of arrays, with a subset of each generation, the parents, being used in creating the following one. In our particular approach, all of the selected

parents are also part of the next generation which they are used to create. This strategy, known as elitism, guarantees that quality of the solution does not decrease between generations. Finally, the individual arrays of the following generation, created from the parent arrays, are known as children.

To begin the algorithmic search, an initial generation of arrays is first created at random (checking to make sure there are no duplicate elements in a single array, and no duplicate arrays in a single generation). Once created, their compound probability, and subsequently KLD, is evaluated as described above. The arrays are then ranked based on the property of interest, typically the KLD or one of the component-KLDs.

In order to create the next generation of arrays, we take a fixed number of the top performing arrays (or bottom two, if seeking the worst performing arrays), along with a fixed number of the remaining arrays at random. These arrays are both part of and parents for the next generation.

There are two approaches which we can use in creating children: crossover and mutation. With crossover, we choose two parents and generate a child from the elements contained in each. With mutation, we choose a single parent and go through each element one at a time. For each element, we generate a random number between 0 and 1. If the number we generate is less than our chosen mutation rate (another number between 0 and 1), than we replace that element with one of the MOFs not currently in the array. If the number we generate is greater than the mutation rate, the element remains in the array. Either one or both of these strategies can be used in creating children for the next generation, however we found that mutation strategies worked best for this application, and thus all of the results presented in the paper use only mutation.

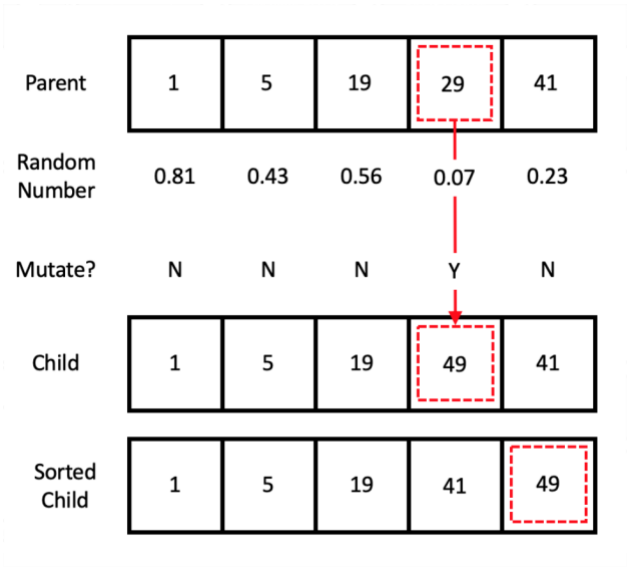

Figure 30. Example decision process for the mutation strategy employed in the genetic algorithm.

This entire process is repeated for the desired number of generations, and typically the genetic algorithm is run multiple times. For the results presented in this paper, the parameters were as follows: 20 arrays per generation, top 2 arrays were used as parents, along with 2 at random, and 200 generations per run. We used a variable mutation rate throughout the process. For the first 25 generations, the mutation rate was 50%, for the next 25 generations it was 25%, the next 50 generations used 10%, followed by another 50 generation at 5%, and lastly 50 generations at 2%. For each array size, the genetic algorithm was run no less than 3 times for seeking both the best and worst arrays, for a minimum of 6 runs.

A flowchart overviewing this process is given below:

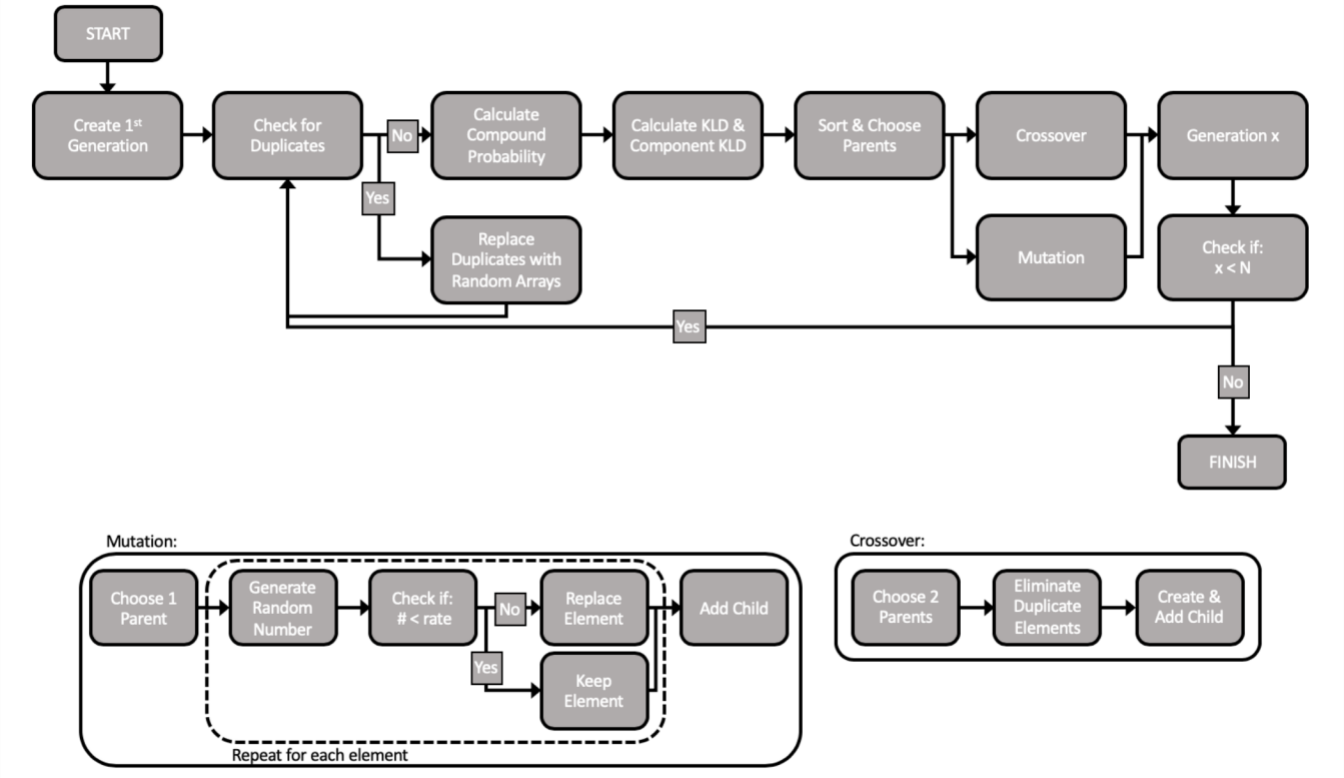

Figure 31. Flowchart overview of the genetic algorithm.

Appendix C.3 Singular Value Decomposition

In past works, array performance was quantified with the KLD score, which examines the difference between any two discrete sets of probabilities. The goal was to ask the question, "How much better are we predicting over random chance?", hence as a reference probability we used a simple uniform distribution.

Although this method worked reasonably well, it had two major limitations: the first is that the system probability was defined with respect to a set of sensor outputs, and thus a different set of sensor outputs would likely result in a different KLD and overall rank; the second is that the sensor

output analysis procedure needed to be run in order to generate the system probabilities, which becomes very time consuming when trying to screen tens of thousands of arrays (and hence the genetic algorithm developed in previous works).

Fortunately, by introducing CLACs, we can screen and rank arrays without the need for a test case and corresponding analysis. In a recent (2019) paper, Sturluson et al. also examined Henry's coefficients of gas adsorption by MOFs with the aim of developing gas sensor arrays, however the key differences being that that their coefficients and arrays were designed for dilute conditions (i.e., low pressures and no air background), and that their analysis procedure is deterministic rather than probabilistic (each of which has its advantages). In their paper, they showed that by performing a singular value decomposition (SVD) on the matrix of Henry's coefficients for each array, one can rank the set of all possible arrays, with the best arrays being those which have the largest minimum sigma value, and the worst being those with the smallest minimum sigma value. Our CLACs are no exception. A brief discussion of the SVD is given below, but for further information, we refer the reader to their paper.

The SVD of the Henry's coefficient matrix, H, can be written as follows:

$$H = U\Sigma V^T \tag{C-1}$$

where H is, again, the Henry's coefficient matrix (or in our case, the CLAC matrix), U is the set of orthogonal unit vectors of sensor response (i.e., mass) space, V is the set of orthogonal unit vectors in gas composition space, and Σ is the matrix of singular values, sigma, which relates sensor response space and composition space. Two key advantages of the SVD are that any matrix can be decomposed without exception, and that the unit vectors of both systems are uniquely determined, which, for the purpose of this work, essentially means that given any set of MOFs and corresponding Henry's coefficients, we can always determine the best array(s).

It is also worth noting that they showed that, with the exception of adding MOFs for which adsorption is no longer in the Henry's regime, adding an additional sensor will always improve the prediction (i.e., increase the value of the smallest singular value, sigma). Thus, for the analyses presented in this work, we chose to limit the array size, since, by their proof, the best array would be the array which uses all applicable MOFs. However, we followed their method exactly for ranking arrays of a given size.

Appendix D Complex Gas Mixture Analysis

Appendix D.1 Combined Linear Adsorption Coefficients (CLACs)

Determining the Henry's coefficients for each of the MOFs/gases is a relatively straightforward procedure. Again, we should note that the CLACS which we evaluate and employ here differ from traditional Henry's coefficients in that they are not for the adsorption of a pure gas species as a function of total pressure, but rather for the adsorption of a trace gas species with a background gas of nitrogen and oxygen as a function of the partial pressure of the trace gas species at a fixed system pressure. This change again begs the question: how does the composition of the background gas impact the predicted CLAC?

To address this, we ran a set of grand canonical Monte Carlo (GCMC) simulations in RASPA in which the mole fraction of the trace gas species was varied from 0-0.05, and the remaining gas was nitrogen and oxygen in each a 3:1, 4:1, and 5:1 ratio. Fortunately, we found that the composition of air (ratio of N_2:O_2) had no impact on the calculated CLAC and thus could be treated as a single component. Knowing this, we then extracted the CLACs from the adsorption data, fitting to all three sets of adsorption data (3:1, 4:1, and 5:1 simultaneously). As mentioned in the paper, we simply fit a line to the adsorbed masses of the trace gas species with the intercept forced to 0, as there is obviously no adsorption of a species not found in the bulk. We then adjusted the width of the dataset, specifically by excluding higher composition data points, until the R-squared value of the fit was greater than 0.95. Then, using the same set of compositions, a line was fit to the adsorbed mass of air, except now with no R-squared cutoff employed, and with the intercept

no longer forced to 0 since there is still adsorbed air in the absence of any trace gas species. By adding the slopes of these fits together, we get the CLACs for the system.

An alternative to the above approach would have been to simply fit a line to the total adsorbed mass and allow for a non-zero intercept. However, there was substantially more nose in the adsorbed massed of nitrogen and oxygen, and thus by fitting to the trace gas species separately, we got a better fit and better result for the pure air mass of each MOF. We did use a weighting scheme to handle the varying amount of error in each of the adsorbed masses. Specifically, weights were assigned inversely proportional to the amount of error. Schemes for non-weighted fitting weighting inversely proportional to squared error were also coded, but not used in this work.

The code which we used is available on GitHub. The table below also provides the CLACs for all MOFs/gases, as well as the pure air mass for each framework.

Table 7. Combined Linear Adsorption Coefficients for all MOFs and Gases.

MOF	K^*, CO_2 [mg / g-framework / mole fraction]	K^*, Argon [mg / g-framework / mole fraction]	K^*, Ammonia [mg/ g-framework / mole fraction]	Pure Air Mass [mg / g-framework]
IRMOF-1	65.764	0.873	33.911	7.882
HKUST-1	756.255	8.189	398.082	16.633
NU-125	235.615	3.571	32.201	14.588
UIO-66	689.234	3.705	17439.557	10.474
ZIF-8	179.071	6.651	155.632	11.019
MgMOF-74	791.004	1.328	2792.297	13.141
MOF-177	77.265	5.239	-13060.475	10.815
NU-100	104.434	7.221	12.689	12.905
MOF-801	58547.931	1.934	None	15.717
ALUKIC	19114.785	7.642	None	14.435
AMIMAL	5878.447	6.746	12208.404	20.13
AXUHEH	1774.947	9.219	148.335	20.482
BAZGAM	56.9	8.326	-0.069	17.355
BIWSEG	1081.739	2.462	None	12.724
EDUVOO	98.813	2.669	38.94	11.95
FIDRIV	2149397.968	6.498	None	21.975
GAGZEV	155.167	6.044	16.623	14.007
GUPBEZ	315.551	5.251	61.999	6.402
HABQUY	132.858	4.624	8.711	13.695
HIFTOG	904.02	8.705	195.056	13.262
JEWCAP	956.107	3.016	214211.898	8.215
KICXAX	167.176	0.398	545.598	1.195
KIFJUF	9640.372	-1.646	None	19.611
KINKAV	1292.102	8.548	142.62	10.15
LODPUQ	1960.705	24.879	99.412	39.135
LOFVUY	2808.085	4.124	None	14.589
MUDTEL	211.765	3.772	39.324	13.773
NAYZOE	102.886	7.502	15.641	12.92
NIBHOW	59.069	2.808	8.07	11.666
NIBJAK	46.734	7.078	3.878	12.214
OFEREX	5091.388	-0.181	None	18.506
RAVXET	550.959	4.414	34011.707	10.935
RAVXIX	199.727	7.671	24.695	15.362
RAVXOD	471.641	7.156	37955.556	14.789
RUTNOK	71.27	3.699	11.804	11.723
SADLEQ	23634.241	-2.752	None	19.7

Table 7 (Continued)

MOF	K^*, CO_2 [mg / g-framework / mole fraction]	K^*, Argon [mg / g-framework / mole fraction]	K^*, Ammonia [mg/ g-framework / mole fraction]	Pure Air Mass [mg / g-framework]
SAPBIW	65.705	3.039	6.274	9.375
SICZOV	254.233	5.34	89899.971	13.549
TOHSAL	401.049	9.461	39.155	21.288
UKUPUL	948237.869	-16.903	None	23.897
VETTIZ	681.192	5.37	151.249	25.703
WIYMOG	62.98	4.175	11.228	9.242
WUNSEE01	2187.015	43.162	177.116	37.272
XAFFAN	114.24	4.641	12453.978	13.083
XAFXOT	55405.68	0.808	None	12.845
XAHQAA	64.245	10.391	7.667	15.282
XALTIP	77.848	4.784	39.476	9.073
XUKYEI	55.169	7.627	9.887	11.779
XUWVUG	28.967	0.411	1.378	0.347
YEQRIV	13991.941	-3.341	None	5.321

*The highlighted MOFs are those which have the necessary Henry's regime (>= 0.05) to be used for the analysis of breath samples.

A more traditional method of calculating Henry's coefficients in GCMC is via Widom insertion, which is an energetic method that returns Henry's coefficients for the adsorption of a pure gas species as a function of system pressure. The primary advantage of the Widom insertion method is speed, however it falls short in other ways. Namely, the Henry's coefficients it returns are not of the type needed for this work. Additionally, it gives us no insight into the width of the Henry's regime, another key metric. Nevertheless, as a test case, we decided to calculate the traditional Henry's coefficients for CO_2, Argon, and Ammonia for NU-100 via Widom insertion and compare them to our CLACs determined via an indirect fitting method.[108] The results comparing the two methods are given in the table below.

Table 8. Henry's Coefficients via Widom insertion vs. CLACs for NU-100 for various gases.

	Widom Insertion (Henry's Coefficient)	Indirect Fitting Method (CLAC)	Percent Difference
CO_2 [mg / g-framework / Pa]	140.2	123.74	11.69
Argon [mg / g-framework / Pa]	19.05	19.02	0.16
Ammonia [mg / g-framework / Pa]	23.21	25.54	10.04

Given that we are comparing non-identical systems, we consider these differences to be acceptable, and further evidence of the minimal impact of air on the adsorption of trace gas species. It is worth noting that the Widom insertion method does provide some energetic information via the heat of adsorption, and thus may be useful in future applications.

Appendix D.2 Simulating Breath Samples

The next thing which we needed to do was generate breath samples. In order to do this, we first generated a set of compositions which reflected a set of both healthy and diseased breath samples, with the compositions of each given as follows:

Table 9. Henry's Coefficients for NU-100 with the direct (traditional) vs. indirect (non-traditional) methods.

Gas	Healthy Range	Diseased Range
CO_2	2-5%	
Argon	0.6-1.2%	
Ammonia	0.49 ± 0.08 ppm	3.32 ± 2.19 ppm
Air	Remainder of Sample, Random ratio of N_2 :O_2 between 3:1-5:1	

*Gases compositions given as a range are chosen at random from a uniform distribution, and gas compositions given as a $\pm$ are chosen at random from a truncated normal distribution.

Initially, we used these compositions to run RASPA simulations in the hope of not making any assumptions about any potential interactions between the trace gas species. Unfortunately, due to the extremely low concentrations of ammonia, an excessively large number of cycles was needed to reduce the error, and even then, the error was often larger than what would be expected of a real device, making the simulation results impractical as a stand in for experimental measurements when trying to detect compounds at ppm concentrations.

As a replacement for simulated breath sample data, we decided to simply generate the masses from the compositions and CLACs, such that the resulting breath sample sensor outputs were "perfect". That is to say that, unless error is introduced, that point in composition space has the highest probability of all possible points. The primary advantage of this choice, aside from

eliminating the need for additional GCMC simulations, is that it allows us to gauge how well behaved the algorithm is. If a quality array cannot accurately predict the composition when the sensor outputs are primed to do so, it is likely that the algorithm (or at least the parameter set) needs to be updated. Additionally, once we are satisfied with the overall algorithm and parameter set for a given array, we can introduce error to the sensor outputs as a way of replicating both measurement error and any unknown trace components, and test how susceptible the system is to these errors and subsequently make adjusts to minimize prediction error.

The set of all compositions for each of the 50 healthy and 50 diseased breath samples is given in the tables below.

Table 10. Healthy Breath Sample Compositions.

Sample Number	Y_{CO2}	Y_{Argon}	$Y_{Ammonia}$	Y_{Air}
1	0.0335	0.0073	3.112×10^{-07}	0.9593
2	0.0395	0.0079	3.428×10^{-07}	0.9526
3	0.0226	0.0100	3.652×10^{-07}	0.9674
4	0.0424	0.0066	3.665×10^{-07}	0.9510
5	0.0201	0.0071	3.847×10^{-07}	0.9728
6	0.0241	0.0099	3.881×10^{-07}	0.9660
7	0.0475	0.0102	4.229×10^{-07}	0.9423
8	0.0285	0.0101	4.239 x 10-^{07}	0.9614
9	0.0414	0.0071	4.286×10^{-07}	0.9515
10	0.0213	0.0062	4.306×10^{-07}	0.9725
11	0.0231	0.0079	4.309×10^{-07}	0.9690
12	0.0322	0.0096	4.340×10^{-07}	0.9582
13	0.0251	0.0085	4.431×10^{-07}	0.9664
14	0.0367	0.0086	4.444×10^{-07}	0.9547
15	0.0400	0.0092	4.459×10^{-07}	0.9508
16	0.0212	0.0105	4.475×10^{-07}	0.9683
17	0.0289	0.0082	4.555×10^{-07}	0.9629
18	0.0486	0.0092	4.613×10^{-07}	0.9422
19	0.0337	0.0080	4.690×10^{-07}	0.9583
20	0.0471	0.0112	4.747×10^{-07}	0.9416
21	0.0499	0.0095	4.750×10^{-07}	0.9406
22	0.0204	0.0099	4.813×10^{-07}	0.9697
23	0.0235	0.0061	4.818×10^{-07}	0.9703
24	0.0294	0.0068	4.929×10^{-07}	0.9639
25	0.0334	0.0069	4.955×10^{-07}	0.9597
26	0.0234	0.0113	4.957×10^{-07}	0.9652
27	0.0477	0.0093	5.003×10^{-07}	0.943
28	0.0230	0.0095	5.123×10^{-07}	0.9676
29	0.0319	0.0110	5.133×10^{-07}	0.9571
30	0.0228	0.0071	5.179×10^{-07}	0.9700
31	0.0473	0.008	5.232×10^{-07}	0.9447
32	0.0471	0.0076	5.249×10^{-07}	0.9453
33	0.0454	0.0066	5.302×10^{-07}	0.9480
34	0.0216	0.0100	5.369×10^{-07}	0.9684
35	0.0403	0.0066	5.448×10^{-07}	0.9531
36	0.0316	0.0073	5.474×10^{-07}	0.9611
37	0.0388	0.0108	5.566×10^{-07}	0.9503
38	0.0268	0.0078	5.716×10^{-07}	0.9654

Table 10 (Continued)

Sample Number	Y_{CO2}	Y_{Argon}	$Y_{Ammonia}$	Y_{Air}
39	0.0479	0.0112	5.729×10^{-07}	0.9410
40	0.0397	0.0113	5.741×10^{-07}	0.9491
41	0.0464	0.0082	5.798×10^{-07}	0.9454
42	0.0458	0.0073	5.821×10^{-07}	0.9469
43	0.0466	0.0099	5.994×10^{-07}	0.9435
44	0.0235	0.0096	6.042×10^{-07}	0.9669
45	0.0322	0.0111	6.081×10^{-07}	0.9568
46	0.0308	0.0077	6.135×10^{-07}	0.9615
47	0.0224	0.0094	6.164×10^{-07}	0.9681
48	0.0472	0.0101	6.205×10^{-07}	0.9427
49	0.0331	0.0120	6.303×10^{-07}	0.9550
50	0.0352	0.0061	6.728×10^{-07}	0.9587

*The highlighted breath samples have a total mole fraction for the trace gas species greater than 0.05. Although this means competitive adsorption may begin if these were real samples, for our computer-generated masses, it simply means our assumption becomes that competitive adsorption does not happen within the range of that specific sample.

Table 11. Diseased Breath Sample Compositions.

Sample Number	Y_{CO2}	Y_{Argon}	$Y_{Ammonia}$	Y_{Air}
1	0.0382	0.0098	1.454×10^{-06}	0.9520
2	0.0372	0.0118	1.626×10^{-06}	0.9510
3	0.0494	0.0079	1.706×10^{-06}	0.9427
4	0.0349	0.0076	1.785×10^{-06}	0.9575
5	0.0499	0.0102	2.025×10^{-06}	0.9399
6	0.0446	0.0111	2.057×10^{-06}	0.9443
7	0.0203	0.0095	2.186×10^{-06}	0.9702
8	0.0226	0.0079	2.237×10^{-06}	0.9695
9	0.0410	0.0066	2.285×10^{-06}	0.9524
10	0.0444	0.0109	2.492×10^{-06}	0.9446
11	0.0215	0.0069	2.496×10^{-06}	0.9716
12	0.0226	0.0069	2.526×10^{-06}	0.9705
13	0.0371	0.0091	2.565×10^{-06}	0.9538
14	0.0393	0.0091	2.931×10^{-06}	0.9515
15	0.0340	0.0078	2.932×10^{-06}	0.9583
16	0.0281	0.0102	2.990×10^{-06}	0.9617
17	0.0273	0.0088	3.101×10^{-06}	0.9639
18	0.0234	0.0065	3.120×10^{-06}	0.9701
19	0.0268	0.0076	3.382×10^{-06}	0.9656
20	0.0494	0.0111	3.398×10^{-06}	0.9395
21	0.0350	0.0074	3.682×10^{-06}	0.9575
22	0.0344	0.0068	3.862×10^{-06}	0.9588
23	0.0290	0.0086	3.908×10^{-06}	0.9624
24	0.0493	0.0063	3.918×10^{-06}	0.9444
25	0.0492	0.0080	3.959×10^{-06}	0.9429
26	0.0310	0.0091	3.967×10^{-06}	0.9599
27	0.0478	0.0092	4.126×10^{-06}	0.9430
28	0.0348	0.0086	4.135×10^{-06}	0.9566
29	0.0274	0.0107	4.172×10^{-06}	0.9619
30	0.0281	0.0073	4.245×10^{-06}	0.9647
31	0.0206	0.0087	4.279×10^{-06}	0.9708
32	0.0354	0.0067	4.321×10^{-06}	0.9578
33	0.0213	0.0092	4.487×10^{-06}	0.9695
34	0.0325	0.0109	4.530×10^{-06}	0.9565
35	0.0208	0.0078	4.571×10^{-06}	0.9714
36	0.0490	0.0090	4.658×10^{-06}	0.9420
37	0.0267	0.0062	4.731×10^{-06}	0.9671
38	0.0490	0.0065	4.873×10^{-06}	0.9445

Table 11 (Continued)

Sample Number	Y_{CO2}	Y_{Argon}	$Y_{Ammonia}$	Y_{Air}
39	0.0448	0.0080	4.946×10^{-06}	0.9472
40	0.0324	0.0097	4.950×10^{-06}	0.9579
41	0.0321	0.0082	4.977×10^{-06}	0.9597
42	0.0255	0.0067	5.324×10^{-06}	0.9678
43	0.0438	0.0099	5.383×10^{-06}	0.9463
44	0.0441	0.0073	5.599×10^{-06}	0.9486
45	0.0274	0.0062	5.713×10^{-06}	0.9664
46	0.0203	0.0079	5.879×10^{-06}	0.9718
47	0.0455	0.0103	6.342×10^{-06}	0.9442
48	0.0489	0.0102	6.377×10^{-06}	0.9410
49	0.0421	0.0118	7.177×10^{-06}	0.9461
50	0.0229	0.0068	8.477×10^{-06}	0.9702

*The highlighted breath samples have a total mole fraction for the trace gas species greater than 0.05. Although this means competitive adsorption may begin if these were real samples, for our computer-generated masses, it simply means our assumption becomes that competitive adsorption does not happen within the range of that specific sample.

Appendix D.3 Composition Prediction Algorithm

Finally, we needed to design and create a process for predicting the composition of a gas sample from a set of sensor output data. The method which we developed in this work is heavily influenced from the method used in previous papers, with the most notable changes being this is an iterative approach and that the masses are not calculated directly from a GCMC simulation, but rather from the CLACs.[34,35,72,73] It is this second change which actually enables us to make this an iterative approach, as theoretically the work from previous papers could have been made iterative if we ran a new set of GCMC simulations between each cycle. Nevertheless, we elaborate on the algorithm presented in the paper here.

We provide a brief overview of the entire process, reproducing Figure 9 from the main body of the paper here as Figure 32.

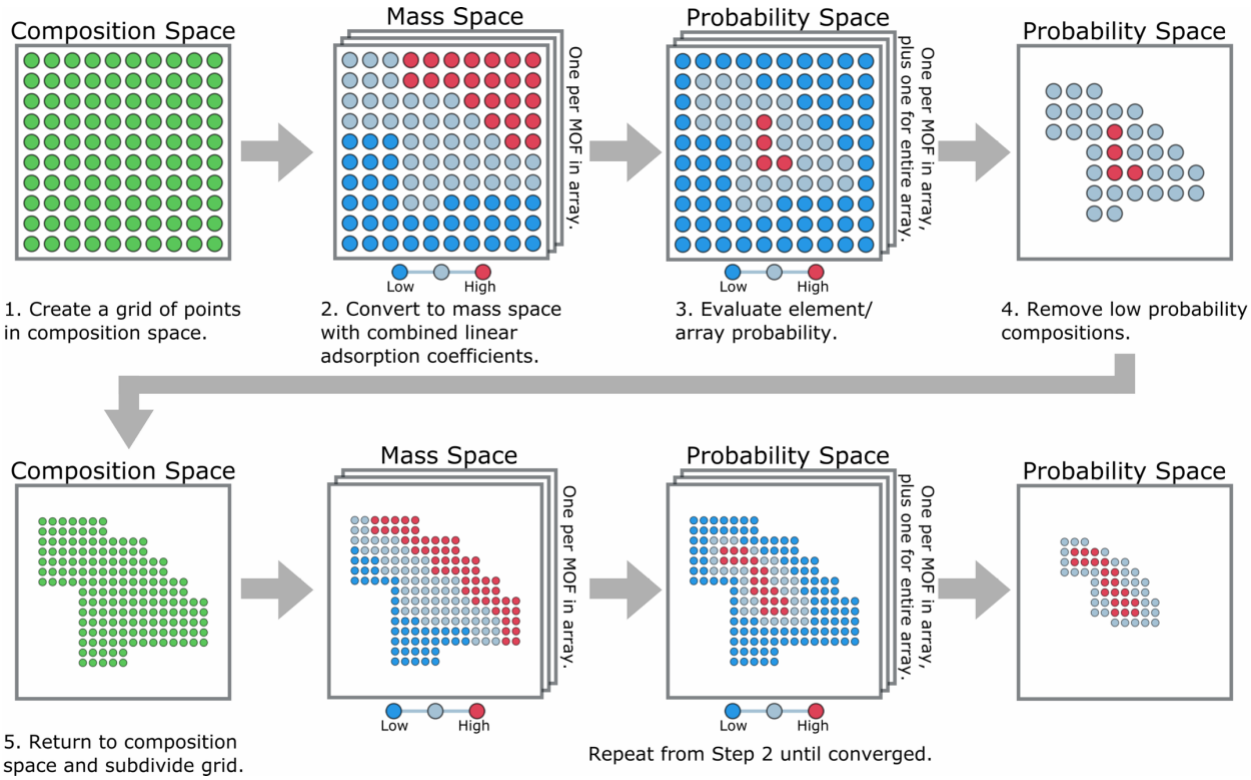

Figure 32. Simplified overview of the algorithm used to predict compositions from a set of sensor data and adsorption coefficients (Copy of Figure 9).

The process begins with the creation of a set of compositions, which are then converted into masses using the CLACs. By comparing the calculated masses to a set of sensor outputs, we assign probabilities to each point, and subsequently filter out the low probability compositions. From there, we return to and subdivide the points in composition space and repeat the process as needed. The code which we used for analysis is also available on GitHub.

Appendix D.3.1 Create Initial Composition Set

The point is largely discussed in the main body of the paper, however one nuance of this point which warrants further discussion is how the initial composition set impacts the prediction. As mentioned, the more finely grained the initial composition space, the better the prediction will be. This is because a fine-grained composition set helps from accidentally filtering out compositions nearest to the answer during early iterations. the A coarsely grained composition set would consequently benefit from keeping more compositions between cycles; however, this can result in an exponential growth in the number of points in the composition space, which in turn can slow down the algorithm and cause memory issues. Thus, a finely grained initial composition set with a small fraction retained is preferrable to the inverse.

Appendix D.3.2 Evaluate Adsorbed Masses

Please refer to main body of the paper.

Appendix D.3.3 Evaluate Element / Array Probabilities

In order to assign probabilities to each point in composition space, we compare the calculated set of masses at that point (one per sensing element) to the set of sensor outputs. This method is the same as is outlined in section S2.1.1, except that composition/mass/probability space is hyperdimensional.

Appendix D.3.4 Filter Low-Probability Compositions

Please refer to main body of the paper.

Appendix D.4 Supplemental Results

Appendix D.4.1 Effect of Array Size and Quality

Although the focus of this paper was on the development of a new methodology, we did also examine the effect of array size and quality on prediction. For this, we first determined a set of parameters via guess-and-check with the best 5-element array. Then using the same set of 50 healthy and 50 diseased breath samples, repeated the analysis with each of the different arrays. All breath samples were "perfect" in that the sensor output masses were calculated using the CLACs and no error was introduced. As a point of clarification, based on the way we assign masses in this work, for a gas sample consisting of N_2, O_2, CO_2, Argon, and Ammonia, a 3-element array would be considered fully determined. This is because N_2 and O_2 are treated collectively as a single component, air, and that the composition of air is implied by the predicted compositions for the other gases in the system, such that the total mole fraction at each point is 1. With this said, we tested the best 1-, 2-, 3-, 4-, 5-, 10-, and 23-element arrays, as well as the worst 5-, and 10-element arrays. Also note that, since only 23 elements met the criteria for this application, there is only one 23-element array, and no larger arrays. For brevity, we show the results for diseased samples only in Figure 33 below.

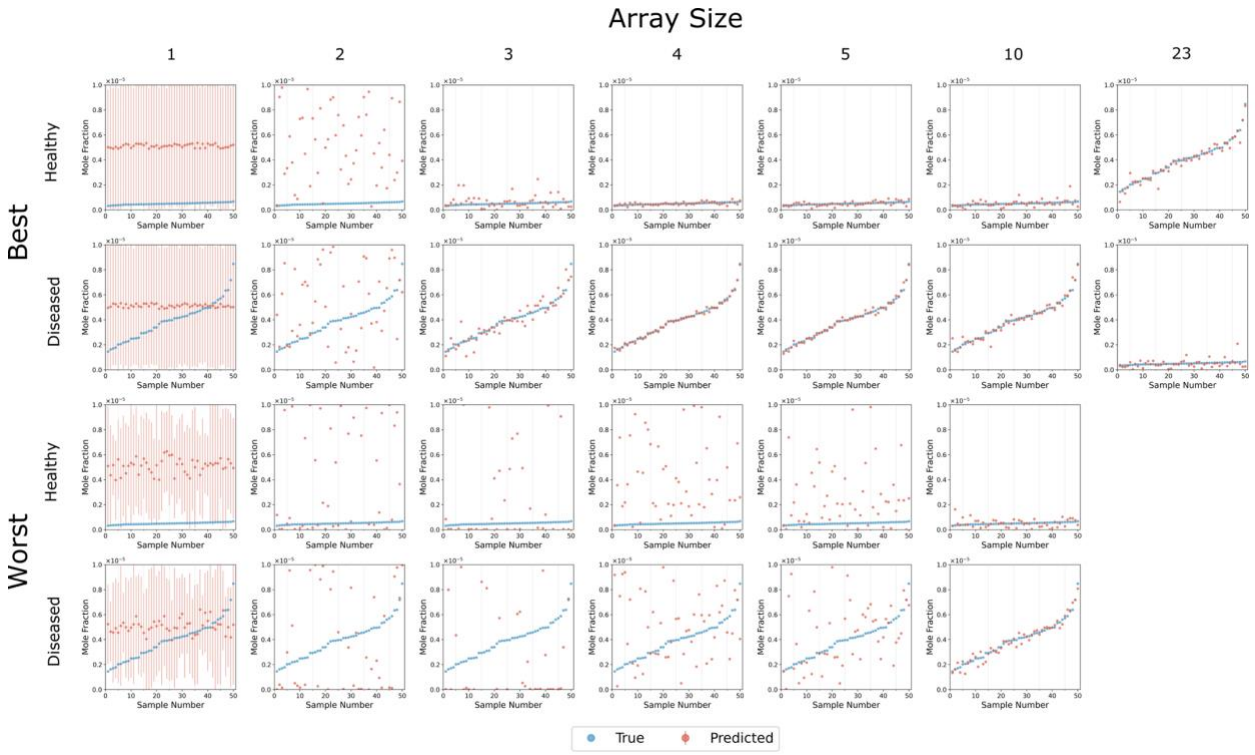

Figure 33. Comparison of the best and worst arrays of various sizes. Note that only the prediction for the concentration of ammonia is shown, as the arrays can reliably predict the concentration of argon and carbon dioxide. Note also that there is only one 23-element

Unsurprisingly, 1- and 2- element arrays struggle to reliably predict the composition of ammonia, with all the 1-element arrays stopping due to reaching the maximum number of cycles. The 2-element arrays converging on a composition, but with a very poor prediction. The 3-element arrays substantially improve prediction quality, but even still, with the chosen parameter set, can lead to poor prediction, such that if this were the result for a real breath sample, there would be several false positive/negatives. The 4-, 5-, 10- and 23- element arrays resolve this problem. The 23-element array is a particularly interesting result, as the prediction quality is noticeably poorer than the best 5- or 10-element arrays, which are both subsets of the best 23-element array. Additionally, since this analysis still uses perfect breath samples, as described in section 4 of this

S.I., the algorithm should be able to predict the composition and we can say confidently that none of the sensors offer contradictory information. Consequently, the poor prediction must be an artifact of the algorithm, and further motivates determining a unique set of parameters for each array/application, not just for each application.

Appendix D.4.2 Effect of Algorithm Parameters

As mentioned, the more finely grained the initial composition space, the better the prediction will be. This is because a fine-grained composition set helps from accidentally filtering out compositions nearest to the answer during early iterations. the A coarsely grained composition set would consequently benefit from keeping more compositions between cycles; however, this can result in an exponential growth in the number of points in the composition space, which in turn can slow down the algorithm and cause memory issues. Thus, a finely grained initial composition set with a small fraction retained is preferrable to the inverse. Figure 34 highlights the effect of initial grid size.

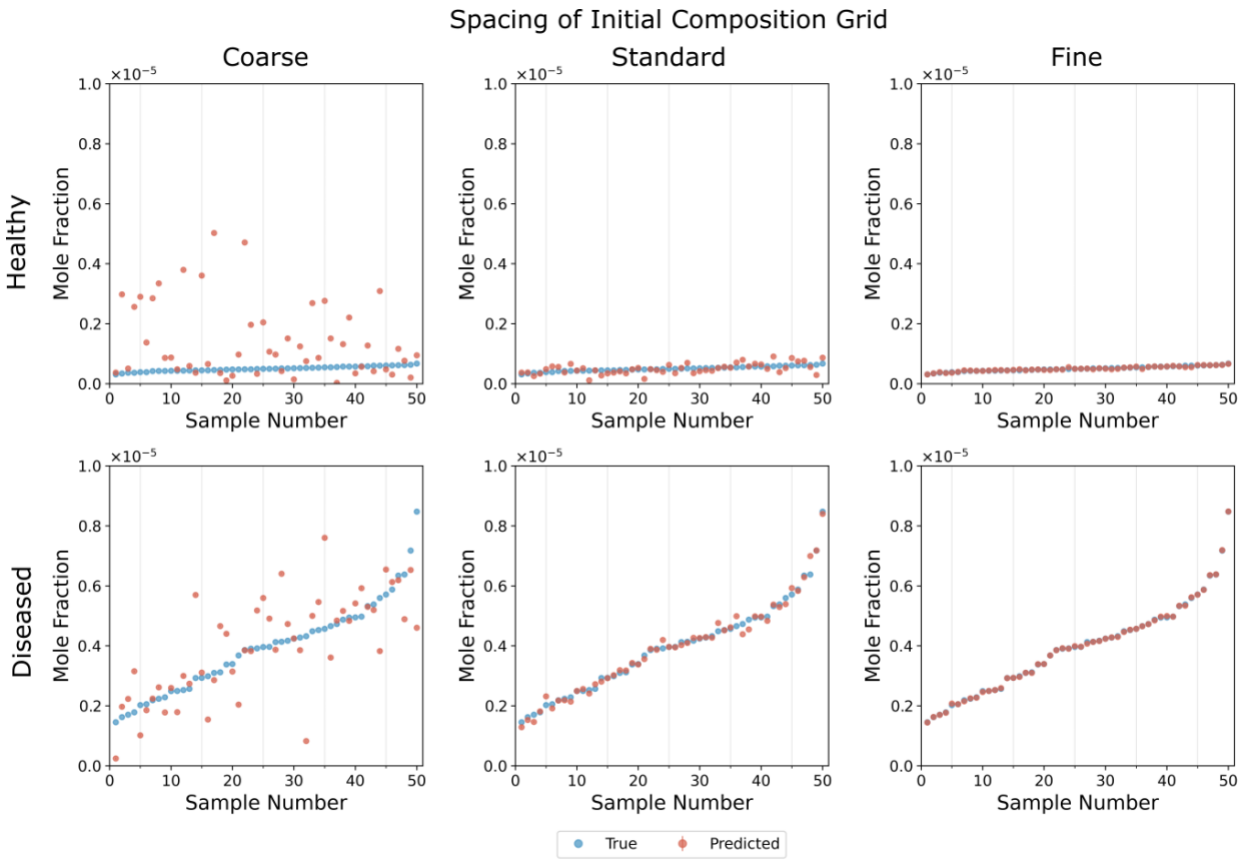

Figure 34. Comparison of the prediction of the ammonia concentration using the best 5-element array as a function of the initial grid spacing. The initial grid for all points spans the following compositions: CO_2: [0.02, 0.05], Argon: [0.0, 0.012], Ammonia: [0.0, 1e-05]. The initial spacing for the coarse grid is CO_2: 2.5e-05, Argon: 0.004, Ammonia: 5e-07; for the standard grid is: CO_2: 1.25e-05, Argon: 0.002, Ammonia: 2.5e-07; and for the fine grid: CO_2: 6.25e-06, Argon: 0.001, Ammonia: 1.25e-07.

Appendix D.4.3 Effect of 'Fraction to Keep'

The fraction of points retained between cycles also impacts the quality of the prediction. Keeping more points improves the prediction but can lead to an explosive growth in the number of compositions considered each cycle, and thus increasing this value should be done carefully, as

it could cause memory issues for the hardware. Figure 35 highlights the effect of fraction of points kept.

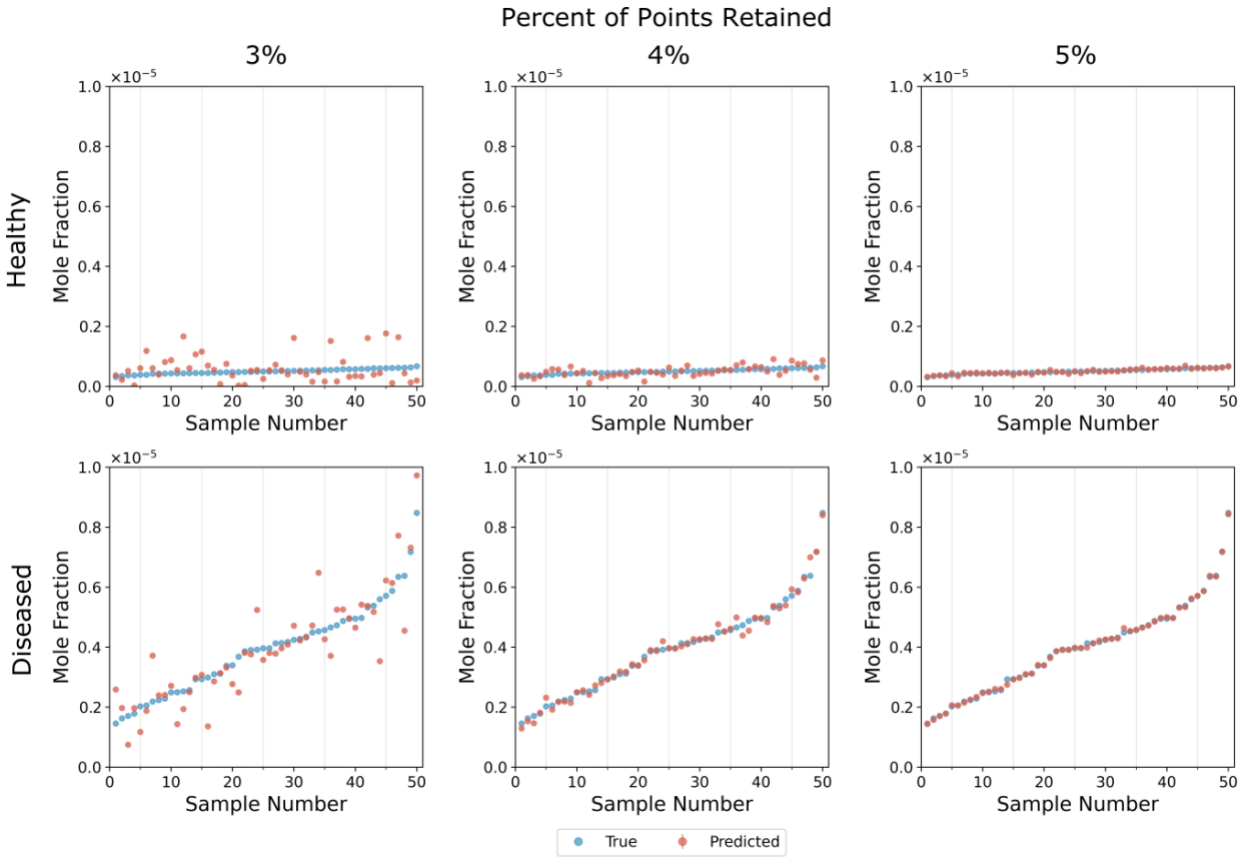

Figure 35. Comparison of the prediction of the ammonia concentration using the best 5-element array as a function of the fraction of points kept.

Appendix E Multiplex Sensing Arrays

The adsorption of the benzene systems in NU-100 was an interesting result in that there did not seem to be a saturation level at lower pressures. Rather, at elevate pressures, there seemed to be a large jump in the total adsorbed mass, following an s-shape if projected down the benzene composition axis. As a result, NU-100 at high pressures improved benzene sensing dramatically. The adsorption of benzene in NU-100 is shown in Figure 36.

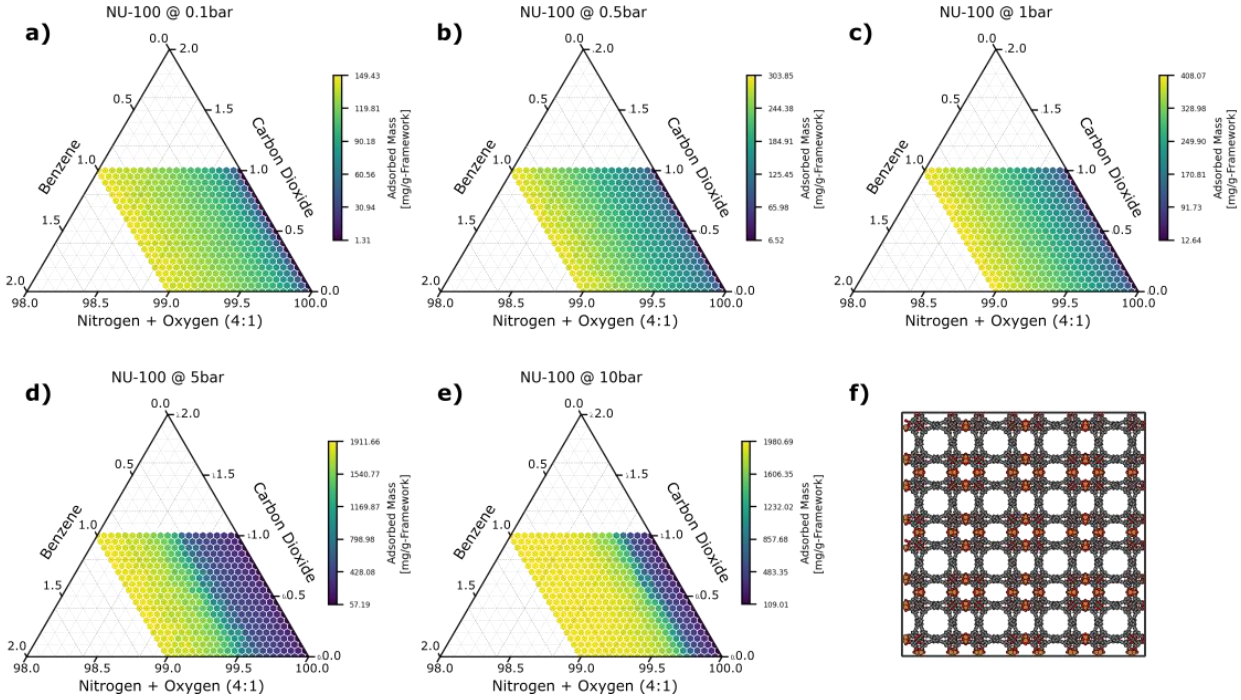

Figure 36. Ternary plots of the adsorbed mass of benzene NU-100 as a function of composition and at the following pressures: a) 0.1 bar, b) 0.5 bar, c) 1 bar, d) 5 bar, and e) 10 bar. f) shows a 2x2x2 unit cell of the MOF projected down the c-axis.

Bibliography

[1] F. Röck, N. Barsan, U. Weimar, Electronic Nose: Current Status and Future Trends, Chem. Rev. 108 (2008) 705–725. https://doi.org/10.1021/cr068121q.

[2] S. Lakkis, R. Younes, Y. Alayli, M. Sawan, Review of recent trends in gas sensing technologies and their miniaturization potential, Sens. Rev. 34 (2014) 24–35. https://doi.org/10.1108/SR-11-2012-724.

[3] J. González-Jiménez, J. Monroy, J.L. Blanco, The Multi-Chamber Electronic Nose—An Improved Olfaction Sensor for Mobile Robotics, Sensors. 11 (2011) 6145–64. https://doi.org/10.3390/s110606145.

[4] A. Amann, W. Miekisch, J. Schubert, B. Buszewski, T. Ligor, T. Jezierski, J. Pleil, T. Risby, Analysis of Exhaled Breath for Disease Detection, Annu. Rev. Anal. Chem. 7 (2014) 455–482. https://doi.org/10.1146/annurev-anchem-071213-020043.

[5] M. Corradi, A. Mutti, Exhaled Breath Analysis: from Occupational to Respiratory Medicine, Acta Bio-Medica Atenei Parm. 76 (2005) 20–29.

[6] C. Di Natale, R. Paolesse, E. Martinelli, R. Capuano, Solid-state gas sensors for breath analysis: A review, Anal. Chim. Acta. 824 (2014) 1–17. https://doi.org/10.1016/j.aca.2014.03.014.

[7] A. Mashir, R.A. Dweik, Exhaled breath analysis: The new interface between medicine and engineering, Adv. Powder Technol. Int. J. Soc. Powder Technol. Jpn. 20 (2009) 420–425.

[8] K.M. Paschke, A. Mashir, R.A. Dweik, Clinical applications of breath testing, F1000 Med. Rep. 2 (2010). https://doi.org/10.3410/M2-56.

[9] Y. Adiguzel, H. Kulah, Breath sensors for lung cancer diagnosis, Biosens. Bioelectron. 65 (2015) 121–138. https://doi.org/10.1016/j.bios.2014.10.023.

[10] E. Aghdassi, J.P. Allard, Breath alkanes as a marker of oxidative stress in different clinical conditions, Free Radic. Biol. Med. 28 (2000) 880–886. https://doi.org/10.1016/S0891-5849(00)00189-1.

[11] G.D. Banik, A. De, S. Som, S. Jana, S.B. Daschakraborty, S. Chaudhuri, M. Pradhan, Hydrogen sulphide in exhaled breath: a potential biomarker for small intestinal bacterial overgrowth in IBS, J. Breath Res. 10 (2016) 026010. https://doi.org/10.1088/1752-7155/10/2/026010.

[12] B. Behera, R. Joshi, G.K. Anil Vishnu, S. Bhalerao, H.J. Pandya, Electronic nose: a non-invasive technology for breath analysis of diabetes and lung cancer patients, J. Breath Res. 13 (2019) 024001. https://doi.org/10.1088/1752-7163/aafc77.

[13] S. Bevc, E. Mohorko, M. Kolar, P. Brglez, A. Holobar, D. Kniepeiss, M. Podbregar, N. Piko, N. Hojs, M. Knehtl, R. Ekart, R. Hojs, Measurement of breath ammonia for detection of patients with chronic kidney disease, Clin. Nephrol. 88 (2017) 14–17. https://doi.org/10.5414/CNP88FX04.

[14] S. Chandrapalan, K. Persaud, R.P. Arasaradnam, Breath diagnostics in the era of SARS-CoV-2—clinical and research arena, J. Breath Res. 14 (2020) 042002. https://doi.org/10.1088/1752-7163/ab924a.

[15] H. Chen, X. Qi, J. Ma, C. Zhang, H. Feng, M. Yao, Breath-borne VOC Biomarkers for COVID-19, MedRxiv. (2020) 2020.06.21.20136523. https://doi.org/10.1101/2020.06.21.20136523.

[16] F.S. Cikach, R.A. Dweik, Cardiovascular Biomarkers In Exhaled Breath, Prog. Cardiovasc. Dis. 55 (2012) 34–43. https://doi.org/10.1016/j.pcad.2012.05.005.

[17] A. Krilaviciute, J.A. Heiss, M. Leja, J. Kupcinskas, H. Haick, H. Brenner, Detection of cancer through exhaled breath: a systematic review, Oncotarget. 6 (2015) 38643–38657.

[18] H.T. Nagle, R. Gutierrez-Osuna, S.S. Schiffman, The how and why of electronic noses, IEEE Spectr. 35 (1998) 22–31. https://doi.org/10.1109/6.715180.

[19] A.D. Wilson, M. Baietto, Applications and Advances in Electronic-Nose Technologies, Sensors. 9 (2009) 5099–5148. https://doi.org/10.3390/s90705099.

[20] R.E. Amor, M.K. Nakhleh, O. Barash, H. Haick, Breath analysis of cancer in the present and the future, Eur. Respir. Rev. 28 (2019). https://doi.org/10.1183/16000617.0002-2019.

[21] T. Chen, T. Liu, T. Li, H. Zhao, Q. Chen, Exhaled breath analysis in disease detection, Clin. Chim. Acta. 515 (2021) 61–72. https://doi.org/10.1016/j.cca.2020.12.036.

[22] M. Gafare, J.O. Dennis, M.H. Md Khir, Detection of ammonia in exhaled breath for clinical diagnosis- A review, AIP Conf. Proc. 1621 (2014) 303–309. https://doi.org/10.1063/1.4898483.

[23] N. Peled, V. Fuchs, E.H. Kestenbaum, E. Oscar, R. Bitran, An Update on the Use of Exhaled Breath Analysis for the Early Detection of Lung Cancer, Lung Cancer Targets Ther. 12 (2021) 81–92. https://doi.org/10.2147/LCTT.S320493.

[24] Lung Cancer Basics, (n.d.). https://www.lung.org/lung-health-diseases/lung-disease-lookup/lung-cancer/basics (accessed July 10, 2022).

[25] P. Quignon, E. Kirkness, E. Cadieu, N. Touleimat, R. Guyon, C. Renier, C. Hitte, C. André, C. Fraser, F. Galibert, Comparison of the canine and human olfactory receptor gene repertoires, Genome Biol. 4 (2003) 1–9. https://doi.org/10.1186/gb-2003-4-12-r80.

[26] T. Onodera, K. Toko, Towards an Electronic Dog Nose: Surface Plasmon Resonance Immunosensor for Security and Safety, Sensors. 14 (2014) 16586–16616. https://doi.org/10.3390/s140916586.

[27] C. Guest, C.M. Otto, Editorial: Canine Olfactory Detection, Front. Vet. Sci. 7 (2020). https://doi.org/10.3389/fvets.2020.00100.

[28] E. Stark, S. Hoover, A. DeCesare, E. Barenholtz, Medicine Has Gone to the Dogs: Deep Learning and Robotic Olfaction to Mimic Working Dogs, IEEE Technol. Soc. Mag. 37 (2018) 55–60. https://doi.org/10.1109/MTS.2018.2876216.

[29] R.P. Arasaradnam, C.U. Nwokolo, K.D. Bardhan, J.A. Covington, Electronic nose versus canine nose: clash of the titans, Gut. 60 (2011) 1768–1768. https://doi.org/10.1136/gut.2011.241216.

[30] D. Grandjean, R. Sarkis, J.-P. Tourtier, C. Julien-Lecocq, A. Benard, V. Roger, E. Levesque, E. Bernes-Luciani, B. Maestracci, P. Morvan, E. Gully, D. Berceau-Falancourt, J.-L. Pesce, B. Lecomte, P. Haufstater, G. Herin, J. Cabrera, Q. Muzzin, C. Gallet, H. Bacqué, J.-M. Broc, L. Thomas, A. Lichaa, G. Moujaes, M. Saliba, A. Kuhn, M. Galey, B. Berthail, L. Lapeyre, O. Méreau, M.-N. Matteï, A. Foata, L. Bey, A.-S. Philippe, P. Abassi, F. Pisani, M. Delarbre, J.-M. Orsini, A. Capelli, S. Renault, K. Bachir, A. Kovinger, E. Comas, A. Stainmesse, E. Etienne, S. Voeltzel, S. Mansouri, M. Berceau-Falancourt, B. Leva, F. Faure, A. Dami, M.A. Costa, J.-J. Tafanelli, J.-B. Luciani, J.-J. Casalot, L. Charlet, E. Ruau, M. Issa, C. Grenet, C. Billy, L. Desquilbet, Detection dogs as a help in the detection of COVID-19 Can the dog alert on COVID-19 positive persons by sniffing axillary sweat samples ? Proof-of-concept study, Animal Behavior and Cognition, 2020. https://doi.org/10.1101/2020.06.03.132134.

[31] E.A. Leighton, E. Hare, S. Thomas, L.P. Waggoner, C.M. Otto, A Solution for the Shortage of Detection Dogs: A Detector Dog Center of Excellence and a Cooperative Breeding Program, Front. Vet. Sci. 5 (2018). https://doi.org/10.3389/fvets.2018.00284.

[32] S.M. Moosavi, A. Nandy, K.M. Jablonka, D. Ongari, J.P. Janet, P.G. Boyd, Y. Lee, B. Smit, H.J. Kulik, Understanding the diversity of the metal-organic framework ecosystem, Nat. Commun. 11 (2020) 4068. https://doi.org/10.1038/s41467-020-17755-8.

[33] J.W. Gardner, P. Boilot, E.L. Hines, Enhancing electronic nose performance by sensor selection using a new integer-based genetic algorithm approach, Sens. Actuators B Chem. 106 (2005) 114–121. https://doi.org/10.1016/j.snb.2004.05.043.

[34] J.A. Gustafson, C.E. Wilmer, Intelligent Selection of Metal–Organic Framework Arrays for Methane Sensing via Genetic Algorithms, ACS Sens. 4 (2019) 1586–1593. https://doi.org/10.1021/acssensors.9b00268.

[35] B.A. Day, C.E. Wilmer, Genetic Algorithm Design of MOF-based Gas Sensor Arrays for CO2-in-Air Sensing, Sensors. 20 (2020) 924. https://doi.org/10.3390/s20030924.

[36] A. Sturluson, R. Sousa, Y. Zhang, M.T. Huynh, C. Laird, A.H.P. York, C. Silsby, C.-H. Chang, C. Simon, Curating Metal-Organic Frameworks to Compose Robust Gas Sensor Arrays in Dilute Conditions, (n.d.). https://doi.org/10.26434/chemrxiv.8179244.v1.

[37] L. Fernández Romero, Understanding the role of sensor diversity and redundancy to encode for chemical information in gas sensor arrays, Ph.D. Thesis, Universitat de Barcelona, 2016. http://www.tdx.cat/handle/10803/395180 (accessed January 9, 2020).

[38] J.W. Gardner, P.N. Bartlett, A brief history of electronic noses, Sens. Actuators B Chem. 18 (1994) 210–211. https://doi.org/10.1016/0925-4005(94)87085-3.

[39] R. Blatt, A. Bonarini, E. Calabro, M.D. Torre, M. Matteucci, U. Pastorino, Lung Cancer Identification by an Electronic Nose based on an Array of MOS Sensors, in: 2007 Int. Jt. Conf. Neural Netw., 2007: pp. 1423–1428. https://doi.org/10.1109/IJCNN.2007.4371167.

[40] Y. Zilberman, R. Ionescu, X. Feng, K. Müllen, H. Haick, Nanoarray of Polycyclic Aromatic Hydrocarbons and Carbon Nanotubes for Accurate and Predictive Detection in Real-World Environmental Humidity, ACS Nano. 5 (2011) 6743–6753. https://doi.org/10.1021/nn202314k.

[41] A. Loutfi, S. Coradeschi, L. Karlsson, M. Broxvall, Putting olfaction into action: using an electronic nose on a multi-sensing mobile robot, 2004 IEEERSJ Int. Conf. Intell. Robots Syst. IROS IEEE Cat No04CH37566. 1 (2004) 337–342 vol.1. https://doi.org/10.1109/IROS.2004.1389374.

[42] D. Karakaya, O. Ulucan, M. Turkan, Electronic Nose and Its Applications: A Survey, Int. J. Autom. Comput. 17 (2020) 179–209. https://doi.org/10.1007/s11633-019-1212-9.

[43] K. Länge, Bulk and Surface Acoustic Wave Sensor Arrays for Multi-Analyte Detection: A Review, Sensors. 19 (2019) 5382. https://doi.org/10.3390/s19245382.

[44] G. Panneerselvam, V. Thirumal, H.M. Pandya, Review of Surface Acoustic Wave Sensors for the Detection and Identification of Toxic Environmental Gases/Vapours, Arch. Acoust. Vol. 43, No. 3 (2018). https://doi.org/10.24425/123908.

[45] D. B. Go, M. Z. Atashbar, Z. Ramshani, H.-C. Chang, Surface acoustic wave devices for chemical sensing and microfluidics: a review and perspective, Anal. Methods. 9 (2017) 4112–4134. https://doi.org/10.1039/C7AY00690J.

[46] J.F. Alder, J.J. McCallum, Piezoelectric crystals for mass and chemical measurements. A review, Analyst. 108 (1983) 1169–1189. https://doi.org/10.1039/AN9830801169.

[47] I. Sayago, M.J. Fernández, J.L. Fontecha, M.C. Horrillo, C. Vera, I. Obieta, I. Bustero, New sensitive layers for surface acoustic wave gas sensors based on polymer and carbon nanotube composites, Sens. Actuators B Chem. 175 (2012) 67–72. https://doi.org/10.1016/j.snb.2011.12.031.

[48] B. Philip, J.K. Abraham, A. Chandrasekhar, V.K. Varadan, Carbon nanotube/PMMA composite thin films for gas-sensing applications, Smart Mater. Struct. 12 (2003) 935–939. https://doi.org/10.1088/0964-1726/12/6/010.

[49] N. Peng, Q. Zhang, C.L. Chow, O.K. Tan, N. Marzari, Sensing Mechanisms for Carbon Nanotube Based NH3 Gas Detection, Nano Lett. 9 (2009) 1626–1630. https://doi.org/10.1021/nl803930w.

[50] C. Cantalini, L. Valentini, I. Armentano, J.M. Kenny, L. Lozzi, S. Santucci, Carbon nanotubes as new materials for gas sensing applications, J. Eur. Ceram. Soc. 24 (2004) 1405–1408. https://doi.org/10.1016/S0955-2219(03)00441-2.

[51] D. J. Wales, J. Grand, V. P. Ting, R. D. Burke, K. J. Edler, C. R. Bowen, S. Mintova, A. D. Burrows, Gas sensing using porous materials for automotive applications, Chem. Soc. Rev. 44 (2015) 4290–4321. https://doi.org/10.1039/C5CS00040H.

[52] K.C. Persaud, Polymers for chemical sensing, Mater. Today. 8 (2005) 38–44. https://doi.org/10.1016/S1369-7021(05)00793-5.

[53] A. Buvailo, Y. Xing, J. Hines, E. Borguet, Thin polymer film based rapid surface acoustic wave humidity sensors, Sens. Actuators B Chem. 156 (2011) 444–449. https://doi.org/10.1016/j.snb.2011.04.080.

[54] J. Gascon, F. Kapteijn, B. Zornoza, V. Sebastián, C. Casado, J. Coronas, Practical Approach to Zeolitic Membranes and Coatings: State of the Art, Opportunities, Barriers, and Future Perspectives, Chem. Mater. 24 (2012) 2829–2844. https://doi.org/10.1021/cm301435j.

[55] K. Sahner, G. Hagen, D. Schönauer, S. Reiß, R. Moos, Zeolites — Versatile materials for gas sensors, Solid State Ion. 179 (2008) 2416–2423. https://doi.org/10.1016/j.ssi.2008.08.012.

[56] X. Xu, J. Wang, Y. Long, Zeolite-based Materials for Gas Sensors, Sensors. 6 (2006) 1751–1764. https://doi.org/10.3390/s6121751.

[57] E. Haldoupis, S. Nair, D.S. Sholl, Pore size analysis of >250 000 hypothetical zeolites, Phys. Chem. Chem. Phys. 13 (2011) 5053–5060. https://doi.org/10.1039/C0CP02766A.

[58] D.M. Ruthven, K.F. Loughlin, K.A. Holborow, Multicomponent sorption equilibrium in molecular sieve zeolites, Chem. Eng. Sci. 28 (1973) 701–709. https://doi.org/10.1016/0009-2509(77)80004-3.

[59] E.-X. Chen, H. Yang, J. Zhang, Zeolitic Imidazolate Framework as Formaldehyde Gas Sensor, Inorg. Chem. 53 (2014) 5411–5413. https://doi.org/10.1021/ic500474j.

[60] O.K. Farha, I. Eryazici, N.C. Jeong, B.G. Hauser, C.E. Wilmer, A.A. Sarjeant, R.Q. Snurr, S.T. Nguyen, A.Ö. Yazaydın, J.T. Hupp, Metal–Organic Framework Materials with Ultrahigh Surface Areas: Is the Sky the Limit?, J. Am. Chem. Soc. 134 (2012) 15016–15021. https://doi.org/10.1021/ja3055639.

[61] Y.G. Chung, J. Camp, M. Haranczyk, B.J. Sikora, W. Bury, V. Krungleviciute, T. Yildirim, O.K. Farha, D.S. Sholl, R.Q. Snurr, Computation-Ready, Experimental Metal–Organic Frameworks: A Tool To Enable High-Throughput Screening of Nanoporous Crystals, Chem. Mater. 26 (2014) 6185–6192. https://doi.org/10.1021/cm502594j.

[62] C.M. Oldenburg, A.J.A. Unger, On Leakage and Seepage from Geologic Carbon Sequestration Sites, Vadose Zone J. 2 (2003) 287–296. https://doi.org/10.2136/vzj2003.2870.

[63] J.L. Lewicki, C.M. Oldenburg, L. Dobeck, L. Spangler, Surface CO2 leakage during two shallow subsurface CO2 releases, Geophys. Res. Lett. 34 (2007). https://doi.org/10.1029/2007GL032047.

[64] K.D. Romanak, P.C. Bennett, C. Yang, S.D. Hovorka, Process-based approach to CO2 leakage detection by vadose zone gas monitoring at geologic CO2 storage sites, Geophys. Res. Lett. 39 (2012). https://doi.org/10.1029/2012GL052426.

[65] S. Srisont, T. Chirachariyavej, A.V.M.V. Peonim, A carbon dioxide fatality from dry ice, J. Forensic Sci. 54 (2009) 961–962. https://doi.org/10.1111/j.1556-4029.2009.01057.x.

[66] K. Permentier, S. Vercammen, S. Soetaert, C. Schellemans, Carbon dioxide poisoning: a literature review of an often forgotten cause of intoxication in the emergency department, Int. J. Emerg. Med. 10 (2017). https://doi.org/10.1186/s12245-017-0142-y.

[67] Satish Usha, Mendell Mark J., Shekhar Krishnamurthy, Hotchi Toshifumi, Sullivan Douglas, Streufert Siegfried, Fisk William J., Is CO2 an Indoor Pollutant? Direct Effects of Low-to-Moderate CO2 Concentrations on Human Decision-Making Performance, Environ. Health Perspect. 120 (2012) 1671–1677. https://doi.org/10.1289/ehp.1104789.

[68] MedlinePlus - Health Information from the National Library of Medicine, (n.d.). https://medlineplus.gov/ (accessed October 8, 2019).

[69] D. Dubbeldam, S. Calero, D.E. Ellis, R.Q. Snurr, RASPA: molecular simulation software for adsorption and diffusion in flexible nanoporous materials, Mol. Simul. 42 (2016) 81–101. https://doi.org/10.1080/08927022.2015.1010082.

[70] C.E. Wilmer, K.C. Kim, R.Q. Snurr, An Extended Charge Equilibration Method, J. Phys. Chem. Lett. 3 (2012) 2506–2511. https://doi.org/10.1021/jz3008485.

[71] M.G. Martin, J.I. Siepmann, Transferable Potentials for Phase Equilibria. 1. United-Atom Description of n-Alkanes, J. Phys. Chem. B. 102 (1998) 2569–2577. https://doi.org/10.1021/jp972543+.

[72] J.A. Gustafson, C.E. Wilmer, Optimizing information content in MOF sensor arrays for analyzing methane-air mixtures, Sens. Actuators B Chem. 267 (2018) 483–493. https://doi.org/10.1016/j.snb.2018.04.049.

[73] J.A. Gustafson, C.E. Wilmer, Computational Design of Metal–Organic Framework Arrays for Gas Sensing: Influence of Array Size and Composition on Sensor Performance, J. Phys. Chem. C. 121 (2017) 6033–6038. https://doi.org/10.1021/acs.jpcc.6b09740.

[74] N.L. Rosi, J. Kim, M. Eddaoudi, B. Chen, M. O'Keeffe, O.M. Yaghi, Rod Packings and Metal−Organic Frameworks Constructed from Rod-Shaped Secondary Building Units, J. Am. Chem. Soc. 127 (2005) 1504–1518. https://doi.org/10.1021/ja045123o.

[75] D. Braga, L. Maini, P.P. Mazzeo, B. Ventura, Reversible Interconversion between Luminescent Isomeric Metal–Organic Frameworks of [Cu4I4(DABCO)2] (DABCO=1,4-Diazabicyclo[2.2.2]octane), Chem. – Eur. J. 16 (2010) 1553–1559. https://doi.org/10.1002/chem.200900743.

[76] G.-L. Zhuang, X.-J. Kong, L.-S. Long, R.-B. Huang, L.-S. Zheng, Effect of lanthanide contraction on crystal structures of lanthanide coordination polymers with 2,5-piperazinedione-1,4-diacetic acid, CrystEngComm. 12 (2010) 2691–2694. https://doi.org/10.1039/C001537G.

[77] D.J.C. MacKay, D.J.C.M. Kay, Information Theory, Inference and Learning Algorithms, Cambridge University Press, 2003.

[78] T. Düren, Y.-S. Bae, R. Q. Snurr, Using molecular simulation to characterise metal–organic frameworks for adsorption applications, Chem. Soc. Rev. 38 (2009) 1237–1247. https://doi.org/10.1039/B803498M.

[79] H. Frost, T. Düren, R.Q. Snurr, Effects of Surface Area, Free Volume, and Heat of Adsorption on Hydrogen Uptake in Metal−Organic Frameworks, J. Phys. Chem. B. 110 (2006) 9565–9570. https://doi.org/10.1021/jp060433+.

[80] A.F. Ismail, K.C. Khulbe, T. Matsuura, Fundamentals of Gas Permeation Through Membranes, in: A.F. Ismail, K. Chandra Khulbe, T. Matsuura (Eds.), Gas Sep. Membr. Polym. Inorg., Springer International Publishing, Cham, 2015: pp. 11–35. https://doi.org/10.1007/978-3-319-01095-3_2.

[81] R. Poloni, K. Lee, R.F. Berger, B. Smit, J.B. Neaton, Understanding Trends in CO2 Adsorption in Metal–Organic Frameworks with Open-Metal Sites, J. Phys. Chem. Lett. 5 (2014) 861–865. https://doi.org/10.1021/jz500202x.

[82] T.-C. Kuo, C.-E. Tan, S.-Y. Wang, O.A. Lin, B.-H. Su, M.-T. Hsu, J. Lin, Y.-Y. Cheng, C.-S. Chen, Y.-C. Yang, K.-H. Chen, S.-W. Lin, C.-C. Ho, C.-H. Kuo, Y.J. Tseng, Human Breathomics Database, Database. 2020 (2020). https://doi.org/10.1093/database/baz139.

[83] L. Zhang, J.I. Siepmann, Development of the trappe force field for ammonia, Collect. Czechoslov. Chem. Commun. 75 (2010) 577–591. https://doi.org/10.1135/cccc2009540.

[84] L. Pauling, A.B. Robinson, R. Teranishi, P. Cary, Quantitative Analysis of Urine Vapor and Breath by Gas-Liquid Partition Chromatography, Proc. Natl. Acad. Sci. U. S. A. 68 (1971) 2374–2376.

[85] T.H. Risby, S.F. Solga, Current status of clinical breath analysis, Appl. Phys. B. 85 (2006) 421–426. https://doi.org/10.1007/s00340-006-2280-4.

[86] W. Filipiak, V. Ruzsanyi, P. Mochalski, A. Filipiak, A. Bajtarevic, C. Ager, H. Denz, W. Hilbe, H. Jamnig, M. Hackl, A. Dzien, A. Amann, Dependence of exhaled breath composition on exogenous factors, smoking habits and exposure to air pollutants, J. Breath Res. 6 (2012) 036008. https://doi.org/10.1088/1752-7155/6/3/036008.

[87] D. Guo, D. Zhang, N. Li, L. Zhang, J. Yang, A Novel Breath Analysis System Based on Electronic Olfaction, IEEE Trans. Biomed. Eng. 57 (2010) 2753–2763. https://doi.org/10.1109/TBME.2010.2055864.

[88] H. Egnér, E. Eriksson, Current Data on the Chemical Composition of Air and Precipitation, Tellus. 7 (1955) 134–139. https://doi.org/10.3402/tellusa.v7i1.8763.

[89] The Composition of Atmospheric Air | SpringerLink, (n.d.). https://link.springer.com/chapter/10.1007/978-1-940033-70-9_1 (accessed May 15, 2020).

[90] J. Canivet, A. Fateeva, Y. Guo, B. Coasne, D. Farrusseng, Water adsorption in MOFs: fundamentals and applications, Chem. Soc. Rev. 43 (2014) 5594–5617. https://doi.org/10.1039/C4CS00078A.

[91] P. Ghosh, K.C. Kim, R.Q. Snurr, Modeling Water and Ammonia Adsorption in Hydrophobic Metal–Organic Frameworks: Single Components and Mixtures, J. Phys. Chem. C. 118 (2014) 1102–1110. https://doi.org/10.1021/jp410758t.

[92] C.E. Ramachandran, S. Chempath, L.J. Broadbelt, R.Q. Snurr, Water adsorption in hydrophobic nanopores: Monte Carlo simulations of water in silicalite, Microporous Mesoporous Mater. 90 (2006) 293–298. https://doi.org/10.1016/j.micromeso.2005.10.021.

[93] A. Nalaparaju, X.S. Zhao, J.W. Jiang, Molecular Understanding for the Adsorption of Water and Alcohols in Hydrophilic and Hydrophobic Zeolitic Metal−Organic Frameworks, J. Phys. Chem. C. 114 (2010) 11542–11550. https://doi.org/10.1021/jp1033273.

[94] C. Wang, X. Liu, N.K. Demir, J.P. Chen, K. Li, Applications of water stable metal–organic frameworks, Chem. Soc. Rev. 45 (2016) 5107–5134. https://doi.org/10.1039/C6CS00362A.

[95] N. Desbiens, A. Boutin, I. Demachy, Water Condensation in Hydrophobic Silicalite-1 Zeolite: A Molecular Simulation Study, J. Phys. Chem. B. 109 (2005) 24071–24076. https://doi.org/10.1021/jp054168o.

[96] G. Peng, E. Trock, H. Haick, Detecting Simulated Patterns of Lung Cancer Biomarkers by Random Network of Single-Walled Carbon Nanotubes Coated with Nonpolymeric Organic Materials, Nano Lett. 8 (2008) 3631–3635. https://doi.org/10.1021/nl801577u.

[97] M. Nishibori, W. Shin, N. Izu, T. Itoh, I. Matsubara, Sensing performance of thermoelectric hydrogen sensor for breath hydrogen analysis, Sens. Actuators B Chem. 137 (2009) 524–528. https://doi.org/10.1016/j.snb.2009.01.029.

[98] C. Grote, J. Pawliszyn, Solid-Phase Microextraction for the Analysis of Human Breath, Anal. Chem. 69 (1997) 587–596. https://doi.org/10.1021/ac960749l.

[99] K.S. Park, Z. Ni, A.P. Côté, J.Y. Choi, R. Huang, F.J. Uribe-Romo, H.K. Chae, M. O'Keeffe, O.M. Yaghi, Exceptional chemical and thermal stability of zeolitic imidazolate frameworks, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 10186–10191. https://doi.org/10.1073/pnas.0602439103.

[100] J. Hafizovic, M. Bjørgen, U. Olsbye, P.D.C. Dietzel, S. Bordiga, C. Prestipino, C. Lamberti, K.P. Lillerud, The Inconsistency in Adsorption Properties and Powder XRD Data of MOF-5 Is Rationalized by Framework Interpenetration and the Presence of Organic and Inorganic Species in the Nanocavities, J. Am. Chem. Soc. 129 (2007) 3612–3620. https://doi.org/10.1021/ja0675447.

[101] L. Ma, J.M. Falkowski, C. Abney, W. Lin, A series of isoreticular chiral metal–organic frameworks as a tunable platform for asymmetric catalysis, Nat. Chem. 2 (2010) 838–846. https://doi.org/10.1038/nchem.738.

[102] A.K. Rappe, C.J. Casewit, K.S. Colwell, W.A. Goddard, W.M. Skiff, UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations, J. Am. Chem. Soc. 114 (1992) 10024–10035. https://doi.org/10.1021/ja00051a040.

[103] S.L. Mayo, B.D. Olafson, W.A. Goddard, DREIDING: a generic force field for molecular simulations, J. Phys. Chem. 94 (1990) 8897–8909. https://doi.org/10.1021/j100389a010.

[104] M.S. Shah, M. Tsapatsis, J.I. Siepmann, Development of the Transferable Potentials for Phase Equilibria Model for Hydrogen Sulfide, J. Phys. Chem. B. 119 (2015) 7041–7052. https://doi.org/10.1021/acs.jpcb.5b02536.

[105] S. Kullback, R.A. Leibler, On Information and Sufficiency, Ann. Math. Stat. 22 (1951) 79–86.

[106] J.M. Joyce, Kullback-Leibler Divergence, in: M. Lovric (Ed.), Int. Encycl. Stat. Sci., Springer, Berlin, Heidelberg, 2011: pp. 720–722. https://doi.org/10.1007/978-3-642-04898-2_327.

[107] H.K. Chae, D.Y. Siberio-Pérez, J. Kim, Y. Go, M. Eddaoudi, A.J. Matzger, M. O'Keeffe, O.M. Yaghi, A route to high surface area, porosity and inclusion of large molecules in crystals, Nature. 427 (2004) 523–527. https://doi.org/10.1038/nature02311.

[108] O.K. Farha, A. Özgür Yazaydın, I. Eryazici, C.D. Malliakas, B.G. Hauser, M.G. Kanatzidis, S.T. Nguyen, R.Q. Snurr, J.T. Hupp, De novo synthesis of a metal–organic framework material featuring ultrahigh surface area and gas storage capacities, Nat. Chem. 2 (2010) 944–948. https://doi.org/10.1038/nchem.834.

[109] J.H. Cavka, S. Jakobsen, U. Olsbye, N. Guillou, C. Lamberti, S. Bordiga, K.P. Lillerud, A New Zirconium Inorganic Building Brick Forming Metal Organic Frameworks with Exceptional Stability, J. Am. Chem. Soc. 130 (2008) 13850–13851. https://doi.org/10.1021/ja8057953.

[110] H. Li, M. Eddaoudi, M. O'Keeffe, O.M. Yaghi, Design and synthesis of an exceptionally stable and highly porous metal-organic framework, Nature. 402 (1999) 276–279. https://doi.org/10.1038/46248.

[111] B. Eckl, J. Vrabec, H. Hasse, An optimised molecular model for ammonia, Mol. Phys. 106 (2008) 1039–1046. https://doi.org/10.1080/00268970802112137.

[112] D.W. Green, R.H. Perry, Perry's Chemical Engineers' Handbook, Eighth Edition, McGraw Hill Professional, 2007.

[113] J. Burkardt, The Truncated Normal Distribution, (n.d.) 35.