ACS SENSORS

pubs.acs.org/acssensors

Article

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

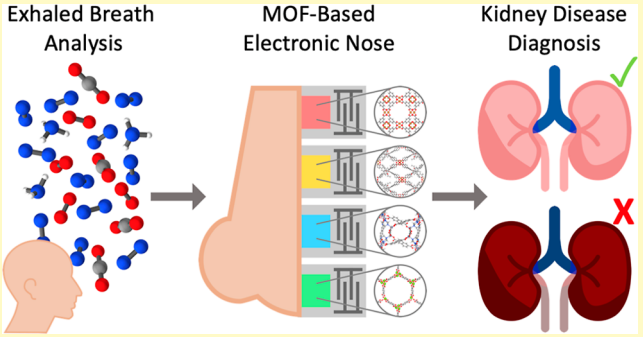
Brian A. Day and Christopher E. Wilmer*

Cite This: *ACS Sens.* 2021, 6, 4425–4434

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The diverse chemical composition of exhaled human breath contains a vast amount of information about the health of the body, and yet this is seldom taken advantage of for diagnostic purposes due to the lack of appropriate gas-sensing technologies. In this work, we apply computational methods to design mass-based gas sensor arrays, often called electronic noses, that are optimized for detecting kidney disease from breath, for which ammonia is a known biomarker. We define combined linear adsorption coefficients (CLACs), which are closely related to Henry's law coefficients, for calculating gas adsorption in metal–organic frameworks (MOFs) of gases commonly found in breath (i.e., carbon dioxide, argon, and ammonia). These CLACs were determined computationally using classical atomistic molecular simulation techniques and subsequently used to design and evaluate gas sensor arrays. We also describe a novel numerical algorithm for determining the composition of a breath sample given a set of sensor outputs and a library of CLACs. After identifying an optimal array of five MOFs, we screened a set of 100 simplified computer-generated, water-free breath samples for kidney disease and were able to successfully quantify the amount of ammonia in all samples within the tolerances needed to classify them as either healthy or diseased, demonstrating the promise of such devices for disease detection applications.



KEYWORDS: *metal–organic frameworks, MOFs, breath, volatile organic compounds, VOCs, diagnostics, kidney disease, computational*

INTRODUCTION

Despite the advances in gas-sensing technologies, clinical use of breath samples for disease detection and monitoring is still very much in its infancy.[1−3] This is due in part to the cost, time, and expertise required of the existing gas-sensing technologies such as gas chromatography and mass spectroscopy and is further complicated by the challenges associated with breath collection and analysis.[1,4−6] The result is that very few breath-based tests are routinely found in practice, with the present examples being for relatively simple yet important applications such as monitoring airway inflammation (asthma) and indirect measurement of blood alcohol content.[2,3] If one could design a device which offered portable, accurate, and real-time sensing for a wide variety of the compounds found in breath, it would allow for a significant leap in noninvasive disease detection.[7,8]

However, disease detection via breath presents many challenges. Although many of the compounds found in breath have been successfully identified, only a subset of these compounds are known biomarkers tied to health status, and of these, only a smaller subset have established concentration ranges associated within healthy and diseased individuals.[9,10] Moreover, the concentration ranges of these compounds can vary due to the factors independent of the health of the patient, such as ambient air composition, activity of the patient before sampling (e.g., diet and exercise), the history of the patient (e.g., smoking habits and pre-existing conditions), the portion of breath sampled (e.g., tidal, alveolar, or full breath), and the specific sampling method, all resulting in seemingly moving targets for normal concentration ranges.[11−16] For example, increased levels of ethane and pentane in breath are believed to be indicative of oxidative stress, but specific concentration ranges have not been established, and some studies even suggest that they are not biomarkers at all but rather reflect the concentration levels in air.[17−19]

Despite these challenges, disease detection and monitoring by breath remains a promising area of research owing to its unique and largely untapped clinical potential. Breath samples are noninvasive and continuously available, allowing for both quick and easy sampling, as well as real-time monitoring.[4] Furthermore, breath tests would be ideal for screening large

Received: August 23, 2021
Accepted: November 15, 2021
Published: December 2, 2021

populations. Herein, we aim to explore the use of gas-sensing arrays, often called electronic noses, for applications of disease detection via breath.

In this work, we chose to target chronic kidney disease (CKD), for which ammonia is a well-established biomarker with known healthy and diseased concentration ranges.[20,21] According to the World Health Organization, in the year 2012, 1.5% of deaths globally was attributed to CKD, and this percentage is expected to rise.[22] Moreover, CKD is caused by diabetes and hypertension, which can also lead to anemia, bone disease, heart disease, and cancer, among numerous other health complications, meaning that adequate detection of CKD could also assist with the treatment of these other diseases.[22]

In our previous works, we used computational methods to rapidly design gas-sensing arrays for the detection of methane and carbon dioxide, using metal−organic frameworks (MOFs) as the sensing elements.[23−26] MOFs are high-surface-area porous crystals that have been widely explored for gas storage and separation[27−30] and more recently have gained attention for gas-sensing applications.[31−35] MOFs offer advantages for gas-sensing that stem from their structural and chemical diversity, tunability, and high internal surface areas, resulting in impressive gas adsorption with varied sensitivity. Furthermore, as crystalline materials, their gas adsorption properties can be predicted using well-established computational techniques.[27,36]

Our previous works, however, were limited in the complexity of the gas mixtures which could be considered due to computational expense limitations, 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, this prior method becomes combinatorically prohibitive. As 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.

Fortunately, for disease detection by breath, the biomarker gases of interest are typically present only in trace quantities, and so their adsorption in MOFs obeys Henry's law (adsorbed concentration is a linear function of the concentration in the bulk mixture outside of the MOF).[37] In this case, it is sufficient to determine only the corresponding Henry's law coefficients for each gas/MOF pair to be able to predict the amount of each trace gas absorbed by the MOFs, thus drastically reducing computational demand. In this work, we evaluated a modified form of Henry's coefficients, which we call combined linear adsorption coefficients (CLACs), which quantify not only the amount of trace gas species adsorbed but also the amount of air displaced by the trace gas, as a function of the trace gas concentration. CLACs were evaluated for three trace gases in 50 MOFs and then were used to design a gas-sensing array which was capable of classifying a set of simplified water-free breath samples as either healthy or diseased for CKD via a newly developed algorithm. These 50 MOFs were originally chosen from the CoRE MOF database to have a diverse set of properties and have been used in several papers by our group for the sake of consistency.[25,26,38]

A similar strategy was first developed by Sturluson et al.[35] 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 nontrace gas species (i.e., gases for which Henry's law would not apply). In our case, the nontrace gases are nitrogen and oxygen (i.e., air), and hence where Sturluson et al. used traditional Henry's coefficients, we use our so-called CLACs. Consequently, the resulting arrays are geared toward different applications. Nevertheless, their work has been influential on ours, and we have adapted their methods liberally, as will be discussed later in the Methods section.

METHODS

The methods employed in this work can be broken into four distinct parts: (1) CLAC evaluation, (2) array design, (3) breath sample generation, and (4) breath sample analysis.

CLAC Evaluation. CLACs are best described as a modification of traditional Henry's coefficients. Consider exposing a MOF to a gas mixture which contains trace quantities of CO_2 dispersed in nontrace quantities of N_2 and O_2. Henry's law states 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 Henry's coefficient does not quantify is how the adsorption of CO_2 impacts the adsorption of N_2 and O_2. Our CLACs address this by quantifying both how much of the trace gas species (i.e., CO_2) is adsorbed and how much of the non-trace gas species (i.e., N_2 and O_2) is displaced, all as a function of the concentration of the trace gas species. Thus, a CLAC is simply the sum of a traditional Henry's coefficient for the trace gas with a displacement correction for the nontrace gases and can be written as follows

$$K_i^* = \frac{\Delta m_{\text{trace gas}}}{\Delta y_{\text{trace gas}}} + \frac{\Delta m_{\text{non-trace gases}}}{\Delta y_{\text{trace gas}}} \qquad \text{for } y_{\text{trace gas}} \leq y_{\text{trace gas,max}} \tag{1}$$

where Δm is the change in adsorbed mass, Δy is the change in concentration, and y_{max} is the maximum concentration for which the trace gas species can be considered dilute. Using CLACs to calculate the total adsorbed mass for a MOF also requires introducing a constant, which is the total adsorbed mass of the nontrace gases, in the absence of any trace gases, at the total pressure and temperature of interest. The resulting model is as follows

$$m_{\text{total}} = m_{\text{non-trace gases}} + \sum_{i=1}^{N} K_i^{**} y_i \tag{2}$$

where m_{total} is the total adsorbed mass, $m_{\text{non-trace gases}}$ 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 Henry's coefficient, which can never have a value less than 0, a CLAC can in principle take on any real value as it is possible that the amount of the nontrace gases displaced in the MOF is greater than the amount of trace gas species adsorbed by the MOF, resulting in a net decrease in mass. For further discussion, please refer to the Supporting Information (Section 2.1).

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 separate CLACs for all distinct majority gas species' concentrations. 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 to 0.05, and the remaining gases were N_2 and O_2 in 3:1, 4:1, and 5:1 ratios.[39] A pressure of 1 bar and a temperature at 298 K were used to replicate ambient conditions. The TraPPE force field was used for all gases, and a combination of the DREIDING and Universal force fields was used for the MOFs.[40−43] Although each force field is well studied, they are quite generic and unlikely to be highly accurate, especially for binding with open metal sites commonly found in MOFs. Nevertheless, the primary focus of this work is the development of a method for computationally designing a MOF-based electronic nose. In future works, we can revisit the force field terms to improve the accuracy of

https://doi.org/10.1021/acssensors.1c01808
ACS Sens. 2021, 6, 4425−4434

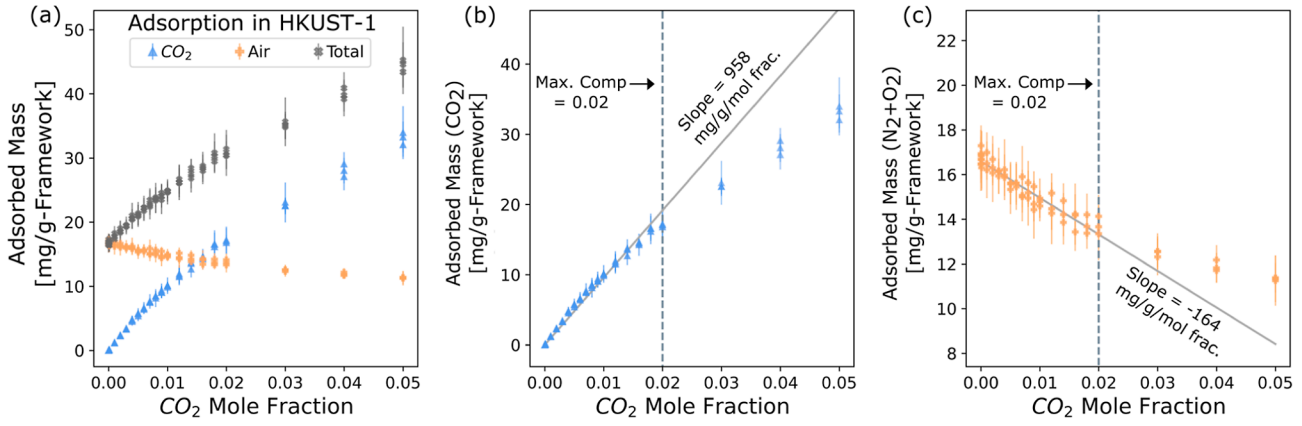


Figure 1. Extraction of CLAC coefficients from the adsorption data (a) in HKUST-1 with a background gas of 3:1, 4:1, and 5:1 N_2:O_2 (i.e., three points for each mole fraction). (b) Linear fit for CO_2 adsorption. (c) Linear fit for air displacement. All fits have an R^2 threshold of 0.95. The resulting CLAC for HKUST-1 is obtained from the sum of the slopes shown in (b,c), that is, 958−164 = 794 mg/g/mole fraction.

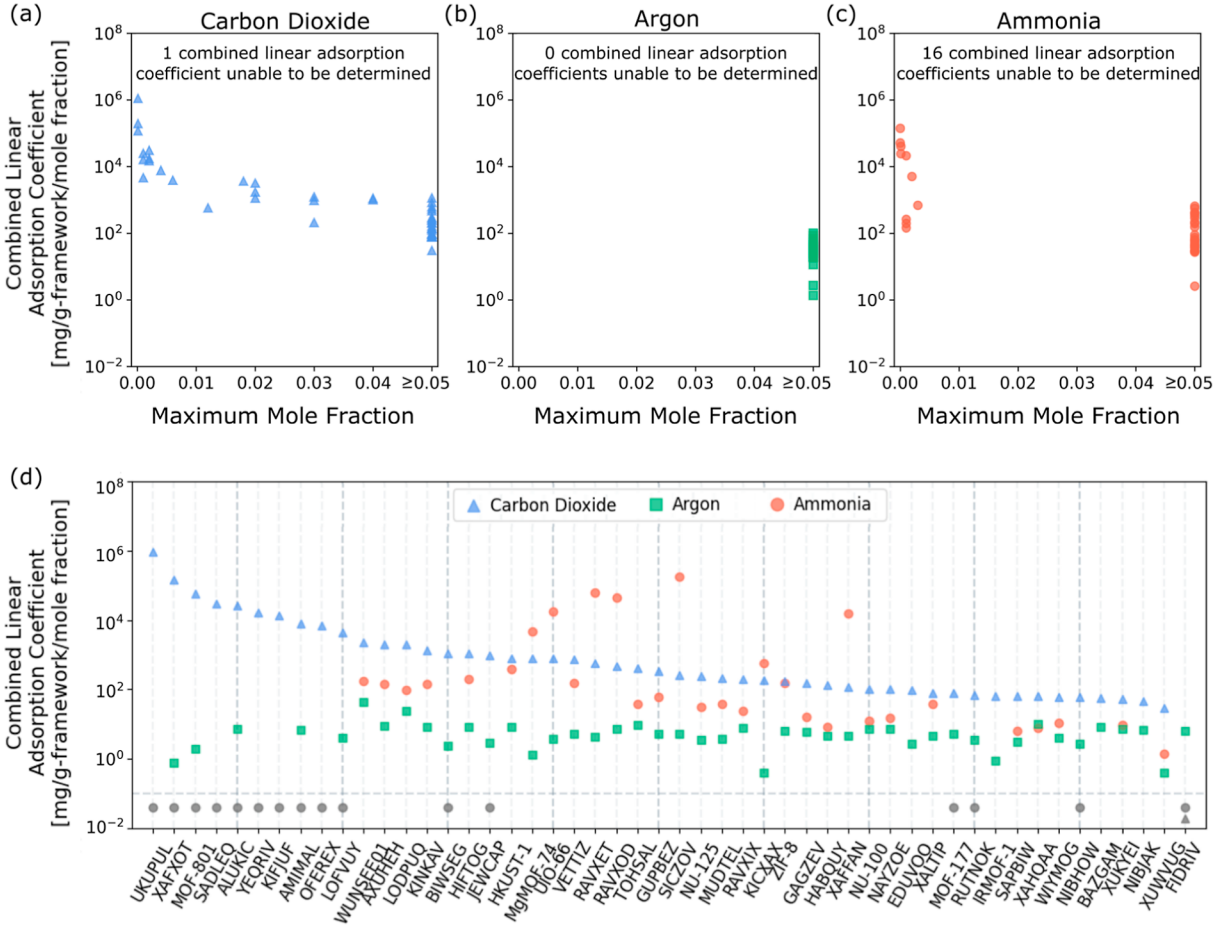


Figure 2. (a−c) CLAC versus maximum concentration for which adsorption of the trace gas species is linear (i.e., end of Henry's regime) for (a) carbon dioxide, (b) argon, and (c) ammonia. A fit was not obtained for all gases/MOFs, either because adsorption was highly nonlinear or because their uncertainty was too high in the adsorbed masses from simulations which prevented fitting with the desired R^2 cutoff. The number of nonfit 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 gray symbol of the same shape is plotted below the horizontal line.

our simulations either via ab initio calculations or experimentally determined CLACs. For further details, please refer to the Supporting Information (Section S2.1).

Conveniently, for almost all gases and MOFs, the adsorbed mass of the trace gas species was independent of the composition of the background gases, as evidenced by the high R^2 values of the resulting fits (see Figure 1). Thus, the evaluated CLACs could 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. This greatly simplified the problem, as these were the only two gases which were present in greater-than-trace amounts, and thus we only needed to determine the constant, $m_{\text{non-trace gases}}$, once for each MOF.

https://doi.org/10.1021/acssensors.1c01808
ACS Sens. 2021, 6, 4425−4434

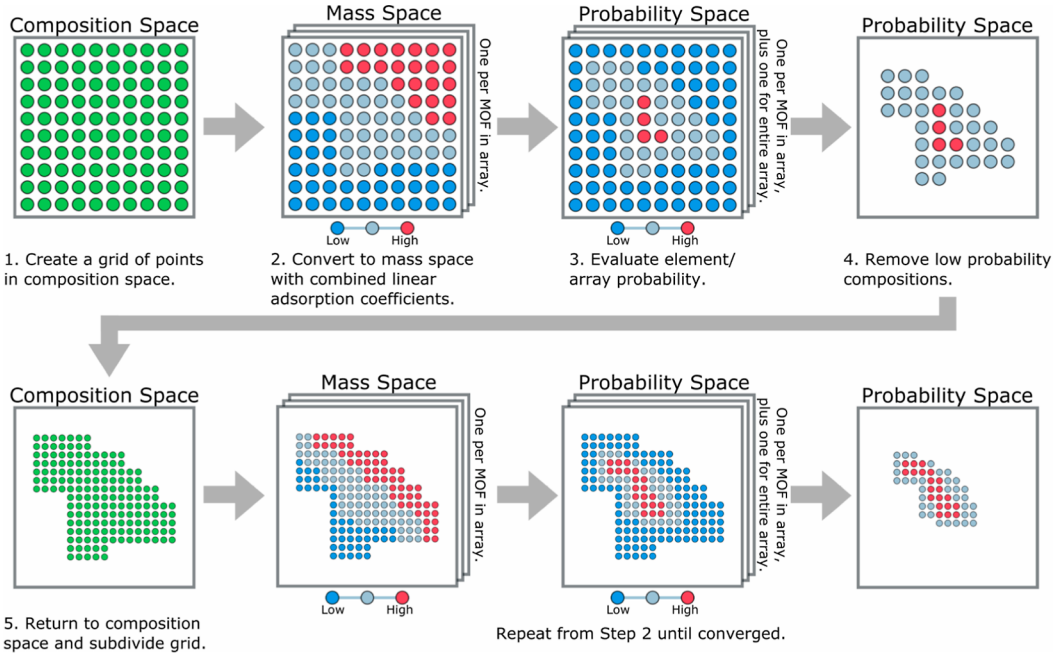


Figure 3. Simplified overview of the algorithm used to determine compositions from a set of sensor data and CLACs. Further details are given in the Supporting Information (Section S5).

To determine the CLACs, we fit a line to the adsorbed masses of the trace gas species for all ratios of air with the intercept forced to 0 (see Figure 1). 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 as 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, with each slope being one of the two terms in eq 1. A detailed description of this method is given in the Supporting Information (Section S2.2). An example of the resulting fits is shown in Figure 1, along with the raw adsorption data. Figure 2 shows the spread of the resulting CLACs for all gas and MOF combinations as a function of the width of their Henry's regime. It should be noted that as the maximum trace gas concentration tested was 0.05 mol fraction, any gas with Henry's regime ending at 0.05 mol fraction may actually exhibit linear adsorption beyond that point, but additional simulations would need to be done to determine this.

Array Design. 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, as quantified by y_{max} in eq 1, such that all trace gas species could be considered dilute and assumed not to interact with each other. Here, that corresponds to a minimum mole fraction of 0.05 for all gases (i.e., if adsorption exhibited a nonlinear behavior below a mole fraction of 0.05, Henry's regime was considered to be too narrow). Although a restrictive cutoff, this was necessary as we treat carbon dioxide as a trace gas despite it being present in mole fractions of upward of 0.05. Fortunately, of the 50 MOFs screened, 23 MOFs met this requirement and could be used to design arrays.

Our approach for array design 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 minimum singular value.[35] 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, as a MOF could in theory exhibit large Henry's coefficients for certain gases but have very low corresponding CLACs if the nontrace gases are displaced in similar amounts to the adsorbed trace gas. 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 toward 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 the accuracy of the composition determined by the algorithm changed with array size and, similarly, how the best and worst arrays of a given size compare. A brief overview of the array design method is given in the Supporting Information (Section S3); however, for a detailed description, we refer the reader to the original paper by Sturluson et al.[35]

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 found in breath, some of which are the byproducts of metabolic processes, while others are simply found in the air we breathe in.[1,10,12,44−46] 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.[47−53] Fortunately, several breath collection methods involve some form of a dehumidification step.[44,54,55]

We identified relevant concentration ranges of ammonia for CKD based on the report of Bevc et al., who showed that concentrations of 0.49 ± 0.08 and 3.32 ± 2.19 ppm corresponded to healthy and diseased individuals, respectively.[20] We chose the other gas components in the simplified breath samples to be as follows: carbon dioxide mole fraction between 2 and 5% (uniform distribution), argon mole fraction between 0.6 and 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.[20] Note that when referring to air in the context of the methods/results, we mean only the nitrogen/oxygen mixture and not any of the other compounds. The exact compositions of all the healthy and diseased breath samples are given in the Supporting Information in Tables S8 and S9, respectively.

https://doi.org/10.1021/acssensors.1c01808
ACS Sens. 2021, 6, 4425−4434

Next, we created a set of corresponding sensor outputs for each breath sample, here envisioning the sensor array to be composed of surface acoustic wave (SAW) devices, each using a different MOF as the sensing material, such that the corresponding sensor output is a measured change in mass due to gas adsorption. A SAW device is a microelectromechanical system that can measure very small changes in mass on its surface, where, for example, a MOF thin film may have been deposited. The "detected" changes in mass for each MOF for each breath sample were calculated using CLACs, as given by eq 2. With no instrument error yet introduced, the mass "detected" by the sensor is exactly what would be calculated by the algorithm for that composition. However, note that this does not necessarily result in a perfect determination of that composition but rather guarantees 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 determine a composition close to the maximumly probable composition, it is poorly behaved.

Breath Sample Analysis. The final step is to determine 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 on the detected mass. Under certain assumptions (see Supporting Information, Section S2, for additional details), the composition can be determined analytically, as outlined in the paper by Sturluson et al.[35] However, we decided to develop and employ a more general numerical algorithm that could be reused in future works where sensing nonlinear adsorption is particularly important, with the only requirement for the algorithm being a way of mapping compositions to adsorbed masses. An outline of this algorithm is depicted schematically in Figure 3.

The first step is to generate the initial set of compositions; there are three aspects to this step. The first aspect is choosing which gas species should be present in the composition space. Real breath samples can contain thousands of different compounds, so knowing which compounds must be included, which compounds can be safely excluded, and which compounds can be grouped together is nontrivial. Fortunately, as we consider simplified breath samples, this is trivial as we know exactly which gases are present. The second aspect is deciding what the minimum and maximum concentration values should be for each of the gas species. For our chosen application of CKD detection, reasonable concentration ranges are already known. Finally, the third aspect is to determine the spacing between points in the initial composition space. This choice can impact the speed of the algorithm and accuracy of the determined composition. Generally, the more finely grained the initial set of compositions is, the better the accuracy of the determined composition will be. This is discussed in more detail in the Supporting Information (Section S6), 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 1 in the Results and Discussion section.

The next step in the 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.[23−26] As a result, considering a finely spaced multicomponent gas mixture for several MOFs would have required significant computing time. Now, masses are determined from eq 2.

Calculating masses in this way enables one to consider all gases individually when using GCMC simulations, dramatically reducing the computational time. Consequently, not only can one evaluate more gases but one can also 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 Henry's regime (i.e., no competitive adsorption).

To illustrate the computational time saved by this change, consider our previous study; 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 to 30% and the concentration of nitrogen ranging from 40 to 100%, all in 1% increments.[26] The result was 48,050 distinct combinations of MOFs and compositions, each requiring a distinct simulation. In contrast, for this study, by considering the trace gas species separately, we evaluated the CLACs for three gases and 50 different MOFs using only 9450 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 all of the thousands of gas species in breath, while still enormously difficult, becomes a more plausible future goal.

Table 1. Compositions and Parameters Used for Evaluating Breath Samples via the Algorithm

(a) Initial Compositions

	initial range (ppm)	initial spacing (ppm)	convergence limits (ppm)
carbon dioxide	20,000−50,000	12.5	1000
argon	0−12,000	2000	1000
ammonia	0−10	0.25	0.1

(b) Algorithm Parameters

Parameter	value
maximum number of iterations	20 cycles
fraction to keep	0.04
standard deviation	0.10 mg/g framework

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 = 10 mg/g framework). Array probabilities are calculated by multiplying all of the element probabilities for each composition which are then normalized.

Finally, the last step is to filter the composition space down to the points which have the highest array probability and check for convergence. Assuming that the algorithm has not converged, we take the remaining highest probability compositions and subdivide the grid in the 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 of that which was used in the previous iteration. The net result is an increasingly narrow and fine-grained set of compositions.

As for the convergence criteria, after each cycle we check whether the range of concentrations for each gas is within its individually set tolerance. For example, in this work, we considered that the determined concentration 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 concentration ranges for the other gases may be far narrower than their specified convergence criteria. Additionally, we set a limit on the maximum number of iterations so that the algorithm will stop if struggling to converge. The resulting parameter set is given in Table 1. Figure 4 shows the evolution of the determined concentration range for a typical breath sample in this work.

https://doi.org/10.1021/acssensors.1c01808
ACS Sens. 2021, 6, 4425−4434

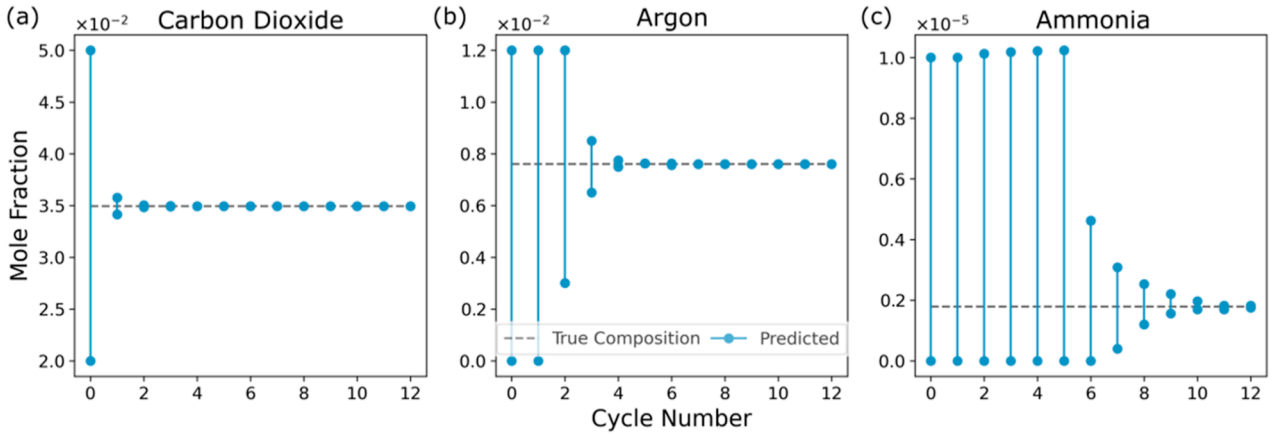


Figure 4. Concentration range for a single breath sample as a function of cycle number for (a) carbon dioxide, (b) argon, and (c) ammonia. The dashed line represents the true concentration of the gas in the breath sample.

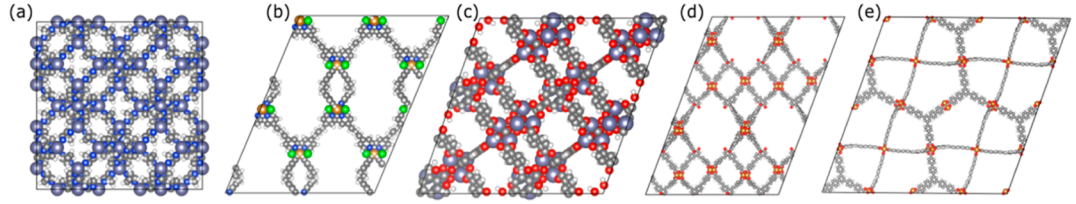


Figure 5. All MOFs which make up the best five-element sensing array, each displayed as a 2 × 2 × 2 unit cell down the crystallographic *a*-axis. The MOFs are: (a) ZIF-8,[56] (b) XUKYEI, (c) twofold interpenetrated MOF-5 (HIFTOG),[57] (d) CMOF-4b (XAHQAA),[58] and (e) MOF-399 (BAZGAM).[58] Note that the common names and first report of MOF are given when known, followed by the CoRe MOF reference code in parenthesis.

RESULTS AND DISCUSSION

Using the array design method outlined in the Methods section, we determined the best five-element array, shown in Figure 5, which includes the following MOFs: ZIF-8,[56] XUKYEI, twofold interpenetrated MOF-5 (HIFTOG),[57] CMOF-4b (XAHQAA),[58] and MOF-399 (BAZGAM).[58]

The results of the breath sample analysis for this array are given in Figure 6, with samples numbered from the lowest to greatest ammonia concentration. Our method could reliably determine the concentrations of each gas, including ammonia, which is the important biomarker for CKD.

We also examined how both array size and array quality impacted the algorithm's performance. To this end, we determined the best and worst one-, two-, three-, four-, five-, and ten-element arrays, as well as the only 23-element array, and again analyzed all 100 samples using the same algorithm parameters. These results are shown in Figure 7.

Unsurprisingly, one- and two-element arrays struggle to reliably determine the concentration of ammonia, with all the one-element arrays stopping due to reaching the maximum number of cycles and all two-element arrays converging with poor accuracy with respect to the true composition. The best three-element array substantially improves the overall accuracy, but there would be several false-positives/negatives (e.g., healthy samples, 6, and diseased samples, 9). The four- and five-element arrays are sufficiently accurate and would not lead to any false-positive/negative cases for any of the 100 breath samples, and the 10- and 23-element arrays lead to very few (e.g., 10-element array healthy samples, 47, and 23-element array diseased sample, 1).

The 10- and 23-element arrays offer particularly interesting results, as the accuracy of the determined compositions is noticeably poorer (i.e., the relative error between the determined and true concentration of ammonia increases) than the best five-element array, which is a subset of both the larger arrays. As this analysis uses computer-generated sensor outputs for the breath samples, and as we have established that there is a composition within the bounds of the initial composition space which uniquely has the highest probability, the algorithm should be able to determine this composition with sufficient parameterization. An additional artifact of using computer-generated sensor outputs is that none of the sensors offer contradictory information. Consequently, the reduction in the accuracy of the compositions determined by the 10- and 23-element arrays must be caused by the algorithm.

Specifically, this behavior is the result of each of the 23 MOFs used in the various arrays having a larger CLAC for CO_2 than for ammonia, such that a small change in the concentration of CO_2 has a stronger impact on the predicted masses than an equally small change in the concentration of ammonia. In some instances, depending on the spacing of the composition grid relative to the true composition, the algorithm accounts for a slight over- or under-prediction of CO_2 with a complementary under- or overprediction of ammonia. Although doing so pushes the algorithm further from the true composition, from a mass perspective, it gets closer to the correct answer. The likelihood of this happening actually increases with additional MOFs, hence the increase in error exhibited by large arrays. That said, the algorithm must be close to the correct set of masses for this behavior to happen, explaining why poor arrays still benefit from additional MOFs; they were never close enough to the correct set of masses to exhibit this behavior in the first place. This problem can be alleviated either by starting with a much finer grid

https://doi.org/10.1021/acssensors.1c01808
ACS Sens. 2021, 6, 4425−4434

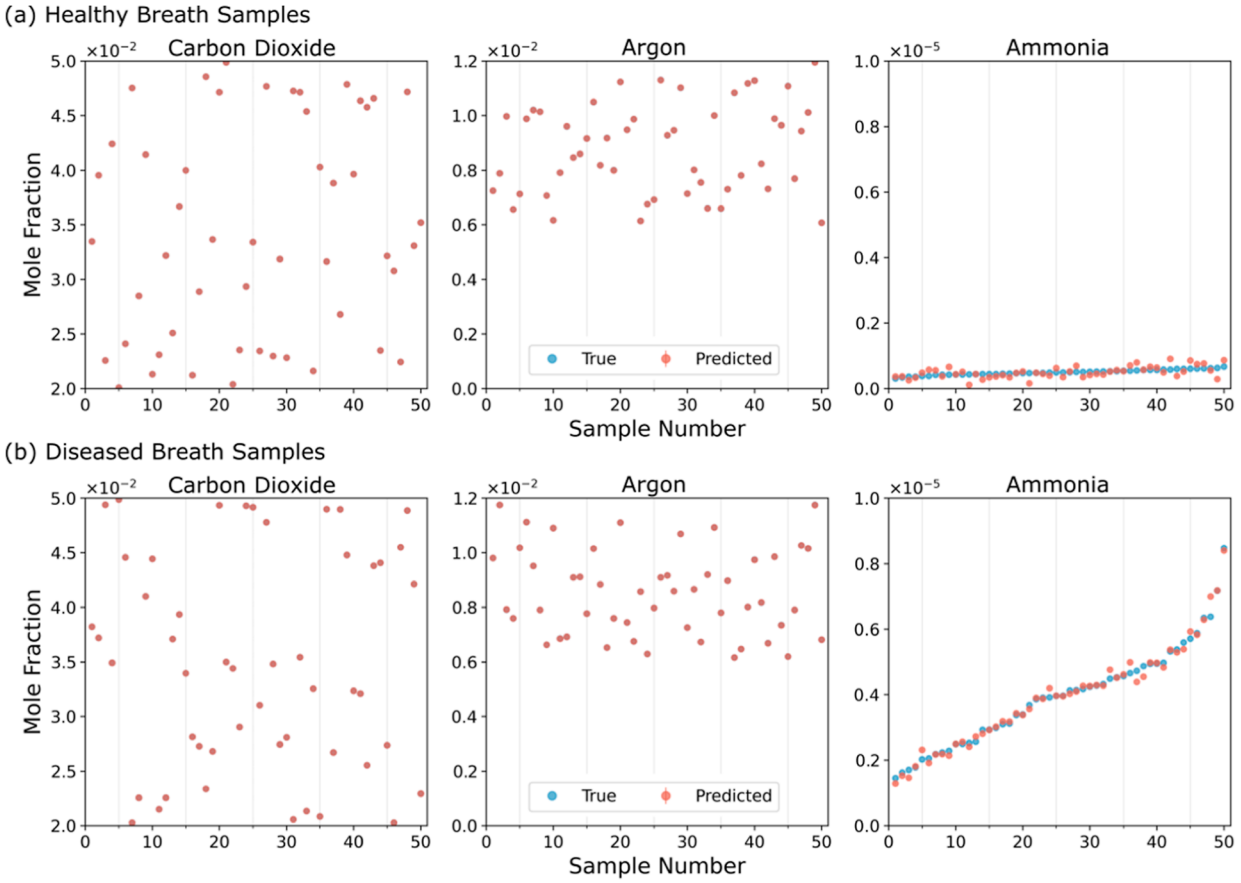


Figure 6. Final determined concentration ranges for the best five-element array for (a) 50 healthy and (b) 50 diseased samples, ordered by increasing ammonia concentration. The upper and lower bounds on each plot correspond to the initial concentration range used in the algorithm.

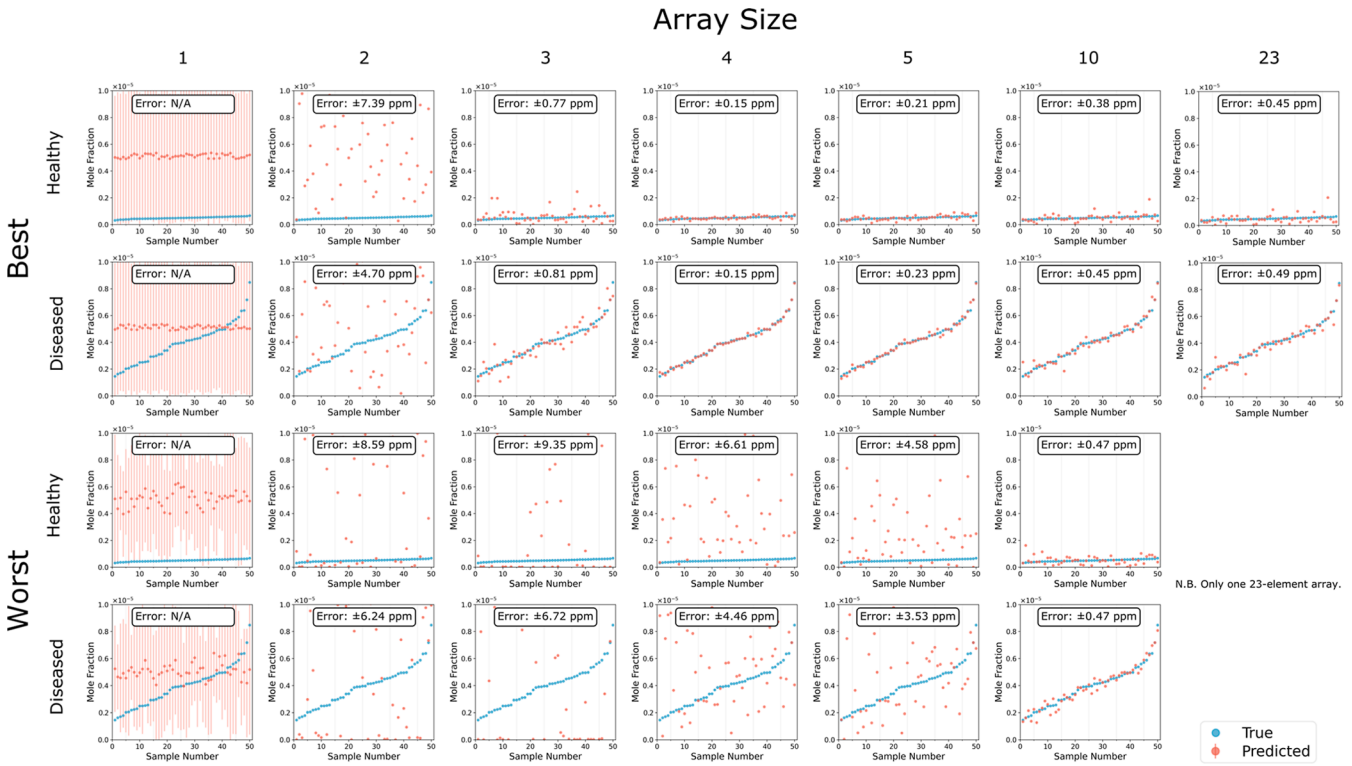


Figure 7. Comparison of the best and worst arrays of various sizes. Note that only the determined concentration of ammonia is shown, as all the arrays reliably determine the concentration of carbon dioxide and argon. Also note that there is only one 23-element array.

https://doi.org/10.1021/acssensors.1c01808
ACS Sens. 2021, 6, 4425−4434

spacing (which can lead to long analysis times) or by adding a sensing element which exhibits more sensitivity to ammonia over CO_2. The former solution motivates the determination of a unique set of algorithm parameters for each array/application, and the latter underscores the importance of both large-scale screening and intelligent selection of materials. For further discussion of this behavior, refer to the Supporting Information (Section S6.1).

As for a comparison between the best and worst arrays, the difference in the accuracy of the determined compositions also highlights that intelligent selection of sensing elements for arrays is still a critical aspect of this electronic nose work, especially considering that almost all real arrays will be underdetermined to some degree given that breath can contain thousands of different components.

Finally, as evidenced by the performance of the 10- and 23-element arrays, the algorithm parameters have a non-negligible impact on the accuracy of the determined compositions. Even if an array which is capable of making accurate determinations is used, in the absence of a sufficient parameter set for the algorithm, the determined compositions will be unreliable. The effects of specific parameters (i.e., initial composition set and fraction to keep) are examined and discussed in Section S6 (Figures S5 and S6) of the Supporting Information. The parameters used within this work were determined in a guess-and-check fashion, but a more systematic way of determining algorithm parameters may be desirable for future work.

CONCLUSIONS

In this study, we developed an updated methodology for designing MOF-based sensor arrays for the detection of dilute gas species in complex gas mixtures based on a modified form of Henry's coefficient which we call a CLAC. We determined the CLACs for a set of gases commonly found in breath for 50 MOFs to design a sensing array for detecting kidney disease, for which ammonia is a well-established biomarker. Then, using the method presented by Sturluson et al., we screened and ranked all arrays of various sizes and selected the best five-element array for testing.[35] Using our newly designed numerical 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 to be done for building a practical MOF-based device for CKD detection, specifically in increasing the number of gases and MOFs considered, accounting for the effects of humidity, accounting for the measurement error, and improving the accuracy of force field parameters, the various methods presented here demonstrate marked efficiency improvements over prior work.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acssensors.1c01808.

Molecular simulation parameters, MOF properties, additional explanation of all methodologies including the CLAC fitting procedure, array design, numerical algorithm, and additional results examining algorithm parameters (PDF)

AUTHOR INFORMATION

Corresponding Author

Christopher E. Wilmer − *Department of Chemical & Petroleum Engineering, Department of Electrical & Computer Engineering, and Clinical and Translational Science Institute, University of Pittsburgh, Pittsburgh, Pennsylvania 15261, United States;* orcid.org/0000-0002-7440-5727; Email: wilmer@pitt.edu

Author

Brian A. Day − *Department of Chemical & Petroleum Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania 15261, United States;* orcid.org/0000-0003-3769-6462

Complete contact information is available at: https://pubs.acs.org/10.1021/acssensors.1c01808

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We are grateful for the support from the National Science Foundation for funding via their EAGER program: CBET-1937179. We are additionally grateful to the Clinical and Translational Science Institute at the University of Pittsburgh for their support via the PInCh award and to the University of Pittsburgh Center for Medical Innovation for providing an early-stage seed grant. We also acknowledge support from the University of Pittsburgh's Swanson School of Engineering and the Center for Research Computing for computational resources.

REFERENCES

(1) Risby, T. H.; Solga, S. F. Current Status of Clinical Breath Analysis. *Appl. Phys. B* **2006**, *85*, 421−426.

(2) Paschke, K. M.; Mashir, A.; Dweik, R. A. Clinical Applications of Breath Testing. *F1000 Med. Rep.* **2010**, *2*. https://doi.org/DOI: 10.3410/M2-56.

(3) Amann, A.; Miekisch, W.; Schubert, J.; Buszewski, B.; Ligor, T.; Jezierski, T.; Pleil, J.; Risby, T. Analysis of Exhaled Breath for Disease Detection. *Annu. Rev. Anal. Chem.* **2014**, *7*, 455−482.

(4) Shirasu, M.; Touhara, K. The Scent of Disease: Volatile Organic Compounds of the Human Body Related to Disease and Disorder. *J. Biochem.* **2011**, *150*, 257−266.

(5) Behera, B.; Joshi, R.; Vishnu, G. K. A.; Bhalerao, S.; Pandya, H. J. Electronic Nose: A Non-Invasive Technology for Breath Analysis of Diabetes and Lung Cancer Patients. *J. Breath Res.* **2019**, *13*, 024001.

(6) Yoon, J.-W.; Lee, J.-H. Toward Breath Analysis on a Chip for Disease Diagnosis Using Semiconductor-Based Chemiresistors: Recent Progress and Future Perspectives. *Lab Chip* **2017**, *17*, 3537−3557.

(7) Nakhleh, M. K.; Amal, H.; Jeries, R.; Broza, Y. Y.; Aboud, M.; Gharra, A.; Ivgi, H.; Khatib, S.; Badarneh, S.; Har-Shai, L.; Glass-Marmor, L.; Lejbkowicz, I.; Miller, A.; Badarny, S.; Winer, R.; Finberg, J.; Cohen-Kaminsky, S.; Perros, F.; Montani, D.; Girerd, B.; Garcia, G.; Simonneau, G.; Nakhoul, F.; Baram, S.; Salim, R.; Hakim, M.; Gruber, M.; Ronen, O.; Marshak, T.; Doweck, I.; Nativ, O.; Bahouth, Z.; Shi, D.-y.; Zhang, W.; Hua, Q.-l.; Pan, Y.-y.; Tao, L.; Liu, H.; Karban, A.; Koifman, E.; Rainis, T.; Skapars, R.; Sivins, A.; Ancans, G.; Liepniece-Karele, I.; Kikuste, I.; Lasina, I.; Tolmanis, I.; Johnson, D.; Millstone, S. Z.; Fulton, J.; Wells, J. W.; Wilf, L. H.; Humbert, M.; Leja, M.; Peled, N.; Haick, H. Diagnosis and Classification of 17 Diseases from 1404 Subjects via Pattern Analysis of Exhaled Molecules. *ACS Nano* **2017**, *11*, 112−125.

(8) Mashir, A.; Dweik, R. A. Exhaled Breath Analysis: The New Interface between Medicine and Engineering. *Adv. Powder Technol.* **2009**, *20*, 420−425.
(9) Corradi, M.; Mutti, A. Exhaled Breath Analysis: From Occupational to Respiratory Medicine. *Acta Bio Med. Atenei Parmensis* **2005**, *76*, 20−29.
(10) Pauling, L.; Robinson, A. B.; Teranishi, R.; Cary, P. Quantitative Analysis of Urine Vapor and Breath by Gas-Liquid Partition Chromatography. *Proc. Natl. Acad. Sci. U.S.A.* **1971**, *68*, 2374−2376.
(11) Lawal, O.; Ahmed, W. M.; Nijsen, T. M. E.; Goodacre, R.; Fowler, S. J. Exhaled Breath Analysis: A Review of 'Breath-Taking' Methods for off-Line Analysis. *Metabolomics* **2017**, *13*, 1.
(12) Filipiak, W.; Ruzsanyi, V.; Mochalski, P.; Filipiak, A.; Bajtarevic, A.; Ager, C.; Denz, H.; Hilbe, W.; Jamnig, H.; Hackl, M.; Dzien, A.; Amann, A. Dependence of Exhaled Breath Composition on Exogenous Factors, Smoking Habits and Exposure to Air Pollutants. *J. Breath Res.* **2012**, *6*, 036008.
(13) Sukul, P.; Trefz, P.; Schubert, J. K.; Miekisch, W. Immediate Effects of Breath Holding Maneuvers onto Composition of Exhaled Breath. *J. Breath Res.* **2014**, *8*, 037102.
(14) Sukul, P.; Trefz, P.; Kamysek, S.; Schubert, J. K.; Miekisch, W. Instant Effects of Changing Body Positions on Compositions of Exhaled Breath. *J. Breath Res.* **2015**, *9*, 047105.
(15) Sukul, P.; Oertel, P.; Kamysek, S.; Trefz, P. Oral or Nasal Breathing? Real-Time Effects of Switching Sampling Route onto Exhaled VOC Concentrations. *J. Breath Res.* **2017**, *11*, 027101.
(16) Sukul, P.; Schubert, J. K.; Kamysek, S.; Trefz, P.; Miekisch, W. Applied Upper-Airway Resistance Instantly Affects Breath Components: A Unique Insight into Pulmonary Medicine. *J. Breath Res.* **2017**, *11*, 047108.
(17) Gorham, K. A.; Andersen, M. P. S.; Meinardi, S.; Delfino, R. J.; Staimer, N.; Tjoa, T.; Rowland, F. S.; Blake, D. R. Ethane and N-Pentane in Exhaled Breath Are Biomarkers of Exposure Not Effect. *Biomarkers* **2009**, *14*, 17−25.
(18) Paredi, P.; Kharitonov, S. A.; Barnes, P. J. Elevation of Exhaled Ethane Concentration in Asthma. *Am. J. Respir. Crit. Care Med.* **2000**, *162*, 1450−1454.
(19) Aghdassi, E.; Allard, J. P. Breath Alkanes as a Marker of Oxidative Stress in Different Clinical Conditions. *Free Radic. Biol. Med.* **2000**, *28*, 880−886.
(20) Bevc, S.; Mohorko, E.; Kolar, M.; Brglez, P.; Holobar, A.; Kniepeiss, D.; Podbregar, M.; Piko, N.; Hojs, N.; Knehtl, M.; Ekart, R.; Hojs, R. Measurement of Breath Ammonia for Detection of Patients with Chronic Kidney Disease. *Clin. Nephrol.* **2017**, *88*, 14−17.
(21) Gafare, M.; Dennis, J. O.; Md Khir, M. H. Detection of Ammonia in Exhaled Breath for Clinical Diagnosis- A Review. *AIP Conf. Proc.* **2014**, *1621*, 303−309.
(22) Webster, A. C.; Nagler, E. V.; Morton, R. L.; Masson, P. Chronic Kidney Disease. *Lancet* **2017**, *389*, 1238−1252.
(23) Gustafson, J. A.; Wilmer, C. E. Computational Design of Metal−Organic Framework Arrays for Gas Sensing: Influence of Array Size and Composition on Sensor Performance. *J. Phys. Chem. C* **2017**, *121*, 6033−6038.
(24) Gustafson, J. A.; Wilmer, C. E. Optimizing Information Content in MOF Sensor Arrays for Analyzing Methane-Air Mixtures. *Sens. Actuators, B* **2018**, *267*, 483−493.
(25) Gustafson, J. A.; Wilmer, C. E. Intelligent Selection of Metal−Organic Framework Arrays for Methane Sensing via Genetic Algorithms. *ACS Sens.* **2019**, *4*, 1586−1593.
(26) Day, B. A.; Wilmer, C. E. Genetic Algorithm Design of MOF-Based Gas Sensor Arrays for CO2-in-Air Sensing. *Sensors* **2020**, *20*, 924.
(27) Getman, R. B.; Bae, Y.-S.; Wilmer, C. E.; Snurr, R. Q. Review and Analysis of Molecular Simulations of Methane, Hydrogen, and Acetylene Storage in Metal−Organic Frameworks. *Chem. Rev.* **2012**, *112*, 703−723.
(28) Düren, T.; Bae, Y.-S.; Snurr, R. Using Molecular Simulation to Characterise Metal−Organic Frameworks for Adsorption Applications. *Chem. Soc. Rev.* **2009**, *38*, 1237−1247.
(29) Li, J.-R.; Ma, Y.; McCarthy, M. C.; Sculley, J.; Yu, J.; Jeong, H.-K.; Balbuena, P. B.; Zhou, H.-C. Carbon Dioxide Capture-Related Gas Adsorption and Separation in Metal-Organic Frameworks. *Coord. Chem. Rev.* **2011**, *255*, 1791−1823.
(30) Zhou, H.-C.; Long, J. R.; Yaghi, O. M. Introduction to Metal−Organic Frameworks. *Chem. Rev.* **2012**, *112*, 673−674.
(31) Kreno, L. E.; Leong, K.; Farha, O. K.; Allendorf, M.; Van Duyne, R. P.; Hupp, J. T. Metal−Organic Framework Materials as Chemical Sensors. *Chem. Rev.* **2012**, *112*, 1105−1125.
(32) Stassen, I.; Burtch, N.; Talin, A.; Falcaro, P.; Allendorf, M.; Ameloot, R. An Updated Roadmap for the Integration of Metal−Organic Frameworks with Electronic Devices and Chemical Sensors. *Chem. Soc. Rev.* **2017**, *46*, 3185−3241.
(33) Campbell, M.; Dincă, M. Metal−Organic Frameworks as Active Materials in Electronic Sensor Devices. *Sensors* **2017**, *17*, 1108.
(34) Lu, G.; Hupp, J. T. Metal−Organic Frameworks as Sensors: A ZIF-8 Based Fabry−Pérot Device as a Selective Sensor for Chemical Vapors and Gases. *J. Am. Chem. Soc.* **2010**, *132*, 7832−7833.
(35) Sturluson, A.; Sousa, R.; Zhang, Y.; Huynh, M. T.; Laird, C.; York, A. H. P.; Silsby, C.; Chang, C.-H.; Simon, C. Curating Metal-Organic Frameworks to Compose Robust Gas Sensor Arrays in Dilute Conditions. *ACS Appl. Mater. Interfaces* **2020**, *12*, 6546.
(36) Sturluson, A.; Huynh, M. T.; Kaija, A. R.; Laird, C.; Yoon, S.; Hou, F.; Feng, Z.; Wilmer, C. E.; Colón, Y. J.; Chung, Y. G.; Siderius, D. W.; Simon, C. M. The Role of Molecular Modelling and Simulation in the Discovery and Deployment of Metal-Organic Frameworks for Gas Storage and Separation. *Mol. Simul.* **2019**, *45*, 1082−1121.
(37) Yu, X.; Choi, S.; Tang, D.; Medford, A. J.; Sholl, D. S. Efficient Models for Predicting Temperature-Dependent Henry's Constants and Adsorption Selectivities for Diverse Collections of Molecules in Metal−Organic Frameworks. *J. Phys. Chem. C* **2021**, *125*, 18046−18057.
(38) Chung, Y. G.; Camp, J.; Haranczyk, M.; Sikora, B. J.; Bury, W.; Krungleviciute, V.; Yildirim, T.; Farha, O. K.; Sholl, D. S.; Snurr, R. Q. Computation-Ready, Experimental Metal−Organic Frameworks: A Tool To Enable High-Throughput Screening of Nanoporous Crystals. *Chem. Mater.* **2014**, *26*, 6185−6192.
(39) Dubbeldam, D.; Calero, S.; Ellis, D. E.; Snurr, R. Q. RASPA: Molecular Simulation Software for Adsorption and Diffusion in Flexible Nanoporous Materials. *Mol. Simul.* **2016**, *42*, 81−101.
(40) Zhang, L.; Siepmann, J. I. Development of the Trappe Force Field for Ammonia. *Collect. Czech. Chem. Commun.* **2010**, *75*, 577−591.
(41) Martin, M. G.; Siepmann, J. I. Transferable Potentials for Phase Equilibria. 1. United-Atom Description of n-Alkanes. *J. Phys. Chem. B* **1998**, *102*, 2569−2577.
(42) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **1992**, *114*, 10024−10035.
(43) Mayo, S. L.; Olafson, B. D.; Goddard, W. A. DREIDING: A Generic Force Field for Molecular Simulations. *J. Phys. Chem.* **1990**, *94*, 8897−8909.
(44) Guo, D. D.; Zhang, D.; Li, N.; Zhang, L.; Yang, J. A Novel Breath Analysis System Based on Electronic Olfaction. *IEEE Trans. Biomed. Eng.* **2010**, *57*, 2753−2763.
(45) Egnér, H.; Eriksson, E. Current Data on the Chemical Composition of Air and Precipitation. *Tellus* **1955**, *7*, 134−139.
(46) Glueckauf, E. The Composition of Atmospheric Air. *Compend. Meteorol.* **1951**, 3.
(47) Canivet, J.; Fateeva, A.; Guo, Y.; Coasne, B.; Farrusseng, D. Water Adsorption in MOFs: Fundamentals and Applications. *Chem. Soc. Rev.* **2014**, *43*, 5594−5617.
(48) Ghosh, P.; Kim, K. C.; Snurr, R. Q. Modeling Water and Ammonia Adsorption in Hydrophobic Metal−Organic Frameworks:

https://doi.org/10.1021/acssensors.1c01808
ACS Sens. 2021, 6, 4425−4434

Single Components and Mixtures. *J. Phys. Chem. C* **2014**, *118*, 1102−1110.
(49) Ramachandran, C. E.; Chempath, S.; Broadbelt, L. J.; Snurr, R. Q. Water Adsorption in Hydrophobic Nanopores: Monte Carlo Simulations of Water in Silicalite. *Microporous Mesoporous Mater.* **2006**, *90*, 293−298.
(50) Nalaparaju, A.; Zhao, X. S.; Jiang, J. W. Molecular Understanding for the Adsorption of Water and Alcohols in Hydrophilic and Hydrophobic Zeolitic Metal−Organic Frameworks. *J. Phys. Chem. C* **2010**, *114*, 11542−11550.
(51) Wang, C.; Liu, X.; Keser Demir, N.; Chen, J. P.; Li, K. Applications of Water Stable Metal−Organic Frameworks. *Chem. Soc. Rev.* **2016**, *45*, 5107−5134.
(52) Desbiens, N.; Boutin, A.; Demachy, I. Water Condensation in Hydrophobic Silicalite-1 Zeolite:A Molecular Simulation Study. *J. Phys. Chem. B* **2005**, *109*, 24071−24076.
(53) Peng, G.; Trock, E.; Haick, H. Detecting Simulated Patterns of Lung Cancer Biomarkers by Random Network of Single-Walled Carbon Nanotubes Coated with Nonpolymeric Organic Materials. *Nano Lett.* **2008**, *8*, 3631−3635.
(54) Nishibori, M.; Shin, W.; Izu, N.; Itoh, T.; Matsubara, I. Sensing Performance of Thermoelectric Hydrogen Sensor for Breath Hydrogen Analysis. *Sens. Actuators, B* **2009**, *137*, 524−528.
(55) Grote, C.; Pawliszyn, J. Solid-Phase Microextraction for the Analysis of Human Breath. *Anal. Chem.* **1997**, *69*, 587−596.
(56) Park, K. S.; Ni, Z.; Cote, A. P.; Choi, J. Y.; Huang, R.; Uribe-Romo, F. J.; Chae, H. K.; O'Keeffe, M.; Yaghi, O. M. Exceptional Chemical and Thermal Stability of Zeolitic Imidazolate Frameworks. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 10186−10191.
(57) Hafizovic, J.; Bjørgen, M.; Olsbye, U.; Dietzel, P. D. C.; Bordiga, S.; Prestipino, C.; Lamberti, C.; Lillerud, K. P. 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.* **2007**, *129*, 3612−3620.
(58) Ma, L.; Falkowski, J. M.; Abney, C.; Lin, W. A Series of Isoreticular Chiral Metal−Organic Frameworks as a Tunable Platform for Asymmetric Catalysis. *Nat. Chem.* **2010**, *2*, 838−846.

https://doi.org/10.1021/acssensors.1c01808
ACS Sens. 2021, 6, 4425−4434