**Naval Research Laboratory**

**Washington, DC 20375-5320**

**NRL/MR/6180--20-10,151**

# Design Theory for Chemical Detection

Kevin J. Johnson

Adam C. Knapp

*Naval Technology Center for Safety & Survivability Branch*
*Chemistry Division*

Month 00, 2020

# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 0704-0188*

| | | |
|---|---|---|
| **1. REPORT DATE** *(DD-MM-YYYY)*<br>00-00-2020 | **2. REPORT TYPE**<br>NRL Memorandum Report | **3. DATES COVERED** *(From - To)*<br>October 2016 – September 2020 |

**4. TITLE AND SUBTITLE**

Design Theory for Chemical Detection

**5a. CONTRACT NUMBER**

**5b. GRANT NUMBER**

**5c. PROGRAM ELEMENT NUMBER**

**6. AUTHOR(S)**

Kevin J. Johnson and Adam C. Knapp

**5d. PROJECT NUMBER**

**5e. TASK NUMBER**

**5f. WORK UNIT NUMBER**
1E39

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**

Naval Research Laboratory
4555 Overlook Avenue, SW
Washington, DC 20375-5320

**8. PERFORMING ORGANIZATION REPORT NUMBER**

NRL/MR/6180--20-10,151

**9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)**

Naval Research Laboratory
4555 Overlook Avenue, SW
Washington, DC 20375-5320

**10. SPONSOR / MONITOR'S ACRONYM(S)**

NRL

**11. SPONSOR / MONITOR'S REPORT NUMBER(S)**

**12. DISTRIBUTION / AVAILABILITY STATEMENT**

**DISTRIBUTION STATEMENT A:** Approved for public release; distribution is unlimited.

**13. SUPPLEMENTARY NOTES**

**14. ABSTRACT**

This report details a rational, quantitative approach for designing and optimizing sensor arrays for chemical detection. The work demonstrates that the degree to which sensor array capabilities and limitations can be understood is unavoidably linked to the degree to which the underlying response mechanism has been characterized, as well as inextricably linked to specific formulations of sensing task. A mathematical framework for describing chemical sensing tasks is presented and device-agnostic quality metrics based on information theoretic measures are derived and interpreted within the context of this framework. An algorithmic approach for sensor array configuration is presented, and it is demonstrated that the derived quality metrics are amenable to convex optimization. Information-theoretic approaches for quantitatively understanding the impact of uncertain chemical backgrounds on chemical detection systems, characterizing the quality of chemical simulants under realistic conditions, and understanding the underlying chemical expressiveness of a given sensor technology are derived from this framework and discussed.

**15. SUBJECT TERMS**

Sensor arrays Design theory Information theory
Chemical detection Machine olfaction

| | | | | | |
|---|---|---|---|---|---|
| **16. SECURITY CLASSIFICATION OF:** | | | **17. LIMITATION OF ABSTRACT** | **18. NUMBER OF PAGES** | **19a. NAME OF RESPONSIBLE PERSON**<br>Kevin J. Johnson |
| **a. REPORT**<br>Unclassified<br>Unlimited | **b. ABSTRACT**<br>Unclassified<br>Unlimited | **c. THIS PAGE**<br>Unclassified<br>Unlimited | Unclassified<br>Unlimited | 85 | **19b. TELEPHONE NUMBER** *(include area code)*<br>(202) 404-5407 |

**Standard Form 298 (Rev. 8-98)**
**Prescribed by ANSI Std. Z39.18**

This page
intentionally
left blank

## CONTENTS

# FIGURES

# TABLES

# EXECUTIVE SUMMARY

This basic research program has sought to develop the mathematical tools and language to refocus sensor array and machine olfaction design away from the predominant ad hoc approach to one based on rigorous statistical estimation. This program has demonstrated that the physics of the underlying sensor mechanisms in concert with statistical models of their attendant noise can rigorously define and assess the sensing capabilities of sensor arrays. These assessments must be made within the context of a systematic description of sensing tasks, which was shown to correspond to specific vectors in a geometric space spanning all potential mixture combinations. The resolving power oriented in the direction of a particular sensing task is governed by the measurement error imparted by the sensors and the projection into chemical stimulus space indicated by the sensor response functions. Figures of merit for multivariate sensor arrays based on general statistical notions and agnostic to sensor-specific quantities were developed using information-theoretic measures, including the Fisher information and Kullback-Leibler information divergences. General measures of array sensitivity, selectivity, and resolving power were derived in this work. Important accompanying investigations regarding the chemical environment in which these sensor arrays might operate were also performed, demonstrating approaches to characterizing the impact of chemical background on sensor performance. Models for realistic chemical background mixtures were developed as well as an information-theoretic approach for assessing the efficacy of chemical simulants under different backgrounds. Finally, this framework was used to explain how the fundamental transduction mechanism associated with a particular type of sensor limits the chemical informing power achievable by arrays of those type of sensors, as well as how the informing power can be expressed and compared between different sensor types. In total, this work indicates that a reliable understanding of sensor array performance is significantly predicated on the extent to which one understands how the sensors function on a molecular basis as well as the extent to which one understands the nature of the sensing tasks that are to be accomplished.

This work has demonstrated a rational, quantitative approach for designing and optimizing sensor arrays for chemical detection. The degree to which sensor array capabilities and limitations can be properly understood is unavoidably linked to the degree to which the underlying response mechanism has been characterized, including a characterization of sensor noise. Further, expressions of sensor array capabilities are inextricably linked to specific formulations of sensing task. In scenarios where detailed knowledge of sensing task is available (i.e. specification desired discernible chemical stimuli and characterization of chemical background) efficient sensor arrays can be designed at the expense of robustness. When the sensing task is poorly understood, additional complexity can be engineered into the detection system to compensate for this uncertainty. This suggests that investigation of the molecular basis of sensor response mechanism and practical investigations into the nature of the sensing tasks that one wishes to address are equally as important as development of the sensor technology itself.

This page
intentionally
left blank

# DESIGN THEORY FOR CHEMICAL SENSING

## 1. INTRODUCTION

Chemical detection and analysis is a critical component of many DoD and government activities and takes place in a wide variety of environments, under a wide range of needs and contexts. For example, both submarine crew and naval aviators need ongoing situational awareness of breathing atmosphere quality, including the ability to detect and identify trace contaminants at particular actionable levels, but on significantly different platform. Warfighters in forward operations may require specific capabilities to detect trace explosives or chemical agents in harsh, low-resource environments, while security agencies may wish to detect the same compounds at domestic checkpoints, but with drastically different available resources, throughput requirements and tolerance for false positives. Trace chemical analysis is needed to validate pristine clean rooms as well as to remediate contaminated sites. Others may require chemical detection systems to ensure critical, chemically complex materials, such as aviation fuel, conform to product specifications prior to use. In some cases, it is critical to quantitatively assess the amount of a particular chemical present, while in others it is sufficient to provide a qualitative assessment of the presence of one of a particular class of compounds. As one might expect, correctly identifying the proper chemical detection strategy is thus essential for successfully support the needs of a given sensing application and context.

In general, chemical measurement encompasses a wide range of qualitatively distinct types of processes that can be loosely indexed by some notion of complexity. At one extreme are relatively simple operations, like target analyte detection in the absence of a chemical background. At the other extreme are more complex problems, such as analytical determination of the content of unknown chemical mixtures. Most chemical sensing applications occupy intermediate locations on this scale, representing varying degrees of complexity. Chemical sensor arrays represent a promising strategy for addressing a variety of resource-constrained, complex chemical sensing tasks [1]. Such devices have generated considerable interest over the past 40 years, both by virtue of their analogy to biological olfaction (in which an array of several hundred different olfactory receptor neurons support perception of thousands of distinct olfactory perceptions) as well as because they represent a promising intermediate between a need to design and field highly specific chemical sensors for each potential sensing task and the costs and logistical burdens associated with using analytical instruments.

Intuitively, and perhaps a bit simplistically, one might expect that a more complicated sensing tasks require a commensurate increase in sensing hardware complexity to be successfully addressed, and conversely, that simpler sensing tasks should require less complex technical solutions. While this intuition may be supported by practical experience in chemical measurement, there is no general principal governing the design of chemical measurement technology that says as much,and there is no general description of what exactly "complexity" means for the purposes of chemical detection system design. There is massive diversity in the scope of possible configurations of arrays of nonspecific chemical sensors, both an advantage and a drawback of such systems. The advantage is that, in theory, instrumental complexity can be tailored to an analytical task of arbitrary scope. However, designing a sensor array configuration that addresses a

---

Manuscript approved TBD.

complex sensing task to a desired performance level can be difficult and there is limited specific guidance and theory as to how such systems should be built. This has doubtless slowed their adoption and commercial prospects in the past few decades [2–6]. Nevertheless, the idea of sensor arrays as a potential alternative to more costly analytical instrumentation remains popular and continues to be an active area of research in the sensing community, where, owing to a persistent desire in most applications to utilize simpler, less costly hardware whenever possible, there remains significant motivation to develop and build sensor arrays for complex sensing tasks.

The underlying thesis of this work is that a sensing technology should be designed so that its information-generating capacity is appropriately matched to the complexity of the sensing task (or tasks) that it is required to accomplish, and thus that design should be informed by an understanding of how information flows through these systems as well as by an understanding of the parameters of the sensing tasks to be accomplished. Doing this in a systematic manner requires a quantitative understanding of how sensor response translates to chemical information for multisensory systems, how this relates to the informational requirements of chemical sensing tasks, and finally how aleatoric uncertainty in sensor response and epistemic uncertainty regarding the sensor task impact this process. We show how concepts and approaches from information theory, a branch of mathematics associated with describing and solving problems in communications, signal processing and cryptography is uniquely suited to address this challenge, enabling derivation of quantitative measures to predict the information-generating capacity of specific chemical measurement technologies under uncertainty.

## 1.1 Sensing Tasks

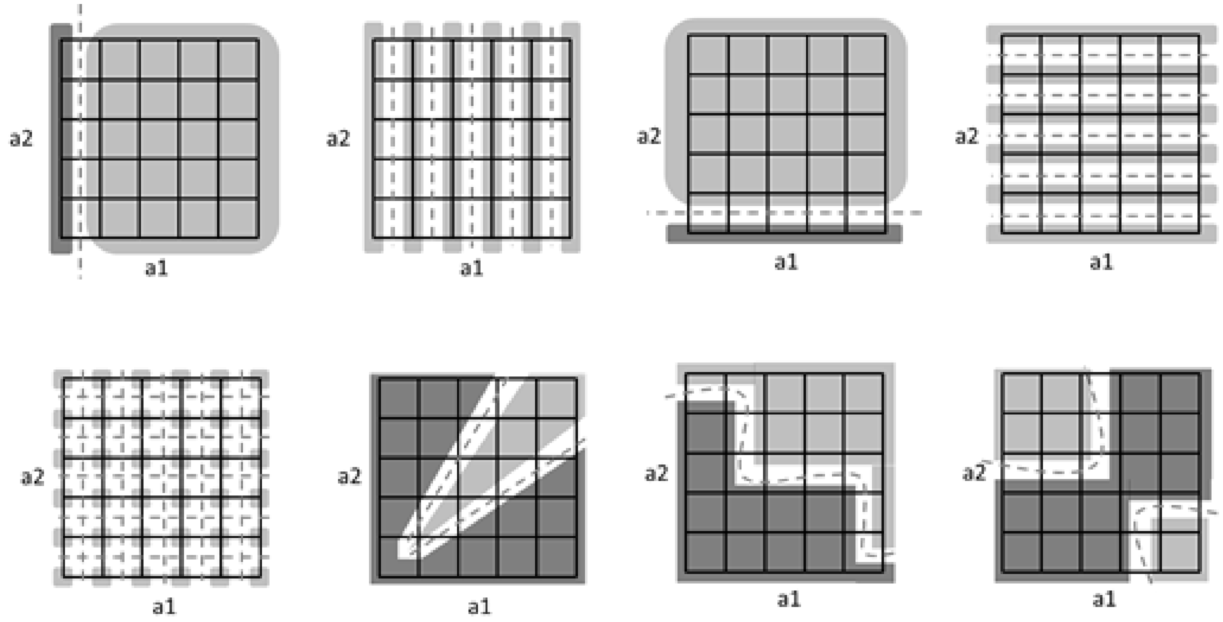

Fig. 1—Examples of analytical tasks involving a two-dimensional space describing all possible mixtures of two analytes, a1 and a2. Top row: detection of a1, quantitation of a1, detection of a2, quantitation of a2. Bottom row: quantification of both a1 and a2, identification of a particular mixture composition, detection of mixtures belonging to contiguous and non-contiguous arbitrary subregions of the space.

Regardless of context or complexity, the fundamental task in any chemical sensing application is to use a measured response from the sensing system to discern between two or more possible chemical stimuli.

The term “chemical stimulus” refers to a specific chemical environment impinging upon the active portion of a chemical sensing system. The term “sensing task” refers to a specification of two or more chemical stimuli that one wishes to discern. It is helpful to think of the range of possible chemical stimuli in a given scenario as occupying a K-dimensional Euclidean space with each axis corresponding to the concentration of one of K possible chemical compounds. Each unique mixture of chemical compounds is represented by a vector of K concentrations, providing Cartesian coordinates in the space. Movement along the direction described by this vector indicates change in the amount of the mixture present, while angular displacement from this vector indicates changes in mixture composition. Thus, the space contains all possible mixtures of the K different compounds in a continuous fashion, with specific mixtures having specific locations, and displacements in particular directions associated with specific changes in mixture composition. Any sensing task can thus be described as a requirement to be able to resolve particular regions of this space from each other, as illustrated in Fig. (1).

Qualitative tasks involve a requirement to discern stimuli from discrete regions of the space, while quantitative tasks involve a requirement to be able to discern incremental displacements in the space in particular directions. More generally, a given analytical task can be described mathematically at each point in the space as a requirement on resolving power associated with incremental displacement in any direction. This provides a means to begin understanding sensing task complexity in a more rigorous manner that is related to the amount of information about the chemical environment it represents. For example, complete quantitative analysis of mixture composition requires every incremental volume of the space to be mutually discernible to the same degree. Alternatively, qualitative detection of one target analyte requires only that points falling on one side of a hyperplane defined by the desired detection threshold be discernible from those falling on the other side. Likewise, identification of a particular chemical mixture from others requires that points that lie along the vector describing that mixture be discernible from those that do not. This geometric interpretation provides a useful framework for objectively assessing chemical sensor system capabilities for specific sensing tasks.

### 1.2 Sensors and Sensor Arrays

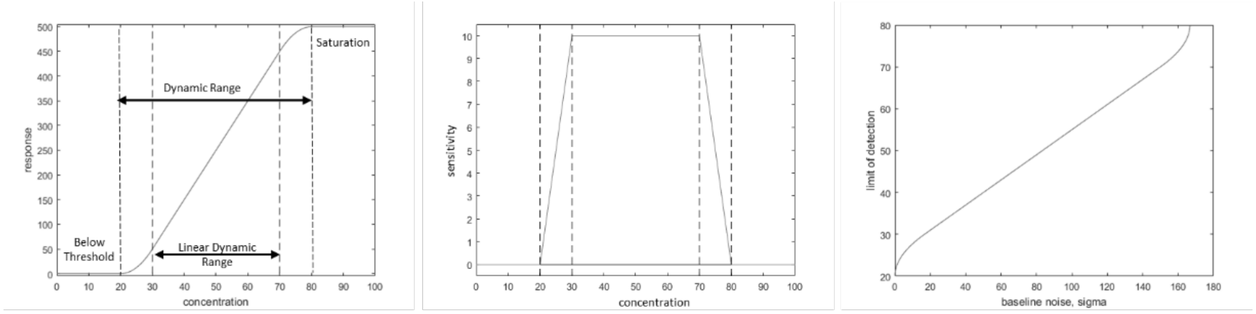

Fig. 2—Example sensor response function and associated metrics. Left: sensor response as a function of concentration, showing dynamic range, linear range, and signal roll-off due to concentrations leading to sensor saturation or below response threshold. Middle: sensitivity of nominal sensor response function at left. Right: limit of detection as a function of baseline noise sigma in sensor response.

Next, we consider complexity of chemical measurement devices. Devices employed for simple sensing tasks, such as detection of specific target analytes, are often categorized as chemical sensors or detectors while those employed to address the most complex sensing tasks are often categorized as instruments. The former tend to provide univariate responses corresponding to individual target analytes, while the latter often provide a multivariate response pattern that changes shape with different target analytes. Here, the term “chemical sensor” will be used as a general term to reference any device that is employed to interrogate a chemical mixture and is capable of transducing a measurable response from a molecular interaction with that mixture, with a characteristic uncertainty according to a particular measurement error distribution. Critical aspects of and individual sensor’s functional relationship are associated with standard figures of merit analytical chemistry, illustrated in Fig. (2).

The derivative of sensor response with respect to stimulus concentration corresponds to analytical sensitivity. The uncertainty associated with sensor measurement error, combined with sensitivity gives rise to characteristic concentrations at which a chemical stimulus can be reliably detected or quantified, as well as to signal-to-noise ratios associated with specific concentrations. Finally, notions of analytical selectivity rest on the degree of similarity between sensor response functions corresponding to different chemical stimuli.

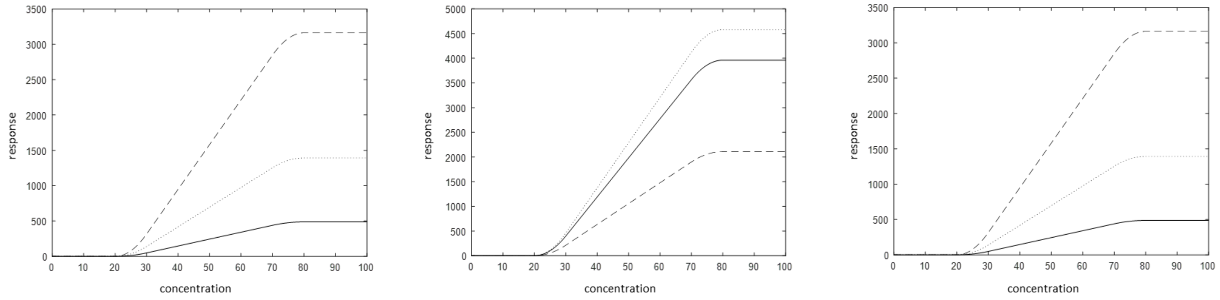

Fig. 3—. Example response functions of three different sensors (left , right and middle) exhibiting cross-sensitivity against three different analytes (solid, dot and dash.)

Sensor arrays consist of multiple different sensors, each interacting with the same chemical stimulus according to its own response function, thus generating a vector response from the array with each measurementas illustrated by the simple example in Fig. (3). Chemical sensors generally exhibit significant sensitivity over a range of chemical compounds (i.e., cross-sensitivity) due to a fundamental trade-off between selectivity and reversibility in sensor transduction mechanisms [7]. Although cross-sensitivity can be an undesirable feature in applications where individual sensors are used, it is an important feature in sensor arrays because it allows each sensor to simultaneously convey information about multiple analytes when the sensor responses are used in aggregate to provide an analytical result. Thus, the goal of sensor array design is to assemble sensors whose individual capabilities will complement each other in an analytically-relevant fashion. A central challenge in realizing the potential of this approach lies in the combinatorial nature of sensor array design, which means that even a modestly sized pool of candidate sensors will result in a substantial number of possible array configurations.

Sensor array complexity is governed by the number of component sensors within the array as well as the diversity in sensor response. Booksh and Kowalski presented a theory of analytical chemistry in which analytical instruments are categorized by the tensor order of the data they produce [8]. In this theory, individual chemical sensors that generate a scalar value per measurement are categorized as “zero order” instruments, while sensor arrays and instruments providing a spectral response are “first order” instruments. This distinction is important, as there is a quantum difference between the capabilities of instruments from each category. In particular, zero order instruments are fundamentally incapable of resolving multicomponent mixtures in which more than one concentration is unknown. In contrast, first order instruments do allow such resolution in principle, provided they exhibit sufficient complexity allow some marginal unique response to each mixture component. Importantly, this means that it is not necessary for a first order instrument to exhibit a one-to-one correspondence between individual sensors and mixture components, but it does indicate that the size of a sensor array and the diversity of sensor response functions are related directly to the scope of sensing tasks it is capable of addressing.

## 2. INFORMATION THEORY

The field of information theory began with the seminal work of Claude Shannon in which he laid out a generalized, probabilistic framework to describe the nature of information and how it can be encoded and transmitted [9]. Shannon's generalized framework considers the propagation of information through five basic components: an information source, a transmitter, a channel, a receiver, and a destination. The information source produces a "message" that is encoded by the transmitter. The channel is medium by which the encoded message is conveyed to the receiver, which decodes the message and delivers it to the destination, who is the recipient of the message. During transmission, the message is subject to corruption with noise. Messages are selected from a set of possible messages, and have meaning. That is, they correlate to different "physical or conceptual entities" and thus the ability to tell one message from another is important. Shannon's insight was that information sources can be characterized as stochastic processes that generate specific messages from the set of n possible discrete messages according to some probability, and that the information content of a source can be characterized by a measure of the form:

$$H = -\sum_{i=1}^{n} p_i \log_2(p_i) \tag{1}$$

This quantity is known as the Shannon entropy, or information entropy, and can intuitively thought of as a measure of the complexity of the messages generated from that source. Sources with higher entropy are associated with a greater span of messages than those with lower entropy. The quantity is expressed in units defined by the base of the logarithm, which provides the combinatorial context for interpreting the value. For example, using base two renders the entropy in "bits," with one bit reflecting the amount of information carried by a boolean variable with uniform probability of each state (i.e., an unbiased coin flip.) Thus, an expression of the information entropy of a source in bits can be interpreted as the number of boolean variables required to accurately represent it. For example, the English alphabet contains 26 letters and thus is theoretically capable of being expressed with 4.7 bits of information per letter, although in practice, computers represent characters with 8 bits.

Information entropy and subsequent measures derived from it for both discrete and continuous distributions provided a quantitative, theoretical understanding of the nature of information and had profound implications in the development of modern communication, computing, and data storage technologies. For instance, Shannon showed that the information entropy of a random variable represented the minimal amount of information required to fully describe it, and thus presented a theoretical limit on the degree to which data can be compressed in a lossless fashion. Continuing the example above, it has been experimentally estimated that the English language only actually conveys 1.3 bits of information per letter, due to the constraints imposed by phonology and grammar [10]. Thus, English text carries a significant degree of redundancy and can be highly compressed using a suitable encoding scheme. Shannon was also able to show that, for a given level of noise, a channel has a theoretical capacity (expressed in units of information per unit time, e.g. bits per second) that reflects the highest rate at which messages can be transmitted losslessly using error-correcting schemes. In other words, he was able to quantitatively express the minimal channel capacity required to noiseless transmit data from a particular information source, as well as the amount by which channel capacity must increase to minimally compensate for increases in channel noise. This can explain, for example, the practical function of the apparent significant redundancy in written English, as this redundancy, in principle, allows a significant measure of error-correcting capability for noisy communication channels. This is evident by the widely observed fact that readers can generally successfully parse English sentences even when the vowels are removed or the middle letters of each word are mixed up. Numerous other information theoretic measures have been derived and explored [11]. Of particular importance is the relative entropy, or Kullback-Liebler (KL) divergence:

$$D_{KL}(p||q) = \int dx p(x) \log \frac{p(x)}{q(x)} \tag{2}$$

which describes the difference in information conveyed by two distributions over random variable X. For example, in coding theory, it can be interpreted as describing the additional amount of information required to represent a message from an information source represented by p with a code optimized for the information source represented by q. KL divergence is greater than zero unless p and q are identical, in which case it is equal to zero. It is also not symmetric, i.e. the KL divergence of p from q is not equal to that of q from p. KL divergence is also closely related to information entropy. The information entropy associated with a distribution p is related to the divergence of p from a uniform distribution over the same support, and thus decreases as p becomes less uniform.

The mutual information between two random variables, X and Y, is defined as:

$$I(x,y) = \int\int dx dy p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \tag{3}$$

and is related to the KL divergence as:

$$I(x,y) = D_{KL}(p(x,y)||p(x)p(y)) \tag{4}$$

Mutual information is thus a measure of correlation between the two variables, in the sense that it reduces to zero if the joint distribution of X and Y is equal to the product of the two marginal distributions, showing the variables are independent and increases as the joint distribution deviates from the product of the marginal distributions. In other words, it provides a general expression of the amount of information one can gain about one variable through measurement of the other.

Information theory has also had significant impact beyond the area of telecommunications and data storage. Information theory has found direct application in cryptography[refs], signal and language processing [11], machine learning and pattern recognition [12], neurobiology [13], bioinformatics [14], and mathematical study of combinatorics and complexity [15]. The apparent analog between information entropy and thermodynamic entropy has provided important insight into thermal and quantum physics, including quantum computing.[refs] Accordingly, one might expect that information theory would also occupy a key role in analytical chemistry and chemical sensing. However, this has not been the case, with relatively little literature focused on this topic. Some 20 years after Shannon's initial publication, Kaiser published a notable exception in which he used concepts from information theory to describe concepts of "informing power of an analytical procedure" and "information required for an analytical problem" in the context of elemental analysis using spectrochemical instrumentation [16, 17]. Critically, Kaiser identified that each can be measured in informational units such as bits and noted that these bits represent a kind of currency that is "earned" by an analytical procedure and "spent" in the resolution of analytical problems. This allows, in principle, an objective assessment of how the two can be matched, although he states that, "there is no simple transformation from the space of problems to the space of procedures leading to the solution of a specific problem. The whole system is a very complicated network through which the analyst has to find his way." And expresses his hope that, "…this rather abstract picture of the two spaces may eventually lead to formal or empirical partial transitions from the space of problems into that region of the space of procedures where the solution of a problem may be found."

Kaiser emphasizes that this approach requires complete numerical characterization of not just the informing power of an instrument but of the information the analytical problem required, something he called "pre-information." Kaiser observes that rigorous descriptions of analytical problems are often neglected in practical situations, it being a "very common offense not to give all informations [sic] about the samples, their origin and history and not to say what knowledge is wanted and what the requirements are as to precision, time, et cetera," and he laments that "In general, it is not realized what an enormous amount of information may be hidden in an exclusion, in particular if this is made as a positive statement." This may offer some explanation for why an information theoretic approach has not gained greater traction in the field of analytical chemistry: computational complexity, and a lack of alignment with the mindset of the typical practitioners of analytical chemistry.

Nonetheless, information theoretics have made increasing appearance in chemical sensing and adjacent fields, particularly in descriptions of neural response to chemical stimuli.[refs] In particular, the utility of mutual information in rigorous assessment of sensor limit of detection has been described [18], as well as some of the implications of sensor array response diversity on information-generating capacity of sensor arrays [19, 20]. Finally, the work of Pierce, et al. [6, 21–28] formally connected earlier geometric notions of the chemical resolving power of sensor arrays to the Fisher information, a quantity closely related to the KL divergence.

## 3. COVARIANCE AND CHEMICAL MEASUREMENT

### 3.1 How Well Can You Know a Measurement? An Introduction to Practical Problems with Chemical Sensing

A persistent problem in chemical sensing lies in developing and implementing sensor systems that are capable of reliably addressing complex sensing tasks and environments. Such conditions are regularly encountered in a highly diverse range of practical endeavors, including environmental monitoring, industrial process control, toxic chemical and fire detection, flavor and fragrance assessment, and medical diagnostics, among other applications. Often, these tasks are centered on the detection of chemical signatures rather than individual chemical compounds. However, even seemingly-basic sensing tasks centered on detection of individual analytes are often complicated significantly by unavoidable environmental conditions and nuances that surround any realistic application, such as backgrounds with relatively dynamic physical conditions and multiple potentially interfering chemical species. This can lead to surprisingly poor performance in "real world" environments after excellent results have been demonstrated in the laboratory [2–4]. Thus, it is not only a matter of knowing and addressing the target analytes themselves, but also an understanding of the surrounding contextual details of a chemical sensing problem that is critical to finding a solution.

Further complicating matters, there is often a considerable degree of uncertainty regarding the underlying parameters of a sensing task. This includes uncertainty regarding which potential analytes and interfering species may or may not be present and what environmental conditions might be encountered, as well as more fundamental problems such as a lack of a priori knowledge as to the identity and number of potential target analytes and interferants. Perniciously, the existence of such uncertainties are often, themselves, poorly understood prior to implementation of a sensor system. Explicit recognition of this situation leads to the logic of an open-world assumption [29], implying that this state of affairs must place some limits on the extent of inference and deduction that may be derived from measured sensor responses.

Importantly, while development of individual sensors is generally driven by a desire to improve selectivity and sensitivity relative to current alternatives, complex environments make these metrics difficult to evaluate. Selectivity, which is defined by IUPAC in strictly qualitative terms, [30] only has meaning in the context of a specific background of potential interferants. In a practical sense, selectivity subsumes the problem of sensitivity, as it reflects the challenge of maximizing sensitivity to the chemical species of interest in the sensing task while minimizing sensitivity to others. Even in a best case scenario where an exhaustive catalog of the potential interferants and target analytes is known, design and implementation of a rigorous, exhaustive series of laboratory experiments for evaluating sensor performance becomes impractical unless there is a relatively small number of analytes and the field environment in which the sensor is to implemented is known to be fairly static. In more typical scenarios, such evaluation is essentially impossible due to the uncertainties described above. Thus, sensing in complex environments can be seen as an exercise in general-purpose chemical analysis of unknown samples rather than straightforward target analyte detection.

### 3.2 Sensor Response Models

#### *3.2.1 Sensor Response Types*

There is significant diversity in the modes of chemical sensor transduction and the corresponding responses that are observed, leading to a wide assortment of candidate sensor technologies that can be brought to bear on complex sensing tasks. Numerous articles and reviews have been published over the past few decades, covering various sensor technologies as well as chemical sensing technology in general [31, 32]. These techniques exist at varying levels of technological maturity, from commercially-available commodity devices to cutting-edge research prototypes and offer widely divergent capabilities, although no single technology presents a definitive solution to general-purpose chemical sensing. While an exhaustive review of all sensor technologies is not within the scope of this work, the sensor types that are most commonly reported in chemical sensing applications can be grouped as belonging to one of three major families according to transduction mechanism: electrochemical, mechanical, or optical.

Electrochemical sensors function by measuring a change induced in the electrical properties of the sensing material induced by analyte vapor. They include chemiresisters based on metal oxide semiconductors [33] conducting polymers [34] metal-organic frameworks [35], and nano-structured materials [36–38] as well as those based on other metal oxide semiconductor field effect transistors (MOSFET) and potentiometric devices [39, 40]. Mechanical sensors function by measuring change in the physical properties of the sensing material induced by analyte vapor [41]. They include devices based on quartz crystal microbalances (QCM) and surface acoustic wave sensors (SAW) [42, 43], and microelectromechanical systems (MEMS) [44]. In both QCM and SAW sensors, this is a change in resonant frequency of an oscillator induced by absorption of analyte vapor into a polymer coating. In MEMS sensors, this is generally either deflection or change in oscillating frequency of a polymer-coated microcantilever. Finally, Optical sensors include those based on materials in which analyte vapor induces a change in the way light is absorbed, emitted, or refracted by the sensing material [45–53]. These include sensors based on vapochromic dyes [47–50], fluorescence [51, 52], chemiluminescence [53, 54], and surface plasmon resonance [55, 56].

Although the specific data processing and feature extraction methods vary with each type of sensor, in each case the goal is to derive a response value that is related to analyte concentration according to some known function. These response curves are evaluated by traditional analytical figures of merit such as sensitivity, limit of detection/signal-to-noise ratio, dynamic range and selectivity [57]. The first three of these all describe sensor performance with respect to individual analytes while selectivity is related to sensor's ability to preferentially respond to one analyte in the presence of one or more others. Thus, any concept or calculation of selectivity is inherently bound to the parameters of a specific sensing task. Specificity is the hypothetical ultimate of selectivity: a "specific" sensor responds only to analytes of interest while a "non-specific" sensor exhibits varying responses to both target analytes and other compounds, i.e. interferants. Often, sensors are designed for particular target analytes for which sensitivity and selectivity are optimized relative to a particular sample matrix. In general, greater sensitivity places greater demands on selectivity while, conversely, greater selectivity places reduced demands on sensitivity. Thus, a significant portion of sensor research and development for complex sensing environments lies in engineering progressively greater selectivity into such devices.

Of course, the end goal of sensor optimization is producing an overall detection system which best provides the information one is seeking to measure. For instance, a device intended for toxic vapor detection might be evaluated in terms of a receiver operating characteristic (ROC) curve, which summarizes the trade-off between detection and false positive rates over a range of decision thresholds applied to the sensor data [58]. Alternatively, sensor data could be used to provide quantitative predictions of analyte vapor concentration, which would typically be evaluated in terms of calibration model fit and estimates of error of prediction [59, 60]. In either case, it is important to understand that the success of the overall detection system rests fundamentally on the capability of the sensor responses to support the informational output that is desired and that this is related to both the sensor response characteristics and the parameters of the analytical task.

Chemical sensors are usually not truly specific, especially for arbitrary, complex detection tasks with uncertain parameters. In the absence of guaranteed specificity, a significant limitation of sensors with univariate output (i.e. those that provide one response value per measurement) for complex sensing tasks, is that they are generally incapable of estimating analyte concentration in the presence of unknown amounts of other compounds and are thus not useful in analysis of mixtures or in the presence of unknown amounts of interfering species [8]. This limitation can be addressed by augmenting the sensor with one or more others to provide complementary information, forming an array of non-specific sensors. From a mathematical standpoint, the data from such an array is thus analogous to a low-resolution version of the multivariate data one might acquire from conventional spectroscopic or mass spectrometric laboratory instrumentation. Identification of pure compounds in these cases is often based on spectral patterns rather than on any one channel within the spectrum, enabling identification of more compounds than the number of channels in the spectrum. Additionally, multivariate regression enables calibration of individual mixture components in the presence of unknown amounts of others [61].

The vertebrate olfaction system is a common inspiration for using arrays of non-specific sensors to solve difficult chemical detection tasks. This system incorporates an array of roughly 100 million olfactory receptor neurons, each expressing one of about 1000 unique receptor types, all exhibiting varying non-specific interactions with potential odorant molecules. Patterns of neural activity within the peripheral olfaction system are then transmitted to the olfactory cortex of the brain where they are mapped in a highly non-linear fashion to enable perception of tens of thousands of unique odors [62]. Despite the equally analogous relationship between spectral analytical instrumentation and its utility in chemical identification, olfaction is arguably the motivating ideal behind much of the chemical sensor array work performed to date, and a reason why such devices have commonly been referred to as ”electronic noses” (or “electric tongues” for liquid sample sensors) and ”machine olfaction” in the technical literature. However, it should be noted that despite considerable overlap, electronic noses are not synonymous with chemical sensor arrays. First, the term “electronic nose” is a generic descriptor for odor-profiling devices, including those based on analytical instrumentation. Second, not all chemical sensing tasks are properly termed as ”olfaction” as this refers specifically to problems involving perception of odors.

In an early review of machine olfaction [27], Pearce discusses the origins of the approach as the result of the confluence of the development of chemical sensors in the 1950s and 1960s, and early work describing signal processing in neural olfaction systems in the late 1960s, which culminated in Persaud and Dodd’s first description of a three-sensor electronic nose in 1982 [63]. Following this, there was a rapid expansion of reports of sensor arrays for chemical detection, for examples, see [64–68].

Since this early work, sensor arrays of nearly every type of chemical sensor technology available have been reported in the literature. For example, Grate et al. have reported extensively on the use of SAW and other sorption-based sensor arrays in detection of chemical vapors [69–72]. Zellers and colleagues have reported microfabricated arrays of chemiresister and hybrid arrays applied to binary and ternary chemical mixtures [73–76]. Suslick and colleagues have reported colorimetric sensor arrays applied to many complex chemical sensing problems [47, 77–79]. Promising work on surface-functionalized nanostructures has recently garnered much attention [80–83] and Marco, et al., have developed a very large-scale array of conducting polymer chemical sensors for biomimetic olfaction [6].

An alternative to a physical array of sensors is the notion of ”adaptable” sensors that are capable of adjusting their operating parameters or environment to alter their response functions during use [84]. In this way, adaptable sensors acquire what is effectively sensor array data with a single sensing element at some cost in data acquisition time over a comparable multisensor array. They also open the potential to build sensor systems that can adapt to changing environments and sensing tasks, gaining efficiency by limiting the number of measurements made to only those that are necessary [85]. For example, microhotplate sensors precisely control the temperature of a metal oxide sensor surface, thus altering the sensor response characteristics in a repeatable, programmable fashion [86, 87]. This approach has been demonstrated to provide enhanced information for complex sensing tasks such as detection of chemical vapors in backgrounds of common commercial cleaning products. In other examples, sensors have been coupled to reagent chemistries, catalysts, or filters to enable variable selectivity [88–93]. Finally, sensors can be coupled to sampling devices that provide spatial or temporal separation of analyte mixtures prior to detection [28].

Numerous books and reviews of sensor array technology have since been authored with at least one published nearly every year since 1997, demonstrating the continued interest in and vitality of the approach [5, 20, 27, 28, 41, 42, 80, 86, 94–118]. The majority of these reviews are focused on describing various sensor technologies and applications for which sensor arrays have been used, while a smaller fraction focus instead on data analysis strategies for sensor arrays. However, only a few provide any focus on optimization strategies for arrays [84, 100, 101, 115, 116]. There have been also been periodic surveys of commercially available sensor-array devices, generally uncovering several new manufacturers with each iteration [5, 28, 105, 108, 112]. Applications of these arrays are frequently reported in the literature where they have been described as successfully applied to a wide variety of tasks. (See, for example those reviewed in [80])) However, it has been noted repeatedly that such devices do not seem to have made significant progress towards widespread commercial use [5, 6]. The reason for this is likely an overestimation of sensor array capabilities combined with an underestimation of sensing task complexity and a lack of sound validation and design principles [2–4]. As is often the case with new technologies, enthusiasm can outpace actual utility, leading to subsequent disappointment that perhaps unfairly clouds any realistic assessment of the capabilities of these devices.

### *3.2.2 Models of Sensor Response: Transduction Models and Sensor Noise*

In order to perform an information-theoretic analysis of a chemical sensor system, a functional model of sensor response must be developed. This model does not necessarily need to fully embody the underlying molecular interaction between the sensor and chemical stimulus, but the more completely it does, the more accurately the capabilities of a sensor system can be predicted and understood. Importantly, information-theoretic measures are generally agnostic with respect to the exact form of functional response, although they do require that the function be differentiable. This allows an information-theoretic approach to be applied across different sensor types and offer directly comparable assessments of sensor array capability. It should be noted that certain functional forms may preclude derivation of analytic expressions for some information theoretic measures, and will require numerical solution instead. A generic chemical sensor can be modeled as a function that transforms a particular chemical stimulus into a response value with a characteristic uncertainty according to a particular error distribution. The actual form of a sensor response function depends on the nature of the molecular interaction between the chemical stimulus and sensor, and thus on molecular properties of both. Thus, the underlying physics of the analyte-sensor interaction drive the nature of the three critical components of a sensor's functional response: mean response with concentration, measurement error distribution, and additivity of response to chemical mixtures. Consider an arbitrary chemical stimulus with k chemical components, represented as a vector of k concentrations,

$$\boldsymbol{c} = [c_1, c_2, c_3, \ldots, c_k,] \tag{5}$$

The observed response, r, of a chemical sensor to this stimulus has a mean that is a function of c, and is distributed according to a measurement error distribution with parameters, $\boldsymbol{\theta}$.

$$r = \mu(\boldsymbol{c}) + \varepsilon(\boldsymbol{\theta}) \tag{6}$$

Many chemical sensors exhibit approximately linear mean response with concentration, at least over particular concentration ranges, which gives the relationship

$$\mu = a\,c \tag{7}$$

where $a$ is a constant corresponding to the sensitivity the sensor exhibits to a particular analyte. For example, with surface acoustic wave sensors (and other MEMS sensors) response is proportional to the mass of analyte absorbed by a polymer coating from vapor, and thus linear with vapor concentration for sensors with thin films and relatively low analyte vapor concentrations [42]. Likewise, sensors based on an integrated optical absorbance measurement are known to provide a linear response with respect to analyte concentration [119].

However, there is often uncertainty regarding the exact form of functional response a given sensor technology exhibits. Assessments of functional response typically come from empirical fitting to measured data and theoretical models of response based on a hypothesized sensor transduction mechanism. For example, consider sensors based on conducting polymers. Ellis, et al. [120] observed that a particular sensor they were investigating exhibited a response that could be approximated as an exponential of the form shown below,

$$\mu = b\,e^{ac} \tag{8}$$

A review of conducting polymers described published work in which conducting polymer sensor response was modeled according to a Langmuir isotherm model of the form,

$$\mu = b\left(\frac{ac}{1+ac}\right) \tag{9}$$

under the assumption that analyte absorption to the polymer was governed by a Langmuir adsorption process [121]. Still other work cited in this review started from an assumption that analyte absorption is governed by a partitioning process between the gas phase and polymer phase, suggesting a linear dependence as shown in eqn. (**??**) instead.

Other types of sensors have been reported to exhibit other non-linear response with concentration. Metal oxide sensors have long been observed to exhibit a power law dependence,

$$\mu = b\,(c^{a}) \tag{10}$$

which has been explained via various theoretical models [122–124].

And biological systems, such as olfactory receptor neurons are often modeled as sigmodal functions, such the logistic function [125]

$$\mu = \left( \frac{a}{1 + e^{-b(c-d)}} \right) \tag{11}$$

or the Hill equation [126],

$$\mu = \left( \frac{a}{1 + (\frac{d}{c})^b} \right) \tag{12}$$

Which are generally interpreted in the context of receptor-ligand binding interactions.

Depending on the mechanism by which it functions, chemical sensors can exhibit various response dependences with other variables related to sensor construction or the physical environment around the sensor. In scenarios where these variables can't be held constant, their effect should also be considered to evaluate their impact in addressing sensing tasks. In some situations, time-dependent variation of sensor parameters can be used to systematically vary analyte response functions, effectively generating multiple distinct responses from a single sensor [86].

Next, we consider the manner in which sensors respond to mixtures. Again, this is highly dependent on the particular mechanism by which the sensor interacts with the chemical stimulus. Generally, if one can assume that the different components of a mixture interact with a sensor in an independent fashion, then mixture response can be modeled as additive, as shown in the example below for a sensor exhibiting linear response to the k different components of a chemical mixture

$$\mu = \sum_{i=1}^{k} a_i c_i \tag{13}$$

For many sensors, this assumption is reasonable for chemical stimuli at relatively low concentrations, becoming less valid at high concentrations as the sensor response saturates and competitive interactions between analytes become more significant. In some systems, interactions between the sensor and different mixture components cannot be considered independent, and specific analyte-analyte dependencies must be explicitly incorporated into the response model. For example, olfactory receptor neurons are known to exhibit non-additive response to chemical mixtures, due to competitive binding and potential masking effects induced by one component of the mixture upon one or more others [127, 128].

The final component of a sensor response model is the measurement error distribution. While this is a critical aspect of observed sensor response, it is also typically one of the least characterized aspects in published work on sensor development, likely due to its perception as a nuisance parameter. However, as implied in the discussion of the previous section, the specific statistical qualities of sensor measurement error can have profound implications in information-theoretic analysis of sensor arrays. In particular, different noise distributions lead to different expressions of information-theoretic measures, and either allow or preclude tractable analytical expressions of these measures for sensor arrays.

In general, sensor measurement error is the result of stochastic fluctuations within the molecular interaction with the sensor, or the transduction process employed by the sensor. Commonly, these are either Gaussian or Poisson processes. For example, all electronic circuits exhibit Johnson-Nyquist, or thermal noise, which exhibits a Gaussian amplitude distribution. Certain detectors, such as photomultiplier tubes, may exhibit Shot noise, which is Poisson-distributed. Other specific processes may exhibit noise distributions that are better modeled by other statistical distributions, such as with systems involving bio or chemiluminescence, which have been observed to exhibit Laplacian-distributed noise [26, 129].

If there is no known noise model for a sensor system, an assumption of Gaussian-distributed measurement error is usually a reasonable approximation. This is because measurement error is generally the aggregate result of a series of stochastic fluctuations from different sources, and thus tends to be normally distributed according to the central limit theorem [130]. Further, from an information-theoretic perspective, Gaussian distributions represent the hypothetical "worst case" scenario for a distribution of a random variable, in that it conveys the least amount of information and thus will provide conservative estimates sensor capability in the event that the measurement error is not actually Gaussian [130]. Conversely, deviations from this common assumption of independent, identically distributed Gaussian noise can be leveraged to provide additional chemical information from sensor measurements. Next, measurement error is often assumed to homogenous, but error models may also explicitly incorporate concentration dependence if necessary. Finally, although measurement error is generally assumed to be independent from sensor to sensor, it is possible for correlation in noise to occur in a sensor array, and this correlation may be significant enough to require explicit consideration in an informatics analysis.

### 3.3 Fisher Information

Both Fisher information (FI) and its generalization to multi-parameter estimation, the Fisher information matrix (FIM), are relevant to the design of statistical estimators (i.e. sensors) as their respective inverses act as lower bounds to the (co)variances of the subject estimator, a property which is referred to as the Cramér-Rao lower bound [11].

To more concretely motivate this assertion, consider a chemical sensor array response, $\boldsymbol{\mu}(\boldsymbol{\theta})+\boldsymbol{\delta}(\boldsymbol{\theta})$, where $\boldsymbol{\mu}(\boldsymbol{\theta})$and $\boldsymbol{\delta}(\boldsymbol{\theta})$are the idealized sensor response vector and noise vector respectively. $\boldsymbol{\theta}$ denotes an external parameter vector which is environmentally dependent. For chemical sensors and sensor arrays, this is typically the analyte concentration vector. Such a sensor array response may then be modeled with a probability density function, $\rho(\boldsymbol{X};\boldsymbol{\mu}(\boldsymbol{\theta}))$ [25], as follows, $\boldsymbol{\mu}(\boldsymbol{\theta}) = \int d\boldsymbol{X}\, \boldsymbol{X}\rho(\boldsymbol{X};\boldsymbol{\mu}(\boldsymbol{\theta}))$, with a covariance matrix given by,

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \int d\boldsymbol{X}\,(\boldsymbol{X}-\mu(\boldsymbol{\theta}))(\boldsymbol{X}-\mu(\boldsymbol{\theta}))^T\rho(\boldsymbol{X};\boldsymbol{\mu}(\boldsymbol{\theta})) \tag{14}$$

A typical goal of sensor array optimization is to minimize the global error of the sensor array. This quantity is captured by det($\boldsymbol{\Sigma}(\boldsymbol{\theta})$), the determinant of the covariance matrix. Since $\boldsymbol{\Sigma}(\boldsymbol{\theta})$is a positive definite matrix, its determinant describes a strictly positive volume that may act as a score or metric for the global error [131]. Thus, from the standpoint of global noise, the goal of the sensor array designer is to minimize this determinant.

Unfortunately, it is often either impractical or computationally intensive to directly calculate $\boldsymbol{\mu}(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ in a way that allows for the analytic optimization and design of arrays, particularly if a complicated estimator is used. It is also worth considering that many different physical sensor system setups or statistical estimators may be constructed for the same system i.e. sensor response probability distribution. This multitude of specific estimator possibilities forces the practitioner to seek a design criterion that is robust in the face of many potentially similar but varying covariance matrices or array responses. Fortunately, via the Cramér-Rao inequality, FI/FIM provide a lower bound in the positive definite sense for the covariance matrix of such a sensor array that is independent of the actual estimator being used. This provides a useful expression of the fundamental analytical potential of the device. Importantly, if the practitioner tunes or re-tunes their setup, this quantity will never change. Thus, we conclude that the FI/FIM provides a robust metric to optimize in the design of chemical sensor arrays.

### *3.3.1 The Cramer-Rao Bound*

Before showing how to utilize the FI/FIM in the context Further manipulation of the integrand yields optimization for sensor selection, it is informative to first derive the FI/FIM relation to the Cramér-Rao lower bound. As a prelude, the FI is defined as [132],

$$f(\mu;\theta) = \int dx \rho(\mu;\theta|x) \left(\frac{\partial \ln(\rho)}{\partial \theta}\right)^2 \tag{15}$$

and each element of the FIM itself is defined as,

$$\boldsymbol{F}(\mu;\boldsymbol{\theta})_{ij} = \int d\boldsymbol{X} \rho(\mu;\boldsymbol{\theta}|\boldsymbol{X}) \left(\frac{\partial \ln(\rho)}{\partial \theta_i}\right)\left(\frac{\partial \ln(\rho)}{\partial \theta_j}\right) \tag{16}$$

with the FIM reducing to the FI in the univariate case. Informally, the FI/FIM may be thought of as conveying how much information an observed random variable, $x$, or set of random variables, $\boldsymbol{X}$, carry about a parameter(s), $\theta$ or $\boldsymbol{\theta}$.

In the event those deterministic parameters $\boldsymbol{\theta}$ are being statistically estimated, the FI/FIM provides a lower bound to their (co)variance independent of the employed statistical estimator(s). Beginning with the univariate case, the FI may be derived [133] by first considering the following expectation value,

$$\boldsymbol{E}[\hat{\theta}(x) - \theta] = \int \left(\hat{\theta}(x) - \theta\right) \rho(x;\theta) = 0 \tag{17}$$

where $\hat{\theta}(x)$ is an unbiased statistical estimator for $\theta$. Next, differentiating by the deterministic parameter yields,

$$\frac{\partial}{\partial \theta} \int dx \left(\hat{\theta}(x) - \theta\right)\rho(x;\theta) = \int dx \left(\hat{\theta}(x) - \theta\right)\frac{\partial \rho}{\partial \theta} - \int dx\rho = 0 \tag{18}$$

Recognizing that since $\rho$ is a probability distribution,

$$\int dx \rho(x;\theta) = 1 \tag{19}$$

and

$$\frac{\partial \rho}{\partial \theta} = \rho \frac{\partial \ln(\rho)}{\partial \theta} \tag{20}$$

which implies

$$\int dx \left(\hat{\theta}(x) - \theta\right) \rho(x;\theta) \frac{\partial \ln(\rho)}{\partial \theta} = 1 \tag{21}$$

Further manipulation of the integrand yields,

$$\int dx \left(\left(\hat{\theta}(x) - \theta\right) \sqrt{\rho}\right) \left(\frac{\partial \ln(\rho)}{\partial \theta} \sqrt{(\rho)}\right) = 1 \tag{22}$$

Applying the Cauchy-Schwartz inequality[1] to this manipulated integrand gives,

$$\begin{aligned} &\left(\int dx \left(\left(\hat{\theta}(x) - \theta\right) \sqrt{\rho}\right) \left(\frac{\partial \ln(\rho)}{\partial \theta} \sqrt{(\rho)}\right)\right)^2 = 1^2 \\ =&1 \leq \left(\int dx \rho \left(\hat{\theta}(x) - \theta\right)^2\right) \left(\int dx \rho \left(\frac{\partial \ln(\rho)}{\partial \theta}\right)^2\right) \end{aligned} \tag{23}$$

After some manipulation of the preceding integrand, the expression resolves itself as,

$$\frac{1}{\left(\int dx \rho \left(\frac{\partial \ln(\rho)}{\partial \theta}\right)^2\right)} \leq \left(\int dx \rho \left(\hat{\theta}(x) - \theta\right)^2\right) \rightarrow \frac{1}{f(\theta)} \leq \left(\mathrm{Var}(\hat{\theta})\right) \tag{24}$$

which is the Cramér-Rao lower bound for the univariate case.

The derivation of the Fisher information matrix (FIM) for the multivariate case [132] is performed in a similar fashion to the univariate case by first considering,

$$\boldsymbol{E}[\hat{\boldsymbol{\theta}}(\boldsymbol{X}) - \boldsymbol{\theta}] = \int d\boldsymbol{X} \rho(\boldsymbol{X}|\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}(\boldsymbol{X}) - \boldsymbol{\theta}) = \boldsymbol{0} \tag{25}$$

And then differentiating this equation so that,

$$\begin{aligned} &\partial_{\boldsymbol{\theta}} \int d\boldsymbol{X} \rho(\boldsymbol{X}|\boldsymbol{\theta}) \left(\hat{\boldsymbol{\theta}}(\boldsymbol{X}) - \boldsymbol{\theta}\right) = \\ &\int d\boldsymbol{X} \left(\partial_{\boldsymbol{\theta}} \rho(\boldsymbol{X}|\boldsymbol{\theta})\right) \left(\hat{\boldsymbol{\theta}}(\boldsymbol{X}) - \boldsymbol{\theta}\right) - \underbrace{(\partial_{\boldsymbol{\theta}} \boldsymbol{\theta})}_{\boldsymbol{I}} \underbrace{\int d\boldsymbol{X} \rho(\boldsymbol{X}|\boldsymbol{\theta})}_{1} = 0 \end{aligned} \tag{26}$$

---

[1] $|\langle \boldsymbol{u}, \boldsymbol{v}\rangle|^2 \leq \langle \boldsymbol{u}, \boldsymbol{u}\rangle \cdot \langle \boldsymbol{v}, \boldsymbol{v}\rangle$ where $\boldsymbol{u}$ and $\boldsymbol{b}$ are vectors with the inner product $\langle \cdot, \cdot \rangle$

where $\partial_{\boldsymbol{\theta}}$ indicates a derivative with respect to the vector $\boldsymbol{\theta}$. Rearranging terms as before, it becomes,

$$\begin{aligned}&\int d\boldsymbol{X}\rho\partial_{\boldsymbol{\theta}}\ln(\rho)\left(\hat{\boldsymbol{\theta}}(\boldsymbol{X})-\boldsymbol{\theta}\right)=\\&\int d\boldsymbol{X}\left((\hat{\boldsymbol{\theta}}(\boldsymbol{X})-\boldsymbol{\theta})\sqrt{\rho}\right)(\sqrt{\rho}\partial_{\boldsymbol{\theta}}\ln(\rho))=\boldsymbol{I}\end{aligned}\tag{27}$$

and applying the Cauchy-Schwartz inequality gives,

$$\boldsymbol{I}\leq\int d\boldsymbol{X}\rho\left(\left(\hat{\boldsymbol{\theta}}(\boldsymbol{X})-\boldsymbol{\theta}\right)\left(\hat{\boldsymbol{\theta}}(\boldsymbol{X})-\boldsymbol{\theta}\right)^T\right)\cdot\int d\boldsymbol{X}\rho\left((\partial_{\boldsymbol{\theta}}\ln(\rho))(\partial_{\boldsymbol{\theta}}\ln(\rho))^T\right)\tag{28}$$

so that the FIM provides a lower bound to the covariance matrix,

$$\boldsymbol{F}(\boldsymbol{\mu};\boldsymbol{\theta})^{-1}\leq\boldsymbol{\Sigma}(\boldsymbol{\theta})\tag{29}$$

with the $\leq$ relation is in the sense of a positive definite matrix and the FIM and covariance matrix ($\boldsymbol{F}$and $\boldsymbol{\Sigma}$ respectively) defined as,

$$\boldsymbol{F}(\boldsymbol{\mu};\boldsymbol{\theta})=\int d\boldsymbol{X}\rho\left((\partial_{\boldsymbol{\theta}}\ln(\rho))(\partial_{\boldsymbol{\theta}}\ln(\rho))^T\right)\tag{30}$$

and

$$\boldsymbol{\Sigma}(\boldsymbol{\theta})=\int d\boldsymbol{X}\rho\left(\hat{\boldsymbol{\theta}}(\boldsymbol{X})-\boldsymbol{\theta}\right)\left(\hat{\boldsymbol{\theta}}(\boldsymbol{X})-\boldsymbol{\theta}\right)^T\tag{31}$$

Clearly, the so-derived FIM also implies the univariate case.

### 3.4 The Fisher Information Matrix Formulated for Chemical Sensing

The literature contains multiple examples regarding explicit treatment of the manner in which sensor measurement error impacts analytical performance of sensor arrays. Gardner and Bartlett described an approach [134] to characterizing sensor array performance based on a ratio of sensor measurement standard error volume to the total volume of the sensor array measurement space, i.e. the geometric n-dimensional space described by the span of all possible response values from an array of n sensors. Performance could then be described in terms of the maximum number of resolvable regions in the measurement space of the array. Pearce and colleagues [25] extended this approach to incorporate the reduction in resolving capacity induced by correlation in response functions among sensors, and then demonstrated an approach to directly propagating sensor measurement error into sample mixture space using an inverse of the geometric transform implied by the sensor array response functions. The sample space is an m-dimensional geometric space describing the span of all possible mixtures containing some ratio of m analytes to be sensed. The sample space and measurement space are related by the analyte-specific sensor response functions. By explicitly casting chemical analysis as a statistical estimation problem, they demonstrated that the Fisher information derived from this set of response functions could be used to provide a lower bound on the expected variance of estimating sample mixture concentration from sensor measurements and demonstrated this approach's connection to the geometric error propagation model considered earlier.

The Fisher information provides a measure of the capability of an observed variable to convey information regarding an unknown parameter to which it is functionally related, in the presence of measurement error. Consider a set of parameters, $\theta$, that are functionally related to a set of measured variables, $X$. Each variable is associated with some type of measurement noise, and is thus characterized by a probability distribution, $p(X|\theta)$. The elements of the Fisher information matrix (FIM) for prediction of $\theta$ are given by eqn. 32.

$$J_{jj'}(\theta) = \int d\mathbf{X} p(\mathbf{X}|\theta) \left(\frac{\partial}{\partial \theta_j} \ln p(\mathbf{X}|\theta)\right) \left(\frac{\partial}{\partial \theta_{j'}} \ln p(\mathbf{X}|\theta)\right) \tag{32}$$

The inverse of the FIM is related to the expected covariance matrix of an unbiased estimator of the parameters, $\hat{\theta}$, via the Cramér-Rao inequality,

$$J^{-1}(\theta) \leq E\left[\left(\theta - \hat{\theta}\right)\left(\theta - \hat{\theta}\right)^T\right] = \mathrm{cov}(\theta), \quad \left|J^{-1}(\theta)\right| \leq |\mathrm{cov}(\theta)| \tag{33}$$

where the matrix inequality is in the positive semidefinite sense. In this way, the Fisher information can be used to provide a lower bound for the variance of any such estimator, and thus quantitative measures regarding the fundamental capability of a given measurement system to estimate a set of parameters. Importantly, the variance bound derived from the Fisher information is independent of the post hoc data analysis techniques that may be applied to generate those estimates, and is thus a measure of the potential analytical capability of the measurement system, rather than a direct prediction of what may be observed in practice. Fisher information matrices are currently used in a wide variety of fields and applications. For example, they are the basis of optimization criteria commonly used in computer-aided design of experiments (DOE) [135]. In both DOE and sensor array design, a varied set of design choices are combined with a set of external constraints which, taken together, provide a design space for optimization. In DOE, the parameters to be estimated are the coefficients of the functional relationship between independent and dependent experimental variables. The goal is to create an optimized experimental design that most efficiently enables accurate and efficient estimation of these coefficients, and thus an understanding of how the variables are related. In sensor array optimization, the parameters to be estimated are chemical concentrations. The sensor response functions to the library of target chemicals are known a priori with the goal being to create an optimized set of sensors that best enables accurate prediction of chemical mixture composition from measured sensor responses.

Pearce proposed using the trace of the inverse of the FIM as a criterion for sensor array optimization. The minimization of this quantity is termed A-optimization in DOE, and can be interpreted as minimizing the average variance of the estimated coefficients. However, care should be taken with this interpretation in a sensor array optimization context. The diagonal of the inverse of the FIM only strictly corresponds to minimum variance bounds for each mixture component concentration when the sensor response vectors to each chemical component are orthogonal. While the rows of a design matrix in DOE are commonly orthogonal by design, this is often not the case with response arrays of non-specific sensors. Another commonly used optimization criterion is the determinant of the inverse FIM. Minimization of this criterion is termed D-optimization in DOE. This results in minimizing the global variance of the estimated coefficients and can lead to non-orthogonal design matrices. As discussed previously, non-orthogonal sensor response vectors are commonly observed and not necessarily problematic for sensor array optimization. Because the determinant criterion convolves variance from all of the estimated parameters, it effectively expresses a global measure of average discernibility among adjacent regions of the analyte mixture space in a signal-to-noise ratio sense. Equivalently, maximization of the determinant of the FIM could be thought of as a maximization of the global sensitivity of the array to displacements in the sample space. However, neither the trace nor the determinant criteria provide an immediate measure of analytical task-specific selectivity for sensor array configurations with non-orthogonal responses.

### 3.5 Chemical Sensing in an "Open World": Environmental Models for Chemical Noise

In this section, we will analyze two different probability distributions, the multivariate normal distribution and the multivariate lognormal distribution These distributions will be used to describe both the responses of our chemical detectors and the concentrations of chemical interferents present in the background environment.

The multivariate normal distribution, $\mathscr{N}(\boldsymbol{X};\boldsymbol{\mu},\boldsymbol{\Sigma})$, is defined as

$$\mathscr{N}(\boldsymbol{X};\boldsymbol{\mu},\boldsymbol{\Sigma}) = |2\pi\boldsymbol{\Sigma}|^{-1/2} e^{-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})} \tag{34}$$

where $\boldsymbol{X}$ is a random vector whose individual members range over $(-\infty,\infty)$, $\boldsymbol{\mu}$ is the vector of parameters which dictate the means of the assi and $\boldsymbol{\Sigma}$ is the covariance matrix. Its moment generating function is given by $M(t) = e^{\boldsymbol{\mu}\cdot\boldsymbol{t}+\frac{1}{2}\boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t}}$ where taking the derivative the number of times equal to the order of the moment desired with respect to the appropriate variables and setting $\boldsymbol{t} = \boldsymbol{0}$ yields the desired value(s).

The multivariate lognormal distribution, $\mathscr{L}\mathscr{N}(\boldsymbol{X};\boldsymbol{\mu},\boldsymbol{\Sigma})$, is defined as

$$\mathscr{L}\mathscr{N}(\boldsymbol{X};\boldsymbol{\mu},\Sigma) = |2\pi\Sigma|^{-1/2}\left(\prod_{i=1}^{d} x_i^{-1}\right) e^{-\frac{1}{2}(\ln X-\boldsymbol{\mu})^T\Sigma^{-1}(\ln X-\boldsymbol{\mu})} \tag{35}$$

$\boldsymbol{X}$ is the random variable ranging over $(\boldsymbol{0},\infty)$, $\boldsymbol{\mu}$ is a parameter vector and $\boldsymbol{\Sigma}$ is a positive definite scale matrix (matrix of parameters). The $\ln X$ term denotes a vector where each element is $\ln X_i$. The mean values of $X_i$ may be calculated following Halliwell [136] as,

$$\langle X_i\rangle = e^{\boldsymbol{e}_i\cdot\boldsymbol{\mu}+\frac{1}{2}\boldsymbol{e}_i^T\Sigma\boldsymbol{e}_i} = e^{\mu_i+\frac{1}{2}\Sigma_{ii}} \tag{36}$$

where $e_i$ is a standard basis vector of appropriate dimension with a scalar 1 in the $i$th dimension and scalar 0 in all other dimensions. The covariance for the multivariate lognormal distribution is given by [136],

$$Cov(X_i,X_j) = \langle X_iX_j\rangle - \langle X_i\rangle\langle X_j\rangle = \langle X_i\rangle\langle X_j\rangle\left(e^{\Sigma_{ij}}-1\right) \tag{37}$$

The parameters $\mu_i$ and $\Sigma_{ij}$ for the multivariate lognormal distribution may be expressed in terms of the desired mean $\langle X_i\rangle$ and the desired covariance values $Cov(X_i,X_i)$ and $Cov(X_i,X_j)$ as

$$\mu_i = \log\left(\frac{\langle X_i\rangle^2}{\sqrt{\langle X_i\rangle^2 + Cov(X_i,X_i)}}\right) \tag{38}$$

and

$$\Sigma_{ij} = \log\left(1 + \frac{Cov(X_i, X_j)}{\langle X_i\rangle\langle X_j\rangle}\right) \tag{39}$$

The concentrations of the chemical interferents in the background environment are described by the multivariate lognormal distribution. This distribution has been repeatedly found to describe lower bounded phenomena like pollutant concentrations, prices, and organism counts [137–141]. Background chemical concentrations will also be modeled using the multivariate normal distribution since this distribution has been shown to be a reasonable when interferents are either unnaturally distributed due to machinery/human intervention or when the concentrations have a empirically measured variance such that the lowest $2\sigma$ boundary is above 0 mols/volume [140].

Using these probability models for detector response and chemical intereferent concentrations in the external, 4 different distributions which jointly describe the detector and environment may be derived. We generically denote these joint distributions by the following,

$$P_{D+E}(\boldsymbol{X}_D, \boldsymbol{X}_E, \boldsymbol{B}_E; \boldsymbol{\mu}_D(\boldsymbol{X}_E, \boldsymbol{b}), \boldsymbol{\mu}_E, \boldsymbol{\Sigma}_D(\boldsymbol{X}_E), \boldsymbol{\Sigma}_E) P_{E,B}(\boldsymbol{B}_E) \tag{40}$$

where $P_{E,B}(\boldsymbol{B}_E)$ is the discrete boolean distribution for the presence and absence of the various chemical interferents. separated out The response of the detector is modeled as the moment $\langle X_{D,i}\rangle_D$ of the detector's probability distribution which represents the response of a sensor or instrument channel $i$. For this paper we assume that this response is linear with respect to the detected chemical concentrations so that

$$\langle X_{D,i}\rangle_D = \sum_j a_j c_j + \sum_k a_k b_k X_{E,k} \tag{41}$$

where the linear response coefficients which convert a concentration into a detection response are given by either $a_j$ or $a_k$, the known chemicals' concentrations are given by $c_j$ and the concentrations of the unknown chemical interferents are given by the random variables associated with the external chemical environment $X_{E,k}$. Finally, $b_k$ are boolean random variables associated with the presence or absence of a chemical interferent. Since $b_k$ are random variables with a discrete distribution, they may be safely ignored in the following derivation which focuses on integrating out the continous random variables. Our final computation will sum over discrete boolean variables.

We assume that the (co)variance of the detector's variables are independent of the concentrations of the chemical interferents. Thus for a chemical detector with a response modeled as a multivariate normal distribution the parameters of that distribution may be modeled as

$$\mu_i = \sum_j a_j c_j + \sum_k a_k b_k X_{E,k} \tag{42}$$

and $\boldsymbol{\Sigma}$ is just a matrix of constant scalars. The 2 possible distributions for the joint response of the detector and environment which are solvable in closed-form for the KLD measures given by eqns. (107) and (108) are given by

$$\mathscr{N}_D(\boldsymbol{X}_D;\boldsymbol{\mu}_D(\boldsymbol{X}_E,\boldsymbol{B}_E),\boldsymbol{\Sigma}_D)\,\mathscr{N}_E(\boldsymbol{X}_E;\boldsymbol{\mu}_E,\boldsymbol{\Sigma}_E)P_{E,B}(\boldsymbol{B}) \tag{43}$$

$$\mathscr{N}_D(\boldsymbol{X}_D;\boldsymbol{\mu}_D(\boldsymbol{X}_E,\boldsymbol{B}),\boldsymbol{\Sigma}_D)\,\mathscr{L}\mathscr{N}_E(\boldsymbol{X}_E;\boldsymbol{\mu}_E,\boldsymbol{\Sigma}_E)P_{E,B}(\boldsymbol{B}) \tag{44}$$

where $\boldsymbol{\mu}_D$ and $\boldsymbol{\Sigma}_D$ are defined as given earlier in this section. The distribution denoted by $\mathscr{P}_{E,B}(\boldsymbol{B})$ is the discrete distribution for the presence or absence of each chemical interferent.

The joint detector and environment distributions which require Monte Carlo evaluation of some of their integrals for the KLD measures given by eqns. (107) and (108) are given by

$$\mathscr{L}\mathscr{N}_D(\boldsymbol{X}_D;\boldsymbol{\mu}_D(\boldsymbol{X}_E,\boldsymbol{B}_E),\boldsymbol{\Sigma}_D(\boldsymbol{X}_E,\boldsymbol{B}_E))\,\mathscr{N}_E(\boldsymbol{X}_E;\boldsymbol{\mu}_E,\boldsymbol{\Sigma}_E)P_{E,B}(\boldsymbol{B}) \tag{45}$$

$$\mathscr{L}\mathscr{N}_D(\boldsymbol{X}_D;\boldsymbol{\mu}_D(\boldsymbol{X}_E,\boldsymbol{B}_E),\boldsymbol{\Sigma}_D(\boldsymbol{X}_E,\boldsymbol{B}_E))\,\mathscr{L}\mathscr{N}_E(\boldsymbol{X}_E;\boldsymbol{\mu}_E,\boldsymbol{\Sigma}_E)P_{E,B}(\boldsymbol{B}) \tag{46}$$

All parameters have the same definitions as given earlier in this section. The evaluation of the suppressed boolean variables which represent the presence or absence of particular chemical interferents will be evaluated by either direct summations or via Monte Carlo simulation depending on the size of the chemical interferent library.

For the purposes of mathematical convenience, we have opted to model the inclusion and exclusion of specific chemical interferents or groupings of chemical interferents as a binary distribution separate from the one that determines the proposed concentration of those chemical interferents. To include this binary distribution, we multiply the concentrations of the chemical interferents in the sensor response term by a boolean variable $b_k$ so that it becomes

$$\langle X_{D,i}\rangle_D = \sum_j a_j c_j + \sum_k b_k a_k X_{E,k} \tag{47}$$

In this paper, we are assuming that the (co)variances of the detector response are independent of the chemical intereferent concentrations, they may be made chemical interferent presence-dependent using a similar approach to that used for the detection response. For eqns. (107) and (108), the KLD measures when evaluated with respect to the binary variables become

$$\langle D_{KLD}(P^{TA}_{D+E}(\boldsymbol{X}_D;\boldsymbol{X}_E)||P^{S}_{D+E}(\boldsymbol{X}_D;\boldsymbol{X}_E))\rangle_B = D_{KLD}(P^{TA}_{E,B}||P^{S}_{E,B}) + \langle D_{KLD}((\mathscr{L})\mathscr{N}^{TA}_D||(\mathscr{L})\mathscr{N}^{S}_D)\rangle_{E,B} \tag{48}$$

where the average over the binary variables may be evaluated via either direct summation or Monte Carlo sampling. $D_{KLD}(P^{TA}_B||P^{S}_B)$, the KLD between the boolean distributions associated with the environments of the TA and the CS, may be evaluated as

$$D_{KLD}(P^{TA}_{E,B}||P^{S}_{E,B}) = \sum_i P^{TA}_{E,B}(B_i)\log\frac{P^{TA}_{E,B}(B_i)}{P^{S}_{E,B}(B_i)} \tag{49}$$

## 4. FIGURES OF MERIT FOR SENSOR ARRAYS

Analytical chemistry uses a variety of figures of merit (FoMs) for quantifying and ranking the quality and capabilities of various aspects of chemical measurement and instrumentation. Among the FoMs which are considered to be basal to analytical chemistry are: sensitivity, selectivity/specificity, limit of detection, limit of recognition, and resolving power. Given this basal nature, it is thus surprising that in their formal IUPAC definitions [142], these FoMs fail to capture multivariate descriptions of their subject quantities and often are only qualitatively described. In the following sections, we briefly describe either available multivariate FoMs for sensor arrays and analytical instruments from the literature or present our own work on the subject.

### 4.1 Sensitivity

Sensitivity is defined and used in multiple, distinct ways in fields that the inexperienced practitioner might assume would share common definitions owing to their similar abstract goals like chemical detection or estimation. Complicating this state of affairs is that analytical chemists occasionally use the term sensitivity as it is defined in other fields without detailing the difference. If we first consider IUPAC's approach [142], sensitivity is defined to be the slope of the sensor response with respect to a particular analyte concentration. This is simply a constant in the case of an individual sensor with linear response, and can be more generally expressed as the derivative of the sensor response function with respect to analyte concentration. A multivariate generalization for sensor arrays may then be defined as a vector of derivatives of sensors/sensor channels with respect to the chemical stimulus. The multivariate sensitivity of a sensor array for multiple analytes thus generalizes as the Jacobian matrix whose elements consist of the partial derivatives of each sensor response function with each analyte. This matrix offers a purely geometric description of the transform between chemical stimulus space and array response space, with sensitivity corresponding to the extent to which responses to adjacent chemical stimuli are separated in the response space. Multivariate sensitivity can be thought of, then, in a global sense as a distortion of the chemical stimulus space as it is projected into the sensor array response space. However, if we want to connect sensitivity to the term's broader usage in fields of study which concern themselves with detection and estimation in areas beyond analytical chemistry, such as signal processing, medicine, and machine learning, we must be cautious, as the term is used with different meaning, owing to its definition by the statistics community [143]. In these fields, sensitivity is generally defined as the true positive rate of a binary classifier, or in terms of chemical detection, the rate at which a chemical is successfully detected when it is present. While at first glance, these two concepts appear completely distinct, there is an underlying connection between the two. As previously discussed, chemical sensing tasks can be systematically framed as binary classification problems in which one seeks to discern one region of the chemical stimulus space from another. When multivariate sensitivity to the chemical stimulus increases, so does the relative displacement in response space to infinitesimal changes in stimulus concentration. Increased displacement along a given axis of the response space associated with a sensing task thus has the impact of reducing the overlap in measurement error distribution between locations associated with that task, necessarily increasing the ability of an optimal classifier to correctly detect one from another. This demonstrates that the two notions are congruent in spirit, although not identical in a chemical sensing context.

The Fisher information itself can be considered as a measure of multivariate sensitivity in a chemical detection context. In particular, it can be derived as the Hessian of the relative entropy of array response distribution as a function of chemical stimulus, and thus expresses the local curvature of this quantity in chemical stimulus space [144]. This corresponds to an informational distance between adjacent locations, which, again, relates directly to the ability of an optimal classifier to discern between them. As discussed previously, the Fisher information is inversely related to a lower bound on the estimation error when an unbiased estimator is used to infer a location in chemical stimulus space from sensor array response. The advantage of using Fisher information as a sensitivity measure is that both sensor error and the geometric projection provided by the sensor array response functions are taken into account by this metric.

### 4.2 Selectivity

Viewed as a chemical detection system, biological olfaction is a large sensor array consisting of hundreds olfactory receptor neurons. These receptor neurons do not generally exhibit specificity with respect to any single chemical compound, but rather provide varying levels of response to multiple compounds. While the precise mechanism for olfactory perception remains unclear, the aggregate responses of this sensor array provide the fundamental chemical information that drives a capability to distinguish tens of thousands of unique odors. This suggests that there might be significant potential in building chemical sensor arrays consisting of collections of non-specific sensors for multiple analyte detection, rather than striving to build fully selective sensors for each [3]. Interestingly, the literature and application of arrays of non-specific sensors for general purpose chemical detection presents an incongruous mixture of apparent enthusiasm for the approach and general disappointment upon implementation. Although such arrays have been frequently reported and have been the subject of numerous reviews over the past 30 years, there remain relatively few examples of successfully commercialized devices. Tellingly, the topic of sensor array design has been infrequently explored, either independently, or in particular depth alongside reports of sensor array development and evaluation [145]. Rather, the focus in such research is often on the individual sensors themselves or on the application for which the sensor array is to be used. While sensor arrays can be empirically evaluated in much the same manner as any analytical device, prediction of their performance for realistic, complex scenarios remains elusive. Thus, sensor array design, optimization, and implementation remains time-intensive and costly, leading to arrays that, when fielded, tend to underperform relative to observed laboratory testing. In this work, we explore the utility of Fisher information-derived metrics for evaluation of analytical capability in arrays of non-specific chemical sensors, directing attention specifically to notions of analytical selectivity and the tradeoffs between selectivity and other sensor array performance measures. First we review contemporary notions of selectivity in analytical chemistry, followed by a discussion of the work of Pearce, et. al [25] in developing Fisher information as a measure of sensor array performance. The relationship between the estimation error bounds provided by Fisher information and selectivity is then developed using the Fisher information Bayes rule analog of Zachariah and Stoica [146] to provide a generalized selectivity measure for sensor arrays that is both congruent with previous selectivity definitions and applicable to a wider variety of potential response functions. Finally, the implications of this selectivity measure on sensor array design are explored and discussed using simulated sensor array examples.

Recently, Cramér-Rao bound analogs of the chain rule and Bayes rule have been described in the literature [146]. The derivation of these analogs involves defining joint and conditional variance bounds, and allows exploration of the impact estimation of one subset of analytes imposes on estimation of another. This enables quantitative assessment of the effect a chemical background imposes on the fundamental analytical capability of a given sensor array, and thus a measure of selectivity that is rooted both in the geometry of sensor response functions and in the measurement error of the sensors.

Zachariah and Stoica derived Cramér-Rao bound analogs of the chain rule and Bayes rule as follows. Defining $CRB(\theta) \triangleq \left|J^{-1}(\theta)\right|$, and dividing the parameter set into two groups $\theta = \left[\theta_\alpha^T \, \theta_\beta^T\right]^T$, the joint bound of $\theta_\alpha$ and $\theta_\beta$ is given by $CRB(\theta)$,

$$CRB(\theta) = CRB(\theta_\alpha, \theta_\beta) = \left| \begin{bmatrix} J_\alpha & J_{\alpha\beta} \\ J_{\beta\alpha} & J_\beta \end{bmatrix}^{-1} \right| \tag{50}$$

which is also the D-optimal, or determinant criterion discussed in the previous section. The bound for $\theta_\alpha$ with known $\theta_\beta$ is given by

$$CRB(\theta_\alpha | \theta_\beta) = \left| J_\alpha^{-1} \right| \tag{51}$$

which is also the bound for estimating $\theta_\alpha$ alone. The variance bound for $\theta_\alpha$ with unknown $\theta_\beta$ (i.e., where both parameter sets must be estimated) is given by

$$CRB(\theta_\alpha) = \left| \left( J_\alpha - J_{\alpha\beta} J_\beta^{-1} J_{\beta\alpha} \right)^{-1} \right| \tag{52}$$

These lead to the Cramér-Rao bound analog of the chain rule,

$$CRB(\theta_\alpha, \theta_\beta) = CRB(\theta_\alpha | \theta_\beta)\, CRB(\theta_\beta) = CRB(\theta_\beta | \theta_\alpha)\, CRB(\theta_\alpha) \tag{53}$$

and the Cramér-Rao bound analog of the Bayes rule,

$$CRB(\theta_\alpha) = \frac{CRB(\theta_\beta)}{CRB(\theta_\beta | \theta_\alpha)} CRB(\theta_\alpha | \theta_\beta) \tag{54}$$

The ratio in eqn. (54) is thus the factor by which the Cramér-Rao bound for $\theta_\alpha$ increases when $\theta_\beta$ must be simultaneously estimated from $X$. Alternatively, the inverse of this ratio is the factor by which the bound for $\theta_\alpha$ is decreased by obtaining knowledge of $\theta_\beta$ through external means. As the authors of [146] note, this measure could be used to provide a quantitative cost versus benefit measure for the prospect of fielding an additional, independent measurement technique for estimating nuisance parameters. This ratio can also serve as a general measure of the analytical selectivity a sensor array provides between defined analyte groups. For instance, if $\theta_\alpha$ is taken as a single target analyte, and $\theta_\beta$ is comprised of the remaining analytes as a potential chemical background, eqn. (51) corresponds to the variance bound one would arrive at with the naïve assumption that the chemical interferences were not present. Eqn. (52) corresponds to the bound one would calculate with the assumption that the chemical interferences are present. The ratio between the two thus expresses the impact that background levies on the analytical capability of the system in terms of lower-bound estimation variance of the target analyte concentration. It is straightforward to show that this ratio is indeed closely related with chemometric notions of selectivity and spectral similarity in multivariate measurement systems. Consider an array of m sensors with linear response functions characterized by a fixed sensitivity value over the sensor's dynamic range for each of n possible analytes and Gaussian noise. The sensor response is

$$X = S\theta + e \tag{55}$$

where $S$ is an m x n matrix of sensor response sensitivities for each sensor/analyte combination and is a n-dimensional multivariate Gaussian noise with covariance matrix $\sigma^2 I$. Dividing the parameter set into target analytes $\theta_\alpha$ and potential interferences $\theta_\beta$,

$$\theta = \begin{bmatrix} \theta_\alpha \\ \theta_\beta \end{bmatrix}, \quad S = [A\ B] \tag{56}$$

the sensor response is

$$X = A\theta_\alpha + B\theta_\beta + e \tag{57}$$

Using the analytic expression derived in R5 for linear sensors with Gaussian noise, the resulting FIM is

$$J(\theta) = \begin{bmatrix} J_\alpha & J_{\alpha\beta} \\ J_{\beta\alpha} & J_\beta \end{bmatrix} = \frac{1}{\sigma^2} \begin{bmatrix} A^T A & A^T B \\ B^T A & B^T B \end{bmatrix} \tag{58}$$

Applying the Bayes rule analog, we get

$$CRB(\theta_\alpha) = \frac{\left| \sigma^2 \left( B^T B - B^T A \left( A^T A \right)^{-1} A^T B \right)^{-1} \right|}{\left| \sigma^2 (B^T B)^{-1} \right|} CRB\left( \theta_\alpha | \theta_\beta \right) \tag{59}$$

which reduces to

$$CRB(\theta_\alpha) = \frac{\left| B^T B \right|}{\left| B^T \Pi_A^\perp B \right|} CRB\left( \theta_\alpha | \theta_\beta \right) \tag{60}$$

or equivalently, through substitution,

$$CRB(\theta_\alpha) = \frac{\left| A^T A \right|}{\left| A^T \Pi_B^\perp A \right|} CRB\left( \theta_\alpha | \theta_\beta \right) \tag{61}$$

The denominator of the ratio in eqn. (61) is directly related to the chemometric concept of net analyte signal (NAS) which is defined as the portion of the signal arising from a given mixture component that is orthogonal to the space spanned by the responses of the other component [147],

$$NAS_{\theta_\alpha} = \sqrt{A^T \Pi_B^\perp A} \tag{62}$$

and, as mentioned previously, forms the basis of many figures of merit in multivariate calibration [148]. Considering again the case with $\theta_\alpha$ as a single target analyte and $\theta_\beta$ the remaining library of potential interferences, the selectivity ratio in eqn. (61) is simply the ratio of the square of magnitude of the response vector to $\theta_\alpha$ to that of the corresponding NAS vector. The inverse of this ratio is directly interpretable as the reduction in signal to noise for the target analyte that is incurred by orthogonal signal correction against the background presented by $\theta_\beta$ and is identical to the multivariate selectivity metric proposed by Lorber in [147]. This ratio is also directly dependent on the angle, $\phi$, between the sensor array response to $\theta_\alpha$ and the subspace spanned by the responses to the elements of $\theta_\beta$, as shown in eqn. (63).

$$\frac{\left|A^T \Pi_B^{\perp} A\right|}{\left|A^T A\right|} = \frac{\|NAS_A\|^2}{\|A\|^2} = \sin^2\phi = 1 - \cos^2\phi \tag{63}$$

Accordingly, the ratio is also closely related to the cosine distance, a commonly used to measure spectral similarity [149],

$$d_{\cos} = 1 - \cos\phi \tag{64}$$

While the selectivity measure in eqn. (63) is applicable to systems with linear response functions only, this is not true for the corresponding ratio in the Cramér-Rao bound analog of Bayes rule postulated by [146], which only requires that the FIM be nonsingular, that $\theta$ be continuous, and that $p(X|\theta)$ be defined according to the measurement error distribution observed for $X$. In this way, the ratio suggested by the Cramér-Rao bound analog for Bayes' rule would appear to provide a reasonable generalized measure of sensor array selectivity, $sel_{\alpha,\beta}$ between analyte subsets $\theta_\alpha$ and $\theta_\beta$, as shown in eqn. (65).

$$sel_{\alpha,\beta} = \frac{CRB(\theta_\alpha)\,CRB\left(\theta_\beta\right)}{CRB(\theta)} \tag{65}$$

The selectivity measure above ranges from zero to one due to the condition

$$CRB(\theta) \geq CRB(\theta_\alpha)\,CRB\left(\theta_\beta\right) \tag{66}$$

(See Appendix for proof A.1.) The above exhibits equality only if the responses due to each analyte set are statistically independent, leading to a selectivity of one. The ratio approaches zero as analyte responses become indistinguishable from each other, and $CRB(\theta)$ grows larger relative to the numerator. Thus, the determinant of the inverse Fisher information matrix, $CRB(\theta)$, reflects both a generalized error volume bound in parameter space and an inverse measure of the global sensitivity of the sensor array to displacements in parameter space. The selectivity measure above is equivalent to or consistent with other selectivity measures and is also directly interpretable as the factor by which the error bound associated with one analyte subset is reduced by removing the requirement to estimate the remaining analytes using the same data. Alternatively, $1 - sel_{\alpha,\beta}$ can be interpreted as the fraction of the global error volume that is attributable to overlap in response between analyte subsets.

### 4.3 Resolving Power

Resolving power is a figure of merit that describes the ability of a sensor, sensor array or other analytical instrument to discriminate between two or more chemical analytes, stimuli or other relevant states. Common expressions of resolving power as a figure of merit in chemical sensing include measures of "limit of detection" and "limit of recognition" for chemical stimuli. As with sensitivity, the concept of "resolution" is used in varied ways in analytical chemistry. It often refers to specific capabilities of specific analytical instruments, which are thus are not agnostic with respect to measurement technique, nor do they necessarily correspond directly to discernment of chemical stimuli implied by sensing tasks. (For example, mass spectrometry, chromatography, and optical spectroscopy are all characterized by particular resolution-based instrumental figures of merit that do not translate to any other type of instrument [150]. For the purposes of this work, the focus will remain on resolving power as it pertains to sensing tasks and discernment of chemical stimuli.

### *4.3.1 Limit of Detection*

Qualitatively, the limit of detection may be thought of as being the lowest detectable concentration of a given analyte against a blank chemical background. IUPAC [142] defines it as

$$x_L = \bar{x}_{bi} + k s_{bi} \tag{67}$$

where $x_L$ is the limit of detection, $\bar{x}_{bi}$ is the average of the blank measurements, $s_{bi}$ is the standard deviation of those blank measurements, and $k$ is the numerical factor associated with the desired confidence level.

A multivariate limit of detection for nonlinear sensor arrays was recently described by Alsaedi et. al. [151]. In that paper, the authors present a Monte Carlo simulation algorithm which incorporates the statistical criterion recommended by IUPAC for a limit of detection measurement: that such a measure incorporate (1) the Neyman-Pearson criterion for hypothesis testing for false positives (Type I errors) and false negatives (Type II errors) and that all known noise sources are included. It is assumed that the noise sources in question are statistical in nature and can be averaged and are not in fact unwanted signal as in the case of chemical “noise.”

### *4.3.2 Limit of Recognition*

Despite regular appearances in the analytical chemistry literature over the past 20 years [75, 152], the IUPAC Gold Book [142], does not appear to contain an official definition of limit of recognition. Qualitatively, limit of recognition refers to the maximum recognizable composition range [75]; in other words, it is a FoM for how well an instrument may ascertain the quantitative composition of a chemical mixture. Unlike, other FoMs, it is inherently a multivariate measurement due to the need to identify and quantitate different chemicals in those mixtures.

Zeller and his co-workers explore the limit of recognition in a way that generalizes limit of detection to a multivariate vector-based context at the expense of narrowly focusing that definition on a surface acoustic wave sensor [152]. However, much of their work is a numerical nature and defines the limit of detection as the lowest concentration at which they can recognize the target chemical. The influence of chemical noise is not explored with mathematical rigor nor is the formal definition of the subject as a statistical quantity.

### *4.3.3 Resolving Power for Arbitrary Sensing Tasks*

In a general sense, resolving power for an arbitrary sensing task can be defined as the ability to distinguish one chemical stimulus from another on the basis of measured sensor data. This connects directly with the previous discussion of sensitivity. A given displacement in chemical stimulus space for which the mean array responses associated with each end of the displacement is not equal can be resolved to the extent allowed by the measurement error distributions of the sensor responses. This can be expressed, for example, as a “signal-to-noise” metric in which the displacement between the two mean response vectors is divided by the average measurement error (e.g. standard deviation multiplied by a scalar value reflecting the degree of certainty required) or as a receiver-operator characteristic (ROC)curve generated from the array response distributions and a particular classifier rule. However, as discussed previously, the Fisher information yields a more direct measure of resolving power by providing a lower-bound estimate of the error in determining a particular location in chemical stimulus. Sensing task-specific resolving power can thus be characterized by the lower-bound error of estimation oriented along the direction of that task in the appropriate region of the chemical stimulus space. This error implies a particular relationship between displacement along this direction and the degree of certainty with which this displacement can be detected that can be expressed in various statistical or decision theoretic terms, such as particular significance tests or ROC curve analysis.

## 5. SENSOR SELECTION VIA CONVEX OPTIMIZATION

### 5.1 Introduction to Sensor Selection

*In situ* chemical sensing for field applications presents significant challenges for the practitioner. The deployed technology must be selective and robust against the chemical background of the subject environment while remaining sensitive to the target chemical species. As almost no sensor technologies present truly target-specific [2] capabilities, the success of a particular sensing strategy in a given application is driven by the complexity and composition of the chemical background in which the sensing is to take place. This observation is at odds with the approach of much of the chemical sensing literature, which tends to prioritize the discussion of target analyte response characteristics under controlled laboratory conditions over consideration of the impact of realistic, complex chemical background conditions. Consequently, the effects of such backgrounds embody combinatorial parameter spaces that are impractical to thoroughly investigate in experimental laboratory testing.

Moreover, the design of chemical sensor arrays from the standpoint of chemical sensor selection and error quantification has historically proceeded as an *ad hoc* process. Frequently, chemical sensors are developed not as general purpose sensing devices, but as analyte or chemical class specific detectors. When such single purpose devices are integrated together as a chemical sensor array, it is unclear *a priori* how well they will function in concert with each other to provide expanded capabilities, an observation that is true of the integration of analytical instruments as well [1]. Further complicating the combination and optimization of these devices is that it is semantically unclear precisely what the combined device or array ought to do. Defining what a combined sensing device ought to do is difficult and highly dependent upon the analytical task the array will be intended to support as well as the specific goals of the array designer.

One commonly proposed way to deal with difficult selectivity requirements is to implement an array of purposely non-specific sensors, with the idea that such systems will provide an aggregate detection capability significantly greater than that of any individual sensor. This approach is rooted in analogy to biological exemplars of selectivity. For instance, vapor sensor arrays have been called “electronic noses,” referring to the array of non-specific olfactory receptor neurons that comprise the olfactory system. The resultant combined output of such a composite array to a stimulus is often referred to as a “fingerprint,” due to the presumed uniqueness of this stimulus pattern.

However, these analogies only go so far. For example, the human olfactory system contains approximately 350 different receptor types, suggesting it is significantly more complex than any electronic nose system that has been constructed to date, where the number of receptors is typically less by an order of magnitude. In terms of information-generating capacity, the olfactory system would, at least superficially, seem to have a greater resemblance to an analytical laboratory instrument than to current multisensor systems. This makes sense, as intuitively one would expect a proportionate relationship between the information-generating capacity of a system and the range of analytical tasks (or degree of uncertainty in a given task) that it can reasonably be expected to support. The central challenge in building a successful sensor array for a complex detection task is thus one of optimizing sufficient information-generating capacity to minimally enable an acceptable likelihood that the system will be able to answer the analytical challenge posed by the uncertainty of the parameters of the task at hand.

---

[2] The term specific is defined as “the ultimate of selectivity,” as recommended by IUPAC in [30]

That said, in the face of an otherwise unspecified sensing task, it is reasonable to assume that the practitioner ought to attempt to minimize the global error of the array, or conversely, to maximize the signal. This is the approach is taken by the authors within this paper. The question remains, however, as to how to best fulfill this objective. While a hypothetical practitioner may be able to take an exhaustive approach to sensor array design by experimentally evaluating all possible sensor combinations, this method quickly becomes infeasible as the number of sensors relative to array slots becomes coequal or large.

In the rare cases when a sensor array optimization has been attempted (as opposed to using whatever sensors were immediately available), it is this aforementioned approach of combinatorial experimentation which historically has typified chemical sensor array design, and thus, severely limited the optimization of sensor arrays. Alternative approaches to array design based on neural networks and machine learning have also been tried e.g. [1, 153–155]. However, due to their opacity, these methods fail to provide significant insight into the chemical detection problem or to suggest subsequent ways to further improve the array design. Consequently, an explicit, precise, and mathematically rigorous approach to chemical sensor array design and optimization is greatly desired.

Given its wide range of applications it is surprising that the literature centered on chemical sensor array optimization strategies is rather sparse, despite the relative frequency of papers describing specific sensor arrays and applications. A notable exception is the Fisher information matrix-based approach proposed by Pearce and Sánchez-Montañes and theoretically applied to simple linear sensor systems with uncorrelated noise [21, 22, 25]. Unfortunately, this methodology has not been greatly developed since its inaugural set of papers. In the view of this paper's authors, this is most likely due to the mathematical complexities and difficulties presented by implementing this program as well as the accompanying change in mentality this forces upon the typical practitioner in the chemical sensing field.

This chapter presents a methodology for designing near-optimal chemical sensor arrays with minimal statistical error by extending Pearce and Sánchez-Montañes's (PSM) Fisher information matrix-based approach [23, 24]. Originally, PSM demonstrated the FIM-based methodology for statistically uncorrelated chemical sensor arrays by selecting optimal sensors by hand. We extend this framework to a more realistic situation by considering statistical correlation amongst the sensor responses and use convex optimization techniques for sensor selection. Using convex optimization for sensor selection is not without precedent, Joshi and Boyd [156] performed sensor selection with convex optimization for a sensor array system with uncorrelated uniform Gaussian noise. Our improvements are made possible by taking advantage of the positive definite nature of the FIM as a rich source of objectives for sensor selection via fast convex optimization techniques without recourse to the combinatorial simulation, so that the possibility of "on-the-fly" chemical sensor array design is apparent.

We further develop the use of the Fisher information matrix as a quantitative descriptor for hypothetical chemical sensor array scenarios in which a collection of co-located sensors respond to chemical mixtures resulting from a pool of possible analytes. We assume that the underlying sensors provide additive linear responses with respect to the system of analytes and that they may exhibit statistically correlated noise. The latter is important as correlated measurement error is realistic, yet frequently unacknowledged in the literature. The former is generally a reasonable assumption in low concentration regimes, which typify the bulk of analytical sensing applications, and present the greatest challenges regarding desired sensitivity and selectivity.This work describes how the positive (semi)definite nature of the Fisher information matrix enables algorithmic chemical sensor array design via convex optimization techniques. This property is a rare bit of mathematical good fortune as the general case of global optimization is generally computationally intractable. The use of elliptically contoured distributions as a general-purpose means of modeling correlated sensor noise is introduced and developed for convex optimization of sensor arrays. Ultimately, this paper presents a theoretical description and practical implementation of this approach to chemical sensor array design and optimization by showing how to (nearly) best select a subset of sensors for a sensor array from a much larger collection while assuming correlated noise and the specific challenges of a chemical environment.

## 5.2 Approaches to Chemical Sensor Array Design

### *5.2.1 Introduction*

The design, development, and implementation of sensor array systems capable of performing robust detection tasks in chemically complicated environments present enormous challenges to practitioners. Among the many design decisions to be considered are: the anticipated operating environment and target set of chemicals or class of chemicals under consideration, the likely chemical interferants in the operating environment, the noise associated with nonchemical sources, the selection of sensor type and number and the likely correlations amongst the chosen sensors, and how to process the generated sensor array data. Moreover, the metrics used to evaluate the quality of a sensor often transfer poorly to a sensor array in a complex chemical environment due to the unknown effects of noise and the biases of chemical system specific phenomena. Consequently, developing task agnostic design criteria for sensor arrays is critically important for their robust design and development as such such a framework inherently provides the metrics for the optimization of such a sensor array.

### *5.2.2 Theoretical Design Criteria for Chemical Sensor Arrays*

A useful way to conceptualize sensing task and sensor array complexity is to describe them as geometric spaces, one describing the span of analyte mixtures to be sensed and the other describing the span of response vectors the array is capable of generating. The sensor response functions provide a transformation from one space to the other. This enables a general geometric description of signal processing based on a linear algebraic model of sensor arrays [21, 25]. For example, assuming an array of sensors with linear response curves and additive response to analyte mixtures, the response, $r$, of a particular sensor to a mixture of $m$ analytes will be:

$$r = \sum_{i=1}^{m} c_i s_i \tag{68}$$

where $c_i$ is the concentration of the *ith analyte* and $s_i$ is the sensitivity of the sensor to the *ith analyte*. Extending this relationship for an array of $n$ sensors leads to a system of linear equations:

$$\boldsymbol{r}^T = \boldsymbol{c}^T \boldsymbol{S}, \tag{69}$$

where the entries, $s_{ij}$, of the sensitivity matrix, $\boldsymbol{S}$, encode the sensitivity of the *jth* sensor to the *ith* analyte and each $r_j$ is the response from the *jth* sensor. The matrix of the sensitivity values, $\boldsymbol{S}$, for a sensor array specifies the geometric transform that maps locations in the sample space to corresponding locations in the measurement space of that array. The characteristics of this geometric relationship directly relate to the analytical performance achievable by the sensor array.

The Fisher information matrix methodology [21, 25, 157] expands the applicability of this geometric framework to array systems with probabilistically defined noise and nonlinear and non-additive sensor responses. It has been used in many applications involving statistical estimation, including quantum chemistry, experimental design, and neural programming [158–161]. For a single sensor, *i*, and concentration vector , $\boldsymbol{c}$, the Fisher information matrix f is defined as,

$$J^i_{jj'}(\boldsymbol{c}) = \int dr_i p(r_i|\boldsymbol{c}) \left(\frac{\partial}{\partial c_j} \ln p(r_i|\boldsymbol{c})\right) \left(\frac{\partial}{\partial c_{j'}} \ln p(r_i|\boldsymbol{c})\right), \tag{70}$$

where $p(r_i|\boldsymbol{c})$ is the conditional probability of a response of the *ith* sensor, given $\boldsymbol{c}$.

Assuming uncorrelated noise, the Fisher information matrix of the array is the sum of the matrices for each sensor. According to the Cramér-Rao bound, the total expected squared reconstruction error for an unbiased estimator across the entire array is related to the Fisher information matrix by eqn. (71).

$$\sigma^2(\hat{\boldsymbol{c}}|\boldsymbol{c}) = \sum_{j=1}^{m} \sigma^2(\hat{\boldsymbol{c}}_j|\boldsymbol{c}) \geq \sum_{j=1}^{m} \sigma^2(J^{-1}(c))_{jj} \equiv \langle \varepsilon^2_{c_j} \rangle \tag{71}$$

Thus, one can derive a measure of how the geometric framework of the detection problem itself imposes a limitation on the certainty to which a target analyte concentration can be estimated.

Using these metrics of optimization, the Pearce *et. al.* [25] were able to illustrate the importance of careful design in sensor array development for a well-defined sensing task. They demonstrated for a simple example that choosing a random sensor configuration would result in significantly sub-optimal performance with relatively high probability. To optimally select sensors for their array, they resorted to brute force search for an optimal solution. Fortunately, other work due to Joshi and Boyd (J-B) [156] has addressed the related issue of sensor selection under Gaussian noise using convex optimization. While J-B did not consider the selection of sensors using the Fisher information matrix, the method they used is applicable to optimizing the Fisher information matrix for a sensor array with multivariate Gaussian noise. This paper generalizes and expands this approach to optimizing a chemical sensor array with noise distributed according to an arbitrary multivariate probability distribution and specifically applies this approach to a sensor array with correlated noise with elliptically contoured distribution.

### 5.3 Convex Optimization of the Fisher Information matrix

Recall the definition of the FIM,

$$\boldsymbol{F}(\boldsymbol{\mu};\boldsymbol{\theta}) = \int d\boldsymbol{X} \rho(\boldsymbol{\mu};\boldsymbol{\theta}|\boldsymbol{X}) \left(\frac{\partial \ln(\rho)}{\partial \boldsymbol{\theta}}\right) \left(\frac{\partial \ln(\rho)}{\partial \boldsymbol{\theta}}\right)^T \tag{72}$$

Due to its structure as matrix defined by an integral over an exterior vector product, the FIM is a positive semidefinite matrix, i .e. $\boldsymbol{a}^T \boldsymbol{F} \boldsymbol{a} \geq 0$, where $\boldsymbol{a}$ is an arbitrary real-valued vector of appropriate dimension. Positive semidefiniteness is crucial as this property allows for the so-described sensor array to be optimized with respect to sensor configuration via convex optimization techniques.

In order to properly implement this idea for sensor array optimization, specifications for of an appropriate objective function as well as a set of constraints are required. To setup this problem, first the objective function will be defined and then the relevant constraints detailed. In the process of setting up the constraints and detailing the supporting mathematical elements for the convex optimization, appropriate sensor responses and noise models for the chemical sensor array will be proposed.

Barring other priorities or specific knowledge of the analytical task, a reasonable design goal for a general purpose chemical sensor array is to minimize the global error (maximize the signal) of the chemical sensor array. A useful measure for this global error is the volume of the ellipsoid cast by the covariance matrix of the relevant estimators since this provides a reasonable metric for the global uncertainty of estimated chemical concentrations and thus for the discernability of similar chemical mixtures. The volume of this error ellipsoid is given by the following expression [131],

$$\boldsymbol{vol}(\boldsymbol{\Sigma}) = \frac{2\pi^{d/2}}{d\Gamma\left(\frac{d}{2}\right)}\det(\boldsymbol{\Sigma})^{1/2} \tag{73}$$

where $\boldsymbol{\Sigma}$ is the covariance matrix and $d$ is the dimension of the volume. Minimizing this volume term, $\boldsymbol{vol}(\boldsymbol{\Sigma})$, ultimately requires the minimization of $\det(\boldsymbol{\Sigma})^{1/2}$ as all other terms for the volume expression are related to the system dimension, which is not subject to optimization.

Since the composition of convex functions are themselves convex and both the square root function of $x > 0$ and the determinant of a positive semidefinite matrix like $\boldsymbol{\Sigma}$ are convex functions themselves, the objective function may be further simplified to$\det(\boldsymbol{\Sigma})$. For reasons of subsequent numerical convenience, this objective function is composed with the natural logarithm to give $\ln(\det(\boldsymbol{X}))$ as the final objective function. It is proven below that this function is concave (convex up) for all positive semidefinite matrices, $\boldsymbol{X}$, by showing that this function satisfies concavity [162].

First consider the following,

$$\begin{aligned} g(t) &= \ln(\det(\boldsymbol{X})) \\ &= \ln(\det(\boldsymbol{Z}+t\boldsymbol{V})) \end{aligned} \tag{74}$$

where $\boldsymbol{X} = \boldsymbol{Z}+t\boldsymbol{V} > 0$. $\boldsymbol{X}$, $\boldsymbol{Z}$, and $\boldsymbol{V}$ are positive definite matrices and $t \geq 0$ is a scalar parameter. Manipulating the matrix function in question to ensure positive definite matrices yields,

$$\begin{aligned} g(t) &= \ln(\det(\boldsymbol{Z}+t\boldsymbol{V})) \\ &= \ln(\det(\boldsymbol{Z}^{1/2}(\boldsymbol{I}+t\boldsymbol{Z}^{-1/2}\boldsymbol{V}\boldsymbol{Z}^{-1/2})\boldsymbol{Z}^{1/2})) \\ &= \ln(\det(\boldsymbol{I}+t\boldsymbol{V}))+\ln(\det(\boldsymbol{Z})) \end{aligned} \tag{75}$$

so that the first and second derivatives of $g(t)$ may be taken as,

$$g\prime(t) = \sum_{i=1}^{n} \frac{\lambda_i}{1+t\lambda_i} \tag{76}$$

and

$$g''(t) = -\sum_{i=1}^{n} \frac{\lambda_i^2}{(1+t\lambda_i)^2} \tag{77}$$

Since $\lambda_i > 0$ due to the definition of positive definite matrices, it follows that $g'(t) > 0$ and $g''(t) < 0$ for $t \geq 0$. This implies that $\ln(\det(\boldsymbol{X}))$ is a convex function for positive definite $\boldsymbol{X}$ [162].

Having shown that this objective function is valid for the optimization problem, it is now important to consider what variables to actually use to optimize the $\ln(\det(\boldsymbol{X}))$ objective. Specifically, it is not the covariance matrix that is being input into the objective function, but the inverse Fisher information matrix. This substitution is justified due to the Cramér-Rao lower bound (just as the Gaussian FIM substitution would be in the case of the upper bound).

Consequently, it is the FIM of a probability distribution and not its covariance matrix which is parametrized for optimization and the objective function thus becomes,

$$\ln(\det(\boldsymbol{C}(\boldsymbol{\theta}))) \geq \ln(\det(\boldsymbol{F}^{-1}(\boldsymbol{\theta};\boldsymbol{s}))) = -\ln(\det(\boldsymbol{F}(\boldsymbol{\theta};\boldsymbol{s}))) \tag{78}$$

where $\boldsymbol{s}$ are the slack variables subject to the optimization. For a given convex optimization in addition to the inequality and equality constraints, the practitioner must supply the convex optimization routine with gradient and Hessian routines for the objective function in the slack variables as well [162].

The gradient for $-\ln(\det(\boldsymbol{F}(\boldsymbol{\theta};\boldsymbol{s})))$, the objective function, is given by,

$$-\nabla \ln(\det(\boldsymbol{F}(\boldsymbol{\theta};\boldsymbol{s}))) = -\sum_{i=1}^{\#(\boldsymbol{s})} \hat{\boldsymbol{e}}_i \mathrm{Tr}(\boldsymbol{F}^{-1}\frac{\partial \boldsymbol{F}}{\partial s_i}) \tag{79}$$

where $\#(\boldsymbol{s})$ is the cardinality of the slack variables, $\boldsymbol{s}$, and $\hat{\boldsymbol{e}}_i$ denote the relevant vector basis set. The matrix elements for the Hessian, $\boldsymbol{h}(\boldsymbol{\theta};\boldsymbol{s})$, are defined by,

$$\begin{aligned} h_{ij}(\boldsymbol{\theta};\boldsymbol{s}) &= -\frac{\partial^2}{\partial s_i \partial s_j} \ln(\det(\boldsymbol{F}(\boldsymbol{\theta};\boldsymbol{s}))) \\ &= \mathrm{Tr}(\boldsymbol{F}^{-1}\frac{\partial \boldsymbol{F}}{\partial s_i}\boldsymbol{F}^{-1}\frac{\partial \boldsymbol{F}}{\partial s_j}) - \mathrm{Tr}(\boldsymbol{F}^{-1}\frac{\partial^2 \boldsymbol{F}}{\partial s_i \partial s_j}) \end{aligned} \tag{80}$$

To evaluate these quantities it is necessary to first choose a noise model, so that the FIM may be properly parametrized for optimization. This matter is discussed in the following section.

## 5.4 Elliptically Contoured Distributions: A Correlated Noise Model for Chemical Sensor Arrays

### *5.4.1 Introduction to Elliptically Contoured Distributions*

Elliptically contoured distributions (ECDs) [163] are a class of statistical model which generalize the multivariate Gaussian and includes many standard statistical models like the multivariate Student's t-distribution.

They are defined as follows

$$\frac{g((\boldsymbol{x}-\boldsymbol{\mu}(\boldsymbol{\theta}))^T\boldsymbol{\Sigma}(\boldsymbol{\theta})(\boldsymbol{x}-\boldsymbol{\mu}(\boldsymbol{\theta})))}{N(\boldsymbol{\theta})} \tag{81}$$

where $g(\cdot)$ is an arbitrary univariate probability distribution, $\boldsymbol{\theta}$ are the external deterministic parameters being bounded by the FIM, $\boldsymbol{\mu}(\boldsymbol{\theta})$ is the mean response function, is a positive definite scale matrix which reduces to the covariance matrix if $g(\cdot)=\exp(-(\cdot))$, and $N(\boldsymbol{\theta})$ is the normalization constant for the probability density function. These distributions are chosen to model the correlated noise of chemical sensor arrays as they can model correlation among sensor responses while remaining both analytically tractable and relatively general.

Examples of FIMs for various well-known probability distributions are given by following: The FIM for the multivariate Gaussian is given by the so-called Slepian-Bangs formula as [164],

$$F_{ij}(\boldsymbol{\theta}) = 2\left(\frac{\partial\boldsymbol{\mu}^T}{\partial\theta_i}\right)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\left(\frac{\partial\boldsymbol{\mu}}{\partial\theta_j}\right) + \mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j) \tag{82}$$

where $\boldsymbol{\Sigma}_i = \frac{\partial\boldsymbol{\Sigma}}{\partial\theta_i}$ and the FIM for the multivariate Student-t distribution [164] is

$$\begin{aligned}F_{ij}(\boldsymbol{\theta}) =& 2\frac{d+M}{d+M+1}\left(\frac{\partial\boldsymbol{\mu}^T}{\partial\theta_i}\right)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\left(\frac{\partial\boldsymbol{\mu}}{\partial\theta_j}\right)\\ &-\frac{1}{d+M+1}\mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i)\mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j)\\ &+\frac{d+M}{d+M+1}\mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j)\end{aligned} \tag{83}$$

where $d$ is the degrees of freedom, a distribution specific quantity, of the Student-t distribution and $M$ is the rank of the scale matrix $\boldsymbol{\Sigma}$.

The FIM for ECDs [164] has been recently derived as a generalization of the Slepian-Bangs formula as

$$\begin{aligned}F_{ij}(\boldsymbol{\theta}) =& 2\frac{\boldsymbol{E}_p[q\phi^2(q)]}{M}\left(\frac{\partial\boldsymbol{\mu}^T}{\partial\theta_i}\right)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\left(\frac{\partial\boldsymbol{\mu}}{\partial\theta_j}\right)\\ &+\left[\frac{\boldsymbol{E}_p[q^2\phi^2(q)]}{M(M+1)}-1\right]\mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i)\mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j)\\ &+\frac{\boldsymbol{E}_p[q^2\phi^2(q)]}{M(M+1)}\mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j)\end{aligned} \tag{84}$$

where $M$ is the scalar dimensionality or rank of the scale matrix, $\boldsymbol{\Sigma} \in \mathbb{R}^{M\times M}$, $\boldsymbol{E}_p[\cdot]$ denotes an expectation value with regard to a probability density,

$$p(q) = \frac{1}{\delta_{M,g}} q^{M-1} g(q) \tag{85}$$

so that,

$$\boldsymbol{E}_p[\cdot] = \frac{1}{\delta_{M,g}} \int_0^\infty dq\,(\cdot)\,q^{M-1} g(q) \tag{86}$$

where

$$\delta_{M,g} = \int_0^\infty dt\; t^{M-1} g(t) \tag{87}$$

and

$$\phi(t) = \frac{g\prime(t)}{g(t)} \tag{88}$$

Using ECDs and their corresponding FIMs as reasonable models for correlated chemical sensor arrays allows the practitioner to propose a specific model for convex optimization.

### 5.4.2 *Gradients and Hessians for the Fisher Information Matrices of Elliptically Contoured Distributions*

Recall from the prior section that the model dependent components of the gradient and Hessian matrix for the convex optimization of the FIM are the matrices $\frac{\partial \boldsymbol{F}}{\partial s_p}$and $\frac{\partial^2 \boldsymbol{F}}{\partial s_p \partial s_q}$. The expressions for the matrix elements of these matrices are developed in the following subsection.

First, express the ECD FIM elements as follows,

$$\boldsymbol{F}_{ECD}(i,j) = \alpha \underbrace{\frac{\partial \boldsymbol{\mu}^T}{\partial \theta_i} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \theta_j}}_{G} + \beta \underbrace{\mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i)\mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j)}_{J}$$

$$+ \gamma \underbrace{\mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j)}_{K} \tag{89}$$

where $\alpha$, $\beta$, and $\gamma$ are distribution specific constants that are not dependent upon the slack variables, $s_p$ and $s_q$, and are given by,

$$\alpha = \frac{2\boldsymbol{E}_p[q\phi^2(q)]}{M} \tag{90}$$

$$\beta = \frac{\boldsymbol{E}_p[q^2\phi^2(q)]}{M(M+1)} - 1 \tag{91}$$

$$\gamma = \frac{\boldsymbol{E}_p[q^2\phi^2(q)]}{M(M+1)} \tag{92}$$

where the subscript $p$ denotes an average with respect to $p(q)$ and $\boldsymbol{\Sigma}_i = \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i}$, and $G$, $J$, and $K$ are so defined to simplify the derivation and presentation.

The derivatives of each of these sub-expressions are given as follows,

$$\begin{aligned}\frac{\partial G}{\partial s_p} &= \frac{\partial^2 \boldsymbol{\mu}^T}{\partial s_p \partial \theta_i}\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\mu}}{\partial \theta_j} + \frac{\partial \boldsymbol{\mu}^T}{\partial \theta_i}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_p}\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\mu}}{\partial \theta_j} \\ &+ \frac{\partial \boldsymbol{\mu}^T}{\partial \theta_i}\boldsymbol{\Sigma}^{-1}\frac{\partial^2 \boldsymbol{\mu}}{\partial s_p \partial \theta_j}\end{aligned} \tag{93}$$

$$\begin{aligned}\frac{\partial J}{\partial s_p} &= \mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_p}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i + \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_p i})\mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j) \\ &+ \mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i)\mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_p}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j + \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_p j})\end{aligned} \tag{94}$$

$$\begin{aligned}\frac{\partial K}{\partial s_p} &= \mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_p}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j + \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_p i}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j) \\ &+ \mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_p}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j + \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_p j})\end{aligned} \tag{95}$$

where $\boldsymbol{\Sigma}_{s_p} = \frac{\partial \boldsymbol{\Sigma}}{\partial s_p}$ and $\boldsymbol{\Sigma}_{s_p i} = \frac{\partial^2 \boldsymbol{\Sigma}}{\partial s_p \partial \theta_i}$.

The Hessian elements for the $G$ and $J$ terms are given by,

$$
\begin{aligned}
\frac{\partial^2 G}{\partial s_p \partial s_q} = {} & \frac{\partial^3 \boldsymbol{\mu}^T}{\partial s_p \partial s_q \partial \theta_i} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \theta_j} \\
& + \frac{\partial^2 \boldsymbol{\mu}^T}{\partial s_p \partial \theta_i} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{s_q} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \theta_j} \\
& + \frac{\partial^2 \boldsymbol{\mu}^T}{\partial s_p \partial \theta_i} \boldsymbol{\Sigma}^{-1} \frac{\partial^2 \boldsymbol{\mu}}{\partial s_q \partial \theta_j} + \frac{\partial^2 \boldsymbol{\mu}^T}{\partial s_q \partial \theta_i} \boldsymbol{\Sigma}^{-1} \frac{\partial^2 \boldsymbol{\mu}}{\partial s_p \partial \theta_j} \\
& + \frac{\partial \boldsymbol{\mu}^T}{\partial \theta_i} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{s_q} \boldsymbol{\Sigma}^{-1} \frac{\partial^2 \boldsymbol{\mu}}{\partial s_p \partial \theta_j} + \frac{\partial \boldsymbol{\mu}^T}{\partial \theta_i} \boldsymbol{\Sigma}^{-1} \frac{\partial^3 \boldsymbol{\mu}}{\partial s_q \partial s_p \partial \theta_j} \\
& + \frac{\partial^2 \boldsymbol{\mu}^T}{\partial s_q \partial \theta_i} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{s_p} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \theta_j} \\
& + \frac{\partial \boldsymbol{\mu}^T}{\partial \theta_i} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{s_q} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{s_p} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \theta_j} + \frac{\partial \boldsymbol{\mu}^T}{\partial \theta_i} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{s_q s_p} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \theta_j} \\
& + \frac{\partial \boldsymbol{\mu}^T}{\partial \theta_i} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{s_p} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{s_q} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \theta_j} \\
& + \frac{\partial \boldsymbol{\mu}^T}{\partial \theta_i} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{s_p} \boldsymbol{\Sigma}^{-1} \frac{\partial^2 \boldsymbol{\mu}}{\partial s_q \partial \theta_j}
\end{aligned}
\tag{96}
$$

$$
\begin{aligned}
\frac{\partial^2 J}{\partial s_p \partial s_q} &= \mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j)\mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_q}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_p}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i + \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_q s_p}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i) \\
&+ \mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j)\mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_p}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_q}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i) \\
&+ \mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j)\mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_p}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_q i} + \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_q}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_p i}) \\
&+ \mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j)\mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_q s_p i}) \\
&+ \mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_p}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i + \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_p i})\mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_q}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j + \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_q j}) \\
&+ \mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_q}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i + \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_q i})\mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_p}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j + \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_q j}) \\
&+ \mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i)\mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_q}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_p}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j + \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_q s_p}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j) \\
&+ \mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i)\mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_p}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_q}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j) \\
&+ \mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i)\mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_p}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_q j} + \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_q}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_p j} + \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_q s_p j})
\end{aligned}
\tag{97}
$$

and the Hessian element for $K$ is setup as follows,

$$
\begin{aligned}
\frac{\partial^2 K}{\partial s_p \partial s_q} &= \frac{\partial}{\partial s_q}\underbrace{\mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_p}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j)}_{A_1} + \frac{\partial}{\partial s_q}\underbrace{\mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_p i}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j)}_{A_2} \\
&+ \frac{\partial}{\partial s_q}\underbrace{\mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_p}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j)}_{A_3} + \frac{\partial}{\partial s_q}\underbrace{\mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{s_p j})}_{A_4}
\end{aligned}
\tag{98}
$$

where the derivatives of each of the sub-terms of the expression are given by the following:

$$\begin{aligned}\frac{\partial A_1}{\partial s_q} &= \mathrm{Tr}(\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{s_q}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{s_p}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{i}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{j}+\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{s_p s_q}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{i}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{j}) \\ &+ \mathrm{Tr}(\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{s_p}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{s_q}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{i}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{j}) \\ &+ \mathrm{Tr}(\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{s_p}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{s_q i}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{j}) \\ &+ \mathrm{Tr}(\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{s_p}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{i}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{s_p}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{j}+\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{s_p}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{i}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{s_q j}) \qquad (99)\end{aligned}$$

$$\begin{aligned}\frac{\partial A_2}{\partial s_q} &= \mathrm{Tr}(\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{s_q}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{s_p i}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{j}+\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{s_p s_q i}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{j}) \\ &+ \mathrm{Tr}(\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{s_p i}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{s_q}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{j}+\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{s_p i}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{s_q j}) \qquad (100)\end{aligned}$$

$$\begin{aligned}\frac{\partial A_3}{\partial s_q} &= \mathrm{Tr}(\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{s_q}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{i}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{s_p}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{j}) \\ &+ \mathrm{Tr}(\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{s_p i}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{s_q}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{j}+\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{i}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{s_q}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{s_p}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{j}) \\ &+ \mathrm{Tr}(\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{i}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{s_q s_p}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{j}) \\ &+ \mathrm{Tr}(\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{i}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{s_p}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{s_q}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{j}+\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{i}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{s_p}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{s_q j}) \qquad (101)\end{aligned}$$

$$\begin{aligned}\frac{\partial A_4}{\partial s_q} &= \mathrm{Tr}(\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{s_q}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{i}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{s_p j}+\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{s_q i}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{s_p j} \\ &+ \mathrm{Tr}(\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{i}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{s_q}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{s_p j}+\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{i}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{s_p s_q j}) \qquad (102)\end{aligned}$$

where $\mathbf{\Sigma}_{ab} = \frac{\partial^2 \mathbf{\Sigma}}{\partial a \partial b}$ and $\mathbf{\Sigma}_{abc} = \frac{\partial^3 \mathbf{\Sigma}}{\partial a \partial b \partial c}$.

## 6. CHARACTERIZING CHEMICAL SIMULANTS

### 6.1 Introduction

Chemical simulants (CSs) are chemical additives and substitutes employed as replacements for targeted analytes (TAs) in situations like a training sessions or an experimental trial when those analytes are too dangerous, rare, illegal, and/or expensive to be actually used. In a qualitative, but technical sense, CSs are chemical stand-ins for TA(s) in the sense that they cause a given chemical sensing system's response to be identical or nearly identical to the response caused by the targeted analytes.

While they are employed in a wide variety of regulatory and security related settings, CSs are, perhaps, most familiar from their usage as canine training aids for drug, explosive, and cadaver recovery detection by security and law enforcement agencies worldwide [165–167]. CSs have also been developed to simulate chemical warfare agents when testing protective equipment [168] and validating the capabilities of environmental remediation techniques for end-of-life hazardous chemical cleanup and disposal [169]. Moreover, CSs are not just relevant to military and security-related settings, but to consumer safety protection as well. The counterfeit ingredients and adulterants often found in fraudulent food [170, 171] and pharmaceutical products [172] are, in essence, CSs intended to fool regulatory tests. Consequently, quantitatively understanding CSs is an important endeavor as the practical impact of their deployment and use is felt broadly throughout society.

Considering these critical applications, it is surprising that while there are end use-specific figures of merit (FOMs), such as those for a lunar regolith simulant [173] or detecting chemical warfare agent (CWA) simulants in consumer grade containers [174], no general, domain transferable, and system agnostic analytical framework for understanding and interpreting the relationships among of CSs, TAs, and analytical instruments is known by the authors to exist. Contributing to this paucity of work is the general lack of explicit recognition that a CS, and thus its FOM, is indivisible from the chemical detection system and the TA(s) to which they are associated. Further complicating the more practical exploration of an already fraught subject is that CSs are often deployed outside of controlled laboratory environments in real world scenarios subject to unknown and uncharacterized chemical interferents. Consequently, any CS FOM should be able to include the effects of the external environment via unknown chemical interferents, a caveat which can significantly increase the technical requirements of any viable, proposed FOM.

Other fields have developed FOMs for defining the quality of simulant-like quantities. In particular, steganography [175], the study of hidden writing, has a now field-standard FOM for theoretically determining the quality of the undetectability of a hidden message embedded in a covertext i.e. an image or similar lossy file. This FOM utilizes the Kullback-Leibler divergence (KLD) [11, 176] to compare the divergence or relative entropy between the probability distributions associated with image containing the hidden message and the reference image. Using this approach, a steganalyst is able to develop a lower bound for the probability of the hidden message's success of not being detected.

An analogous approach for a CS FOM is developed in this paper using an information theoretic inequality in this paper. This result is expanded upon as bounds are given both for the probability of detection of the chemical simulant as well as in terms of a ROC curve which bounds the quality of the CS FOM under a variety of detection scenarios. We further generalize this approach by including the effects of the external chemical environment in the KLD FOM. This addition allows us to describe bounds on various ROC curves in a variety of environmental backgrounds and to tune our CS to specific analytical task requirements. Finally, simulations of CSs and their accompanying ROC curves are provided for a variety of scenarios using both synthetic and experimental data.

## 6.2 Prior Work on Chemical Simulants

As has already been mentioned, CSs have a relatively sparse literature and little readily available quantitative data. Nonetheless, CSs have a long, practical history due to their undeniable utility. Much of that which is available is found in patents, corporate white papers, and government reports and *not* in the peer-reviewed literature. To reiterate what was previously mentioned: CSs are often deployed for training purposes, such as training canines to detect narcotics or cadavers, or design, such as designing and testing defenses against CWAs as well as surrogates for toxic chemicals in chemical remediation scenarios. In a rather less benign context, CSs make an appearance as cheap substitutes for more expensive ingredients in fraudulent consumer products or as adulterants specifically intended to defeat regulatory tests.

When designing a CS, some approaches in the literature attempt to replicate key chemical structures or functional groups of the simulated analyte and its possible breakdown products [177] or target the response(s) of the chemical detection system by replicating the key physical aspects of the such as the headspace or chemical vapor profile [ 178]. Materials intended to train military and police working dogs for scent detection or canine training aids are a large CS class, typically meant to replicate the scents of explosives, narcotics and human cadavers. Often, their specific composition is proprietary and quantitative information, not typically part of the sales literature, is often unavailable.

Nonetheless, due to their importance for investigative forensics, several peer-reviewed studies exist that detail the chemical composition and quality of these aids [179, 180]. In the instance of human remains-related training aids [179], concerns were raised by researchers as to the relevance of many of the simulant's constituent compounds to canine remains detection. Simulants for explosives are similarly analyzed by Harper et. al. [180]. Studies and articles associated with CWA CSs [169, 181–184] provide a useful perspective concerning how CSs themselves are used and reported in the literature; often they are employed for the research itself, but the actual subject is the TA with relatively little effort spent describing how the CS differs from the TA.

Perhaps unsurprisingly given these applications and oversights, if one is specifically interested in quantitative measurements of the quality of a CS, only a single case of a materials specific physico-mechanical FOM for lunar regolith colloquially known as "moon soil" is known by the authors to exist. Various lunar regolith simulants (LRSs) have been developed in academic labs and government agencies throughout the world [185–190]. These particular simulants attempt to exactly replicate the physical and material composition of lunar regolith, the moon's functional equivalent to terrestrial soil [190], for the purpose of designing exploratory equipment and developing lunar relevant studies and experiments. Interestingly, a FOM has been developed for quantifying the quality of LRSs [173]. This FOM characterizes LRS by their composition, granular size and shape assortment, and density. This LRS FOM is narrowly focused on criteria relevant to its application and presents itself as a scalar quantity from 0 to 1 with 0 being no correlation to lunar regolith and 1 being a perfect match. While this FOM is not chemical detection-oriented, it is the only quantitative FOM for a CS or CS-adjacent quantity found by the authors in their literature search.

## 6.3 Developing Figures of Merit: From Steganography to Chemical Simulants

### *6.3.1 Introduction and Overview*

We begin the development of our approach to CS quantitation via analogy with the development of a FOM for steganography, the study of information hiding to inform our ultimate goal of creating a system-agnostic FOM for CSs. FOMs for steganography provide a direct analogue to CSs due to the functional similarity of a steganographic payload and a CS. Steganography has a well-developed information theoretic FOM based on the KLD for assessing the quality of the steganographic scheme in question. After presenting that steganographic FOM scheme, we consider the analytical task of a CS in the presence of background/environmental chemical interferents and show how to analogously develop such an FOM to the measure the quality of a CS. Using an information processing inequality [11], we relate the KLDs for a CS and its corresponding TA to the decision concerning the presence or absence of the chemical simulant using a testing framework. We develop receiver operating characteristic (ROC) curve boundaries which help define CS quality and constrain the maximal effectiveness of that CS with a particular KLD pair.

### *6.3.2 A Figure of Merit from Steganography for the Quality of Information Hiding*

Steganography is a cryptographic adjacent field of study concerning itself with hiding data in an otherwise innocuous and ultimately fungible file called a *covertext*, often a lossy image. The archetypical example for the application of steganography is Simmons' "Prisoners' Problem" [191], which involves 2 prisoners attempting to secretly communicate under the watchful eye of an antagonistic warden. C. Cachin has provided an information-theoretic interpretation and FOM for this problem and for steganography itself by a mathematically formalized interpretation of what this problem requires to declare a participant "successful." We briefly sketch and define the mathematical details of Cachin's approach and refer the interested reader to [175, 192].

A stegosystem is a communication protocol between two *users*, Alice and Bob, in which Alice wants to send a seemingly innocuous message with hidden information through an open channel to Bob. This channel is subject to the scrutiny of an *adversarial* third party, Eve. Alice and Bob are deemed successful if Eve, who has access to all content transmitted through that open channel, is unable to detect that hidden messages are being transmitted. We call Alice *inactive* when she sends a message without hidden content. We refer to such a message as a *covertext*, $C$, which may be modeled as being generated by a probability distribution $P_C$. We call her *active* if she sends a message with an *embedded message*, $E$, from a message space, $\mathscr{E}$, encoded into the covertext with an *embedding function*, $\mathscr{F}$. The resultant covertext with an embedded message is referred to in totality as a *stegotext*, and modeled by the distribution $P_S$. Moreover, an independent *private random source* $R$ is available solely to Alice and the independent, random *secret key*, $K$, is available to both Alice and Bob, but not Eve. Likewise, Bob has an *extracting algorithm*, $\mathscr{G}$, which he uses to extract a decision value $\hat{E}$ from $S$ using $K$ with the hope that this procedure imparts information about $E$. We assume that Bob has access to some oracle that tells him whether or not Alice is active. This is not an implausible assumption since if Bob in actuality were to try to extract a hidden message from a covertext without an embedded message he would extract meaningless garbage with very high probability.

Cachin defined the security of this stegosystem by the KLD or relative entropy between the distribution for the covertext and the stegotext as,

$$\begin{aligned} D(P_C||P_S) &= \sum_i P_C(i) \log \frac{P_C(i)}{P_S(i)} \quad \text{(discrete)} \\ &= \int dX\, P_C(X) \log \frac{P_C(X)}{P_S(X)} \quad \text{(continuous)} \qquad (103) \end{aligned}$$

A stegosystem is deemed to be *perfectly secure* if $\lim_{n\to\infty} \frac{1}{n} D(P_C||P_S)$ tends to 0 over $n$ text transmissions, which means Eve has no way of detecting the difference between an empty covertext and a stegotext and

$\varepsilon$*-secure* if $\lim_{n\to\infty}\frac{1}{n}D(P_C||P_S) \leq \varepsilon$ over $n$ text transmissions. Since Eve is successful if she can detect the presence of a hidden message in a stegotext, $\varepsilon$-secure suggests the existence of a probabilistic, KLD-related, upper bound to Eve's successful detection of a stegotext. Ultimately this measure of Eve's success may be clarified and formalized using methods drawn from hypothesis testing.

Hypothesis testing is a binary decision rule for discriminating between a hypothesis $H_0$ and another hypothesis $H_1$. For this steganographic problem, we let $H_0$ signify the hypothesis that just a covertext is transmitted between Alice and Bob, and we let $H_1$ signify the hypothesis that Alice has transmitted a stegotext to Bob. The probability of a *type I error*, accepting $H_1$ when $H_0$ is true, is denoted by $\alpha$ and the probability of a *type II error*, accepting $H_0$ when $H_1$ is true, is denoted by $\beta$ [175].

As a consequence of the data-processing inequality [11], no amount of post collection information processing, which does not add additional assumptions to the data, can increase the KLD or relative entropy between $P_C$, the probability associated with measuring a covertext, and $P_S$, the probability associated with measuring a stegotext. Consequently, this inequality implies $D(P_{H_0}||P_{H_1}) \leq D(P_{Q_0}||P_{Q_1})$ where $P_H$ the probability of the hypotheses are calculated from the measurement probabilities $P_Q$. Formally we define the function $f : \mathscr{Q} \to \mathscr{H}$, which maps the space of measurements, $\mathscr{Q}$, to the space of binary hypotheses, $\mathscr{H}$, such that $H_0 = f(Q_0)$ and $H_1 = f(Q_1)$.

Since the decision process for determining between $H_0$ and $H_1$ is a binary decision subject to type I and type II errors, we may set $D(P_{H_0}||P_{H_1}) = D(P_\alpha||P_\beta)$, which is explicitly expressed as $D(P_\alpha||P_\beta) = \alpha\log_2\frac{\alpha}{1-\beta} + (1-\alpha)\log_2\frac{1-\alpha}{\beta}$, where this expression follows the binary log convention for the KLD used by Cachin [175]. Inserting this result into the information processing inequality yields $\alpha\log_2\frac{\alpha}{1-\beta} + (1-\alpha)\log_2\frac{1-\alpha}{\beta} \leq D(P_{Q_0}||P_{Q_1})$ which may used to directly relate the probabilities of the type I and type II errors and the KLD of the measurement probabilities. For instance, if we assume that Eve never makes a type I error ($\alpha = 0$) i.e. never falsely believes an unloaded covertext to be a stegotext, Eve's probability for making a type II error may be bounded from below as $\beta \geq 2^{-D(P_{Q_0}||P_{Q_1})}$ [175].

### 6.3.3 *Figures of Merit for Chemical Simulants in the Presence of Background Chemical Interferents: Models, Figures, and Intrepretations*

#### **Introduction**

Adapting Cachin's KLD-based approach to a FOM for steganography to a FOM for CSs in an idealized, interferent-free environment is relatively straightforward: The probability distributions of the covertext and the loaded text are replaced with the probability distributions of the detection responses of the TA and of the CS respectively. Afterwards, all subsequent calculations proceed in a similar fashion to that followed by Cachin.

However, when one moves from the idealized case to a CS in the presence of chemical noise in a realistic environment, a more nuanced approach focused on the practitioners' end goals must be developed. In a laboratory environment, one may safely assume that the majority of a detector's response is due to the TA, some well-characterized spectator analytes, and the inherent zero-mean noise of the detector. In most operational settings, however, a significant fraction of that signal will be due to chemical noise representing uncharacterized and unwanted signal whose origins lie in the presence of unknown chemical interferents. As it is a legitimate, if unwanted, signal, chemical noise does not average out like true random noise.

### A Model for the Chemical Environment for Figure of Merit Analysis

Since much of the novelty of this approach resides in the probabilistic description of the external chemical environment, the remainder of this subsection will focus on the formulation of $P_E\left(\boldsymbol{X}_E;\boldsymbol{\theta}_E\right)$, the probability distribution associated with the chemical intereferents present in the external chemical environment and their concentrations. This external chemical environment model generates multiple random variables which define the identities of the unknown chemical interferents as well as their concentrations. Our probabilistic model for the external chemical environment may be expressed as a probability distribution,

$$P_E\left(\boldsymbol{X}_E;\boldsymbol{\theta}_E\right)=P_{\boldsymbol{c}}\left(\boldsymbol{c}|\boldsymbol{a}\right)P_{\boldsymbol{a}}\left(\boldsymbol{a}\right) \tag{104}$$

This expression generates variables which determine the number, identity, and concentrations of the chemical unknowns present in the external environment. $P_{\boldsymbol{a}}(\boldsymbol{a})$ is the probability distribution which generates the number and identity of the chemical interferents present in the environmental background. This distribution accomplishes this by generating a vector of boolean variables $\boldsymbol{a}$ of total length $N$ where $N$ is the total number of compounds being considered as chemical interferents. $P_{\boldsymbol{c}}\left(\boldsymbol{c}|\boldsymbol{a}\right)$ is the probability distribution which generates the concentration vector $\boldsymbol{c}$ associated with $\boldsymbol{a}$. Both of these vectors $\boldsymbol{c}$ and $\boldsymbol{a}$ are of length $N$ which means that concentrations of excluded intereferents are calculated but since they are multiplied by 0, the boolean value of their presence, they are effectively excluded. This somewhat artificial formalism is used for the dual purposes of mathematical and computational convenience.

The probabilistic responses of the chemical detection system defined by the probability distribution $P_D(\boldsymbol{X}_D|\boldsymbol{X}_E;\boldsymbol{\theta}_D)$ are dependent upon the chemical interferents generated by $P_E(\boldsymbol{X}_E;\boldsymbol{\theta}_E)$. The specifics of this preconditioning are dependent upon the sensing system in question, the known chemical spectators and the TA itself. Our explicit consideration of them is deferred until a later section.

### Figures of Merit for Chemical Simulants

The interactions between the CS(s) or the TA(s) and chemical noise present in the background environment are central to our approach to a rigorous FOM. Mathematically, we describe them by a two part statistical model for a joint chemical detector/environmental system: The first piece is a model for the response of the chemical detector conditionally dependent upon the internal noise of the detector, the TAs and/or the CSs and the known chemical background. The second piece is a model for the environment comprised of the unknown chemical interferents and concentration. The joint distribution of these models is given by the following equation:

$$P_{D+E}\left(\boldsymbol{X}_{D+E};\boldsymbol{\theta}_{D+E}\right)=P_D\left(\boldsymbol{X}_D|\boldsymbol{X}_E;\boldsymbol{\theta}_D\right)P_E\left(\boldsymbol{X}_E;\boldsymbol{\theta}_E\right) \tag{105}$$

where $\boldsymbol{X}_{D+E}$, $\boldsymbol{X}_D$, and $\boldsymbol{X}_E$ are the random variables generated by the combined detector and environment, detector, and environment respectively. Likewise, $\boldsymbol{\theta}_{D+E}$, $\boldsymbol{\theta}_D$, and $\boldsymbol{\theta}_E$ denote the parameters for the whole system, the detection system, and chemical environment respectively.

Using this detector/environment system model, a CS FOM may be formulated in a variety of different ways depending on the specific goal(s) of the practitioner(s). The following divergences may be used to assess various CS design scenarios given below: Using this detector/environment system model, a CS FOM may be formulated in a variety of different ways depending on the specific goal(s) of the practitioner(s). The following divergences may be used to assess various CS design scenarios given below:

$$D(P^{TA}_{D|E}||P^{S}_{D|E}) = \int d\boldsymbol{X}_D \, P^{TA}_{D|E}(\boldsymbol{X}_D;\boldsymbol{X}_E) \log \frac{P^{TA}_{D|E}(\boldsymbol{X}_D;\boldsymbol{X}_E)}{P^{S}_{D|E}(\boldsymbol{X}_D;\boldsymbol{Y}_E)} \tag{106}$$

This first FOM, eqn. (106), measures the divergence between a TA and a CS detector system such that the external chemical environments of both detector systems are fixed but necessarily equivalent. This FOM allows practitioners to investigate how a CS fares in an external environment different from that of the TA so that they may investigate the covering capacity of chemical noise.

$$\begin{aligned}\langle D(P^{TA}_{D}||P^{S}_{D})\rangle_E = &\int d\boldsymbol{X}_E P^{TA}_{E}(\boldsymbol{X}_E)\cdot \\ &\left(\int d\boldsymbol{X}_D \, P^{TA}_{D}(\boldsymbol{X}_D|\boldsymbol{X}_E) \log \frac{P^{TA}_{D}(\boldsymbol{X}_D|\boldsymbol{X}_E)}{P^{S}_{D}(\boldsymbol{X}_D|\boldsymbol{X}_E)}\right)\end{aligned} \tag{107}$$

This second FOM, eqn. (107), measures the divergence between a detector's responses when subjected to a TA and a CS subject to the same external chemical environment. The influence of this external environment is then averaged out to yield an average KLD for the detection system. This FOM allows practitioners to design the best CS on average for a system subjected to a set of possible external chemical environments.

$$D(P^{TA}_{D+E}||P^{S}_{D+E}) = \int d\boldsymbol{X} \, P^{TA}_{D+E}(\boldsymbol{X}) \log \frac{P^{TA}_{D+E}(\boldsymbol{X})}{P^{S}_{D+E}(\boldsymbol{X})} \tag{108}$$

Finally, this third FOM, (108), measures the divergence between joint detector-environment systems in the presence of a TA and a CS while assuming that all of the background interferents are the same for each instance of the system. This FOM is targeted toward assessment of the CS if the practitioner wants it to truly simulate the TA response even for unknown chemical interferents.

## 7. INTERPRETING FIGURES OF MERIT FOR CHEMICAL SIMULANTS: DERIVING INEQUALITIES FOR ROC CURVE ANALYSIS

Going beyond the interpretation of the CS FOMs as non-metric KLD-based measures of relative worth in a fashion appropriate for practical effect requires more work. While Cachin developed an inequality relating his FOM for steganography to the probability of falsely declaring an embedded message not to be present, we improve on this approach both in the quality of the bound and in the ease-of-use to the end user. Like Cachin, we combine hypothesis testing methods with the information processing inequality, but instead of weakening bounds in order to get an analytic result, we use numerical solving techniques to draw receiver operating characteristic (ROC) curves. These ROC curves provide both equivalent to or tighter bounds than analytic result while providing practitioners with relevant, interpretable diagrams.

We utilize the information processing inequality [11, 176] to relate the various KLD quantities given by eqns. (106), (107), and (108),

$$D(H^{TA}_{D:E}(\boldsymbol{X})||H^{S}_{D:E}(\boldsymbol{X})) \leq D(P^{TA}_{D:E}(\boldsymbol{X})||P^{S}_{D:E}(\boldsymbol{X})) \tag{109}$$

to decisions concerning hypotheses denoted by $H^{TA}_{D:E}$ and $H^{S}_{D:E}$. Probability distributions for Two hypotheses $H$ defines a hypothesis which has binary testing function operating on the random variable $\boldsymbol{X}$ generated by the probability distribution denoted by the super and subscripts on both the binary testing function and the relevant probability distributions and the ":" operator is meant to stand in for "+", "|" or the implication "marginalized over $E$" as appropriate. Inequalities relevant to applications and practitioners may then be derived from the KLDs for binary hypothesis testing.

To help set up the decision structure of our CS FOM, $H^{TA}_{D:E}$ acts as our null hypothesis and asserts the presence of the TA and not the CS; likewise, $H^{S}_{D:E}$ asserts the presence of the CS and not the TA. Following the analogous approach given by Cachin, the respective binary distributions for the Type I and Type II errors are given by $(\alpha, 1-\alpha)$ and $(\beta, 1-\beta)$ where $\alpha$ is the probability of falsely accepting $H^{S}_{D:E}$ (Type I error) and $\beta$ is the probability of falsely accepting $H^{TA}_{D:E}$ (Type II error).

The KLD for these decision probabilities is given by

$$\begin{aligned} D(H^{TA}_{D:E}(\boldsymbol{X})||H^{S}_{D:E}(\boldsymbol{X})) &= D(P(\alpha)||P(\beta)) \\ &= \alpha \log \frac{\alpha}{(1-\beta)} + (1-\alpha) \log \frac{(1-\alpha)}{\beta} \end{aligned} \tag{110}$$

which is just the discrete KLD for a binary decision between the presence or absence of a CS with the absence signified by $H^{TA}_{D:E}$ acting as the null hypothesis. Using this relation we may derive bounds on the probability $\beta$ of detecting a CS. Beginning with the following inequality,

$$\begin{aligned} D(P^{TA}_{D:E}(\boldsymbol{X})||P^{S}_{D:E}(\boldsymbol{X})) &\geq D(H^{TA}_{D:E}(\boldsymbol{X})||H^{S}_{D:E}(\boldsymbol{X})) \\ &= \alpha \log \frac{\alpha}{(1-\beta)} + (1-\alpha) \log \frac{(1-\alpha)}{\beta} \end{aligned} \tag{111}$$

we may regroup terms on the rhs so that this expression becomes,

$$\begin{aligned} D(P_{D:E}^{TA}(\boldsymbol{X})||P_{D:E}^{S}(\boldsymbol{X})) \geq\ & \alpha \log \alpha + (1-\alpha)\log(1-\alpha) \\ & -\log\beta + \alpha\log\beta - \alpha\log(1-\beta) \end{aligned} \tag{112}$$

Recognizing that $\mathscr{H}_\alpha = -\alpha\log\alpha - (1-\alpha)\log(1-\alpha)$ is the information theoretic entropy for the $\alpha$'s binary distribution, we rewrite the inequality as

$$D(P_{D:E}^{TA}(\boldsymbol{X})||P_{D:E}^{S}(\boldsymbol{X})) \geq -\mathscr{H}_\alpha - \log\beta + \alpha\log\beta - \alpha\log(1-\beta) \tag{113}$$

which may be re-expressed as $\beta^{(1+\alpha)}(1-\beta)^\alpha \geq e^{-D(P_{D:E}^{TA}(\boldsymbol{X})||P_{D:E}^{S}(\boldsymbol{X}))-\mathscr{H}_\alpha}$.

Recognizing that $1 \geq (1-\beta)^\alpha$ and then rearranging the inequality yields the lower bound $\beta \geq e^{-\frac{1}{(1+\alpha)}(D(P_{D:E}^{TA}(\boldsymbol{X})||P_{D:E}^{S}(\boldsymbol{X}))-\mathscr{H}}$ Similarly, recognizing that $1 \geq \beta^{(1+\alpha)}$ allows for an upper bound $\beta \leq 1 - e^{-\frac{1}{\alpha}\left(D(P_{D:E}^{TA}(\boldsymbol{X})||P_{D:E}^{S}(\boldsymbol{X}))-\mathscr{H}_\alpha\right)}$.

Combining these expressions yields,

$$e^{-\frac{1}{(1+\alpha)}\left(D(P_{D:E}^{TA}(\boldsymbol{X})||P_{D:E}^{S}(\boldsymbol{X}))-\mathscr{H}_\alpha\right)} \leq \beta \leq 1 - e^{-\frac{1}{\alpha}\left(D(P_{D:E}^{TA}(\boldsymbol{X})||P_{D:E}^{S}(\boldsymbol{X}))-\mathscr{H}_\alpha\right)} \tag{114}$$

which allows us to bound the probability of acceptance of a CS in lieu of a TA (Type II error) in terms of the KLD of the chemical detection scenarios.

As an aside, it is worth noting that the entirety of the preceding derivation applies to eqn. (107) as well since an average may be considered to be independent of the variables $\alpha$ and $\beta$ and are hence immaterial to the derivation of these relationships.

### Interpreting Figures of Merit for Chemical Simulants: ROC Curve Analysis

While inequalities like those given by eqn. (114) are useful analytical results, a more direct and immediately useful approach is available. Returning our attention to the processing inequality given by eqn. (113), we may change the inequality to an equality to define a soluble expression for a boundary line for the inequality as

$$D(P_{D:E}^{TA}(\boldsymbol{X})||P_{D:E}^{S}(\boldsymbol{X})) = -\mathscr{H}_\alpha - \log\beta + \alpha\log\beta - \alpha\log(1-\beta) \tag{115}$$

Assuming that a KLD value is available, we may then use a nonlinear numerical solver to solve for either $\alpha$ or $\beta$ as a function of $\beta$ or $\alpha$ respectively. The $\alpha$ and $\beta$ values may then be used to plot a ROC curve for determining the presence or absence of a CS.

A ROC curve is a plot which allows a practitioner to easily and rapidly evaluate the diagonostic ability of a binary classifier, e.g. a hypothesis test, as its discrimination threshold is v aried [193]. Specifically, a ROC curve plots the true positive rate (the probability a positive result is actually positive) vs the false positive rate (the probability a positive result is actually negative). Ideally this plot results in a concave curve in the upper left hand quadrant of the chart with a dashed line often included along the diagonal stretching from the lower left hand corner to the upper right hand corner. This dashed line represents the binary classifier attributed to a "50:50" coin flip or a random guess.

For typical binary classifiers, a ROC curve closer to the upper left hand corner of the chart is better than one closer to the diagonal. However, the ROC curve for a CS is read differently. Since in this case a CS is intended to replicate the TA in a detection scenario so a ROC curve closer to the diagonal line is preferred as it indicates that the CS is hard to distinguish from the TA.

$\alpha$ and $\beta$ may be used to define the axes of the ROC curve. $\alpha$ is the false positive rate since it is the probability of falsely accepting the CS as the TA, likewise $\beta$ is the false negative rate so $1-\beta$ is the true positive rate. This allows the end user to plot an upper ROC curve boundary under which the true ROC curve must exist. If the role of the $P^{TA}_{D:E}(\boldsymbol{X})$ and $P^{S}_{D:E}(\boldsymbol{X})$ in the KLD is switched a second boundary curve may be found as follows:

$$D(H^{S}_{D:E}(\boldsymbol{X})||H^{TA}_{D:E}(\boldsymbol{X})) = \beta \log \frac{\beta}{(1-\alpha)} + (1-\beta)\log \frac{(1-\beta)}{\alpha} \tag{116}$$

The intersection of these two curves indicates the best available boundary curve with this technique. For a CS, the true ROC curve for this detection system lies to the right of this boundary curve and the diagonal line representing a random guess between the hypotheses. The area beneath this ROC boundary defines the probability that a detector will be able to distinguish between the presence of a CS and a TA. If that probability is 0.5, the detector is totally unable to distinguish between them as it is not better than randomly guessing [193].

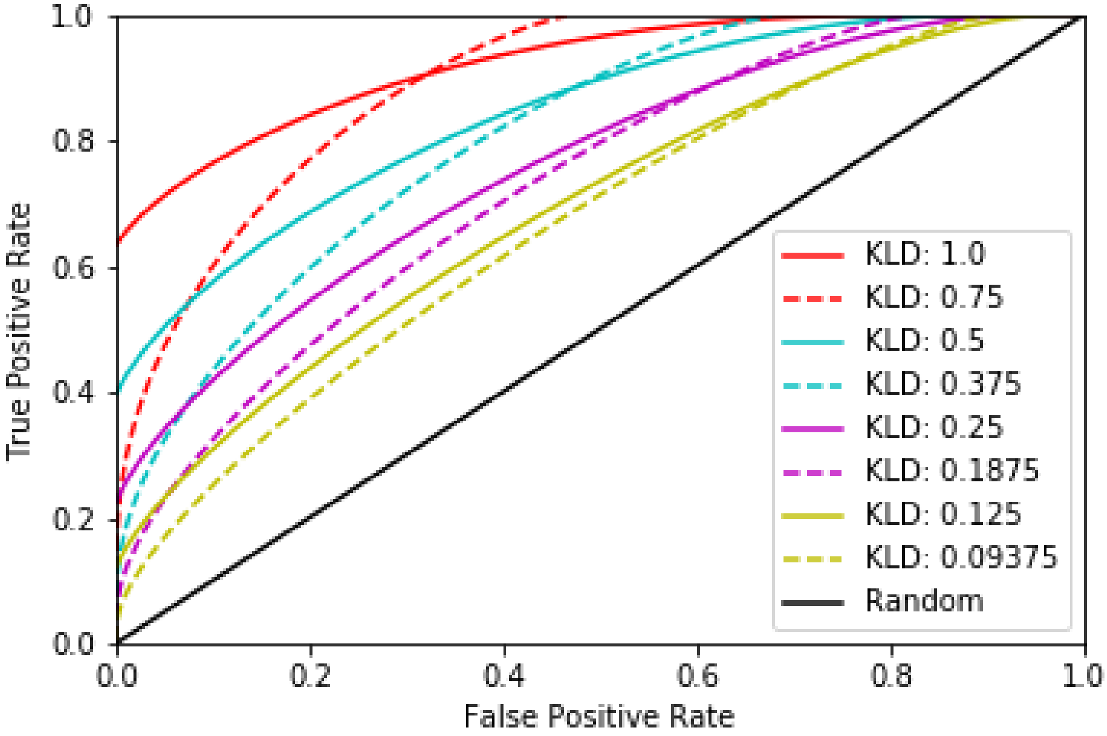

Fig. 4—This image depicts ROC boundary curves calculated with various KLD values. The solid lines depict curves calculated with the eqn. (115) and the dashed lines depict curves calculated with eqn. (116). If one presumes that each color represents a single detection/environment system then the boundary line made by joining them represents the upper limit for the location for the location of the true ROC curve for that detector system.

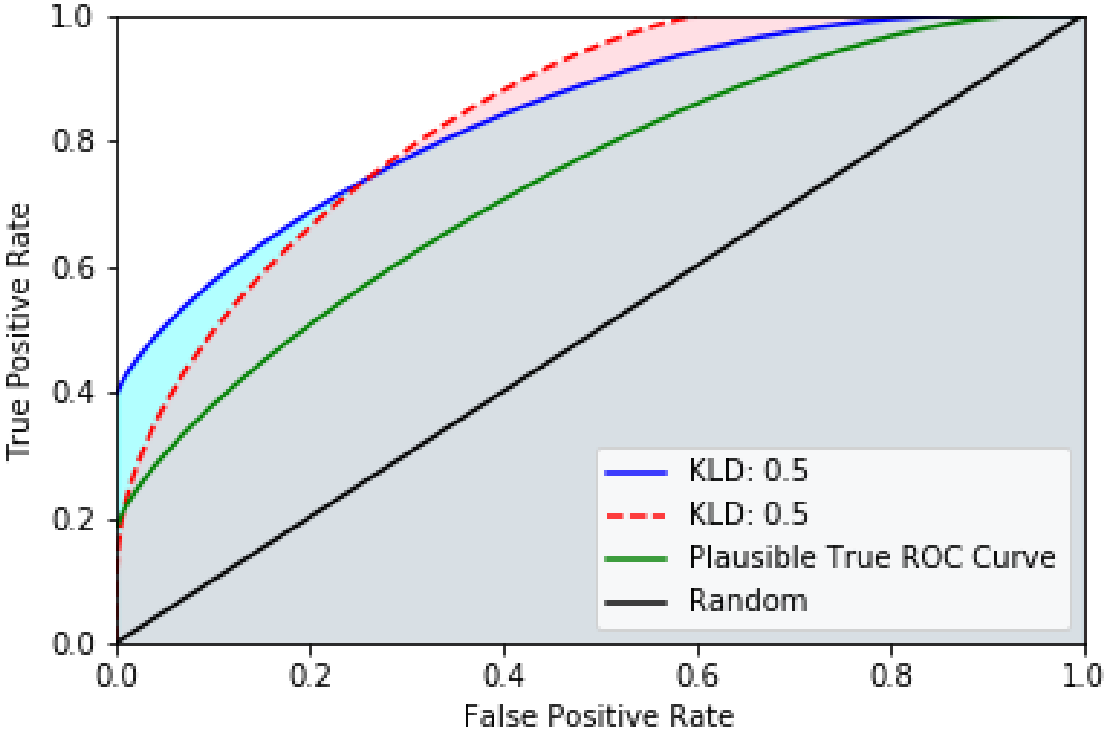

Fig. 5—This image depicts the filled area beneath a ROC boundary curve described as the joint line formed by a solid line calculated with the eqn. (115) and a dashed line calculated with eqn. (116). The shaded area is equivalent to the maximum probability that a detector with the displayed KLDs could distinguish between a CS and a TA. The green line is a plausible true ROC curve described by the boundary cuves.

## 8. ROBUSTNESS OF DETECTION IN AN OPEN WORLD: THE IMPACT OF EXTERNAL KNOWLEDGE ON SENSOR CAPABILITY

### 8.1 Introduction

Electronic nose (e-nose) technology attempts to mimic mammalian olfaction through combined cross-correlated sensor inputs and post-processing, in the ideal becoming a system greater than the sum of its parts. Unfortunately the result is often less than satisfying, leading to perceptions of the e-nose as the perpetual sensor system of tomorrow [3]. A principal reason for this is that it is often unclear how to quantitatively predict the analytical capability of an e-nose for realistic chemical sensing scenarios, particularly those involving partially unknown chemical backgrounds, sometimes referred to as "chemical noise." Consequently, it is often possible to significantly over- or under- estimate the value of a given e-nose system in a new sensing context.

Practitioners typically already possess some knowledge regarding the subject chemical environment and the parameters describing the e-nose's interactions with it. For example, there might be a particular class of chemical likely to be present as unknown interferents, leading to a bounded interval of possible sensor response characteristics to the chemical noise that are of particular concern. A method is needed that allows practitioners to incorporate this type of information in a systematic fashion to predict the quantitative impact they will have on e-nose capability for targeted sensing tasks. Typical pattern recognition or machine learning strategies used for post-processing of e-nose data [106, 194] do not enable this and can even undermine a quantitative and causal understanding of the information flow through an e-nose system if they are implemented in a "black box" fashion.

In this chapter, we take a more statistically basal approach to e-nose quality using a Fisher information-based metric. A quantitative approach to assessing the quality of an e-nose for a particular sensing task is presented. The approach uses experimentally accessible values, chemical sensing databases, and qualitative knowledge and assumptions about the sensing environment and the e-nose itself. Moreover, the methodology can be inverted to assess not only the value of the e-nose itself but the value of the chemical database and qualitative knowledge/assumptions as well.

Specifically, we propose using an empirically calculable lower bound to the Fisher information which allows us to avoid having an analytical model for the statistical error of the e-nose. Next, by using quantitative and qualitative external knowledge, like sensor response type, chemical databases, the likely class of chemical unknown, manufacturing process, analytical task etc., we show that the quality of the e-nose for a particular sensing task may be inferred via a Bayes'-type rule for the Cramér-Rao bound. Equivalently, we may invert this procedure and use it to estimate the quality of our external knowledge against a set of internal knowns. Additionally, once calculated these values may be used as "prior"-like values for a Bayesian approach to Cramér-Rao bounds and additional experimental and qualitative information may be used to sequentially update our understanding of the quality of our e-nose for a particular sensing task.

### 8.2 Quantitative Metrics and Methodology for E-Nose Quality

#### *8.2.1 Overview*

Among the many challenges in quantitatively assessing an e-nose is the metric associated with its global quality for a specific task. Typically within analytical chemistry there are three central figures of merit for an analytical instrument: precision, sensitivity and selectivity. An ideal measure for an e-nose should encompass precision and sensitivity; selectivity inherently quantifies the degree of independence in the data and is thus unsuited for a sensor array or e-nose being used to detect an unknown quantity as this task requires correlation to be detectable [195]. An efficient measure for the capability of an e-nose for a particular task is given by the Cramér-Rao lower bound (CRB), a quantity which bounds the global error $\det(\boldsymbol{\Sigma}(\boldsymbol{\theta})) \leq CRB(\boldsymbol{\theta}) = \det(\boldsymbol{F}^{-1}(\boldsymbol{\theta}))$ where $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is the covariance matrix for a statistical estimator vector and its determinant describes a strictly positive volume that may act as a score or metric for the global error [131] and $\boldsymbol{F}(\boldsymbol{\theta})$ is the Fisher information matrix for the covariance matrix.

### 8.2.2 The Fisher Information Matrix and its Lower Bound

In statistical estimation systems like an e-nose, it is often useful to have estimator-independent lower bounds for the (co)variance of its estimates like the inverse of the Fisher information (matrix) (FI/FIM) which is given by

$$\boldsymbol{F}(\boldsymbol{\theta}) = \int d\boldsymbol{X} \rho(\boldsymbol{X};\boldsymbol{\theta}) \left( (\partial_{\boldsymbol{\theta}} \ln(\rho)) (\partial_{\boldsymbol{\theta}} \ln(\rho))^T \right) \tag{117}$$

and

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \int d\boldsymbol{X} \rho(\boldsymbol{X};\boldsymbol{\theta}) \left( \hat{\boldsymbol{\theta}}(\boldsymbol{X}) - \boldsymbol{\theta} \right) \left( \hat{\boldsymbol{\theta}}(\boldsymbol{X}) - \boldsymbol{\theta} \right)^T \tag{118}$$

such that the FIM, $\boldsymbol{F}(\boldsymbol{\mu};\boldsymbol{\theta})$, provides the aforementioned CRB [11] to the covariance matrix, $\boldsymbol{\Sigma}(\boldsymbol{\theta})$,

$$\boldsymbol{F}(\boldsymbol{\theta})^{-1} \leq \boldsymbol{\Sigma}(\boldsymbol{\theta}) \tag{119}$$

where the $\leq$ relation is in the sense of a positive definite matrix. Since the CRB is independent of an estimator's specifics, it acts as a global lower bound for that estimate.

Unfortunately, a precise statistical model is often not available or practical to calculate for an FIM. Nonetheless, particularly for e-nose designers, it is very useful to have a way to rigorously quantify the quality of an e-nose using experimentally obtained values. A lower bound to the FIM may be obtained by considering the reparametrized form of the FIM [196]

$$\boldsymbol{F}(\boldsymbol{\theta}) = \left( \frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\theta}} \right) \boldsymbol{F}(\boldsymbol{\eta}(\boldsymbol{\theta})) \left( \frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\theta}} \right)^T \tag{120}$$

and then substituting via the CRB for a known vector estimator that may be transformed into the desired estimator via the Jacobian, $\left( \frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\theta}} \right)$,

$$\boldsymbol{F}(\boldsymbol{\theta}) \geq \left( \frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\theta}} \right) \boldsymbol{\Sigma}^{-1}(\boldsymbol{\eta}(\boldsymbol{\theta})) \left( \frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\theta}} \right)^T \tag{121}$$

which provides a lower bound to the Fisher information matrix (LBF) of interest via an experimentally accessible covariance matrix and a sensor response dependent Jacobian; an alternate derivation for eqn.

(121) is provided by Stein *et. al.* [197]. This result shows that the FIM for a system with parameter independent Gaussian noise is the global lower bound to any sensing device and thus a conservative approach for analysis and optimization which may be subsequently improved upon with more careful study.

### 8.2.3 A Bayes'-Type Theorem for the Cramér-Rao Lower Bound

Given eqn. (121), a LBF for an arbitrary e-nose in the presence of known chemicals may be calculated and a corresponding global error bound or capability metric for that e-nose for an underlying analytical task calculated. This known estimate may then be combined with either rigorously known or qualitatively known responses and correlations from a database of chemical sensor responses to give an estimate of the quality of the e-nose in presence of a previously unanalyzed chemical (class).

Bayes'-type reasoning for the Cramér-Rao bound may be formulated via the following expressions [146] where,

$$\boldsymbol{F}(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{F}_{\boldsymbol{\alpha}} & \boldsymbol{F}_{\boldsymbol{\alpha\beta}} \\ \boldsymbol{F}_{\boldsymbol{\beta\alpha}} & \boldsymbol{F}_{\boldsymbol{\beta}} \end{bmatrix} \tag{122}$$

is the FIM for an vector estimator $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta})$, where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are subvectors that divide $\boldsymbol{\theta}$ into 2 blocks.

The CRB for an FIM $\boldsymbol{F}(\boldsymbol{\theta})$ is given by

$$CRB(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \left| \begin{bmatrix} \boldsymbol{F}_{\boldsymbol{\alpha}} & \boldsymbol{F}_{\boldsymbol{\alpha\beta}} \\ \boldsymbol{F}_{\boldsymbol{\beta\alpha}} & \boldsymbol{F}_{\boldsymbol{\beta}} \end{bmatrix}^{-1} \right| \tag{123}$$

The CRB for an FIM where there is no uncertainty for $\boldsymbol{\beta}$ is given by

$$CRB(\boldsymbol{\alpha}|\boldsymbol{\beta}) = |\boldsymbol{F}_{\boldsymbol{\alpha}}^{-1}| \tag{124}$$

and the CRB for just the $\boldsymbol{\alpha}$ component which excludes $\boldsymbol{\beta}$ while still being influenced by it is given by

$$CRB(\boldsymbol{\alpha}) = |(\boldsymbol{F}_{\boldsymbol{\alpha}} - \boldsymbol{F}_{\boldsymbol{\alpha\beta}} \boldsymbol{F}_{\boldsymbol{\beta}}^{-1} \boldsymbol{F}_{\boldsymbol{\beta\alpha}})^{-1}| \tag{125}$$

Using the Schur decomposition, these equations may be related by a chain rule,

$$CRB(\boldsymbol{\alpha}, \boldsymbol{\beta}) = CRB(\boldsymbol{\alpha}|\boldsymbol{\beta}) CRB(\boldsymbol{\beta}) \tag{126}$$

for the CRB and by a Bayes'-type rule,

$$CRB(\boldsymbol{\alpha}) = \frac{CRB(\boldsymbol{\beta})}{CRB(\boldsymbol{\beta}|\boldsymbol{\alpha})} CRB(\boldsymbol{\alpha}|\boldsymbol{\beta}) \tag{127}$$

for the CRB [146].

### *8.2.4 Estimating the Quality of an E-Nose and the Value of External Quantitative and Qualitative Knowledge*

Using the equations from the preceding subsections, we are now equipped to estimate the quality of an e-nose in the presence of a chemical unknown(s) or conversely to quantitatively value our external knowledge of a chemical sensing situation. To do this, we set up an approximate covariance matrix,

$$\widetilde{\boldsymbol{\Sigma}}(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{\alpha}}(\boldsymbol{\theta}) & \widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{\alpha\beta}}(\boldsymbol{\theta}) \\ \widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{\beta\alpha}}(\boldsymbol{\theta}) & \widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}}(\boldsymbol{\theta}) \end{bmatrix} \tag{128}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}(\boldsymbol{\theta})$ is an experimentally measured covariance matrix for known relevant chemical components, the tilde indicates an approximated quantity i.e. $\widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{\alpha\beta}}(\boldsymbol{\theta}), \widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{\beta\alpha}}(\boldsymbol{\theta}), \widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}}(\boldsymbol{\theta})$, the subscript $\boldsymbol{\alpha}$ denotes known measured chemical quantities and the subscript $\boldsymbol{\beta}$ denotes the unknown chemical quantities, and the subscripts $\boldsymbol{\alpha\beta}$ and $\boldsymbol{\beta\alpha}$ denote the covariances caused by the interaction between the known and unknown chemicals.

While $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}(\boldsymbol{\theta})$ is either just the measured covariance matrix for the chemical quantities in question or may be derived using eqn. (121), $\widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}}(\boldsymbol{\theta})$ must be estimated from known qualitative data. Despite not knowing the exact chemicals present as either interferents or as targets, a chemical analyst frequently knows what class(es) of chemicals might be present or at least are of interest for the analytical task at hand. He or she may then use sensor or instrument specific chemical databases, quantitative structure–activity relationship models (QSAR) [198] or related semi-empirical methods to model the chemical-sensor interaction and response as reflected in the FIM or covariance matrix.

In doing so, we are specifically interested in probing how the error term given by the combination of the experimental LBF and the numerical fitting of the qualitative knowledge changes with different assumptions for our unknowns or knowns depending upon the analytical task. From a mathematical standpoint, the volumetric portion of the CRB for the unknowns $\beta$ given an imprecisely known $\alpha$ are calculated as

$$CRB(\boldsymbol{\beta}) = \frac{CRB(\boldsymbol{\alpha})}{CRB(\boldsymbol{\alpha}|\boldsymbol{\beta})} CRB(\boldsymbol{\beta}|\boldsymbol{\alpha}) \tag{129}$$

for the covariance form of the CRB. Similarly, the total CRB is given as

$$CRB(\boldsymbol{\alpha},\boldsymbol{\beta}) = \begin{vmatrix} \boldsymbol{\Sigma}_{\boldsymbol{\alpha}} & \widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{\alpha\beta}} \\ \widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{\beta\alpha}} & \widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}} \end{vmatrix} \tag{130}$$

and the CRB for an imprecisely known $\alpha$ with an imprecisely known $\beta$ is given as

$$CRB(\boldsymbol{\alpha}|\boldsymbol{\beta}) = |\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}| \tag{131}$$

Finally, a direct calculation of the error attributable to an imprecisely known $\alpha$ and $\beta$ respectively are given by

$$CRB(\boldsymbol{\alpha}) = |\boldsymbol{\Sigma}_{\boldsymbol{\alpha}} - \widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{\alpha\beta}} \widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}}^{-1} \widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{\beta\alpha}}| \tag{132}$$

$$CRB(\boldsymbol{\beta}) = |\widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}\boldsymbol{\alpha}}\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^{-1}\widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{\alpha}\boldsymbol{\beta}}| \tag{133}$$

The preceding equations provide a quantitative methodology for evaluating and accounting for the error given by known and unknown quantities using strictly empirical methods and qualitative assumptions or data for the analytical situation. Just as importantly the resulting $CRB(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is useful as a sort of "prior" in a Bayesian approach to updating this LBF quantity [146].

### 8.3 Conclusion

This chapter has sought to show that since the chemical unknowns or chemical noise are frequently "known" in a qualitative fashion when an e-nose is used to sense a chemical environment that this additional information may be used and leveraged in modeling the capability of an e-nose for a particular analytical task. Often due to the analytical task or the manufacturing and refinement processes, the chemical unknowns belong to a likely class or classes of known chemicals. This is useful information as it shows how well a new detection strategy for a chemical unknown may integrate into a previously developed e-nose system. These quantitative inputs for qualitative knowns may be developed by fitting to trends presented in chemical databases for various chemical correlations amongst and within sensors. Ultimately, by utilizing the Bayes' rule setup given by eqns. (126) and (127), it is possible to quantify how much value our qualitative knowledge may provide for developing our e-nose system by formally analyzing its influence upon the error bound for the target knowns and unknowns. Depending upon the immediate goals and desires of the analytic practitioner, the resulting information may be used to make quality judgments upon the e-nose itself, the assumptions made, or as a prior for a global error bound to be subsequently updated through experiment.

## 9. CHEMICAL GAMUT: UNDERSTANDING THE ”EXPRESSIVENESS” OF A SENSING TECHNIQUE

In sensor array design, one is generally concerned with optimization of sensor array configuration against a specific set of sensing tasks. However, it is occasionally useful to consider the capability of a chemical measurement strategy in the absence of a specific sensing application or task family. In particular, this becomes more important as the span of potential sensing tasks of concern, and thus the potential sensing task complexity becomes very large. As with sensor array configuration, the complexity of potential sensing tasks is subject to combinatorial explosion, with the number of potential mixtures increasing exponentially with the number of potential independently-varying chemical compounds. In the logical extreme, one is left with the intractable problem of an essentially infinite-dimensional sample space encompassing all possible chemicals. Obviously, no finite-dimensional sensor array can be configured to “solve” what amounts to a completely generalized mixture characterization problem in analytical chemistry. However, it is also true that a given sensor array will be capable of addressing various subspaces within this infinite-dimensional problem. The span of sensing tasks addressable by a given measurement strategy form what might called the “chemical gamut” of that strategy, in analogy to color theory in which color gamut expresses the portion of the full color space that can be represented, or reproduced by a system. This can be thought of as a measure of the general chemical informing power of a given sensor array. An assessment of this measure could provide a useful criterion for selecting one measurement strategy over another or, alternatively, for selecting one to augmenting another, when resources are limited and there is an expectation of a need to be able to address a wide range of sensing tasks.

Consider a nominal, hypothetical linear sensor system exhibiting additive response. The mean response vector, $\boldsymbol{\mu}$, from an array of $n$ sensors to a chemical stimulus $\boldsymbol{c}$ with $m$ components can be expressed as a system of linear equations:

$$\boldsymbol{\mu} = A\boldsymbol{c} \tag{134}$$

Where $A$ is an $n$ by $m$ matrix with elements $a_{i,j}$ corresponding the linear sensitivity of the $i$th sensor to the $j$th chemical component. In this example, the role of the molecular interaction between the sensor and chemical stimulus is clearly observed, with the matrix of sensitivities providing a linear transform from chemical stimulus space to sensor array response space. The concentrations of the components of the chemical stimulus can be calculated from the measured sensor response vector as:

$$\boldsymbol{c} = A^{-1}\boldsymbol{\mu} \tag{135}$$

provided the left inverse of $A$ exists, which requires $m \leq n$ (i.e. more sensors than chemical mixture components) and for the rank of $A$ to be equal to $m$. To examine sensor array performance when these conditions do not hold, consider two chemical stimuli, $\boldsymbol{c}_1$ and $\boldsymbol{c}_2$. These stimuli will result in two corresponding response vectors, $\boldsymbol{\mu}_1$ and $bs\mu_2$. The difference between these chemical stimuli, $\Delta\boldsymbol{c}$ is related to the difference in response,$\Delta bs\mu$ by:

$$\Delta \boldsymbol{c} = A^{-1} \Delta \boldsymbol{\mu} \tag{136}$$

If the sensor array produces identical responses for two chemical stimuli the following system of homogenous linear equations results:

$$A^{-1} \Delta \boldsymbol{\mu} = \mathbf{0} \tag{137}$$

Thus, if matrix $A$ has a non-empty null space, then any difference vector within this space indicates that the corresponding stimuli are not distinguishable. This condition arises when the rank of A is exceeded by the number of chemical components describing the stimulus space and reflects the underlying ambiguity of projecting from a high-dimensional space to a lower-dimensional space. In practical terms, this means that a given sensor array exhibits an inability to address the span of sensing tasks described by the basis set of the null space of $A$, or in other words, will be "blind" to variation along specific directions in the space of possible chemical stimuli.

Likewise, any difference between chemical stimuli that is within the row space of matrix $A$ will result in a non-zero difference in array response vector. Thus, the row space of A provides a direct expression of chemical gamut, in that, in principle, the sensing tasks represented within this row space are addressable by the sensor array. However, sensor measurement error will degrade the ability to resolve adjacent locations in the space of potential array responses along the response difference vector, $\Delta bs\mu$, and the projection performed by $A$ on will vary the magnitude of the response difference observed across different differences in chemical stimulus.

These concepts connect directly to the variance bound derived from Fisher information. For systems with a non-empty null space, the geometric limitation described above leads to an infinite error bound, interpretable as an inability to project data from a lower dimensional space to a higher-dimensional space without ambiguity. For systems with a full-rank row space, the variance bound will reflect the impact of measurement error on inferring location in chemical stimulus space, given matrix $A$. Thus, displacements in chemical stimulus space associated with smaller response difference and/or increased measurement error will exhibit greater marginal error bounds.

Next, we consider the contributions of sensor array and chemical stimulus on the sensitivity matrix, and their impact on design of sensor arrays for general-purpose analysis of chemical mixtures. As discussed earlier, sensor response is dependent on the nature of the molecular interaction between the sensor and chemical stimulus. This means that, in theory, such interactions can be parameterized by molecular properties of both the potential chemical stimuli and the sensor. For example, consider a hypothetical sensor response mechanism in which sensitivity of a given sensor to a specific analyte is given by the sum of the products of k matched pairs of analyte and sensor parameters.

$$a = \sum_{i=1}^{k} p_{s,i} p_{a,i} \tag{138}$$

Where $p_{a,i}$ and $p_{s,i}$ are the $i$th matched pair of analyte and sensor parameters, respectively. Thus, the matrix $A$ is equal to the product of matrices Ps and Pa, whose rows and columns contain the $k$ parameters for each of the $n$ sensors and $m$ analytes, respectively:

$$A = P_s P_a \tag{139}$$

Giving array response model:

$$\boldsymbol{\mu} = P_s P_a \boldsymbol{c} \tag{140}$$

Matrix multiplication by the left inverse of $P_s$ gives:

$$P_s^{-1} \boldsymbol{\mu} = P_a \boldsymbol{c} \tag{141}$$

In practical terms, the matrix of sensor parameters, $P_s$, depends only on the sensor array configuration while $P_a$ is defined in the context of the chemical compounds over which $\boldsymbol{c}$ is defined. The equivalence above shows that the informational linkage between chemical stimulus and sensor array response lies in an intermediate $k$-dimensional parameter space, owing to the $k$ independent parameters describing the molecular interaction between the sensors and chemical stimuli. Thus, intuitively, one can view a chemical sensor array as a physical instantiation of a data compression algorithm that is governed by the molecular complexity of the interaction between the sensors and chemical stimuli.

The left hand side of the equation allows consideration of the impact of sensor array configuration. Sensor responses are projected into this intermediate parameter space via the inverse of the sensor parameter matrix and analyte concentrations via the analyte parameter matrix. The sensor parameter matrix needs to be invertible for this equivalence to hold, implying that sensors should be chosen to span the parameter space, i.e., the sensor parameter matrix should be of rank k, ensuring an empty null space. This suggests that sensor array size should not be less than $k$ and also that there is limited utility in configuring sensor arrays with a number of unique sensors that greatly exceeds k, as the rank is limited by the molecular interaction complexity. The relative capabilities of two sensor array configurations can be compared by examining their respective sensor parameter matrices and evaluating Fisher information metrics related to variance bound associated with estimating locations in the intermediate parameter space. Sensor array optimization is thus straightforward and independent of chemical stimulus.

The right hand side of equation the above equation allows consideration of the ability of the sensor transduction mechanism to address specific chemical sensing tasks. The complexity of this molecular interaction will limit the number of chemical components that can be independently determined to $k$. Highly complex chemical stimuli will have greater that $k$ chemical components, and thus the analyte parameter matrix will exhibit a non-empty null space with a basis set that encompasses all of the sensing tasks in the chemical stimulus space which result in no corresponding displacement in the intermediate parameter space. Importantly, the inability to address these tasks is not a function of sensor array configuration, but rather a limitation on the sensing technique itself. The row space of the analyte parameter matrix encompasses the span of sensor tasks that produce a non-zero displacement in the parameter space, and thus a fundamental measure of the span of sensing tasks achievable by the sensing technique within the context of a chemical stimulus space of arbitrary dimension. The row space will also describe the relative magnitude of displacement in the parameter space that will be associated within the span of addressable chemical sensing tasks in the chemical stimulus space, and together with the sensor parameter matrix, will enable calculation of variance bounds for these tasks via the Fisher information. Two different sensing techniques can thus be compared to one another by comparing the null and row spaces of their respective analyte parameter matrices, defined on an arbitrary chemical stimulus space of interest. The impact of fusing two sensor techniques can be evaluated by examining the augmented analyte parameter matrix they form, showing additional capability as specific reduction in the span of the null space and increase in the span of the row space.

Conceptually, the intermediate parameter space “throttles” the amount of information that can be conveyed through the sensor array: first, by limiting the total number of chemical stimuli that can be described at once, and second, by compressing the representations of the chemical stimulus space according similarity in molecular interaction parameters among the chemical components, resulting in increased estimation error bounds, as implied by the relevant Fisher information metrics. The total amount of chemical information a given sensor array can convey can be extended by implementing operations that reduce a complex chemical sensing task to a series of simpler tasks performed in series. For example, a chromatographic separation performed prior to the sensor array will decrease the sensing task complexity associated with a chemical mixture to a series of one-dimensional tasks performed in series, assuming each mixture component is fully resolved. Even under conditions where only partial chromatographic resolution is achieved, a significant reduction in mixture complexity as a function of time can be provided, rendering the underlying sensing tasks within the capability of a sensing technique.

The example shown in this section is limited to a nominal array of linear sensors, with a sensor response mechanism that allows direct expression via linear algebra. However, the general concept of this analysis should hold for non-linear response functions and more complex sensor response mechanisms. Namely, all sensor response mechanisms will depend on molecular properties of both the sensor and the chemical with which it interacts. Provided this interaction is separable into components specific to each, sensor array configuration optimization against the intermediate parameter space can be considered independently using Fisher information metrics. Likewise, the generalized analytical capability of a sensing technique can be assessed through analysis of the span of molecular properties it interrogates in chemical stimuli through explicit calculation of the span of sensing tasks within the chemical stimulus space that result in zero displacement in the intermediate molecular parameter space. The finite span of possible molecular interactions will provide a finite span of addressable sensing tasks, or chemical gamut, that is intrinsic to the sensing technique.

## 10. CONCLUSIONS

This work has demonstrated that an information-theoretic approach can provide quantitative guidance in the design and evaluation of sensor arrays for chemical detection. Such arrays have been (and continue to be) of significant interest because they mirror evolution-optimized biological solutions to chemical perception, as well as because they present the possibility of relatively low-cost, low-complexity field-portable hardware for detection and analysis of chemical mixtures, relative to the analytical instrumentation that is currently used for such tasks in the laboratory. However, despite the excitement engendered by this approach, sensor arrays tend to offer disappointing performance in real applications and have been relegated to niche applications. One likely reason for this is that sensor arrays are generally constructed in an ad hoc, intuitive manner rather than by rational design, and thus the capabilities and limitations of sensor arrays are poorly understood prior to use.

A central tenet of this work is that sensor array design is optimized by properly specifying the sensing tasks the sensor array will be required to accomplish and then engineering appropriate complexity into the sensor array to match those tasks. While this may sound obvious, sensing tasks are often poorly specified or overly simplified in literature, and almost never defined in rigorous mathematic terms. Here, we provide a systematic approach for mathematical expression of sensing tasks. In this approach, the span of the n possible chemical stimuli that may impinge upon the sensor array form an n-dimensional Euclidean space in which each axis describes the concentration of a particular compound. Sensing tasks are essentially represented as vectors within this space that describe the difference between locations in the space. A sensing task can be successfully addressed if the sensor array responses due to each location can be discerned from one another with a specified level of accuracy. Thus, sensor array capability ultimately resides in the array's ability to provide a bijective mapping between the span of possible array response vectors and the locations in the chemical stimulus space that are relevant to the required sensing tasks as well as the degree of measurement error that is oriented along the vectors describing the sensing tasks.

As described earlier by Pearce, et al., the Fisher information can be used to provide lower-bound on the error of inferring chemical concentration from sensor array measurements, via the Cramer-Rao bound [21–23, 25, 27, 28, 199]. Specifically, the inverse of the Fisher information matrix provides a lower-bound estimate of the covariance in chemical stimulus space associated with a given set of sensor response functions and sensor noise models. This bounded estimate does not provide knowledge of the error associated with a particular device, but it does provide a measure of the theoretical limits to device capability, making it useful in optimizing sensor array design. Pearce centered discussion on use of the diagonal elements of the matrix as error bounds associated with individual components within a chemical stimulus. This work expanded beyond Pearce's to further develop sensor array optimization metrics and place them within a more rigorous framework for sensing tasks and for use in more complex sensing scenarios.

As it is directly calculated from the chemical stimulus-specific sensitivities provided by each element of the sensor array together with a statistical distribution modeling sensor noise, the Fisher information matrix intrinsically captures a notion of aggregate sensor array sensitivity relative to specific directions in chemical stimulus space. More specifically, it embodies a notion of a "signal-to-noise" type measure as greater aggregate sensitivity to a given sensing task under constant sensor measurement error leads to smaller estimation error bounds in chemical stimulus space. The lower-bound covariance matrix derived from the Fisher information matrix can thus be used to assess sensing task-specific sensitivity or signal-to-noise by calculating the marginal variance associated along the direction in chemical stimulus space associated with that task. Alternatively, if general capability against a specific chemical stimulus space is desired, the determinant of the bounded covariance matrix can be used as a global error criterion that can be minimized in sensor array configuration problems.

A sensor array-specific notion of selectivity that is congruent with classic notions of analytical selectivity was also derived from Fisher information. Qualitatively, all common notions of selectivity pertain to the degree of shared response a sensor (or sensor array) exhibits between different chemical stimuli. A significant consequence of this is that selectivity in sensor arrays can be characterized quantitatively as the degree of correlation induced in bounded estimation error in chemical stimulus space, and thus will exhibit specific impacts on specific sensing tasks. For example, a sensor array with lower selectivity between two chemical components will exhibit greater estimation error bounds for sensing tasks involving discernment of different mixtures of those components, but also an improved capability to discern between different concentrations of a given mixture of those compounds.

A central problem in sensor array design is its combinatorial nature, which can render array selection an impractical (if not intractable) problem for even modest array sizes with a large number of potential sensor variants to select from. Fisher information-derived metrics were shown to be positive definite, allowing the use of convex optimization algorithms for efficient searching of potentially large configuration spaces. Theoretically, techniques in algorithmic differentiation can be used to speed Fisher information calculation in optimization algorithms, further improving optimization efficiency. It was also demonstrated that correlated noise in sensor arrays (for example, via noise introduced by the sensor transduction electronics) can be incorporated to an algorithmic optimization framework using elliptically-contoured distribution models.

The information-theoretic approach to sensor array design can also provide a powerful framework for assessing and optimizing sensor arrays against uncertain and complex backgrounds. An information divergence-based measure for evaluating chemical simulant efficacy was developed and linked to receiver-operator characteristic (ROC) curves for easy interpretability. This measure allows generalized consideration of the ability of one chemical stimulus to represent another under varying chemical backgrounds. This has significant implications in both the design of chemical simulants for developing new sensor technologies of hazardous materials, as well as in the evaluation of likely interferences for new and existing sensor technologies. The ability to consider varying backgrounds, together with the development of statistical models of chemical background variation can reduce the risk of surprise in real-world performance relative to laboratory conditions. The capabilities of sensor arrays for particular sensing tasks under different, uncertain chemical backgrounds can be assessed by Shurr decomposition of the underlying Fisher information problem, which allows separation of the chemical stimulus space into analyte and background subspaces such that capabilities can be probed against varying background composition, as well as quantitative evaluation of the potential capability gain of acquiring external knowledge of the background.

Finally, an analytical approach for understanding the fundamental chemical expressiveness of a sensor type was described. As sensor response is the result of molecular interactions between the chemical stimulus and the sensor, this interaction should be parameterized according to properties of both. This carries important implications. First, it implies that the manner in which the sensor technology interacts with the chemical environment presents a critical informational bottleneck that fundamentally limits the span of addressable sensor tasks, both through limitations on the ability of arrays to provide a bijective mappings between sensor arrays and chemical stimulus spaces relevant to particular sensing tasks, as well as though the degree of array selectivity the allow, which limits the capabilities of the technique in addressable sensing tasks. Second, it implies that sensor design considerations can be globally optimized against this bottleneck, as opposed to against a vast, and potentially unknowable chemical stimulus space. Last, it shows the theoretical utility of hybrid arrays that combine multiple types of sensors using fundamentally different transduction mechanisms, as this approach offers the potential to widen the span of potential information that can be conveyed across a sensor array beyond what can be achieved by merely expanding the size of an array utilizing only one sensor technology. This work has demonstrated a rational, quantitative approach for designing and optimizing sensor arrays for chemical detection. The degree to which sensor array capabilities and limitations can be properly understood is unavoidably linked to the degree to which the underlying response mechanism has been characterized, including a characterization of sensor noise. Further, expressions of sensor array capabilities are inextricably linked to specific formulations of sensing task. In scenarios where detailed knowledge of sensing task is available (i.e. specification desired discernible chemical stimuli and characterization of chemical background) efficient sensor arrays can be designed at the expense of robustness. When the sensing task is poorly understood, additional complexity can be engineered into the detection system to compensate for this uncertainty. This suggests that investigation of the molecular basis of sensor response mechanism and practical investigations into the nature of the sensing tasks that one wishes to address are equally as important as development of the sensor technology itself, although neither typically receives as much attention in the literature of chemical sensors.

## (U) ACKNOWLEDGMENTS

The authors would like to thank the Office of Naval Research for funding.

## (U) REFERENCES

1. K. J. Johnson and S. L. Rose-Pehrsson, “Sensor Array Design for Complex Sensing Tasks*,” *Annual Review of Analytical Chemistry* **8**, 287–310 (2015).
2. S. Marco, “The need for external validation in machine olfaction: emphasis on health-related applications,” *Analytical and bioanalytical chemistry* **406**(16), 3941–3956 (2014).
3. P. Boeker, “On ‘electronic nose’methodology,” *Sensors and Actuators B: Chemical* **204**, 2–17 (2014).
4. K. L. Goodner, J. G. Dreher, and R. L. Rouseff, “The dangers of creating false classifications due to noise in electronic nose and similar multivariate analyses,” *Sensors and Actuators B: Chemical* **80**(3), 261–266 (2001).
5. E. Zubritsky, “Product review: e-noses keep an eye on the future,” 2000.
6. S. Marco, A. Gutiérrez-Gálvez, A. Lansner, D. Martinez, J. P. Rospars, R. Beccherelli, A. Perera, T. Pearce, P. F. Verschure, and K. Persaud, “A biomimetic approach to machine olfaction, featuring a very large-scale chemical sensor array and embedded neuro-bio-inspired computation,” *Microsystem technologies* **20**(4-5), 729–742 (2014).
7. R. A. Potyrailo, “Multivariable sensors for ubiquitous monitoring of gases in the era of internet of things and industrial internet,” *Chemical reviews* **116**(19), 11877–11923 (2016).
8. K. S. Booksh and B. R. Kowalski, “Theory of analytical chemistry,” *Analytical Chemistry* **66**(15), 782A–791A (1994).
9. C. E. Shannon, “A mathematical theory of communication,” *The Bell system technical journal* **27**(3), 379–423 (1948).
10. C. E. Shannon, “Prediction and entropy of printed English,” *Bell system technical journal* **30**(1), 50–64 (1951).
11. T. M. Cover and J. A. Thomas, *Elements of information theory* (John Wiley & Sons, 2012).
12. D. J. MacKay and D. J. Mac Kay, *Information theory, inference and learning algorithms* (Cambridge university press, 2003).
13. N. M. Timme and C. Lapish, “A tutorial for information theory in neuroscience,” *eneuro* **5**(3) (2018).
14. S. Vinga, “Information theory applications for biological sequence analysis,” *Briefings in bioinformatics* **15**(3), 376–389 (2014).
15. H. Aydinian, F. Cicalese, and C. Deppe, *Information theory, combinatorics, and search theory* (Springer, 2013).
16. H. KAISER, “QUANTITATION IN ELEMENTAL ANALYSIS,” *ANALYTICAL CHEMISTRY* **42**(2), A24+ (1970), ISSN 0003-2700, doi:10.1021/ac60284a022.

17. H. Kaiser, "Report for analytical chemists. II. Quantitation in elemental analysis," *Analytical Chemistry* **42**(4), 26A–59a (1970).

18. J. Fonollosa, L. Fernández, R. Huerta, A. Gutiérrez-Gálvez, and S. Marco, "Temperature optimization of metal oxide sensor arrays using mutual information," *Sensors and Actuators B: Chemical* **187**, 331–339 (2013).

19. T. K. Alkasab, J. White, and J. S. Kauer, "A computational system for simulating and analyzing arrays of biological and artificial chemical sensors," *Chemical Senses* **27**(3), 261–275 (2002).

20. A. Hierlemann and R. Gutierrez-Osuna, "Higher-order chemical sensing," *Chemical reviews* **108**(2), 563–613 (2008).

21. M. A. Sánchez-Montañés and T. C. Pearce, "Fisher information and optimal odor sensors," *Neurocomputing* **38**, 335–341 (2001).

22. M. A. Sanchez-Montanes, J. W. Gardner, and T. C. Pearce, "Spatio-temporal information in an artificial olfactory mucosa," *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **464**(2092), 1057–1077 (2008).

23. T. C. Pearce and M. A. Sánchez-Montañes, *Chemical Sensor Array Optimization: Geometric and Information Theoretic Approaches*, chapter 14, pp. 347–375 (Wiley-VCH, 2004), ISBN 9783527601592, doi:10.1002/3527601597.ch14. URL `http://dx.doi.org/10.1002/3527601597.ch14`.

24. M. A. S. M. nes and T. C. Pearce, "Fisher information and optimal odor sensors," *Neurocomputing* **38–40**, 335 – 341 (2001), ISSN 0925-2312, doi:http://dx.doi.org/10.1016/S0925-2312(01)00364-2. Computational Neuroscience: Trends in Research 2001.

25. T. C. Pearce and M. Sanchez-Montanes, "Chemical sensor array optimization: geometric and information theoretic approaches," *Handbook of Artificial Olfaction Machines* pp. 347–376 (2003).

26. T. Pearce, P. Verschure, J. White, and J. Kauer, "Robust stimulus encoding in olfactory processing: hyperacuity and efficient signal transmission," in *Emergent neural computational architectures based on neuroscience*, pp. 461–479 (Springer, 2001).

27. T. Pearce, "Computational parallels between the biological olfactory pathway and its analogueThe Electronic Nose':: Part I. Biological olfaction," *BioSystems* **41**(1), 43–67 (1997).

28. T. C. Pearce, S. S. Schiffman, H. T. Nagle, and J. W. Gardner, *Handbook of machine olfaction: electronic nose technology* (John Wiley & Sons, 2006).

29. P. Smets, "Belief functions," *Non-standard logics for automated reasoning* pp. 253–286 (1988).

30. J. Vessman, R. I. Stefan, J. F. Van Staden, K. Danzer, W. Lindner, D. T. Burns, A. Fajgelj, and H. Müller, "Selectivity in analytical chemistry (IUPAC Recommendations 2001)," *Pure and Applied Chemistry* **73**(8), 1381–1386 (2001).

31. J. Janata, *Principles of chemical sensors* (Springer Science & Business Media, 2010).

32. D. M. Wilson, S. Hoyt, J. Janata, K. Booksh, and L. Obando, "Chemical sensors for portable, handheld field instruments," *IEEE sensors journal* **1**(4), 256–274 (2001).

33. C. Wang, L. Yin, L. Zhang, D. Xiang, and R. Gao, "Metal oxide gas sensors: sensitivity and influencing factors," *sensors* **10**(3), 2088–2106 (2010).

34. J. Janata and M. Josowicz, "Conducting polymers in electronic chemical sensors," *Nature materials* **2**(1), 19–24 (2003).

35. L. E. Kreno, K. Leong, O. K. Farha, M. Allendorf, R. P. Van Duyne, and J. T. Hupp, "Metal–organic framework materials as chemical sensors," *Chemical reviews* **112**(2), 1105–1125 (2012).

36. R. M. Penner, "Chemical sensing with nanowires," *Annual review of analytical chemistry* **5**, 461–485 (2012).

37. H. Haick, "Chemical sensors based on molecularly modified metallic nanoparticles," *Journal of Physics D: Applied Physics* **40**(23), 7173 (2007).

38. J. W. Kemling, A. J. Qavi, R. C. Bailey, and K. S. Suslick, "Nanostructured substrates for optical sensing," *The journal of physical chemistry letters* **2**(22), 2934–2944 (2011).

39. A. Sibbald, "A chemical-sensitive integrated-circuit: The operational transducer," *Sensors and Actuators* **7**(1), 23–38 (1985).

40. J. Janata, "Potentiometric microsensors," *Chemical Reviews* **90**(5), 691–703 (1990).

41. J. W. Grate, "Hydrogen-bond acidic polymers for chemical vapor sensing," *Chemical reviews* **108**(2), 726–745 (2008).

42. J. W. Grate, "Acoustic wave microsensor arrays for vapor sensing," *Chemical Reviews* **100**(7), 2627–2648 (2000).

43. V. B. Raj, H. Singh, A. Nimal, M. Sharma, and V. Gupta, "Oxide thin films (ZnO, TeO2, SnO2, and TiO2) based surface acoustic wave (SAW) E-nose for the detection of chemical warfare agents," *Sensors and Actuators B: Chemical* **178**, 636–647 (2013).

44. I. Dufour, F. Josse, S. M. Heinrich, C. Lucat, C. Ayela, F. Ménil, and O. Brand, "Unconventional uses of microcantilevers as chemical sensors in gas and liquid media," *Sensors and Actuators B: Chemical* **170**, 115–121 (2012).

45. A. Cusano, F. J. Arregui, M. Giordano, and A. Cutolo, *Optochemical nanosensors* (CRC Press, 2012).

46. H. H. Qazi, A. B. b. Mohammad, and M. Akram, "Recent progress in optical chemical sensors," *Sensors* **12**(12), 16522–16556 (2012).

47. N. A. Rakow and K. S. Suslick, "A colorimetric sensor array for odour visualization," *Nature* **406**(6797), 710–713 (2000).

48. M. K. LaGasse, J. M. Rankin, J. R. Askim, and K. S. Suslick, "Colorimetric sensor arrays: Interplay of geometry, substrate and immobilization," *Sensors and Actuators B: Chemical* **197**, 116–122 (2014).

49. M. R. R. Khan, B. H. Kang, S. W. Lee, S. H. Kim, S. H. Yeom, S. H. Lee, and S. W. Kang, "Fiber-optic multi-sensor array for detection of low concentration volatile organic compounds," *Optics express* **21**(17), 20119–20130 (2013).

50. T. A. Dickinson, K. L. Michael, J. S. Kauer, and D. R. Walt, "Convergent, self-encoded bead sensor arrays in the design of an artificial nose," *Analytical chemistry* **71**(11), 2192–2198 (1999).

51. H. Sohn, M. J. Sailor, D. Magde, and W. C. Trogler, “Detection of nitroaromatic explosives based on photoluminescent polymers containing metalloles,” *Journal of the American Chemical Society* **125**(13), 3821–3830 (2003).

52. J. M. Sutter and P. C. Jurs, “Neural network classification and quantification of organic vapors based on fluorescence data from a fiber-optic sensor array,” *Analytical chemistry* **69**(5), 856–862 (1997).

53. G. E. Collins and S. L. Rose-Pehrsson, “Chemiluminescent chemical sensors for inorganic and organic vapors,” *Sensors and Actuators B: Chemical* **34**(1-3), 317–322 (1996).

54. Z. Zhang, S. Zhang, and X. Zhang, “Recent developments and applications of chemiluminescence sensors,” *Analytica Chimica Acta* **541**(1-2), 37–46 (2005).

55. A. Shalabney and I. Abdulhalim, “Sensitivity-enhancement methods for surface plasmon sensors,” *Laser & Photonics Reviews* **5**(4), 571–606 (2011).

56. J. Homola, S. S. Yee, and G. Gauglitz, “Surface plasmon resonance sensors,” *Sensors and actuators B: Chemical* **54**(1-2), 3–15 (1999).

57. C. M. Riley, “Statistical parameters and analytical figures of merit,” in *Progress in Pharmaceutical and Biomedical Analysis*, volume 3, pp. 15–71 (Elsevier, 1996).

58. C. D. Brown and H. T. Davis, “Receiver operating characteristics curves and related decision measures: A tutorial,” *Chemometrics and Intelligent Laboratory Systems* **80**(1), 24–38 (2006).

59. K. Faber and B. R. Kowalski, “Propagation of measurement errors for the validation of predictions obtained by principal component regression and partial least squares,” *Journal of Chemometrics: A Journal of the Chemometrics Society* **11**(3), 181–238 (1997).

60. J. F. Pierna, L. Jin, F. Wahl, N. Faber, and D. Massart, “Estimation of partial least squares regression prediction uncertainty when the reference values carry a sizeable measurement error,” *Chemometrics and Intelligent Laboratory Systems* **65**(2), 281–291 (2003).

61. W. P. Carey, K. R. Beebe, B. R. Kowalski, D. L. Illman, and T. Hirschfeld, “Selection of adsorbates for chemical sensor arrays by pattern recognition,” *Analytical chemistry* **58**(1), 149–153 (1986).

62. L. B. Buck, “Unraveling the sense of smell (Nobel lecture),” *Angewandte Chemie International Edition* **44**(38), 6128–6140 (2005).

63. K. Persaud and G. Dodd, “Analysis of discrimination mechanisms in the mammalian olfactory system using a model nose,” *Nature* **299**(5881), 352–355 (1982).

64. S. Zaromb and J. Stetter, “Theoretical basis for identification and measurement of air contaminants using an array of sensors having partly overlapping selectivities,” *Sensors and Actuators* **6**(4), 225–243 (1984).

65. S. L. Rose-Pehrsson, J. W. Grate, D. S. Ballantine, and P. C. Jurs, “Detection of hazardous vapors including mixtures using pattern recognition analysis of responses from surface acoustic wave devices,” *Analytical chemistry* **60**(24), 2801–2811 (1988).

66. W. P. Carey, K. R. Beebe, and B. R. Kowalski, “Multicomponent analysis using an array of piezoelectric crystal sensors,” *Analytical chemistry* **59**(11), 1529–1534 (1987).

67. C. Hierold and R. Müller, "Quantitative analysis of gas mixtures with non-selective gas sensors," *Sensors and Actuators* **17**(3-4), 587–592 (1989).

68. K. Ema, M. Yokoyama, T. Nakamoto, and T. Moriizumi, "Odour-sensing system using a quartz-resonator sensor array and neural-network pattern recognition," *Sensors and Actuators* **18**(3-4), 291–296 (1989).

69. J. W. Grate and M. H. Abraham, "Solubility interactions and the design of chemically selective sorbent coatings for chemical sensors and arrays," *Sensors and Actuators B: Chemical* **3**(2), 85–111 (1991).

70. J. W. Grate, S. J. Patrash, and M. H. Abraham, "Method for estimating polymer-coated acoustic wave vapor sensor responses," *Analytical Chemistry* **67**(13), 2162–2169 (1995).

71. J. W. Grate, B. M. Wise, and M. H. Abraham, "Method for unknown vapor characterization and classification using a multivariate sorption detector. Initial derivation and modeling based on polymer-coated acoustic wave sensor arrays and linear solvation energy relationships," *Analytical Chemistry* **71**(20), 4544–4553 (1999).

72. J. W. Grate, S. J. Patrash, S. N. Kaganove, M. H. Abraham, B. M. Wise, and N. B. Gallagher, "Inverse least-squares modeling of vapor descriptors using polymer-coated surface acoustic wave sensor array responses," *Analytical chemistry* **73**(21), 5247–5259 (2001).

73. M. D. Hsieh and E. T. Zellers, "Limits of recognition for simple vapor mixtures determined with a microsensor array," *Analytical Chemistry* **76**(7), 1885–1895 (2004).

74. C. Jin, P. Kurzawski, A. Hierlemann, and E. T. Zellers, "Evaluation of multitransducer arrays for the determination of organic vapor mixtures," *Analytical chemistry* **80**(1), 227–236 (2008).

75. C. Jin and E. T. Zellers, "Limits of recognition for binary and ternary vapor mixtures determined with multitransducer arrays," *Analytical chemistry* **80**(19), 7283–7293 (2008).

76. F. I. Bohrer, E. Covington, C. Kurdak, and E. T. Zellers, "Characterization of dense arrays of chemiresistor vapor sensors with submicrometer features and patterned nanoparticle interface layers," *Analytical chemistry* **83**(10), 3687–3695 (2011).

77. L. Feng, C. J. Musto, J. W. Kemling, S. H. Lim, W. Zhong, and K. S. Suslick, "Colorimetric sensor array for determination and identification of toxic industrial chemicals," *Analytical chemistry* **82**(22), 9433–9440 (2010).

78. P. J. Mazzone, X. F. Wang, Y. Xu, T. Mekhail, M. C. Beukemann, J. Na, J. W. Kemling, K. S. Suslick, and M. Sasidhar, "Exhaled breath analysis with a colorimetric sensor array for the identification and characterization of lung cancer," *Journal of Thoracic Oncology* **7**(1), 137–142 (2012).

79. Y. Zhang, J. R. Askim, W. Zhong, P. Orlean, and K. S. Suslick, "Identification of pathogenic fungi with an optoelectronic nose," *Analyst* **139**(8), 1922–1928 (2014).

80. J. R. Askim, M. Mahmoudi, and K. S. Suslick, "Optical sensor arrays for chemical sensing: the optoelectronic nose," *Chemical Society Reviews* **42**(22), 8649–8682 (2013).

81. A. Lichtenstein, E. Havivi, R. Shacham, E. Hahamy, R. Leibovich, A. Pevzner, V. Krivitsky, G. Davivi, I. Presman, R. Elnathan, et al., "Supersensitive fingerprinting of explosives by chemically modified nanosensors arrays," *Nature communications* **5**(1), 1–12 (2014).

82. N. J. Kybert, M. B. Lerner, J. S. Yodh, G. Preti, and A. C. Johnson, "Differentiation of complex vapor mixtures using versatile DNA–carbon nanotube chemical sensor arrays," *ACS nano* **7**(3), 2800–2807 (2013).

83. K. Scholten, L. K. Wright, and E. T. Zellers, "Vapor discrimination with single-and multitransducer arrays of nanoparticle-coated chemiresistors and resonators," *IEEE Sensors Journal* **13**(6), 2146–2154 (2013).

84. R. Gutierrez-Osuna and A. Hierlemann, "Adaptive microsensor systems," *Annual Review of Analytical Chemistry* **3**, 255–276 (2010).

85. A. Vergara, J. L. Ramírez, and E. Llobet, "Reducing power consumption via a discontinuous operation of temperature-modulated micro-hotplate gas sensors: Application to the logistics chain of fruit," *Sensors and Actuators B: Chemical* **129**(1), 311–318 (2008).

86. D. C. Meier, B. Raman, and S. Semancik, "Detecting chemical hazards with temperature-programmed microsensors: overcoming complex analytical problems with multidimensional databases," *Annual Review of Analytical Chemistry* **2**, 463–484 (2009).

87. B. Raman, D. C. Meier, J. K. Evju, and S. Semancik, "Designing and optimizing microsensor arrays for recognizing chemical hazards in complex environments," *Sensors and Actuators B: Chemical* **137**(2), 617–629 (2009).

88. T. Eklöv and I. Lundström, "Distributed sensor system for quantification of individual components in a multiple gas mixture," *Analytical Chemistry* **71**(16), 3544–3550 (1999).

89. S. R. Morrison, "Selectivity in semiconductor gas sensors," *Sensors and actuators* **12**(4), 425–440 (1987).

90. M. Nakagawa, S. Kawabata, K. Nishiyama, K. Utsonomiya, I. Yamamoto, T. Wada, Y. Yamashita, and N. Yamashita, "Analytical detection system of mixed odor vapors using chemiluminescence-based gas sensor," *Sensors and Actuators B: Chemical* **34**(1-3), 334–338 (1996).

91. H. Lin, M. Jang, and K. S. Suslick, "Preoxidation for colorimetric sensor array detection of VOCs," *Journal of the American Chemical Society* **133**(42), 16786–16789 (2011).

92. J. P. Viricelle, A. Pauly, L. Mazet, J. Brunet, M. Bouvet, C. Varenne, and C. Pijolat, "Selectivity improvement of semi-conducting gas sensors by selective filter for atmospheric pollutants detection," *Materials Science and Engineering: C* **26**(2-3), 186–195 (2006).

93. W. R. Collin, G. Serrano, L. K. Wright, H. Chang, N. Nunovero, and E. T. Zellers, "Microfabricated gas chromatograph for rapid, trace-level determinations of gas-phase explosive marker compounds," *Analytical chemistry* **86**(1), 655–663 (2014).

94. M. A. Ryan, A. V. Shevade, C. J. Taylor, M. Homer, M. Blanco, and J. R. Stetter, *Computational methods for sensor material selection* (Springer, 2010).

95. C. Ziegler, W. Göpel, H. Hämmerle, H. Hatt, G. Jung, L. Laxhuber, H. L. Schmidt, S. Schütz, F. Vögtle, and A. Zell, "Bioelectronic noses: a status report. Part II," *Biosensors and Bioelectronics* **13**(5), 539–571 (1998).

96. H. T. Nagle, R. Gutierrez-Osuna, and S. S. Schiffman, "The how and why of electronic noses," *IEEE spectrum* **35**(9), 22–31 (1998).

97. E. J. Cho and F. V. Bright, “Pin-printed chemical sensor arrays for simultaneous multianalyte quantification,” *Analytical chemistry* **74**(6), 1462–1466 (2002).

98. P. C. Jurs, G. Bakken, and H. McClelland, “Computational methods for the analysis of chemical sensor array data from volatile analytes,” *Chemical Reviews* **100**(7), 2649–2678 (2000).

99. K. J. Albert, N. S. Lewis, C. L. Schauer, G. A. Sotzing, S. E. Stitzel, T. P. Vaid, and D. R. Walt, “Cross-reactive chemical sensor arrays,” *Chemical reviews* **100**(7), 2595–2626 (2000).

100. D. Wilson, S. Garrod, S. Hoyt, S. McKennoch, and K. Booksh, “Array optimization and preprocessing techniques for chemical sensing microsystems,” *Sensors Update* **10**(1), 77–106 (2002).

101. B. Snopok and I. Kruglenko, “Multisensor systems for chemical analysis: state-of-the-art in Electronic Nose technology and new trends in machine olfaction,” *Thin Solid Films* **418**(1), 21–41 (2002).

102. R. Gutierrez-Osuna, “Pattern analysis for machine olfaction: a review,” *IEEE Sensors journal* **2**(3), 189–202 (2002).

103. J. R. Stetter and W. R. Penrose, “Understanding chemical sensors and chemical sensor arrays (electronic noses): Past, present, and future,” *Sensors update* **10**(1), 189–229 (2002).

104. S. Ampuero and J. Bosset, “The electronic nose applied to dairy products: a review,” *Sensors and Actuators B: Chemical* **94**(1), 1–12 (2003).

105. K. Arshak, E. Moore, G. Lyons, J. Harris, and S. Clifford, “A review of gas sensors employed in electronic nose applications,” *Sensor review* (2004).

106. S. M. Scott, D. James, and Z. Ali, “Data analysis for electronic nose systems,” *Microchimica Acta* **156**(3-4), 183–207 (2006).

107. P. Ciosek and W. Wróblewski, “Sensor arrays for liquid sensing–electronic tongue systems,” *Analyst* **132**(10), 963–978 (2007).

108. F. Röck, N. Barsan, and U. Weimar, “Electronic nose: current status and future trends,” *Chemical reviews* **108**(2), 705–725 (2008).

109. M. Peris and L. Escuder-Gilabert, “A 21st century technique for food control: Electronic noses,” *Analytica chimica acta* **638**(1), 1–15 (2009).

110. A. Berna, “Metal oxide sensors for electronic noses and their application to food analysis,” *Sensors* **10**(4), 3882–3910 (2010).

111. O. R. Miranda, B. Creran, and V. M. Rotello, “Array-based sensing with nanoparticles:‘chemical noses’ for sensing biomolecules and cell surfaces,” *Current opinion in chemical biology* **14**(6), 728–736 (2010).

112. A. D. Wilson and M. Baietto, “Advances in electronic-nose technologies developed for biomedical applications,” *Sensors* **11**(1), 1105–1176 (2011).

113. M. Del Valle, “Sensor arrays and electronic tongue systems,” *International Journal of Electrochemistry* **2012** (2012).

114. H. Smyth and D. Cozzolino, “Instrumental methods (spectroscopy, electronic nose, and tongue) as tools to predict taste and aroma in beverages: advantages and limitations,” *Chemical reviews* **113**(3), 1429–1440 (2013).

115. A. Vergara and E. Llobet, “Sensor selection and chemo-sensory optimization: toward an adaptable chemo-sensory system,” *Frontiers in Neuroengineering* **4**, 19 (2012).

116. S. Marco and A. Gutierrez-Galvez, “Signal and data processing for machine olfaction and chemical sensing: A review,” *IEEE Sensors Journal* **12**(11), 3189–3214 (2012).

117. N. S. Ramgir, “Electronic nose based on nanomaterials: Issues, challenges, and prospects,” *International Scholarly Research Notices* **2013** (2013).

118. A. D. Wilson, “Diverse applications of electronic-nose technologies in agriculture and forestry,” *Sensors* **13**(2), 2295–2348 (2013).

119. T. G. Mayerhöfer, A. V. Pipa, and J. Popp, “Beer’s Law-Why Integrated Absorbance Depends Linearly on Concentration,” *ChemPhysChem* **20**(21), 2748–2753 (2019).

120. D. L. Ellis, M. R. Zakin, L. S. Bernstein, and M. F. Rubner, “Conductive polymer films as ultrasensitive chemical sensors for hydrazine and monomethylhydrazine vapor,” *Analytical chemistry* **68**(5), 817–822 (1996).

121. H. Bai and G. Shi, “Gas sensors based on conducting polymers,” *Sensors* **7**(3), 267–307 (2007).

122. Z. Hua, Y. Li, Y. Zeng, and Y. Wu, “A theoretical investigation of the power-law response of metal oxide semiconductor gas sensors I: Schottky barrier control,” *Sensors and Actuators B: Chemical* **255**, 1911–1919 (2018).

123. V. Gaman, “Physical principles of operation of oxidizing gas sensors based on metal oxide semiconductors,” *Russian Physics Journal* **54**(12), 1364–1371 (2012).

124. O. Anisimov, V. Gaman, N. Maksimova, S. Mazalov, and E. Chernikov, “Electrical and gas-sensitive properties of a resistive thin-film sensor based on tin dioxide,” *Semiconductors* **40**(6), 704–709 (2006).

125. J. P. Rospars, P. Lánskỳ, P. Duchamp-Viret, and A. Duchamp, “Spiking frequency versus odorant concentration in olfactory receptor neurons,” *Biosystems* **58**(1-3), 133–141 (2000).

126. M. Wachowiak and L. B. Cohen, “Representation of odorants by receptor neuron input to the mouse olfactory bulb,” *Neuron* **32**(4), 723–735 (2001).

127. G. Reddy, J. D. Zak, M. Vergassola, and V. N. Murthy, “Antagonism in olfactory receptor neurons and its implications for the perception of odor mixtures,” *Elife* **7**, e34958 (2018).

128. V. Singh, N. R. Murphy, V. Balasubramanian, and J. D. Mainland, “Competitive binding predicts nonlinear responses of olfactory receptors to complex mixtures,” *Proceedings of the National Academy of Sciences* **116**(19), 9598–9603 (2019).

129. A. Hassibi, C. Contag, M. O. Vlad, M. Hafezi, T. H. Lee, R. W. Davis, and N. Pourmand, “Bioluminescence regenerative cycle (BRC) system: theoretical considerations for nucleic acid quantification assays,” *Biophysical chemistry* **116**(3), 175–185 (2005).

130. P. Billingsley, *Convergence of probability measures* (John Wiley & Sons, 2013).

131. S. Joshi and S. Boyd, "Sensor selection via convex optimization," *IEEE Transactions on Signal Processing* **57**(2), 451–462 (2009).

132. C. Arndt, *Information measures: information and its description in science and engineering* (Springer Science & Business Media, 2012).

133. B. R. Frieden, *Science from Fisher information: a unification* (Cambridge University Press, 2004).

134. J. W. Gardner and P. N. Bartlett, "Performance definition and standardization of electronic noses," *Sensors and Actuators B: Chemical* **33**(1), 60–67 (1996).

135. D. C. Montgomery, *Design and analysis of experiments* (John Wiley & Sons, 2008).

136. L. J. Halliwell, "The lognormal random multivariate," Proceedings of the Casualty Actuarial Society E-Forum, Spring, volume 5.

137. E. Limpert, W. A. Stahel, and M. Abbt, "Log-normal distributions across the sciences: Keys and clues," *BioScience* **51**(5), 341–352 (2001).

138. R. B. Bergmann and A. Bill, "On the origin of logarithmic-normal distributions: An analytical derivation, and its application to nucleation and growth processes," *Journal of Crystal Growth* **310**(13), 3135–3138 (2008).

139. C. G. Enke and L. J. Nagels, "Undetected components in natural mixtures: how many? What concentrations? Do they account for chemical noise? What is needed to detect them?," *Analytical chemistry* **83**(7), 2539–2546 (2011).

140. W. R. Ott, "A physical explanation of the lognormality of pollutant concentrations," *Journal of the Air & Waste Management Association* **40**(10), 1378–1383 (1990).

141. A. S. Kao and S. K. Friedlander, "Frequency distributions of PM10 chemical components and their sources," *Environmental science & technology* **29**(1), 19–28 (1995).

142. A. McNaught, A. Wilkinson, I. U. of Pure, and A. Chemistry, *Compendium of Chemical Terminology: IUPAC Recommendations*, IUPAC Chemical Data Series (Blackwell Science, 1997), ISBN 9780865426849. URL `https://books.google.com/books?id=dO5qQgAACAAJ`.

143. J. Yerushalmy, "Statistical problems in assessing methods of medical diagnosis, with special reference to X-ray techniques," *Public Health Reports (1896-1970)* pp. 1432–1449 (1947).

144. S. Kullback, *Information Theory and Statistics*, A Wiley publication in mathematical statistics (Dover Publications, 1997), ISBN 9780486696843. URL `https://books.google.com/books?id=luHcCgAAQBAJ`.

145. K. J. Johnson and S. L. Rose-Pehrsson, "Sensor Array Design for Complex Sensing Tasks," *Annual Review of Analytical Chemistry* **8**(1), 287–310 (2015), doi:10.1146/annurev-anchem-062011-143205. URL `http://dx.doi.org/10.1146/annurev-anchem-062011-143205`, PMID: 26132346.

146. D. Zachariah and P. Stoica, "Cramer-Rao Bound Analog of Bayes' Rule [Lecture Notes]," *IEEE Signal Processing Magazine* **32**(2), 164–168 (2015).

147. A. Lorber, "Error propagation and figures of merit for quantification by solving matrix equations," *Analytical chemistry* **58**(6), 1167–1172 (1986).

148. J. Ferré and N. K. M. Faber, "Net analyte signal calculation for multivariate calibration," *Chemometrics and Intelligent Laboratory Systems* **69**(1), 123–136 (2003).

149. R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification* (John Wiley & Sons, 2012).

150. D. Skoog, F. Holler, and S. Crouch, *Principles of Instrumental Analysis* (Cengage Learning, 2017), ISBN 9781305577213. URL `https://books.google.com/books?id=n1zEDQAAQBAJ`.

151. B. S. Alsaedi, C. M. McGraw, T. M. Schaerf, and P. W. Dillingham, "Multivariate limit of detection for non-linear sensor arrays," *Chemometrics and Intelligent Laboratory Systems* p. 104016 (2020).

152. E. T. Zellers, J. Park, T. Hsu, and W. A. Groves, "Establishing a limit of recognition for a vapor sensor array," *Analytical chemistry* **70**(19), 4191–4201 (1998).

153. K. E. Kramer, S. L. Rose-Pehrsson, K. J. Johnson, and C. P. Minor, "Hybrid arrays for chemical sensing," in *Computational Methods for Sensor Material Selection*, pp. 265–298 (Springer, 2009).

154. M. Pardo and G. Sberveglieri, "Classification of electronic nose data with support vector machines," *Sensors and Actuators B: Chemical* **107**(2), 730–737 (2005).

155. R. E. Shaffer, S. L. Rose-Pehrsson, and R. A. McGill, "A comparison study of chemical sensor array pattern recognition algorithms," *Analytica Chimica Acta* **384**(3), 305–317 (1999).

156. S. Joshi and S. Boyd, "Sensor Selection via Convex Optimization," *Signal Processing, IEEE Transactions on* **57**(2), 451–462 (Feb 2009), ISSN 1053-587X, doi:10.1109/TSP.2008.2007095.

157. K. J. Johnson and S. L. Rose-Pehrsson, "Sensing Strategies for Complex Environments," *Annual Review of Analytical Chemistry* **8**(1) (2015).

158. S. D. Wilke and C. W. Eurich, "Representational accuracy of stochastic neural populations," *Neural Computation* **14**(1), 155–189 (2002).

159. W. M. Brown and A. Bäcker, "Optimal neuronal tuning for finite stimulus spaces," *Neural computation* **18**(7), 1511–1526 (2006).

160. R. F. Nalewajski, "Use of Fisher information in quantum chemistry," *International Journal of Quantum Chemistry* **108**(12), 2230–2252 (2008).

161. A. Emery and A. V. Nenarokomov, "Optimal experiment design," *Measurement Science and Technology* **9**(6), 864 (1998).

162. S. Boyd and L. Vandenberghe, *Convex optimization* (Cambridge university press, 2004).

163. F. Kai-Tai and Z. Yao-Ting, *Generalized multivariate analysis* (Science Press, 1990).

164. O. Besson and Y. I. Abramovich, "On the Fisher information matrix for multivariate elliptically contoured distributions," *IEEE Signal Processing Letters* **20**(11), 1130–1133 (2013).

165. B. M. Dobratz, "Properties of chemical explosives and explosive simulants," comp. and ed.; California Univ., Livermore (USA). Lawrence Livermore Lab., 1972.

166. "Canine Training Aids Sigma Pseudo Scents," `http://www.sigmaaldrich.com/analytical-chromatography/analytical-products.html?TablePage=9620026`. Accessed: 2017-07-31.

167. "ScentLogix K9 Scent Detection Training Aids," `http://scentlogix.com/s/?page_id=137#K9ScentDetectionTrainingAids`. Accessed: 2017-07-31.

168. R. Kaiser, A. Kulczyk, D. Rich, R.J. Willey, J. Minicucci, and B. MacIver, "Effect of pore size distribution of commercial activated carbon fabrics on the adsorption of CWA simulants from the liquid phase," *Industrial & Engineering Chemistry Research* **46**(19), 6126–6132 (2007).

169. S.L. Bartelt-Hunt, D.R. Knappe, and M.A. Barlaz, "A review of chemical warfare agent simulants for the study of environmental behavior," *Critical reviews in environmental science and technology* **38**(2), 112–136 (2008).

170. "Melamine Pet Food Recall of 2007," `https://www.fda.gov/animalveterinary/safetyhealth/recallswithdrawals/ucm129575.htm`. Accessed: 2017-08-01.

171. "Latest test results on melamine," `http://www.cfs.gov.hk/english/whatsnew/whatsnew_fstr/whatsnew_fstr_Test_results_of_dairy_product_samples.html`. Accessed: 2017-08-01.

172. "Counterfeit Medicine," `https://www.fda.gov/drugs/resourcesforyou/consumers/buyingusingmedicinesafely/counterfeitmedicine/`. Accessed: 2017-08-01.

173. C.M. Schrader, D.L. Rickman, C.A. Mclemore, J.C. Fikes, D.B. Stoeser, S.J. Wentworth, and D.S. McKay, "Lunar regolith characterization for simulant design and evaluation using figure of merit algorithms," Proceedings of the Proc., 47th AIAA Aerospace Sciences Meeting. ISO, volume 690, 2009.

174. N.J. Galan-Freyle, A.M. Figueroa-Navedo, Y.C. Pacheco-Londoño, W. Ortiz-Rivera, L.C. Pacheco-Londoño, and S.P. Hernández-Rivera, "Chemometrics-enhanced fiber optic Raman detection, discrimination and quantification of chemical agents simulants concealed in commercial bottles," *Analytical Chemistry Research* **2**, 15–22 (2014).

175. C. Cachin, "An information-theoretic model for steganography," *Information and Computation* **192**(1), 41–56 (2004).

176. R. Blahut, *Principles and Practice of Information Theory*, Addison-Wesley series in electrical and computer engineering (Addison-Wesley Publishing Company, 1987), ISBN 9780201107098. URL `https://books.google.com/books?id=qotQAAAAMAAJ`.

177. J. Lavoie, S. Srinivasan, and R. Nagarajan, "Using cheminformatics to find simulants for chemical warfare agents," *Journal of Hazardous Materials* **194**, 85 – 91 (2011), ISSN 0304-3894.

178. D. Adebimpe, "Methods of Producing Pseudoscent Compositions of Narcotic Materials and Compositions Thereof," Oct. 23 2014. URL `http://www.google.com/patents/US20140311218`, U.S. Patent App. 14/217,431.

179. S. Stadler, P.H. Stefanuto, J.D. Byer, M. Brokl, S. Forbes, and J.F. Focant, "Analysis of synthetic canine training aids by comprehensive two-dimensional gas chromatography–time of flight mass spectrometry," *Journal of chromatography A* **1255**, 202–206 (2012).

180. R. J. Harper, J. R. Almirall, and K. G. Furton, "Identification of dominant odor chemicals emanating from explosives for use in developing optimal training aid combinations and mimics for canine detection," *Talanta* **67**(2), 313–327 (2005).

181. J. McKenna, E. S. Dhummakupt, T. Connell, P. S. Demond, D. B. Miller, J. M. Nilles, N. E. Manicke, and T. Glaros, "Detection of chemical warfare agent simulants and hydrolysis products in biological samples by paper spray mass spectrometry," *Analyst* **142**(9), 1442–1451 (2017).

182. D. Tušek, D. Ašperger, I. Bačić, L. Ćurković, and J. Macan, "Environmentally acceptable sorbents of chemical warfare agent simulants," *Journal of materials science* **52**(5), 2591–2604 (2017).

183. D. Matatagui, J. Martí, M. Fernández, J. Fontecha, J. Gutiérrez, I. Gràcia, C. Cané, and M. Horrillo, "Chemical warfare agents simulants detection with an optimized SAW sensor array," *Sensors and Actuators B: Chemical* **154**(2), 199–205 (2011).

184. T. James, S. Wyke, T. Marczylo, S. Collins, T. Gaulton, K. Foxall, R. Amlôt, and R. Duarte-Davidson, "Chemical warfare agent simulants for human volunteer trials of emergency decontamination: A systematic review," *Journal of Applied Toxicology* **38**(1), 113–121 (2018).

185. L. A. Taylor, C. M. Pieters, and D. Britt, "Evaluations of lunar regolith simulants," *Planetary and Space Science* **126**, 1–7 (2016).

186. H. Kanamori, S. Udagawa, T. Yoshida, S. Matsumoto, and K. Takagi, "Properties of lunar soil simulant manufactured in Japan," in *Space 98*, pp. 462–468, 1998.

187. Y. Zheng, S. Wang, Z. Ouyang, Y. Zou, J. Liu, C. Li, X. Li, and J. Feng, "CAS-1 lunar soil simulant," *Advances in Space Research* **43**(3), 448–454 (2009).

188. Y. Li, J. Liu, and Z. Yue, "NAO-1: Lunar highland soil simulant developed in China," *Journal of Aerospace Engineering* **22**(1), 53–57 (2009).

189. D. S. McKay, J. L. Carter, W. W. Boles, C. C. Allen, and J. H. Allton, "JSC-1: A new lunar soil simulant," *Engineering, construction, and operations in space IV* **2**, 857–866 (1994).

190. L. Sibille, P. Carpenter, R. Schlagheck, and R. A. French, "Lunar regolith simulant materials: recommendations for standardization, production, and usage (2006).

191. G. J. Simmons, "The prisoners' problem and the subliminal channel," Proceedings of the Advances in Cryptology (Springer), 1984, pp. 51–67.

192. J. Fridrich, *Steganography in Digital Media: Principles, Algorithms, and Applications* (Cambridge University Press, 2010), ISBN 9780521190190. URL `https://books.google.com/books?id=wcAZ-QEthqkC`.

193. T. Fawcett, "An introduction to ROC analysis," *Pattern recognition letters* **27**(8), 861–874 (2006).

194. S. E. Stitzel, M. J. Aernecke, and D. R. Walt, "Artificial Noses," *Annual Review of Biomedical Engineering* **13**(1), 1–25 (2011), doi:10.1146/annurev-bioeng-071910-124633. URL `http://dx.doi.org/10.1146/annurev-bioeng-071910-124633`, PMID: 21417721.

195. A. C. Knapp and K. J. Johnson, "Selectivity Measure for Arrays of Non-Specific Sensors," submitted to Sensors and Actuators B: Chemical (November 2016).

196. E. L. Lehmann and G. Casella, *Theory of point estimation* (Springer Science & Business Media, 2006).

197. M. Stein, A. Mezghani, and J. A. Nossek, "A lower bound for the Fisher information measure," *IEEE Signal Processing Letters* **21**(7), 796–799 (2014).

198. K. Roy, S. Kar, and R. Das, *A Primer on QSAR/QSPR Modeling: Fundamental Concepts*, SpringerBriefs in Molecular Science (Springer International Publishing, 2015), ISBN 9783319172811. URL `https://books.google.com/books?id=FFcSCAAAQBAJ`.

199. T. Pearce, "Computational parallels between the biological olfactory pathway and its analogue 'The Electronic Nose': Part II. Sensor-based machine olfaction," *Biosystems* **41**(2), 69–90 (1997).

# Appendix A

## A.1 Chain Rule Theorem

First consider the selectivity $sel_{\alpha,\beta}$,

$$1 \geq sel_{\alpha,\beta} = \frac{\det J_\alpha^{-1} \det J_\beta^{-1}}{\det J^{-1}} = \frac{\det J}{\det J_\alpha \det J_\beta} > 0 \tag{A1}$$

Given

$$J = \begin{pmatrix} J_\alpha & J_{\alpha\beta} \\ J_{\beta\alpha} & J_\beta \end{pmatrix} \tag{A2}$$

where $J_{\alpha\beta}$ is a matrix, $J_{\alpha\beta} = (J_{\beta\alpha})^T$, and $J^{(n+m)\times(n+m)}$, $J_\alpha^{n\times n}$, and $J_\beta^{m\times m}$ are positive definite matrices. It may be shown that $\det J_\alpha \cdot \det J_\beta \geq \det J > 0$ and hence $1 \geq \frac{\det J}{\det J_\alpha \cdot \det J_\beta} > 0$.

First consider the matrix equation

$$\begin{pmatrix} J_\alpha & J_{\alpha\beta} \\ J_{\beta\alpha} & J_\beta \end{pmatrix} \begin{pmatrix} I_n & 0 \\ -J_\beta{}^{-1} J_{\beta\alpha} & I_m \end{pmatrix} = \begin{pmatrix} J_\alpha - J_{\alpha\beta} J_\beta{}^{-1} J_{\beta\alpha} & 0 \\ J_{\beta\alpha} & J_\beta \end{pmatrix} \tag{A3}$$

Taking the determinant of both sides using the Leibniz block matrix determinant formula yields

$$\det J = \det J_\beta \cdot \det(J_\alpha - J_{\alpha\beta} J_\beta{}^{-1} J_{\beta\alpha}) \tag{A4}$$

Since

$$\begin{aligned} J_{\alpha\beta} J_\beta{}^{-1} J_{\beta\alpha} &= J_{\alpha\beta} J_\beta{}^{-1/2} J_\beta{}^{-1/2} J_{\beta\alpha} \\ &= J_{\alpha\beta} J_\beta{}^{-1/2} (J_{\alpha\beta} J_\beta{}^{-1/2})^T \\ &= |J_{\alpha\beta} J_\beta{}^{-1/2}| \end{aligned} \tag{A5}$$

$J_{\alpha\beta} J_\beta{}^{-1} J_{\beta\alpha}$is a positive semidefinite matrix. Since $J$ and $J_\beta$ are positive definite matrices, $\det J > 0$,$\det J_\beta > 0$, and $\det(J_\alpha - J_{\alpha\beta} J_\beta{}^{-1} J_{\beta\alpha}) > 0$ as well. So, $J_\alpha \succ J_{\alpha\beta} J_\beta{}^{-1} J_{\beta\alpha} \succeq 0$, which implies $\det J_\alpha \geq \det(J_\alpha - J_{\alpha\beta} J_\beta{}^{-1} J_{\beta\alpha})$. Consequently, $\det J_\alpha \cdot \det J_\beta \geq \det J$which proves the starting proposition,

$$1 \geq sel_{\alpha,\beta} = \frac{\det J_\alpha^{-1} \det J_\beta^{-1}}{\det J^{-1}} = \frac{\det J}{\det J_\alpha \det J_\beta} > 0 \tag{A6}$$