Chemometric Methods for the Determination of Volatile Organic Compounds with Microsensor Arrays

by

Chunguang Jin

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Industrial Health)
in The University of Michigan
2008

Doctoral Committee:

Professor Edward T. Zellers, Chair
Professor Michael D. Morris, Cognate
Research Associate Professor Timothy D. Johnson
Professor James H. Vincent

DEDICATION

To

Xiaojing Xin, my loving wife.

ACKNOWLEDGMENTS

This dissertation would not exist without the support from my advisor Prof. Edward Ted Zellers. I am grateful of all the people who have helped me in different ways during my doctoral study from September 2003 to September 2008 in the University of Michigan.

This dissertation was successfully defended on July 2nd, 2008.

I thank you for reading this dissertation.

TABLE OF CONTENTS

LIST OF FIGURES

Figures

LIST OF TABLES

Tables

ABSTRACT

This research concerns the application of chemometric methods to responses obtained from arrays of planar microfabricated chemical sensors designed to respond reversibly to multiple volatile organic compounds (VOC). It addresses critical modeling and data analysis functions needed to guide the design and implementation of novel meso-scale and micro-scale instrumentation incorporating such arrays, which is intended for use in monitoring human exposures to complex VOC mixtures in occupational, residential, and ambient environments. Upstream chromatographic separation facilitates vapor recognition and quantification from the pattern of responses generated by a sensor array, and the work described is predicated on this pretreatment. The first issue addressed relates to the fidelity of the response pattern generated from a chromatographically resolved analyte to that in a calibration library. A statistically rigorous decision rule was developed that accounts for the inherent variability in the array signals and permits assessments of pattern fidelity at a known rate of error. Building on this first study, a more sophisticated and robust approach to pattern fidelity and peak purity was developed on the basis of fixed-size moving window factor analysis. A simulated embedded peak with an area 0.5% of the embedding peak is detected at a peak signal/noise ratio of 20:1. To address problems involving partial overlap of chromatographic peaks, a self-modeling curve resolution method was applied, which entails an alternating least squares algorithm coupled with evolving factor analysis. This method was tested on binary co-elutions to

assess the ability to extract the component response patterns from their mixture responses as a function of random noise, chromatographic resolution, pattern similarity and relative composition. By using this hybrid algorithm on simulated Gaussian peaks, a correlation coefficient as high as 0.98 between resolved and measured pattern is achieved for noisy sensor with a peak signal/noise ratio of 20:1. In a separate series of studies, the advantages of using multi-transducer sensor arrays instead of single-transducer arrays for vapor recognition were examined. Starting with a database of sensitivities to 11 vapors from 15 microsensors, it was shown by Monte Carlo simulation and principal component regression modeling that optimal MT arrays consistently outperform optimal ST arrays of similar size, and that with judiciously selected 5-sensor MT arrays one-third of all possible ternary vapor mixtures are reliably discriminated from their individual components and binary component mixtures, whereas none are reliably determined with any of the ST arrays. Quaternary mixtures could not be analyzed effectively with any of the arrays. Using the same database, the so-called limits of recognition were determined for various mixtures, revealing that, in general, mixtures cannot be recognized at relative concentration ratios exceeding 20:1 between and two components. Collectively, the research reported here has served to help define the limits of performance and interpret the output of microsensor arrays as components of microanalytical systems.

CHAPTER 1

Introduction

1.1 Research Overview

This dissertation focuses on chemometric analysis and modeling of a sensor-array based analytical system designed for determination of volatile organic compounds (VOCs). Potential applications of such analytical system include, but not limited to, indoor-air quality assessments, personal-exposure monitoring, breath analysis, ambient air pollution mapping, and in-situ assessments of the barrier effectiveness of polymeric chemical protective clothing. The research presented in this dissertation is part of a large-scale interdisciplinary project funded through the NSF Engineering Research Center for Wireless Integrated Microsystems (WIMS) that aims to develop a microfabricated gas chromatograph.[1] This analytical system couples chromatographic separation with microsensor array detector. Complex vapor mixture is separated into components prior to detection by a chemical sensor array.[2] To allow for qualitative and quantitative VOC analysis, chemometrics is introduced as an essential part of the disciplines involved in the instrumentation of such an analytical system. This dissertation presents a series of projects that aim to develop chemometric methods to address critical data analysis function needed for both instrument development and field applications. The key issues addressed in this dissertation include the following:

- Towards developing a sensor array of high selectivity to common VOCs in indoor air, the use of two and more types of vapor transducers in the same sensor array, as compared to the single-transducer array, is explored for the enhanced orthogonality among the sensors and improved performance in analysis of individual vapors and simple vapor mixtures. The recognizable concentration range of a vapor mixture also needs to be quantified.
- Since additional information can be obtained for the chemical problems under study by hyphenating chromatography with sensor array detection, numerical method can be developed to judge the purity of an isolated chromatographic peak and the fidelity of its measured response pattern to the reference pattern. Pattern fidelity assessment is necessary when variability of sensor response imposes uncertainty on peak identity assignment by pattern matching. For overlapped peaks of co-eluting components, numerical curve resolution can be used to separate them instead of changing chromatograph conditions. The ability to resolve numerically overlapped peak is crucial when fast gas chromatograph (GC) separation and analysis of complex vapor mixture is needed.

To address these issues, a series of projects were performed that aimed at providing chemometric solutions to uncertainty assessment in vapor identity assignment by pattern matching, performance evaluation of multi-transducer arrays, limit of recognition for simple vapor mixture analysis, and self-modeling curve resolution of overlapped chromatographic peaks. The profound implication of chemometrics to modern analytical instrumentation is explicitly manifested throughout this dissertation.

The contents and structure of this dissertation are described below. Chapter 1 provides the background information of this research. Analytical instruments used for exposure assessment of VOCs and their trends are reviewed. Microsensor, a key contributor to instrument miniaturization, is reviewed with an emphasis on two types of chemical sensors: polymer-coated surface acoustic wave (SAW) device and chemiresistor with gold-thiolate monolayer-protected nanoparticle (MPN) absorptive interfaces, since these are the two major types of microsensors that have been employed in this research. Chemometrics is introduced and its implication to modern analytical science is reviewed. Key aspects of chemometric analyses specific to the sensor-array based chromatography are emphasized.

Chapter 2 describes a novel, yet simple statistical decision rule for chemical identity assignment by pattern matching. Variation in the response pattern caused by the uncertainty in the sensor array measurement is accounted for when vapor identity is assigned by matching measured pattern to its calibrated pattern in a reference library. In case all of the vapors could be chromatographically resolved, the confirmation of vapor identity reduces to one of assessing the fidelity of the measured response pattern to that in the calibration library. In this chapter the goodness of fit of a unknown sample to its calibration set is tested in order to conclude with confidence that the resolved peaks observed with the array are indeed attributable to the vapor expected to elute at the given retention time. This decision rule is based on a principal-component classification model constructed from a calibration data set. The threshold of maximum distance from a subsequent sample to the centroid of the calibration set in multi-dimensional space (i.e., the Mahalanobis distance) is calculated at a certain significance level in terms of the

residual error after projection of the sample to the principal component axes. If a new sample falls within the boundary established by this threshold, its identity will be assigned to the vapor corresponding to that model at a predefined confidence level. Otherwise it is rejected. Application of this decision rule to pattern fidelity assessment in a SAW sensor array stability study is presented. The method presented in Chapter 2 was published in the Journal of Environmental Monitoring.[3]

In Chapter 3, a study of vapor recognition and quantification by polymer-coated multi-transducer (MT) arrays is described. The primary data set consists of experimentally derived sensitivities for 11 organic vapors obtained from 15 microsensors comprising five cantilever, capacitor, and calorimeter devices coated with five different sorptive-polymer films. These are used in Monte Carlo simulations coupled with principal component regression models to assess expected performance. Recognition rates for individual vapors and for vapor mixtures of up to four components are estimated for single-transducer (ST) arrays of up to five sensors and MT arrays of up to 15 sensors. Recognition rates are not significantly improved by including more than five sensors in an MT array for any specific analysis, regardless of difficulty. Optimal MT arrays consistently outperform optimal ST arrays of similar size, and with judiciously selected 5-sensor MT arrays one-third of all possible ternary vapor mixtures are reliably discriminated from their individual components and binary component mixtures, whereas none are reliably determined with any of the ST arrays. Quaternary mixtures could not be analyzed effectively with any of the arrays. A ‘universal’ MT array consisting of eight sensors is defined, which provides the best possible performance for all analytical

scenarios. Accurate quantification is predicted for correctly identified vapors. Chapter 3 was published in Analytical Chemistry.[4]

In Chapter 4, the discrimination of simple vapor mixtures from their components with polymer-coated multi-transducer (MT) arrays as a function of the absolute and relative concentrations of those components is explored. The data set consists of calibrated responses to 11 organic vapors from arrays of five or eight microsensors culled from a group of five cantilever, five capacitor, and five calorimeter transducers coated with one of five different sorptive-polymer films. Monte Carlo methods are applied to simulate error-enhanced composite responses to all possible binary and ternary mixtures of the 11 vapors, and principal component regression models are established for estimating expected rates of recognition as a function of mixture composition. The limit of recognition (LOR), defined as the maximum recognizable mixture composition range, is used as the metric of performance. With the optimal 8-sensor MT array, 19 binary and three ternary mixtures could be discriminated from their components with < 5% error. The binary-mixture LORs are shown to decrease with increases in the baseline noise levels and random sensitivity variations of the sensors, as well as the similarity of the vapors. Importantly, most of the binary LOR contours are significantly asymmetric with respect to composition, and none of the mixtures could be recognized with < 5% error at component relative concentration ratios exceeding 20:1. Discrimination of ternary mixtures from their components and binary sub-component mixtures is possible only if the relative concentration ratio between any two of the components is <5:1. In comparing binary LORs for the best 5-sensor single-transducer (ST) array to those of the best 5-sensor MT array, the latter were larger in nearly all cases. The implications of these

results are considered in the context of using such as arrays as detectors in analytical systems with upstream chromatographic modules. Chapter 4 is an expansion of the paper we recently submitted to Analytical Chemistry for publication.[5]

In Chapter 5, a hybrid self-modeling curve resolution method that incorporates an iterative alternating least squares (ALS) algorithm into non-iterative evolving factor analysis (EFA) is developed for deconvolution of chromatographic peaks measured with a miniaturized gas chromatograph equipped with a chemiresistor array detector. The advantage of this hybrid algorithm over EFA is illustrated in terms of the quality of recovered response pattern that is measured by the correlation coefficient r between estimated and measured response patterns. Response patterns estimated by EFA are further refined in ALS through an iterative procedure to produce a better resolution of co-eluting components. Based on sensor sensitivities measured for eight VOCs, pairs of overlapped Gaussian peaks are simulated to explore the capacity limits of EFA-ALS in peak deconvolution as a function of chromatographic resolution R, response pattern similarity ρ, signal-to-noise ratio S/N, and quantitative composition. For the vapor pairs with dissimilar patterns (ρ is close to 0), the quality of resolved pattern deteriorates at R<0.4; for vapor pairs with similar patterns (ρ is close to 1), the quality of resolved pattern deteriorates at R<0.8. Poor curve resolution due to pattern similarity is compensated by good chromatographic separation. At S/N=100:1, satisfactory curve resolution (r>0.95) requires pattern similarity ρ<0.7. The individual response patterns derived from EFA-ALS are dominated by the component that has a higher concentration. Peak purity assessment by Fixed-Sized Moving Window EFA is illustrated with an example. The EFA-based curve resolution methods and their applications to chemical

sensor signal resolution are discussed in the context of developing an analytical system for fast and in-situ VOC analysis.

Conclusions obtained from these chapters, along with the most important findings and their implications for VOC monitoring with sensor-array based GC, are summarized in Chapter 6. Future work is also proposed in order to fully explore the potential of chemometrics towards the field implantation of sensor-array based GC.

1.2 VOC Exposure Measurement

Although the precise definition of volatile organic compounds (VOCs) is codified by regulators such as the US Environmental Protection Agency (EPA) and the European Parliament, such definitions are more a matter of policy than a matter of science. For example, the US EPA defines VOCs as "any compound of carbon, excluding carbon monoxide, carbon dioxide, carbonic acid, metallic carbides or carbonates, and ammonium carbonate, which participates in atmospheric photochemical reactions," but also includes a list of dozens of exceptions for compounds "determined to have negligible photochemical reactivity".[6] Under European law, the definition of VOC is based on evaporation into the atmosphere, rather than its reactivity: VOC as any organic compound having an initial boiling point less than or equal to 250°C measured at a standard atmospheric pressure of 101.3 kPa.[7] World Health Organization has a similar definition: organic compounds includes all organic compounds (substances made up of predominantly carbon and hydrogen) with boiling temperatures below 250-260°C, excluding pesticides.[8] Generally it is well accepted that VOCs are a class of hydrocarbons with a vapor pressure greater than 0.1 mmHg at 25°C.[8] This means that VOCs are present in vapor or gas phase in normal ambient temperatures. VOCs include

alkanes (or paraffins), alkenes (or olefins), saturated and unsaturated alkyl halides, carbonyls, alcohols, aromatic and halogenated aromatic hydrocarbons.[8] VOCs are often present as complex mixtures of these different organic compounds in outdoor or indoor environments.

1.2.1 Exposure Assessment. In outdoor ambient air, VOCs are released into atmosphere from both biogenic (mainly vegetation) and anthropogenic sources (e.g, vehicle emissions, the manufacture and use of petroleum products, biomass burning, landfills and sewage treatment plants) where they may undergo various chemical degradation processes.[9] Biogenic emissions account for over 90% of total VOCs entering the atmosphere.[10] Vehicle emission, both exhaust and evaporative, accounts for the majority of VOCs in urban areas.[11] Ambient air levels of VOCs are monitored mainly because of their role in observed increases in levels of tropospheric (ground-level) ozone and decreases in levels of stratospheric ozone which deteriorates ecosystems and can harm lung function and irritate the respiratory system. Methane is known as a very effective greenhouse gas that contributes to global warming.

However, epidemiology studies that correlate ambient air VOCs to morbidity and mortality are mostly not convincing because of the low level of VOCs in ambient air and the lack of reliable exposure data. Instead, adverse health effects of VOCs supported by epidemiologic evidence in occupational populations have been well established.

Indoor air pollution is a compelling environmental issue in which VOCs plays a significant role. Formaldehyde, an irritant in adhesives used widely in furniture manufacturing, is the source of many indoor air quality complaints. Releases of VOCs from building materials, carpets, paints, furnishings, photocopiers, cosmetics, household

chemicals, as well as tobacco smoke are related to sick building syndromes.[12] Indoor levels of VOCs may reach 1,000 times that of the outside air due to certain activities and low air exchange rate.[13]

The highest VOCs levels have been observed in industrial settings. Manufacture of chemicals and pharmaceuticals releases man-made chemicals into the atmosphere. Oils, gasoline, industrial solvents, paints, and dyes are the major sources of VOCs in the workplace. High temperature and improper manufacturing or operational practice causes rapid evaporation of liquid organics into air. Lack of efficient administrative or engineering pollution control brings about amassment of high level of VOCs in the workplace within a short time. For industrial hygiene purposes, VOCs and vapors are used interchangeably in this dissertation (If the substance is normally a liquid (or solid) at normal temperature and pressure, then the gaseous component in equilibrium with its liquid (or solid) state is called a vapor. Carbon tetrachloride, formaldehyde, and benzene are examples of compounds that are present in the vapor state).

People are exposed to VOCs by inhaling the contaminated air, which is the dominant route of VOC exposure. Other potential exposure route may include ingestion through contact with VOC-enriched airborne particles. The health effects depend on the specific composition of the VOCs, the concentration, and the frequency and length of exposure. High concentrations of some compounds could have serious health effects. General effects include eye, nose and throat irritation, headaches, loss of coordination, nausea, damage to the liver, kidneys and central nervous system and some are suspected or known to cause cancer in humans.[14] While the toxicology of many VOCs is relatively well understood when they are considered individually or in very simple mixtures, very

little is known about the toxic effects of complex mixtures. The Occupational Safety and Health Association (OSHA) has formulated permissible exposure limit (PEL) and immediately dangerous to life of health (IDLH) for separate organic compounds but not for total VOC exposure. The American Conference of Governmental Industrial Hygienists (ACGIH) has established Threshold Limit Value-Time Weighted Average (TLV-TWA) values and Threshold Limit Value, Short Term Exposure Limit (STEL) for single compounds but does not take into account the effects of simultaneous or serial exposure to complex mixtures and has not established values for total VOC exposure. The World Health Organization has not established guidelines for total VOC exposure. A lot of research has been done with regard to the toxicity of some specific individual VOCs, whereby the dose-effect relationships have been established in occupational settings. However the health risk of complex VOC mixtures, the prevailing form of presence of VOCs in the environment, is not clear. Interest in VOCs in air has evolved from a focus on relatively high levels of a limited number of compounds known to be important health risk factors, to the analysis of complex mixtures of VOCs at increasingly lower levels.[15] Therefore exposure assessment of *complex mixture* of VOCs at ppb (part per billion) or sub-ppb level is necessary, especially in epidemiologic study of health effects of VOCs in ambient air.

Like other air pollutants, one of the most significant characteristics of VOC exposure is the great variability of the concentration levels observed in both occupational and non-occupational (environmental) settings. The concentration of a VOC in the space of a workplace varies with time over both short and long periods. Moreover, workers move in varying patterns through an environment where the VOC concentration varies

with location, and the actions of the workers themselves may cause the concentration to vary. All of these sources of variability lead to an exposure distribution which is usually best described statistically by the logarithmic normal distribution and that typically has geometric standard deviations from two to five or more.[16] This variability has to be taken into consideration of estimating the exposure of a group of workers having differing exposures to find the most exposed workers, or forming similarly exposed worker group by assigning the same exposure level to a group of workers who have similar work patterns. Such variability is even more notable in exposure assessment of VOC in ambient air. It was reported that a typical environmental study would record exposure varying by about 100-fold from one subject to another, while a typical occupational study would find only about 15-fold variation.[17] The great exposure variability entails a careful planned sampling program for VOC measurement. Implications of the exposure variation can be summarized below for VOC measurement:

- Compared to this environmental variability, the variability introduced by the sampling and analytical error is small. But it does not dispense with accurate exposure measurement. Interpretation of measurement data must rely on the context of sampling such as the location, time, subject, duration of sampling, as well as worker practice and if any exposure control is in place. These observations must be combined with measurement data to understand how to interpret the analysis.
- Complex vapor mixture should be sampled and analyzed for its components. Risk analysis of VOCs exposure is only a fair approximation without detailed information about the composition of VOC mixtures and its concentration.

- Ability to measure the real-time concentration becomes essential. To estimate the exposure profile for a worker during an eight-hour workday, it is necessary to employ an analytical method that is capable of monitor the concentration variation during the work, especially for the VOC that has short term exposure limit due to its acute toxicity.
- The analytical method should be able to provide a wide response range that is linear to the concentration of VOC of interest. Ideally the analytical method is capable of measuring a concentration from as low as ppb level to as high as percentage level.
- Personal monitor can provide a much more accurate exposure assessment than area monitoring. Personal exposure data should be collected to supplement the area measurement data. Combination of personal exposure data with personal health information provides a reliable approach to the epidemiologic efforts trying to establish dose-effect relationship. Unfortunately such personal exposure data are not available in most epidemiologic studies.[18] To avoid intruding the worker's practice, such personal monitor should be small and easy to wear.

1.2.2 VOC Measurement. In industrial hygiene VOC measurement usually involves sample enrichment on an appropriate adsorbent like charcoal, followed by solvent or thermal desorption with subsequent measurement by gas chromatography. [19] From the earliest days of industrial hygiene, breathing zone ambient air was collected using personal air samplers carried by workers. Sampling and analytical methods used

can be found in the U.S. National Institutes of Occupational Safety and Health's Manual of Analytical Methods.[20] In the case of VOCs, the most commonly used method continues to be adsorption of contaminated air onto activated charcoal tubes clipped to the worker's collar and connected to a small pump carried at the waist. This active sampling system provides a constant flow and known volume of air, containing typically ppm levels of toxic contaminants, to the sampler during an 8-hour workday. In the early 1980s, passive badges employing activated charcoal were developed for use in occupational sampling. The badges operate on the principle of diffusion and are often operated over an 8-hour workday to provide an integrated average exposure for comparison to the occupational standards (e.g, the threshold limit value). Although such newer sampling methods have since been adopted by NIOSH and the U.S. Occupational Safety and Health Administration (OSHA), the fundamental sampling strategy of collecting breathing zone samples was sound and forms the basis of industrial hygiene practice today.[21] Sampling using activated charcoal tubes is still a mainstay in industrial hygiene sampling of VOCs and the dominant analytical method is still gas chromatography usually equipped with a flame ionization detector.[20] Obviously such active or passive sampling methods are not suitable for accurate measurement of short-term VOC concentrations, not to mention on-site real-time VOC measurement.

The same "VOC trap in field then GC analysis in lab" strategy is also used in ambient VOC measurement, but the sampling methodologies developed for industrial hygiene are not suitable for the environmental monitoring of ambient air, where exposure levels were generally three to four orders of magnitude lower than those found in the workplace. During 1970s an alternative sorbent Tenax® (poly-2,6-diphenylphenylene

oxide) was widely adopted in ambient air sampling of VOCs in the wake of an initiative taken by EPA to develop methodology for collection and analysis of VOCs at ppb levels [22] . Compared to charcoal used in industrial hygiene, Tenax® has low background contamination and is stable at temperatures up to 250ºC, allowing thermal desorption instead of solvent desorption. In the case of thermal desorption, it is sometimes necessary to add an intermediate cryogenic preconcentration step in order to improve the chromatographic separation. Tenax® is now widely used in a multisorbent bed in which Tenax acts as the first sorbent and newer types of activated charcoal (Spherocarb® and Carbosieve®) as the second, or backup, sorbent. Tenax® collects the bulk of the VOCs, and the activated charcoal collects those more volatile VOCs that break through the Tenax®.[23] The EPA developed its own methods capable of achieving much greater sensitivity than those developed by NIOSH. EPA methods include the TO-1, TO-2, and TO-17 protocols for sampling using Tenax, carbon molecular sieve, and multibed tubes, respectively.[24] The Summa™ canister method, which use evacuated Summa canister to directly sample whole air and enrich VOCs by a sryogenic preconcentrator,[25,26] was adopted by the EPA in 1988 as the EPA TO-14 method (revised in 1997 as the TO-14A method) targeting 40 VOCs, and, in 1999, as the TO-15 method targeting 97 VOCs. [27,28,29]

In occupational and environmental studies, VOC samples are usually analyzed by first separating the mixture into individual components using GC. Three GC detection methods in common use are flame ionization (FID), electron capture, and mass spectrometry (MS). Only GC/MS has the ability to unambiguously identify many chemicals. Neither GC/FID nor GC/electron capture detection is able to separate

chemicals that coelute (emerge from the chromatographic column at the same time). Also, GC/FID response is depressed by chlorine and other halogens, so it is not suitable for samples containing halogens. MS, by breaking chemicals into fragments and then identifying these fragments, is often capable of differentiating between coeluting chemicals. However, since chemicals are identified by comparing these mass fragment spectra to existing libraries, and the libraries are incomplete or inaccurate, GC/MS identifications are often tentative or mistaken. Regardless, due to EPA-mandated use of mass spectrometric detection for the TO-15 method, there has been an increase of adoption of GC/MS for VOCs analysis.

Because of the complexity of indoor or outdoor air samples, and the inaccuracy of VOC measurement, an alternative to chemical measurement methods has arisen in recent years—use of a trained panel to judge the possible health or comfort effects of an air sample directly. Human nose is very sensitive to odorous VOC. The presence of odor usually indicates the exceeding of VOC level over threshold limits. This olfactory analysis method, pioneered by Fänger of Denmark,[30] employs panels of 6–10 people who have been previously trained by sampling known mixtures of odorous compounds. When exposed to a test atmosphere, the judges provide an instantaneous estimate of its pollution potential, measured in units called decipols. One decipol is equivalent to the amount of pollution (body odor) produced by one person in a room ventilated at one air change per hour. The method is capable of both predicting how people will react to air of a given quality and estimating the relative contribution of various sources (e.g, ventilation system, office machines, employees) to indoor air quality.

However, reliance on human nose for VOC detection is problematic. Odor itself is not a good indication if something is hazardous. Some dangerous chemicals have no odor (e.g, carbon monoxide) or what some would consider a pleasant odor (vinyl chloride), while some safe substances may have a very offensive odor.[31] Human nose can only smell a substance beyond a certain concentration level. Below this odor threshold VOC can no longer be perceived. Even for a concentration beyond the odor threshold, human nose can easily come into a state of "olfactory fatigue":[32] human nose comes to a point that it is no longer able to smell the odor while the odor still maintains a high concentration level. Due to these limitations a lot of studies have been performed trying to use chemical sensors to replace human nose for odor detection. A class of analytical device called "electronic noise" has emerged. An excellent review of electronic noise was recently given by Weimar et al. [33] By analogy to human noise which perceives complex odor composed of multiple odorant molecules by activating several odorant receptors simultaneously, electronic noise device uses chemical gas sensor array to generate the "flavor fingerprint" that is characteristic of the odor "smelled".[34] The application of chemical sensor and sensor array for VOC detection will be reviewed below.

1.3 Advances in VOC Analysis and Analytical system

As mentioned earlier, measurement of VOCs of great variability in the workplace or general environment requires an analytical instrument that has excellent sensitivity and selectivity and also is capable of in-situ real-time measurement of complex VOC mixtures. It sounds like a magic machine as the "tricorder" in the movie Star Trek. Nevertheless, this magic analytical instrument is becoming a reality day by day with the

progress in analytical technology. Advance in sensing material, optical sensor, and microfabrication technology all contributes to the development of an analytical device that is small and has high analytical performance in terms of sensitivity and selectivity.[35] Current status and future trends in VOC measurement methods will be reviewed here.

There are a vast array of sampling and analytical methods available for qualitative and quantitative analysis of VOCs in air, According to Climent et al,[36] the selection of an appropriate method depends on several criteria that are similar to the aforementioned measurement requirements. These criteria include a wide concentration range, high sensitivity and high selectivity, portability, and cost. Methodological developments in the field of sampling and analysis of VOCs in air have been reviewed by Dewulf and Van Langenhove.[37] By far the most commonly used methods for the determination of VOCs are sorbent- (in occupational settings) or canister-based (in environmental settings) sampling followed by laboratory-based capillary gas chromatography using flame ionization (FID), electron capture (ECD), or photoionization detection (PID). Gas chromatography/mass spectrometry (GC/MS) is becoming increasingly common in environmental laboratory due to its high sensitivity when operated in selected ion monitoring (SIM) mode, and its unique ability of chemical structure analysis.[38-41] As VOCs are found in trace concentration in air, and the sensitivity of common analytical methods is insufficient, it is often impossible to measure low concentration directly and a preliminary enrichment step is required.

Gas chromatography used for VOC detection is extensively studied and two excellent review articles exist.[42, 43] Some new progresses in chromatography with regards to VOC detection are noteworthy, among which is the development of multidimensional

GC like GC×GC which use a second column to separate the co-eluting compounds in the first column.[44, 45] As a result the capability of separating complex sample and the resolution of GC column is greatly increased. Besides separation, positive compound identification is made by coupling the GC to a mass spectrometer, usually amass selective detector (MSD), and other multi-channel detectors like spectrophotometer and chemical sensor array. Analysis time is generally 10–60 minutes with conventional capillary columns (i.d. 0.25–0.32 mm, 1–2 mL/min),[46] depending on the complexity of sample and the experimental conditions. Such long analytical time, plus the time used for sampling preparation, makes the traditional GC unsuitable for simultaneous analysis that is required in cases like in-situ exposure profile monitoring or on-line process monitoring. Accordingly, another tendency in modern gas chromatography development is towards shorter separation time and fast GC.

GC instruments developed during the 1990s use shorter, narrower columns and higher carrier gas velocity to achieve less analysis times, on the order of minutes or even milliseconds. The use of "fast GC" (1–3 s), "very fast GC" (30–200 ms), and "ultrafast GC" (5–30 ms) for the analysis of VOCs and semi-volatile VOCs (SVOCs) has also been reviewed.[47] Analysis using fast capillary GC (i.d. 0.1 mm) can give peak widths at half height of 0.2–3 s and separation efficiencies (plate number) equal to or greater than conventional GC but 5–30 times faster.[47] The acquisition rates of scanning mass spectrometers in full scan mode typically range from 10–20 spectra per second. The GC peaks must, therefore, have a width of at least 0.5 s, which is within the research scope of fast GC instruments. The advantages of fast GC for routine analysis are obvious: (1) reduced time needed to obtain analytical results; (2) greater throughput; (3) feasibility of

more replicate analyses resulting in greater precision; (4) lower carrier gas consumption; (5) reduced costs, and (6) applicability to field-portable instruments.[48] In temperature-programmed GC, on the other hand, there is no advantage if the run time is determined by the time it takes for the oven to reach the temperature needed to elute the all components. Very fast GC results in peak widths at half height of 30–200 ms and efficiencies of only 25,000 plates, and is suitable for the analysis of simple mixtures only. Currently, ultrafast GC has no useful application since, although peak widths are 5–30 ms, its efficiency is only about 7000 plates.[47] Future developments that include fast GC coupled with time-of-flight mass spectrometry (TOF-MS) having high-speed acquisition rates (500 scan/s) make it likely to detect ppq (parts-per-quadrillion) level in a routine way in the future.

Miniaturization is one of the most noteworthy features of modern analytical instrumentation.[48] GC is not an exception. The benefits of such miniaturized GC are obvious: portable, saving of material and power without sacrificing analytical performance. One of the efforts towards GC miniaturization is the development of portable GC (transportable, or handheld). Portable GC, as well as the miniaturization of GC instruments has been reviewed by Yashin et al.[49] In that review a classification of portable GC was provided. That is shown in Table 1-1. In most of the portable GC developed so far, bulky gas cylinder was discarded and filtered air is used as carrier gas. FID in traditional GC is replaced with thermal conductivity detector or photoionization detector. Micro fabricated mass spectrometer was also used as the detector thanks to the development of new type of mass spectrometer that can used in normal atmosphere.[50, 51] Therefore both qualitative and quantitative analysis can be provided with such portable

GC/MS system.[49, 52] Generally these portable devices can provide near real-time fast analysis of complex VOC mixtures. However the fast separation is achieved at the cost of analytical performance. Most portable GC can only perform simple analysis in laboratory. Field application is seldom reported.[53-56] In particular, a portable GC device that used a chemical sensor array as detector was developed by Zellers et al.[57-62] This portable GC instrument uses filter ambient air as the carrier gas to avoid the need for on-board gases, and incorporates a multi-stage adsorbent preconcentrator/focuser (PCF), a tandem-column separation module with independent 'at-column' temperature-programming capabilities and mid-point pressure/flow control, and a detector comprising an array of either microfabricated surface acoustic wave gas sensor,[57, 58] or chemiresistor (CR) sensors with nanoparticle interface layers.[63] A schematic of the analytical layout of this portable GC is illustrated in Figure 1-1. This device is capable of near real-time analysis of complex vapor mixture and has been demonstrated for environmental tobacco smoke detection,[63] breath analysis and contraband currency detection.[64]

Cutting edge manufacture technologies have enabled producing silicon chip-based chromatographs and micromachined chromatographs. Microelectromechanical systems (MEMS) technology based on the dielectric properties of pure silicon has matured. Using MEMS technology, one can fabricate miniature analytical devices.[65] Fabrication includes several steps: polishing of a silicon plate, oxidation to SiO_2, deposition of a photoresist, photolithography, etching of silicon oxide, the removal of the photoresist, etching in the silicon microstructure, and others.[66] This technology can be used to create any three dimensional item in silicon. The technology seems to be most promising for manufacturing micromachined gas chromatographs. In 1975 researchers at Stanford

University, who set up the company Microsensor Technology, pioneered the development of the first silicon chip-based miniaturized gas chromatograph based on silicon technology. This chromatograph was equipped with a katharometer (Merriam-Webster definition: an apparatus for determining the composition of a gas mixture by measuring thermal conductivity that has been used to determine the basal metabolic rate by measuring the rate of production of carbon dioxide based on the composition of expired air). It was intended for air analysis in spacecrafts.[67-69]

Based on the work of Terry and Reston and Kolesar et al,[70,71] several reports have also appeared on micromachined chromatographic separation channels.[72-78] At first, such micromachined silicon gas chromatographs operated at room temperature and were used for the analysis of gases and volatiles compounds. One of such silicon gas chromatographs was able to separate a15-component mixture for 10 s at 28°C using a capillary column (200 cm ×0.135 mm) and a DB 1701 film liquid phase.[79, 80] An improved micromachined chromatograph was described.[81] The dimensions of this device are 23 cm^2 with a height of 2.5 mm. A spiral capillary column (0.9 mm in length) was coated with copper phthalocyanine as a stationary phase (film thickness 0.2 mm). Applications of early-stage micromachined gas chromatographs for analyzing volatile pollutants were reported in Reference 82-86 .

Although these pioneering works focused on column miniaturization, they manifest the promising potential of MEMS technology in analytical instrumentation. In particular, Sandia National Laboratories reported an integrated micro GC, μChemLab®, aiming at chemical warfare agent detection.[77] This GC system is a hyphenation of micromachined separation columns with an integrated surface acoustic wave microsensor

array. Notably, a microfabricated preconcentrator was also monolithically integrated into this analytical system.[77, 78] A complete analytical system is being developed by Zellers et al under the name of WIMS MicroGC.[87] This system has been demonstrated for its capability of trapping, separating, identifying and quantifying complex environmental vapor mixtures.[87, 88] The WIMS MicroGC contains the similar components as that found in the portable GC developed by the same group. These components in the WIMS MicroGC have the same analytical function as that in portable GC but of micro scale. These components are shown in Figure 1-2 including: a sample inlet with particulate filter, on-board calibration-vapor source, multi-stage preconcentrator/focuser (mPCF), dual-column separation module with pressure- and temperature-programmed separation tuning, an array of microsensors for analyte recognition and quantification, and pumps and valves to direct sample flow. Several of the key components employed with the WIMS MicroGC have been reported. [88-105] Pressure and temperature sensors are to be included at various locations throughout the system for climate control. The whole system is powered by battery and analytical signal can be transmitted wirelessly so that remote monitoring and control is possible. A built-in microprocessor controls the instrument and processes the data. MEMS technologies are being used to fabricate the system with the ultimate goal of creating a fully operational micro-instrument that occupies only 1–2 cm^3, requires an average power of just a few mW per run, provides rapid determinations of mixtures of at least 30 vapors of arbitrary composition at low- or sub-part-per-billion (ppb) levels.[87] Resolving complex vapor mixture into individual components or simple co-eluting mixtures chromatographically and then coupling pattern

recognition analysis with retention time information numerically represents a powerful analytical system for vapor identification and quantification.

Although miniaturized GC (portable or micromachined) instrument are currently still not comparable to traditional capillary GC, the gap is decreasing. Analytical capacities of miniaturized GC keep improving. It seems the future belongs to the miniaturized GC represented by portable GC and sensor-array based GC.[106]

1.4 Chemical Sensor and Sensor Array

Besides the traditional VOC measurement methods like flame ionization detector, electron capture detector and mass spectrometry, application of chemical sensor to VOC measurement holds more promise for sensitive and selective analysis. Chemical sensor is an active research field, [107, 108] and many microsensor technologies have been developed for analyzing gases and vapors in both a qualitative and quantitative way.[109] A brief review of vapor sensors and sensor arrays is given in this section. The focus will be put on these types of sensors that hold promise for use as detectors in the sensor-array based GC systems. These vapor transducers to be reviewed include thermal sensors (i.e., calorimeter sensor), mass sensor (surface acoustic wave sensor, cantilever sensor), electrochemical sensor (chemiresistor sensor), and optical sensor.[107, 108]

A chemical sensor provides information about the chemical composition of its environment at low cost. Its most common configuration consists of a physical transducer and a chemically selective sensing material that is coated on the surface of the physical transducer.[107] Once analyte (in liquid or gas phase, seldom in solid state) is selectively absorbed into the coating layer/film, the interface material will change in its property

(change in electronic property like current, potential, resistance and capacitance, physical shape, frequency shifts, modulation of light, temperature or produced heat change, and even mass). Such change can be transformed in measurable electronic signal by some form of transduction. Based on the principle of transduction, chemical sensors are divided into four fundamental groups: thermal sensor, mass sensor, electrochemical sensor and optical sensor.

The key feature of these devices for purposes of gas-phase sensing is that measurable characteristics (mass, temperature, resistance, light absorbance) are altered upon sorption of an analyte onto the surface of the device (adsorption), or into the bulk of an applied thin film (absorption). Thus, the sensor response consists of a sorption step, involving phase transfer equilibria and kinetics, and a transduction step, leading to an analytical signal.[110] Absorption of vapor molecules from the gas phase into a sorbent thin film on a sensor device is shown schematically in Figure 1-3. This figure shows the sorption and desorption accompanying interaction between analyte and sensor interface layer. In most sensor applications, the interaction between analyte and interface layer is reversible, i.e., after absorption into the interface layer analyte will be desorbed rapidly once it is removed from air. At stable ambient concentration, the sorption and desorption will quickly reach an equilibrium with the analyte concentration in the interface layer proportional to its ambient concentration.

The magnitude of sensor response at equilibrium reflects the amount of analyte in the ambient air. Such reversible interaction as sorption and desorption can only happen if the interaction does not involve chemical reaction but the relatively weak non-bonding intermolecular force like London dispersion, dipole-dipole, or hydrogen bonding.[111]

Because intermolecular force is universally present and generally not selective in terms of analyte-interface layer selection, chemical sensors suffer from poor selectivity, which remains as the bottleneck of sensor development and application. Reversible interactions between analyte and interface layer is necessary in most applications because when the analyte concentration varies in the sample under investigation, the sensor should faithfully follow such change in the concentration. On the other hand, when the sensor is intended for detection rather than quantification, irreversible interaction between analyte and interface layer is allowed. An example is metal oxide gas sensors which convert analyte into other species through chemical reaction. Since sensor response depends on the irreversible chemical changes, sensitivity of such sensor device can be adjusted by choosing suitable reactive layer material and operation temperature. Due to this limitation metal oxide gas sensor is mainly used in direct-reading instrument for combustible gas (like methane and alcohols) detection and alarming. Typical gases detected by metal oxide sensor (or semiconducting devices) include oxidizable substances such as hydrogen, hydrogen sulfide, carbon monoxide, and alkanes (SnO_2, ZnO, etc.), as well as reducible gases like chlorine, oxygen, and ozone (NiO, CuO).[112-114] One of the earliest metal-oxide chemical sensor, SnO_2 sensors, designed by N. Taguchi and referred to as the "Figaro sensor" or Taguchi sensor make it possible to detect as low as 0.2ppm carbon monoxide (Permissible Exposure Level =50ppm regulated by OSHA). [115]

The sensing material of sensor interface layer on the surface of a chemical sensor determines the maximal sensitivity that can be achieved. Some lists of materials sensitive to detectable gaseous compounds together with the corresponding operating temperatures are provided in Reference 116–120. Judicious choice of sensing material can also

improve partial selectivity of a chemical sensor to the analyte of interest. There are two types of sensing materials extensively reported in literature that are also used in this dissertation. One is organic polymer that has been used extensively as chemical sensing materials.[121] Of special interest is polypyrrole. Chemiresistors coated with polypyrrole have shown high sensitivity towards such gases as nitrogen dioxide,[122] ammonia,[123] and methanol.[124] A wide variety of polymers of this type is available, including substituted polypyrroles, polythiophenes, polyindoles, and polyanilines.[125] These polymers have diverse sensitivities which are leading the way to combining more of these sensors in sensor arrays and using chemometrics to VOC analysis. The major advantage of polymer materials is that they interact reversibly with analyte and they can be easily modified chemically, which favors their use in developing selective sensor array composed of two or more partially selective sensors. The major disadvantage is that most polymers are not thermally stable and only be used at room temperature. One exception is phthalocyanines that constitute one group of organic materials with thermal stability >400ºC.[126, 127] Another type of chemical sensing material comprises nanometer-sized conductive core (carbon or metal like copper and gold) surrounded by organic monomolecular layer (e. g, thiolate monolayer).[128-129] These conductive nanoparticles interact reversibly with analyte and their sensitivity can be adjusted by bonding different chemical functional groups to the ligand coordinated to the conductive core. These materials have been applied to chemiresistor for humidity [130] and VOC detection.[131-133] The transduction principle of such conductive nanoparticle-coated chemical sensor can be explained tentatively as follow: Carbon or metal powder of very small particle size is conductive. The resistance observed depends on the distance and the number of particles

that are in contact with each other. Uptake of water, VOCs and other analyte of interest leads to swelling of the polymer and disturbance of the ohmic contacts between conducting particles, which in turn increases the resistance. For this reason the resistance correlates with the humidity or concentration of VOCs.

As mentioned earlier, chemical sensors for purpose of gas/vapor detection can be classified into four basic categories according to the principle of vapor transduction: mass sensor, thermal sensor, electrometric sensor and optical sensor.[107, 108] Each type comprises a variety of different sensors. In this dissertation no intention is given to the full coverage of the galaxy of chemical sensors. Several key gas/vapor sensors, however, will be reviewed for their principles and potentials to use as detector of sensor-array based GC for VOC analysis.

1.4.1 Surface Acoustic Wave Sensor. The surface acoustic wave (SAW) sensor is one type of mass sensor that is sensitive to mass change of sensor device. The SAW sensor has been employed for measurement of force, acceleration, hydrostatic pressure, electric field strength, dew point, and gas concentration.[134] There have been many reviews on acoustic wave chemical sensors for gas- and liquid-phase sensing.[135-151,158] SAW sensors have also been the subject of at least three books that cover various aspects of SAW sensor devices.[151-153] Grate et al have published a series of studies about acoustic wave device types, operating instrumentation, and the physical mechanisms for their responses,[148-152] the applications of linear solvation energy relationships to chemical sensors and arrays,[153, 156-158] glass transition behaviors,[150] and early studies in acoustic wave biosensors.[152] The rationale of SAW sensor for VOC analysis can be referred to

Figure 1-4 and summarized accordingly.[159] The operating principle of an SAW for use with gases is conceptually quite simple. Like the well-known quartz microbalance (QCM, another type of mass-sensitive device that is commercialized for film thickness measurement. [160]), SAW devices are all based on the piezoelectricity of solids, [161] a phenomenon that permits the generation of vibrations or the propagation of waves. A piezoelectric substrate, usually a quartz plate, is stimulated to vibration (deformations of the crystal) by an oscillating electric circuit. A surface acoustic wave, a periodic deformation perpendicular to the quartz surface, is transmitted across the surface of quartz solid. The most stable frequencies of SAW are the characteristic resonance frequencies. Such a resonance frequency is itself highly dependent on the mass of the oscillating plate. A slight change in the mass of the quartz plate, as a result of analyte sorption into the sensing material coated on the surface of quartz plate, will result in a considerable change in the resonance frequency, which can be measured very accurately as Δf_v indicated below: [148, 149, 163, 164]

$$\Delta f_v = n\Delta f_s C_v K/\rho_s \qquad \text{Eq. 1-1}$$

where the multiplier n is equal to 1 for mass-loading responses. If swelling induced modulus changes increase the responses, then n will be equal to that amplifying factor. The sensor's response to the mass of vapor absorbed, a frequency shift denoted by Δf_v, is dependent on the frequency shift due to the deposition of the film material onto the bare sensor substrate (a measure of the amount of sensing material usually polymer on the

sensor surface), Δf_s, the vapor concentration C_v, the density of the sorbent phase ρ_s, and the partition coefficient K that is defined by Eq. 1-2.

$$K=C_s/C_v \qquad \text{Eq. 1-2}$$

The partition coefficient K quantifies the equilibrium distribution of vapor from the gas phase into a sorbent phase and thus indicates the effect (sorption) of all the interactions between the vapor and the sorbent. The partition coefficient gives the ratio of the concentration of the vapor in the sorbent phase, C_s, to the concentration of the vapor in the gas phase, C_v. The importance of K to the response of SAW sensor [165, 166] and other mass-sensitive chemical sensor [167-169] is well studied.

Like any other mass sensitive device used for chemical detection, SAW sensor must be coated with a layer of appropriate sensing (ideally, chemically selective) material capable of interacting with the proposed vapor. The applications of organic polymers as interface layer of SAW sensors have been explored by Zellers et al for their potential to use with the detector of a portable GC. [170-176]

1.4.2 Chemiresistor. Chemiresistor is one type of electrochemical sensors that responds to analyte by measuring the change of resistance or conductivity of sensing layer. Chemiresistors have been extensively reported (some under the name conductometric sensor) for sensitive measurement of organic or inorganic analyte in liquid phase where the presence of analyte changes the conductivity of electrolyte between two electrodes. There are also numerous reports about measurement of inorganic

analyte in gaseous phase, like H_2O, H_2, O_2, CO, CO_2, H_2S, NH_3, NO_2, Cl_2 and HCL. Refer to the review articles Ref. 107 and 108 for a complete coverage of these reports. This dissertation provides a short review of chemiresistor for purpose of VOC analysis.

Inorganic semiconductor oxides have been traditionally used as sensing materials of chemiresistor. The most commonly used material is tin oxide SnO_2 ("Taguchi sensors"),[177] which has been commercialized for a wide range of industrial and domestic applications, including gas monitors, leak detectors, and alarm systems for toxic or inflammable gases. However, their uses are restricted to the detection of gases that are redox reactive. Many substances cannot be detected using the semiconductor material because they either are too inert (as many halogenated species) or do not react electrochemically at limited potentials. New materials have emerged, including the organic semiconducting polymer mentioned earlier. The greatest disadvantage of organic materials in gas detection is that they are usually very poor conductors, and conductivity measurements would be correspondingly difficult. One of the ways to improve conductivity of organic polymer is to add into the polymer carbon, copper or gold powder that has excellent conductivity. Carbon-doped polymers were reported by Lewis et al for VOC detection.[178] Researchers at Sandia National Laboratory also reported a similar polymer/carbon particle composite. [179-181] Transduction principle of these chemiresistor seems simple. If a polymer/conductive particle composite increases its volume by swelling when absorbing a chemical, the electrical resistance increases due to the breaking of some of the conductive pathways through the film. The expansion can produce large increases in resistance if the polymer volume is changed close to the percolation threshold. This threshold concentration has been found to be between 20 and

40% by volume of the conductive particles.[182-184] Sensitivity and selectivity can be tailored by changing the polymer. A schematic of such chemiresistor is shown in Figure 1-5.

Gold-thiolate monolayer-protected nanoclusters (MPNs) have emerged as new interfacial materials of chemiresistor sensors. Wohltjen and Snow first reported chemiresistor of this type for analysis of toluene, perchloroethylene, 1-propanol, and water vapors.[185] Gold-thiolate MPNs are synthesized by the reduction of Au^{3+} ions in the presence of organic thiols.[186, 187] The resulting structure is a gold core of nano-dimensions, typically 2-10 nm, surrounded by a self-assembled thiolate monolayer (See Figure 1-6). The monolayer provides stabilization, allowing the formation of the thermodynamically unfavorable nano-sized gold cores. By changing the composition of thiolate monolayer, sensitivity and selectivity can be tailored for optimal performance. MPN-coated chemiresistor has been employed as the detector of sensor-array based GC and shows superior sensitivity as compared to other chemical sensors.[187-189] These devices have been shown to provide improved detection limits and simplified system-level and electrical requirements over other, polymer-based sensors. Response mechanism of MPN-coated chemiresistor is complicated. A model is derived that allows predictions of MPN-coated chemiresistor (CR) responses from vapor-film partition coefficients, and analyte densities and dielectric constants.[189] Other nanoparticle-based sensing material like carbon nanotube (CNT) has also been reported with chemiresistor showing fast response and wide linear concentration range.[190, 191] For a complete coverage of chemiresistor and its application, refer to Aswal and Gupta's book.[192]

1.4.3 Capacitive Sensor. Capacitive sensor (sometimes called chemocapacitors or dielectrometers) is also one type of electrochemical sensor. The capacitive sensor consists of two sets of interdigitated electrode structures, which correspond to the two plates of a standard capacitor (Figure 1-7a). The sensor monitors changes in the dielectric coefficient of the polymer between electrodes upon analyte absorption. Primary application areas of chemocapacitors have been in humidity sensing using polyimide films,[193-199] because water has a high dielectric constant of 76.6 (in liquid state at 303 K), leading to large capacitance changes. More recent applications include VOC detection using polymeric layers.[200-205] Since water is omnipresent in environment and evokes large capacitive sensor signals, the inference to VOC detection from water must be considered. Kummer et al explored the possibility of minimizing the influence of analyte having high dielectric constant by adjusting the thickness of coating polymer.[206]

Although capacitive sensor responds to capacitance changes, the output signal is converted to a frequency change that is read out as a differential signal generated using a Sigma-Delta-modulator circuitry between a passive reference and a polymer-coated sensing capacitor.[207, 208] The response model for the CAP sensor is given by the following expression, which assumes a thick polymer layer (The polymer layer is chosen to be considerably thicker than the electric field extension so that the probed volume is completely filled with polymer) :[206, 208]

$$S = \frac{\Delta C}{\Delta c_a} \propto G_{cap} K V_m \Delta \varepsilon \qquad \text{(Eq. 1-3)}$$

where S is the sensitivity (pF/ppm), ΔC is the change in capacitance (pF) per unit change in vapor concentration Δc_a (ppm), G_{cap} is a constant determined by the geometric features of the capacitor device, V_m is the molar volume of the vapor in the condensed (liquid) state, and $\Delta\varepsilon$ is the difference between dielectric constants of the vapor (ε_a) and polymer (ε_p) indicated by:

$$\Delta\varepsilon = \varphi_A(\varepsilon_A - \varepsilon_{polymer}) \qquad \text{(Eq. 1-4)}$$

where φ_A is the amount of absorbed analyte expressed as volume fraction. The equation shows that sensor signals are positive for analytes with a dielectric constant larger than that of the polymer and negative otherwise.

1.4.4 Cantilever Sensor. Cantilever sensors commonly employed in atomic force microscopy [209] have been reported to measure a variety of different quantities like temperature, [210] magnetic fields, [211] viscosity, [212] and surface stress. [213] As one type of promising mass-sensitive chemical sensor, cantilever device coated with chemically sensitive polymers have been used for VOC detection. [214-221] The cantilever base is firmly attached to the silicon support. The free-standing cantilever end is coated with a sensitive layer that selectively absorbs analyte molecules from the gas phase (Refer to Figure 1-7b). The change of the resonance frequency upon mass-loading is measured. Similar to other well-established mass sensitive gas sensors such as thickness shear mode resonators (TSMR, quartz microbalances) and surface acoustic (SAW) device, [223] the

operation resonant cantilevers are based on the change of resonance frequency after mass-loading of analyte into coating polymer layer.

The response model for the CAN sensor is: [208]

$$S = \frac{\Delta f_o}{\Delta c_a} \propto G_{can} \cdot h \cdot K \cdot M \qquad \text{(Eq. 1-5)}$$

where S is the sensitivity (Hz/ppm) derived from the calibration curve, f_0 is the resonant frequency of the cantilever (Hz), c_a is the air concentration of the analyte vapor (ppm), G_{can} is a composite term related to the structure and mechanical properties of cantilever, h is the thickness of polymer layer deposited on sensor surface, K is the partition coefficient, and M is molecular mass of the vapor. Swelling effects and analyte-induced changes in the elastic modulus of the polymer have been neglected since these effects are only significant at high analyte concentrations. [224]

1.4.5 Calorimeter Sensor. Most chemical as well as enzyme- catalyzed reactions are accompanied by changes in enthalpy. For this reason, calorimetric transducers represent a universally applicable approach to chemical or biochemical sensors. One well-known example is the so called “pellistor” for reducible gas detection in the presence of a heated catalyst. [225] Resistor made of platinum has a significant increase of resistance with increasing temperature caused by the combustion of gas on the surface. The change of resistance is directly proportional to the concentration of gas present. Semiconductive thermistor is also used as thermal transducer for biochemical sensing. [226-229] Thermal devices based on thermocouples have been investigated for measurement

of glucose. [230, 231] Thermocouples operate on the basis of the Seebeck effect. If wires fabricated from two different metals or semiconductors are soldered together to form a circuit, any temperature difference that exists between the joined points leads to a measurable potential difference, the magnitude of which depends on the extent of the temperature difference and the materials involved. The use of several thermocouples connected in series (thermopiles) increases the sensitivity of the sensor. Application of thermocouple to VOC detection has also been reported. [232-234] The thermopile detects enthalpy changes produced by the absorption (heat of condensation) or desorption (heat of vaporization) of analyte molecules in the polymer film (see Figure 1-7c).

The response model for the CAL sensor can be formulated as below: [208]

$$S = \frac{\Delta \int U_{th} \cdot dt}{\Delta c_a} \propto G_{cal} \cdot h \cdot K \cdot \Delta H_S \qquad \text{(Eq. 1-7)}$$

where S is the sensitivity (mV-s/ppm), U_{th} is voltage generated by the thermopile during sorption and desorption (mV), integrated over time, t (s), G_{cal} combines the Seebeck coefficient and the geometric features of the CAL sensor, and ΔH_S is the sorption enthalpy.

Thermal sensors detect heat change as a result of chemical or physical process in sensing layer, and consequently, an important disadvantage is the requirement for thermostatic control of the device if strict measurement is essential. Regardless, universal applicability and new techniques for fabricating microthermosensors suggest that termal sensor may find wider use in the future.

1.4.6 Optical Sensor. Optical sensors represent a great variety of spectroscopic technologies that hold great promise for chemical analysis. Optical sensors are based on the interaction of electromagnetic radiation with matter. Such interaction changes the optical properties of analyte and results in light absorbance, transmission and reflectance that contains rich information about the chemical composition of the environment. Besides, quantification is possible by measure the intensity of radiation that is correlated to the concentration of analyte in accordance with Beer-Lambert's law. In a general arrangement of optical sensor, the monochromatic radiation passes through a sample and its properties are examined at the output. Alternatively, the sample may respond with a secondary radiation (induced luminescence) which is also measured. Applications of FTIR were reported for remote VOC sensing. [235-237] Another spectroscopic technique used for remote sensing of VOC in the atmosphere was reported by Ritcher et al. [238] Near-infrared (NIR) and middle-infrared (MIR) regions have been historically explored for detection of hydrocarbons in the air. However, the characteristic frequency region of hydrocarbon absorption is usually buried under the strong wide absorption bands of CO_2 and H_20 in the atmosphere. Although the interference can sometimes be reduced by using selective sensing layer on the optical fiber or choosing selective frequency, there are still many other factors that impedes optical sensor from using as a microsensor in VOC analysis. Compared to other vapor transducers, the sensitivity of most optical sensors is usually low (fluorescence might be an exception). Unlike other sensors mentioned above that can be miniaturized by using MEMS techniques, optical sensor can hardly microfabricated because it requires certain path length to achieve sensitivity. Besides,

optical sensors also need light source, monochromator, light detector and other accessory optical parts, as compared to other vapor transducers that need only silicon substrate and supporting electronic circuit. It is noteworthy that the emerging modern optical technologies like laser and more sensitive diode array detector have successfully helped reduce the size of optical device from bench-scale instrument to hand-held spectrophotometer. If this trend continues, it is believed that optical sensors can be used in analytical system in the near future.

The past decades witnessed the microcomputer revolution, which has provided sensor researchers with inexpensive but powerful tools and techniques for monitoring and assessing sensors. This, coupled with exciting advances in microsensor fabrication technology will undoubtedly lead to a rapid expansion in the use of chemical sensors and sensor arrays for many analytical measurements and, in particular, for those which require high-quality real time measurements.

1.5 Sensor Array Detector

1.5.1 Sensor Array. The biggest challenge to chemical sensor is the presence of interference. All chemical sensors suffer from a lack of selectivity which would permit their wide applications without detailed knowledge regarding the matrix composition of the samples under investigation. Although extensive efforts have been made to improve the selectivity of chemical sensor, an absolutely selective chemical sensor does not exist. The development of a single sensor to distinguish different vapors is difficult. However, sensor arrays with several partially selective sensor elements can be used to sense a wide variety of vapors. Probably the most important trend in chemical sensor research and

development is the sensor array and higher-order sensor. “Higher order” means that more than one transduction principle is applied to the same selective sensing material. For example, a selective layer on a surface acoustic wave sensor can be simultaneously used as coating material on a chemiresistor or an optical sensor. This higher-order sensor is the subject of **multiple transducer array** (or, “hybrid” array) studies and is elaborated in Chapter 3 of this dissertation. On the other hand, an array of different selective layers used in the same transduction principle forms a sensor array of one order. Both the order and the number of sensing layers in the array increase the number of data acquisition channels and thus the information content. [239]

It is essential that the different layers and higher-order sensors are orthogonal in their response. Otherwise the information provided by an additional sensor is redundant and will not contribute to the analytical performance.

Theoretically, and under ideal conditions, a sensor array containing several sensors with slightly different selectivity and sensitivity towards the analyte and its main interferences should make it possible to compensate for errors caused by the limited selectivity of each individual sensor. The task is analogous to the solving several unknowns in a mathematical linear equation system. There must be as many independent equations as unknowns so that the unknowns can be uniquely determined. In sensor array applications this means that one additional but different sensor is required for each interfering compound. The task then becomes one of solving n equations for n unknowns, which can be accomplished with chemometrics using some matrix algebra algorithms. Enhanced selectivity offered by combing different sensors into a sensor array makes it possible to develop a chemical sensor array system, mostly under the name of “electronic

noise". Importantly, modern micro fabrication technique makes it possible to fabricate sensor array system in a compact size. An important aspect of fabrication of sensing arrays is the ability to deposit multiple selective layers in defined areas of the chemical sensor. The ultimate limits of miniaturization of chemical sensors and its complications have been discussed by Williams and Pratt. [240] The setup of a sensing system based on sensor array is illustrated in Figure 1-8. This setup also constitutes the basic configuration of electronic nose device. [33] For the purposes of this review, a complete sensor array system will be defined as consisting of the following basic components: [241]

1. An array of three or more electronically independent sensors;
2. Analogue circuitry;
3. Adata acquisition (I/O) card;
4. Microcomputer or microprocessor control;
5. Software for control of data acquisition and processing

After passing through the required analogue signal processing circuitry (impedance conversion, current-voltage conversion, amplification, offset, and filter), signals from the array are digitized using a multichannel DAC (I/O) card and passed to the on-board computer (microprocessor) where they are displayed, processed, and stored according to the software used. With the I/O card connected to various actuators (valves, pumps, heaters, etc.), the software can also maintain control of the system under study and apply digital signal processing (gain, offset, filter, etc.) on-line. The user interacts with the system via a graphical front panel interface. [241] Chemical sensing devices fabricated from silicon are commonly created with MEMS (micro-electromechanical systems) technologies, which integrate mechanical elements, sensors, actuators, and

electronics into complete lab-on-a-chip systems. The electronic components are fabricated using integrated circuit (IC) processes (e.g., CMOS, bipolar, or BICMOS processes), and the micromechanical components are fabricated using compatible "micromachining" processes that selectively etch away parts of the silicon wafer or add new structural layers to form the mechanical and electromechanical devices. In the end, MEMS enable the development of devices with the computational performance of microelectronics and the perception and control capabilities of microsensors and microactuators. [33]

1.5.2 Multi-transducer Array. One field of sensor array study of special interest is the high-order multi-transducer (MT) array mentioned earlier that uses two and more transduction principles in the same array. Including various vapor transducer principles in the same array leads to greater flexibility in the choice of sensors and materials. The most notable merit of an MT array is the improved orthogonality among the sensors in the array. Such merits have been claimed by several researchers. [242-245] A monolithic MT array was reported by Hierlemann and Baltes et al with a micro-fabricated CMOS single-chip gas detection system. [222] Unfortunately most work published on hybrid array is limited in scope by focusing on characterizing odors or flavors rather than discriminating among analytes. A comprehensive study of MT array as compared to traditional single-transducer array is presented in Chapter 3 of this dissertation.

1.5.3 Electronic Nose. Electronic nose is an imposing example of successful application of sensor array system to characterization of odors or identification of fruit, wine, beverage and cheese. [246-258] In the first electronic nose report ever published, three

different metal oxide gas sensors were employed to capture the "flavor fingerprint" of several substances. [259] Until now the classical electronic nose, consisting of an array of sensors, is still the most common approach, although new technologies like optical sensor, mass spectrometry and ion mobility spectrometry have recently entered this field. [33]

However, literature examples of successful sensor-array applications should not be extrapolated to the quantitative analysis of a single analyte. Considerable caution is warranted with respect to this sensor array approach to increasing the selectivity of chemical sensors and correcting for analytical errors caused by interference. First, the interfering components must be known in advance so that one can prepare a set of calibration mixtures for the sensor array. Complete information regarding potential interferents within the sample in question is therefore required. This may prove applicable in certain well-defined analytical situations, but certainly not in cases where the sample matrix is not sufficiently known a priori, as in most environmental and some clinical analyses. Such precautions cannot be circumvented even with the use of the most sophisticated modern chemometric approaches to pattern recognition analysis. [260]

To address the complex mixture analysis issue with a sensor array, application of a simple and rapid preliminary separation step will be more productive with respect to the goal of achieving a reliable analytical result. The hyphenation of separation technology with sensor array detector greatly relieve sensor array from analyzing complex mixtures. With the use of upstream chromatographic separation, complex vapor mixture is separated into individual vapors or simple vapor mixtures. This not only enhances the selective measurement by sensor array but also greatly reduce the laborious calibration needs and the computational cost of subsequent data analysis. Several researchers have

utilized preceding chromatographic separation in detection of complex vapor mixtures. [175, 261-264] In particular, Zellers et al. performed extensive work in developing a sophisticated yet portable gas chromatograph with tunable retention and detection with microsensor array. [175] A facile hyphenation of gas chromatography and a microcantilever array sensor was reported by Sepaniak et al. [263] Zampolli et al also reported a selective hybrid microsystem consisting of a zero grade air unit, a commercial minipump, a minivalve, a silicon micromachined packed GC column, and a metal oxide sensor as the detector. [264]

Compared to other traditional sophisticated analytical methods, the sensor array system or electronic nose devices are intent on fast analysis, easy operation and small size. Consequently GC entered in this field in the fast or ultrafast mode. To increase the separation speed during analysis, different fast GC techniques have to be adapted. For gas/liquid chromatography this can be an increase of the carrier gas flow rate, an increase of the temperature-program heating rates, a reduction of the column length, a reduction of the column diameter, a reduction of the thickness of the stationary phase, and the use of a faster carrier gas. Fast separation also brings out such side effects as lower resolution and smaller sample capacity, or both. It is also important to note that these fast GC techniques increase the demands on the detector technology used in terms of sensitivity, speed, and dead volume. [33]

1.6 Chemometrics

The WIMS MicroGC belongs to a group of hyphenated analytical instruments. Hyphenated (or multidimensional) instruments employ two or more analytical instrumental techniques either sequentially or in parallel. Hence, one can have

multidimensional separations (e.g., GC/GC, HPLC/GC), identifications (e.g., MS/MS), or separations/identifications such as GC/MS. [265-283] The types of analytical instruments that can be joined are very large depending only upon the non-destruction of samples after the initial analytical procedure and the ability of the manufacturer to interface the instrumental techniques. [284] The advantage of interfacing two or more analytical instruments is obvious: the analytical information is exponentially increased while data acquisition time is considerably reduced. [285] As compared to the traditional single-channel analytical detector that produces a signal response value for a sample (a scalar, or zero-dimensional data), and the multi-channel detector (e.g, spectrometer and sensor array) that produces a multivariate response for a sample (vector, or one-dimensional data), modern hyphenated instrument can easily generate higher-order data. For example, a two-dimensional response per sample (in form of matrix) can be generated by GC/sensor array system, with one dimension corresponding to different sensor channels and another dimension corresponding to retention time. Thus three-way data can be also be easily generated for more than two samples by the GC/sensor array system. Because hyphenated systems generate an immense amount of data within a short time, it is compelling to use data processing techniques for rapid automatic data mining, analysis and interpretation. Chemometrics was introduced since the late 1960s to address the data analysis needs that come with modern analytical instruments. [286, 287] With the availability of personal computers, large databases and complicated statistical computations can be easily handled. Chemometrics is rather new, but it has already had a huge impact on the modern analytical instrumentation, especially on spectroscopic field to such an extent that chemometrics software is nowadays integrated with spectroscopic laboratory and process

instrumentation as a standard. [288, 289] Traditional univariate statistics are not sufficient to describe the multivariate data generated from modern measurements. From the beginning chemometrics goes beyond the limitations of univariate statistics by using methods from multivariate statistics, mathematics and computer science. Today chemometrics is still evolving in order to include new development. The domain of chemometrics is very broad covering topics from experimental design to artificial intelligence and overlaps with bioinformatics, genomics, proteomics and metabonomics. [290] This dissertation does *not* intend for a detailed review of every aspect of chemometrics that have been the subject of several excellent monographs. [291-302] Several key chemometric methods, however, will be introduced briefly in the context of sensor-array based GC development.

1.6.1 Data Acquisition and Pre-processing. Vapor sensor detects the chemical composition of its environment and converts this information into measurable electronic signals. Data acquisition is the first step for data analysis of this electronic signal. Data acquisition hardware in the sensor array system collects the signal and converts it into a digital signal that is more suitable for computer processing. [302] Each sensor in the sensor array is a separate data acquisition channel. The response from each sensor can be put together into a response vector. If sensors in an array are orthogonal to each other, i.e., the response of one sensor is independent of the others, which can often realized by employing different sensing materials, the response vector contains useful information about the identity and quantity of the sample under investigation, and forms the basis of employing chemometrics for the purpose of vapor identification and quantification. While the magnitude of the sensor response is correlated to vapor concentration by sensor

calibration, the concentration may have a scaling effect on the patterns of the sensors, especially when the sensor responses are heterogeneous in the case of a multi-transducer array which utilizes different transduction principles in the same array. While a pattern recognition algorithm normally examines the differences between the patterns, a scaling effect may mask the interrelations between them. Data preprocessing is used to remove the scaling effect. A common practice is the data normalization. [303] Two data normalization methods used in this dissertation will be discussed. The first one is relative scaling by dividing each element in the response vector by the maximal response. This relative scaling compresses response values with a maximum value of 1, resulting in a relative response pattern that is independent of concentrations and can be characteristic "fingerprint" of the analyte under study if the sensor array is selective.

Given a response vector X_i measured with a sensor array for a sample i (individual vapor or vapor mixture) whose element is indicated as x_{ij} (j=1 to k. k is the number of sensors in the array), normalization of sensor response to a maximum of 1 is expressed as:

$$Y_i = X_i / \max(X_i) \qquad \text{Eq. (1-8)}$$

where Y_i is the normalized response vector with element y_{ij} having value between 0 and 1. This relative scaling method is used whenever response pattern is needed for qualitative analysis. This relative scaling method is preferred for its simplicity even though there exist other relative scaling methods.

In cases where the variance of responses measured for a sensor across many samples may distort the results by imposing inappropriate weight on the data analysis,

autoscaling is used to force the variance of each sensor to a value of 1. Given the number of samples measured with the sensor array is *n*, autoscaling of response x_{ij} (i=1 to n, j=1 to k) becomes:

$$y_{ij} = \frac{x_{ij} - \frac{\sum_{i=1}^{n} x_{ij}}{n}}{\sqrt{\frac{\sum_{i=1}^{n} (x_{ij} - \frac{\sum_{i=1}^{n} x_{ij}}{n})^2}{n}}} \qquad \text{(Eq.1-9)}$$

Autoscaling is frequently used in chemometric analysis when it is suspected that the sensors are measured on different scales and the variance may affect the correlation among the patterns. [304]

1.6.2 Extended Disjoint Principal Component Regression (EDPCR) and Monte Carlo Simulation Model. Pattern recognition techniques have been used to correlate the response pattern measured by a sensor array to the identity of the sample under investigation. [304] Typically, principal component analysis (PCA) or cluster analysis is performed on normalized response patterns obtained for the analyte of interest and its interfering analytes. The selectivity of the sensor array to the analyte, or the uniqueness of its response pattern is evaluated for the discrimination among the analyte and its interference by calculating the dissimilarity among the response patterns. Once it is determined that the response pattern of the analyte is unique enough that the clustering of response patterns measured for analyte during calibration can be sufficiently

discriminated from other interfering analytes, a classification model can be established and used for identification of future "unknown" response pattern. The "unknown" sample will be classified to analyte if its response pattern falls within the cluster of analyte. Criterion of classification can be established by using supervised leaning methods like nearest neighborhood, linear discrimination, and many others. [296] Once the identity of the "unknown" sample is determined, its concentration can be correlated to the magnitude of sensor responses by using multivariate calibration. [305] Multivariate linear regression is seldom used in sensor array calibration because of collinearity that is commonly observed among sensors. Latent variable based calibration methods like principle component regression (PCR), partial least square (PLS) regression are good alternatives.

Extended disjoint principal component regression (EDPCR) is based on the classical disjoint principal component classification model and incorporates the regression feature of PCR. [306,307] Therefore it can perform classification and calibration simultaneously. The concept of disjoint principal component classification model is simple: a principal component model is developed for each of the individual species to be discriminated. Classification of an unknown is based on the goodness of fit of its response vector to each of the PCA models. [307] EDPCR takes advantage of the integration of the qualitative and quantitative contents of the sensor responses. Unlike other pattern recognition method based on normalized array responses, information on the vapor concentrations is retained in EDPCR. Misclassification can be minimized and estimation of vapor concentrations is facilitated. [306] For a detailed description of the algorithms used in EDPCR, refer to Zellers et al. [306] By computing classification error

and accuracy of concentration estimation, the analytical performance of a sensor array can be evaluated with EDPCR.

A bottleneck in evaluating the performance of sensor array is the availability of a large amount of array response measurements required both for training and testing of a robust EDPCR model. Due to the cost of experiment setup and laborious sample preparation and sensor calibration, large amount of measurements are hard to obtain. The drift of sensor response makes experiment even more difficult. Most sensors can provide linear and additive responses at low concentration ranges. As long as their response behavior is reproducible, systematic and random errors in sensor response can be characterized, a simulation model can developed to synthesize sensor responses based on sensor calibrations. Monte Carlo simulations of sensor response were developed by Zellers et al. as a powerful tool of sensor array performance assessment.[173] As long as the sensor response errors are well characterized, it has been shown that the synthesized sensor response is comparable to measurement.[308] The synthetic sensor responses can then be analyzed by EDPCR for vapor recognition and concentration estimation. Monte Carlo simulation coupled with EDPCR provides an efficient way to explore the analytical performance of various sensor array combinations in analyzing a wide array of individual vapors and vapor mixtures of various complexities that would be difficult to accomplish experimentally. [173, 176, 308]

In Monte Carlo simulation of sensor response, random and systematic errors are applied to the calibrated responses to the vapor under study assuming a Gaussian error distribution. The population of error-enhanced synthetic responses is sampled iteratively, and each sample is treated as an unknown that is then assigned an identity and

concentration by EDPCR models established on calibration data. The number and nature of recognition errors observed from a large sample set (e.g., 500 trials) are logged and evaluated with respect to the average rate of recognition, vapor-specific rates of recognition, and the identities of any incorrect assignments. The simulated response data are generated according to the following error model: [173]

$$r_{ij}' = r_{ij}(1 + k_1\alpha + k_2\beta_j) + k_{3j}\gamma_j + k_{4j}\delta \qquad \text{Eq. 1-10}$$

Here, r_{ij}' is the synthetic response to vapor i from coated sensor j; r_{ij} is the starting-point response value generated by randomly selecting a point along the regression line derived from calibration; k_1 is the relative standard deviation (RSD) that accounts for variability in sample delivery to the sensor array; k_2 is the RSD of the sensitivity estimate obtained from repeated calibrations; k_{3j} is the root-mean-square (RMS) error in the baseline for each sensor due to random noise; k_{4j} is the RMS error attributable to the combination of inherent baseline drift and fluctuations in the response to residual water vapor; α, β, γ and δ and are independent normally distributed variables with zero mean and unit standard deviation. The error model accounts for different factors that can contribute to the variation of sensor responses. There might be other unknown significant factor that is not included. On the other hand some factors may not be significant and will also not be included, depending on specific sensor array application.

1.6.3 Limit of Recognition of a Sensor Array. For a single chemical sensor, its sensitivity is always limited. Therefore its application is mostly restricted to detection of a wide range of analytes without discrimination. Its ability to detect a vapor is commonly measured by the metric limit of detection (LOD). When sensors of different selectivity (often also of different sensitivity) are combined into a sensor array, vapor recognition becomes possible because the combination of partially selective sensors in the array is able to provide a characteristic response pattern that can be correlated to the vapor identify. For an analyte of low concentration it may be detectable by some of sensors in the array but not by others due to the different sensitivities. One analyte can only be recognizable if its concentration is high enough to be detectable by most sensors in the array. So the limit of concentration of a vapor or the range of concentrations of a vapor mixture beyond which the analyte is deemed "recognizable" is essential for evaluating the capacity of sensor array. Realizing the difficulty in recognizing vapor at low concentrations, Zellers et al. defined the limit of recognition (LOR) as an alternate performance criterion for sensor arrays. [309-311] LOR of an individual vapor is defined as the concentration below which the vapor cannot be recognized with an acceptably low error (e.g., smaller than 5%). LORs were determined for 16 individual vapors using an array of four polymer-coated SAW sensors. [311] LOR of a vapor mixture is defined as the continuous combinations of component concentrations beyond which the mixture can be recognized with an error rate smaller than 5%. LOR of a vapor mixture is a function of the absolute and relative concentrations of the components. A 6-sensor SAW array was evaluated for its LOR determined with different binary and ternary mixtures. [173]

1.6.4 Self-Modeling Curve Resolution. The purpose of curve resolution is threefold: 1) determines the number of components in a mixture, 2) estimate the concentration profiles of each component, and 3) estimate the characteristic pure response profile (Pure spectrum in spectroscopy, or response pattern in sensor array. Response pattern generated by a sensor array resembles spectrum in spectroscopy. For simplicity, "spectra" and "response pattern" are used interchangeably in this section and Chapter 5). [312-314] Curve resolution offers insight into the composition of a mixture, the relative concentration of mixture components, and more importantly the clue to the identity of these mixture components. It is assumed that sensor response is additive. Given n samples of different mixture proportions each measured with k sensors, the mixture response forms a $n \times k$ matrix $\boldsymbol{X}$ (n rows and k columns) (Note: Unlike the uppercase X used earlier to denote a vector, Boldface capital $\boldsymbol{X}$ is used to denote a matrix. Lowercase letters are used to denote a scalar). In a hyphenated GC/sensor array system like sensor-array based GC, the n samples are ordered according to their sequence in retention time. Each row constitutes a spectrum and each column constitutes a chromatogram. Starting from matrix $\boldsymbol{X}$ it is possible to calculate the pure spectra and to determine the proportions of each in each mixture. Curve resolution is very useful in situations where there are co-eluting components and the overlapped peaks need to be quickly resolved numerically. The equation for curve resolution is as follow:

$$\boldsymbol{X} = C_1 S_1^{'} + C_2 S_2^{'} + \ldots + C_a S_a^{'} + \boldsymbol{E} \qquad \text{Eq. (1-11)}$$

where the C_a are concentration profiles for the a constituents in the mixture and the S_a are spectra of the chemicals for unit concentrations (standardized spectra). S' means the transpose of S. $\boldsymbol{E}$ is the matrix of residuals of the same size as $\boldsymbol{X}$ representing random noise left after extracting a components from $\boldsymbol{X}$. The C_A can come from different unordered mixtures, but also from sequential elution profiles in chromatography. The goal is to have the calculated spectra resemble the pure constituent spectra as much as possible. Concentration profiles C_1, C_2,… and C_a can be put into a matrix $\boldsymbol{C}$. Similarly S_1, S_2,…S_a can also put into a matrix $\boldsymbol{S}$. Then curve resolution model can be denoted in matrix format as follow:

$$\boldsymbol{X}=\boldsymbol{C}\boldsymbol{S}'+\boldsymbol{E} \qquad \text{Eq. 1-12}$$

The $n\times a$ matrix $\boldsymbol{C}$ contains a concentration profiles as its columns. The $k\times a$ matrix $\boldsymbol{S}$ contains a standardized spectra as its columns. This bilinear model forms the basis of self-modeling curve resolution (SMCR).

When the pure spectra for all components are known as a priori, the solution of the equation above is a least square problem. With non-overlapping spectral peaks and noise-free data, the equation seems trivial, but overlapping peaks and noise in the data make the task rather awkward, especially when there is no information about the composition of the mixture. The philosophy of SMCR is to use the bilinear decomposition, a mathematical deconvolution approach, also called factor analysis, in place of chemical or physical separation techniques, to cope with the instrumentally unresolved mixture signals and elucidate the pure variables of overlapped components.

[315] As the terminology "self-modeling" implies, SMCR does not require *a priori* any specific information concerning the data to resolve the pure variables. The only premises are the bilinear model for the data and certain generic knowledge about the pure variables such as non-negativity and unimodality that can be naturally satisfied for two-way data obtained from multivariate measurements.

Mathematically there are infinite solutions to the decomposition of $\boldsymbol{X}$ in to $\boldsymbol{C}$ and $\boldsymbol{S}$. That is the rotational ambiguity problem in employing factor analysis to resolve a system into underlying latent factors. [294] To solve this problem chemical expertise must be combined with mathematical deconvolution to generate solutions that are both mathematically and chemically sound. Relevant information about the problem under study is incorporated into the curve resolution in the forms of different constraints. The commonly used constraints to curve resolution solution are non-negativity of concentrations and unimodality of chromatographic peaks. Other constraints like the sign of sensor response (negative or positive) can also be employed. These constraints are used in many techniques that have been proposed for resolving $\boldsymbol{X}$ in to $\boldsymbol{C}$ and $\boldsymbol{S}$. [316-328] A detailed review of these techniques is not the intention of this dissertation. A monograph published by Malinowski is available for that purpose. [294]

A special SMCR method, evolving factor analysis, is worthy further discussion because of its compatibility with sensor-array based GC. Sensor-array based GC generates sequential response vectors that are ordered according to the progress of chromatographic elution. The intrinsic order of the mixture response vector provides a unique way for SMCR solutions. By exploring the local rank information in the mixture response matrix $\boldsymbol{X}$, the points of emergence and degradation of mixture components can

be estimated. Selective region of each component can then be used for unique determination of $\boldsymbol{C}$ and $\boldsymbol{S}$ in factor analysis. The estimated $\boldsymbol{C}$ and $\boldsymbol{S}$ can also be further refined by alternating least square (ALS) algorithm. Detailed discussion of this hybrid EFA/ALS solution and its application to peak deconvolution are discussed in Chapter 5 for enhancing analytical capacity of sensor-array based GC.

Figures and Tables

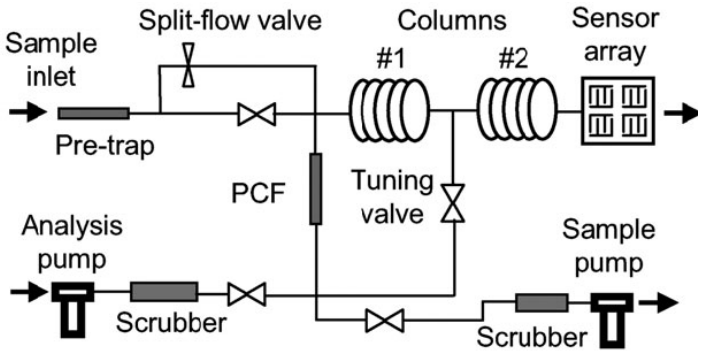

Figure 1-1. Schematic of fluidic sub-system of portable GC (PCF: preconcentrator/focuser). (Ref. 63)

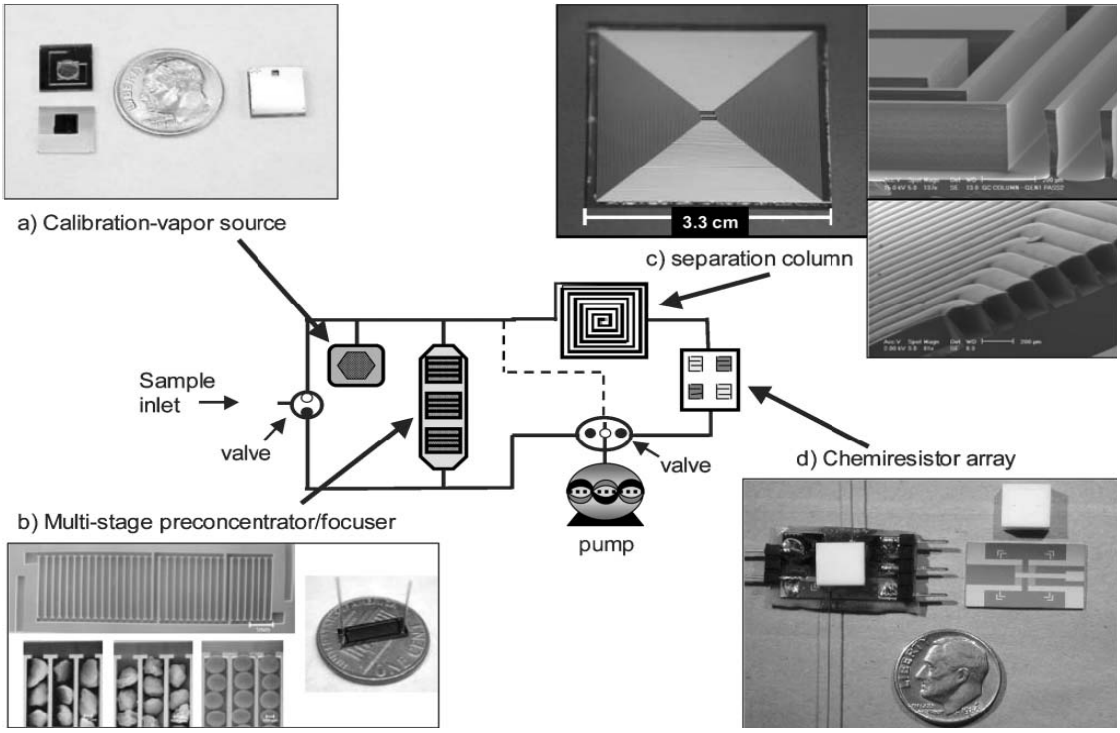

Figure 1-2. Block diagram of the prototype MEMS sensor-array based GC analytical system: (a) calibration-vapor source before (left) and after (right) assembly; diffusion channel and headspace aperture can be seen in the top section and macro-PS reservoir can be seen in the bottom section; (b) 3-stage adsorbent μPCF prior to loading and sealing (top left), with close-up SEM images of each section loaded with adsorbents (lower left) and assembled structure with capillary interconnects on a U. S. penny; (c) 3 m separation-column chip (left) with close up views of the channel crosssections prior to (top right) and after (lower right) sealing; (d) detector assembly with 4-chemiresistor array chip (right), Macor® lid (white square structure), and sealed detector with connecting capillaries mounted on a custom mounting fixture (left). The dashed line is a flow-splitter (Ref. 87)

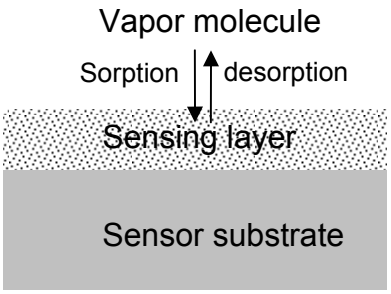

Figure 1-3. Reversible vapor-sensor interaction as a result of sorption in the sensing layer.

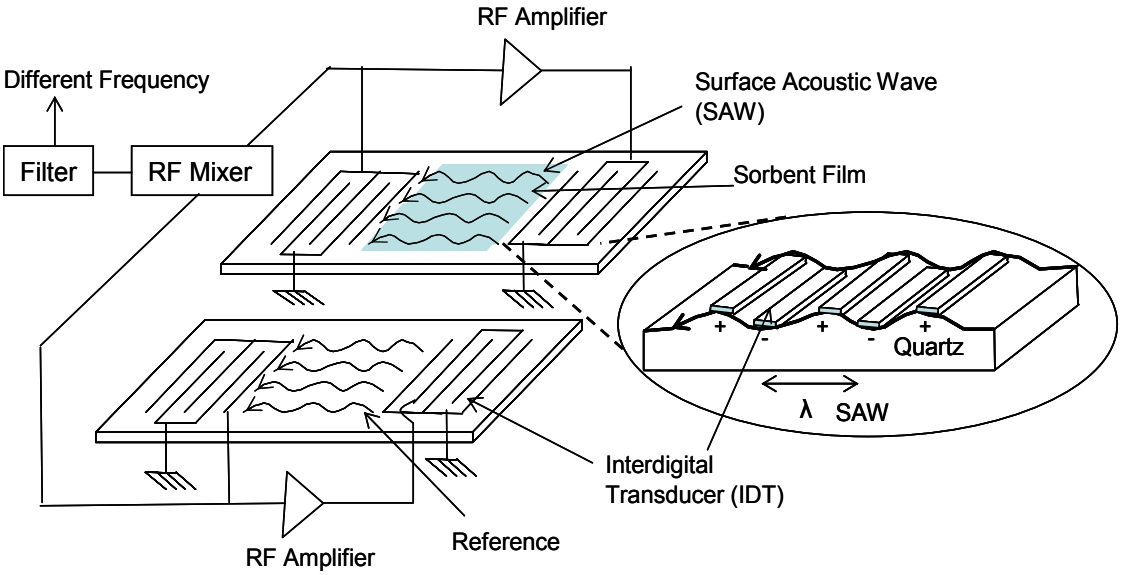

Figure 1-4. Schematic of SAW sensor in a dual delay-line oscillator configuration. A SAW device is produced by using microfabrication techniques to deposit a pair of metal interdigital transducers (IDT) on the surface of a thin, piezoelectric quartz plate. An input radio-frequency (RF) voltage is applied to the input IDT, which results in the launching of acoustic waves across the substrate that are received by the output IDT. When the output IDT receives this surface acoustic wave, it converts it back into RF voltage. An RF amplifier is connected to the input and output IDT electrodes to create a feedback loop circuit. As the SAW device works, a stable resonant frequency (f_0) is oscillated within this loop. The resonant frequency within the loop is determined by the SAW device and will change as the property of the sensor surface is altered (e.g. applying polymer coating). The net frequency shifts of a SAW device during sensing layer coating or vapor detection are monitored by electronically subtracting the resonant frequency from another bare SAW device through an RF mixer. A frequency-selective filter (i.e, low-pass filter) is placed in line with the SAW device to isolate different frequency. The difference frequency is measured using a frequency counter (Ref 151).

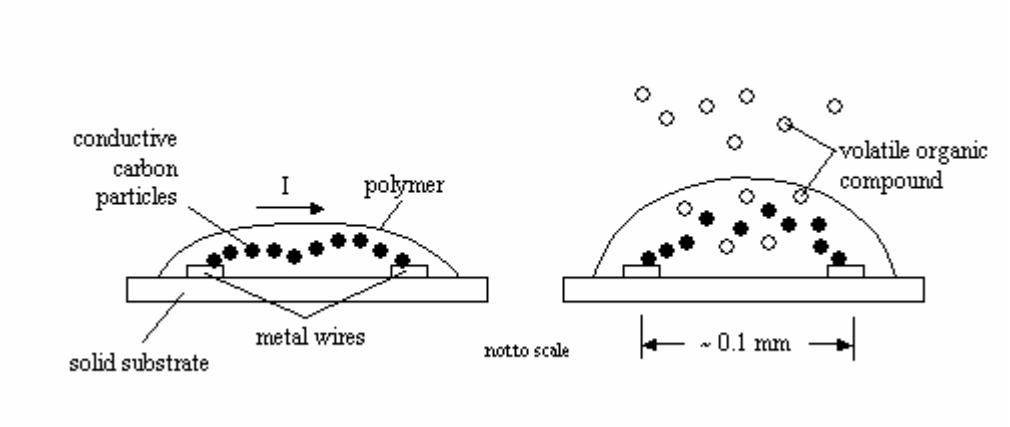

Figure 1-5. Change of polymer/carbon composite as a result of chemiresistor sorption of vapor. Carbon-loaded polymer is deposited on metal wires, creating a thin-film resistor when a voltage is applied across the wires. Electrical current is carried through the carbon particles. Polymer swells as vapors are absorbed. Resistance to current increases as carbon particles becomes separated. Resistance is proportional to concentration and the process is reversible. (Ref. 180)

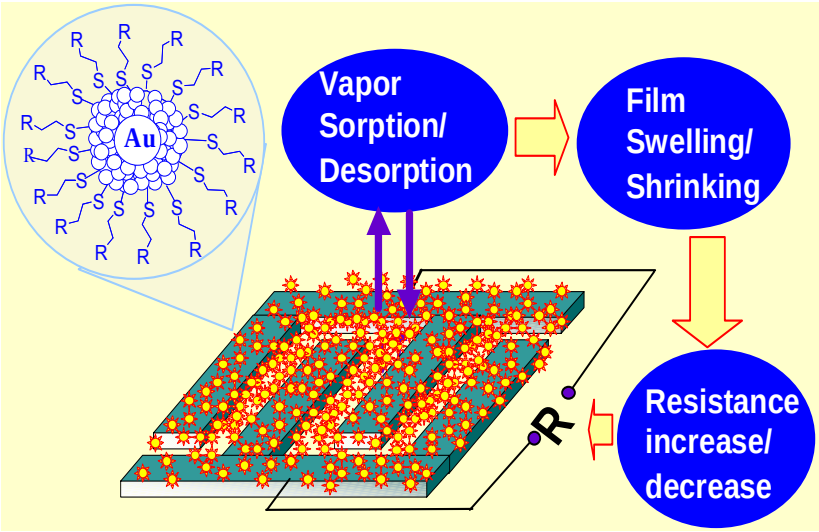

Figure 1-6. Transduction principle of MPN-coated chemiresistor. The net resistance change of MPN material as a result of sorptions is proportional to vapor concentration. (Ref. 189)

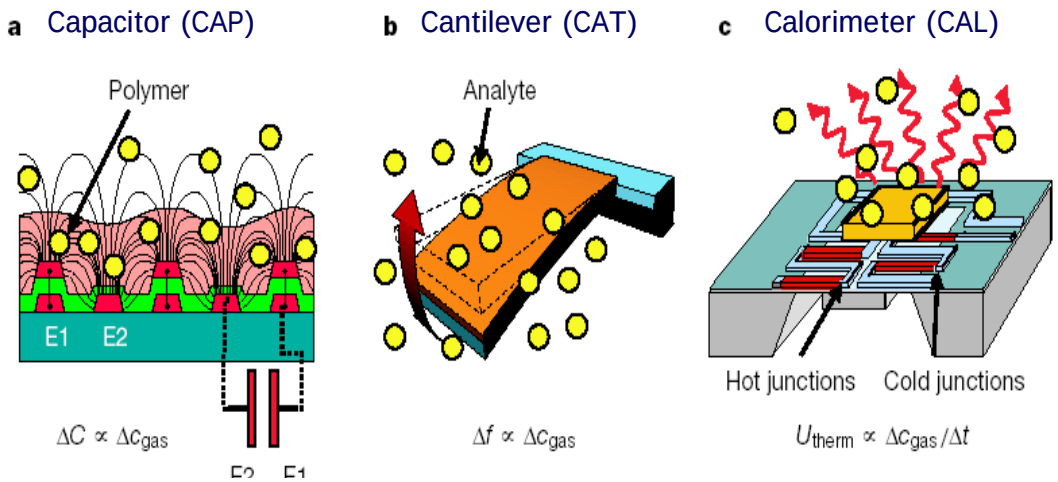

Figure 1-7. Schematics of three polymer-coated vapor transducers. (Ref. 222)

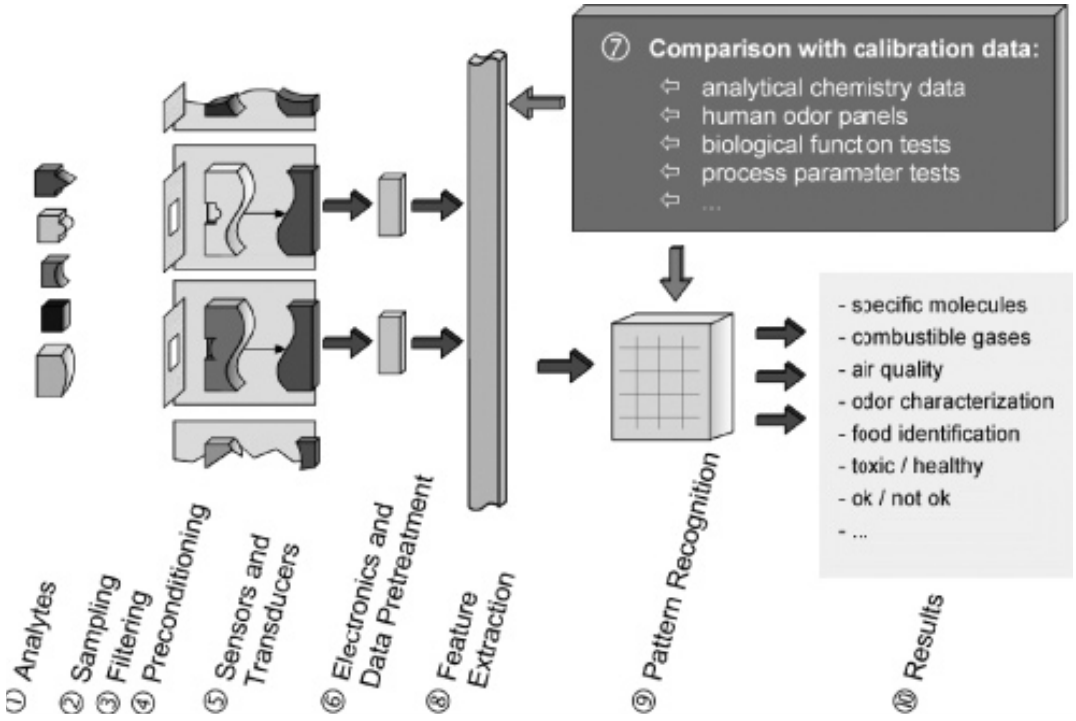

Figure 1-8. Schematic setup of a sensor system. Via sampling, filtering, and reconditioning the analytes are led to the sensing elements. These consist of a sensitive layer and a transducer to transform the chemical information into an electrical one. After the signal is recorded, data pretreatment, and feature extraction, pattern recognition evaluates the data using the calibration data (Ref. 241)

Table 1-1. Classification of portable chromatographs (Ref. 49)

Type	Purpose	Advantages, capabilities
Compact	For mobile and stationary laboratories	Saving of costs, power, materials, and space with analytical characteristics similar to those of stationary chromatographs, weight 10–25 kg
Portable, transportable, field	For on-site analysis	Small weight, rapid analysis, gas and power self-supporting, weight 5–15 kg
Chip-based chromatographs ,sili-con micromachining technology,, handheld, personal, pocket	For on-site analysis, handheld	For the fast resolution of relatively simple analytical problems, fully self-supporting, restricted analytical capabilities, weight 0.2–3 kg
Specially designed chromatographs, micro chromatographs	For space investigations	Automated analysis, small weight, resistant to impact and shaking

Reference:

1. Wise, K. Najafi, D, K, Sacks, R. D, Zellers, E. T, "A Wireless Integrated Microsystem for Environmental Monitoring," Digest IEEE International Solid-State Circuits Conference, San Francisco, CA, February 9-12, 2004, pp. 434-435.

2. Lu,C.-J, Steinecker, W. H, Tian, W.-C, Agah, M, Potkay, J. A, Oborny, M. C, Nichols, J, Chan, H. K. L, Driscoll, J, Sacks, R. D, Pang, S. W, Wise, K. D, Zellers, E. T, *Lab On A Chip*, 2005, 5, 1123-1131.

3. Lu, C.-J, Jin, C, and Zellers, E. T, *J. Environ. Monit*, 2006, 8, 270–278.

4. Jin, C, Kurzawski, P, Hierlemann, A, Zellers, E. T, *Anal.Chem*, 2008, 80(1), 227-236.

5. Jin, C, Zellers, E. T, "Limits of Recognition for Binary and Ternary Vapor Mixtures Determined with Multi-Transducer Arrays", submitted to *Analytical Chemistry*
6. United States Government Printing Office, "Title 40: Protection of Environment, Part 51-Requirements for Preparation, Adoption, and Submittal of Implementation Plans, Subpart F—Procedural Requirements, 51.100: Definitions." GPO Access: Electronic Code of Federal Regulations (e-CFR), www.vlex.com/vid/19784887, retrieved April 2008

7. European Union Publications Office, "Directive 2004/42/CE of the European Parliament and of the Council", EUR-Lex, www.eur-lex.europa.eu/LexUriServ, retrieved April 2008

8. World Health Organization Regional Office for Europe, Air quality guidelines for Europe, 2nd ed. Copenhagen, 2000 (WHO Regional Publications, European Series, No. 91)

9. Atkinson, R, Arey, J, *Chem. Rev.* 2003,103, 4605–4638

10. Greenberg, J. P, Guenther, A, Zimmerman, P, Baugh, W, Geron, C, Davis, K, Helmig, D, Klinger, L.F, *Atmos. Environ*. 1999, 33, 855–867

11. Rogak, SN, Pott, U, Dann, T, Wang, D, *J. Air. Waste Manage Assoc*, 1998, 48, 604–615

12. Molhave, L, Nielsen, GD, *Indoor Air*, 1992, 2, 65–77

13. Van Winke, M. R, Scheff, P. A, *Indoor Air*, 2001, 11, 49-64

14. Salvatore R. Dinardi (Editor), The Occupational Environment: Its Evaluation, Control, and Management, Second Edition, AIHA, 2003

15. Wang, D. K. W, Austin, C. C, *Anal Bioanal Chem*, 2006, 386,1089–1098

16. Leidel, N. A, Busch, K. A, in L. J. and L. V. Cralley, eds, Patty's Industrial Hygiene and Toxicology, Vol. 3a, 2nd ed, John Wiley & Sons, Inc, New York, 1985.

17. Rappaport, S. M, Kupper, L. L,*Journal of Exposure Analysis and Environmental Epidemiology*, 2004, 14, 92–107

18. Paustenbach, D J, *Journal of toxicology and environmental health*. Part B, Critical reviews, 2000, 3(3),179-291

19. DiNardi, Salvatore R, The Occupational Environment: Its Evaluation, Control, and Management, AIHA Press, 2003

20. NIOSH, NIOSH Manual of Analytical Methods (NMAM). Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, Division of Physical Sciences and Engineering, Cincinnati, OH, see http://www.cdc.gov/niosh/nmam/, retrieved April 2008

21. Mulhausen, JR, Damiano, J (eds), A strategy for assessing and managing occupational exposures. American Industrial Hygiene Association, Fairfax, VA, 1998

22. Krost, KJ, Pellizzari, ED, Walburn, SG, Hubbard, SA. Collection and analysis of hazardous organic emissions. *Anal. Chem*. 1982, 54, 810–17

23. Hodgson, AT, Girman, JR, Binenboym, J. A multi-sorbent sampler for volatile organic compounds in indoor air. 1986. Presented at 79th Annu. Meet. Air Pollut. Control Assoc, June, Minneapolis, MN

24. EPA (1999) Compendium of methods for the determination of toxic organic compounds in ambient air. Center for Environmental Research Information, US Environmental Protection Agency (see http://www.epa.gov/ttn/amtic/airtox.html, last accessed April 2008)

25. Hoyt, SD, Rasmussen, RA (1985) Determining trace gases in air and seawater. In: Mapping strategies in chemical oceanography (ACS Advances in Chemistry Series 209). American Chemical Society, Washington, DC

26. Hoyt, SD (1982) Air-sea exchange of atmospheric trace gases. In: Environmental science and engineering. Oregon Graduate Center, Beaverton, OR

27. EPA (1999) Compendium method TO-15: Determination of volatile organic compounds (VOCs) in ambient air using specially prepared canisters with subsequent analysis by gas chromatography. In: Compendium of methods for the determination of toxic organic compounds in ambient air, 2nd edn. US Environmental Protection Agency (see http://www.epa.gov/ttn/amtic/files/ambient/airtox/to-15r.pdf, last accessed April 2008)

28. EPA (1988) Compendium method TO-14: The determination of volatile organic compounds (VOCs) in ambient air using summa passivated canister sampling and gas chromatographic analysis. In: Compendium of methods for the determination of toxic organic compounds in ambient air. US Environmental Protection Agency

29. EPA (1999) Compendium method TO-14A: Determination of volatile organic compounds (VOCs) in ambient air using specially prepared canisters with subsequent analysis by gas chromatography. In: Compendium of methods for the determination of toxic organic compounds in ambient air, 2nd edn. US Environmental Protection Agency

30. Fänger, O. 1987. A solution to the sick building mystery. In Indoor Air '87 Proc. 4th Int. Conf. Indoor Air, ed. B Seifert, 4:49–57. Berlin: Inst.Water Soil Air Hyg.

31. Amoore, J. E., Hautala, E., J*ournal of Applied Toxicology*, 1983, 3(6), 272 – 290

32. Leonardos, G., Kendall, D., Barnard, N., *Journal of the Air Pollution Control Association*, 1969,19, 91-95

33. Röck, F, Barsan, N, Weimar,U, Electronic Nose: Current Status and Future Trends, *Chem. Rev*. 2008, 108, 705-725

34. Halova, J.; Strouf, O.; Zak, P.; Sochozova, A.; Uchida, N.; Yuzuri, T.; Sakakibara, K.; Hirota, M. *Quant. Struct.-Act. Relat*. 1998, 17, 37

35. Luque de Castro M. D, Gámiz-Gracia L, *Analytica Chimica Acta*,1997, 351(1-3), 23-40

36. Aragon, P, Atienza, J, Climent, MD, *Crit Rev Anal Chem*, 2000, 30, 121–151

37. Dewulf, J, Van Langenhove, H, *J Chromatogr* A, 1999, 843, 163–177

38. Santos, FJ, Galceran, MT, *J Chromatogr* A, 2003, 1000,125–151

39. Santos, FJ, Galceran, MT, *Trends Anal Chem*, 2002, 21,672–685

40. Helmig, D, *J Chromatogr* A, 1999, 843,129–146

41. Yashin, YI, Yashin, AY, *J Anal Chem*, 2001, 56,200–213

42. Santos, FJ, Galceran, MT, *Trends Anal Chem*, 2002, 21,672–685

43. Helmig, D *J Chromatogr* A, 1999, 843,129–146

44. Bertsch, W, J, *High Resol. Chromatogr*, 1999, 22, 647-665

45. Phillips, J. B, Xu, J, *J. Chromatogr*. A,1995, 703, 327-334

46. Helmig, D , *Journal of Chromatography* A , Volume 843, 1 - 2 , 129 - 146
47. Tannerj RL, *Atmos Environ*, 2003, 37,1271–1276

48. Matisová, E, Dömötörová, M, *Journal of Chromatography* A, 2003, 1000(1-2), 199-221

49. Yashin, YI, Yashin, AY, *J Anal Chem*, 2001, 56,794–805

50. Puskar, M. A, Plese, M. R, *Am. Ind. Hyg. Assoc. J*, 1996, 57, 843–848.

51. Spangler, G. E, Miller, R. A, *Int. J. Mass Spectrom*, 2002, 214,95–104.

52. Matisova, E, Domotorova, M, *J Chromatogr* A, 2003, 1000,199–221

53. Harris, C. M, *Anal. Chem*, 2002, 74(21), 585A–589A.

54. Eiceman, G. A, Gardea-Torresdey, J, Dorman, F, Overton, E, Bhushan, A. and Dharmasena, H. P, *Anal. Chem*, 2006, 78(12), 3985–3996.

55. Smith, P. A, Sng, M. T, Eckenrode, B. A, Leow, S. Y, Koch, D, R. P. Erickson, C. R. Jackson Lepage and G. L. Hook, *J. Chromatogr*, A, 2005, 1067, 285–294.

56. Liu, X, Pawliszyn, J, *Int. J. Environ. Anal. Chem*, 2005, 85(15),1189–1200.

57. Lu, C.-J, Zellers, E. T, *Anal. Chem*, 2001, 73, 3449–3457.

58. Lu, C.-J, Zellers, E. T, *Analyst*, 2002, 127, 1061–1068.

59. Grall, A. J, Zellers, E. T, Sacks, R. D, *Environ. Sci. Technol*,2001, 35, 163.

60. Whiting, J. J, Lu, C.-J, Zellers, E. T. and Sacks, R. D, *Anal. Chem*,2001, 73(19), 4668–4675.

61. Whiting, J. J, Sacks, R. D, *Anal. Chem*, 2002, 74, 246.

62. Lu, C.-J, Whiting, J. J, Sacks, R. D. and Zellers, E. T, *Anal. Chem*, 2003, 75(6), 1400–1409.

63. Zhong Q, Veeneman RA, Steinecker, H, Jia C, Batterman S. and ET. Zellers, *J. Environ. Monit*, 2007, 9, 440 – 448

64. Q. Zhong, PhD dissertation, University of Michigan, 2008

65. Angell, J.B, Terry, S.C, and Barth, P.W, *Sci. Am*, 1983, 248, 44.

66. Regnier, F.E, He, B, Lin, S, and Busse, J, *Trends Biotechnol*,1999,17, 101.

67. Elphick, M, *High Technol*, 1982, 2, 30.

68. Mowery, R.A, *ISA Trans*, 1986, 25, 46.

69. Lem, J.D, *Aerosp. Med*, 1967, 1110.

70. Terry, S. C, Jerman, J. H, Angell, J. B, *IEEE Trans. Electron Devices*, 1979, 26, 1880.

71. Reston, R. R, Kolesar, E. S, *J. Microelectromech. Syst*, 1994, 3, 134.

72. Yu, C. M, Lucas, M, Koo, J. C, Stratton, P, DeLima, T. and Behmeyer, E, Proc. ASME Micro-Electro-Mechanical Systems, Anaheim, CA, November 1998, pp. 481–486.

73. Lehmann, U, Krusemark, O. and Müller, J, Proceedings MicroTotal Analysis Systems Workshop, Enschede, Netherlands, May 2000, pp. 167–170.

74. SLS Microtechnology website, http://www.sls-micro-technology.de/unternehmen/. Retrieved April 2008

75. Noh, H, Hesketh, P. J. and Frye-Mason, G. C, *J. Microelectromech. Syst,* 2002, 11, 718.

76. Spengler, G. E, *Anal. Chem*, 1998, 70, 4805

77. Frye-Mason, G, Kottenstette, R, Mowry, C, Morgan, C, Manginell, R, P. Lewis, C. Matzke, G. Dulleck, L. Anderson and D. Adkins, in Proceedings Micro Total Analysis Systems Workshop, ed. J. M. Ramsey and A. van den Berg, Kluwer, Dordrecht, The Netherlands, May 2001, pp. 658–660.

78. Manginell, R. P, Okandan, M, Kottenstette, R. J, P. R. Lewis, D. R. Adkins, J. M. Bauer, R. G. Manley, S. Sokolowski and R. J. Shul, in Proc. 7th International Conf. on Miniaturized Chemical and Biochemical Analysis Systems – mTAS ’03,

Squaw Valley, CA, October 2003, pp. 1247–1250.

79. Product Brochure, Microsensor Technology, Fremont, 1983.

80. Saadat, S. and Terry, S.C, *Am. Lab*. (Shelton, Conn.),1984, 16, 90.

81. Reston, R.R. and Kolesar, E.S, *J. Microelectromech. Syst*, 1994, 3, 134.

82. Mainga, A, Avail. Univ. Micro Films Int. Order, no. DA9 301 081, 1992, p. 237.

83. Bruns, M.W, Erdoel Kohle, Erdgas, *Petrochem*, 1994, 47, 80.

84. Overton, E.B. et al, *Field Anal. Chem. Technol*, 1996, 1, 87.

85. Overton, E.B. et al, Proc. Symp. Field Screening Methods Hazard. Wastes Toxic Chemicals, Las Vegas, 1995, p. 207.

86. Overton, E.B, Stewart, M, Corney, K.R, *J. Hazard Mater*, 1995, 43, 77.

87. Lu, C.-J, Steinecker, W.H, Tian, W.-C, Agah, M, Potkay, J.A, Oborny, M.C, Nichols, J, Chan, H.K.L, Driscoll, J, Sacks, R.D, Pang, S.W, Wise, K.D. and Zellers, E.T. (2005). First generation hybrid MEMS gas chromotograph. *Lab on A Chip*, 5, 1123-1131.

88. Wise, K. D, Najafi, K, Aslam, D. M, Brown, R. B, Giachino, J. M, L. C. McAfee, C. T.-C. Nguyen, R. O. Warrington and E. T. Zellers, in Technical Digest Sensors Expo, Chicago, IL, June 2001, pp. 175–182.

89. Zellers, E. T, Steinecker, W. H, Lambertus, G. R, Agah, M, Lu, C.-J, H. K. L. Chan, J. A. Potkay, M. C. Oborny, J. M. Nichols, A. Astle, H. S. Kim, M. P. Rowe, J. Kim, L. W. da Silva, J. Zheng, J. J. Whiting, R. D. Sacks, S. W. Pang, M. Kaviany, P. L. Bergstrom, A. J. Matzger, C¸ . Kurdak, L. P. Bernal, K. Najafi and K. D. Wise, Proceedings Solid-State Sensor, Actuator, and Microsystems Workshop, Transducers Research Foundation, Inc, Hilton Head, SC, June 6–10, 2004, pp. 61–66.

90. Lu, C.-J, Tian, W.-C, Steinecker, W. H, Guyon, A, M. Agah, M. C. Oborny, R. D. Sacks, K. D. Wise, S. W. Pang and E. T. Zellers, in Proc. 7th International Conf. on Miniaturized Chemical and Biochemical Analysis Systems – mTAS ’03, Squaw Valley, CA, October 2003, pp. 411–415.

91. Steinecker, W. H, Rowe, M. P, Matzger, A. J. and Zellers, E. T, in Proc. 12th International Conference on Solid-State Sensors, Actuators and Microsystems – Transducers ’03, Boston, MA, June 2003, IEEE, Boston, pp. 1343–1346.
92. Potkay, J. A, Wise, K. D, Proceedings of the IEEE International Conference on Micro Electro Mechanical Systems (MEMS), Miami, Florida, January 2005, pp.

415–418.

93. Agah, M, Potkay, J. A, Elstro, A. L, Lambertus, G. R, R. D. Sacks and K. D. Wise, Technical Digest Solid-State Sensor and Actuator Workshop, Hilton Head SC, June, 2004, pp. 302–305.

94. Lambertus, G, Elstro, A, Sensenig, K, Potkay, J. A, Agah, M, Scheuering, S, K. D. Wise, F. Dorman and R. D. Sacks, *Anal. Chem*, 2004, 76, 9, 2629–2637.

95. Lambertus, G. R, Sacks, D, *Anal. Chem*, 2005, 77, 2078–2084.

96. Whiting, J, Sacks, R. D, *J. Sep. Sci*, 2006, 29 (2) 193–328

97. Tian, W. -C, Pang, S. W, Lu, C. -J. and Zellers, E. T, *J. Microelectromech. Syst*, 2003, 12, 264.

98. Tian, W.C, Chan, H. K. L, Lu, C. -J, Pang, S. W. and Zellers, E. T, *J. Microelectromech. Syst*, 2005, 14, 498–507.

99. Chan, H. K. L, Takei, M, Pang, S. W, Veeneman, R. A. and Zellers, E. T, *J. Microelectromech. Syst*, submitted.

100. Astle, A, Paige, A, Bernal, L. P, Munfakh, J, Kim, H. and Najafi, K, Proc. ASME Internat. Mechanical Engineering Congress & Expos, New Orleans, LA, Nov. 17–22, 2002.

101. Kim, H. S, Najafi, K, Washabaugh, P. D. and Bernal, L. P, Proc. of the 12th International Conference on Solid-State Sensors, Actuators and Microsystems – Transducers ’03, Boston, MA, June 8–12, 2003, IEEE, Boston, pp. 794–797.

102. da Silva, L. W. and Kaviany, M, *Int. J. Heat Mass Transfer*, 2004, 47, 2417.

103. Oborny, M. C, Zheng, J, Nichols, J. M, Lu, C.-J, Bergstrom, P. L, R. P. Manginell, G. C. Frye-Mason and E. T. Zellers, Proceedings of the Seventh International Conference on Miniaturized Chemical and Biochemical Analysis Systems – mTAS ’03, Squaw Valley, CA, October 5–9, 2003, pp. 1243–1246.

104. Cai, Q. Y. and Zellers, E. T, *Anal. Chem*, 2002, 74, 3533.

105. Zheng, J, Christophersen, M. and Bergstrom, P. L, Extended Abstracts 4th Int. Conf. on Porous Semicon. Sci. Tech, Valencia, Spain, March 14–19th, 2004, pp. 42–43.

106. Dolnik, V, Liu, S, and Jovanovich, S, Electrophoresis (Weinheim, Fed. Repub. Ger.), 2000, 21, 41.

107. Janata, J, Bezegh, A, *Anal. Chem*. 1988. 60. 62R-74R

108. Janata, J, Josowicz, M, *Anal. Chem*. 1998, 70, 179R-208R

109. Aragón, P, Atienza, J, Climent, M. D, *Critical Reviews in Analytical Chemistry*, 2000, 30(2&3),121–151

110. Grate, J. W, *Chem. Rev*. 2000, 100, 2627-2648

111. Stone, A. J, The Theory of Intermolecular Forces, 1996, Clarendon Press, Oxford

112. Yamazoe, N, Krakowa, Y, Seiyama, T, *Sens. Actuators*, 1983, 4, 283

113. Matsushima, S, Teraoka, Y, Miura, N, Yamazoe, N, *Jpn. J. Appl. Phys*. 1988, 27, 1798.

114. Nitta, M, Kanefusa, S, Ohtani, S, Haradome, M, *J. Electr. Mater*. 1984, 13, 15.

115. Chiba, A, "Development of the TGS Gas Sensor," in S. Yamauchi (ed.), Chemical Sensor Technology, vol. 4, Elsevier, Amsterdam 1992, pp. 1 – 18.

116. Göpel, W, Hesse, J, Zemel, J. N, Sensors, vol. 2, VCH Verlagsgesellschaft, Weinheim 1991.

117. Göpel, W, Hesse, J, Zemel, J. N, Sensors, vol. 3, VCH Verlagsgesellschaft, Weinheim 1991.

118. Moseley, P. T, Tofield, B. C, Solid State Gas Sensors, Adam Hilger, Bristol 1987.

119. Moseley, P. T, Norris, J. O.W, Williams, D. E, Techniques and Mechanisms in Gas Sensing, Adam Hilger, Bristol 1991.

120. Lee, D.-D, Choi, D.-H, *Sens. Actuators* B, 1990, 1, 231

121. Tieman, R. S, Heineman, W. R, Johnson, J, Seguin, R, *Sens. Actuators* B, 1992, 8, 199.

122. Hanawa, T, Kuwabata, S, Yoneyama, H, *J.Chem. Soc. Faraday Trans.* 1988, 84,1587.

123. Nylander, C, Armgarth, M, Lundström, I, *Anal. Chem. Symp. Ser*. 1983, 17 203.

124. Bartlett, P.N, Archer, P. B. M, Ling-Chung, S.K, *Sens. Actuators*,1989, 19, 125.

125. Bartlett, P. N, Archer, P. B. M, Ling-Chung, S.K, *Sens. Actuators*,1989, 19 141.

126. Sadaoka, Y, Jones, T. A, Göpel, W, *Sens.Actuators* B, 1990, 1, 148.

127. Dogo, S, Germain, J. P, Maleysson, C, Pauly, A, *Sens. Actuators* B, 1992, 8 257.

128. Snow, A. W.; Wohltjen, H.; Jarvis, N. L. Abstracts of Papers of the American Chemical Society 2001, 221, 324-IEC

129. Wohltjen, H.; Snow, A. W. *Analytical Chemistry* 1998, 70, 2856-59.

130. Sakai, Y, Sadaoka, Y, Fukumoto, H, *Sens. Actuators* 1988, 13, 243.

131. Severin, E. J.; Lewis, N. S. *Anal. Chem*. 2000, 72(9), 2008-2015

132. Steinecker, W.H, Rowe, W.P. and Zellers, E.T. (2007). Model of Vapor-Induced Resistivity Changes in Gold-Thiolate Monolayer-Protected Nanoparticle Sensor Films. Analytical Chemistry, 79(13), 1164-1172.

133. Rowe, M.P. Steinecker, W.H. and Zellers, E.T. , *Analytical Chemistry*, 2006, 79(3), 1164-1172.

134. Khlebarov, Z. P, Stoyanova, A. I, Topalova, D. I, *Sens. Actuators* B,1992, 8, 33 – 40.

135. Goepel M, *Sens. Actuators*, B 1995, 26-27, 126-134.

136. Auge, J.; Hauptmann, P.; Hartmann, J.; Roesler, S.; Lucklum, R. *Sens. Actuators*, B 1995, 26, 181-186.

137. Ulmer, H.; Mitrovics, J.; Noetzel, G.; Weimar, U.; Gopel, W. *Sens. Actuators*, B 1997, B43, 24-33.

138. Di Natale, C.; macagnano, A.; Davide, F.; D'Amico, A.; Paolesse, R.; Boschi, T.; Faccio, M.; Ferri, G. *Sens. Actuators*, B 1997, B44, 521-526.

139. Alder, J. F.; McCallum, J. J. Analyst (London) 1983, 108, 1169-1189. Acoustic Wave Microsensor Arrays for Vapor Sensing Chemical Reviews, 2000, Vol. 100, No. 7 2645 19, 1-28.

140. D'Amico, A.; Verona, E. *Sens. Actuators* 1989, 17, 55-66.

141. Bastiaans, G. J. In Chemical Sensors; Blackie, Glasgow, 1988; pp 295-319.

142. Nieuwenhuizen, M. S.; Venema, A. *Sens. Mater*. 1989, 5, 261-300.

143. Ballantine, D. S.; Wohltjen, H. *Anal. Chem.* 1989, 61, 704A-715A.

144. McCallum, J. J. *Analyst*,1989, 114, 1173-1189.

145. Frye, G. C.; Martin, S. *J. Appl. Spectrosc. Rev*. 1991, 26, 73-149.

146. Thompson, M.; Kipling, A. L.; Duncan-Hewitt, W. C.; Rajakovic,L. V.; Cavic-Vlasak, B. A. *Analyst* , 1991, 116, 881-890.

147. Nieuwenhuizen, M. S.; Venema, A. In Sensors A Comprehensive Survey; Goepel, W, Jones, T. A, Kleitz, M, Lundstrom, J, Seiyama, T, Eds.; VCH, Weinheim, 1991; Vol. 2; pp 647-680.

148. Grate, J. W.; Martin, S. J.; White, R. M. *Anal. Chem.* 1993, 65, 940A-948A.

149. Grate, J. W.; Martin, S. J.; White, R. M. *Anal. Chem*. 1993, 65, 987A-996A.

150. Grate, J. W. In Assignment of the Glass Transition, ASTM STP 1249; Seylor, R. J, Ed.; ASTM, Philadelphia, 1994; pp 153-164.

151. Grate, J. W.; Frye, G. C. In Sensors Update; H. Baltes, W. Goepel and J. Hesse, Ed.; VSH, Weinheim, 1996; Vol. 2; pp 37-83.

152. Grate, J. W.; Baer, R. L. In Current Topics in Biophysics, Biosensors; Frangepol, P. T, Nilolelis, D. P, Eds.; Iasi University Press, Iasi, Romania, 1996; pp 51-78.

153. Grate, J. W.; Abraham, M. H. *Sens. Actuators*, B 1991, 3, 85-111.

154. Ballantine, D. S.; White, R. M.; Martin, S. J.; Ricco, A. J.; Zellers, E. T.; Frye, G. C.; Wohltjen, H. Acoustic Wave Sensors. Theory, Design, and Physico-Chemical Applications; Academic Press, New York, 1997.

155. Thompson, M. H.; Stone, D. C. Surface-Launched Acoustic Wave Sensors, Chemical Sensing and Thin-Film Characterization; John Wiley and Sons, New York, 1997.

156. Grate, J. W.; Abraham, M. H.; McGill, R. A. In Handbook of Biosensors, Medicine, Food, and the Environment; Kress-Rogers, E, Nicklin, S, Eds.; CRC Press, Boca Raton, FL, 1996; pp 593- 612.

157. McGill, R. A.; Abraham, M. H.; Grate, J. W. *CHEMTECH* 1994, 24(9), 27-37.

158. Grate, J. W, *Chem. Rev*. 2000, 100, 2627-2648

159. Hsiu, M.-D, Doctoral dissertation, University of Michigan, 2003

160. Pulker, H. K, Decostered, J. P. in C. Lu, A.W. Czanderna (eds.), "Applications of Piezoelectric Quartz Crystal Microbalances," Methods and Phenomena, vol. 7, Elsevier, Amsterdam 1984, Chap. 3.

161. Heising, R. A, Quartz Crystal for Electrical Circuits – Their Design and Manufacture, van Nostrand, New York 1947, p. 16.

162. Grate, J. W.; Snow, A.; Ballantine, D. S.; Wohltjen, H.; Abraham, McGill, M. H.; Sasson, R. A.; *Anal. Chem*. 1988, 60, 869-875.

163. Wohltjen, H. *Sensors and Actuators* 1984, 5, 307-25.

164. Snow, A.; Wohltjen, H. *Anal. Chem*. 1984, 56, 1411-1416.

165. Grate, J. W.; Snow, A.; Ballantine, D. S.; Wohltjen, H.; Abraham,M. H.; McGill, R. A.; Sasson, P. *Anal. Chem*. 1988, 60, 869-875.

166. Janghorbani, M.; Freund, H. *Anal. Chem*. 1973, 45, 325-332.

167. Edmunds, T. E.; S.West, T. *Anal. Chim. Acta* 1980, 117, 147-157.

168. McCallum, J. J.; Fielden, P. R.; Volkan, M.; Alder, J. F. *Anal. Chim. Acta* 1984, 162, 75-83.

169. Zellers, E. T. G. Z. Zhang, *Analy. Chim. Acta* 1993, 280, 1-13

170. Patrash, S. J. and Zellers, E. T, *Anal. Chem*. 1993, 65, 2055-2066.

171. Zellers, E. T, Batterman, S. A, Han, M, and Patrash, S. J, *Anal. Chem*. 1995, 67, 1092-1106.

172. Park, J, Groves, W. A, and Zellers, E. T, *Anal. Chem*. 1999, 71, 3877-3886

173. Hierlemann, A, Zellers, E. T, Ricco, A. J, *Anal. Chem*. 2001, 73, 3458-3466

174. Lu, C, Whiting, J. J, Sacks, R. D, and E. T. Zellers, *Anal. Chem.* 2003, 75(6), 1400-1409

175. Hsieh, M. D. and Zellers, E. T, *Anal. Chem*. 2004, 76, 1885-1895

176. Watson, J, *Sens. Actuators* B, 1992, 8, 173.

177. Severin, E. J.; Lewis, N. S. *Anal. Chem*. 2000, 72(9), 2008-2015.

178. Eastman, M. P.; Hughes, R. C.; Yelton, W. G.; Ricco, A. J.; Patel, S. V.; Jenkins, M. W. , *J. Electrochem. Sot.* 146,3907-3913 (1999).

179. Hughes, R. C.; Eastman, M. P.; Yekon, W. G.; Ricco, A. J.; Patel, S. V.; Jenkins, M. W. Tech. Digest of the 1998 Solid-State Sensor and Actuator Workshop, Transducers Research Foundation, Cleveland, 1998, pp. 379-382.

180. Patel, S. V.; Jenkins, M. W.; Hughes, R. C.; Yelton, W. G.; Ricco, A. *J. Anal. Chem*. 2000, 72(7), 1532-1542.

181. Heaney, M. B. *Appl. Phys. Lett*. 1996,69,2602.

182. M. B. Heaney, *Physics A*, 1997,241,296.

183. Viswanathan, R.; Heaney, M. B. *Phys. Rev. Lett*. 1995, 75, 4433.

184. Wohltjen, H.; Snow, A. W. *Anal. Chem*. 1998, 70(14), 2856-2859.

185. Brust, M.; Walker, M.; Bethell, D.; Schiffrin, D. J.; Whyman, R. J. *Chem. Soc, Chem. Commun*. 1994, (7), 801-2.

186. Rowe, M. P.; Plass, K. E.; Kim, K.; Kurdak, C.; Zellers, E. T.; Matzger, A. J. *Chem. Mater*. 2004, 16(18), 3513-3517.

187. Rowe, M. P, Steinecker, W. H, and Zellers, E. T, *Anal. Chem*, 2007, 79, 1164-1172.

188. Steinecker, W. H, Rowe, M. P, and E. T. Zellers, *Anal. Chem*, 2007, 79, 4977-4986.

189. Jose, K A.; Biju, P; Ashwin, W.; *Smart materials and structures*, 2004, 13, 1045-1049

190. Wang, F, Gu, H, Swager. TM, *J. Am. Chem. Soc*, 2008, 130 (16), 5392–5393.

191. Aswal, DK, Gupta, SK. (Author), Science and Technology of Chemiresistor Gas Sensors, Nova Science Publishers; 1 edition, 2006

192. Sheppard, N. F.; Day, D. R.; Lee, H. L.; Senturia, S. D. *Sens. Actuators* 1982, 2, 263-274.

193. Senturia, S. D. *Tech. Dig. - Transducers* 1985; 198-201.

194. Glenn, M. C.; Schuetz, J. A. *Tech. Dig. - Transducers* 1985; 217-219.

195. Denton, D. D.; Senturia, S. D.; Anolick, E. S.; Scheider, D. *Tech. Dig. - Transducers* 1985; 202-205.

196. Delapierre, G.; Grange, H.; Chambaz, B.; Destannes, L. *Sens. Actuators* 1983, 4, 97-104.

197. Boltshauser, T.; Baltes, H. *Sens. Actuators*, A 1991, 26, 509-512.

198. Boltshauser, T.; Chandran, L.; Baltes, H.; Bose, F.; Steiner, D. *Sens. Actuators*, B 1991, 5, 161-164.

199. Cornila, C.; Hierlemann, A.; Lenggenhager, R.; Malcovati, P.; Baltes, H.; Noetzel, G.; Weimar, U.; Göpel, W. *Sens. Actuators*, B 1995, B25, 357-361.

200. Steiner, F. P.; Hierlemann, A.; Cornila, C.; Noetzel, G.; Bachtold, M.; Korvink, J. G.; Göpel, W.; Baltes, H. *Tech. Dig. - Transducers* 1995, 2, 814-817.

201. Hagleitner, C.; Koll, A.; Vogt, R.; Brand, O.; Baltes, H. *Tech. Dig. -Transducers* 1999, 2, 1012-1015.

202. Koll, A.; Kummer, A.; Brand, O.; Baltes, H. *Proc. SPIE* 1999, 3673, 308-317.

203. Josse, F.; Lukas, R.; Zhou, R.; Schneider, S.; Everhart, D. *Sens. Actuators*, B 1996, 36, 363-369.

204. Domansky, K.; Liu, J.; Wang, L. Q.; Engelhard, M. H.; Baskaran, S. J. *Mat. Res*. 2001, 16 (10), 2810-2816.

205. Kummer, AM, Hierlemann, A, Baltes, H, *Anal. Chem*. 2004, 76, 2470-2477

206. Koll, A.; Kummer, A.; Brand, O.; Baltes, H. *Proceedings of the SPIE* 1999, 3673, 308-317.

207. Kurzawski, P, Hagleitner C, Hierlemann A, *Anal. Chem*. 2006, 78, 6910-6920

208. Berger, R.; Gerber, Ch.; Lang, H. P.; Gimzewski, J. *J. Microelectron. Eng*. 1997, 35, 375-379.

209. Barnes, J. R.; Stephenson, R. J.; Welland, M. E.; Gerber, Ch.; Gimzewski, J. K. *Nature* 1994, 372, 79-81

210. Leichlé, T. C.; von Arx, M.; Allen, M. G. Proc. IEEE MEMS 2001, Interlaken, Switzerland, 2001; pp 274-277.

211. Oden, P. I.; Chen, G. Y.; Steele, R. A.; Warmack, R. J.; Thundat, T. *Appl. Phys. Lett.* 1996, 68, 3814-3816.

212. Fritz, J.; Baller, M. K.; Lang, H. P.; Rothuizen, H.; Vettiger, P.; Meyer E.; Güntherodt H. J.; Gerber, Ch.; Gimzewski J. K. *Science* 2000, 288, 316-318.

213. Hierlemann, A.; Lange, D.; Hagleitner, C.; Kerness, N.; Koll, A.; Brand O.; Baltes, H. *Sens. Actuators*, B 2000, 70, 2-11.

214. Lange, D.; Hagleitner, C.; Brand, O.; Baltes, H. Digest of Technical Papers TRANSDUCERS 1999, Sendai, Japan, 1999; pp 1020-1023.

215. Maute, M.; Raible, S.; Prins, F. E.; Kern, D. P.; Ulmer, H.; Weimar, U.; Göpel, W. *Sens. Actuators*, B 1999, 58, 505-511.

216. Thundat, T.; Chen, G. Y.; Warmack, R. J.; Allison, D. P.; Wachter, E. A. *Anal. Chem*. 1995, 67, 519-521.

217. Lange, D.; Koll, A.; Brand, O.; Baltes, H. *Proc. SPIE* 1998, 3224, 233-243.

218. Lange, D.; Hagleitner, C.; Brand, O.; Baltes, H. Proc. IEEE MEMS 2000, Myazaki, Japan, 2000; pp 547-552.

219. Senesac, L.; Dutta, P.; Datskos, P. G.; Sepaniak, M. *Anal. Chim. Acta* 2006, 558, 94-101.

220. Sepaniak, M.; Datskos, P.; Lavrik, N.; Tipple, C. *Anal. Chem*. 2002, 74, 568A-575A.

221. Hagleitner, C.; Hierlemann, A.; Lange, D.; Kummer, A.; Kerness, N.; Brand, O.; Baltes, H. *Nature* 2001, 414, 293-296.

222. Brand, O.; Baltes, H. Micromachined resonant sensors. In Sensors Update; Baltes, H, Göpel, W, Hesse, J, Eds.; VCH, Weinheim, Germany, 1998; Vol. 4, p 36.

223. Lange, D.; Hagleitner, C.; Hierlemann, A.; Brand, O.; Baltes, H. *Anal. Chem*. 2002, 74, 3084-3095.

224. Baker, A. R, GB Pat. 892 530, 1962.

225. Mosbach, K. et al, *Biochim. Biophys. Acta*, 1974, 364, 140.

226. Mattiasson, B. et al, *Anal. Lett*. 1976, 9, 217.

227. Danielsson, B. et al, *Appl. Biochem. Bioeng*. 1981, 3, 103.

228. Hundeck, H. G. et al, *GBF Monogr. Ser*. 1992, 17, 321.

229. Guilbeau, E. J. et al, *Trans. Am. Soc. Artif. Intern. Organs* 1987, 33, 329.

230. Steinhage, G, Dissertation, University of Muenster, 1996.

231. Lerchner, J.; Seidel, J.; Wolf, G.; Weber, E. *Sens. Actuators*, B 1996, 32,71-75.

232. Bataillard, P.; Steffgen, E.; Haemmerli, S.; Manz, A.; Widmer, H. M. *Biosens. Bioelectron*. 1993, 8, 89-98.

233. Kerness, N.; Koll, A.; Schaufelbuhl, A.; Hagleitner, C.; Hierlemann, A.; Brand, O.; Baltes, H. N-well Based CMOS Calorimetric Chemical Sensors. Proceedings of IEEE Thirteenth Annual International Conference on Micro Electro Mechanical Systems, Miyazaki, Japan, 2000.

234. Bangalore, A.S.; Small, G.W.; Combs, R.J.; Knapp, R.B.; Kroutil, R.T.; Traynor, C.A.; Ko, J.D. *Anal. Chem*, 1997, 69, 118–129.

235. Chu, P.M.; Rhoderick, G.C.; Van-Vlack, D.; Wetzel, S.J.; Lafferty, W.J.; Guenther, F.R. Fresenius J. *Anal. Chem*, 1998, 360, 426–429.

236. Johnson, S.A.; Pham, N.H.; Novick, V.J.; Maroni, V.A, *Appl. Spectrosc*, 1997,51, 1423–1426.

237. Richter, P.I.; Peczeli, I.; Boeroecz, S.; Gazdag, L.; Leonelli, J.; Dolash, T. *Field Anal. Chem. Technol*, 1996, 1, 103–107.

238. Wise, B. M.; Janata, J. Encyclopedia of Energy Technology and the Environment; J. Wiley & Sons, New York 1995; pp 1247-58.

239. Williams, D. E.; Pratt, K. F. E. *J. Chem. Soc, Faraday Trans*.1995, 91(13), 1961-6

240. Diamond, Dermot, *Electroanatysiq* ,1993, 5, 795-802

241. Snow, A. W.; Barger, W. R.; Klusty, M.; Wohltjen, H.; Jarvis, N. L. *Langmuir* 1986, 2, 513- 519

242. Schierbaum, K. D, Gerlach, A, Haug, M, Gopel, W, *Sens. Actuator* A, 1992, 31, 130-137

243. Mitrovics, J, Ulmer, H, Weimar, U, Gopel, W, *Acc. Chem. Res*, 1998, 31, 307-315

244. Hierlemann, A, Baltes, H, *ANALYST* 2003, 128 ,1,15-28

245. Mitrovics, J.; Ulmer, H.; Weimar, U.; Göpel, W. Acc. Chem. Res.1998, 31, 307.
246. Wold, S. et al, Food Research and Data Analysis, Applied Sciences, Barking 1983, pp. 147 – 188.

247. Nakamoto, T, Fukunishi, K, Moriizumi, T, *Sens. Actuators* B, 1989, 1, 473.

248. Ema, K. et al, *Sens. Actuators* B, 1989, 1, 291.

249. Stetter, J. R, Jurs, P.C, Rose, S. L, *Anal. Chem*. 1986, 58, 860.

250. Weimar, U. et al, *Sens. Actuators* B, 1990, 1, 93.

251. Carey, W. P. et al, *Sens. Actuators*,1986, 9, 223.

252. Wienke, D, Danzer, K, *Anal. Chim. Acta* 1986,184, 107.

253. Kowalski, B. R, *Anal. Chem*. 1975, 47, 1152.

254. Kowalski, B. R, *Anal. Chem*. 1980, 52 ,112 R.

255. Frank, E, Kowalski, B. R, *Anal. Chem*. 1982, 54 , 232 R.

256. Beebe, K. R, Kowalski, B. R, *Anal. Chem*. 1987, 59 , 1007 A.

257. Osten, D.W, Kowalski, B. R, *Anal. Chem*. 1985, 57 , 908.

258. Persaud, K. C.; Dodd, G. *Nature*, 1982, 299, 352.

259. Karl Cammann, Bernd Ross, Andreas Katerkamp, Jörg Reinbold4, Bernd Gründig, Reinhard Renneberg, Chemical and Biochemical Sensors, in Ullmann's Encyclopedia of Industrial Chemistry, 2002, Wiley

260. Rocha, Santos, T. A. P.; Duarte, A. C.; Oliveira, J. A. B. P, *Talanta* 2001,54, 383-388.

261. Williams, D.; Pappas, G. *Field Anal. Chem. Technol*. 1999, 3, 45-53.

262. Chapman, PJ, Vogt, F, Dutta, P, Panos G. Datskos, Gerald L. Devault, and Michael J. Sepaniak, *Anal. Chem*. 2007, 79, 364-370

263. Zampolli, S.; Elmi, I.; Sturmann, J.; Nicoletti, S.; Dori, L.; Cardinali, C. *Sens. Actuators*, B 2005, 105, 400.

264. Grayson, M. A, *J. Chromatog. Sci*. 1986, 24, 529

265. Ashcroft, A. E. et al, *Am. Lab*, 1985, Aug, 59

266. Slack, R. W, Heim, A. C, *Am. Lab*, 1986, Aug, 80

267. Smith, S. L, *J. Chromatog. Sci*. 1984, 22, 143.

268. Lee, E. D, Henion, J. D, *J. Chromatog. Sci*. 1985, 23, 253

269. Alcock, N. J, Biomed. *Mass Spectros*. 1982, 9, 499.

270. Selby, M. and Hieftje, G. M, *Am. Lab*. 1987, Aug, 16.

271. Houk, R. S, *Anal. Chem*. 1986, 58, 97A

272. Davies, I. L. et al, *Anal. Chem*. 1988, 60, 683A

273. Smith, R. D, Barinaga, C. J, and Udseth, H. R, *Anal. Chem*. 1988, 60, 1948.

274. Goodall, D. M, Lloyd, D. K, and Williams, S. J, *Liq. Chromatog. Gas Chromatog*. 1989, 8, 788.

275. Uden, P. C, *Spectroscopy* 1989, 4, 44.

276. Armstrong, J.T *Anal. Chem*. 1988, 60, 1159A.

277. Dorn, H. C, *Anal. Chem*. 1984,56, 747A.

278. Bruno, TJ. *Separation and Purification Methods*, 2000, 29(1), 63-89.

279. Sweedler, JV. *Analytical and Bioanalytical Chemistry*, 2002, 373(6), 321-322.

280. Fetzer, JC. Hyphenated techniques for chromatographic detection. Handbook of Spectroscopy,2003, 2, 381-440.

281. Meuzelaar, Henk L. C.; Mcclennen, William H.; Dworzanski, Jacek P.; Sheya, Sue Anne; Snyder, A. Peter; Harden, Charles S.; Arnold, Neil S. Hyphenated techniques; Next generation of field-portable analytical instruments? Field Screening Methods for Hazardous Wastes and Toxic Chemicals, Proceedings of an International Symposium, Las Vegas, Nev, Feb. 22-24, 1995, 38-47.

282. Haseth, De, James, A. *Spectroscopy*, United States,1987, 2 (10), 14-16.

283. Hirschfeld, T, *Anal. Chem*. 1980, 52, 297A.

284. Moody, JR, *Anal. Chem*. 1985, 57, 1374A.

285. Geladi, Esbensen, P, K, *J. Chemometr*, 1990, 4, 337-354

286. Esbensen, K, Geladi, P, *J. Chemometr*, 1990, 4, 389-412

287. Geladi, P, *Spectrochimica Acta Part* B, 2003, 58, 767-782

288. Mark, H, *Analytica Chimica*, 1989, 223, 75-93

289. Lavine, B, *Anal. Chem,* 2000, 72, 91R-97R

290. Jurs, P, Isenhour, T, Chemical Applications of Pattern Recognition, Wiley, New York, 1975.

291. Kowalski, B,Ed,, Chemometrics, Theory and Application, American Chemical Society, Washington, DC, 1977.

292. Massart, D, Dijkstra, A, Kaufman, L, Evaluation and optimization of laboratory methods and analytical procedures, A Survey of Statistical and Mathematical Techniques, Elsevier, Amsterdam, Netherlands, 1978.

293. Malinowski, E, D. Howery, Factor Analysis in Chemistry, Wiley, New York, 1980.

294. Varmuza, K, Pattern Recognition in Chemistry, Springer, Berlin, 1980.

295. Kowalski, B,Ed,, Chemometrics. Mathematics and Statistics in Chemistry, Kluwer Academic, Dordrecht, 1984.

296. Sharaf, M, Illman, D, Kowalski, B, Chemometrics, Wiley, New York, 1986.

297. Deming, S, Morgan, S, Experimental Design, A Chemometric Approach, second ed, Elsevier, Amsterdam, 1993.

298. Carroll, D, Green, P, Chaturvedi, A, Mathematical Tools for Applied Multivariate Analysis, Academic Press, San Diego, 1997.

299. Adams, M, Chemometrics in Analytical Spectroscopy, The Royal Society of Chemistry, Cambridge, UK, 1995.

300. Brereton, R, Chemometrics, Data Analysis for the Laboratory and Chemical Plant, Wiley, Chichester, 2003.

301. Young, Simon S. Computerized Data Acquisition and Analysis for the Life Sciences. Cambridge University Press. 2001

302. Gardner, JW, Bartlett, P N ,1999, Electronic noses, principles and applications. Oxford University Press, Oxford

303. Scott, SM, James, D, Ali, Z, *Microchim Acta* 2007, 156, 183–207

304. Martens, H, Næs, T, Multivariate Calibration, John Wiley & Sons, 1992

305. Zellers, ET, Pan Tin-Su, Patrash, Samuel J, Mingwei Han and Stuart A. Batterman, *Sensors and Actuators* B, 12 ,1993, 123-133

306. Wold, S, Pattern recognition by means of disjoint principal components models, *Pattern Recognition*, 8 ,1978, 127 -139.

307. Park, J.; Zhang, G. Z.; Zellers, E. T. *American Industrial Hygiene Association Journal* 2000, 61, 192-204.

308. Grate, J. W. *Chemical Reviews* 2000, 100, 2627-47.

309. Carey, W. P.; Beebe, K. R.; Sanchez, E.; Geladi, P.; Kowalski, B. R. *Sensors and Actuators* 1986, 9, 223-34.

310. Zellers, E. T.; Park, J.; Hsu, T.; Groves, W. A. *Analytical Chemistry* 1998, 70, 4191-201.

311. Karjalainen, E, Karjalainen, U, Data Analysis for Hyphenated Techniques, Elsevier, Amsterdam, Netherlands, 1996.

312. Martens, H, Næcs, T, Multivariate Calibration, Wiley, Chichester, 1989.

313. Beebe, K, Pell, R, Scasholtz, M.-B, Chemometrics, A Practical Guide, Wiley, New York, 1998.

314. Jiang, Jian-Hui, Yuki hiro Ozaki, *APPLIED SPECTROSCOPY REVIEWS* Vol. 37, No. 3, pp. 321–345, 2002

315. Maeder, M. *Anal. Chem*. 1987, 59, 527–530.

316. Malinowski, E.R. *J. Chemom*. 1992, 6, 29–40.

317. Kvalheim, O.M.; Liang, Y.Z. *Anal. Chem*. 1992, 64, 936–946.

318. Liang, Y.Z.; Kvalheim, O.M. *Chemom. Intell. Lab. Syst*. 1993, 20, 115–125.

319. Manne, R.; Shen, H.L.; Liang, Y.Z. *Chemom. Intell. Lab. Syst*. 1999, 45, 171–176.

320. Jiang, J.-H.; Ozaki, Y. *Anal. Sci*. 2001, 17, i705–708.

321. Sanchez, F.C.; van den Bogaert, B.; Rutan, S.C.; Massart, D.L. *Chemom. Intell. Lab. Syst*. 1996, 34, 139–171.

322. Karjlainen, E.J. *Chemom. Intell. Lab. Syst*. 1989, 7, 31.

323. Tauler, R.; Casassas, E. *Chemom. Intell. Lab. Syst*. 1992, 14, 305.

324. Windig, W.; Guilment, *J. Anal. Chem*. 1991, 63, 1425–1432.

325. Gemperline, P.J. *J. Chem. Info. Computer Sci*. 1984, 24, 206–212.

326. Vandeginste, B.G.M.; Derks, W.; Kateman, G. *Anal. Chim. Acta* 1985, 173, 253–264.

327. Paatero, P. *Chemom. Intell. Lab. Syst.* 1997, 37, 23–35.

CHAPTER 2

A Decision-making Rule for Automated Vapor Identification by Reference Library Search

2.1 Introduction

Chemical sensors are generally characterized by high sensitivity, low cost, low power consumption, small size and great potential to combining into a sensor array detector for selective analysis.[1] All these features make the chemical sensor array an ideal candidate for detector in a microfabricated VOC analyzer and explain its wide use in "electronic nose" devices.[2] The hyphenation of chemical sensor array with chromatography holds the most promise for developing novel microanalytical systems that are capable of fast and in-situ analysis of complex VOC mixture.[3, 4] Like mass spectrometer, Infrared (IR), and other spectroscopic detectors commonly employed with hyphenated chromatographic method, sensor array also belongs to the type of multiple-channel detector. One important difference is that the number of independent data acquisition channels in a sensor array is much smaller than that of a spectroscopic detector which measures light absorbance or emission at hundreds of wavelengths simultaneously. Spectroscopic detector, sensor array detector, and other multiple-channel detectors share the same feature: the multivariate data generated in the form of spectrum or response pattern contain rich information that can be used for chemical structure elucidation.

Spectroscopic techniques, particularly mass spectrometry, Infrared spectroscopy, and nuclear magnetic resonance, have been widely employed as the major qualitative analysis tools for material identification in organic chemistry, biology, forensics, and environmental science. Development and application of novel sensor array technology will definitely benefit from the understanding of the chemical structure elucidation in spectroscopy.

Spectrum evaluation for the purpose of chemical structure elucidation can be divided into two categories: identification of a compound assuming its spectrum is already known; Interpretation of spectral data in terms of its chemical structure when the spectrum of the unknown is not available. Accordingly spectral interpretation can be divided into direct and indirect database approaches. Direct database methodology involves file search or database search in which the spectrum of an unknown compound is compared against a library of reference spectra of *known* compounds. After several decades of extensive research massive spectral databases have been complied for mass spectrometry,[5, 6] Infrared,[7-11] NMR,[12] and even for UV/Vis.[13] A basic assumption in library search systems is the hypothesis "similar spectra correspond to similar structure" [14]. Generally a search for spectra in a library is guided by numerical similarity between two spectra used to find the reference spectra most similar to the spectrum of the unknown. The output usually comprises a list of spectra ("hit list") that are most similar to that of the unknown compound, together with a metric indicating the degree of similarity in each case. The idea of matching unknown spectra against a spectral library is illustrated in Figure 2-1. The search is automatically programmed and performed in a computer. In practice, however, it is almost unlikely that an exact match can be found between the unknown spectrum and a

reference spectrum. Exact reproduction is always difficult, if not possible, not to mention there is not a database that can include reference spectra of all unknown compounds. Systematic and random variations from sample, instrument, operator and environment are more or less a common phenomenon that causes the difference in spectra of a compound calibrated and measured on the same detector. Therefore, the provision of a list of candidates instead of the single best match can somewhat mitigate the uncertainty caused by the spectral variations mentioned above. In general, the effectiveness of a spectral library search depends on the quality of the given spectrum, the content and quality of spectral library, and the similarity criterion employed for the search. Of course the fundamental restriction is the limited structural information present in the spectrum, e.g., isomers often can not be distinguished from each other by their mass spectra.

Compared to the direct library search for chemical structure elucidation, indirect database search for spectral interpretation is used much less frequently due to its requirement for expert knowledge to correlate spectral details to chemical sub-structures. If the unknown compound has never been identified there will be no reference spectrum that can match to the spectrum of the unknown. In this case, it may be possible to use the interpretative system to deduce chemical sub-structures by finding sub-spectra in the library that are similar to features in the spectrum of the unknown. The chemical structure information in a sensor array response pattern is barren compared to the information contained in IR, NMR or mass spectrum. Thus, interpretative library search can hardly be applied to sensor array and will not be mentioned any more in this dissertation. Focus will be placed on direct library search for unknown compound identification. For a detailed review of indirect computer-aided chemical structure elucidation in spectroscopy, refer to

the articles published by Warr and Richert. [15,16] Although spectroscopy tends to provide more qualitative information, mass spectrometry will be reviewed for spectral matching algorithms because of its dominant role in VOC monitoring and analysis (see Chapter 1).

2.1.1 Matching Mass Spectra. For the last several decades massive mass spectrum databases have been complied and sophisticated computer searching programs have been developed for the purpose of identifying components of complex mixtures.[17, 18] Nowadays spectral searching software and libraries of mass spectra are available for nearly all GC/MS systems.[19] As mentioned earlier, a hit list for each unknown component is generated after the search, usually arranged in descending order according to the degree of agreement between the mass spectrum of the unknown and the mass spectrum from the reference library. The degree of agreement is measured by some similarity parameter or fit parameter. The fit parameters are normally scaled in percent so that when perfect agreement exists between the unknown and reference mass spectra, a fit parameter of 100% is reported. It should be noted that the fit parameters of hit list should not be taken seriously. Instead it is wise to look down the list where it may well be possible to find a better candidate than that the computer has picked out.

Of the numerous mass spectrum matching algorithms developed for spectral retrieval, the well-known Cornell algorithm will be reviewed because of its intensive commercial implementation.[17, 20] Cornell algorithm was developed by McLafferty et al in Cornell University in early 1970s.[21] It includes probability-based matching (PBM) and self-training interpretive and retrieval system (STIRS) and has been improved continuously over the last three decades. [22-24] Like other searching algorithms, PMB uses well-designed and powerful algorithms that are very fast considering the size of reference

library through which they have to search instantly. To speed up the search, reference data will be pre-filtered by a screening procedure, and an abbreviated reference data and a small fraction of unknown spectrum will be generated for an extensive similarity match. [25] The PMB method uses significant peaks. Due to the intrinsic nature of mass spectrometry, those that occur less frequently are more diagnostic and more confidence can be placed on a match of these significant peaks. These peaks tend to locate on the rear of a mass spectrum and have a large mass/charge value. To reflect the added importance of peaks at large mass, spectra are first weighted by the addition of the mass to the abundance of each peak.

The PMB can perform forward and reverse search for pure analyte and mixture analysis. A forward search looks for the unknown spectrum in the reference database and the reverse search looks to find the best reference spectrum in the unknown. The forward search is faster: reference spectrum that does not contain peaks of the unknown can be eliminated from candidate pool. But it demands unknown spectra of pure compounds to produce good results. Forward search will not work correctly if the unknown spectrum contains peaks from unwanted impurities or of a mixture. Reverse search strategy can handle this mixture spectrum by matching reference spectra to the unknown, whereby the hit list shows the best reference spectrum in the library as found in the unknown data. Peaks in the unknown that do not appear in the reference spectrum do not downgrade the fit parameter as they would in the forward search because they may come from impurities and other compounds in a mixture. Once the most significant component has been identified, its reference spectrum can be subtracted from the unknown mixture and the reverse search starts again to identify other components in the unknown mixture.

Instead of using similarity criteria based on correlation coefficient or Euclidean distance.[26] PMB uses similarity criterion that is based on so-called uniqueness values for masses and abundances. The uniqueness of a signal is equivalent to the information content as defined in information theory and is given by the negative logarithm of the probability of the occurrence of this signal. This means that a low probability corresponds to a high uniqueness. Probabilities for masses and intensity intervals have been estimated from the database used. Matching spectra is performed by using selected peaks that exhibit highest uniqueness values for mass and intensity.

To handle the variability problem in spectrum measurement, PMB uses peak flagging (stepwise removing peaks of low abundance) and abundance scaling to permit matching of an unknown spectrum against a large database not restricted to spectra taken under the same experimental condition. More details of PMB will not be iterated here. Refer to Reference 21 for more information.

It should be reminded that PMB is not the only searching algorithm available. Its implementations rely on the commercialization effort from GC/MS instrument manufacturers that try to bind a mass spectrum database with their instruments. The abilities of the PBM and other spectrum matching algorithms to provide correct identifications of unknown mass spectra were evaluated by Stein and Scott.[27] This evaluation, which is the most complete and exhaustive comparison to date, [28] evaluated five different searching algorithms, and used a large test set of 12,592 alternate spectra and a reference library of 62,235 compounds. Both sets of mass spectra were obtained from the 1992 release of the National Institute of Standards and Technology/Environmental Protection Agency/National Institutes of Health mass spectral database (NIST database).

The algorithms tested were probability-based matching (PBM), dot-product,[28] similarity index,[29] Euclidean distance, and absolute value distance. Most algorithms were optimized by varying their mass weighting and intensity scaling factors. Rank in the list of candidate compounds was used as the criterion for accuracy. The best performing algorithm (75% accuracy for rank 1) was the dot-product function that measures the cosine of the angle between spectra represented as vectors. Other methods in order of performance were the Euclidean distance (72%), absolute value distance (68%), PBM (65%), and Hertz et al. (64%).

These various search algorithms consist of a data preprocessing step before spectral similarity calculation. Data preprocessing is necessary in order to compensate for spectral variations. Data preprocessing consists of peak selection, peak intensity scaling, and weighting of peak intensity by mass position. Obviously these techniques are very exclusive and restricted only to mass spectrum where selectivity can be achieved at high mass values. None of them can be adapted for use in spectroscopic or sensor array field. Each sensor in a sensor array is an integral part of the sensor array and contributes equally to the response pattern. The limited number of sensors in an array makes it very hard to exclude some from pattern comparison. Consequently, PMB, as a spectrum matching algorithm designed specifically for mass spectral library search, cannot be applied to sensor array study.

2.1.2 Pattern Fidelity. In most sensor array applications, an unknown vapor is identified by reference library search for its measured response pattern. Variation in response pattern impedes direct comparison of measured response pattern to its reference pattern. Thus, variability of response pattern must be considered before identity of the

unknown can be assigned by pattern matching. Accounting for pattern variation in pattern matching is especially important when vapor identity can be prescreened by using retention information for a chromatographically resolved peak (Retention index should be used to accommodate the fluctuation of retention time). Instead of search through every reference pattern in the library, pattern recognition can be narrowed down to the few candidate vapors whose retention indices fall into the elution window of the chromatographically resolved unknown peak. Therefore the effectiveness of pattern recognition can be greatly enhanced by combining retention information into qualitative analysis. The inclusion of retention time into pattern recognition and the subsequent retention time window approach was reported by Zellers et al. with a portable GC.[30] Although it is possible to use retention index alone to identify a chromatographically resolved single peak, pattern recognition is still indispensable in order to handle the situations where co-elution with an unknown and uncalibrated interference might occur. Pattern recognition should be made by taking into the consideration of pattern variation caused by interference in the sample, or disturbance in instrumentation and environment. Frequently the pattern is changed so much that the peak is misidentified. Therefore, we need to have a decision rule of vapor identity assignment and to set up a threshold of pattern variation beyond which vapor identity cannot be reliably assigned. With this decision rule one would be able to judge whether the peak under study is due to the expected analyte, and accordingly correct peak identification would be allowed.

2.1.3 Mahalanobis Distance. One of the common metrics used to measure similarity between spectra is Euclidean distance. In geometric view, a sensor array response vector corresponds to a point in a multi-dimensional space whose dimensions correspond to the

sensors in the array. Euclidean distance is the point-to-point straight line distance which can be easily calculated from the difference of two response vectors. Obviously the variation of sensor responses at each dimension is not considered in such simple dissimilarity metrics. That is the reason PBM and other spectral matching algorithms rely on a data preprocessing step to compensate for variations. A better statistic distance, Mahalanobis distance, is used here because of its ability to take into consideration the variability of sensor responses. To understand how this works, consider that, when using Euclidean distance, the set of points equidistant from a given location is a sphere. The Mahalanobis distance stretches this sphere to correct for the respective scales of the different variables, and to account for correlation among variables.[31] Given the i^{th} sample (i=1,…n) of a vapor calibrated on the j^{th} sensor (j=1,…k) of a k-sensor array, a calibration matrix $\boldsymbol{X}$ can be formed by including n *normalized* response vectors as its row (response pattern will be generated as a result of normalization of response vector). Each row of $\boldsymbol{X}$ (X_1, X_2…X_n) corresponds to a point in a k-dimensional hyperspace; the variations of response pattern measured for the vapor can be visually displayed by the distribution of n points in the hyperspace. The location of the vapor is indicated by the centroid of $\boldsymbol{X}$, i.e., the mean of the rows of $\boldsymbol{X}$, $\overline{X}$. Mahalanobis distance of the i^{th} sample to the vapor represented by $\overline{X}$ is formulated as below:

$$d_i^2 = (X_i - \overline{X})' \boldsymbol{M}^{-1} (X_i - \overline{X}) \qquad \text{Eq. 2-1}$$

$\boldsymbol{M}$ is the variance-covariance matrix of $\boldsymbol{X}$, which is calculated as below:

$$\boldsymbol{M}=\frac{1}{n}(\boldsymbol{X}\text{-}\mathbf{1}_{\mathrm{n}}\overline{X})'(\boldsymbol{X}\text{-}\mathbf{1}_{\mathrm{n}}\overline{X}) \qquad \text{Eq. 2-2}$$

where $\mathbf{1}_{\mathrm{n}}$ is a sum vector of dimension *n* which is a column vector composed of *n* elements equal to 1. Mahalanobis distance from each calibration sample to the mean response vector is calculated as a measure of the similarity between the sample and the vapor. Based on the distribution of all these distance values, the variation of the vapor's response pattern can be quantified. A threshold distance value can thereby be established for this vapor. The identity of an unknown pattern will not be assigned to the vapor if the Mahalanobis distance of the unknown response pattern from the centroid of the vapor is larger than this threshold distance established. Otherwise it is classified to the vapor. The application of Mahalanobis distance to capture sensor response variation and the threshold used for binary classification is illustrated in Figure 2-2. Intuitively, Hotelling's T-square distribution seems applicable to the description of the Mahalanobis distances calculated above for a vapor. But it demands multivariate normal distribution of sensor array responses which is not guaranteed in practice. Chemometric equivalence, the residual of principal component model that can be described by F distribution,[32] will be employed instead in this study. I will introduce the principle of the decision-making rule developed for response pattern matching in the next section, followed by its applications to pattern fidelity evaluation for a sensor array measured under different environmental conditions.

2.2 Computational Method

The use of an upstream separation column limits the number of vapors presented to the sensor array simultaneously. This, in turn, reduces the problem of vapor recognition to a series of simpler analyses applied to each retention time window individually. For this

study, all of the vapors could be chromatographically resolved by carefully adjusting the pressure tuning and temperature programming parameters of the separation module. Therefore the confirmation of vapor identity reduces to one of assessing the fidelity of the measured response pattern to that in the calibration library. That is, the goodness of fit of a unknown sample to its calibration set must be tested in order to conclude with confidence that the resolved peaks observed with the array are indeed attributable to the vapor expected to elute at the given retention time.

The approach can be illustrated by a one-principal-component classification model as shown in Figure 2-3, constructed from a calibration data set using principal components analysis. The threshold of maximum distance from a subsequent sample to the centroid of the calibration set in multi-dimensional space (i. e., the Mahalanobis distance) can be computed at a certain significance level in term of the residual error after projection of the sample to the principal component (PC) axes. If a new sample falls within the boundary established by this threshold, the sample is assigned with confidence the identity of the vapor corresponding to that model. Otherwise it is rejected.

We start with a calibration data set for a target vapor that can be expressed in matrix form as follows:

$$\underset{n\times m}{X} = \begin{bmatrix} x_{11} & x_{12} & . & x_{1m} \\ x_{21} & x_{22} & . & x_{2m} \\ . & . & . & . \\ x_{n1} & x_{n2} & . & x_{nm} \end{bmatrix} \qquad \text{Eq. 2-3}$$

where x_{nm} is the response of the m^{th} sensor to the n^{th} calibration sample of the target vapor.

PC modeling of this data set separates the data into a model matrix and a residual matrix. The latter is used to calculate a residual standard deviation (RSD, geometrically the Mahalanobis distance) when dealing with vector quantities, which is used for determining the range of allowable variability in assessing the response pattern fidelity of subsequent test samples.

Cross-validation was used to obtain models with maximum predictive ability. A common cross-validation technique, which is well suited for small data sets, is the leave-one-out procedure in which one row at-a-time is deleted from the data matrix *X* and the PC(s) that best account for the variance in the remaining data are calculated. The resulting *n* PC models (the number of models is the same as the number of samples) are used to predict the response patterns for the left-out samples (models may contain one or more PCs). A metric called the predicted residual error sum of squares (PRESS) measures the prediction error of each PC model. The deviation between the actual and predicted values is used to estimate an overall prediction error. The model providing the minimum prediction error is finally calculated with all samples included. By this procedure, all samples are utilized both for calculating and for validating the model.

To assess the fidelity of a subsequent (unknown) sample to the established model, the following equation is used:

$$e_i(v) = X_i' - \bar{X}'(v) - \sum_{a=1}^{A} t_{ia} p_a' \qquad \text{Eq. 2-4}$$

where $e_i(v)$ is the residual (vector) of the unknown sample after being fitted to the model for vapor v; X_i' is the transpose of the unknown sample response vector of m elements (m is

$\bar{X}'(v)$

the number of sensors in the array); is the transpose of the mean vector of the vapor *v*; *A* is the dimension of the model (i.e., the optimal number of principal components) determined by cross-validation; t_{ia} is the "score" of sample *i* on the principal component *a*, and p_a is the corresponding loading vector (weighting of all variables on principal component *a*).

The Mahalanobis distance S_i is the distance of an unknown sample *i* to the model under consideration, which can be calculated via:

$$S_i = \sqrt{e_i'(v)\, e_i(v) / (m - A)} \qquad \text{Eq. 2-5}$$

where *m* is the number of sensors, $e_i(v)$ is the residual vector of the unknown sample *i* after fitting to class for vapor *v* and $e_i'(v)$ is the transposed vector of $e_i(v)$. The division by *(m-A)* provides a distance measure that is independent of the number of variables and corrected for the loss of freedom due to the fitting of *A* principal components.

The S_i values for all *n* samples in the calibration set can also be calculated using Eq. 2-5 and collected in a distance vector *S* of dimension *n×1*. The mean *RSD,* S_v, of the model for vapor class *v* is defined by the following equation:

$$S_v = \sqrt{S'S / (n - A - 1)} \qquad \text{Eq. 2-6}$$

where S' is the transpose of distance vector *S*. The division by (*n*-*A*-1) gives a scale that is also independent of the number of samples and corrected for the loss in degrees of freedom due to the mean-centering and fitting of *A* principal components in Eq. 2-4.

Comparison of the RSD for an unknown sample *i* (calculated by Eq. 2-5) to the

mean RSD for the class *v* (calculated by Eq. 2-6) gives a direct measure of its similarity to the model. An *F* statistic was used for the comparison of S_i^2 and S_v^2. The degrees of freedom used to obtain the critical F-value are (*m*-*A*) and (*m*-*A*)(*n* -*A*-*1*), respectively, for S_i^2 and S_v^2. The upper limit for the RSD for samples that belongs to the class can thus be calculated:

$$S_{\max}^2 = S_v^2 F_{crit} \qquad \text{Eq. 2-7}$$

F_{crit} is usually determined at a significance level of 1% or 5% (i.e., p = 0.01 or 0.05). The latter level is a more stringent boundary, allowing 1 out of 20 samples that fit the model, on average, to be mis-identified as outliers. In the latter case, only 1 out of 100 samples that fit the model is rejected.

So far a decision rule of identity assignment was developed for a target vapor. If S_i of an unknown response pattern is less than or equal to S_{max}, then this sample is assigned the identity of the vapor described by that particular model. If $S_i > S_{max}$, then the sample is rejected indicating that the response pattern has become distorted enough to suggest that the target vapor is contaminated with another vapor or that the response pattern is not due to the target vapor.

2.3 Experimental Section

Experimental data were generated by C.-J. Lu et al. in a chamber study of a portable GC device equipped with a surface acoustic wave sensor array detector. The experiment was introduced briefly below in order to illuminate the practical application of

the decision rule for response pattern fidelity evaluation. For a detailed description of the experiment, refer to Reference 36. The sensor array was designed and fabricated by Sandia National Laboratories (Albuquerque, NM) and consists of four surface acoustic wave (SAW) sensors.[33] One of the sensors is used as a reference and is uncoated, while the other three were coated, respectively, with polyisobutylene (PIB), ethyl cellulose (ECEL), and a co-polymer of polydimethylsiloxane with hexafluorobisphenol A (BSP3).[34, 35] To evaluate the influence of humidity and temperature on sensor responses, the sensor array was put in a stainless steel chamber so that the sensor array can be exposed to the testing atmosphere under controlled temperature and humidity. The experimental setup can be referred to Figure 2-4.[36] In order to generate low concentrations of multiple vapors, a high-concentration (i.e., ppm-range) vapor mixture was prepared in a Tedlar® bag and delivered into a purified air stream (~20 L/min) at a relatively low continuous flow rate (5-20 mL/min), and then got diluted to ppb levels.

The air source was compressed air passing through a series of traps to remove moisture and trace organic species. The stream was then divided into two branches, each controlled by a rotameter equipped with a needle valve. One branch was routed through two Greenburg-Smith impingers (connected in tandem each containing 500 mL of distilled water) and then returned to the dry air stream. Humidity level in the chamber downstream was adjusted by changing flow rates and monitored by a humidity/temperature meter probe (Oakton 35612, Cole-Parmer, Vernon Hills, IL). The other branch of purified air was routed into a sealed plastic container and created a pressure that pressed a Tedlar® bag containing testing atmosphere of target vapor. A deactivated fused silica tube (1.5-m long, 0.52 mm i.d., Supelco, PA) was inserted into the Tedlar® bag and connected to the main

dilution flow stream that eventually passed through the sensor array. The target vapors include 15 volatile organic contaminants commonly found in indoor air quality studies. These compounds are listed in Table 2-1.

Data acquisition was facilitated by the industrial standard RS-232 interface and a virtual instrument programmed in LabVIEW 5.0 (National Instrument Inc., Austin, TX). Data were then imported to Grams 32 (Thermogalactics, Inc. Salem, NH) for the peak integration. Sensor response vectors were preprocessed to remove magnitude effect before fidelity test can be performed. Response pattern of each target vapor was obtained by normalizing response vector to a sum of 1. The principal component classification model was established in Matlab 7 (Natick, MA). The Matlab codes used for decision-making were attached in the Appendix of this chapter.

2.4 Results and Discussion

2.4.1 Reproducibility of Response Pattern. Short-term variability of response pattern generated by the 3-sensor SAW array was evaluated for each of the 15 target vapors. For each of the vapors, the sensor array was calibrated by using five to seven vapor samples spanning a concentration ranging from 10ppm to 170ppm. Then, eight samples of testing atmosphere containing all 15 compounds at fixed concentrations were introduced into the test chamber and a set of eight consecutive measurements were collected over a 1-hour period. During the course of this test, the chamber temperature was kept constant at 25ºC (with a fluctuation ±1 ºC) and the relative humidity was 0% (< 4% RH, referred to 0%). Principal component classification models were setup for each of vapors based on their calibration data. The mean standard deviation (MSD) of each vapor and their

threshold Mahalanobis distance are tabulated in Table 2-1. The degree of response pattern variability can be judged by the magnitude of the MSD values. Generally the pattern variability of these 15 vapors are comparable, with MSD ranging from 0.565 for 2-methylheptane to 0.944 for perchloroethylene. Smaller MSD corresponding to lower variability indicates a low level of variability observed for 2-methylheptane. Threshold Mahalanobis is also calculated for each vapor based on the number of calibration samples and F distribution of the RSD values of individual samples. This threshold is set at a significance level of 0.05, meaning there may be one error in every twenty classifications. If the MSD is large, a high threshold, i.e., a smaller significance value like 0.001 should be employed.

These thresholds were used to test the short-term reproducibility of sensor array by evaluating the fidelity of the eight testing response patterns measured for each vapor to its calibrated patterns. The testing results are listed in Table 2-3. For every vapor, its identity assignment rule was tested using both the training data and test data. The Actual Error Rate, which is determined by the proportion misclassified in the testing sample, is 13%. Among the 120 test samples (8 replicates of 15 vapors) 111 were correctly recognized and there were at most two errors in recognition for any vapor (7 of the 15 vapors were recognized without error). The Apparent Error Rate, which is defined as the fraction of observations in the training set that are misclassified by the established rule, is as low as 2% (only two of the total 91 training samples were misclassified). The result of the testing on training set is obviously optimistic because the data used to build the rule are also used to evaluate it. The low error rate of fidelity testing shows excellent short-term response pattern reproducibility for all vapor investigated. This finding is essential because it means the

pattern disturbance is ignorable during a normal period of sensor array calibration and measurement. The fidelity of response pattern will be retained during this period and the classification model can be safely used for vapor identification.

2.4.2 Humidity Effect on Pattern Fidelity. The sensor array was challenged with the 15-vapor mixture at different relative humidity (RH) level: 0, 45, and 90% RH, while the temperature was kept constant at 25ºC. The effect of humidity on response pattern was evaluated for each of the 15 vapors by testing the fidelity of pattern measured at 45% and 90% RH to the patterns at 0% RH. Four duplicate measurements were made at 0% RH for the 15-vapor mixture, and principal component classification model was separately established for each vapor. Each classification model was then applied to the duplicate patterns measured at higher RH. To more rigorously assess the consistency of the response patterns as a function of humidity, fidelity tests were run using data derived from the pooled set of array responses at 0% RH as a training set and response patterns obtained at 45 and 90 %RH as test sets. F-tests performed on the 30 response patterns at the higher RH levels (average of duplicate exposures at each elevated RH level for each vapor) indicated that 29 had no significant difference in pattern (the exception was α-pinene at 90 %RH).[35] The independence from humidity effects observed in this data set is consistent with results reported from other studies of polymer-coated SAW-sensor based instruments employing adsorbent preconcentrators without a separation column.[36]

2.4.3 Temperature Effect on Pattern Fidelity. The effect of temperature on response patterns were studied on 15 vapors by comparing the response patterns measured at 30 ºC to their patterns measured at 25 ºC. Duplicate patterns measured at 30 ºC for each vapor were averaged to get a representative mean response pattern. For most compounds

the average response pattern measured at 30 ºC was fitted quite well at significance level of 0.05 to their class models established for a temperature of 25 ºC, which means that temperature imposes no effect on their response patterns. The only exception is vapor D5, a component of anit-perspirants, that can't be fitted even at a significance level of 0.001. Score plot of principal component analysis of D5 measurements shows that the training data are clustered together while the testing data scatter loosely around the training data (plot not shown). These findings are also consistent with those reported elsewhere demonstrating the independence of response patterns for most vapors from polymer-coated SAW sensor arrays over modest temperature ranges.[37, 38]

2.5 Conclusion

Mass spectrum, IR/Raman and NMR spectrum are used extensively for unknown chemical identification because they convey rich information about chemical structure. Massive spectral databases are available for material identification. As a miniaturized "spectrophotometer", a chemical sensor array produces a characteristic response pattern that is also multivariate in nature. Although the amount of qualitative information in a sensor array response pattern is not yet comparable to that of a spectrum, and the response pattern generated from a simple sensor array cannot provide sub-structural information about a chemical, it does not impede chemical sensor array from wide application in electronic nose devices and other microanalytical system. As compared to spectroscopy, sensor array can easily achieve selective analysis by judicious sensing material selection. With more sensitive and selective sensing material available, and more independent sensor channels to be included in a microfabricated sensor array, it is not a wild imagination that

in the near future a microsensor array can be as powerful as a spectrometer in terms of chemical identification, which also entails a massive reference response pattern library as well as a powerful library search algorithm.

To implement automated matching of an unknown pattern against a large selection of reference patterns, it is necessary to adopt a pattern matching algorithm that is capable of fast searching through a large database and generating the results by taking into consideration the uncertainty of spectrum measurement. In mass spectrometry, the dominating algorithm is the so-called Cornell algorithm that has been commercialized by Agilent™ and bound with its MS instruments. This algorithm handles spectrum variations by using different data preprocessing techniques before actual spectral matching. Unfortunately Cornell algorithm is exclusive and cannot be adapted to other domains. Alternative search algorithms have already emerged and showed lower misidentification rates.

In this chapter I propose a decision-making rule for vapor identity assignment by pattern library search. Instead of the simple similarity metrics like the correlation coefficient and Euclidean distance used in spectroscopy, Mahalanobis distance is used here to measure the similarity of a measured response pattern to its reference pattern. The merit of this similarity metric is its ability to account for the variability of sensor responses. While this merit distinguishes Mahalanobis distance from other metrics, statistical reference of Mahalanobis distance requires multivariate normal distribution of sensor responses which is not usually abided in practice. Because of this, residuals of principal component class model are quantified as the Mahalanobis distance. Although they are different terminologies, they produce equivalent results on normally distributed data, only

that the residual of principal component model does not require any form of data distribution. Importantly, the residuals can be described by F distribution by choosing suitable degrees of freedom. Threshold of the allowable distance can be obtained at a pre-defined significance level and used as a criterion of vapor identity assignment. It should be noted that this Mahalanobis-distance based decision rule is flexible. If one believes the variability of sensor array response is low, a lower threshold can be adopted by using a higher significance level. Otherwise a lower significance level should be used if the variability of sensor response is high.

Appendix

A MATLAB function pafi is coded for computation of the mean residual standard deviations and the threshold distance. The usage of this function is explained in the codes. Help can be accessed in MATLAB interface by typing “help pafi”.

```
function [threshold,RSD_sample] =pafi(nfac,data,sample,alpha)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%  [threshold,RSD_sample] =pafi(nfac,data,sample,alpha)
% This function pafi is used to judge unknown sample's belonging to a group
% at a significance level alpha.
% nfac is the number of factors used to build the principal componet model
% based on calibration data.
% data is the calibration data on which a class model is build via PCA.
% sample is the unknown sample whose belonging is to be classified to the
% class model.
% alpha is the significance level usually 0.05.
%threshold is the output of maximal distance built for the class model at a
%significnce level alpha.
%RSD_sample is a matrix output of unknown samples' distance to the class
%model, the first column is sample number, the second is distance, the
%third is adjusted distance used when the model is evaluated on itself.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
A=nfac; % number of principal components
n=size(data,1); %number of row (response vectors)
m=size(data,2); % number of column (sensors)
%rows normalized to 100%
for x=1:n
   aa=sum(data(x,:));
   data_ed(x,:)=data(x,:)/aa;
end
data_ed %%%%%%%%%%%%%%% percentage normalized data
%Leave-one-out cross validation
press= ;
for a=1:(min(m,n)-1)
```

```
for x=1:size(data_ed,1)
     pp=0;
   unknown=data_ed(x,:);
   cali=data_ed;
   cali(x,:)= ;
   mean_cali=mean(cali);
   std_cali=std(cali);
   cali_ed=zscore(cali);
    coeff1,score1,latent1 =princomp(cali_ed);
   unknown_ed=(unknown-mean_cali)./std_cali;
   %e(1,:)=unknown_ed-mean(cali_ed);
   score_un=unknown_ed*coeff1;
   e=unknown_ed-score_un(:, 1:a )*coeff1(:, 1:a )';
   pp=pp+dot(e,e);
 end
 ppp= a,pp ;
 press= press;ppp ;
end
press % output PRESS to choose a
%column standardization
data_std=(data_ed-ones(n,1)*mean(data_ed))./(ones(n,1)*std(data_ed));
coeff,score,latent .= PRINCOMP(data_std);
coeff
score
latent
%%%%%A=1;
g_distance= ;
%residual= ;
%for x=1:n
   %model= 0 0 0 0 0 ;
%%%%%%%
   %for y=1:A
    %   model=model+SCORE(x,y)*COEFF(:,y)';
   %end
   model=score(:, 1:A )*coeff(:, 1:A )';
```

```
mean_rsd=sqrt(distance'*distance/(n-A-1));
f_sta=finv(1-alpha,m-A,(m-A)*(n-A-1));
threshold_rsd=sqrt(mean_rsd^2*(f_sta));%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%
df_num=m-A
df_denum=(m-A)*(n-A-1)
f_sta
threshold=threshold_rsd;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Fit an unknown sample to the PCA class model  built above with A PCs.
un=size(sample,1);
g_mean=mean(data_ed);% mean of the class
g_std=std(data_ed);% Std of the class
for x=1:un
    aa=sum(sample(x,:));
    sample_ed(x,:)=sample(x,:)/aa;
end
sample_ed
RSD_sample= ;
for j=1:un
    x=sample_ed(j,:); % unknown sample (normalized to 100%)
    x_std=(x-g_mean)./g_std;% standardizing the unknown sample according to the class
%parameters
   % end
    score_x=x_std*coeff;
    residual_x=x_std-score_x(:, 1:A )*coeff(:, 1:A )';
    rsd=sqrt(residual_x*residual_x')/(m-A);
   rsd_adjusted=sqrt(rsd^2*(n-1)/(n-A-1));
   result= j,rsd,rsd_adjusted ;
    RSD_sample= RSD_sample;result ;
end
%%%%%%END of the function%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

Figures and Tables

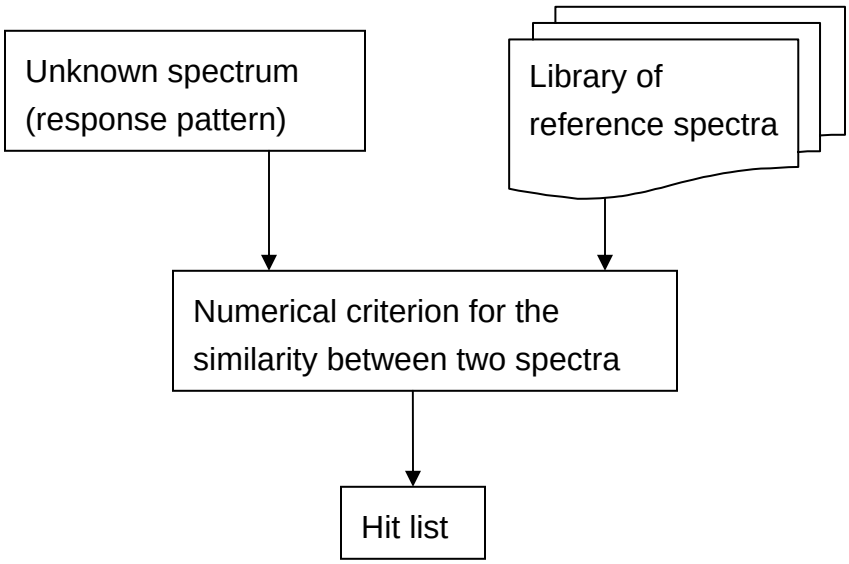

Figure 2-1. Flow chart of a spectral library search. The output hit list is an order list of reference spectra most similar to the spectrum of the unknown compound.

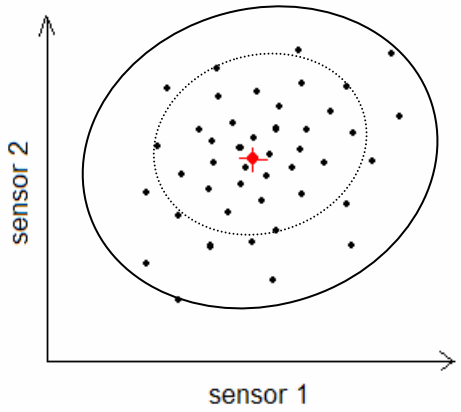

Figure 2-2. Principle of Mahalanobis distance based pattern recognition illustrated for a vapor. Black dots represent samples calibrated on the sensor array. The centroid of these samples is marked by a cross. The mean distance of all samples to their centroid is shown by the equidistant curved ellipse. A threshold distance is thereby defined by the distribution of Mahalanobis distances calculated from all calibrated samples.

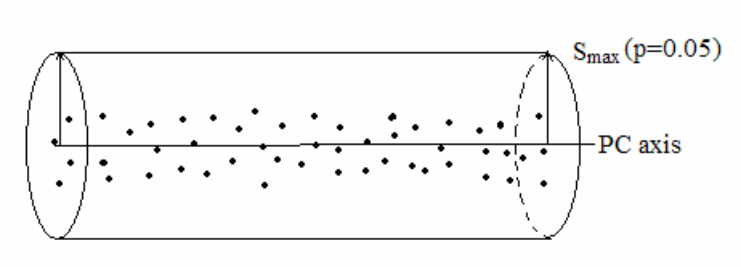

Figure 2-3. Graphical representation of the threshold Mahalanobis distance defining vapor identity. A single principal component (PC) model is assumed for illustration and the threshold is established at a predefined significance level (e.g., 0.05) on the basis of response patterns determined during calibration (i.e., training). The solid line is the PC axis, surrounded by calibration samples denoted by solid triangles. The contour of the constant distance from the PC axis defined by S_{max} is in a shape of cylinder.

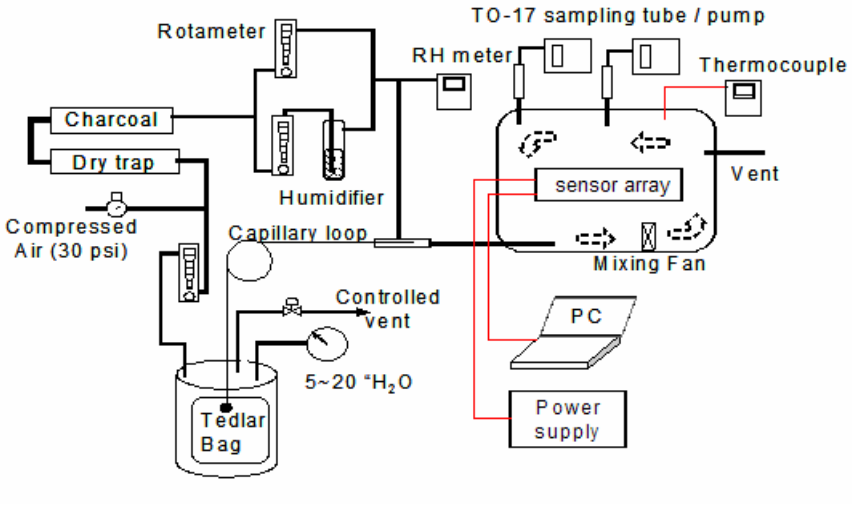

Figure 2-4. Schematic experimental setup for a chamber study of a surface acoustic wave sensor array. (Ref. 36)

Table 2-1. Threshold distance of the classification model established for each of the 15 vapors

Class	MSD	F Critical [a]	Threshold
trichloroethylene	0.868	5.14(2,6)	1.97
2,4-dimenthylhexane	0.805	5.14(2,6)	1.83
tolune	0.816	4.46(2,8)	1.72
2-methylheptane	0.565	5.14(2,6)	1.28
perchloroethylene	0.944	4.10(2,10)	1.91
chlorobenzene	0.619	4.10(2,10)	1.25
ethylbenzene	0.809	4.10(2,10)	1.64
m-xylene	0.832	4.10(2,10)	1.69
nonane	0.847	4.10(2,10)	1.72
styrene	0.822	4.10(2,10)	1.67
α-pinene	0.735	4.10(2,10)	1.49
mesitylene	0.794	4.10(2,10)	1.61
3-octanone	0.736	4.10(2,10)	1.49
d-limonene	0.816	4.10(2,10)	1.65
D5 [b]	0.578	4.10(2,10)	1.17

[a]: F statistic value at significance level of 0.05. The two numbers in the parenthesis are respectively the degree of freedom at the numerator and denominator.

[b]: D5 is the abbreviation of decamethylcyclopentasiloxane.

Table 2-2. Fidelity test of the reproducibility data

	Calibration Training set		Reproducibility Test set	
Compound	Sample number	Misclassification	Sample number	Misclassification
trichloroethylene	5	0	8	1
2,4-dimenthylhexane	5	0	8	1
tolune	6	0	8	1
2-methylheptane	5	0	8	1
perchloroethylene	7	0	8	1
chlorobenzene	7	0	8	2
ethylbenzene	7	0	8	0
m-xylene	7	0	8	0
nonane	7	1	8	1
styrene	7	0	8	1
α-pinene	7	0	8	0
mesitylene	7	1	8	0
3-octanone	7	0	8	0
d-limonene	7	0	8	0
D5	7	0	8	0
TOTAL	91	2	120	9

Reference

1. Janata, J., Josowicz, M., *Anal. Chem*. 1998, 70, 179R-208R

2. Röck, F., Barsan, N., Weimar,U., *Chem. Rev*. 2008, 108, 705-725

3. Lu, C.-J., Steinecker, W. H., Tian, W.-C., Agah, M., Potkay, J. A., Oborny, M. C., J. Nichols, H. K. L. Chan, J. Driscoll, R. D. Sacks, S. W. Pang, K. D. Wise, E. T. Zellers, Lab On A Chip, 2005, 5, 1123-1131.

4. Zampolli, S., Elmi, I., Sturmann, J., Nicoletti, S., Dori, L., Cardinali, C. *Sens. Actuators*, B 2005, 105, 400.

5. McLafferty, F. W.; Stauffer, D. B. Registry of Mass Spectral Data, 6th electronic ed.; Wiley: New York, 1994.

6. NIST/EPA/NIH Mass Spectral Database; National Institute of Standards and Technology: Gaithersburg, MD, 1992, 1998.

7. Pouchert, C. J., The Aldrich Library of Infrared Spectra, Aldrich Chemical Co., Milwaukee, WI 1981

8. Simons, W. W., The Sadtler Handbook of Infrared Spectra, Sadtler Research Laboratory, Philadelphia, PA1978

9. Schrader, B., Raman/Infrared Atlas of Organic compounds, VCH, Weinheim 1989

10. Hummel, D.O., Atlas of Polymer and Plastic Analysis, VCH, Weinheim 1991

11. Pachler, K. G., Matlok, R. F., Gremlich, H.-U., Merck FT-IR Atlas, Wiley-VCH, Weinheim, 1993

12. http://www.lohninger.com/spectroscopy/dbsurvey.html, retrieved in May 2008

13. The Sadtler Handbook of Ultraviolet Spectra, Sadtler Research Laboratory, Philadelphia, PA1979

14. Davies, A.N. in: G. Gauglitz, T. Vo-Dinh (Eds.), Handbook of Spectroscopy, vol. 2, Wiley-VCH, Weinheim, 2003, pp. 488–504.

15. Warr, W. A., *Perspective in information Management*, 1993, 3(2), 95-118

16. Will, M., Fachinger, W., Richert, J. R. *J. Chem. Inf. Comput. Sci*. 1996, 36, 221-227
17. McLafferty, F. W., Stauffer, D. A., Loh, S. Y., Wesdemiotis, C. *J. Am. Soc. Mass Spectrom.* 1999,10, 1229.

18. Ausloos, P. et al., *J. Am. Soc. Mass Spectrom.* 1999, 10, 287.

19. Milman, B. L. *TrAC Trends in Analytical Chemistry*, 2005,24(6), 493-508

20. Warr, W. A., *Anal Chem*., 1993, 65(23), 1045R-1051R

21. Pesyna, G. M.; Venkataraghavan, R.; Dayringer, H. E.; McLafferty, F. W. *Anal. Chem*. 1976, 48, 1362–1368.

22. Stauffer, D. B., McLafferty, F.W., Ellis, RD, Peterson, DW, *Anal Chem*, 1985, 57:771–773

23. McLafferty, F. W.; Stauffer, D. B. *J. Chem. Inf. Comput. Sci*. 1985, 25, 245–252.

24. McLafferty, F. W.; Zhang, M.-Y.; Stauffer, D. B.; Loh, S. Y. *J. Am. Soc. Mass Spectrom*. 1998, 9, 92–95.

25. Davis, A. N. *Spectrosc. Eur.*, 1993, 5(1),34-38

26. Demuth, W., Karlovits, K., Varmuza, K., *Analytica Chimica Acta*, 2004, 516, 75-85

27. Silva-Wilkinson, R. A., Burkhard, L. P., Sheedy, B. R., DeGraeve, G. M., Lordo, R. A. *Arch. Environ. Contam. Toxicol*. 1999, 36, 109–114

28. Sokolow, S.; Kamofsky, J.; Gust&on, P. The Finnigan Library Search Program; Finnigan Application Report 2; Finnigan Corp.; San Jose, CA, March 1978.

29. Hertz, H. S.; Hites, R. A.; Biemann, K. *Anal. Chem*. 1971, 43, 681-691.

30. Lu, C.-J.; Whiting J. J.; Sacks, R. D.; Zellers, E. T.; *Anal. Chem*. 2003**,** *75(6)*, 1400-1409.

31. Bryan F. J. Manly, Multivariate Statistical Methods: A Primer, CRC Press, 2004

32. Vandeginste, B. G. M., Massart, D. L., Buydens, L. M. C., S. De Jong, P. J. Lewi and J. Smeyers-Verbeke, *Handbook of chemometrics and Qualimetrics: Part B*, Elsevier, Amsterdam, 1998, ch. 33, pp. 228–232.

33. Frye-Mason, G.; Kottenstette, R. J.; Lewis, P. R.; Heller, E. J.; Manginell, R. P.; Adkins, D. R.; Dullock, D.; Martinez, D.; Sasaki, D.; Mowry, C.; Matzke, C.; Anderson, L. Proceedings of Micro Total Analysis Systems 2000, Kluwer Academic Publishers, Dordrecht, Netherlands, 2000, pp. 229-232.

34. Grate, J. W.; Patrash, S. J.; Kaganove, S. N. *Anal. Chem.*, 1999, 71, 1033.

35. Grate, J. W. Kaganove, S. N. Patrash, S. J.; Craig, R.; Bliss, M. *Chem. Mater.*, 1997, 5, 1201.

36. Lu, C.J., Jin, C. and Zellers, E.T., *Journal of Environmental Monitoring*, 2006, 8, 270-278.

37. Park, J.; Zellers, E.T. *Analyst*, 2000,*125*, 1775-1782.

38. Zellers, E. T.; Han, M. *Anal. Chem.* 1996, *68*, 2409-2418.

CHAPTER 3

Evaluation of Multi-Transducer Arrays for the Determination of Organic Vapor Mixtures

3.1 Introduction

In most reports on the application of microfabricated sensor arrays to multi-vapor analysis, the devices employed operate on the same transduction principle. Examples include surface-acoustic-wave (SAW) resonators,[1-3] cantilevers,[4-5] capacitors,[6-9] calorimeters,[10,11] and chemiresistors.[12-15] The sensors in such single-transducer (ST) arrays are typically coated with different sorptive-polymer interface layers. Reversible, partially selective responses to a given vapor can be obtained by virtue of differences in the magnitude of partitioning and the corresponding changes in the properties of the interface layer to which the underlying transducer is sensitive. Recognition (discrimination) of as many as 15-20 individual vapors is possible with ST arrays of just 2-6 sensors by use of statistical classification methods that compare each measured array response pattern to those stored in a calibration library.[1,3,15-20] Recognition of the components of mixtures containing more than two vapors, on the other hand, has proven to be quite difficult with ST arrays.[2,17-22]

The difficulty in quantitatively analyzing vapor mixtures can be ascribed to 1) the limited range of non-bonding interactions that occur between vapors and polymers, which constrains the extent of differential partitioning, and 2) the limited range of properties probed by the sensors in an ST array. These two factors will dictate the diversity of responses achievable with such arrays, regardless of the transducer employed. It stands to reason that arrays incorporating sensors that operate on different transduction principles should enhance response diversity by probing different aspects of the vapor-interface interaction. This, in turn, should afford a greater amount of uncorrelated information about a set of vapors and, thereby, improve discrimination.

The notion of combining different microfabricated transducers in an array or operating a given transducer in different modes, is not new; reports on multi-transducer (MT) arrays of gas/vapor microsensors, often referred to as "hybrid arrays", date back to the mid-1980's.[23] Perhaps the most extensive effort to develop MT-array technologies has been that of Göpel et al., who explored combinations of transducers (and interface materials) that probe changes in capacitance, heat flow, refractive index, thickness, mass, resistance, and electrochemical activity.[24] The MOSES II modular sensor system is a practical embodiment of the MT-array concept.[25] In a parallel and similarly sustained effort, Baltes et al. have worked on integrating different vapor-sensitive transducers with on-chip circuitry to improve signal quality and reduce size.[26] A monolithic realization of such an MT-array chip was reported in 2001[27] and characterized more recently.[28] A number of other reports on MT-array performance have been published.[29-33]

Although it has been established that discrimination of vapor-phase analytes with an MT array can be superior to that with an ST array, comparative studies have been

narrow in scope, generally relying on empirical data drawn from a few cases that afford little or no insight about the mechanistic features of the discrimination or the ultimate boundaries on performance. Approaches to determining the optimal number of sensors to include in arrays for multi-vapor discriminations have been described in several reports,[3,20,29,30,34,35] but only one of these explores this topic for MT arrays.[30] The important question of whether MT arrays can provide quantitative analysis of multi-vapor mixtures of greater complexity than those that can be analyzed by ST arrays remains unanswered.

This article is a follow-up to a recent article by two of the authors (P. K. and A. H.) documenting that capacitive, calorimetric, and cantilever sensors provide responses that are generally consistent with their respective transduction mechanisms.[28] It was shown that the relative response patterns among a set of vapors varied with molecular mass for cantilevers, dielectric constant for capacitors, and heat of sorption for calorimeters after accounting for the degree of partitioning. Here, we use calibration data generated with the same set of MT chips as in that study. By use of various metrics of selectivity and Monte Carlo simulations coupled with pattern recognition algorithms we evaluate the expected performance of such MT arrays, and compare the performance of MT and ST arrays as a function of the size (i.e., the number of sensors) of the arrays. An emphasis is placed on determining the complexity of mixtures that can be quantitatively analyzed with such arrays.

3.2 Experimental

3.2.1 Data Set Description. The data set employed for this study consists of the slope sensitivities of 15 sensors to each of 11 organic vapors. Five sets of integrated 3-transducer chips were used to gather the data, each having a capacitor, cantilever, and calorimeter (referred to henceforth as CAP, CAN, and CAL, respectively). The three transducers on each chip were spray-coated simultaneously with one of the following polymers: ethyl cellulose (EC), poly(cyanopropylmethylsiloxane) (PCPMS, 10% cyano groups), poly(dimethylsiloxane) (PDMS), poly(epichlorohydrin) (PECH), and poly(etherurethane) (PEUT). The 11 vapors included in this series of analyses are: n-octane (OCT), n-heptane (HEP), toluene (TOL), methanol (MOH), ethanol (EOH), 1-propanol (POH), ethyl acetate (EAC), chloroform (CHL), carbon tetrachloride (CCL), trichloroethylene (TCE), and perchloroethylene (PCE). Details of the devices, polymer structures, film deposition, vapor generation, and sensor calibration can be found in reference 28.

Most of the 165 sensitivity values included in the data base were determined by linear regression of experimental sensor responses onto vapor concentration with a forced-zero intercept. For 12 of the vapor-sensor combinations, however, low signal-to-noise ratios in the target concentration range (methanol, ethanol) or other data acquisition difficulties precluded accurate determinations of sensitivity values. In order to have a coherent data set for this study, these 12 values were imputed using established response models. A summary of our use of these models for imputing missing values and the accuracies of the models for this data set have been reviewed in Chapter 1. Tabulations of sensitivity values and the corresponding limits of detection are presented in the Table 3-1 and Table 3-2.

3.2.2 Monte Carlo Simulations and EDPCR Models. Although sensors were integrated in groups of three on each array chip, all five arrays (i.e., 15 sensors) were exposed simultaneously to all test atmospheres, and we have treated each sensor as an independent element for the purposes of exploring the various combinations of sensors that might be assembled into an ST or MT array for vapor analysis. To assess the performance of all possible combinations of sensors and vapors, we have employed Monte Carlo simulations in conjunction with classification models derived using extended disjoint principal components regression (EDPCR). Details of this approach to sensor-array evaluation can be found in previous articles.[2,3,18]

For each simulated analysis, the slope sensitivity for each vapor-sensor pair is used as the starting point, and a response value is calculated from the regression equation for a vapor concentration randomly selected from within a pre-defined range (see below). Error is then superimposed on this response value as a way of simulating realistic variations in output signals during actual operation. Two sources of error were included in the error model employed: random slope variation and baseline noise. Both are assumed to be Gaussian and zero-centered. Estimates of the slope variation and baseline noise were obtained for each of the 15 sensors from experimental data and are summarized in Tables 3-3 and 3-4, respectively. The average relative standard deviation (RSD) of the slope sensitivity from the 11 calibrations generated for a given sensor was determined from the residual errors of the linear regression models. These ranged from 3.7 to 13.3% among all of the sensors. For simplicity, the grand average of these, 7%, was used as the basis for the error model in the Monte Carlo simulations. The average

RMS baseline noise values used are 2 Hz, 1 Hz and 5.5 mV, respectively, for the CAP, CAN, and CAL sensors.

For each simulated exposure, an error-enhanced *response* is obtained independently for each sensor in the array under consideration (note: no pre-processing of the data was performed). This leads to an error-enhanced *response pattern* that is then classified via EDPCR by comparison with patterns constructed from the calibration slopes and stored in a library. By iteratively generating error-enhanced response patterns and classifying them via EDPCR, probabilistic estimates of recognition are obtained. Each test consisted of 500 simulations, and the recognition rate (RR) was used to assess performance.

For problems involving the discrimination of individual vapors, the 95% confidence interval (CI_{95}) calculated around each average RR reflects the *distribution of individual-vapor RR values* for a given array, where the number of samples is equal to the number of vapors being discriminated. For problems involving the discrimination of mixtures from their components, the CI_{95} calculated around each average RR reflects the *distribution of mixture RR values* for a given array. In these cases, the number of samples is equal to the number of mixtures tested (note: for reference, there are 55 possible binary mixtures, 165 possible ternary mixtures, and 330 possible quaternary mixtures of the 11 vapors in the primary data set). This interval is used to assess the statistical significance of differences in RR among different arrays. The number and nature of recognition errors observed from a large sample set can be logged into a recognition matrix and evaluated with respect to the average or vapor-specific rates of recognition and the identities of any incorrect assignments. By repeating a given analysis several times it is possible to derive estimates of the CI_{95} of the *distribution of average RR values*. This

interval is used in this investigation to determine the statistical significance of differences in RR values obtained with arrays of different sizes performing the same analysis (e.g., discrimination of several individual vapors or a mixture from its components).

Previous studies of polymer-coated SAW sensors have shown that for individual-vapor concentrations below ~5 times the LOD the RR varies directly with the vapor concentration.[2,36] If the vapor concentration is >5×LOD, the RR becomes independent of concentration.[36] A similar threshold has been reported for binary and ternary mixture determinations.[2] This follows from the fact that at higher concentrations the predominant contributor to pattern variability is the slope error, which is presumed to be proportional to concentration. Note that the LOD is defined here in terms of the array as a whole such that it is determined by the highest LOD value among the individual sensors in that array.

Defining vapor concentrations on the basis of the LOD is useful for modeling, particularly in this case where transducers with different units of measure are employed. As part of our preliminary data screening, we performed Monte-Carlo-EDPR analyses of each 5-sensor ST-array for individual-vapor recognition of the 11-vapor data set at discrete concentration intervals ranging from 1-20×LOD. Results (not shown) confirm that the average RR increases with the assumed vapor concentration up to ~5LOD and then remains relatively constant at higher concentrations. Therefore, for this study we have confined consideration to vapor concentrations in the range of 5-10×LOD. While this guarantees that there are measurable signals from all sensors in the array, this also means that tests within a given series of simulations are performed at different absolute concentrations, and that for many tests the concentration is well above the 5-10×LOD range for the more sensitive sensors in the array under consideration.

Vapor mixtures of two, three, and four components were tested. For these analyses, each mixture is considered separately, and the RR is defined in terms of the ability to discriminate the mixture from its individual-vapor components and any lower-order mixtures of those components. The composite response to a mixture was simulated by summing the modeled responses to the component vapors. Ample evidence to support the assumption of response additivity for resonant mass sensors (e.g., cantilevers, SAW sensors, TSMRs) can be found in the literature,[17,21,37] and although it is reasonable to assume, as we have done here, that composite responses to mixtures of vapors from polymer-coated capacitive and calorimetric sensors are also additive, experimental evidence is rather sparse.[38,39] Error is then superimposed on the composite response for each sensor in the array under consideration, and the response pattern is compared to the previously defined EDPCR models for that mixture, its component vapors, and their lower-order mixtures (for ternary and quaternary mixtures). As with the individual-vapor cases, 500 simulations are performed for each mixture.

It has been shown that the relative concentrations of the components of a mixture can have a dramatic effect on performance and that, in general, the capability to recognize the components of the mixture is never higher than when all the components are present at the same multiple of their respective LODs (i.e., at the same signal-to-noise ratio, S/N).[2] For this initial study of MT array performance, we consider mixtures where all components are present at the same S/N (i.e., the most favorable case).

For a representative subset of analyses, quantification accuracy is assessed by comparing concentration estimates for the correctly recognized cases to those derived from the calibration data using the corresponding EDPCR model.

3.2.3 Sensor Selection. For most of the cases considered here, all possible combinations of sensors are considered for each simulated analysis (exceptions noted). For the ST arrays, which are limited to a maximum of five sensors, this requires consideration of 5, 10, and 10 subsets, corresponding to arrays containing four, three, and two sensors, respectively. For the MT arrays, where up to 15 sensors could be included, the number of possible combinations is considerably larger. For example, 3003 different 5-sensor arrays can be selected from these 15 sensors. Although we consider MT arrays containing from 2 to 15 sensors for selected analyses, we have focused on MT arrays containing five sensors for comparison with the ST arrays and for additional reasons that will be explained below.

For each mixture a separate determination of the best MT array of a given size is made on the basis of the highest RR. For the binary mixtures, this means that there could be as many as 55 different "optimal" 5-sensor MT arrays (MT-O). The average of the distribution of RR values obtained from these MT-O arrays has been used to represent the best possible 5-sensor MT-array performance. From a practical standpoint, it was also of interest to identify the single 5-sensor MT array that provides the best overall performance for a given set of analyses, that is, the highest average RR value among all the mixtures of a given complexity (i.e., binary, ternary, or quaternary). We refer to this as the "global" MT array (MT-G). Finally, we sought to determine the size and composition of the MT array that would provide optimal performance for all possible analyses, which we refer to as the "universal" MT array (MT-U). This was achieved through an assessment of the frequencies with which each of the 15 sensors appears in the 5-sensor MT-G and MT-O arrays and the dependence of performance on array size.

Monte Carlo simulations and EDPCR routines were performed in Excel using macros written in Visual Basic. Other computations were performed using Matlab 7 (MathWorks, Inc., Natick, MA).

3.3 RESULTS AND DISCUSSION

3.3.1 Individual-Vapor Recognition and Array Size. Differentiating individual vapors in a data set is one of the simplest types of vapor recognition problems one can address because it does not consider the possibility of mixtures. Table 3-1 summarizes the results of attempting to discriminate among all 11 individual vapors with MT and ST arrays composed of different numbers of sensors. The average recognition rates provided by the three 5-sensor ST arrays are 91%, 83%, and 97% for the CAP, CAN, and CAL array, respectively (note: recognition matrices for these analyses are presented and are discussed below).-

The three alcohols are difficult to discriminate for all three ST-arrays. Given the differences in dielectric constant, it is somewhat surprising that the CAP array does not provide better performance, however, the order of dielectric constant values tracks that of the vapor pressures, which determine (primarily) partition coefficients, and it appears that these two countervailing influences on the magnitude of responses with the CAP sensors effectively reduce the ability to discriminate among the alcohols. While recognition errors for the alcohols with the CAP array are generally confined to other alcohols, recognition errors are more widely distributed with the other two arrays. The alkanes, n-heptane and n-octane, are consistently discriminated at a high rate, though there is persistent low-level confusion between these two vapors with all three types of arrays.

For the CAL array, RR values are uniformly high, with the only notable error arising in the confusion of PCE with toluene and vice versa. The average RR value for the CAN array is surprisingly low and the nature of the errors is difficult to rationalize on the basis of physicochemical interactions. The electro-thermal actuation mechanism used in these devices results in an elevation of the CAN temperature by up to 19 ºC with respect to the ambient,[4c] which leads to some loss of sensitivity but should not affect response patterns greatly. Migration of the polymer films under the oscillatory motion of the cantilever might alter the mass sensitivity, which could be contributory. Although such phenomena would mitigate general conclusions about the contribution of mass sensitive sensors to MT-array discrimination, the data obtained here can still be considered typical of the cantilever technology, and analyses have been performed under that assumption.

Also shown in Table 3-5 are results of analyses with ST arrays containing fewer than five sensors. The arrays presented are those providing the highest 11-vapor average RR of all possible subsets of a given size. In all cases, as the number of sensors is reduced the RR also declines, but the dependence of the RR on array size is not particularly acute for n = 3-5 sensors, and the performance of the 3-sensor CAP and CAL arrays remains fairly good (i.e., RR > 89%). These data can be compared to those reported by Park using a polymer-coated SAW sensor to analyze 16 vapors.[3] In that study higher rates of recognition were generally observed and it was found that as few as two sensors could provide recognition rates as high as 95% for individual-vapor recognition. The generally lower rates observed here can be ascribed to the higher slope error (i.e., 7% was used here whereas 3% was used in that study) as well as differences in

the vapors and polymer interfaces. If the applied slope error is reduced to 3% for all three ST arrays, the RR values converge and approach those reported for SAW sensors (i.e., > 90% even for 2-sensor arrays, data not shown).[3] Thus, for individual-vapor recognition, it appears that the type of transducer employed is not highly critical, provided that the variation in response is kept low. As variation increases, performance differences among the transducer types become apparent.

Considering the MT array data, Table 3-5 shows that the optimal MT array of any given size outperforms the corresponding optimal ST arrays in all cases. However, the margin of difference between the best MT array and the best ST array is not very large, consistent with the relatively facile discrimination problem being addressed. In fact, for arrays of two, three, or five sensors the differences are not statistically significant (i.e., the 95% CIs overlap). A high RR value (i.e., > 94%) is predicted even for the optimal 3-sensor MT array in spite of the high slope error assumed. It is worth noting that, while the optimal 5-sensor MT array provides an RR of 99%, there are 245 other 5-sensor MT arrays (out of 3003) that provide RR values > 95%. At the same time, the lowest RR value found among the 5-sensor MT arrays is 61% and there are 578 5-sensor MT arrays that provide RR values less than the RR provided by the 5-sensor CAN array (i.e., < 83%). Thus, judicious sensor selection is essential for obtaining good performance from an MT array.

Figure 3-1a summarizes the individual-vapor performance data by presenting the average RR, CI_{95} (error bars), and the range of RR values (vertical lines) for each type of 5-sensor array. There is considerable spread in the RR values. The RRs from the CAL

and CAP arrays are not statistically significantly different, although the CAL array performs better on average.

Figure 3-2a presents the RR values obtained with optimal MT arrays of different sizes for individual-vapor recognition of three sets of vapors: the entire 11-vapor set, a subset of five vapors from different functional-group classes, which can be considered relatively easy to discriminate, and a 5-vapor subset comprising two alkanes and three alcohols, which contains the vapors that are the most difficult to discriminate. There are several noteworthy features to these plots. First, the performance improves fairly sharply as the number of sensors increases from two to four or five. It then shows little or no change until the number of sensors reaches somewhere between 9 and 13, depending on the difficulty of the analysis, at which point the performance starts to decline with additional sensors.

The CI_{95} values span a range of up to ±4% and the difference of RR values on going from 5 to 11 sensors is ≤ 4%. Thus, including more than five sensors in the MT array provides no advantage for individual-vapor recognition and eventually degrades performance due to the effects of redundancy and noise. As a result, the performance of the best 3-sensor MT array exceeds that of the entire 15-sensor MT array (RR = 93%, Table 3-1) for all three analyses considered here. The trends are similar for all three data sets, despite the difference in the maximum RR achievable, although the onset of the performance decline occurs at a smaller number of sensors with the more difficult analysis. The plateau in RR observed for >4 sensors is consistent with previous reports on polymer-coated acoustic-wave and chemiresistor sensors.[3,16,20,35] The onset and extent

of performance degradation for larger arrays will depend on the degree of collinearity (redundancy) among the sensors employed and the nature of the discrimination problem.

3.3.2 Local and Global Metrics of Selectivity. Although the approach used here of simulating exposures and estimating recognition rates affords a detailed assessment of performance, it is computationally intensive and it would be useful to have summary measures of performance with which to pre-screen arrays. Common metrics of discrimination or selectivity that can be derived for specific vapor pairs include the Euclidean distance, the pair-wise correlation coefficient (r), and the so-called 'resolution factor', which is the ratio of the average difference in responses between two vapors to the variations inherent in their individual responses.[16,35a,35c]

Local selectivity metrics such as these, when determined with respect to the *vapors,* are problematic for comparison among different MT arrays because the component sensors have different units of response. Any sensor-wise normalization or autoscaling procedure used to address the discrepancy in sensor units will corrupt the inherent between-vapor correlations, leading to principal components models that are not meaningful. Local selectivity metrics can also be derived for the *sensors* in an array on the basis of their contribution to the differentiation of vapors in a data set. For the purposes of comparing the performance of various MT arrays this approach is preferred; it is not subject to the problems arising from differences in the units of measure because between-sensor correlations are not affected by sensor-wise normalizations among different transducers. Therefore, we have restricted consideration to selectivity metrics derived from the correlations among the sensors.

Local selectivity metrics such as these are constrained to vapor-specific analyses. To derive a global selectivity metric useful for summarizing array performance in multi-vapor discriminations, the most straightforward approach is to take the average or sum of the r values or Euclidean distances determined from all possible pair-wise comparisons. Alternatively, one can calculate the condition number (CN) of the correlation matrix comprising the sensor pair-wise correlation coefficients.[40] The CN is defined as the ratio of the maximum-to-minimum singular values (i.e., square root of the respective eigenvalues) of the data matrix and can be derived from principal components analysis (PCA). The CN can be calculated directly from the PCA of the sensitivity matrix. However, in this case it reflects the composite influences of sensitivity and selectivity within the data matrix.[41] By calculating the CN from the sensor-wise correlation matrix, it is strictly a function of selectivity (collinearity).[40]

Although the CN has been used in spectroscopic chemometrics to summarize collinearity among the spectral features associated with chemical species in mixtures,[41,42] we have found no prior reports on its application to sensor array data. In fact, we found only one report on vapor sensor arrays that employed any global selectivity metric at all.[34] In that study, the sum of the Euclidean distances (SED) was used as a guide for down-selecting sensors into an optimal array, but its correlation with rates of misclassification was not explored.

The pair-wise correlation matrix for the sensors comprising each of the three ST arrays as determined on the basis of their sensitivities to the 11 individual vapors in the data set is presented in Table 3-6. The average (i.e., root-mean-square) r value (r_{rms}) for each ST array is presented in Table 3-5. The r_{rms} values are also presented for the MT

arrays listed in Table 3-5. For each array, r_{rms} is a summary measure of correlation among the constituent sensors. The trend in r_{rms} values is consistent with the trend in RR values for the ST arrays and confirms that the CAL arrays are the most diverse, followed by the CAP and then the CAN arrays. The SED values are highly correlated with the r_{rms} values of the arrays, as expected, since they are similar measures of between-sensor vector separation.

The r_{rms} value *between* two arrays is determined by averaging the individual pair-wise r values determined among all sensor pairs in the two arrays (one from each different array, Table 3-6) and reflects the net (i.e., global) correlation between the two arrays. The inter-array r_{rms} values found in this data set are 0.50, 0.72, and 0.49 for the cantilever-capacitor, cantilever-calorimeter, and capacitor-calorimeter pairs, respectively (note: correlations derived from a canonical analysis of the sensor pairs are consistent with these values). Ideally, these values would be close to zero. Thus, although each type of sensor measures a different property of the vapor-polymer interaction,[28] the responses provided are far from independent, most likely due to the direct or indirect influence of the extent of vapor-polymer partitioning on all sensor responses. The dominant influence of the nonspecific van der Waals dispersive interactions on vapor sorption in polymers has been described before.[1]

The CN values calculated from the 11-vapor data set for the 5-sensor CAP, CAN, and CAL arrays are 741, 1665, and 176, respectively. The lower value of the CN for the CAL array indicates a higher degree of selectivity. This is consistent with the other selectivity metrics and the high RR values presented in Table 3-5. The trend in CN

values among the ST arrays of a given size is also consistent with the trend in RR values.[43]

The CN values for the optimal MT arrays in Table 3-5 follow the expected trends, in general. There are two 5-sensor MT arrays, however, that produce an RR value of 98.8%, with one yielding a CN value of 109 and the other a CN value of 200. The latter CN value is higher than that for the CAL array and is therefore inconsistent with the general correlation of RR and CN values. None of the other optimal MT arrays present this anomaly. That is, for optimal MT arrays of four, three, and two sensors, the CN values shown in Table 3-5 are consistent with the RR values and are all lower than the corresponding CN values for the ST arrays of the same size.

In an attempt to extend the use of the CN as a tool for pre-screening arrays, RR values were plotted against CN values for the 11-vapor individual-vapor discrimination problem for all 3003 5-sensor MT arrays. Curiously, there was no significant correlation between RR and CN among these sensors (i.e., $r<0.50$). A similar lack of correlation was found between RR and SED. Since the average RR value for this problem is generally inflated by (dominated by) the more successful recognitions, we also tried to correlate 1-RR with CN. However, a strong correlation still could not be found. Although we have no definitive explanation for the lack of correlations among these global metrics (note: the average RR value can also be considered a global selectivity metric), it appears as if the subtle differences in performance among this large group of arrays arise from a few vapor-specific (i.e., local) errors in recognition that are not effectively captured by the global metrics. The inherent insensitivity of global metrics to local selectivity changes has been noted.[41] Thus, while the data in Table 3-5 suggest that the CN provides an

effective tool for ranking arrays, further study is warranted before an unconditional endorsement can be made of its use in this application.

3.3.3 Binary Mixture Recognition. The next series of analyses considered the problem of differentiating a mixture from its components and, in the case of ternary and quaternary mixtures, also from the lower-order mixtures of those components. The initial analyses considered MT arrays consisting of five sensors to facilitate comparisons with the ST arrays. Results for the three 5-sensor ST arrays are summarized in Figure 3-1b. As indicated by the maximum and minimum RR values, a considerable range in performance is exhibited by all three arrays. The best performance, on average, is obtained from the CAP array (RR = 76%), followed in order by the CAL array (RR = 70 %) and CAN array (RR = 49%). The difference in performance between the CAP and CAL arrays is not statistically significant but both of these arrays are significantly better than the CAN array. Still, only 11 out of the 55 binary mixtures (22%) could be analyzed with <5% error with the CAP (i.e., best ST) array.

According to the recognition matrices for the individual-vapor discriminations (Tables 3-7 through 3-10), one would expect greatest difficulty in discriminating the alcohols with all three arrays. Indeed, the RR values for the binary mixtures of these three alcohols are quite low even for the CAP array (i.e., RR values of 17, 17, and 13% are obtained for mixtures of MOH+EOH, MOH+POH, and EOH+POH, respectively).

Analysis of the binary (and other) mixtures with the 5-sensor MT arrays took two forms. In one case, a new array was selected for each analysis and the RR values from the best array for each of the 55 mixtures were compiled. The distribution of RR values for this set of optimal MT arrays (MT-O) is presented in Figure 3-1b. Then, these data

were analyzed to find the single best 5-sensor MT array. This is designated as 'MT-G' in Figure 3-1b. The sensors comprising the MT-G array for the binary mixture case are presented, along with those of the MT-G array for the individual-vapor analysis, in Table 3-11. The compositions of these two MT-G arrays are quite similar – they have four sensors in common.

With RR values of 84% and 93%, respectively, the MT-G and MT-O arrays outperform significantly all three of the ST arrays in recognizing binary mixtures. This is particularly evident for the problematic alcohol mixtures, for which the MT-O arrays provided an average RR value 3-5 times higher than those of the best ST array (i.e., RR values of 50, 66, and 68% for MOH+EOH, MOH+POH, and EOH+POH, respectively) and for which the MT-G array provided an RR value 2-4 times higher (i.e., 36.4, 50.2, and 56.0%, respectively). Overall, 32 of the 55 binary mixtures (60%) could be analyzed with <5% error with the analyte-specific MT-O arrays, and 25 of the 55 binary mixtures (45%) could be analyzed with <5% error with the MT-G array.

To explore the general question of how many sensors are required for effective binary mixture analyses with an MT array, a subset of three binary mixtures was chosen; one each from the upper, middle, and lower quartiles of the RR distribution obtained for the recognition of the 55 binary mixtures with 5-sensor MT arrays. For each mixture, the optimal MT array was determined for array sizes of 2-15 sensors. The dependence of the RR on the number of sensors in the optimal MT array is plotted in Figure 3-2b. Trends similar to those found for the individual-vapor recognition problems are observed. That is, a plateau is reached with a relatively small array, and the addition of sensors has little or no effect on performance up to ~10-13 sensors beyond which a significant decline in

RR is observed. For the relatively easy mixture only two sensors are required for an RR >99%. For the moderately difficult mixture the RR reaches a maximum value at ~6 sensors, but the difference in RR between 4 and 11 sensors is not significant (see error bars in Figure 3-2b). For the most difficult mixture the RR does not improve significantly beyond an MT array of three sensors. The trends are consistent regardless of the difficulty of the analysis. A typical binary-mixture recognition matrix is presented in Table 3-12.

3.3.4 Ternary and Quaternary Mixture Recognition. Ternary mixtures were then addressed. Results are summarized in Figure 3-1c. Of the 165 possible ternary mixtures, none could be analyzed at an RR > 90% with any of the 5-sensor ST arrays. In fact, the highest ST-array RR value, obtained with the CAP array, is 73.4%. With 5-sensor MT-O arrays, RR values > 90% are achieved for 54 ternary mixtures, or one-third of the total. This capability for ternary mixture analysis is unprecedented.[2,19] The single best MT array (MT-G) provides an average RR of 58.4% (range = 8 to 98%), with RR values of >90% achieved for 14 ternary mixtures (~8% of the total).

The composition of the MT-G array for the ternary mixtures is shown Table 3-11, and a recognition matrix obtained using the MT-O array for this mixture is presented in Table 3-13. Figure 3-1c shows that while the overall performance is rather low for all arrays, the ratio of average RR values for the MT and ST arrays increases compared to the individual-vapor and binary-mixture cases. This follows from the increased difficulty of ternary mixture determinations.

The question of how many sensors are required for ternary mixture analyses was also explored for a subset of three ternary mixtures, one each from the upper, middle, and

lower quartiles of the RR distribution determined with the 5-sensor MT-O arrays. For each of these mixtures, the optimal MT array was determined for array sizes of 3-15 sensors. As shown in Figure 3-2c, the sensitivity to the number of sensors is similar to that for the individual-vapor and binary-mixture cases. That is, beyond a few sensors there is little or no improvement in performance with increases in the number of sensors added to the array, regardless of the maximum RR value.

Results of analyses of quaternary vapor mixtures are summarized in Figure 3-2d, which shows that the best 5-sensor MT arrays outperform the 5-sensor ST arrays by a considerable margin, with an average RR of 41% for the MT-O arrays compared to 5%, 2%, and 7% for the 5-sensor CAP, CAN, and CAL arrays, respectively. The highest RR achievable with an ST array is 23% (CAL array). With 5-sensor MT arrays, the highest RR is only 62%. The single MT-G array provides an average RR of 31% and a maximum of 56%. The composition of the MT-G array for the quaternary mixtures is shown Table 3-2.

3.3.5 Constructing a Universal MT Array. The analyses described above beg the question of whether a 'universal' MT array (MT-U) could be constructed – one that would provide optimal performance for any type of analysis. Certain sensors are selected repeatedly into the most effective 5-sensor MT arrays, while others are selected rarely. From Table 3-11, we see that all of the MT-G arrays can be constructed from eight sensors: the three CAP sensors coated with PCPMS, PDMS, and PEUT; the CAN sensor coated with PEUT; and the four CAL sensors coated with EC, PDMS, PEUT, and PECH. Consistent with the relative order of individual-vapor ST-array recognition rates (Table 3-5), the MT-G arrays contain more CAP and CAL sensors than CAN sensors.

The columns on the right-hand side of Table 3-11 show the RR values obtained with the MT-G arrays for each type of analysis, confirming that none of these arrays provides universally excellent performance for all types of analyses. That is, for example, the MT-G array selected on the basis of the binary-mixture analyses gives an average RR value for the ternary analyses (52.8%) that is significantly lower than that provided by the MT-G array selected on the basis of the ternary analyses (58.4%). It follows that an array of eight sensors from which all of the 5-sensor MT-G arrays could be constructed should provide the best overall performance for all possible analyses. This approach would ensure that the critical set of set of five sensors needed for a given recognition problem is available, and it assumes that including an additional set of three sensors will not degrade performance, as suggested by the trends shown for the representative cases in Figure 3-2a-c.

This 8-sensor MT-U array provides RR values of 99.6% (CI_{95} = 0.2%), 85.1%, (CI_{95} = 4.4%) and 60.0% (CI_{95} = 3.5%) for the 11 individual vapors, 55 binary mixtures, and 165 ternary mixtures, respectively. These RRs are slightly better than the corresponding values obtained from the 5-sensor MT-G arrays (98%, 84%, and 58%, respectively, Figure 3-1a-c), confirming that there is no cost to having additional "non-essential" sensors in the array beyond the five sensors comprising the MT-G array for any particular type of analysis, and considerable benefit from being able to access those sensors required for any possible analysis.

At the same time, the MT-U RR values are considerably lower than those obtained from the 5-sensor MT-O arrays (Figure 3-1). Although this is expected because the MT-O arrays are selected on a case-by-case basis, it begs the question of whether an

MT-U array could be constructed more effectively by consideration of the MT-O arrays. An alternative means of determining an MT-U array was therefore explored whereby the sensors were ranked according to their frequency of inclusion in the 5-sensor MT-O arrays whose RR distributions are shown in Figure 3-1a-d. Those sensors appearing most often in the MT-O arrays were culled out.[44] Arrays containing from 5-11 sensors were constructed in this manner and their performance was assessed for the individual-vapor, binary-mixture, and ternary-mixture analyses (note: given the low RR values for the quaternary mixture analyses, even with the MT-O arrays, they were not pursued further).

The results, summarized in Table 3-14, show that as the number of sensors in the array increases from five to eight there is a slight increase in RR across the board. The 9-sensor array provides the same performance as the 8-sensor array, and the 10-sensor and 11-sensor arrays show a decline in performance, most likely due to the negative influence of redundancy for certain discriminations outweighing the positive influence of added discrimination power for certain other discriminations. Thus, the 8-sensor MT-U array provides the best performance with the fewest possible sensors.[45] Not surprisingly, seven of the eight sensors selected by this approach coincide with those selected on the basis of the MT-G arrays. However, the one different sensor produces an incremental, if not statistically significant, improvement in the performance, i.e., RRs of 99.8% (CI_{95} = 0.1%), 90.4% (CI_{95} = 4.8%), and 64.2% (CI_{95} = 4.6%), respectively, for analyses of the individual vapors, binary mixtures, and ternary mixtures, compared to the corresponding RRs of 99.6%, 85%, and 60% provided by the 8-sensor MT-U array constructed from the sensors in Table 3-11.

3.3.6 Quantification. Up to this point we have considered only recognition. Although correct recognition generally leads to accurate quantification,[3,18,36,46] we set out to confirm this for the MT arrays considered here. Analyses were performed for three arbitrarily chosen problems, using the 5-sensor MT-O array in each case. Once the identity of the vapor or vapor mixture is established, the concentration of each vapor is determined by linear regression of the principal-component score for that sample onto its concentration. For a vapor mixture, the composite response vector is decomposed into its constituent vectors prior to quantifying the contribution of each mixture component to the net response.

Table 3-15 presents the summary statistics. In all cases, the net bias is quite low (i.e. < 1%, data not shown). For the 11 individual vapors, average absolute errors are consistently low, ranging from 2.8 to 6.4%, and the maximum errors are acceptable. Errors for the binary mixture of EAC and TCE, and the ternary mixture of HEP, EOH, and CCL are a bit higher (i.e., average errors are <17% and maximum errors are <72%), though still generally acceptable. There is always a tendency toward higher errors with more complex mixtures due to the inherent imprecision involved in decomposing the composite response pattern into its constituent vectors. Of course, the quantification accuracy is strongly influenced by the slope error assumed in the Monte Carlo simulations, and even better performance would be expected with sensors exhibiting lower noise and less slope variation.

3.4 CONCLUSIONS

On the basis of this assessment of MT arrays for multi-vapor determinations, it is clear that inclusion of judiciously selected sensors that operate by different transduction mechanisms in an array of vapor sensors enhances performance relative to arrays composed of sensors operating on a single transduction mechanism. MT arrays can provide recognition rates of > 95% for problems involving the determination of individual vapors and numerous binary and ternary mixtures. The degree to which performance is improved over arrays employing sensors operating on the same transduction principle (ST arrays) increases with the difficulty or complexity of the recognition problem.

This is the first report to compare the performance of ST arrays and MT arrays in a comprehensive and quantitative manner with respect to quantitative analysis of multi-vapor mixtures. The capability for accurately determining the composition of a fairly large number of ternary mixtures, in spite of a relatively large degree of variability assumed in the sensitivity values of the sensors in the arrays, is unprecedented. With better control on response variability, performance is expected to improve to some extent. However, it was also shown that quaternary mixtures could not be analyzed with acceptably low error rates, indicating that the improvement in performance afforded by combining sensors that operate on different principles is finite. Although the limitations on mixture complexity are related to the extent of correlation among the sensors, and there was significant correlation among these sensors, it seems highly unlikely that any array of sensors employing reversible, sorptive interface layers will be capable of consistently analyzing mixtures of more than three components. Thus, ternary mixtures

analysis would appear to be the upper boundary on performance with this class of sensor arrays.

These results add further support to the growing body of evidence that there is no advantage to including a large number of partially selective sensors in arrays of vapor sensors employing sorptive interface layers for any specific analysis. Furthermore, even for generalized applications, the advantage of adding sensors to address multiple discrimination problems is apparently quickly offset by the disadvantage associated with redundancy – in the series of cases studied here, a maximum of eight or nine sensors could be tolerated without a net loss in performance.

Several local and global selectivity metrics were examined as potential tools for summarizing or pre-screening the performance of ST and MT arrays. Correlations with the recognition rates determined by Monte Carlo simulations coupled with EDPCR analysis were generally good. The condition number, which has apparently not been used before in the context of assessing sensor array performance, was one of the global metrics explored here for quantifying the selectivity of an array. If applied to the correlation matrix derived from a set of vapor response data, it is a pure measure of collinearity. Furthermore it has the advantage of being applicable to underdetermined systems – i.e., where the number of sensors is less than the number of vapors under consideration. Analyses revealed that while the CN values generally correlated with RR values, there were exceptions occurred. We speculate that local (i.e. vapor-specific) errors affect RR values to a greater extent than CN values, making the latter insensitive to important factors affecting array performance. However, further study is warranted.

For this initial study of MT array performance, we considered mixtures where all components are present at the same S/N and where all vapors are at a concentration of at least 5×LOD. Performance degrades as absolute concentrations (or sensor sensitivities) decrease and as the difference in concentrations (or, more accurately relative S/N ratios) among mixture components increases.[2] Although the most promising application of MT or ST arrays is in microsystems with upstream preconcentration (and separation) modules that can address the limited sensitivities of the sensors,[46-48] future studies should be directed at exploring the effects of absolute and relative vapor concentrations on MT array performance.

Finally, a limited study of previously published response models of polymer-coated CAP, CAN, and CAL arrays yielded mixed results. While model performance was quite good for some transducer-polymer combinations, for several it was only fair and for some it was not acceptable. Cases of poor performance were confined to CAP sensors and could be ascribed to properties of the polymer coatings that violate assumptions about free volume inherent in the models[28] or to data sets with a preponderance of low responses.

Acknowledgments

This research was supported through the Michigan Center for Wireless Integrated Microsystems (WIMS) by the Engineering Research Centers Program of the National Science Foundation under Award No. ERC-9986866.

Tables and Figures

Table 3-1: Experimentally derived sensitivity values for 11 vapors calibrated from the polymer-coated CAP, CAN, and CAL sensors.

Vapor	CAP (Hz/ppm $\times 10^{-2}$)					CAN (Hz/ppm $\times 10^{-2}$)					CAL (mV·s/ppm $\times 10^{-2}$)				
	EC	PCPMS	PDMS	PECH	PEUT	EC	PCPMS	PDMS	PECH	PEUT	EC	PCPMS	PDMS	PECH	PEUT
HEP	17.2	-19.5	-16.2	17.8	-9.08	-2.31	-2.08	-0.500	-0.280	-1.49	3.28	1.08	6.82	0.554[a]	1.34
OCT	36.4	-49.9	-41.8	50.7	-32.0	-5.53	-7.88	-1.71	-0.790	-3.54	10.1	5.45	34.0	1.51	7.73
TOL	67.8	-5.45	-4.23	206	-21.0	-6.46	-3.58	-0.730	-1.66	-5.74	12.1	2.71	12.4	5.25	12.3
MOH	38.8	15.6	1.83	25.7	40.0	-0.730	-0.330	-0.0717[a]	-0.192[a]	-0.380	2.69	0.715	0.354[a]	0.466[a]	1.82
EOH	62.6	25.5	4.66	38.8	73.8	-1.27	-0.645	-0.110	-0.243[a]	-0.850	4.11	1.02	0.943[a]	0.826[a]	2.92
POH	194	67.5	13.9	107	199	-4.59	-1.41	-0.170	-0.460	-2.91	14.4	3.16	2.00	2.02	18.6
EAC	81.4	17.7	18.9	105	32.1	-2.98	-1.31	-0.230	-0.470	-1.48	4.69	1.58	2.11	1.67[a]	2.48
CHL	67.1	23.0	11.5	66.4	72.8	-5.64	-1.47	-0.390	-0.800	-4.21	8.20	1.17	1.73	1.06	6.42
CCL	9.42	-6.74	-2.72	24.0	-5.61	-2.37	-1.78	-0.700	-0.350	-3.20	2.65	0.945	4.62	0.946[a]	3.42
TCE	85.4	10.9	11.1	98.1	45.2	-7.13	-2.35	-0.480	-0.830	-5.03	12.1	1.48	5.34	1.87[a]	6.83
PCE	86.2	-28.6	-11.4	168	-32.2	-2.31	-2.08	-0.500	-0.280	-1.49	22.0	4.09	18.8	4.96[a]	17.8

[a] These values were imputed using response models described above (see text).

Table 3-2. Limits of detection (LOD, ppm) for the 11 test vapors with each sensor.[a]

	CAP					CAN					CAL [b]				
Vapor	EC	PCPMS	PDMS	PECH	PEUT	EC	PCPMS	PDMS	PECH	PEUT	EC	PCPMS	PDMS	PECH	PEUT
HEP	54	37	27	49	63	73	79	150	560	72	340	2300	340	2100	900
OCT	26	14	11	17	18	31	21	46	200	30	110	450	68	770	150
TOL	14	130	100	4	27	26	46	100	95	19	95	910	180	220	97
MOH	24	46	240	34	14	230	500	1100	820	280	420	3400	6500	2500	650
EOH	15	28	95	23	8	130	250	710	650	130	280	2400	2400	1400	400
POH	5	11	32	8	3	37	110	460	340	37	79	780	1200	570	64
EAC	12	41	23	8	18	57	120	340	330	72	240	1500	1000	700	480
CHL	14	31	38	13	8	30	110	200	190	25	130	2100	1300	1100	180
CCL	99	110	160	37	100	72	93	110	450	33	430	2600	500	1200	350
TCE	11	66	40	9	13	24	70	160	190	21	94	1600	430	620	170
PCE	11	25	39	5	18	13	25	44	82	8	52	600	120	230	67

[a] Determined using 3σ/sensitivity, where σ = RMS baseline noise (see Tables S-5 and S-6).
[b] LOD calculated by assuming a peak width of 2 seconds at the base.

Table 3-3. Variation in calibrated slope sensitivities for all sensors on the basis of responses to all 11 test vapors.

Transducer	Avg. RSD of Sensitivity (%)[a]					Transducer Avg.
	EC	PCPMS	PDMS	PECH	PEUT	
CAP	12 (9.0)	8.6 (10)	7.6 (5.7)	8.0 (8.5)	5.2 (5.5)	8.2
CAN	5.0 (2.8)	9.0 (9.6)	10 (8.4)	12 (11)	3.3 (5.3)	7.9
CAL	7.2 (3.4)	7.4 (3.9)	5.7 (3.6)	2.7 (4.6)	6.7 (6.8)	6.0
Coating Avg.	8.0	8.3	7.9	7.5	5.0	Grand Avg. = 7.0

[a] standard deviation (%) around the average RSD is given in parentheses.

Table 3-4. RMS baseline noise for each sensor.

Transducer (units)	RMS noise					Transducer Avg.
	EC	PCPMS	PDMS	PECH	PEUT	
CAP (Hz)	3.1	2.4	1.5	2.9	1.9	2.4
CAN (Hz)	0.57	0.55	0.26	0.53	0.36	0.45
CAL (mV)	3.8	8.2	7.7	3.9	4.0	5.5

Table 3-5. Comparison of ST and MT arrays of different sizes for individual-vapor recognition of 11 vapors in terms of predicted recognition rates and various global selectivity metrics.[a]

Array size	Array type	Array composition					RR (CI_{95})	CN	r_{rms}	SED
15	MT	-	-	-	-	-	93.0 (2.0)	-	0.61	424
5	ST	CAP-EC	CAP-PCPMS	CAP-PDMS	CAP-PECH	CAP-PEUT	91.0 (7.7)	740	0.67	26
5	ST	CAN-EC	CAN-PCPMS	CAN-PDMS	CAN-PECH	CAN-PEUT	82.9 (7.3)	1660	0.80	19
5	ST	CAL-EC	CAL-PCPMS	CAL-PDMS	CAL-PECH	CAL-PEUT	96.5 (1.6)	176	0.60	30
5	MT	CAP-PCPMS	CAP-PDMS	CAL-EC	CAL-PDMS	CAL-PEUT	98.8 (1.2)[b]	109	0.57	40
4	ST	CAP-EC	CAP-PCPMS	CAP-PDMS	CAP-PEUT	-	90.2 (4.0)	245	0.74	13
4	ST	CAN-PCPMS	CAN-PDMS	CAN-PECH	CAN-PEUT	-	78.9 (6.7)	247	0.78	12
4	ST	CAL-EC	CAL-PCPMS	CAL-PDMS	CAL-PEUT	-	93.3 (2.0)	123	0.69	15
4	MT	CAP-PCPMS	CAP-PDMS	CAL-EC	CAL-PEUT	-	98.2 (1.2)	70.4	0.52	21
3	ST	CAP-EC	CAP-PDMS	CAP-PEUT	-	-	89.5 (5.8)	11.7	0.65	8.0
3	ST	CAN-PDMS	CAN-PECH	CAN-PEUT	-	-	72.8 (7.0)	30.5	0.78	6.2
3	ST	CAL-EC	CAL-PDMS	CAL-PECH	-	-	91.3 (3.0)	10.1	0.57	8.9
3	MT	CAP-PCPMS	CAP-PDMS	CAL-PEUT	-	-	94.1 (2.2)	3.5	0.50	10
2	ST	CAP-PECH	CAP-PEUT	-	-	-	74.5 (7.0)	1.2	0.09	4.7
2	ST	CAN-PDMS	CAN-PEUT	-	-	-	53.0 (9.1)	6.5	0.73	2.3
2	ST	CAL-PDMS	CAL-PECH	-	-	-	68.6 (8.0)	2.3	0.39	3.5
2	MT	CAP-PDMS	CAL-PEUT	-	-	-	75.8 (6.8)	1.1	0.03	4.5

[a] For all ST arrays and all MT arrays containing < 15 sensors, the array presented is that providing the highest recognition rate (RR, %) based on 500 trials; CI_{95} is the width of the 95% confidence interval (%) around each RR value; r_{rms} is the RMS average pair-wise correlation coefficient for each array; SED is the sum of the Euclidean distances among the sensors in each array.

[b] Two MT arrays give the same RR value of 98.8%; one yields a CN value of 109 and the other a CN value of 200 (see text).

Table 3-6. Pair-wise (Pearson) correlation coefficients (r) among the 15 sensors.[a]

	A1	A2	A3	A4	A5	B1	B2	B3	B4	B5	C1	C2	C3	C4	C5
A1	1.00														
A2	0.69	1.00													
A3	0.49	0.85	1.00												
A4	0.49	0.00	0.14	1.00											
A5	0.77	0.93	0.65	-0.09	1.00										
B1	-0.31	0.33	0.18	-0.73	0.26	1.00									
B2	0.09	0.77	0.78	-0.37	0.57	0.70	1.00								
B3	0.19	0.81	0.76	-0.32	0.65	0.72	0.97	1.00							
B4	-0.17	0.39	0.22	-0.86	0.42	0.91	0.67	0.68	1.00						
B5	-0.19	0.38	0.20	-0.69	0.35	0.96	0.66	0.73	0.91	1.00					
C1	0.58	-0.10	-0.11	0.76	0.02	-0.93	-0.62	-0.60	-0.83	-0.88	1.00				
C2	0.28	-0.45	-0.62	0.44	-0.21	-0.62	-0.90	-0.81	-0.57	-0.53	0.69	1.00			
C3	-0.19	-0.80	-0.87	0.23	-0.59	-0.51	-0.97	-0.91	-0.51	-0.46	0.45	0.88	1.00		
C4	0.31	-0.23	-0.10	0.95	-0.29	-0.79	-0.54	-0.52	-0.94	-0.80	0.79	0.53	0.39	1.00	
C5	0.73	0.12	-0.03	0.73	0.25	-0.75	-0.47	-0.43	-0.69	-0.71	0.92	0.67	0.32	0.74	1.00

[a] A, B, and C denote CAP, CAN, and CAL sensors, respectively; 1, 2, 3, 4, and 5 denote the polymers EC, PCPMS, PDMS, PECH, and PEUT, respectively.

Table 3-7. Recognition matrix for the 5-sensor CAP array generated from Monte Carlo/EDPCR analyses for discrimination among the 11 individual vapors.[a]

Assigned identity	Actual identity										
	HEP	OCT	TOL	MOH	EOH	POH	EAC	CHL	CCL	TCE	PCE
HEP	483	19	0	0	0	0	0	0	0	0	0
OCT	17	481	0	0	0	0	0	0	2	0	0
TOL	0	0	493	0	0	0	0	0	0	0	0
MOH	0	0	5	426	39	28	0	0	0	2	0
EOH	0	0	0	44	340	153	0	0	0	0	0
POH	0	0	1	29	120	318	0	0	0	0	0
EAC	0	0	0	0	0	0	491	0	0	5	0
CHL	0	0	0	0	1	1	2	497	0	0	0
CCL	0	0	0	0	0	0	0	0	493	0	5
TCE	0	0	1	0	0	0	7	0	0	493	2
PCE	0	0	0	1	0	0	0	0	5	0	493
RR (%)	96.6	96	99	85	68	64	98	99	99	99	99

[a] n = 500 trials for each vapor; the average RR is 91%.

Table 3-8. Recognition matrix for the 5-sensor CAN array generated from Monte Carlo/EDPCR analyses for discrimination among the 11 individual vapors. [a]

Assigned identity	Actual identity										
	HEP	OCT	TOL	MOH	EOH	POH	EAC	CHL	CCL	TCE	PCE
HEP	480	19	0	0	1	0	0	0	2	0	0
OCT	14	480	0	0	0	3	0	0	0	0	0
TOL	0	0	408	4	78	0	1	0	0	0	17
MOH	0	0	2	409	4	0	2	0	0	0	0
EOH	4	0	66	48	334	1	73	2	0	5	11
POH	0	0	0	0	2	420	4	29	1	120	0
EAC	2	0	0	38	74	1	419	0	0	11	0
CHL	0	0	0	0	0	3	0	404	0	55	0
CCL	0	1	1	0	0	0	0	0	476	0	44
TCE	0	0	2	0	2	69	1	58	0	274	21
PCE	0	0	21	0	5	0	0	7	21	35	407
RR (%)	96	96	82	82	67	84	84	81	95	55	81

[a] n = 500 trials for each vapor; the average RR is 82%.

Table 3-9. Recognition matrix for the 5-sensor CAL array generated from Monte Carlo/EDPCR analyses for discrimination among the 11 individual vapors. [a]

Assigned identity	Actual identity										
	HEP	OCT	TOL	MOH	EOH	POH	EAC	CHL	CCL	TCE	PCE
HEP	482	7	0	0	0	1	0	0	0	0	0
OCT	15	485	0	0	0	0	0	0	0	0	0
TOL	0	0	461	0	0	0	0	0	11	0	36
MOH	0	0	0	475	20	0	0	0	0	0	0
EOH	0	0	0	17	476	0	1	3	0	1	0
POH	0	0	0	1	0	488	0	10	0	0	0
EAC	3	0	1	0	1	0	497	0	2	0	0
CHL	0	0	0	7	3	8	0	483	0	0	0
CCL	0	0	8	0	0	0	0	0	487	0	1
TCE	0	8	0	0	0	3	0	4	0	489	6
PCE	0	0	30	0	0	0	0	0	0	10	457
RR (%)	96	97	92	95	95	98	99	97	97	98	91

[a] n = 500 trials for each vapor; the average RR is 96%.

Table 3-10. A recognition matrix generated from Monte Carlo/EDPCR analysis of the 5-sensor optimal MT array for discrimination among 11 single vapors.[a,b]

	HEP	OCT	TOL	MOH	EOH	POH	EAC	CHL	CCL	TCE	PCE
HEP	498	3	0	0	0	0	0	0	0	0	0
OCT	2	497	0	0	0	0	0	0	0	0	1
TOL	0	0	497	0	0	0	0	0	0	0	0
MOH	0	0	0	486	22	0	0	0	0	0	0
EOH	0	0	0	10	469	9	0	2	0	0	0
POH	0	0	0	1	1	486	0	0	0	0	0
EAC	0	0	0	2	2	3	498	0	0	0	0
CHL	0	0	0	0	0	1	1	498	0	0	0
CCL	0	0	0	0	0	0	0	0	495	0	9
TCE	0	0	0	1	0	1	1	0	0	499	0
PCE	0	0	3	0	0	0	0	0	5	0	490
RR (%)	99.6	99	99	97	94	97	100	100	99	100	98

[a] n = 500 trials for each vapor; the average RR is 98%.
[b] refer to Table 2 in the text of the article for the composition of this (optimal) MT array.

Table 3-11. Composition and overall performance of 5-sensor global MT arrays (MT-G) selected on the basis of performance in a specific type of analysis.

Basis for Sensor Selection	Array composition[a]					Avg. Recognition Rate (%)			
						11 Individual vapors	55 Binary mixtures	165 Ternary mixtures	330 Quaternary mixtures
11 Individual vapors	CAP-PCPMS	CAP-PDMS	CAL-EC	CAL-PDMS	CAL-PEUT	98.8	76.2	53.0	18.0
55 Binary mixtures	CAP-PCPMS	CAP-PDMS	CAN-PEUT	CAL-EC	CAL-PEUT	98.2	84.3	52.8	27.4
165 Ternary mixtures	CAP-PCPMS	CAP-PDMS	CAN-PEUT	CAL-PDMS	CAL-PEUT	95.6	77.9	58.4	23.9
330 Quaternary mixtures	CAP-PDMS	CAP-PEUT	CAN-PEUT	CAL-PDMS	CAL-PECH	91.2	78.4	48.2	30.8

[a] These arrays are the 5-sensor MT-G arrays selected on the basis of the type of analysis in the first column. The performance (average recognition rate) of each array for each type of analysis is presented in the last four columns.

Table 3-12. Recognition matrix from a Monte Carlo/EDPCR analysis of a binary mixture of toluene and perchloroethylene with its optimal 5-sensor MT array [a]

Assigned identity	Actual identity		
	TOL	PCE	TOL+PCE
TOL	487	6	75
PCE	10	492	20
TOL+PCE	3	2	405
RR (%)	-	-	81

[a] This MT array is composed of CAP-PCPMS, CAP-PECH, CAT-EC, CAT-PEUT and CAL-PDMS sensors.

Table 3-13. Recognition matrix from a Monte Carlo/EDPCR analysis of a ternary mixture of toluene, ethanol, and trichloroethylene with its optimal 5-sensor MT array[a]

Assigned identity	Actual identity						
	TOL	EOH	TCE	TOL+EOH	TOL+TCE	EOH+TCE	TOL+EOH +TCE
TOL	498	0	0	0	2	0	0
EOH	0	489	1	9	0	47	24
TCE	0	6	499	5	27	0	0
TOL+EOH	2	2	0	484	0	5	32
TOL+TCE	0	0	0	0	471	0	0
EOH+TCE	0	3	0	2	0	446	37
TOL+EOH+TCE	0	0	0	0	0	1	401
RR (%)	-	-	-	-	-	-	80.2

[a] This MT array is composed of CAP-PCPMS, CAT-PDMS, CAT-EC, CAL-PCPMS, and CAL-PDMS sensors.

Table 3-14. Performance of universal MT array (MT-U) candidates comprising from 5-11 sensors selected on the basis of their ranking among all 5-sensor optimal MT arrays (MT-O).

Number of sensors[a]	Avg. Recognition Rate (%)		
	11 Individual vapors	55 Binary mixtures	165 Ternary mixtures
5	95.8	86.2	51.7
6	98.8	85.6	50.0
7	98.9	86.9	52.5
8[b]	98.5	90.4	64.2
9	98.6	90.4	64.2
10	95.3	85.4	59.8
11	94.6	84.8	59.5

[a] Composition of each array is based on the rank of each sensor as determined by the frequency with which it was included in the 5-sensor optimal arrays (MT-O) for all analyses. See text for ranked frequencies of the sensors.

[b] This 8-sensor array provides the best overall performance as a universal MT array (MT-U).

Table 3-15. Accuracy of quantification for representative individual-vapor, binary-mixture, and ternary-mixture determinations. [a]

Vapors	RR (%)	Absolute error (%)	
		Avg. [b]	Max
individual vapors			
HEP	99.2	3.5 (2.7)	14
OCT	100	3.3 (2.4)	12
TOL	99.8	3.0 (2.2)	11
MOH	97.4	5.2 (3.9)	21
EOH	95.6	5.2 (3.8)	18
POH	96.4	4.7 (3.5)	19
EAC	98.8	3.8 (2.9)	16
CHL	99.6	4.2 (3.1)	19
CCL	100	3.3 (2.5)	13
TCE	99.8	2.8 (2.1)	12
PCE	99.8	3.0 (2.1)	11
binary mixture			
EAC	92.6	6.2 (4.4)	20
TCE	92.6	11 (7.6)	34
ternary mixture			
HEP	82.0	4.6 (3.6)	19
EOH	82.0	4.0 (3.0)	15
CCL	82.0	17 (13)	71

[a] Obtained with the corresponding 5-sensor MT-G arrays in Table 2.

[b] Values in parentheses are standard deviations (%).

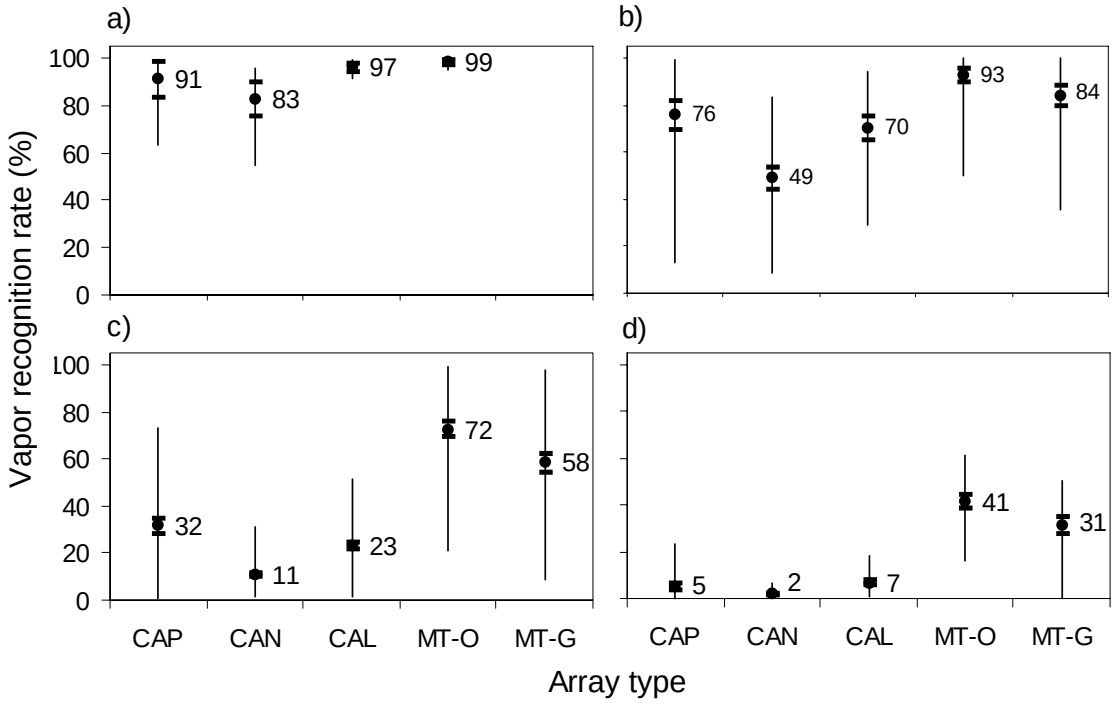

Figure 3-1. Comparison of 5-sensor ST arrays to 5-sensor global and optimal MT arrays (MT-G and MT-O, respectively, see text for explanation) for the recognition of a) 11 individual vapors, b) 55 binary mixtures, c) 165 ternary mixtures and d) 330 quaternary mixtures. Confidence interval (95%) around mean recognition rate is shown for each array type, along with the range (minimum to maximum) of recognition rates calculated for each array type.

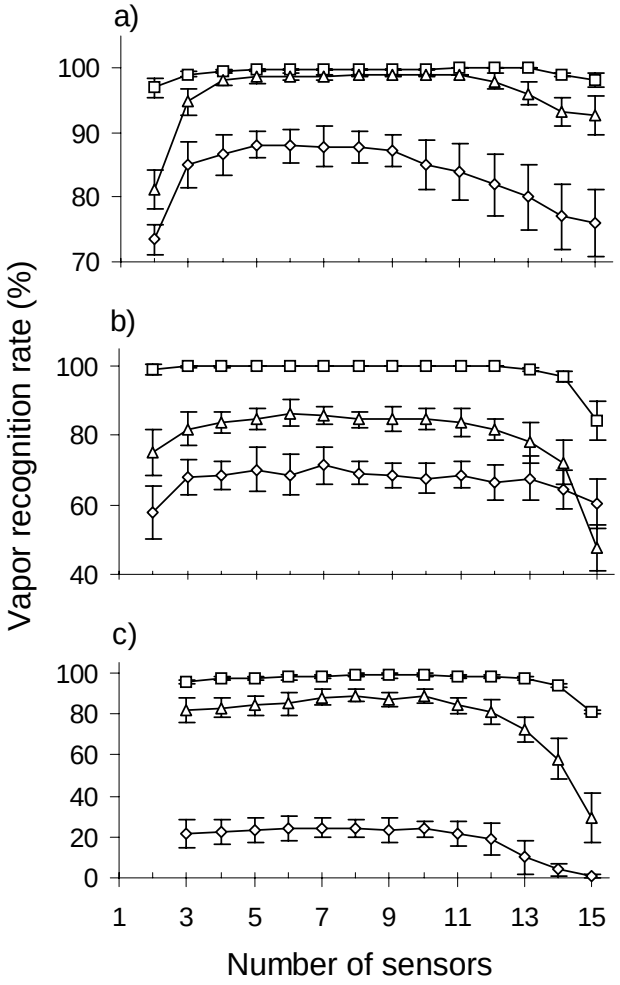

Figure 3-2. Effect of array size on performance for optimal MT arrays containing from 2-15 sensors for: a) individual-vapor recognition of the entire 11-vapor set (triangles), a subset of five vapors of diverse structures considered to be relatively easy to discriminate (HEP, TOL, POH, EAC, CHL) (squares), and a subset of five vapors of similar chemical structure considered relatively difficult to discriminate (HEP, OCT, MOH, EOH, POH) (diamonds); b) recognition of three binary mixtures, in order of increasing difficulty, HEP+MOH (squares), TOL+PCE (triangles), and HEP+OCT (diamonds); c) recognition of three ternary mixtures, in order of increasing difficulty, OCT+EAC+CCL (squares), TOL+EOH+TCE (triangles), and MOH+EOH+POH (diamonds). Error bars represent the CI_{95} for each average RR value, determined on the basis of the number of vapors for the individual-vapor cases and on the basis of five iterations for the binary and ternary cases.

Reference

1. a) Grate, J. W. *Chem. Rev.* 2000, *100*(7), 2627-2647; b) Grate, J. W.; Rose-Pehrsson, S. L.; Venezky, D. L.; Klusty, M.; Wohltjen, H. *Anal. Chem.* 1993, *65*, 1868-1881.

2. Hsieh, M.-D.; Zellers, E. T. *Anal. Chem.* 2004, *76*, 1885-1895.

3. Park, J.; Groves, W. A.; Zellers, E. T. *Anal. Chem*. 1999, *71*, 3877-3886.

4. a) Chapman, P. J.; Vogt, F.; Dutta, P.; Datskos, P. G.; Devault, G. L.; Sepaniak, M. J. *Anal. Chem*. 2007, *79*(1), 364-370; b) Kim, B. H.; Prins, F. E.; Kern, D. P.; Raible, S.; Weimar, U. *Sens. Actuators B* 2001, *78*(1-3),12-18.; c) Vančura, C.; Rüegg, M.; Li, Y.; Hagleitner, C.; and Hierlemann, A. *Anal. Chem.*, 2005, 77, 2690-2699.

5. a) Dutta, P.; Senesac, L. R.; Lavrik, N. V.; Datskos, P. G.; Sepaniak, M. J. *Sens. Lett.* 2004, *2*, 238-245; b) Senesac, L. R.; Dutta, P.; Datskos, P. G.; Sepaniak, M. J. *Anal. Chim. Acta* 2006, *558*, 94-101.

6. Cornila, C.; Hierlemann, A.; Lenggenhager, R.; Malcovati, P.; Baltes, H.; Noetzel, G.; Weimar, U.; Göpel, W. *Sens. Actuators B* 1995, *24-25*, 357-361.

7. Kummer, A. M.; Hierlemann, A.; Baltes, H. *Anal. Chem.* 2004, *76*, 2470-2477.

8. Patel, S. V.; Mlsna, T. E.; Fruhberger, B.; Klaassen, E.; Cemalovic, S.; Baselt, D. R. *Sens. Actuators B* 2003; *96*(3); 541-553.

9. Lange, D.; Hagleitner, C.; Hierlemann, A.; Brand, O.; Baltes, H., *Anal. Chem.* 2002, *74*, 3084-95.

10. Lerchner, J.; Seidel, J.; Wolf, G.; Weber, E. *Sens. Actuators B* 1996,*32*, 71-75.

11. Lerchner, J.; Caspary, D.; Wolf, G. *Sens. Actuators B* 2000, *70*, 57-66.

12. a) Swann, M. J.; Glidle, A.; Cui, L.; Barker, J. R.; Cooper,J. M. *Chem. Commun.* 1998, *24*, 2753-2754; b) Severin, E. J.; Lewis, N. S. *Anal. Chem.* 2000, *72*(9), 2008-2015.

13. Patel, S. V.; Jenkins, M. W.; Hughes, R. C.; Yelton, W. G.; Ricco, A. J. *Anal. Chem.* 2000, *72*(7), 1532-1542.

14. Persaud, K. C.; Travers, P. J. in *Handbook of Biosensors and Electronic Noses*; E. Kress-Rogers, Ed.; CRC Press: Boca Raton, FL, 1997; pp. 563-592; b) Slater, J. M.; Watt, E. J.; Freeman, N. J.; May, I. P.; Weir, D. J. *Analyst*,1992, *117*, 1265-1270; c) Bartlett, P. N.; Ling-Chung, S. K. *Sens. Actuators* 1989, *20*, 287-292.

15. a) Shi, X.; Wang, L.; Kariuki, N.; Luo, J.; Zhong, C. J.; Lu, S. *Sens. Actuators B* 2006, *117*, 65-73; b) Cai, Q. Y.; Zellers, E. T. *Anal. Chem*. 2002, *74*, 3533-3539.

16. Doleman, B. J.; Lonergan, M. C.; Severin, E. J.; Vaid, T. P.; Lewis, N. S. *Anal. Chem.* 1998, *70*, 4177-4190

17. a) Cai, Q. Y.; Heldsinger, D.; Hsieh, M. D.; Park, J.; Zellers, E. T. *Sen. and Actuators B* 2000, *62*, 121-130; b) Park, J.; Zhang, G. Z.; Zellers, E. T. *AIHAJ* 2000, *61*,192-204.

18. Zellers, E. T.; Batterman, S. A.; Han, M.; Patrash, S. J. *Anal. Chem*. 1995, *67*, 1092-1106.

19. Hsieh, M. D.; Zellers, E. T. *J. Occup. Environ. Hyg*. 2004, *1*, 149-60.

20. a) Ricco, A. J.; Crooks, R. M.; Osbourn, G. C. *Acc. Chem. Res*. 1998, *31*(5), 289-296; b) Osbourn, G. C.; Bartholomew, J. W.; Ricco, A. J.; Frye, G. C. *Acc. Chem. Res.* 1998, *31*(5), 297-305; c) Osbourn, G.C.; Martinez, R.F.; Bartholomew, J.W.; Yelton, W.G.; Ricco, A.J. *Proc. Electrochem. Soc.* 1999, *99*, 127–131; d) Park J.; Zellers E.T. *Proc. Electrochem. Soc.* 1999, *99*, 132–137.

21. Then, D.; Vidic, A.; Ziegler, Ch. *Sen. Actuators B* 2006, *B117*(1), 1-9.

22. Hierlemann, A.; Weimar, U.; Kraus, G.; Schweizer-Berberich, M.; Göpel, W. *Sens. Actuators B* 1995, *26-27*, 125-134.

23. a) Snow, A. W.; Barger, W. R.; Klusty, M.; Wohltjen, H.; Jarvis, N. L. *Langmuir* 1986, *2*, 513- 519; b) Bott B.; Jones T. A. *Sens. Actuators* 1986, *9*, 19-25.

24. a) Schierbaum, K. D.; Gerlach, A.; Haug, M.; Göpel, W. *Sens. Actuators A* 1992, *31*, 130-137; b) Haug, M.; Schierbaum, K. D.; Nahm, W.; Gauglitz, G.; Göpel, W. *Sens. Actuators B* 1993, *11*, 383-391; c) Zhou, R.; Hierlemann, A.; Weimar, U.; Göpel, W. *Sen. Actuators B* 1996, *34*, 356-360; d) Ulmer, H.; Mitrovics, J.; Noetzel, G.; Weimar, U.; Göpel, W. *Sens. Actuators B* 1997, *43*, 24–33.

25. Mitrovics, J.; Ulmer, H.; Weimar, U.; Göpel, W. *Acc. Chem. Res.* 1998, *31*, 307-315; b) Stetter, J. R.; Strathmann, S.; McEntegart, C.; Decastro, M.; Penrose, W. R. *Sens. Actuators B* 2000, *69*, 410-419.

26. Hierlemann, A.; Baltes, H. *Analyst,* 2003**,** *128* (1), 15-28.

27. Hagleitner, C.; Hierlemann, A.; Lange, D.; Kummer, A.; Kerness, N.; Brand, O.; Baltes, H. *Nature* 2001, *414*(15), 293-296.

28. Kurzawski, P.; Hagleitner, C.; Hierlemann, A. *Anal. Chem*. 2006, *78*(19), 6910-6920.

29. Holmberg, M.; Winquist, I.; Gardner, J.W.; Hines, E.L. *Sens. Actuators B* 1995, *27*, 246–249.

30. Pardo, M.; Kwong, L.G.; Sberveglieri, G.; Brubaker, K.; Schneider, J.F.; Penrose, W.R.; Stetter, J.R. *Sens. Actuators B* 2005, *106*, 136–143.
31. Sauter, D.; Weimar, U.; Noetzel, G.; Mitrovics, J.; Göpel, W. *Sen. Actuators B* 2000, *69*, 1-9.

32. Ulmer, H.; Mitrovics, J.; Weimar, U.; Göpel, W. *Sens. Actuators B,* 2000, *65*, 79–81.

33. a) Göpel, W. *Sens. Actuators* B 1998, *52*, 125-142; b)Weimar, U., Göpel, W. *Sens. Actuators* B 1998, *52*, 143-161

34. Chaudry, A. N.; Hawkins, T. M.; Travers, P. J. *Sens. Actuators* B 2000, *69*, 236-242.

35. Burl, M. C., Sisk, B. C., Vaid, T. P., Lewis, N. S. *Sens. Actuators B* 2002, *87*,130-149.; b) Polikar, R.; Shinar, R.; Udpa, L.; Porter, M. D. *Sen. Actuators B* 2001, *80*, 243-254.; c) Wilson, D. M.; Garrod, S. D. *IEEE Sensors Journal* 2002, *2*(3), 169-178.; d) Gardener, J. W.; Boilot, P.; Hines, E. L. *Sen. Actuators B* 2005, *106*, 114-121; e) Eklov, T.; Martensson, P.; Lundstrom, I. *Anal. Chim. Acta* 1999, *381*, 221-232.

36. Zellers, E. T.; Park, J.; Hsu, T.; Groves, W. A. *Anal. Chem.* 1998, *70*, 4191-4201.

37. Patrash, S. J.; Zellers, E. T. *Anal. Chem.* 1993, *65*, 2055-2066.

38. Kummer, A. D., Burg T. P., Hierlemann A. *Anal. Chem*. 2006, *78*, 279-290.

39. Siperstein F., Gorte, R. J., Myers, L. *Langmuir* 1999, *15*, 1570-1576.

40. Meloun, M.; Militky, J.; Hill, M.; Brereton, R. G. *Analyst* 2002, *127*, 433-450

41. Kalivas, J. H.; Lang, P. M. *Chemom. Intell. Lab. Syst.* 1996, *32*, 135-149

42. Faber N. M.; Ferre´, J.; Boque´, R.; Kalivas, J. H. *Trends Analy. Chem.* 2003, *22* (6), 352-361.

43. Since the CN measures the magnitude of correlations among the sensors, the dimension of the correlation matrix directly affects the values of the CN, with more sensors corresponding to a higher range of CN values. This is evident in the CN values for the arrays of different sizes in Table 3-1 – the relative order of the CN values among the ST arrays is maintained but the values decrease with the size of the array under consideration.

44. In order to eliminate the inherent bias arising from the fact that there are different numbers of binary, ternary, and quaternary mixtures, the frequency at which each sensor appeared in the MT-O arrays for a given type of analysis was converted to a percentage value, and then the percent frequencies were summed across all types of

analyses.

45. The component sensors and their frequency of occurrence in the 5-sensor MT-O arrays are: CAP-PDMS (16.7%), CAP-PCPMS (12.1%), CAN-PEUT (10.1%), CAL-PEUT (9.4%), CAL-PDMS (8.6%), CAL-EC (6.7%), CAN-EC (6.2%), and CAL-PECH (5.9%).

46. Lu, C.-J.; Jin, C.; Zellers, E. T. *J. Environ. Monitor.* 2006, *8*, 270-278.

CHAPTER 4

Limits of Recognition for Binary and Ternary Vapor Mixtures Determined with Multi-Transducer Arrays

4.1 Introduction

In the vast majority of studies on the application of microfabricated sensor arrays to analyses of volatile organic compounds (VOC) the sensors employed in the arrays operate on the same transduction principle.[1-5] With most of these single-transducer (ST) arrays, a thin interfacial film of a sorptive polymer serves to reversibly concentrate vapors near the surface of each sensor. Although the differential sorption that occurs among transducers coated with different polymer films gives rise to array response patterns that can be used to differentiate one vapor from another, it is not generally possible to determine the components of mixtures of more than two VOCs using an ST array.[6-12] This is because of the inherent limitations on both the range of vapor-polymer interactions and the features of those interactions probed by the underlying transducer.[6,13]

The use of arrays of transducers operating on different principles, which we refer to as multi-transducer (MT) arrays, should enhance the capability for vapor recognition by probing different aspects of the vapor polymer interactions.[13-17] In the context of VOC-mixture analyses, MT arrays are expected to afford better discrimination of

mixtures from their components when compared to ST arrays, and to extend the range of component-vapor concentrations over which such mixtures could be effectively analyzed.

In a recent study, we used a data base of experimental responses to 11 individual vapors in Monte Carlo simulations to compare the expected performance of polymer-coated ST arrays and MT arrays for multi-vapor analysis.[17] Sensors consisted of polymer-coated capacitors, cantilevers, and calorimeters. Assuming additivity in constructing the simulated composite responses to binary, ternary and quaternary mixtures, optimally composed MT arrays outperformed optimally composed ST arrays having the same number of sensors in all cases. For problems involving the discrimination of individual vapors from other individual vapors in the calibration library, the difference in performance was not particularly large, owing to the relative simplicity of the problems. However, for binary and ternary mixture analyses there was a more significant advantage observed for the better MT arrays. Using an 8-sensor MT array that was determined to provide the best overall performance among all possible combinations of sensors under consideration, the 11 individual vapors could be differentiated from one another with an average error of only 1.2%, and 19 of the 55 possible binary mixtures that could be constructed from those 11 vapors could be discriminated from their component vapors with < 5% error. Only three of the 165 possible ternary mixtures and none of the 330 possible quaternary mixtures could be accurately determined with this MT array.

The results of that study were obtained under the constraints that all of the components of the mixtures were present at the same signal-to-noise ratio and that the concentrations of the vapors were all greater than five times their respective limits of detection (LOD). Since the response patterns used to discriminate vapor mixtures from

their components will generally degrade at low concentrations due to increased noise in the sensor signals, it is important to consider the range of absolute concentrations over which reliable recognition is possible. In addition, since recognition will also degrade as the relative concentration of the minority component(s) decrease, it is important to consider performance over a wide range of relative concentrations.

The "limit of recognition" (LOR) was first introduced in 1998 as an additional metric by which microsensor arrays could be assessed.[18] For individual-vapor recognition problems the LOR is defined as the lowest concentration at which a vapor can be reliably recognized among a pre-defined set of interfering vapors on the basis of its sensor array response pattern. It complements the LOD as a performance assessment parameter for microsensor arrays. Extension of the LOR concept to vapor mixture analyses was reported subsequently, and the methods for defining and assessing mixture LORs were illustrated using a 16-vapor data set obtained with an ST array of six polymer-coated surface acoustic wave (SAW) sensors.[6]

Briefly, since mixture recognition varies as a function of both the absolute and relative concentrations of the components, the mixture LOR is defined as an *m*-dimensional region, where *m* is the number of components in the mixture, whose boundary corresponds to the threshold between recognizable and unrecognizable mixtures based on some predetermined level of acceptable recognition error (e.g., 5%). For a binary mixture, the LOR is defined as the area of a planar region; for a ternary mixture, it is defined as the volume of a polyhedron; and so on. The size of the region reflects the range of component concentrations over which a mixture can be analyzed. As the concentration ratio of the mixture components becomes larger, it becomes more

difficult to differentiate patterns of the mixture from patterns of the component vapors (binary and higher) and their lower-order mixtures (ternary and higher). Comprehensive analysis generally demands that the LORs be evaluated probabilistically, and, by convention a default threshold recognition rate (RR) of 95% is used.[6,18] The corresponding LOR is designated as the LOR_{95}.

In this article, we extend our recent study of MT-array performance to address the influence of changes in absolute and relative vapor concentrations on the ability of an MT array to recognize binary and ternary mixtures by use of the LOR. Using the optimal 8-sensor MT array identified in the original study, the LORs for binary mixtures are evaluated and discussed. The effects of baseline sensor noise and random sensitivity fluctuations on the LOR are examined for representative binary mixtures. The asymmetry of the LOR profiles with respect to the component-vapor concentrations is then characterized and rationalized. Ternary-mixture LORs determined with this MT array are then addressed briefly followed by a comparison between an optimal 5-sensor MT array and the best of the 5-sensor ST arrays in the data set with regard to their resilience to concentration variations in binary mixtures. The practical implications of the results for vapor mixture analyses by microanalytical systems that might incorporate MT arrays as detectors downstream from a chromatographic separation module are discussed.

4.2 Experimental

4.2.1 Data set description. The primary data set consists of calibrated responses to the vapors of 11 VOCs from 15 sensors: five capacitors (CAP), five cantilevers (CAN), and five calorimeters (CAL). One of each type of transducer was coated with one of the

following polymers: ethyl cellulose (EC), poly(cyanopropylmethylsiloxane) (PCPMS, 10% cyano groups), poly(dimethylsiloxane) (PDMS), poly(epichlorohydrin) (PECH), and poly(etherurethane) (PEUT). Sensor responses were collected for the following 11 VOCs over a range of concentrations: n-heptane (HEP), n-octane (OCT), toluene (TOL), methanol (MOH), ethanol (EOH), 1-propanol (POH), ethyl acetate (EAC), chloroform (CHL), carbon tetrachloride (CCL), trichloroethylene (TCE), and perchloroethylene (PCE). All of the calibration curves were linear and the sensitivity of each vapor-sensor pair determined from the slope was used as the starting point for all simulations. Details about the transducers, polymer structures, film deposition, vapor generation, and sensor calibrations can be found in references 13 and 17.

4.2.2 Monte Carlo Simulations and EDPCR Models. To assess the expected performance of different sensor arrays for different vapor mixtures, Monte Carlo simulations were coupled with extended disjoint principal components regression (EDPCR) classification models. In this approach, measurement error is superimposed on calibrated response values as a way of simulating the variations in sensor output signals expected during actual operation. Error-enhanced response patterns for individual vapors and their mixtures from each array under consideration are then compared to stored library patterns using EDPCR to determine if the vapor or vapor mixture has been correctly recognized. For each mixture, 500 simulated response patterns were generated, from which statistical estimates of the RR (%) were derived. Further details of this approach to sensor-array evaluation can be found elsewhere.[6,8,17,19]

Two sources of error were included in the Monte Carlo error model employed in this study: random sensitivity variations and baseline noise. The former was estimated for

each of the 165 vapor-sensor combinations from the relative standard deviation (RSD) of the slope of each calibration curve.[17] Values ranged from 4-13%, and for the simulations performed here the average of 7% was used in all cases. The average root-mean-square (RMS) baseline noise values were determined experimentally as 2 Hz, 1 Hz, and 5.5 mV for the coated CAP, CAN, and CAL transducers, respectively.

For all simulations, vapor concentrations were expressed as multiples of the LOD (see reference 17 for tabulations of the LOD values). The LOD is defined for the array as a whole, meaning that it is determined by the highest LOD value among the individual sensors in that array. Specifying the absolute and relative concentrations as multiples or fractions of the vapor LODs facilitates comparisons across all vapor mixtures among arrays constructed from transducers with different units of measure. For this data set, LODs for a given vapor among the transducers of a given type coated with different polymers varied by ≤ 17-fold, while those for a given vapor among different transducers with the same polymer coating varied by as much as two orders of magnitude.

For a given binary mixture, each analysis entailed discriminating the response pattern of the mixture from those of the constituent vapors. The absolute concentrations of the components ranged from 0.1 – 1500 × LOD. The concentration of one component at a time was varied in seven discrete increments (0.1, 0.5, 1, 5, 10, 15, and 20 × LOD) while the concentration of the second component was stepped through discrete increments corresponding to relative concentration ratios (RCR) of 1:75 to 75:1 (again, expressed as multiples of the component-vapor LODs) with respect to the first component. For the ternary cases, each analysis entailed discriminating the response pattern of the mixture from those of the constituent vapors and their binary mixtures. The

concentration of one component at a time was varied in five discrete increments (1, 5, 10, 15, 20, and 30 × LOD) while the concentrations of the second and third components were individually stepped through twenty discrete increments between ratios of 1:10 to 10:1 (in terms of the LOD of each vapor) with respect to the first component. For the purposes of these analyses, it is assumed that sensor responses remain linear and that composite responses to mixtures are equivalent to the sum of the responses to the component vapors at concentrations exceeding those examined during calibration, in spite of expected deviations at the higher concentrations examined.

The manner in which we have formulated the recognition problems in this study is relevant to the situation where the sensor array is used as the detector for a microsystem with an upstream separation column. Differential retention reduces the number of vapors simultaneously presented to the sensor array and retention times can be used *a priori* to constrain the subset of possible vapors subject to consideration at any given point in the analysis.[20] This justifies limiting the mixture-analysis problem to one where 1-3 specific vapors might be present simultaneously within a retention-time window and the identity (and quantity) of each of those that are actually present needs to be determined. For a binary mixture the problem is simply to distinguish the mixture from its two constituent vapors. For a ternary mixture, the problem is to distinguish the mixture from the three individual vapors as well as the three binary mixtures that could be composed from those vapors.

4.2.3 MT and ST Arrays. From our previous analyses of these data, the best overall performance, in terms of the recognition rates for all possible individual vapors and mixtures, was provided by an 8-sensor MT array of the following sensors: CAP-

PCPMS, CAP-PDMS, CAN-EC, CAN-PEUT, CAL-EC, CAL-PDMS, CAL-PECH, and CAL-PECH.[17] This array was considered to be 'universal' and is therefore abbreviated as the MT-U array. Out of the 55 possible binary mixtures one can create from 11 vapors, 19 could be recognized by the MT-U array with an RR ≥95% (assessed over an absolute concentration range of 5–20LOD and an RCR of 1:1), and most of the analyses described below focus on these binary mixtures or subsets thereof. Out of 165 ternary mixtures that could be created from 11 vapors, an RR ≥95% could be achieved (and the LOR_{95} defined) for only three mixtures by the MT-U array for the same absolute and relative concentrations as mentioned above. Therefore, in order to increase the number of mixtures that could be considered, the threshold RR value was reduced to 90%, which yielded nine ternary mixtures with finite LOR_{90} values.

Among the 5-sensor ST arrays, the CAP array provided the best performance, with RR values >95% for 11 of the 55 binary mixtures (5-20LOD; RCR = 1:1). For the purpose of comparing the performance of ST and MT arrays, an optimal 5-sensor MT array, consisting of those sensors that yield the highest average recognition rate among all of the 55 binary mixtures, was selected. It consists of the following sensors: CAP-PCPMS, CAP-PDMS, CAN-PEUT, CAL-EC, and CAL-PEUT.[17] This array was capable of achieving an RR > 95% for 12 binary mixtures, 11 of which were the same as those for which the CAP array provided finite LOR_{95} values.

4.2.4 Recognition Surfaces and LOR Contours. Vapor RR values calculated over all compositions of a given mixture were used to construct three-dimensional recognition surfaces from which the LOR could be derived. For each recognition-surface plot, the mixture recognition rate (vertical z-axis) is presented as a function of the

absolute concentration of the minority component (y-axis) and the RCR (x-axis). Absolute concentrations are presented in multiples of the LOD of the minority mixture component, spanning a range of 0.1-20LOD, with recognition assessed at seven absolute concentration values within this range. Beyond 20LOD the LOR boundary does not change significantly so the plots are truncated at this value of absolute concentration. For the left-hand side of the plots (i.e., RCRs of 1:1-75:1), the absolute concentrations on the y-axis correspond to one of the mixture components and for the right-hand side (i.e., RCRs of 1:1-1:75) they correspond to the other component. The plots are labeled with the names of each vapor in a manner that assists with interpretation. Each surface consists of 119 coordinate points that define the RR for the absolute and relative concentrations of interest. By connecting the points corresponding to an RR of 95% on the z-axis, a boundary is defined corresponding to the LOR_{95} compositions for the mixture. The area of the contour projected on the x-y plane provides a convenient means of summarizing the LOR_{95} values of binary mixtures (see Figure 4-1).

To quantify the area of the irregular LOR_{95} contour, the coordinates of the x and y axes are transformed in a manner that preserves the spatial relationships of the data but permits interpolation between the discrete points at which the RRs are evaluated.[21] Since the units of the LOR_{95} become arbitrary, we have adopted the convention of expressing the LOR_{95} without units.

Although the binary LOR_{95} can be used to rank different arrays in terms of their resilience to composition changes, the quantity itself is rather intangible. It is useful therefore to summarize further the LOR_{95} by selecting a single absolute concentration value and to calculate the magnitude of the limiting RCR along this cross section (y-axis

coordinate), which gives the range of relative concentrations over which an RR = 95% can be achieved. Since the 'depth' of the LOR_{95} has a maximum value of 20 (i.e., 1-20LOD), and the RR does not usually change radically in the range of 5-20LOD, the binary-mixture LOR_{95} values will generally vary more due to differences in this limiting RCR than due to differences in the absolute concentration range. Thus, this is a reasonable way of reducing the results to facilitate interpretation. We have chosen the 10LOD coordinate for this cross-sectional metric, because 10LOD is often used as the 'limit of quantitation' in analytical measurements. Accordingly, the limiting RCR is designated as RCR_{10}.

Two approaches to quantifying the RCR_{10} could be used. In one, the length on either side of the 1:1 RCR point can be calculated separately. In the other, the total length can be calculated by merely dividing the RCR value on one side of the 1:1 RCR point by that on the other side (e.g., 5:1/1:2 = 10 for the case where the LOR has RCR_{10} boundaries at 5LOD:1LOD and 1LOD:2LOD). Although the latter approach is conceptually sound, from a practical standpoint it is problematic because, by definition, the minority component changes as the 1:1 RCR point is crossed, and it takes on the value of 10LOD. However, to the extent that it portrays the 'fold-range' over which a mixture can be recognized at ≥ 95%, this approach to evaluating the RCR_{10} serves its purpose. As it turns out, there is a strong linear correlation ($r^2 = 0.96$) between the values of RCR_{10} and LOR_{95} for the 19 mixtures analyzed by the MT-U array in this data set, which supports its use as done here.

For ternary mixtures, it is not possible to capture all of the variables affecting recognition in a single contour plot because four dimensions are required. In the approach

taken here, two-dimensional sections are created at a fixed, discrete level (see above) of absolute concentration for the first component and over a range of RCRs for the second and third components. The average of the LOR_{95} contour areas over these discrete absolute concentration increments gives a score for the ternary mixture that varies in proportion to the total range of reliably analyzed concentrations. This way, a ternary mixture analysis is reduced to a series of binary mixture analyses and an average LOR score can be calculated accordingly to assess array performance.

Monte Carlo simulations and EDPCR routines were performed in Excel using macros written in Visual Basic. Calculations of LOR contour areas, plots of the recognition surfaces, and other computations were performed using Matlab 7 (MathWorks, Inc., Natick, MA).

4.3 Results and Discussion

4.3.1 LORs for Binary Mixtures. The LOR_{95} values determined with the MT-U array for all 19 binary mixtures are presented in Table 4-1. They span a 7.4-fold range, from 16 to 120. The corresponding RCR_{10} values, presented in Table 4-2, range from 5 to 20, indicating that the minimum relative concentration range over which a binary mixture can be analyzed is 5-fold and the maximum is 20-fold. At larger RCR values, the mixture is confused for one of the two mixture components.

In those cases where the LODs differ between the two components, the RCR_{10} expressed in terms of actual vapor concentrations (e.g., part-per-million, ppm) will differ proportionally. Among the 19 mixtures, the largest component-LOD ratio is 8:1 (for n-octane:methanol) and the next-largest ratio is 6:1 for chloroform:perchloroethylene. By

evaluating the LODs for each of the components of the mixture, it is possible to define the RCR_{10} ppm-wise. These RCR_{10} ranges are also presented in Table 4-2. The mixture of heptane + chloroform, which has one of the largest values of RCR_{10} in terms of signal-to-noise ratio (i.e., 20), also has the largest RCR_{10} in terms of ppm (i.e., 16). In contrast, the mixture of carbon tetrachloride + trichloroethylene, which has a small RCR_{10} when defined in terms of signal-to-noise ratio (i.e., 5), has one of the largest RCR_{10} values when defined by terms of ppm (i.e., 20).

Figure 4-1 shows two representative recognition-surface and LOR_{95} contour plots. In each case, the RR never exceeds that at an RCR = 1:1. This holds for all 19 binary mixtures and confirms the intuitive expectation that recognition becomes more difficult as the composition (or, more accurately, the array response vector) of the mixture becomes dominated by either component. This is consistent with findings from our previous study of binary LORs with polymer-coated SAW sensors.[6]

The rate of decline in RR on either side of the line corresponding to RCR=1:1 is determined by the similarity in response patterns for the components, and the asymmetry in the decline in RR as a function of relative composition is related to the differences in the LODs and array sensitivities for each component. The surfaces do not change much as a function of absolute concentration above ~5LOD (minority component). In contrast, as the concentration of the minority component falls below the LOD, there is a precipitous decline in RR commensurate with the loss of sensors making a contribution to the response pattern and with the increasing influence of baseline noise on the response patterns. All of these issues are addressed below.

4.3.2 Response Pattern Similarity. The correlation of the response-pattern similarity between the components of a binary mixture with the magnitude of the LOR_{95} was evaluated using the Euclidean distance, E_{AB}, as the similarity metric. Values of E_{AB} were determined on the basis of normalized response patterns, wherein each sensor response was divided by the sum of all sensor responses to an arbitrary concentration of the vapor. The magnitude of E_{AB} decreases with the similarity in the response patterns of the two vapors and, for the 19 mixtures with finite LOR_{95} values, ranges from 1.6 to 2.0 (Table 4-1).

Figure 4-2 shows the correlation of the LOR_{95} with the component-vapor E_{AB} values. Data for all 55 binary mixtures in the original data set have been included to show that there is a threshold E_{AB} value of ~1.6 below which analysis is not possible (i.e., $LOR_{95} = 0$). Above this value the LOR_{95} increases rapidly with E_{AB} in a roughly linear manner; however, the correlation is not strong enough to establish any sort of predictive model. Furthermore, even beyond the threshold E_{AB}, several mixtures cannot be analyzed. These anomalous cases are most likely attributable to the magnitude of the sensitivity error used in the Monte Carlo simulations and to the fact that neither the variability of responses nor the relative or absolute vapor concentrations are accounted for in the E_{AB} calculation.

The recognition surface plots and LOR_{95} contours presented in Figure 4-1 are for the binary mixtures with the largest and smallest E_{AB} values among the 19 vapors with finite values of LOR_{95}: n-octane + 1-propanol and trichloroethylene + perchloroethylene, respectively. The large difference in LOR_{95} is apparent from a visual inspection of the

surface plots. It is also apparent that above a minority-component concentration of 5LOD, the RCR dictates the LOR_{95}, not the absolute concentration.

4.3.3 Sensitivity Variations and Baseline Noise. The ability to identify a vapor on the basis of its response pattern relies implicitly on the fidelity of that pattern over time which, in turn, relies on the reproducibility of the responses from the sensors in the array. Considered in the context of a micro-analytical system, the reproducibility of sample volumes, flow rates, preconcentrator capture and desorption efficiencies, chromatographic retention, and temperature control must be considered. Fluctuations in any or all of the processes involved in signal generation can contribute to perturbations in response patterns that may undermine recognition via matching to library patterns.

The effect of sensitivity variations on the LOR_{95} was explored for several mixtures while the baseline noise levels of all sensors were maintained at their experimental values. Figure 4-3 shows the recognition surfaces and LOR_{95} contours for a representative binary mixture of n-heptane + ethyl acetate for sensitivity errors of 3, 5, and 10% applied randomly to the sensors in the MT-U array. This mixture gives a mid-range LOR_{95} value and thus represents a recognition problem of moderate difficulty. The corresponding LOR_{95} values are 210, 88, and 0, respectively. The effect is quite significant – a 1.7-fold increase in sensitivity error (i.e., from 3 to 5%) leads to a proportional reduction in the LOR_{95} of 58%, and if the RMS sensitivity variation is allowed to increase to 10%, then the mixture cannot be analyzed at any composition. These results serve to illustrate that stable (reproducible) sensor responses are essential for satisfactory recognition of mixtures over relevant concentration ranges.

To explore the effect of baseline noise on mixture recognition, analyses were confined to a concentration range of 0.1-5LOD where such noise has its greatest impact on response patterns.[6,17] As noted above, below 1LOD at least one of the sensors will not be contributing a statistically significant response to the pattern produced by the array. Although high values of RR can be maintained for some mixtures, continued reductions in concentration inevitably lead to a decline in performance.[18]

The influence of baseline noise on the recognition surface and LOR_{95} is shown in Figure 4-4, again, for the binary mixture of n-heptane + ethyl acetate. As the noise level increases there is a commensurate reduction in RR and LOR_{95} values. The rate of reduction in RR is greater at lower concentrations (i.e., < 1LOD) but with increasing noise levels significant reductions are also observed in the 1-5LOD range. Note that the LOR_{95} contour shrinks along both the absolute and relative concentration axes, and for the example shown in Figure 4-4 a 10-fold increase in noise leads to ~4-fold decrease in LOR_{95} (i.e., from 56 to 13). The general trend holds for other mixtures as well, although the dependence on noise levels it is more acute for mixtures whose LOR_{95} values are initially smaller (i.e., for mixtures whose components are more difficult to discriminate). In none of the cases considered did the noise level affect RR values above an absolute concentration of 5LOD, leading to the practical recommendation that sensors be operated above this concentration to avoid any possible influence of baseline noise on pattern fidelity.

4.3.4 Compositional Asymmetry in the LOR Contours. In nearly all cases, the LOR_{95} contours are significantly asymmetric with respect to composition. This is shown in Table 4-2 and it is what gives rise to the marked skew of the surfaces and contours in

the example shown in Figures 4-1a. Taking the ratio of the LOR_{95} areas on each side of the 1:1 composition line, $lor_{95\text{-}A}$ and $lor_{95\text{-}B}$, respectively, provides a way of quantifying the asymmetry (see Figure 4-1a and Table 4-1). We refer to this ratio as *A* in Table 4-1 and employ it here as the index of asymmetry. In only two cases does the value of *A* fall within the interval of 0.5-2. The greatest degree of asymmetry occurs for the mixture of chloroform + perchloroethylene with $A = 0.045$ (i.e., 1/22).

The primary factor contributing to the asymmetry in the recognition contours is the difference in the sensitivity of the array to each component, *S*. However, since we define concentration in terms of the LOD, differences in the component LODs will also contribute. Indeed, the ratio of the LOD-*S* product, *P*, of each component in the mixture is strongly correlated with the degree of asymmetry, and regressing *A* onto $(P_A/P_B)^{-1}$ yields a line with an $r^2 = 0.986$. In physical terms, the product of LOD and *S* corresponds to the increase in the length of the array response vector per unit increase in concentration, where concentration is measured in multiples of the LOD (i.e., 'xLOD').

Note that the array sensitivity to each vapor corresponds to the Euclidean vector norm for that vapor as determined from the net contributions of all of the sensors. Mathematically, it is the square root of the sum of squares of each sensor's sensitivity to that vapor. The LOD, however, is determined by the single sensor in the array exhibiting the largest minimum response for the vapor. Thus, a vapor may have a high LOD, but at the same time may have a high *S* value, since the two are not necessarily correlated.

The *P* values for the components of each mixture along with their ratio are compiled in Table 4-1. The recognition surface is always higher toward the component with the lower value of *P*. For the example shown in Figure 4-1a, $P = 630$ for n-octane

and P = 2,500 for 1-propanol; giving a ratio of about 1:4. Accordingly, A = 4 and the RR remains higher for mixtures rich in n-octane. For the mixture of n-heptane and ethyl acetate shown in Figure 4-3a, P_A/P_B is only 1.3:1 and the surface and LOR_{95} are correspondingly more symmetric (A = 0.46), with a slight preference for mixtures rich in ethyl acetate.

The need to account for the LOD in assessing asymmetry is an artifact of expressing concentration in these analyses as multiples of the LOD for each vapor. If concentrations were expressed in units of mass/volume or ppm, then the sensitivity ratio alone would determine the asymmetry. This is evident in Table 4-1 for the mixture of chloroform and carbon tetrachloride, where the LODs differ by < 10% but $P_A/P_B > 7$ and A = 0.053. It is also evident in several entries in Table 4-2, where the direction of asymmetry actually reverses when the RCR_{10} values are evaluated in terms of ppm rather than 'xLOD'.

The Euclidean distance between the mixture components can also be partitioned into the segments corresponding to the distances between each component and the mixture with an RCR = 1:1. Here again, account must be taken again of the differences in LOD values and array sensitivities of the mixture components. If the mixture components are present at equivalent signal-to-noise ratios then the composite response vector falls at the midpoint between the two individual component vectors in hyperspace. If the contribution of each component to the composite response is multiplied by its P value, then the resulting composite vector is shifted in the direction of the vapor with the larger value of P. After applying this weighting factor to each normalized response vector, the partitioned Euclidean distances between each individual-vapor vector and the

1:1 mixture can be determined. These are denoted here as e_A and e_B. By definition, the e_A/e_B ratio is directly proportional to P_A/P_B and therefore it should also be inversely related to the asymmetry index, A.—

Accordingly, only small increases in the relative concentration of the vapor with the larger value of P in the mixture can be tolerated before the composite response pattern for the mixture can no longer be differentiated from that of the individual component. This is illustrated by the sharp decline in RR shown in Figures 4-1a and 4-3a as the composition of the mixture shifts from 1:1 toward compositions rich in this component. The fact that shifts in the other direction, toward compositions rich in the other component, permit higher RR values to be achieved over a wider RCR range is consistent with the greater Euclidean distance that exists between the vector corresponding to the individual component and that for the 1:1 mixture.

The practical implications of the asymmetry are essentially superimposed on the overall constraint imposed by the LOR_{95}. As stated above, the maximum ratio of concentrations of mixture components that can be tolerated is about 20:1. Beyond this range the minority component will not be recognized. Due to asymmetry, however, this range is not equally applicable to each component, and when the component to which the array is more sensitive becomes the majority component the RCR over which recognition is possible is much smaller than when it is the minority component. In a microanalytical system where a chromatographic band is delivered to the array, these results have direct implications for assessing peak purity -- if a minority component to which the array is less sensitive is embedded within the peak corresponding to the majority component, it will not likely be detected.

4.3.5 LORs for Ternary Mixtures. Figure 4-5 shows cross-sectional LOR contours for the mixture of n-heptane + ethyl acetate + chloroform, which is one of the three ternary mixtures for which average RR values > 95% were obtained (5-20LOD at an RCR = 1:1:1). The panels show the iso-recognition contours at concentrations of 5LOD and 20LOD for each of the components. The coordinate axes are the RCRs for the other two components relative to the one whose concentration is specified (i.e., the minority component). In contrast to the general trend observed for binary mixtures, the recognition rate for the ternary mixtures continues to increase in the range of 5-20LOD, but does not increase significantly at higher minority-component concentrations. Interestingly, at the 90% contour, the concentration dependence is reduced. None of the mixtures could be recognized at a minority-component concentration < 5LOD, meaning that the LOR places a greater constraint on the analysis than does the LOD.

The LOR_{95} score (evaluated from 1-20LOD) for the mixture shown in Figure 4-5 is 4.0 and those for other two ternary mixtures are both 3.5, reflecting the small range of concentrations over which they can be recognized. Referring to Figure 4-5, the largest RCR within the 95% contours is ~8 (i.e., from 4:1 to 1:2) regardless of which vapor is held constant, and the largest ratio of concentrations that can be recognized, in units of ppm, is ~5:1. The largest RCR values for the other two mixtures are comparable to these values.

The asymmetric distribution of the iso-recognition contours with respect to an RCR=1:1 for the two components whose concentrations are allowed to vary in these cross sections is correlated with P_A/P_B, as observed with the binary mixtures. For the example shown, the values of P for n-heptane, chloroform, and ethyl acetate are,

respectively, 590, 1000, and 460, which results in $P_A/P_B = 0.6$ for n-heptane:chloroform, 0.77 for ethyl acetate:n-heptane, and 0.45 for ethyl acetate:chloroform (see Table 4-1). Accordingly, the LOR_{95} contours in Fig. 4-5 are skewed with respect to the RCR=1:1 point toward ethyl acetate to the greatest extent, followed by n-heptane. That is, higher RR values are possible over wider concentration ranges when these two vapors are majority components. Similar correlations are found for the other two ternary mixtures with finite LOR_{95} values.

If the threshold RR is reduced to 90%, nine ternary mixtures give finite LOR_{90} values and trends can be explored. Table 4-3 presents the LOR_{90} scores for these mixtures as a function of the absolute concentration of the first component listed for each mixture. It is worth noting that the three mixtures with finite LOR_{95} values can be recognized at concentrations < 1LOD using the 90% recognition criterion (i.e., LOR_{90}), while the other ternary mixtures cannot. Regardless, the cross-sectional LOR_{90} values do not increase significantly beyond a minority-component concentration of about 5LOD. The LOR_{90} scores (from 1-20LOD) range from 4.7 to 18 and show a strong correlation with the average RR value determined at an RCR of 1:1:1 over the range of 5-20LOD.

4.3.6 Binary LORs for 5-sensor ST and MT arrays. The performance of the 5-sensor CAP ST array and the optimal 5-sensor MT array was assessed in the same manner as the MT-U array. Detailed results are tabulated in the Supporting Information accompanying this article (Tables 4-4 through 4-7). Key points are summarized here as they relate to the relative advantages of using MT vs. ST arrays for multi-vapor determinations.

The fact that only 12 binary mixtures could be analyzed with an average RR ≥95% (5-10LOD at an RCR of 1:1) with this 5-sensor MT array, compared to 19 binary (and 3 ternary) mixtures for the 8-sensor MT-U array, attests to the value of the added sensors in the latter array. Seven of these 12 vapors were recognizable by both MT arrays at an RR ≥95%. The 5-sensor CAP array was capable of analyzing 11 of the 12 mixtures analyzed by the 5-sensor MT array.

In most cases the average RR (RCR = 1:1, 5-10LOD) is higher for the MT array among these mixtures, but the differences are generally small. In all cases, the LOR_{95} with the MT array is at least as large as that with the ST array, and is a much as 2-fold larger. However, a t-test of the average LOR_{95} values from each array showed the difference not to be statistically significant. RCR_{10} values are comparable for the two arrays, though about 10% larger on average for the MT array. Interestingly, the largest RCR_{10} observed among these data is 36 (i.e., from 3:1 to 1:12), for the mixture of n-heptane + trichloroethylene (also seen for n-octane + trichloroethylene) with the 5-sensor MT array, which is somewhat larger than the largest RCR_{10} observed with the MT-U array. The extent of asymmetry observed is similar for both 5-sensor arrays and is comparable to that for the MT-U array. The strong correlations observed between values of *A* and the LOR-*S* product ratios for the MT-U array are also observed for the 5-sensor MT and ST arrays, as expected. The one marked difference between the ST and MT arrays is the lower LODs provided by the ST array (range = 30 to 160 ppm), which are generally about an order of magnitude lower than those of the MT array (range = 70 to 890 ppm). Thus, for the binary mixtures analyzed in this data set, the performance of ST and MT arrays is quite similar in most respects. The advantage in terms of LOR_{95} areas

provided by the MT array is offset to some extent by the advantage of the ST array with regard to sensitivity.

4.4 Conclusions

A number of conclusions can be drawn from this study related to the use of MT arrays as detectors in microanalytical systems. First, above the LOD, the relative concentration ratios of the vapors are much more important determinants of performance than are their absolute concentrations, and for binary mixtures the relative concentration range over which components can be reliably recognized is typically 5-20. When the purpose of the analysis is to determine all vapors in a complex mixture, this poses a significant constraint on trace analysis, because chromatographically unresolved binary mixture components at low relative concentrations cannot be detected reliably. The problem is even more severe for unresolved ternary mixtures. Furthermore, the analysis of binary (or ternary) mixtures whose components yield very different array sensitivities will be further constrained, favoring mixtures rich in the component for which the array is less sensitive. That is, on the basis of the data generated here, if the concentration of the component to which the array is more sensitive exceeds that of the vapor to which the array is less sensitive by 2-fold or more, the minority component will not be detectable.

If, on the other hand, the purpose is to perform targeted-vapor analysis and the target vapor cannot be resolved from interfering compounds chromatographically, as long as the sensitivity to that vapor is greater than its co-eluting interferences, it should be recognizable over a wide range of concentrations even if it is the minority component. If it is the majority component, and it is present in 2-fold or greater excess, its recognition

will be facilitated by the inability to detect the presence of the co-eluting interference. Note, however, that quantification may be affected, since the presence of the interference, though not recognized, will contribute to the net response and thereby tend to positively bias estimates of target-vapor concentrations.

The impact on recognition of the inherent variability in the sensitivity and baseline noise of the sensors in the array was found to be quite significant. Without tight controls on the sources of variation in responses and response patterns, the advantages of using sensor arrays to confirm the identities of vapors eluting from an upstream chromatographic column at a particular retention time will be lost. In practice, this demands low-noise circuitry with drift compensation and minimizing fluidic and thermal variables that may affect sensor responses directly or indirectly, such as through the delivery of samples to the sensor array.

The brief comparison of 5-sensor MT and ST arrays performed here with respect to LOR values revealed that the general behavior is quite similar to that of the 8-sensor MT array in terms of the range of recognizable compositions and the degree of asymmetry. As expected, the 5-sensor MT array provided somewhat larger LOR95 values, but the sensitivity advantage of the ST array was a notable extenuating factor.

This is only the second report to consider the systematic analysis of mixtures by arrays of microsensors using the LOR. This metric of performance permits detailed insights into array performance. The magnitude of the LOR facilitates ranking mixtures in terms of the ease with which they can be recognized and differentiated from their constituents and identifying mixtures where such discriminations are not possible. As

the trend toward an increasing emphasis on mixture analyses with microsensor arrays continues, this metric should find wider application.

Acknowledgments

The authors are indebted to Prof. Andreas Hierlemann of ETH, Zurich and to Dr. Petra Kurzawski of ETH, Zurich (currently at Silicon Microstructures Inc., Milpitas, CA) for providing the data used in this study and for helpful technical assistance. This research was supported through the Michigan Center for Wireless Integrated Microsystems (WIMS) by the Engineering Research Centers Program of the National Science Foundation under Award No. ERC-9986866.

Figures and Tables

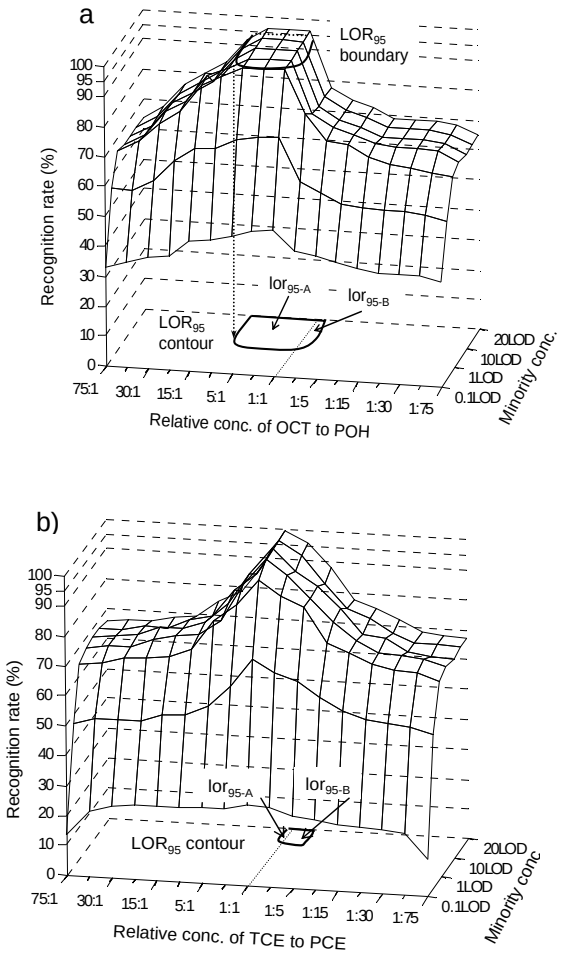

Figure 4-1. Recognition surfaces and projected LOR_{95} contours for two binary mixtures derived from simulated responses from the 8-sensor MT-U array: (a) n-octane + 1-propanol (LOR_{95}= 120); (b) trichloroethylene + perchloroethylene (LOR_{95} =16).

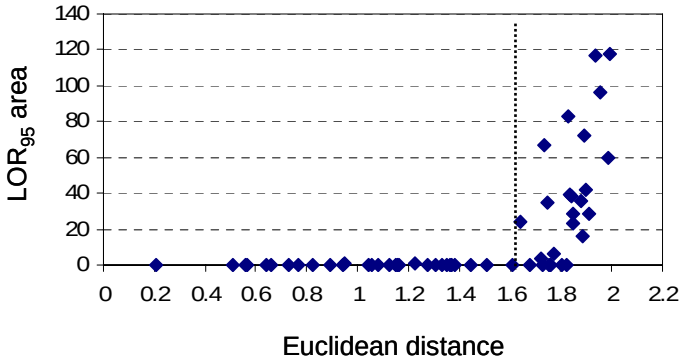

Figure 4-2: Relationship between the LOR95 and the Euclidean distance between the vectors of each mixture component for 55 binary mixtures. The vertical dotted line denotes the threshold Euclidean distance below which recognition is not possible.

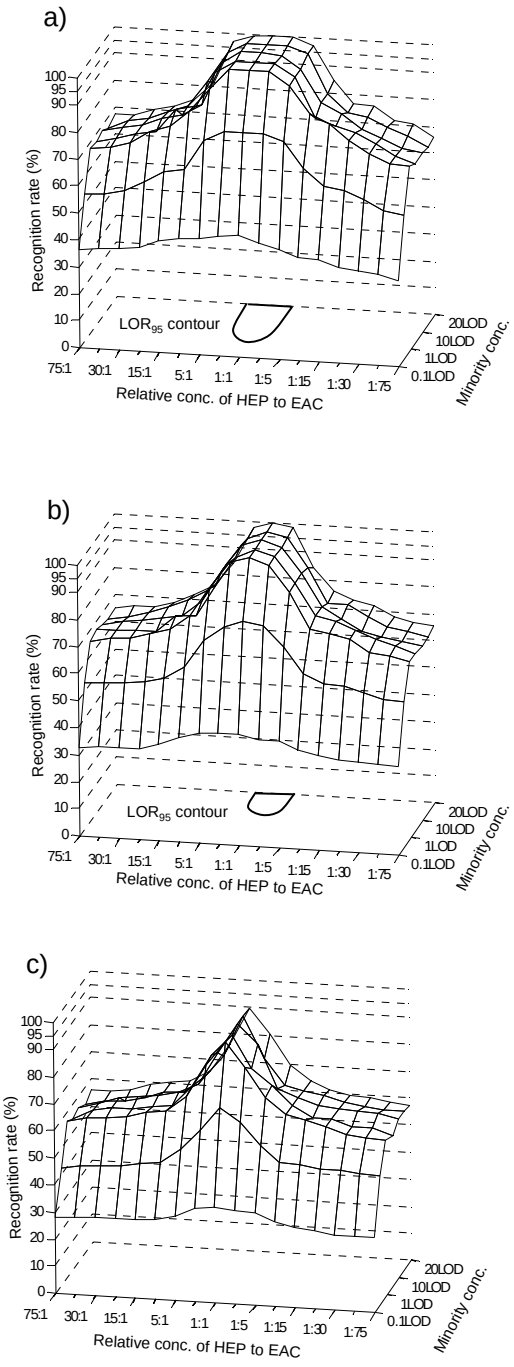

Figure 4-3: Recognition surfaces and LOR_{95} contours for the binary mixture of n-heptane and ethyl acetate as a function of assumed sensitivity error: a) 3% (LOR_{95} = 210); b) 5% (LOR_{95} = 88); and c) 10% (LOR_{95} = 0).

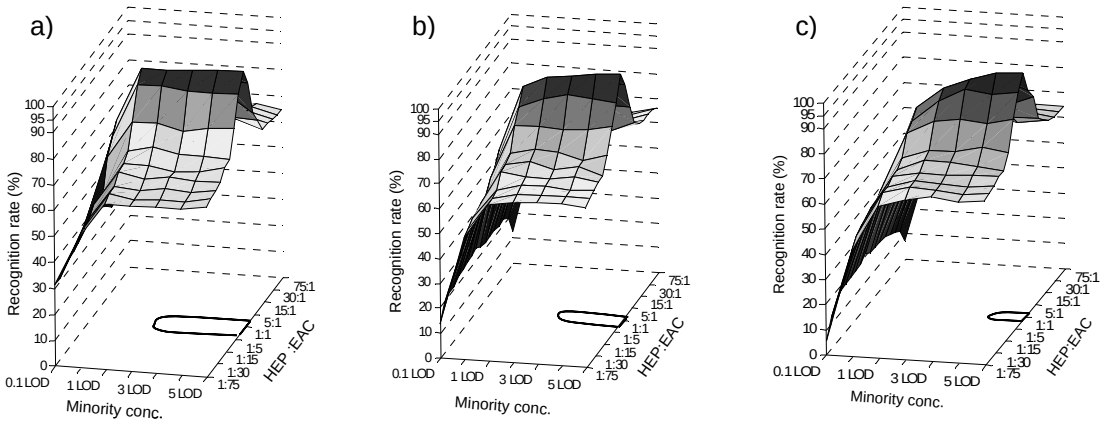

Figure 4-4: Effect of baseline noise levels on the LOR_{95} for the binary mixture of n-heptane + ethyl acetate: a) average experimental RMS noise (LOR_{95} = 56); b) 5 × RMS noise (LOR_{95} = 40); and c) 10 × RMS noise (LOR_{95} = 13). . Concentration range shown spans from 0.1-5LOD, but LOR_{95} is calculated from 1-5LOD.

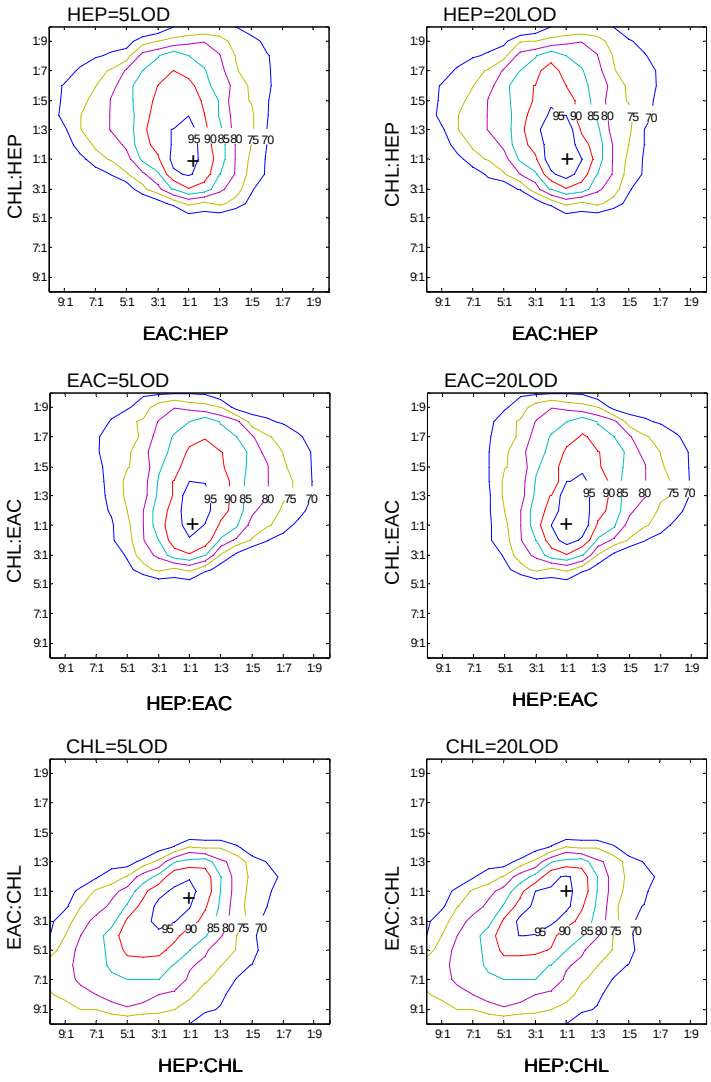

Figure 4-5. Cross-sectional LOR contours for the ternary mixture of n-heptane + chloroform + ethyl acetate at 5LOD and 20LOD for each of the three components. The x- and y-axes are the relative concentration ratios (in multiples of the LOD) of the indicated (varying) mixture component to the fixed component.

Table 4-1. LOR evaluations of 19 binary mixtures using simulated data from the 8-sensor MT-U array.

Components													
A	B	RR(%)[a]	LOR_{95}	LOD_A[b]	LOD_B[b]	S_A[c]	S_B[c]	E_{AB}[d]	e_A/e_B[e]	A[f]	P_A[g]	P_B[g]	P_A/P_B
HEP	MOH	97.4	42	2100	6500	0.28	0.43	1.9	1.6	6.2	590	2800	0.21
HEP	EOH	99.8	36	2100	2400	0.28	0.78	1.9	1.8	4.0	590	1900	0.31
HEP	POH	99.2	24	2100	1160	0.28	2.1	1.8	2.1	7.1	590	2460	0.24
HEP	EAC	100	56	2100	1100	0.28	0.42	1.6	0.85	0.43	590	460	1.3
HEP	CHL	99.8	120	2100	1300	0.28	0.78	1.9	1.7	1.5	590	1000	0.56
HEP	TCE	100	40	2100	620	0.28	0.50	1.8	0.64	0.23	590	300	1.9
OCT	MOH	98.2	24	770	6500	0.81	0.43	1.6	2.5	9.1	630	2800	0.22
OCT	EOH	99.6	24	770	2400	0.81	0.78	1.7	2.5	3.3	630	1900	0.33
OCT	POH	99.2	120	770	1200	0.81	2.1	2.0	1.6	4.0	630	2460	0.26
OCT	EAC	100	38	770	1100	0.81	0.42	1.8	0.72	0.29	630	460	1.4
OCT	CHL	99.8	28	770	1300	0.81	0.78	1.9	1.1	1.2	630	1000	0.60
OCT	TCE	99.8	66	770	620	0.81	0.50	1.7	0.46	0.18	630	300	2.0
TOL	TCE	98.2	28	220	620	0.32	0.50	1.8	2.1	5.6	70	300	0.22
EAC	CCL	99.8	60	1100	1200	0.42	0.12	1.9	0.39	0.15	460	140	3.2
EAC	PCE	99.8	72	1100	230	0.42	0.58	1.9	0.27	0.10	460	140	3.4
CHL	CCL	98.4	96	1300	1200	0.78	0.12	1.9	0.14	0.047	1000	140	7.3
CHL	PCE	97.8	84	1300	230	0.78	0.58	1.8	0.12	0.045	1000	140	7.7
CCL	TCE	100	34	1200	620	0.12	0.50	1.7	4.0	2.1	140	310	0.46
TCE	PCE	99.4	16	600	230	0.50	0.58	1.6	0.48	0.25	310	140	2.3

[a] average recognition rate determined at an RCR = 1:1 over a concentration range of 5-10LOD for each component; [b] units = ppm; [c] array sensitivity in units of response/ppm; [d] E_{AB} is the Euclidean distance between normalized vectors for vapors A and B; [e] e_A and e_B are the Euclidean distances between the vectors for components A and B and the vector for their 1:1 mixture, respectively; [f] *A* is an index of asymmetry defined as the ratio of $lor_{95\text{-}A}/lor_{95\text{-}B}$ (see Figure 1a); [g] *P* is the product of the LOD and S for a given mixture component.

Table 4-2: Range and asymmetry of relative concentration ranges evaluated at a minority-component concentration of 10LOD (RCR_{10}) over which the recognition rate of the mixture is ≥95%.[a]

Mixture Components		RCR_{10}	Range of RCR_{10}
A	B		10:1 9:1 8:1 7:1 6:1 5:1 4:1 3:1 2:1 1:1 1:2 1:3 1:4 1:5 1:6 1:7 1:8 1:9 1:10
HEP	MOH	6	(2:1) — (1:3)
HEP	EOH	5	(4:1) — (1:1)
HEP	POH	5	(9:1) — (1:2)
HEP	EAC	9	(6:1) — (1:2)
HEP	CHL	20	(16:1) — (1:1)
HEP	TCE	6	(3:1) — (1:2)
OCT	MOH	5	(2:1) — (1:9)
OCT	EOH	5	(2:1) — (1:3)
OCT	POH	20	(7:1) — (1:3)
OCT	EAC	8	(1:1) — (1:6)
OCT	CHL	5	(3:1) — (1:2)
OCT	TCE	12	(3:1) — (1:5)
TOL	TCE	5	(2:1) — (1:3)
EAC	CCL	12	(2:1) — (1:7)
EAC	PCE	14	(9:1) — (1:2)
CHL	CCL	18	(2:1) — (1:8)
CHL	PCE	16	(4:1) — (1:1)
CCL	TCE	5	(10:1) — (1:2)
TCE	PCE	5	(3:1) — (1:2)

[a] the column of entries for RCR_{10} is the fold-range of relative concentration ratios for each mixture for which RR ≥95%. Each line spans that RCR range and shows the degree of asymmetry relative to the RCR = 1:1 value when concentrations are expressed as multiples of the LOD for each component. Values in parentheses at the end of a line are the corresponding RCR values in units of ppm.

Table 4-3: LOR evaluations of the nine recognizable ternary mixtures at an RR threshold of 90%.

Mixture Components			RR(%)[a]	Cross-sectional LOR_{90} at Six Concentrations of A[b]						LOR_{90} Score [c]
A	B	C		0.1	0.5	1	5	10	20	
OCT	POH	EAC	91.0	0	0	1.2	5.5	6.1	6.0	4.7
HEP	EOH	TCE	91.2	0	0	4.8	5.0	6.0	6.1	5.5
HEP	POH	EAC	92.0	0	0	2.5	4.4	5.4	5.2	4.4
OCT	EOH	TCE	92.0	0	0	5.0	5.3	7.4	7.2	6.2
OCT	MOH	EAC	92.8	0	1.3	12	15	16	16	15
HEP	MOH	EAC	93.4	0	1.5	14	14	16	15	15
HEP	EAC	CHL	95.2	0	4.5	10	13	13	13	12
HEP	EOH	EAC	95.4	0	4.7	13	16	17	16	16
OCT	EOH	EAC	96.2	0	11	14	18	19	20	18

[a] average recognition rate determined at an RCR = 1:1:1 over a concentration range of 5-10LOD for each component; [b] concentration is defined in multiples of the LOD; [c] The score is the average cross-sectional LOR_{90} contour area determined from 1-20LOD.

Table 4-4: LOR evaluations of 19 binary mixtures using simulated data from the 8-sensor MT-U array.

Components A	B	RR(%)[a]	LOR_{95}	LOD_A[b]	LOD_B[b]	S_A[c]	S_B[c]	E_{AB}[d]	e_A/e_B[e]	A[f]	P_A[g]	P_B[g]	P_A/P_B[h]
HEP	CHL	99.8	84	890	190	0.26	0.28	0.91	0.14	0.16	230	52	4.4
OCT	CHL	99.8	80	150	190	0.66	0.28	1.2	0.25	0.2	100	52	2.0
OCT	EAC	99.6	46	150	480	0.66	0.27	1.0	2.3	2.5	102	130	0.8
HEP	TCE	98.4	96	890	180	0.26	0.21	0.91	0.09	0.13	230	37	6.2
OCT	TCE	98.2	100	150	180	0.66	0.21	1.1	0.22	0.26	100	37	2.8
EOH	CCL	98.0	52	410	430	0.26	0.09	0.48	0.26	0.21	110	39	2.8
HEP	EAC	97.8	56	890	480	0.26	0.27	0.60	0.11	0.15	230	130	1.8
CHL	CCL	97.0	56	190	430	0.28	0.09	0.86	0.35	0.38	52	39	1.3
EOH	PCE	96.8	76	410	70	0.26	0.43	0.95	0.22	0.18	110	29	3.7
CHL	PCE	96.6	70	190	70	0.28	0.43	0.86	0.18	0.22	52	29	1.8
TOL	EOH	96.4	10	130	410	0.20	0.26	0.94	5.3	7.5	26	110	0.2
POH	PCE	95.0	74	80	70	0.73	0.43	0.66	0.40	0.41	58	29	2.0

[a] average recognition rate determined at an RCR = 1:1 over a concentration range of 5-10LOD for each component;

[b] units = ppm; [c] array sensitivity in units of response/ppm; [d] E_{AB} is the Euclidean distance between normalized vectors for vapors A and B; [e] e_A and e_B are the Euclidean distances between the vectors for components A and B and the vector for their 1:1 mixture, respectively; [f] *A* is an index of asymmetry defined as the ratio of $lor_{95\text{-}A}/lor_{95\text{-}B}$ (see Figure 1a in text);

[g] *P* is the product of the LOD and S for a given mixture component. [h] The ratio of P_A/P_B is reversely correlated to asymmetry A with a r_2=0.835.

Table 4-5: Range and asymmetry of relative concentration ranges evaluated at a minority-component concentration of 10LOD (RCR_{10}) for 5-sensor MT array determined with 12 binary mixtures over which the recognition rate of the mixture is ≥95%.[a]

Binary Mixtures		RCR_{10}	Range of Limiting RCR																			
A	B		8:1	7:1	6:1	5:1	4:1	3:1	2:1	1:1	1:2	1:3	1:4	1:5	1:6	1:7	1:8	1:9	1:10	1:11	1:12	
HEP	CHL	22						(2:1)	—	—	—	—	—	—	—	—	—	—	—	—	—	(1:10)
OCT	CHL	20						(2:1)	—	—	—	—	—	—	—	—	—	—	—	(1:12)		
OCT	EAC	5			(2:1)	—	—	—	—	—	(1:3)											
HEP	TCE	36					(2:1)	—	—	—	—	—	—	—	—	—	—	—	—	—	—	(1:15)
OCT	TCE	36					(3:1)	—	—	—	—	—	—	—	—	—	—	—	—	—	—	(1:14)
EOH	CCL	5							(1:1)	—	—	—	—	(1:5)								
HEP	EAC	6							(2:1)	—	—	—	—	—	(1:3)							
CHL	CCL	6							(2:1)	—	—	—	—	—	(1:14)							
EOH	PCE	18					(1:1)	—	—	—	—	—	—	—	(1:18)							
CHL	PCE	16						(3:1)	—	—	—	—	—	—	—	—	—	(1:6)				
TOL	EOH	2						(2:1)	—	—	(1:3)											
POH	PCE	18						(2:1)	—	—	—	—	—	—	—	—	—	—	(1:7)			

[a] the column of entries for RCR_{10} is the fold-range of relative concentration ratios for each mixture for which RR ≥95%. Each line spans that RCR range and shows the degree of asymmetry relative to the RCR = 1:1 value when concentrations are expressed as multiples of the LOD for each component. Values in parentheses at the end of a line are the corresponding RCR values in units of ppm.

Table 4-6: LOR evaluations of 12 binary mixtures by a 5-sensor CAP ST array. These mixtures are those deemed recognizable by the 5-sensor MT array at a threshold RR of 95%.

Components A	Components B	RR(%)[a]	LOR_{95}	LOD_A[b]	LOD_B[b]	S_A[c]	S_B[c]	E_{AB}[d]	e_A/e_B[e]	A[f]	P_A[g]	P_B[g]	P_A/P_B[h]
HEP	CHL	97.4	70	60	40	0.73	0.60	0.85	0.10	0.13	46	23	2.0
OCT	CHL	97.4	64	30	40	1.8	0.66	1.0	0.16	0.17	46	25	1.8
OCT	EAC	99.0	44	30	40	2.2	0.61	1.1	0.32	0.31	57	25	2.3
HEP	TCE	96.8	76	60	70	1.4	0.35	0.94	0.09	0.09	91	23	4.0
OCT	TCE	95.2	78	30	70	3.5	0.38	1.0	0.26	0.1	91	25	3.6
EOH	CCL	93.6	0	90	160	0.46	0.64	0.69	5.3	6.9	44	100	0.4
HEP	EAC	98.4	44	60	40	0.90	0.56	0.98	0.35	0.35	57	23	2.5
CHL	CCL	96.4	50	40	160	1.2	0.29	0.81	0.93	1.1	44	46	1.0
EOH	PCE	96.0	76	90	40	0.80	2.6	0.98	3.1	4.9	76	100	0.7
CHL	PCE	97.2	38	40	40	2.0	1.2	0.96	0.64	0.89	76	46	1.7
TOL	EOH	95.0	5.2	130	90	0.78	3.0	1.1	3.7	5.3	100	280	0.4
POH	PCE	95.2	56	30	40	2.4	2.5	0.82	2.5	3.0	76	98	0.8

[a] average recognition rate determined at an RCR = 1:1 over a concentration range of 5-10LOD for each component;

[b] units = ppm; [c] array sensitivity in units of response/ppm; [d] E_{AB} is the Euclidean distance between normalized vectors for vapors A and B; [e] e_A and e_B are the Euclidean distances between the vectors for components A and B and the vector for their 1:1 mixture, respectively; [f] *A* is an index of asymmetry defined as the ratio of $lor_{95\text{-}A}/lor_{95\text{-}B}$ (see Figure 1a in text);

[g] *P* is the product of the LOD and S for a given mixture component. [h] The ratio of P_A/P_B is reversely correlated to asymmetry A with a r_2=0.894.

Table 4-7: Range and asymmetry of relative concentration ranges evaluated at a minority-component concentration of 10LOD (RCR_{10}) for 5-sensor CAP ST array determined with 12 binary mixtures.[a]

Binary Mixtures		RCR_{10}	Range of Limiting RCR
A	B		12:1 11:1 10:1 9:1 8:1 7:1 6:1 5:1 4:1 3:1 2:1 1:1 1:2 1:3 1:4 1:5 1:6 1:7 1:8 1:9 1:10
HEP	CHL	24	(20:1) — (1:1)
OCT	CHL	20	(7:1) — (1:3)
OCT	EAC	8	(6:1) — (1:2)
HEP	TCE	27	(9:1) — (1:3)
OCT	TCE	30	(8:1) — (1:4)
EOH	CCL	0	
HEP	EAC	8	(12:1) — (1:2)
CHL	CCL	15	(1:1) — (1:4)
EOH	PCE	27	(7:1) — (1:4)
CHL	PCE	16	(1:1) — (1:5)
TOL	EOH	5	(2:1) — (1:1)
POH	PCE	18	(1:1) — (1:7)

[a] the column of entries for RCR_{10} is the fold-range of relative concentration ratios for each mixture for which RR ≥95%. Each line spans that RCR range and shows the degree of asymmetry relative to the RCR = 1:1 value when concentrations are expressed as multiples of the LOD for each component. Values in parentheses at the end of a line are the corresponding RCR values in units of ppm.

Reference

1. a) Grate, J. W. *Chem. Rev.* 2000, *100*, 2627-2647; b) Park, J.; Groves, W. A.; Zellers, E. T. *Anal. Chem*. 1999, *71*, 3877-3886.

2. a) Senesac, L. R.; Dutta, P.; Datskos, P. G.; Sepaniak, M. J. *Anal. Chim. Acta* 2006, *558*, 94-101; b) Vančura, C.; Rüegg, M.; Li, Y.; Hagleitner, C.; and Hierlemann, A. *Anal. Chem.*, 2005, *77*, 2690-2699.

3. a) Kummer, A. M.; Hierlemann, A.; Baltes, H. *Anal. Chem.* 2004, *76*, 2470-2477; b) Mlsna, T. E.; Cemalovic,S.; Warburton, M.; Hobson,S. T.; Mlsna, D. A.; Patel, S. V. *Sens. Actuators B*, 2006, *116*, 192-201.

4. Lerchner, J.; Seidel, J.; Wolf, G.; Weber, E. *Sens. Actuators B* 1996, *32*, 71-75.

5. a) Severin, E. J.; Lewis, N. S. *Anal. Chem.* 2000, *72*, 2008-2015; b) Patel, S. V.; Jenkins, M. W.; Hughes, R. C.; Yelton, W. G.; Ricco, A. J. *Anal. Chem.* 2000, *72*, 1532-1542.

6. Hsieh, M.-D.; Zellers, E. T. *Anal. Chem.* 2004, *76*, 1885-1895.

7. a) Hsieh, M. D.; Zellers, E. T. *J. Occup. Environ. Hyg*. 2004, *1*, 149-60; b) Park, J.; Zhang, G. Z.; Zellers, E. T. *Am. Ind. Hyg. Assoc. J.* 2000, *61*,192-204.

8. Zellers, E. T.; Batterman, S. A.; Han, M.; Patrash, S. J. *Anal. Chem*. 1995, *67*, 1092-1106.

9. Pinnaduwage, L. A.; Zhao, W.; Gehl, A. C.; Allman, S. L.; Shepp, A.; Mahmud, K. K.; Leis, J. W. *Appl. Phys. Lett.* 2007, *91*, 044105-1:044105-3.

10. Ricco, A. J.; Crooks, R. M.; Osbourn, G. C. *Acc. Chem. Res*. 1998, *31*, 289-296.

11. Then, D.; Vidic, A.; Ziegler, Ch. *Sens. Actuators B* 2006, *117*, 1-9.

12. Hierlemann, A.; Weimar, U.; Kraus, G.; Schweizer-Berberich, M.; Gopel, W. *Sens. Actuators B* 1995, *26-27*, 125-134.

13. Kurzawski, P.; Hagleitner, C.; Hierlemann, A. *Anal. Chem*. 2006, *78*, 6910-6920.

14. Ulmer, H.; Mitrovics, J.; Weimar, U.; Göpel, W. *Sens. Actuators B* 2000, *79*, 79-81.

15. Schierbaum, K. D.; Gerlach, A.; Haug, M.; Göpel, W. *Sens. Actuators A* 1992, *31*, 130-137.

16. Mitrovics, J.; Ulmer, H.; Weimar, U.; Göpel, W. *Acc. Chem. Res.* 1998, *31*, 307-315.

17. Jin C., Kurzawski P., Hierlemann A., Zellers E. T. *Anal. Chem*. 2008, *80*, 227-236.

18. Zellers, E. T.; Park, J.; Hsu, T.; Groves, W. A. *Anal. Chem.* 1998, *70*, 4191-201.

19. Park, J.; Groves, W. A.; Zellers, E. T. *Anal. Chem*. 1999, *71*, 3877-3886.

20. Lu, C. J.; Whiting, J.; Sacks, R. D.; Zellers, E. T. *Anal. Chem*. 2003, *75*, 1400-1409.

21. The LOR_{95} is determined on the basis of RR values at rather widely spaced discrete coordinates along both the RCR and LOD axes, raising a question about the precision of the area calculation. However, a series of comparisons was made to show that the area calculated this way differed by only 3%, on average, from the area calculated on the basis of RR values at unit resolution on both axes.

CHAPTER 5

Multivariate Curve Resolution of Chemical Sensor Array Signals

5.1 Introduction

5.1.1 Peak overlapping and resolutions. A more or less universal concern in the practice of chromatography is the high probability of overlapping peaks. [1, 2] This problem is almost inevitable considering one or two of the following common situations: 1) when a complex mixture is involved; 2) fast separation is valued and short column is employed. Davis and Giddings have proven that theoretically a chromatogram must be more than 95% vacant in order to provide greater than 90% confidence that a component of interest will appear as an isolated peak. [3] A closely related issue is peak purity which is complicated by the embedding of minor peak under an isolated peak due to co-elution. The accuracy of qualitative and quantitative chromatographic analyses relies on pure peaks for accurate results. Frequently, peak overlapping can be so severe that no obvious evidence of contamination may be present. [3-6] Peak overlapping problem can be addressed via hardware or software. The former improves chromatographic resolution by optimizing conditions, e.g., by changing column, increasing column length, programming oven temperature, or decreasing the flow rate of mobile phase. Obviously it takes time, effort and experience to reach an optimal separation and the outcome may not be a

satisfactory resolution. The other approach is numerical deconvolution by means of software. This software approach resolves overlapped peaks from each other mathematically.

The traditional peak deconvolution method is peak fitting, which is commonly used with single-channel chromatographic detectors like flame ionization detector (FID) or electron capture detector (ECD). [7, 8] Peak fitting attempts to achieve a non-linear least square fit of the chromatographic data by assuming a peak shape function. Many peak shape functions have been developed on the basis of the Gaussian distribution, but none can provide a peak shape function of general applicability to describe the peaks of different analytes in a real chromatogram. [9] The number of functions, i.e., the number of analytes attributable to the overlapped peaks has to be also known a priori. The more functions that are used in the fitting, the better the fitting will turn out. Due to these facts, peak deconvolution provided by peak fitting is at most a rough approximation to the underlying analytes.

Multivariate curve resolution (MCR) based on factor analysis is capable of deconvoluting seriously overlapped peaks without any assumptions concerning peak shape. The only requirement is a multivariate response vector generated per sample by a multi-channel detector like spectroscopic diode array detector, mass spectrometer and microsensor array. Unlike the classical single-channel detector that yields one-dimensional data per sample (a vector consisting of one response at a time along the retention time direction), multi-channel detector provides a data matrix per sample composed of responses of all detector channels along the retention time direction. This second dimension requires chemometric data analysis and enables the application of

MCR to deconvolution of overlapped peaks. [11,12] MCR has been applied extensively to two-dimensional (or two-way) data obtained from chromatography hyphenated with multi-channel detection such as GC-MS and HPLC-diode array detection (DAD) , [13-17] as well as spectroscopy such as infrared (IR), Raman, and near-infrared (NIR) spectroscopy. [18-28] The application of MCR to microsensor array based GC detection systems has not yet been reported.

Although the majority of MCR research and applications are performed in the spectroscopy field, MCR can be applied to sensor array data analysis with few modifications. The spectrum recorded by a spectrometer is analogous to the response pattern measured by a sensor array detector. Therefore the terminology "spectrum" and "response pattern" are used interchangeably here. The only difference is the availability of independent detector channels. Spectroscopic data may contain thousands of wavelengths and usually there is a large degree of collinearity among these wavelengths. On the contrary, the number of sensors in a sensor array detector is small and the inter-sensor dependence can be minimized by careful selection of sensing materials.

5.1.2 Factor analysis. As Malinowski said, "One of the most difficult problems concerns the chemical analysis of mixtures containing an unknown number of unknown compounds." [10] Deconvolution of overlapped chromatographic peaks is one of such mixture resolution problems that can be addressed by multivariate curve resolution. The use of factor analysis for multivariate curve resolution can be well summarized by "The ultimate goal of curve resolution would be able to determine the number of components in an overlapping chromatographic peak as well as the spectrum and concentration profile

of each component, without assumption regarding peak shape, location or identity." [29] Mathematically factor analysis reduces matrices of data to their *lowest* dimensionality by the use of orthogonal factor space and transformations that yield recognizable factors. [10] In matrix format the factor analysis model can be expressed as:

$$\boldsymbol{X}=\boldsymbol{C}\boldsymbol{S}'+\boldsymbol{E} \qquad \text{Eq. 5-1}$$

Matrix $\boldsymbol{X}$ ($n\times k$) contains response data of k sensors measured for n mixtures. Each mixture consists of the same a components and their concentrations change from mixture to mixture. Concentration matrix $\boldsymbol{C}$ ($n\times a$) has the concentration profile of each component as its columns. Sensitivity matrix $\boldsymbol{S}$ ($k\times a$) contains the sensitivities of k sensors to all a components. $\boldsymbol{E}$ is the residual matrix containing the random residual noise left after extracting a components from the mixture data.

For an overlapping peak of a components in a chromatogram recorded by a sensor array, if n response vectors (or spectra) are measured with k sensors as a function of elution time, they can be arranged row by row into an $n\times k$ data matrix $\boldsymbol{X}$. The rows of $\boldsymbol{X}$ are n ordered response vectors measured at regular time intervals and the k columns are chromatograms measured at different sensor channels. Accordingly one may consider this data matrix to be composed of a number of sequentially recorded response vectors (or spectra). Plotting any row in $\boldsymbol{X}$ displays the mixture response pattern (or spectrum) of the analytes that co-elute at that specific time. Alternatively it may be seen as a collection of chromatograms measured at different sensor channels. Plotting any column displays the chromatogram of the co-eluting analytes at that specific sensor. Residual matrix $\boldsymbol{E}$ is

of the same dimension as $\boldsymbol{X}$, consisting of random noise of each sensor measured at each time point. The product of $\boldsymbol{C}$ and $\boldsymbol{S}$ contains the systematic structural information in $\boldsymbol{X}$ corresponding to the responses caused by a overlapping components. A fair assumption made here about the signal produced by sensors in the array for any mixture is that the signal is additive and bilinear. Additive means the systematic part of the measured signal can be regarded as individual contributions from each component. Let $\boldsymbol{X}_i$ be the contribution from component i (i=1,2,…,a), the additive assumption for Equation 5-1 may be written as:

$$X = X_1 + X_2 + ... + X_a + E = \sum_{i=1}^{a} X_i + E \qquad \text{Eq. 5-2}$$

The bilinearity assumption implies that the signal from any chemical component can be seen as the outer product of two vectors: C_i describing the concentration profile and S_i describing the spectral profile (response pattern) for the component i (i=1,2,…,a), which is summarized as:

$$\boldsymbol{X}_i = C_i S_i' \quad (i=1,2,..,a) \qquad \text{Eq. 5-3}$$

Both assumptions lead to another formation of Equation 5-1:

$$\boldsymbol{X} = C_1 S_1' + C_2 S_2' + ... + C_a S_a' + E = \sum_{i=1}^{a} C_i S_i' + E \qquad \text{Eq. 5-4}$$

The matrix $\boldsymbol{C}$ in Equation 5-1, with C_i as its columns implied above, contains the concentration profiles of all components contributing to $\boldsymbol{X}$. The matrix $\boldsymbol{S}$ contains, with S_i as columns, the spectral profiles (or response patterns) of all components. Due to the random experimental noise there is always a slight difference between $\boldsymbol{X}$ and $\boldsymbol{CS}'$. This difference is referred to residual matrix $\boldsymbol{E}$.

Naturally one can relate Equation 5-4 to Beer-Lambert's law in spectrometry. Thereby the physical significance of factor analysis model is obvious. It has to be noted at this point that the factor analysis for multivariate curve resolution is not restricted to mixture resolution in analytical spectrometry as implied by Beer-Lambert's law. The application of factor analysis has been expanded to receptor model in environment, [30,31] process control and monitoring, [32,33] kinetic and equilibrium study in chemistry, [34-36] and mixture resolution in biology. [37,38] In this study we will focus on resolution of overlapping peaks in chromatography hyphenated with sensor array detector.

The first application of factor analysis to chemical problem can be tracked to the pioneering work of Lawton and Sylvestre under the name of self-modeling curve resolution. [39] They approached peak deconvolution by taking advantage of the evident fact of non-negativity of the absorption spectra and concentration profiles, yielding a range of possible solutions which are narrow enough to be chemically useful. The big disadvantage of the self-modeling curve resolution is its intrinsic limitation to two-component systems. Although there had been a few following endeavors to study three components, [40-44] the additional assumptions incurred make this approach unrealistic and too cumbersome to resolve complex mixtures. The geometric approach originally proposed for curve resolution is seldom used nowadays, but the terminology "self-modeling curve resolution" still survives and applies to a variety of factor-analysis based curve resolution methods. As its name implies, self-modeling curve resolution does not require assumptions about peak shape or spectra a priori, except some general chemical information about the system under investigation like non-negativity of concentrations and unimodality of peaks.

Another group of multivariate curve resolution methods includes iterative target transformation factor analysis (ITTFA) and its derivatives. [45-47] These methods have no general restriction about the maximal number of components. They rely on an initial estimate of concentration profiles of all components and calculate spectra profiles in an iterative way. First, initial estimates for the concentration profiles ***C*** for all components have to be made, and then they are refined by iterative calculations of the target transform of ***C***. Convergence to an acceptable solution is usually slow and sometime no convergence can be made at all, depending on the choice of starting concentration profiles. Different techniques such as needle search, [45] or varimax rotation of the eigenvectors, [46] have been explored to give starting concentrations of less ambiguity.

5.1.3 Evolving Factor Analysis. All these aforementioned problems have been largely settled by the method of evolving factor analysis (EFA) when it is applied to hyphenated chromatographic data. EFA has been shown to provide the exact region of existence for each component in a complex mixture presented in the form of overlapping chromatographic peaks. This information is subsequently exploited to generate a unique solution of concentration profiles ***C*** and spectra profile ***S***. Initially EFA was developed to determine equilibrium constants from spectrometric titration curves. [48] The method also proved useful in equilibrium studies, chromatography and flow injection analysis. [48-50] One can consider two steps in EFA application, the first step being the detection of the number of compounds present and their regions of existence in the chromatogram, and the second step being the determination of the profiles (concentration profile and spectral profile). Therefore EFA can help with peak purity evaluation by judging the number of

components contributing to the peak. However, a better peak purity assessment method, fixed-size moving window evolving factor analysis (FSMW-EFA), a variant of EFA to be discussed later, will be used instead for peak purity evaluation purpose.

In a multi-channel chromatogram, the spectra are recorded in the sequence of elution. The intrinsic order in which the rows of $\boldsymbol{X}$ are arranged contains additional information which is not exploited in traditional factor analysis. The fundamental idea of EFA is to follow the change in the rank of the data matrix $\boldsymbol{X}$ as a function of retention time, which is done by principal component analysis (PCA) on an increasing data matrix. It works as follows. Starting from the first row of $\boldsymbol{X}$, i.e., the first spectrum or response vector measured, the eigenvalues (EVs) are successively calculated for all the sub-matrices $\boldsymbol{X}(i)$ which are formed by the first i (i=1,2,…,n) rows, according to a standard principal component analysis (PCA) model:

$$\boldsymbol{X}(i)=\boldsymbol{TP}+\boldsymbol{F} \qquad \text{Eq. 5-5}$$

where $\boldsymbol{T}$ ($i\times a$) is the score matrix and $\boldsymbol{P}$ ($a\times a$) is the matrix of loadings. $\boldsymbol{F}$ is the residual matrix after extracting a principal components from $\boldsymbol{X}(i)$. EVs are not a direct measure of the concentrations. They are strongly depending on the dissimilarity of the elution and/or spectra profiles. The appearance of each new component at the detector is intrinsically associated with the increase of the rank of matrix $\boldsymbol{X}(i)$ by one. There has to be a minimal difference in the retention times and the spectra have to be linearly independent, so that the evolution of rank can be accurately monitored. For the first sub-matrix $\boldsymbol{X}(1)$ which is built by the first row of $\boldsymbol{X}$, EVs can be calculated by Equation 5-5. A second row is then

added and the EVs are calculated again. Likewise, a row is added and the EVs are calculated until PCA is performed on the whole matrix of $\boldsymbol{X}$. Since the process of factor analysis evolves forward on the time direction, it is called forward EFA. The evolution of EVs with time can be observed by plotting logarithms of EVs against retention time (also known as EFA plot). At the time point where an EV rises out of the noise level, a new component appears, leading to an increase of the rank by one. The point where the second EV rises out of the noise indicates the presence of the second component, and so on. In fact, the number of components can be easily determined by the visual inspection of EFA plot: the number of EVs above the noise level, i.e., the rank of $\boldsymbol{X}$, equals the number of underlying components. In analogy to forward EFA, one can also perform factor analysis starting from the last row of $\boldsymbol{X}$; this is called backward EFA. Starting with the last row measured, the EVs are again determined by means of Equation 5-5 for all sub-matrices $\boldsymbol{X}(i)$, formed by the last i (i=1,2,…n) rows of $\boldsymbol{X}$. The logarithms of EVs can then be plotted against time from which one can determine disappearance of compounds. An increase of an EV above the noise level indicates when a compound leaves the detector. Thus EFA provides information on both the appearance and the disappearance of the components along retention time in the chromatogram. When forward EFA and backward EFA are plotted together, a rank map can be produced to help illustrate the evolution of system rank with the progress of elution. One example of EFA rank map is illustrated in Figure 5-1 for a binary mixture.

In order to correctly associate the appearance of a given component with its disappearance, i.e., to determine its region of existence in the chromatogram, it is assumed that the first compound present in the system will be the first to disappear. In

good chromatography where column is not overloaded, peak width monotonically increases with retention time. Therefore the first component will be the first to disappear and the second component will be the second to disappear, etc. Thus a concentration window (or elution window) is generated for the i^{th} component from the time point where the rank of $\boldsymbol{X}$ rises to i in the forward calculations to the time point in the backward EFA plot where the rank rises to $n+1-i$. Outside the window the concentration is known to be zero. In Figure 5-1, the first component starts elution around retention time 16 and then disappears around time 62. Outside the elution window of 16-62 its concentration is zero. The second components starts elution around retention time 26 and disappear around time 73. Its concentration is zero outside the elution window 26-73.The zero-concentration (or zero-component) regions containing only noise possess rank of zero. Note that the noise level here is defined by the level of non-chemical variation in the data due to random measurement errors. Selective or single component regions have rank of one. Non-selective regions, single component regions with non-linear relation between measured response and concentration, or single analyte regions with heteroscedastic noise, have rank equal to or larger than 2, ideally depending on the number of components present [32]. This knowledge of elution windows will be used for calculation of spectra and elution profiles in the subsequent resolution step.

It is critical to determine the noise level so that the emerging of a new component can be judged by its eigenvalues rising above the noise. The noise level is normally estimated as the largest EV which is caused by noise and obtained when PCA is performed on the whole data matrix $\boldsymbol{X}$, which also corresponds to the singular value of the sub-matrix consisting of the zero-component region in a rank map. Once the noise

level is determined correctly, any EV that is beyond this level is deemed as a new component and thereby the number of underlying component in the mixture is also determined. For the purpose of peak purity assessment (EFA has been further developed into fixed size moving window evolving factor analysis (FSMW-EFA) which will be introduced later).

Based on the rank map information provided above, a non-iterative calculation proposed by Maeder was used for numerical resolution of $\boldsymbol{X}$ into elution profile matrix $\boldsymbol{C}$ and spectra matrix $\boldsymbol{S}$. [32] Matrix $\boldsymbol{X}$ can be decomposed either according to Equation 5-1 with real factors of physical meaning or according to Equation 5-5 with abstract eigenvectors for the whole data set $\boldsymbol{X}$. Since the abstract eigenvectors need to be transformed into real factors, the goal is to find a rotation matrix $\boldsymbol{R}$ ($a \times a$) which permits the recalculation of the matrix of column factors $\boldsymbol{C}$ from the scores $\boldsymbol{T}$ by means of:

$$\boldsymbol{C=TR} \qquad \text{Eq. 5-6}$$

Combining Equation 5-1 and Equation 5-5 yields:

$$\boldsymbol{CS=TP} \qquad \text{Eq. 5-7}$$

Multiply from right with the orthonormal loading matrix $\boldsymbol{P}'$, one has:

$$\boldsymbol{CSP'=TPP'=T} \qquad \text{Eq. 5-8}$$

which can also be written as:

$$\boldsymbol{C}=\boldsymbol{T}(\boldsymbol{SP}')^{-1} \qquad \text{Eq. 5-9}$$

Compare Equation 5-6 and 5-9, one gets:

$$\boldsymbol{R}=(\boldsymbol{SP}')^{-1} \qquad \text{Eq. 5-10}$$

Once the sensitivity matrix $\boldsymbol{S}$ is known, the rotation matrix $\boldsymbol{R}$ can be easily calculated by Equation 5-10 via least squares regression. However, both $\boldsymbol{C}$ and $\boldsymbol{S}$ are not known in practice. $\boldsymbol{R}$ has to be found via another way from experimental data $\boldsymbol{X}$. To find $\boldsymbol{R}$ one decomposes Equation 5-6 into a equations

$$C_i=\boldsymbol{T}R_i \qquad \text{Eq. 5-11}$$

where C_i is the i^{th} column of $\boldsymbol{C}$, R_i is the i^{th} column of $\boldsymbol{R}$ (i=1,2,...a). By itself Equation 5-11 cannot be solved. However, if only those rows of C_i with zero elements are extracted and if the corresponding rows of $\boldsymbol{T}$ are combined into matrix $\boldsymbol{T}_\mathrm{i}^0$ (h×a, h is the number of rows in C_i with zero elements), one obtains a set of equations which can be solved:

$$C_\mathrm{i}^0=\boldsymbol{T}_\mathrm{i}^0R_\mathrm{i} \qquad \text{Eq. 5-12}$$

where C_i^0 is a zero vector and h is the number of rows where component i is not present. The idea of this procedure is to find consequently one column of $\boldsymbol{R}$ after the other. By extraction of zero vector C_i^0 the corresponding vector R_i can be obtained.

Equations 5-6 to 5-12 are depicted schematically in Figure 5-2, using a method for the representation of the matrix operations introduced by Maeder. [32] Each column of $\boldsymbol{C}$ may be seen as a concentration profile of a pure component. The unshaded area in each column indicates the presence of the component, i.e. its concentration window. Accordingly, the shaded area represents the absence of the substance (Fig. 5-2a). Fig. 5-2b depicts Equation 5-11 and shows that the shaded part of C_i is a linear combination of the shaded part of $\boldsymbol{T}$. The shaded parts of C_i are combined into the zero vector C_0 and the corresponding ones of $\boldsymbol{T}$ into $\boldsymbol{T}_0$ (Fig. 5-2c). Since component i is not present in the column vector C_0 this component is not contributing to $\boldsymbol{T}_0$, and the rank of $\boldsymbol{T}_i^0$ is a- 1.

Equation 5-12 is not of full rank, so it has the trivial solution $R_i = 0$ and also an indefinite number of non-trivial solutions. Although quantification is not possible, one is still able to obtain the relative factor profiles (either concentration profile or spectrum/response pattern). In hyphenated chromatography/sensor array system, this will lead to the response patterns and the peak shapes of overlapping components. These non-trivial solutions can be obtained by assigning an arbitrary value to one of the coefficients r (one of elements of R_i) and computing the other values by a simple regression. Having found the elements of $\boldsymbol{R}$, the matrix $\boldsymbol{C}$ can easily be recalculated by means of Equation 5-6. Thus, concentration profile matrix $\boldsymbol{C}$ is calculated non-iteratively and without any assumptions about its shape.

The last step of EFA is the calculation of $\boldsymbol{S}$ by a simple linear regression. Given $\boldsymbol{X}$ and $\boldsymbol{C}$ in Equation 5-1, the sensitivity matrix $\boldsymbol{S}$ can now be obtained via the pseudo-inverse of $\boldsymbol{C}$:

$$\boldsymbol{S}=(\boldsymbol{C}'\boldsymbol{C})^{-1}\boldsymbol{C}'\boldsymbol{X} \qquad \text{Eq. 5-13}$$

This non-iterative method relies on the combination of small elution sections of the data set to obtain either the elution profiles or the pure spectra of the components. These small sections have special properties and they are built from the knowledge of the elution windows of the different components provided by EFA rank map. Once the profiles in the ***C*** or in the ***S*** matrix are recovered, the complementary matrix (***S*** or ***C***, respectively) is estimated through a single least-squares step as above in Equation 5-13. Other non-iterative methods based on EFA include window factor analysis (WFA) by Malinowski and coworkers, [51,52] subwindow factor analysis (SFA) by Manne and co-workers, [53,54] and heuristic evolving latent projections (HELP) by Kvalheim and co-workers. [55,56] Recent algorithms developed from these parent approaches include orthogonal projection resolution (OPR) from WFA, [57] or parallel vector analysis (PVA) from SFA. [58]

It should be noted at this point that the determination of absolute concentrations of components is not possible without calibration. It is believed that as long as a component can be resolved from the others, it presents with a concentration above the limit of detection of the detector and therefore can be quantified with careful calibration. Concentration profiles in matrix ***C*** and spectra in ***S*** are normalized. While the quantitative information will be lost after normalization, this operation will not affect qualitative analysis for analyte identification purpose.

5.1.4 Fixed Size Moving Window-Evolving Factor Analysis. An important family of methods derived from EFA performs PCA on windows of fixed size that are

moved along the data set. The most widely used approach of this kind is fixed size moving window-evolving factor analysis (FSMW-EFA), developed by Keller et al. [59,60] As shown in Fig. 5-3b, PCA analyses are performed on *fixed size* windows moved row by row downwards along the elution direction of the data set. The FSMW-EFA plot shows the eigenvalues obtained in all the PCA analyses as it was done in EFA. This representation contains the information on the component overlap in the elution direction. Thus, elution ranges where two eigenvalue lines arise from the noise level indicate the presence of two overlapped compounds. In general, the number of compounds overlapping in a certain time range equals the number of eigenvalue lines above the noise level.

FSMW-EFA is particularly useful for the detection of selective elution regions for the different compounds, i.e., zones where only one compound is present. When such ones are present, obtaining the response pattern of the related components is straightforward. FSMW-EFA is better suited for detecting impurities or minority components than is EFA due to the local analysis of small elution windows. We already saw in Fig. 5-1 that the EVs associated with a chemical substance show a sharp increase as a function of time due to the elution of a compound, while those caused by noise increase only slightly. The reason for the increase of the latter is that PCA is performed on a sub-matrix comprising an increasing number of spectra, and noise therefore accounts for more and more of the variability in the data. In EFA the EVs due to noise increase with the size of sub-matrix, making it difficult to distinguish between an EV from noise and one from a minor component, and imposing limitations on applications of EFA. This problem is largely settled by FSMW-EFA. Since the number of spectra analyzed in

FSMW-EFA is fixed, the noise will also be constant and the EVs due to noise are represented as a horizontal line as seen in Fig. 5-3b. Refer to Figure 5-3 for a comparison of EFA and FSMWEFA. [61] This fixed-size window operation makes FSMW-EFA sensitive to the detection of minor components. This sensitivity explains the superiority of this method to address peak purity problems.[59, 62-64] To maximize the sensitivity of FSMW-EFA to minority components, the size of window that moves downwards the data set must be chosen via a trial and error process, despite a size of seven recommended in some spectroscopic analyses. [59, 60]

At this point, it is important to be reminded that the estimation of the number of factors or significant eigenvalues in data sets and the identification with the number of co-eluting components may present difficulties in some situations, when there is a strong heteroscedastic noise, [65] distortion of the linear sensor response, [66] or non-chemical contributions to the sensor signal such as drift of baselines and other unidentified systematic variations. The visual output of these EV-based methods can warn about the presence of these artifacts and data preprocessing procedures can help solve many of these problems. [67]

5.1.5 Alternating Least Squares. EFA represents one group of non-iterative MCR methods that finds a unique solution without any iterative process. Another group of multivariate curve resolution methods uses alternating least squares (ALS) method to optimize resolution via an iterative algorithm. Although a unique solution can be found by EFA it is not necessarily the optimal one. The solution provided by non-iterative EFA curve resolution may be further refined by ALS.

Iterative resolution approaches to MCR have the flexibility to cope with many kinds of data structures and chemical problems and the ability to accommodate external information in the resolution process. [68-72] All of them share a common step of optimization (of $\boldsymbol{C}$ and/or $\boldsymbol{S}$ matrices) that starts from initial estimates of $\boldsymbol{C}$ or $\boldsymbol{S}$ and evolves in each iteration cycle to yield optimal concentration profiles and response patterns, tailored according to chemical or mathematical information included in the iterative optimization process in the form of constraints. [73-75] The iterative optimization ends when there is no further improvement in the profiles or patterns recovered or when the reproduction of $\boldsymbol{X}$ from the product $\boldsymbol{CS}'$ is satisfactory or a convergence criterion is met.

ALS was developed by Tauler and co-workers [68-70] based on alternating regression [76, 77] and was among the first iterative approaches proposed, and is still the most common MCR algorithm used in chemometric software applications. ALS is based on iterating between the following two equations:

$$\boldsymbol{C}=\boldsymbol{XS}(\boldsymbol{S}'\boldsymbol{S})\boldsymbol{S}^{-1} \qquad \text{Eq. 5-14}$$

$$\boldsymbol{S}'=(\boldsymbol{C}'\boldsymbol{C})^{-1}\boldsymbol{C}'\boldsymbol{X} \qquad \text{Eq. 5-15}$$

Equation 5-14 is the least squares solution of the equation $\boldsymbol{X}=\boldsymbol{CS}'$ with regard to $\boldsymbol{C}$, provided that $\boldsymbol{S}$ is known. Equation 5-15 is the least squares solution to the same equation assuming $\boldsymbol{C}$ is known. Of course, neither $\boldsymbol{C}$ nor $\boldsymbol{S}$ is known prior to curve resolution. By iterating between Equation 5-14 and 5-15, starting from an initial estimate of either $\boldsymbol{C}$ or $\boldsymbol{S}$, one approaches an optimized solution. Karjalainen suggests using random numbers for

starting estimates. [77] While this may be the best starting point in a strict statistical sense, in as much as it provides an unbiased solution, other approaches have surfaced. Among these are needle vectors [45] and key profiles [78] mentioned earlier with ITTFA. If external estimates of the elution profile or response pattern of any of the components are available, these estimates should of course be used. Response patterns can be established through selective regions found with EFA, or using other pure variable selection methods like SIMPLSMA (simple-to-use interactive self-modeling analysis) [79, 80] or orthogonal projection approach. [81] Once the initial estimates are selected, Equation 5-13 provides the concentration profiles corresponding to the starting estimate of $\boldsymbol{S}$. Because the initial response patterns may be poor estimates of the true underlying spectra, the concentration profiles provided by Eq. 5-13 often contain regions of unlikely features, such as negative regions, multi-modal profiles for a chromatographic peak, and so forth. The unlikely features are removed from the profiles during ALS iterations. As a general rule, the solution improves as the number of constraints increases.

The common constraints in hyphenated chromatographic techniques that are also used in this study are illustrated in Figure 5-4. Apart from the application of non-negativity to concentration profiles and response patterns, as imposed on many other chemical problems, the particular properties of the evolutionary elution allow for the introduction of constraints related to the peak shape or to the sequential elution pattern of compounds. Thus, unimodality can be selected to preserve the presence of only one maximum in each chromatographic profile and the knowledge of the elution windows can help to set selective elution regions (time channels where only one compound elutes) or to introduce local rank information (time channels where one or several compounds are

absent). For some particular sensor channels whose response may be negative, ALS can be tailored accordingly by including negativity constraint to these sensors in ALS iterations. Additional peak shape constraints can also be employed by fitting chromatographic data to a particular shape defined by a mathematical function. After the concentration profiles have been adjusted according to the constraints, Equation 5-15 is used to calculate the spectral profiles corresponding to the adjusted concentration profiles. Again, the spectral profiles are adjusted according to similar constraints. This cycle continues until convergence, which means that the calculated profiles satisfy all constraints. It may be necessary to go back and check the initial estimates in case convergence is not reached or the convergence is slow.

The criterion of convergence of ALS algorithm is usually determined by the changes of the residuals after every iteration cycle. The residual is defined by SSR, sum of square of residuals:

$$\text{SSR}=\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{k} e_{ij}^{\,2}} \qquad \text{Eq. 5-16}$$

where e_{ij} is the element of residual matrix $\boldsymbol{E}$ as in Equation 5-1. If the change of residual (SSR_{new}-SSR_{old}) is smaller than a predefined threshold value (usually 10^{-6}), the ALS optimization stops. It has been observed that ALS can converge very quickly to a minimum after only four rounds of iterations. [82]

5.1.6 A Hybrid EFA-ALS Algorithm. A problem with the iterative MCR approach is that the use of the common constraints like unimodality and non-negativity is not enough to ensure that correct solution can be found if the initial estimates of $\boldsymbol{C}$ are far from being accurate. Frequently there are an infinite number of sets of profiles that all

fulfill these constraints and fit the data equally well. This is referred as the fundamental uncertainty of curve resolution. [83] The band of solutions can be narrowed by employing more constraints, but a unique solution is often difficult to obtain through the use of iterative methods alone. As mentioned earlier, non-iterative EFA gives a unique solution by taking advantage of local rank information. The concentration profiles $\boldsymbol{C}$ estimated by EFA can be used as the initial estimates of iterative ALS for further refinement under additional constraints. Therefore a hybrid algorithm that combines EFA and ALS is proposed in this study for multivariate curve resolution of chemical sensor signals. The flow diagram of this hybrid algorithm is illustrated in Figure 5-5. As far as we know, the use of ALS for refinement of EFA solution has not yet been reported.

All the multivariate curve resolution methods described above have been successfully used with hyphenated chromatographic techniques to improve the interpretation of complex chromatographic data, [11-17] but no one has reported its application to sensor-array based GC system. Towards rapid VOC analysis using sensor-array based GC, it is important to apply self-modeling curve resolution method to peak deconvolution so that the potential of array-based GC can be fully exploited for complex vapor mixture analysis. EFA is especially suitable for peak deconvolution with sensor-array based GC due to the intrinsic evolutionary nature of chromatographic elution.

By contrast to spectroscopic applications where an analyte is routinely measured for light absorbance/transmittance at hundreds of wavelengths, the number of detector channels in a sensor array is much smaller. Only four nanoparticle-coated sensors are used in the chemiresistor array in this study. While this small number of detector channels puts constraint on the information available for curve resolution, it provides a

chance to study the capacity limit of MCR in peak deconvolution. In theory, the maximal number of components in a mixture resolvable numerically is equal to the number of independent sensors in an array detector. In practice, however, the number of sensors required for resolution of a mixture should be higher than the number of components in the mixture, considering the measurement noise in mixture response data. In this study binary mixtures presented in the form of two overlapping peaks are analyzed with a four-sensor array for peak deconvolution by EFA. It is deemed safe to use four variables to resolve binary mixtures. Analysis of ternary mixtures with a four-sensor array will not be attempted, since a three-principal component model built on four variables is most likely over fitted by including measurement errors into the model.

One advantage of the sensor-array based GC analytical system is that column separation reduces complex vapor mixtures into individual vapors or simple vapor mixtures in the form of overlapping peaks in a chromatogram prior to detection. Under normal conditions, it is expected that the analytical challenge faced by the sensor array is mostly binary or ternary vapor mixtures. Our previous study has shown that binary mixtures, and even some ternary mixtures, can be reliably recognized by a sensor array with as few as four or five sensors. [84-86] Quaternary and more complex mixture cannot be analyzed and chromatographic separation is mandated before MCR can be safely employed for peak deconvolution. In the past, we have used extended disjoint principal component regression (EDPCR) model to quantify the ability of a microsensor array to discriminate one mixture from its components and/or sub-component mixtures. [84-87] EDPCR searches through reference pattern library for analyte identification and uses principal component regression for analyte quantification. MCR is a mixture resolution

approach and does not involve pattern recognition. The resolved overlapping components can be delivered to EDPCR for pattern recognition and analyte quantification. In this sense, MCR serves as an important complement to EDPCR for the purpose of VOC identification and quantification.

In this study, the capacity of hybrid EFA-ALS algorithm is explored by using simulated 2-component chromatographic peaks. Its ability to recover response patterns is evaluated as a function of different factors including chromatographic resolution, spectra similarity, noise level, and relative composition. Application of FSMW-EFA to peak purity assessment is also studied.

5.2 Experiments and Simulations

5.2.1 Data Set Description. Experimental data were provided by Q. Zhong using a prototype chromatographic/sensor array system. [88] The response patterns used in the simulations were registered by means of a chemiresistor (CR) array which consists of four sets of interdigital Au/Cr electrodes patterned on a single oxide-coated silicon substrate. Each CR sensor is coated with a different layer of gold–thiolate monolayer protected nanoparticle (MPN). [89] MPNs derived from the following thiols were used in this sensor array: n-octanethiol (C8), 1-mercapto-6-phenoxyhexane (OPH), 7-mercaptoheptanitrile (CCN), methyl 6-mercaptohexanoate (HME), and 4-mercaptodiphenylacetylene (DPA). [90] The structures of the thiolates are illustrated in Fig. 5-6.

In the CR array, the reversible sorption of vapors into each MPN film cause the coating layer to swell, which changes the electron tunneling barrier and thereby the film resistance which is measured as sensor signal [91] . The affinities of each sensor to a given

vapor differ as a result of the structure difference of the MPN ligands, and consequently the array of CRs produces a different set of responses for each vapor. The collective responses can then be analyzed by pattern recognition techniques for the purpose of vapor identification, [92] and the magnitudes of the responses can be used collectively to quantify the vapor concentrations by means of multivariate calibration. [93]

Out of the numerous vapors whose response patterns were registered due to their potential roles as vapor-phase biomarkers or contaminants of environmental tobacco smoke, [88] eight vapors are chosen for data simulations because they can be combined into vapor pairs in binary mixtures that have different spectra similarity. Identities of these vapors, as well as their response patterns are illustrated in Figure 5-7. The similarity between response patterns of any vapor pair is measured by the correlation coefficient ρ between the component response patterns. Thus the higher the value ρ_{12} is, the more similar one pattern is to the other. The pairs of components combined in an overlapping chromatographic profile and the corresponding spectral similarity expressed by means of correlation coefficient are listed in Table 5-1. Seven vapor pairs, each constituting 2-component mixture co-eluting in a chromatographic profile, span a large degree of spectral dissimilarity with ρ_{12} ranging from 0.0234 for methyl isobutyl ketone and ethylbenzene to 0.999 for n-undecane and n-dodecane.

5.2.2 Chromatographic Data Simulations. Simulations of the chromatographic separations are performed in order to determine the influence of the random noise, chromatographic resolution, response pattern similarity, and concentration on the results of curve resolution. The influence of each of these factors on curve resolution is indicated by the quality of response pattern recovered that changes with each factor. Although

many criteria may exist for the quality of a curve resolution task, for qualitative analysis purpose we adopt response pattern recovered by curve resolution as the criterion of resolution quality. Curve resolution is successful only if the resolution algorithm is able to estimate component response patterns accurately. Such accuracy is measured by the correlation coefficient between estimated and measured response patterns. Once the effects of these various factors are known, these factors can be adjusted so that an optimal curve resolution can be achieved. Apparently it would be a daunting job if all these factors needed to be elucidated experimentally. In case the sensor signal drifts and retention time fluctuates, it is left no other option but data simulations.

Simulation of mixture response data $\boldsymbol{X}$ is performed according to Equation 5-1. Sensitivity data $\boldsymbol{S}$ are measured. Concentration data $\boldsymbol{C}$ are simulated by assuming Gaussian elution. It should be noted that curve resolution does not require any information about peak shape. We adopt Gaussian profile only for simplicity purpose. The simulations consist of the combination of sensitivity data of two components using Gaussian elution profiles in a time interval of 100 seconds. Every two-component peak cluster was analyzed on the basis of 100 response vectors evenly spread over the 100-second interval (one response vector per second). The Gaussian function used for simulation of a chromatographic peak is described as follows: [94]

$$G(t) = \frac{A}{\sigma\sqrt{2\pi}} \exp[-(\frac{t - t_R}{\sqrt{2}\sigma})^2] \qquad \text{Eq. 5-17}$$

where G(t) is the value of the Gaussian function at time t, the A is the area of the Gaussian peak, σ is the standard deviation of the Gaussian and t_R is the center of the peak (retention time). By adjusting t_R and σ for adjacent peaks, different degree of peak overlapping can be obtained that is measured by chromatographic resolution (see below).

By changing peak area A, the relative concentration ratio of mixture components can be varied to study the effect of mixture composition on curve resolution, as well as the ability of FSMW-EFA to detect an impurity under an isolated chromatographic peak.

Multiplication of $\boldsymbol{C}$ and $\boldsymbol{S}$ produces the mixture response data $\boldsymbol{X}$. Random noise was added to $\boldsymbol{X}$ to simulate sensor responses in real applications. Gaussian noise (zero mean, varying standard deviation) is assumed. Different signal-to-noise ratios are achieved by changing the standard deviation of Gaussian noise equivalent to a fraction of the maximal sensor response in $\boldsymbol{X}$. Once $\boldsymbol{X}$ is simulated, it is subject to curve resolution using EFA/ALS. The estimated response pattern $\hat{\boldsymbol{S}}$ is then compared to the true pattern $\boldsymbol{S}$ for quantitative assessment of resolution quality. The refinement of resolved response pattern by using EFA/ALS, as well as peak purity assessment by using FSMW-EFA is also presented.

5.2.3 Software and Calculations. The software used for the multivariate analysis is programmed in Matlab 7 (Mathworks, Natick, MA). The built-in singular value decomposition Matlab function is used for extracting principal components from mixture data $\boldsymbol{X}$. Matlab codes for Gaussian peak simulation, implementation of EFA/ALS algorithm, and result presentation and plotting are listed in the Appendix of this chapter. All computations are performed in Matlab with a personal computer.

5.3 Results and Discussion

5.3.1 Improvement of EFA Resolution by Alternating Least Squares. The advantage of the ALS algorithm rests on its great flexibility to bring external constraints into iterative optimization of elution profiles or response patterns. Each cycle of iterative

calculations is accompanied by modification of the results according to the external constraints. The imposition of external constrains to ALS is significant: the outcome is forced towards the desired direction and the optimization procedure is accelerated. As a result, the calculation converges quickly towards the optimal solutions. Improved resolution by adding ALS algorithm into EFA is exemplified by using one representative vapor pair in Table 5-1, styrene and n-dodecane, which has a spectral correlation coefficient ρ_{12}=0.556 and represents a moderate spectra similarity. Chromatogram of overlapping styrene and n-dodecane peaks is shown in Figure 5-8(a). Peak areas are the same, thus the relative concentration ratio of styrene and dodecane is 1:1. The superposition of these two overlapping peaks gives rise to an isolated composite peak that might be regarded as a single elution peak measured at each sensor channel. This composite peak is resolved first by EFA alone and the results are illustrated in Figure 5-8(b) and 5-8(c), as compared to the resolution by EFA-ALS in Figure 5-8(d) and 5-8(e). The contrast between EFA and EFA-ALS is striking. While EFA wrongly estimates the response pattern of n-dodecane, with the HME sensor mistakenly assigned a negative response, EFA-ALS reproduces correctly the response patterns of both components, with the estimated pattern in excellent agreement with their true patterns. The advantage of ALS algorithm is more obvious in the estimation of concentration profiles. The serrated noisy elution profiles estimated by EFA are smoothed by ALS iterations. The agreement between estimated and true concentration profiles offered by EFA-ALS is quantified as *r* values as high as *r*=0.9984 for styrene and *r*=0.9993 for n-dodecane, as compared to the corresponding lower *r* values *r*=0.8194 and *r*=0.864 given by EFA alone. The optimization progress of iterative process is illustrated in Figure 5-9. The ALS algorithm

converges to minimum after 30 iterations. Residual error, indicated by SSR, is reduced from 0.061 to 0.055. The advantage offered by the hybridization of EFA and ALS is not incidental. By using additional knowledge of the mixture and analytical system, external constraints can be readily integrated into the iterative refinement of curve resolution and force the algorithm to produce optimal results. The more knowledge one possesses about the system under investigation, the less ambiguous the analytical results will be.

5.3.2 Peak Purity Assessment by FSMW-EFA. The first step of any factor analysis is the determination of the number of underlying factors in the system. In the case of peak deconvolution, the number of factors is also the number of co-eluting components. While it is generally accepted that a chromatographic peak is pure if the underlying factor is one, otherwise it is a composite peak of two and more co-eluting components, the choice of the suitable factor number still remains the most intractable part of factor analysis. Fewer factors will lose useful information. More factors will results in over-fitting of the data and a useless model for future prediction. Historically many approaches have been proposed to help determine the suitable number of factors to be included in the optimized model, but none can be used safely without careful discretion of specific application. [10] The lack of objective criterion makes the determination process mostly a subjective trial-and-error experiment. However, this difficulty of determining suitable number of factors is largely lessened in the hyphenated chromatographic domain where sequential response data occurs. The evolutionary information contained in the mixture response vectors ordered according to their sequential elution from chromatographic column provides unique local rank information about the number of co-eluting components. EFA-based methods have been proven a

valuable tool to tackle peak purity problems. [59, 62-64] Due to its superior sensitivity to detecting minor components embedded in a major peak, FSMW-EFA is employed here and illustrated for its application to peak purity assessment with sensor-array based GC system. Before its use to detect minor contaminant in a peak, we should be reminded that FSMW-EFA is not a cure for all. Like other factor analysis, the capability of FSMW-EFA in mixture resolution relies on the difference between the components in terms of their elution profiles and spectra. If the embedded peak is collinear to its embedding peak and both have the same retention time, as the case illustrated in Figure 5-10(a), then there is no numerical method that can be used to separate them. Obviously the degree of separation between adjacent peaks, indicated by the chromatographic resolution *R*, is a major player in peak purity assessment by FSMW-EFA (They will be studied in detail later). On the other hand, the merit of FSMW-EFA is manifest in the fact that it is capable of finding minor contaminant in a chromatographic peak as long as it still has some separation from the major component, as the case illustrated in Figure 5-10(b). It is noteworthy that this capacity is also affected by the noise in the mixture responses. If the sensor noise is high and the signal to noise (S/N) ratio is low for the components under study, the minor component will be buried in the baseline noise and it will be difficult for EFA to pick up analytical signal from its noisy background. In this study we will use simulations to explore the ability of FSMW-EFA to detect inner peak in a major chromatographic peak under different noise levels. Again, co-eluting styrene and n-dodecane are used as a representative vapor pair to illustrate the potential of FSMW-EFA in peak purity evaluation, with styrene being the embedding peak and n-dodecane being the embedded peak. To avoid the bias caused by magnitude difference between the

sensitivities of the two components, the sensitivities the CR array measured for styrene and n-dodecane are all normalized to a sum of unity. Proportion of the minor component varies from 1% to 10% by adjusting the peak areas of these two superimposed peaks. Different noise levels, in terms of signal-to-noise ratio at the maximal mixture response, are also applied for each mixture composition.

FSMW-EFA provides a visual display of the number of components contributing to a peak. If a principal component rises above its baseline, it indicates the presence of an analyte inside of the peak. If there are two and more principal components rising above the baselines, we can conclude that the peak is not pure and the peak is a composite peak of more than one co-eluting analytes. When the signal-to-noise ratio is 100:1 (as mentioned earlier, it means that standard deviation of Gaussian noise added is 1% of the maximal mixture response in $\boldsymbol{X}$), minor component n-dodecane of different proportions is added into styrene and co-elute with styrene with a retention time difference as small as 0.1 (it corresponds to a chromatographic resolution R=0.033. Definition of R is provided later). As shown in Figure 5-11, inner peak of n-dodecane with a proportion of 1% cannot be detected by FSMW-EFA at S/N ratio=100:1, until its proportion increases to 5% and beyond. Inner peak of n-dodecane with a proportion of 1% can be easily detected by FSMW-EFA when the S/N ratio increases to 1000:1. It is obvious that noise level has a remarkable influence on the ability of FSMW-EFA to detect impurity. At high noise level, a small amount of impurity might not be detectable. On the other hand, minor impurity can be detected by FSMW-EFA when the sensitivity is high and/or the sensor noise is low, as long as the two peaks are not completely overlapping (i.e., the retention time is different).

It should be noted that, while it is possible to judge whether there is an inner peak or not in a chromatographic peak, it does not mean the inner peak can be resolved by numerical resolution. In the cases of embedded peaks, most frequently it is required that chromatographic conditions should be adjusted so that the inner peak can be separated as far as possible from its embedding peak. Only if selective regions exist for both co-eluting components can it be possible some curve resolution method be applied unambiguously.

5.3.3 Influence of Chromatographic Resolution. The capacity of factor analysis in extracting meaningful factors from mixture data relies on the difference between the factors in terms of their elution profiles and response pattern. If the elution profiles or the spectra are the same for these factors, there is no way we can separate the overlapping components from each other. Thus, the ability of EFA-ALS is compromised by the similarity of elution profiles of mixture components, as well as the spectra similarity. In this section we will discuss the collinearity issue in $\boldsymbol{C}$ and $\boldsymbol{S}$ and their influence on the quality of curve resolution results. As usual, the quality of resolution results is measured by the correlation coefficient *r* between measured and estimated response patterns.

Similarity between elution profiles of two co-eluting analytes is directly determined by the degree of separation between these two overlapping peaks that is measured by a common metric, chromatographic resolution *R*, defined as below:

$$R = \frac{t_{R2} - t_{R1}}{2(\sigma_1 + \sigma_2)} \qquad \text{Eq. 5-18}$$

where t_{R1} and t_{R2} are respectively the retention times of the two adjacent peaks, and σ_1 and σ_2 are standard deviations of their Gaussian peaks. Apparently different degree of peak overlapping, as exemplified in Figure 5-12, can be obtained by adjusting these parameters.

By keeping the noise level to S/N=100:1 and relative concentration ratio to 1:1, the effects of elution profile similarity are studied by plotting *r* as a function of chromatographic resolution *R* for all of the seven pairs of vapors listed in Table 5-1. The results are plotted in Figure 5-13. As expected, the more separated the co-eluting peaks are from each other, the more accurately the response pattern can be reproduced. This trend becomes more obvious for the vapor pairs whose spectra are similar. In other word, when the chromatographic resolution remains the same, the vapor pair with dissimilar spectra will be more accurately resolved from each other. If an *r* value at least of 0.9 is deemed equivalent to acceptable curve resolution, a chromatographic resolution R>0.8 must be achieved for vapor pair with similar response patterns indicated by $\rho_{12}>0.7$ for the vapor pairs n-undecane and n-dodecane ($\rho_{12}=0.999$), 1,2,4-trimethylbenzene and 1,4-dichlorobenzene ($\rho_{12}=0.842$), and vapor pair m-xylene and 1,4-dichlorobenzene ($\rho_{12}=0.760$). For vapor pairs with dissimilar patterns ($\rho_{12}<0.7$), a chromatographic resolution R>0.4 must be achieved. The trends observed in Figure 5-13 are expected because the collinearity in either C or S equally brings about the deterioration of factor analysis. The implication of this finding to mixture analysis by array based GC system is obvious. No matter if the co-eluting vapors have dissimilar response patterns or not, every effort should be made to optimize the chromatographic conditions and to separate

their peak as much as possible. No curve resolution can be achieved for completely overlapped peaks even using the most advanced MCR algorithm.

5.3.4 Influence of noise. We have observed that the noise level may compromise the resolution results by blurring the demarcation between systematic and random contributions to mixture responses. The effects of noise in mixture response data are explored by studying the quality of resolution results as a function of noise levels varying from S/N=100:1 to 100:10. As mentioned earlier, Gaussian noise is added to mixture response with standard deviation equal to a fraction of the maximal mixture response. Gaussian noises of different levels are applied to each of the seven vapors pairs. To make the analysis consistent, the relative concentration ratio of the vapor pair is kept to 1:1 and chromatographic resolution is set to a moderate value of 0.5. The results are presented in Figure 5-14. For all vapor pairs, the quality of resolved pattern decreases monotonically with increasing noise. The vapor pair with similar spectra is mostly affected. In other words, resolution of mixture with similar pattern is more subject to noise effect. Similar to the finding in Figure 5-13, the more dissimilar spectra the vapor pair has, the more accurate the resolution will be. If an *r* value of 0.9 is designated as the threshold of acceptable and unacceptable curve resolution, an S/N ratio value of 100:5 must be maintained for the mixture under investigation that have dissimilar response patterns ($\rho_{12}<0.7$). By contrast, for vapor pair n-undecane and n-dodecane whose ρ_{12} is as high as 0.999, the best resolution achievable is still well below $r<0.8$ when it occurs at S/N value 100:1. When the noise rises to S/N=100:10, this vapor pair can no longer be resolvable.

5.3.5 Quantitative Mixture Compositions. It is conceivable that the result of multivariate curve resolution is heavily influenced by the relative concentration ratio of mixture components. If one of the components constitutes the majority of a mixture, the mixture will be dominated by that component and the mixture behaves more like a single analyte. Consequently the dominating component in the mixture will dictate the curve resolution results. To confirm the influence of quantitative composition on curve resolution, a series of simulations are performed where the peak area of the first component in Table 5-1 is set to 100, while the peak area of the second component is changed so that its proportion in the mixture varies from 0.01 to 0.99. To avoid confounding factors, the chromatographic resolution *R* is set to 0.5 and the noise level is set to S/N=100:1 for each simulation. Three vapor pairs with different spectra similarities are selected for the study. The results in Figure 5-15 show the quality of resolved response patterns as a function of the proportion of the second component in the mixture. For the dissimilar vapor pair in Fig. 5-15 (ρ_{12}=0.023), it will be relatively easier for MCR to resolve its components. When the proportion of the second component increases from 1% to 99%, the resolved pattern of the first component maintains at high quality until its proportion is overwhelmed by the second component. Similarly, the quality of resolved pattern of the second component does not improve until its proportion rises to 20%. The extent of the influence of mixture composition changes with vapor pairs, with the vapor pair being more affected if its components have more similar patterns. With increasing spectra similarity, the fraction threshold beyond which *r* is close 1 can be achieved also increases. When ρ_{12} increases from 0.556 to 0.841 the fraction threshold increases from

about 0.5 to 0.9. This can be attributed to the difficulty in resolving mixtures whose components are collinear in terms of their spectra similarity.

The results of curve resolution is dictated by whichever component that have a higher fraction in the mixture. Since the resolved pattern is the estimate of the pattern of the dominating component, the *r* value at the extreme fractions is actually the correlation coefficient between the response patterns of the corresponding vapor pair, which can be indicated by the r values at which the curves in Fig. 5-15 cross the ordinates when the fraction values approach 0 or 1.

5.4 Conclusion

In this study a hybrid multivariate curve resolution (MCR) algorithm is described and applied to a sensor-array based GC detection system for deconvolution of simulated chromatographic peaks that overlapped together as a result of co-elution. By comparison to traditional non-iterative evolving factor analysis (EFA), the advantage of the hybrid EFA/ALS algorithm is exemplified by resolution of co-eluting vapor pair styrene and n-dodecane. As a result of iterative refinement of EFA solution using ALS, the sum of squares of residuals is reduced from 0.061 to 0.055, a 10% decrease.

Fixed-side moving window EFA (FSMW-EFA) is also described for its application to peak purity assessment in a special situation of peak overlapping when a peak is contaminated by minor component in the form of embedded inner peak. Minority contaminant n-dodecane with a proportion of 1% in styrene can be easily detected by FSMW-EFA when the S/N ratio is as high as 1000:1.

Since the ability of MCR in resolving vapor mixture relies on the difference of the mixture components in terms of their elution profiles and response patterns, the quality of resolved response patterns is evaluated as a function of chromatographic resolution R, response pattern similarity ρ, signal-to-noise ratio S/N, and quantitative composition. For the vapor pairs with dissimilar patterns (ρ is close to 0), the quality of resolved pattern deteriorates at $R<0.4$; for vapor pairs with similar patterns (ρ is close to 1), the quality of resolved pattern deteriorates at $R<0.8$. Poor curve resolution due to pattern similarity is compensated by good chromatographic separation. At S/N=100:1, satisfactory curve resolution ($r>0.95$) requires pattern similarity $\rho<0.7$. The individual response patterns derived from EFA-ALS are dominated by the component whichever has a higher concentration.

The self-modeling curve resolution algorithms developed in this study aims at providing fundamental methodology for mixture analysis by sensor-array based GC analytical system. The hyphenation of sensor array detector with chromatography makes it possible to resolve co-eluting components without hardware separation. The ability of numerical resolution of overlapped peaks becomes more important when rapid separation and fast analysis is valued. Hyphenated analytical techniques commonly used in modern laboratory and field monitoring provide additional dimensions of analytical information that can be used to gain insight into the chemical composition of complex mixture. Factor analysis is a unique chemometric technique that be used with hyphenated analytical instrumentation for mixture analysis. Its capacity in illuminating the composition of complex mixture is only restricted by the independent detector channels that are available in the analytical instrument. Theoretically the maximal order of a mixture (i.e., number of

its constituents) that can be resolved numerically is equivalent to the number of independent detector channels. Unfortunately collinearity among the detector channels is a very common phenomenon and in practice the number of independent variables required for successful mixture resolution is frequently tens of times of the order of the mixture. In instrument hyphenated with a spectrometric detector, hundreds of wavelengths are routinely measured even for the simplest sample. In the sensor-array based GC system that employs a chemical sensor array detector, sensor collinearity can be reduced to minimum by judicious choosing of sensing materials, although the choice is often limited. On the other hand, due to the use of upstream chromatographic column separation, it is possible to reduce complex mixture into individual components or simpler mixtures in case of co-elution. With the order of mixture reduced to a smaller number, a large sensor array can thus become unnecessary. However, it should be stressed that in any case the number of sensors to be used in a sensor array detector should accommodate the complexity of mixture under investigation. A rule of thumb is that the number of sensors should be at least two or three times of the order of mixture.

As compared to other hyphenated techniques, the most noteworthy feature of hyphenated chromatographic technique is that the spectra or response vectors are sequentially recorded by the multivariate detector in accordance with the progressive elution in the column. The evolutionary elution of spectra contains additional information about the elution window of co-eluting components. This unique feature can be explored by local rank analysis of sub-matrices of mixture response data using evolving factor analysis. Unlike global rank analysis of the whole mixture data, local rank analysis is able to provide directly the number of independent components in the mixture, as well as the

start and end point of each component in the chromatogram. Thus the elution window can be uniquely defined by EFA and the selective regions can be directly used to estimate the pure spectra. However the solution of EFA still has room for improvement considering the fluctuation of baseline level used to define the emergence of a new component, the uncertainty in the definition of elution windows, and the rank deficiencies in least square calculation of spectra. By including external constraints about the non-negativity of concentration profiles and spectra, unimodality of chromatographic peak, zero concentration outside of elution window, alternating least square (ALS) can be subsequently used to refine the EFA resolution towards an optimized result. The flexibility of ALS in utilizing various constraints makes it a popular algorithm in self-modeling curve resolution.

We have shown that the hybridization of EFA and ALS provide a powerful tool for deconvolution of overlapped peaks measured with sensor-array based GC system. This serves as another example of how chemometrics can help enhance performance of an analytical instrument. The ability of numerical resolution of overlapped peaks is essential especially when fast separation of complex mixture is emphasized for the field implementation of sensor-array based GC system. However, such numerical ability does not spare the requirement for high column efficiency and powerful sensor array detector. EFA-ALS can only work fast towards a more accurate resolution with the least ambiguity if the mixture is already reduced into the simplest binary or ternary forms and the sensor array detector is capable of generating specific response pattern for each component. By incorporating advanced chemometric algorithm into the high-standard instrumental

hardware, an integrated analytical system that combines analytical power with intelligent user interface can be developed for optimal sample analysis and data interpretation.

As a part of the effort performed in this study, the application of Fixed-Sized Moving Window Evolving Factor Analysis has been exemplified to the peak purity assessment with sensor-array based GC system. The sensitivity of FSWM-EFA to minor contaminant embedded in a chromatographic makes it a special tool for peak purity evaluation that should be employed with hyphenated chromatographic techniques. FSMW-EFA scans through the evolution of eigenvalues of the mixture data and find clues about the systematic structure under the peak. Obviously any other factor that cases fake systematic changes in eigenvalues may be mistakenly interpreted as a contaminant to the peak under study. These compromising factors may include non-linearity of sensor response or heteroscedastic noise. Therefore it is suggested that some exploratory analysis like principal component analysis of the whole data set should be performed prior to peak purity assessment. Abnormality of the data set is usually explicitly indicated by the minor principal components showing some regular patterns whereas they should represent random noise only. FSMW-EFA has not been fully explored here. The limited number of sensors used in this study makes it a tentative and worthwhile trial.

Acknowledgement

Dr. Qiongyan Zhong provided CR sensor calibration data used in the simulations. This study is funded by National Science Foundation through WIMS Engineering Research Center in the University of Michigan under award number ERC-9986866.

Appendix

Main Matlab codes used in this study are listed below. Annotations are added after "%"where it is necessary in the codes.

Gaussian Peak Simulations

```
array={'c8','OPH','HME','CCN'};
A_sim(1,:)= 0.0436 0.0628 0.0427 0.0490 .;    % component spectra
A_sim(2,:)= 0.1950 0.1422 0.0957 0.0904 .;

t=(40:0.1:60)';
rt= 50,52 .;
va= 1,1 .;
C_sim(:,1)=Gaussian(t,100,rt(1),va(1));        % elution profiles
C_sim(:,2)=Gaussian(t,100,rt(2),va(2));

Y=C_sim*A_sim;
randn('seed',0);
noise=max(max(Y))*0.1;
Y=Y+noise*randn(size(Y));

 C,A .=ALS_EFA(2,Y);

 C_est_n,A_est_n .=norm_max(C,A);       % normalisation of calc. C and A
 C_sim_n,A_sim_n .=norm_max(C_sim,A_sim); % norm. of true C and A
A_est_n=A_est_n./repmat(max(A_est_n')',1,nl);
A_sim_n=A_sim_n./repmat(max(A_sim_n')',1,nl);
%subplot(4,1,1)
%subplot(3,1,1)
   plot(t,C_sim(:,1),'*',t,C_sim(:,2),'+',t,sum(C_sim')');
   title('True concentration profiles');
   %ylabel('Concentration');xlabel('Time');
%subplot(4,1,2);
figure;
  %plot(log10(SSR/50000));
  %xlabel('Iteration');ylabel('Log(SSR)');
 % title('Convergence of ALS iterations');
  %xlim( 0 100 .);
  %axis( 0 100 -3 0 .);   xlabel('Time');ylabel('Response');
%subplot(4,1,3)
%subplot(3,1,2)
  %bar( A_est_n,A_sim_n .);
  %xlabel('Sensor');ylabel('Relative response');
 % title('Estimated vs.True response pattern');
%subplot(4,1,4)
%subplot(3,1,3)
 % plot(t,C_est_n,'-',t,C_sim_n,'.');
  %xlabel('Time');ylabel('Relative concentration');
  %title('Esitimated vs.True concentration profiles (normalized to maximum)');
```

```
function z=gauss(n,mu,sigma,options)
% Compute Gaussian profiles
% Input:
%       n (1x1) number of points in z
%       mu (objectsx1) first moments of gaussian distribution
%       sigma (ogjectsx1) second moments of gaussian distribution
%       options (1x1) triggers N(0,s) noise addition
% output:
%       z(objectsxn) data table with gaussian curves

if (nargin<3)
   help gauss
   return
elseif (nargin==3)
   options=0;
end

nz=length(mu);
if nz~=length(sigma)
   error('error:Number of entries doesnt match')
end

x=1:n;
for a=1:nz
   z(a,:)=exp(-0.5*((x-mu(a))/sigma(a)).^2)./(sigma(a)*(2*pi)^0.5);
end
if options(1)
   nmax=max(z(:))*options(1);
   z=z+randn(size(z))*nmax;
end%function t,array,Y,C,A.=bi_mixture(a2)

% simulated binary mixture response matrix Y
% t is the retention time range
% array is a cell containing name of sensor elements
% C is the concentration profile matrix (true, simulated Gaussian peaks)
% A is sensor sensitivity vector measured for 4 CR sensors
a2= 1.0101 2.0408 5.2632 11.111 42.857 100 233.33 400 900 1900
    9900 .;
array={'c8','OPH','HME','CCN'};
A0(1,:)= 0.2930 0.2456 0.2845 0.1879 .;    % component spectra
A0(2,:)= 15.4000 2.8906 1.2050 1.4287 .;

t=(1:0.1:20)';
rt= 10,12 .;
va= 1,1 .;
C0(:,1)=Gaussian(t,100,rt(1),va(1));        % elution profiles
C0(:,2)=Gaussian(t,a2(11),rt(2),va(2));

%Y=C0*A0;
%randn('seed',0);
%noise=max(max(Y))*0.05;
%YY=Y+noise*randn(size(Y));
%plot(C);
r_a2=r_a2(1,2);
r_A= r_a1,r_a2 .;
```

```
function [EFA_f,EFA_b]=EFA(Y,ne)

% changed from Marcel Maeder & Yorck-Michael

[ns,nl]=size(Y);
EFA_f=NaN(ns,ne);
EFA_b=NaN(ns,ne);

for i=1:ns
  s_f=svd(Y(1:i,:));                                    % forward SV
  s_b=svd(Y(ns-i+1:ns,:));                              % backward SV
  EFA_f(i,1:min(i,ne))=s_f(1:min(i,ne))';
  EFA_b(ns-i+1,1:min(i,ne))=s_b(1:min(i,ne))';
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%function [C,A]=Non_it_EFA(nc,Y)
array={'C8','OPH','HME','CCN'};
A_sim(1,:)=[0.289812067 0.242927794 0.28140455 0.185855589];     % component spectra
A_sim(2,:)=[0.735986389 0.138145601 0.057588545 0.068279465];
t=(40:0.1:60)';
rt=[50,52];
va=[1,1];
C_sim(:,1)=Gaussian(t,100,rt(1),va(1));        % elution profiles
C_sim(:,2)=Gaussian(t,100,rt(2),va(2));

Y=C_sim*A_sim;
randn('seed',0);
noise=max(max(Y))*0.05;
Y=Y+noise*randn(size(Y));

[ns,nc]=size(C_sim);
ne=nc;

% [EFA_f,EFA_b]=EFA(Y,ne+1);      % perform EFA
%ne=ne+1;
[ns,nl]=size(Y);
EFA_f=NaN(ns,ne+1);
EFA_b=NaN(ns,ne+1);

for i=1:ns
  s_f=svd(Y(1:i,:));                                    % forward SV
  s_b=svd(Y(ns-i+1:ns,:));                              % backward SV
  EFA_f(i,1:min(i,ne+1))=s_f(1:min(i,ne+1))';
  EFA_b(ns-i+1,1:min(i,ne+1))=s_b(1:min(i,ne+1))';
end
%sig_level=0.01;               % define cut-off level
sig_level=EFA_f(ns,ne+1); % baseline level                        % build window matrix
C_window=EFA_f(:,1:nc)>sig_level & fliplr(EFA_b(:,1:nc))>sig_level;
[U,S,Vt]=svd(Y,0);
U_bar=U(:,1:ne);

T=ones(nc,nc);
for i=1:nc
   U_i_0=U_bar(~C_window(:,i),:);

%A=C\Y;
```

```
s_b=svd(Y(ns-i+1:ns,:));                                % backward SV
  EFA_f(i,1:min(i,ne+1))=s_f(1:min(i,ne+1))';
  EFA_b(ns-i+1,1:min(i,ne+1))=s_b(1:min(i,ne+1))';
end
%sig_level=0.01;              % define cut-off level
sig_level=EFA_f(ns,ne+1); % baseline level                    % build window matrix
C_window=EFA_f(:,1:nc)>sig_level & fliplr(EFA_b(:,1:nc))>sig_level;
[U,S,Vt]=svd(Y,0);
U_bar=U(:,1:ne);

T=ones(nc,nc);
for i=1:nc
   U_i_0=U_bar(~C_window(:,i),:);
   T(2:nc,i)=-U_i_0(:,2:nc)\U_i_0(:,1);
end
C=U_bar*T;

neg=-min(C)>max(C);          % sign of conc profiles
C=C*diag((-1).^neg);         % reverse if negative
%A=C\Y;
A=pinv(C)*Y;

[C_est_n,A_est_n]=norm_max(C,A);       % normalisation of calc. C and A
[C_sim_n,A_sim_n]=norm_max(C_sim,A_sim); % norm. of true C and A
A_est_n=A_est_n./repmat(max(A_est_n')',1,nl);
A_sim_n=A_sim_n./repmat(max(A_sim_n')',1,nl);
subplot(3,1,1)
   plot(t,C_sim);
   title('True concentration profiles');
   xlabel('Time');ylabel('Concentration');
subplot(3,1,2)
  bar([A_sim_n,A_est_n]);
  xlabel('Sensor');ylabel('Relative response');
  title('Estimated vs.True response pattern');
subplot(3,1,3)
  plot(t,C_est_n,'-',t,C_sim_n,'.');
  xlabel('Time');ylabel('Relative concentration');
  title('Estimated vs.True concentration profiles (normalized to maximum)');

  array={'c8','OPH','HME','CCN'};
A_sim(1,:)=[0.289812067 0.242927794 0.28140455 0.185855589];     % component spectra
STYRENE
A_sim(2,:)=[0.735986389 0.138145601 0.057588545 0.068279465]; %DODECANE
t=(40:0.1:60)';
rt=[50,52];
va=[1,1];
C_sim(:,1)=Gaussian(t,100,rt(1),va(1));        % elution profiles
C_sim(:,2)=Gaussian(t,100,rt(2),va(2));

Y=C_sim*A_sim;
randn('seed',0);
noise=max(max(Y))*0.01;
Y=Y+noise*randn(size(Y));
% [t,lam,Y,C_sim,A_sim]=Data_Chrom2a;

% [EFA_f,EFA_b]=EFA(Y,ne+1);      % perform EFA
```

```
array={'c8','OPH','HME','CCN'};
 t=(0:0.1:20)';
 rt= [10,12];
 va= [1,1];
 C_sim(:,1)=Gaussian(t,100,rt(1),va(1));        % elution profiles
 C_sim(:,2)=Gaussian(t,100,rt(2),va(2));


for i=1:size(CR,1)/2
   A_sim1=CR((i-1)*2+1,:);     % component spectra
   A_sim2=CR(i*2,:);
   A_sim= [A_sim1;A_sim2];
   Y=C_sim*A_sim;
   noise=0.01:0.01:0.1;
   for ii=1:size(noise,2)
      % ii=ii+1;
      randn('seed',0);
       noised=max(max(Y))*noise(ii);
       Y_ed=Y+noised*randn(size(Y));
        [C,A]=Non_it_EFA(2,Y_ed);
        [C_est_n,A_est_n]=norm_max(C,A);        % normalisation of calc. C and A
        [C_sim_n,A_sim_n]=norm_max(C_sim,A_sim); % norm. of true C and A
       %A_est_n=A_est_n./repmat(max(A_est_n')',1,nl);
       %A_sim_n=A_sim_n./repmat(max(A_sim_n')',1,nl);
       c1_r=corrcoef(C_est_n(:,1),C_sim_n(:,1));
       c2_r=corrcoef(C_est_n(:,2),C_sim_n(:,2));
       a1_r=corrcoef(A_est_n(1,:)',A_sim_n(1,:)');
       a2_r=corrcoef(A_est_n(2,:)',A_sim_n(2,:)');
       r_C(i,ii,:)= [c1_r(1,2) c2_r(1,2)];
       r_A(i,ii,:)= [a1_r(1,2) a2_r(1,2)];
        [t1_C,index1_C]=max(C(:,1));
        [t2_C,index2_C]=max(C(:,2));
       t1_delta(i,ii,:)= [rt(1) index1_C];
       t2_delta(i,ii,:)= [rt(2) index2_C];
   end
end
array={'c8','OPH','HME','CCN'};
 t=(0:0.1:20)';
 %rt= [10,12];
 rt1=10;
 rt2= [10.4 10.8 11.2 11.6 12 12.4 12.8 13.2 13.6 14];
 va= [1,1];

 %plot(t,C_sim(:,1)',t,C_sim(:,2)',t,sum(C_sim'))
 %xlabel('Time');
% ylabel('Sensor response');
 %ylim( [0 80]);
 %Title('R=1')

   for i=1:size(CR,1)/2
   A_sim1=CR((i-1)*2+1,:);     % component spectra
   A_sim2=CR(i*2,:);
   A_sim= [A_sim1;A_sim2];

   %noise=0.01:0.01:0.1;
```

```
    for i=1:size(CR,1)/2
    A_sim1=CR((i-1)*2+1,:);     % component spectra
    A_sim2=CR(i*2,:);
    A_sim= A_sim1;A_sim2 .;
   %areas= 1.0101 3.0928 5.2632 7.5269 9.8901 12.36 14.943 17.647 20.482 23.457
           26.582 29.87 33.333 36.986 40.845 44.928 49.254 53.846 58.73 63.934
           69.492 75.439 85.185 92.308 100 108.33 112.77 122.22 127.27 138.1
           143.9 156.41 163.16 177.78 185.71 203.03 212.5 233.33 244.83 270.37
           284.62 316.67 334.78 376.19 400 455.56 488.24 566.67 614.29 733.33
           809.09 1011.1 1150 1566.7 1900 3233.3 4900 9900 .,
   areas= 1.0101 2.0408 5.2632 11.111 25 42.857 66.667 100 150 233.33 400
           900 1900 9900 .;
    %noise=0.01:0.01:0.1;
    %for ii=1:size(noise,2)
      % ii=ii+1;
     for j=1:size(areas,2)
       C_sim(:,1)=Gaussian(t,100,rt(1),va(1));         % elution profiles
       C_sim(:,2)=Gaussian(t,areas(j),rt(2),va(2));
       Y=C_sim*A_sim;
      noise=0.01;
      randn('seed',0);
       noised=max(max(Y))*noise%(ii);
       Y_ed=Y+noised*randn(size(Y));
        C,A .=Non_it_EFA(2,Y_ed);
        C_est_n,A_est_n .=norm_max(C,A);       % normalisation of calc. C and A
        C_sim_n,A_sim_n .=norm_max(C_sim,A_sim); % norm. of true C and A
       %A_est_n=A_est_n./repmat(max(A_est_n')',1,nl);
       %A_sim_n=A_sim_n./repmat(max(A_sim_n')',1,nl);
       c1_r=corrcoef(C_est_n(:,1),C_sim_n(:,1));
       c2_r=corrcoef(C_est_n(:,2),C_sim_n(:,2));
       a1_r=corrcoef(A_est_n(1,:)',A_sim_n(1,:)');
       a2_r=corrcoef(A_est_n(2,:)',A_sim_n(2,:)');
       r_C(i,j,:)= c1_r(1,2) c2_r(1,2) .;
       r_A(i,j,:)= a1_r(1,2) a2_r(1,2) .;
        t1_C,index1_C .=max(C_est_n(:,1));
        tt1_C,index11_C .=max(C_sim_n(:,1));
        t2_C,index2_C .=max(C_est_n(:,2));
        tt2_C,index22_C .=max(C_sim_n(:,2));
       t1_delta(i,j,:)= index11_C index1_C .;
       t2_delta(i,j,:)= index22_C index2_C .;
    end
end
array={'c8','OPH','HME','CCN'};
 t=(0:0.1:20)';
 rt= 10,12 .;
 %rt1=10;
 %rt2= 10.4 10.8 11.2 11.6 12 12.4 12.8 13.2 13.6 14 .;
 va= 1,1 .;
  %plot(t,C_sim(:,1)',t,C_sim(:,2)',t,sum(C_sim'))
 %xlabel('Time');
% ylabel('Sensor response');
 %ylim( 0 80 .);
 %Title('R=1')
```

```
t=(40:0.1:60)';
rt=[50,52].;
va=[1,1].;
C_sim(:,1)=Gaussian(t,100,rt(1),va(1));        % elution profiles
C_sim(:,2)=Gaussian(t,100,rt(2),va(2));

Y=C_sim*A_sim;
randn('seed',0);
noise=max(max(Y))*0.01;
Y=Y+noise*randn(size(Y));

%[t,lam,Y,C_sim,A_sim].=Data_Chrom2a;
[ns,nc].=size(C_sim);
ne=nc;

%[EFA_f,EFA_b].=EFA(Y,ne+1);      % perform EFA
%ne=ne+1;
[ns,nl].=size(Y);
EFA_f=NaN(ns,ne+1);
EFA_b=NaN(ns,ne+1);

for i=1:ns
   s_f=svd(Y(1:i,:));                                          % forward SV
   s_b=svd(Y(ns-i+1:ns,:));                                    % backward SV
   EFA_f(i,1:min(i,ne+1))=s_f(1:min(i,ne+1))';
   EFA_b(ns-i+1,1:min(i,ne+1))=s_b(1:min(i,ne+1))';
end
%sig_level=0.01;                % define cut-off level
sig_level=EFA_f(ns,ne+1); % baseline level                          % build window matrix
C_window=EFA_f(:,1:nc)>sig_level & fliplr(EFA_b(:,1:nc))>sig_level;
[U,S,Vt].=svd(Y,0);
U_bar=U(:,1:ne);

T=ones(nc,nc);
for i=1:nc
    U_i_0=U_bar(~C_window(:,i),:);
    T(2:nc,i)=-U_i_0(:,2:nc)\U_i_0(:,1);
end
C=U_bar*T;

neg=-min(C)>max(C);            % sign of conc profiles
C=C*diag((-1).^neg);           % reverse if negative
%A=C\Y;
for it=1:100 % ALS interation
    %C=norm_max(C);
     [C,A].=constraints_nonneg_unimod(Y,C);
     [C,A].=constraints_nonneg(Y,C);
     %C=C.*C_window;
     R=Y-C*A;
    SSR(it)=sum(sum(R.*R));
end
[C_est_n,A_est_n].=norm_max(C,A);        % normalisation of calc. C and A
[C_sim_n,A_sim_n].=norm_max(C_sim,A_sim); % norm. of true C and A
A_est_n=A_est_n./repmat(max(A_est_n')',1,nl);
A_sim_n=A_sim_n./repmat(max(A_sim_n')',1,nl);
```

```
  plot(log10(SSR/100));
   xlabel('Iteration');ylabel('Log(SSR)');
   title('Convergence of ALS iterations');
   %xlim( 0 100 .);
  %axis( 0 100 -1 0 .);
   %xlabel('iteration');ylabel('Response');
subplot(4,1,3)
%subplot(3,1,2)
   bar( A_est_n,A_sim_n .);
   xlabel('Sensor');ylabel('Relative response');
   title('Estimated vs.True response pattern');
subplot(4,1,4)
%subplot(3,1,3)
   plot(t,C_est_n,'-',t,C_sim_n,'.');
   xlabel('Time');ylabel('Relative conc.');
   title('Esitimated vs.True concentration profiles (normalized to maximum)');
```

FSMW-EFA and plotting

```
clear
A_sim(1,:)= 0.28981 0.24293 0.2814 0.18586 .;     % component spectra STY
%A_sim(2,:)= 0.22009  0.31701 0.21555 0.24735 .;
A_sim(2,:)= 0.73599 0.13815 0.057589 0.068279 .;%DOD
%0.36391 0.2371          0.23056           0.16843
%0.28963 0.28998 0.22312 0.19726
t=(0:0.1:20)';
rt= 10,10.5 .;
va= 1,0.5 .;
%areas= 1.0101 2.0408 5.2632 11.111 25 42.857 66.667 100 150 233.33 400 900 1900 9900 .;
%noise_level=0.01:0.01:0.01;
noise_level=0.001;
%for ii=1:size(areas,1)

    C_sim(:,1)=Gaussian(t,100,rt(1),va(1));         % elution profiles
    C_sim(:,2)=Gaussian(t,10,rt(2),va(2));
    %C_sim(:,2)=Gaussian(t,5.2632,rt(2),va(2));
    Y=C_sim*A_sim;
    %randn('seed',0);
    noise=max(max(Y))*noise_level;
    YY=Y+noise*randn(size(Y));
  ns,nc .=size(YY);

size_w=7;                       % small windows, commonly used
%EFA_w=zeros(ns-size_w+1,size_w);
for i=1:ns-size_w+1
   s_w=svd(YY(i:i+size_w-1,:));
   EFA_w(i,:)=s_w';
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

Figures and Tables

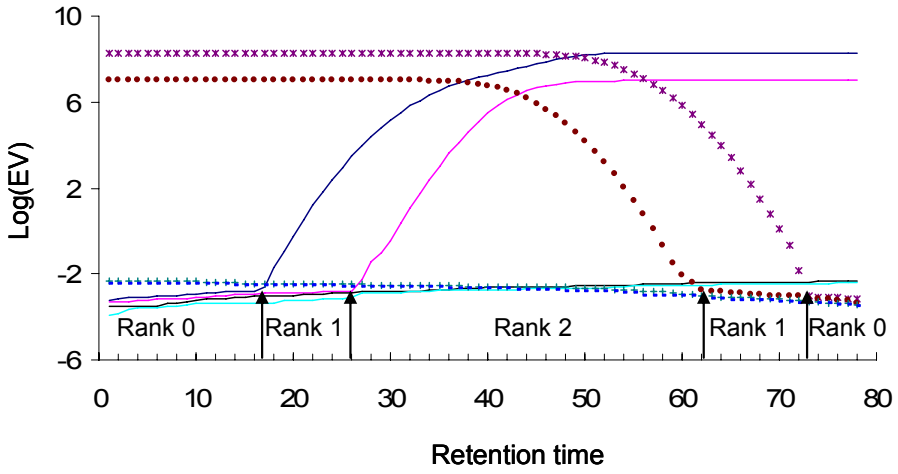

Figure 5-1. Rank map for a two-component mixture. The rise of EVs above baseline indicates the start (in the forward EFA plot, solid curves) or end (in the backward EFA plot, dotted curves) of elution of a component. Concentration window for the first component is defined by the rise of the first eigenvalues (around retention time 16) in the forward EFA plot and rise of the second engenvalue in the backward EFA plot (around retention time 62). Concentration window for the second component is defined by the rise of the second eigenvalues in the forward EFA plot (around retention time 26) and the rise of the first eignevalue in the backward EFA plot (around retention time 73).

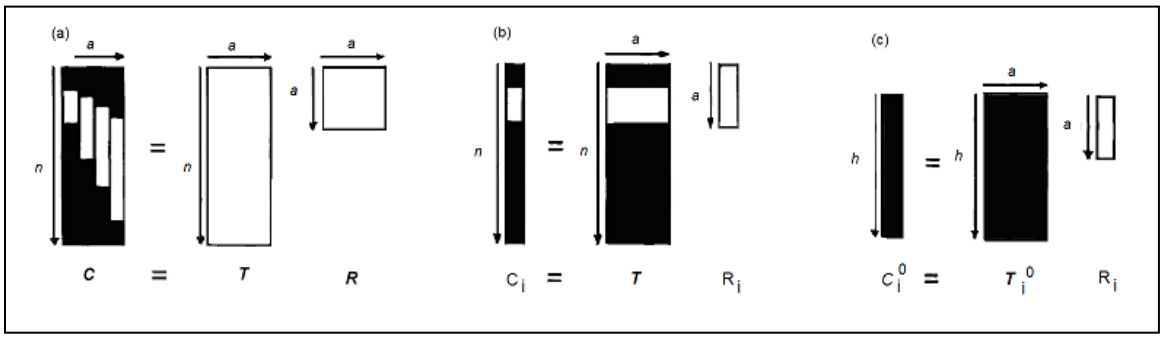

Figure 5-2. Schematic illustration of how the rotation matrix R can be found (Equation 5-12). The matrix of column factors C (with the white areas indicating the concentration windows) can be viewed as the product of T and R (see (a)); the shaded areas in C_i are a linear combination of the corresponding areas in T (see (b)); (c) represents how a zero vector C_0 is extracted from C_i and how T_0 is extracted from T. (Ref. 32)

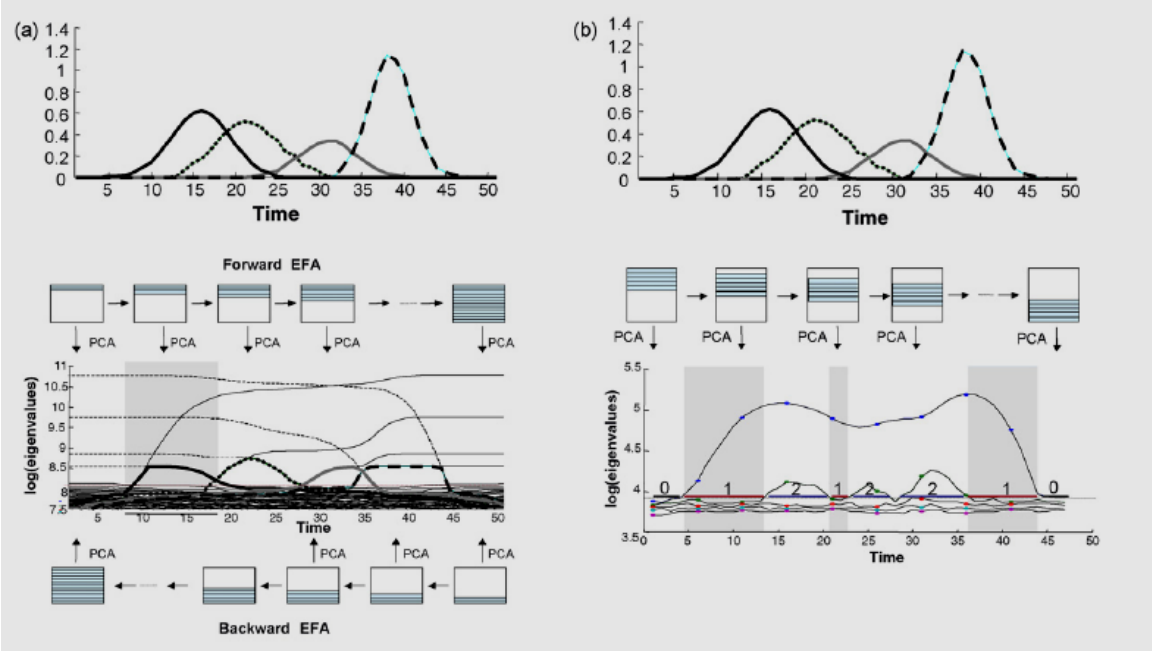

Figure 5-3. (a) Top plot: concentration profiles of an HPLC–DAD data set. Bottom plot: Information derived from the data set in top plot by evolving factor analysis (EFA) scheme of PCA analyses performed. Combined forward EFA (solid black lines) and backward EFA (dashed black lines) plot shows elution regions. The thick lines overlaid are the derived elution profiles. The shaded zone marks the elution window for the first eluting compound. (b) Top plot: as in (a). Bottom plot: information derived from the data set in top plot by fixed size moving window-evolving factor analysis (FSMW-EFA) scheme of the PCA analyses performed. The straight lines and associated numbers mark the different windows along the data set as a function of their local rank (number). The shaded zones mark the elective windows in the data set (rank 1). (Ref. 61)

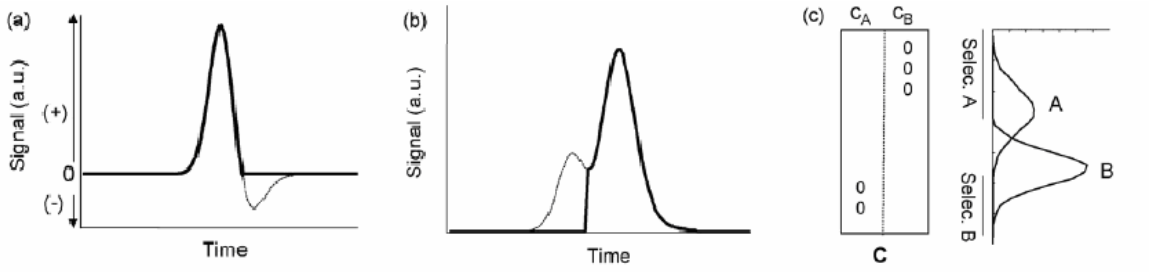

Figure 5-4. Common constraints applied to chromatographic elution profiles. (a)Non-negativity, (b)unimodality. Thick profiles in (a) and (b) are constrained profiles; thin dashed lines represent elution profiles before being constrained (the non-visible part of the unconstrained profiles overlaps with the constrained profiles). (c) Selectivity, included as null concentration for the absent compounds.

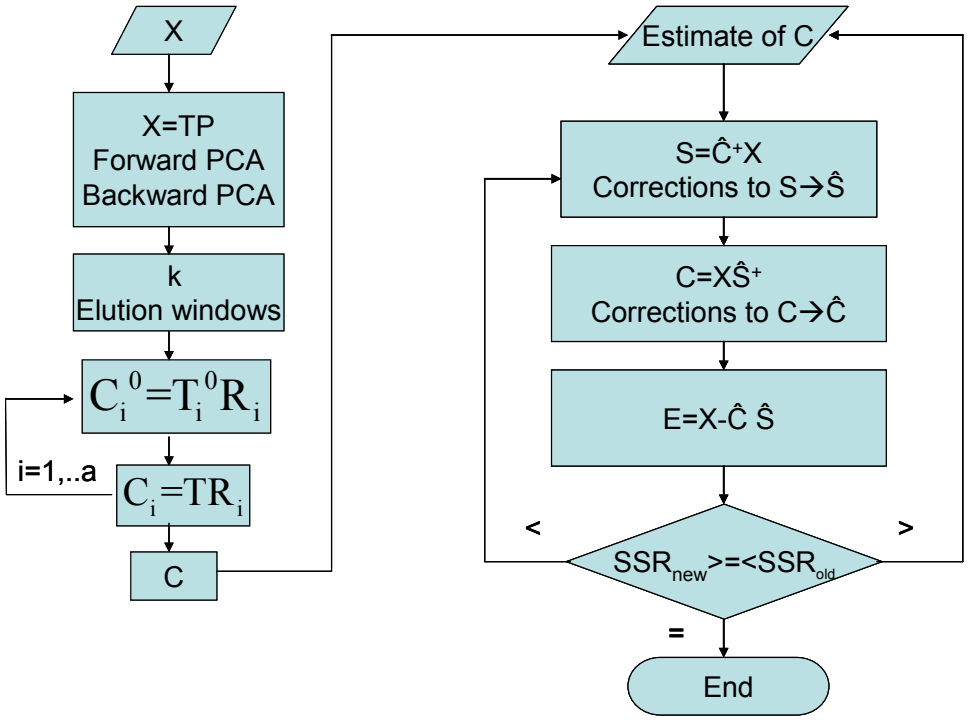

Figure 5-5. Flow chart of the EFA-ALS algorithm. The flow diagram starts from mixture response matrix X. Sequential response vectors are explored for local rank analysis by EFA, which results in elution profile C. The initial estimate of C is iteratively optimized by imposing additional constraints in ALS. See text for the notations.

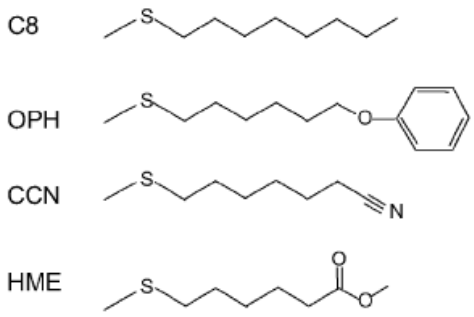

Figure 5-6. Chemical structures of the thiolate monolayers on the MPNs.

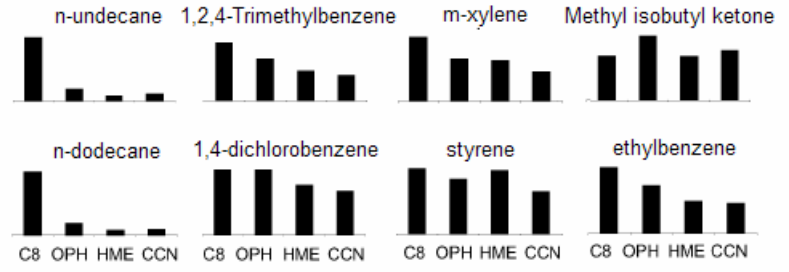

Figure 5-7. Response patterns of eight organic compounds that may present in indoor air quality study. Response pattern is derived by dividing each element in a response vector by the maximal response in the vector. (Ref. 88)

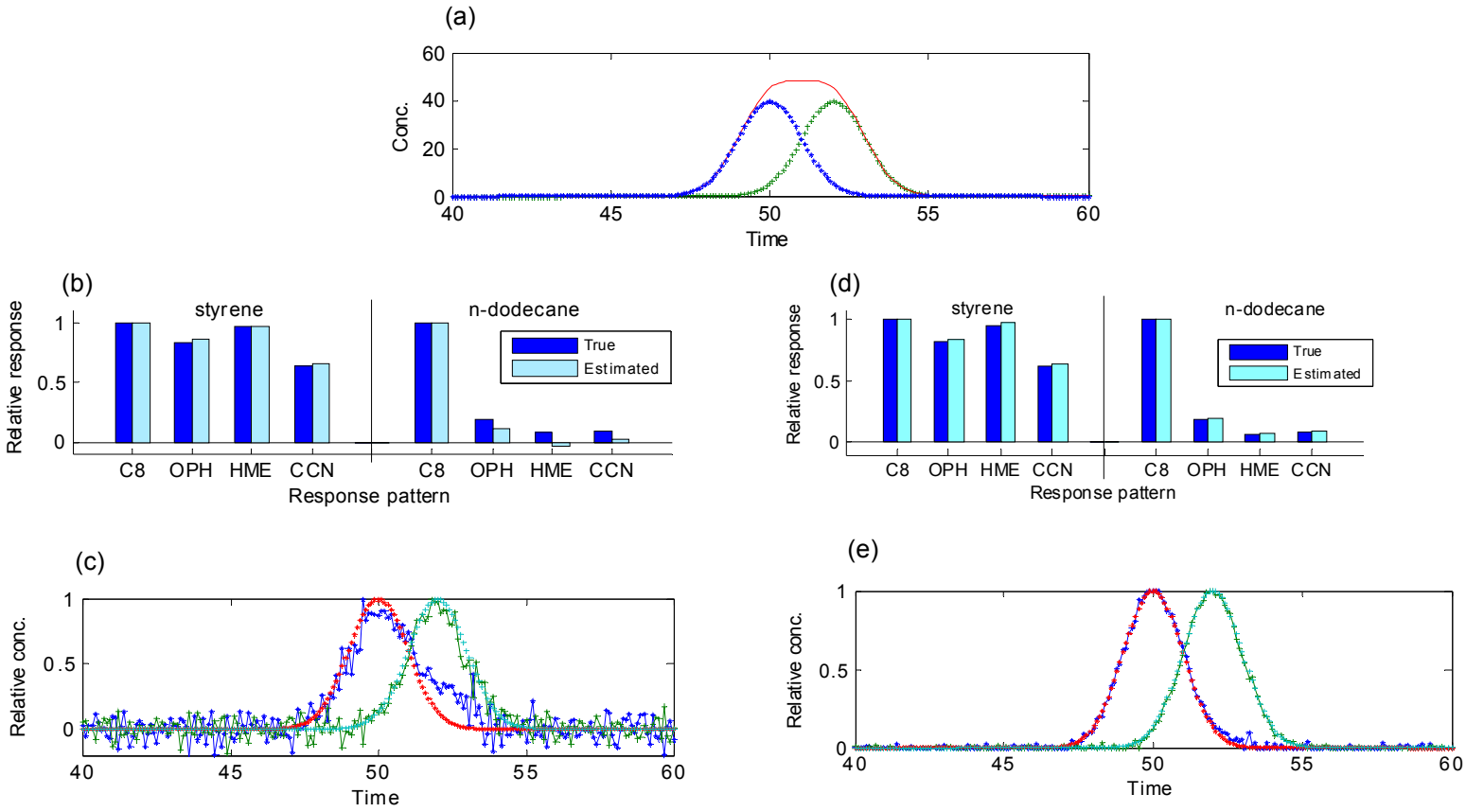

Figure 5-8. Curve resolution improvement by ALS refinement of EFA calculations. (a) co-elution of styrene ("*", A=100, t_g =50, σ=1) and n-dodecane ("+", A=100, t_g =52, σ=1). Deconvolution of styrene and n-dodecane peaks is performed separately by EFA and EFA-ALS. (b) Response patterns estimated by EFA as compared to their true patterns. (c) Elution profiles estimated (continuous curves) by EFA as compared to their true profiles (broken curves). (d) Response patterns of the same components reproduced by EFA-ALS as compared to their true patterns. (e) Elution profiles of the same components reproduced (continuous curves) by EFA-ALS as compared to their true profiles (broken curves).

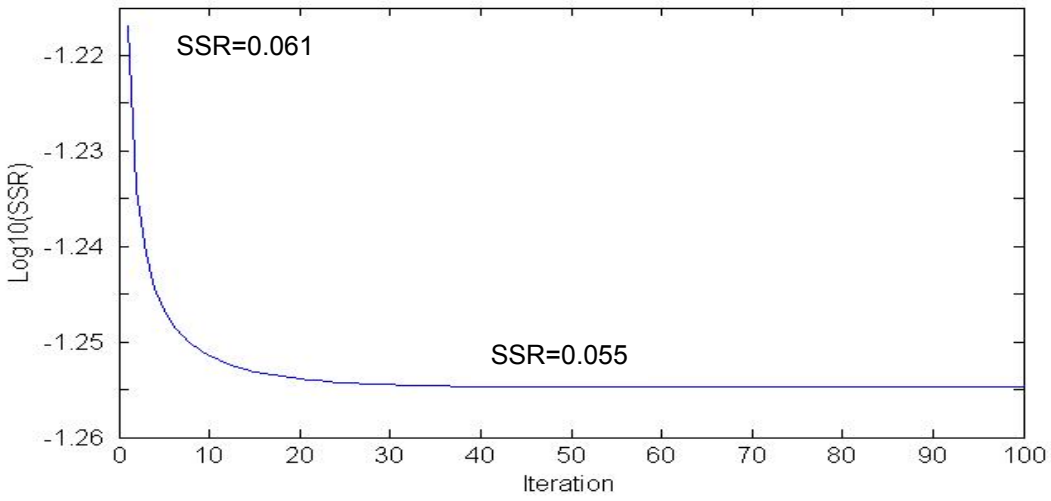

Figure 5-9. Convergence of ALS algorithm used for resolution of mixture of styrene and n-dodecane.

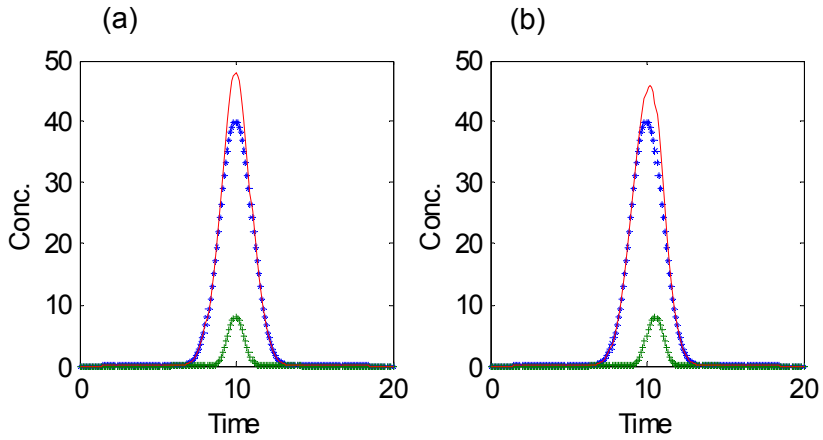

Figure 5-10. Chromatographic peak of styrene ("*") with an n-dodecane peak ("+") embedded inside. The sold curve is the composite peak of styrene and n-dodecane. (a) Styrene peak (A=100, t_R=10, σ=1) is completely overlapped with n-dodecane (A=10, t_R=10, σ=0.5). t_R(n-dodecane) –t_R(styrene) =0. (b) Styrene peak (A=100, t_R=10, σ=1) is partially overlapped with n-dodecane (A=10, t_R=10.5, σ=0.5). t_R(n-dodecane) -t_R(styrene) =0.5.

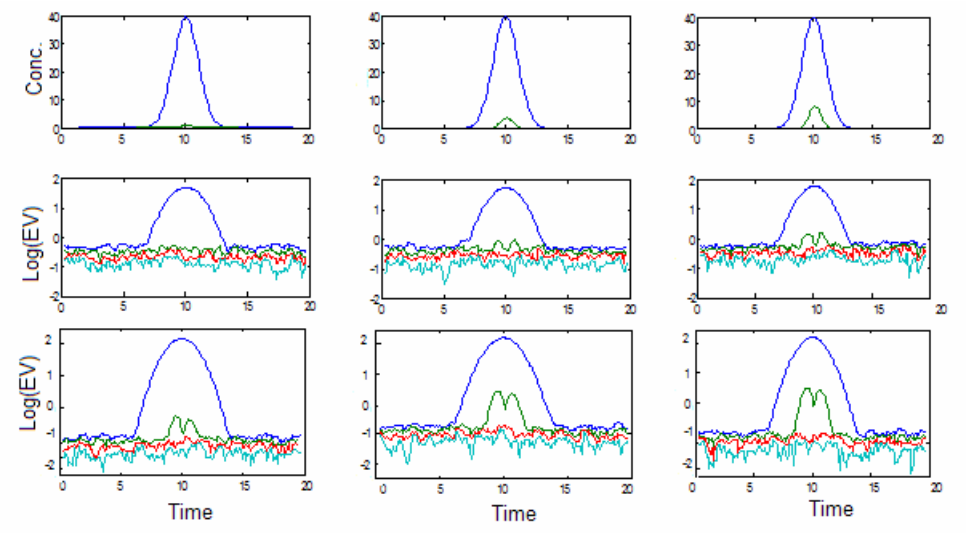

Figure 5-11. Capacity of FSMW-EFA in detecting impurity n-dodecane with a proportion 1% (top left), 5% (top middle) and 10% (top right) in styrene at S/N=100:1 (the middle three figures) and S/N=1000:1 (the bottom three figures).

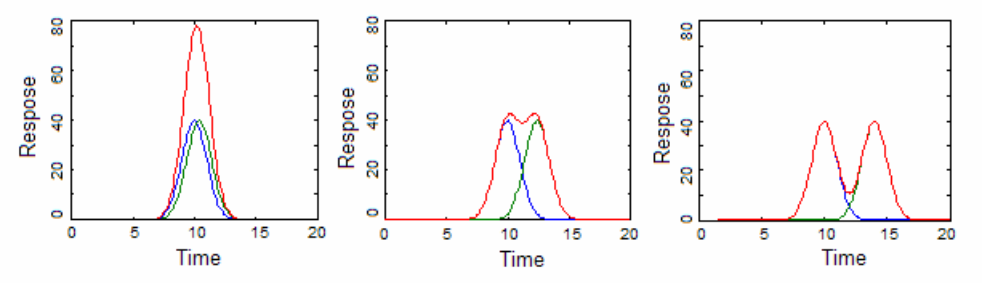

Figure 5-12. Two co-eluting peaks of different degree of overlapping. Left: *R*=0.1; Middle: *R*=0.5; Right: *R*=1.0.

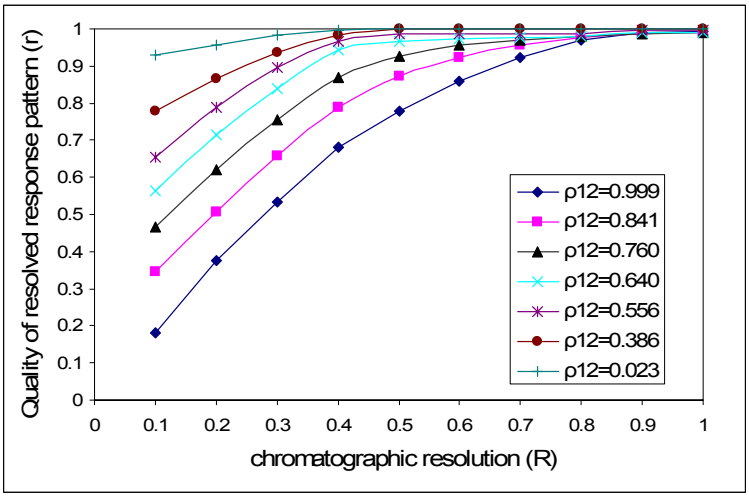

Figure 5-13. Change of resolution quality as a function of chromatographic resolution for seven vapor pairs with different spectra similarity. One of the mixture components is plotted. The same applies to the other one.

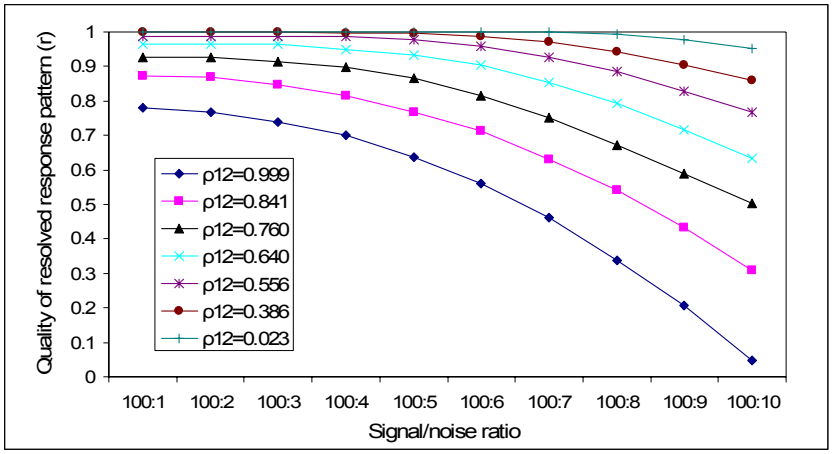

Figure 5-14. Dependence of curve resolution quality on measurement noise.

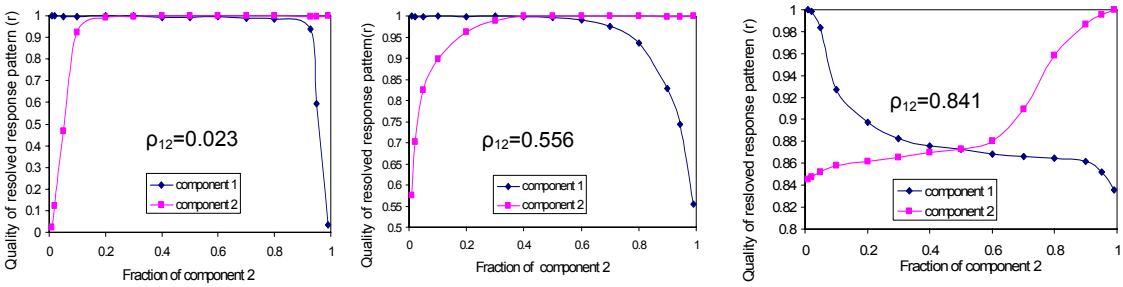

Figure 5-15. Quality of curve resolution as a function of relative concentration plotted for three vapor pairs each with a different spectra similarity.

Table 5-1: Identities of the vapor pairs combined in the simulations and their pattern similarity

Component 1	Component 2	ρ12
n-undecane	n-dodecane	0.999
1,2,4-Trimethylbenzene	1,4-dichlorobenzene	0.842
m-xylene	1,4-dichlorobenzene	0.760
1,4-dichlorobenzene	n-dodecane	0.641
styrene	n-dodecane	0.556
methyl isobutyl ketone	1,4-dichlorobenzene	0.386
methyl isobutyl ketone	ethylbenzene	0.023

Reference

1. Dieter Neue, U., *Journal of Chromatography* A, 2008, 1184(1-2), 107-130

2. Frans, S. D., McConnell, M. L., Harris, J. M., *Anal. Chem*., 1985, 57(8), 1552 – 1559

3. Davis, J. M., Glddings, J. C., *Anal. Chem*. 1983, 55, 418-424

4. Gemperline, P. J., *Anal. Chem*. 1986, 58, 2656-2663

5. Herman, D. P., Gonnord, Marie-France, Guiochon,G., *Anal. Chem*. 1984, 56, 995-1003

6. Felinger, A., Pietrogande, M. C., *Anal. Chem.*, 2001, 1, 618A-626 A

7. Chesler, S. N., Cram,S. P., *Anal. Chem.*,1973, 45(8), 1354-1359.

8. Goodman, K. J., Brenna, J. T., *Anal. Chem*., 1994, 66(8), 1294-1301.

9. Nikitas, P., Pappa-Louisi, A., Papageorgiou, A., *Journal of Chromatography* A, 2001, 912, 13–29

10. Malinowski, E. R., Factor Analysis in Chemistry, 2nd ed., Wiley, New York, 1991

11. Kowalski, B. R.; Sharaf, M. A. *Anal. Chem.* 1982, 54, 1291-1296.

12. Kowalski, B. R.; Osten, D. W. *Anal. Chem*. 1984, 56, 991-995.

13. Borgen,O.S., Kowalski, B.R., *Anal. Chim. Acta*, 1985, 174, 1 –26.

14. Karjalainen, E.J., Karjalainen, U.P., Data Analysis for Hyphenated Techniques, Elsevier, Amsterdam, 1996.

15. Keller,H.R., Massart,D.L., *Anal. Chim. Acta* 1991,246, 379.

16. Sanchez,F.C., Rutan,S.C., Garcia, M.D.G., Massart, D.L., *Chemom. Intell. Lab. Syst*. 1997, 36, 153.

17. Tauler, R., Smilde,A.K., Kowalski,B.R., *J. Chemom*. 1995, 9, 31– 58.

18. Libnau, F.O., Toft, J., Christy,A.A., Kvalheim,O.M., *J. Am. Chem. Soc*. 1995,116, 8311.

19. Forland,G.M., Libnau,F.O., Kvalheim,O.M., Hoiland,H., *Appl. Spectrosc.* 1996, 50, 1264.

20. Libnau,F.O., Christy,A.A., Kvalheim,O.M., *Vib. Spectrosc*. 1995,7, 139.

21. Libnau,F.O., Christy,A.A., Kvalheim,O.M., *Appl. Spectrosc*. 1995, 49,1431.

22. Nodland,E., Libnau, F.O., O.M. Kvalheim, H.J. Luinge, P. Klaeboe, *Vib. Spectrosc*. 1996, 10, 105.

23. Tauler, R., Kowalski,B.R., Fleming,S., *Anal. Chem*. 1993, 65, 2040– 2046.

24. R. Tauler, Smilde, A.K., Henshaw, J.M., Burgess,L.W., Kowalski,B.R., *Anal. Chem*. 1994, 66, 3337–3342.

25. Sasic,S., Ozaki,Y., Olinga,A., Siesler,H.W., *Anal. Chim. Acta* 2002, 452,265–276.

26. Sasic,S., Amari,T., Siesler,W., Ozaki,Y., *Appl. Spectrosc*. 2001, 55, 1181–1191.

27. Shen,H.L., Stordrange,L., Manne,R., Kvalheim,O.M., Y.Z. Liang, *Chemom. Intell. Lab. Syst*. 2000, 51, 37– 42.

28. Libnau,F.O., Kvalheim,O.M., Christy,A.A., Toft,J., *Vib. Spectrosc*. 1994, 7, 243–250.

29. Delaney, M. F., *Anal. Chem*., 1984, 56, 261R

30. Johnson, K., De Juan, A., Rutan, S., *J. Chemometrics* 1999, 13, 331–334.

31. Kim, B.M., Henry, R.C., *Atmos. Environ*., 2000, 34,1747–1759.

32. Maeder, M., *Anal. Chem*. 1987, 59, 527–530.

33. De Braekeleer, K., de Juan, A., Massart, D.L., *J. Chromatogr*. 1999, 832, 67–86.

34. Saurina, J., Hernandez-Cassou, S., Tauler, R., Izquierdo-Ridorsa, A., *Anal. Chem*. 1999, 71, 126–134.

35. Malinowski, E.R., *J. Chemometrics* 1996, 10, 273–279.

36. Saurina, J., Hernandez-Cassou, S., Tauler, R., *Anal. Chem*. 1997, 69, 2329–2336.

37. Johnson, P., Johansson, A.I., Gullberg, J., Trygg, J., Grung, J.A.B. S. Marklund, M. Sjöström, H. Antti, T. Moritz, *Anal. Chem.* 2005, 77, 5635.

38. Jonsson, P., Johansson,E.S., Wuolikainen,A., Lindberg, J., I. Schuppe-Koistinen, M. Kusano, M. Sjöström, J. Trygg, T. Moritz, H. Antti, *J. Proteome Res*. 2006, 5, 1407.

39. Lawton,W.E., Sylvestre, E.A., *Technometrics*, 1971,13, 617.

40. Ohta, N., *Anal. Chem*. 1973, 45, 553-557.

41. Chen, J-H., Hwang, L.-P. *Anal. Chim. Acta* 1981, 733, 271-279.

42. Vandeginste, B., Essers, R., Bosman. T., Reijnen, J., Kateman, G. *Anal. Chem*. 1985, 57, 971-985.

43. Meister, A. *Anal. Chim. Acta* 1984, 167, 149-161.

44. Borgen, O. S., Kowalski, B. R. *Anal. Chim. Acta* 1985, 774, 1-26.

45. Gemperline, P. J. *J. Chem. Inf. Comput. Sci*. 1984, 24, 206-212.

46. Vandeginste, B. G. M., Derks, W., Kateman, G. *Anal. Chim. Acta* 1985, 173, 253-264.

47. Lacey, R. F. *Anal. Chem.* 1988, 58, 1404-1410.

48. Gampp, H., Maeder, M., Meyer. C. J., Zuberbuehier, A. D. *Talanfa* 1985, 32, 1133-1139.

49. Gampp, H., Maeder. M., Meyer, C. J., Zuberbuehler, A. D. *Chimia*. 1985, 39. 315-317.

50. Maeder, M., Zuberbuehier, A. D *Anal. Chim. Acta* 1986, 787, 287-291.

51. Malinowski, E.R. *J. Chemom*. 1992, 6, 29.

52. Schostack, K.J., Malinowski, E.R., *Chemom. Intell. Lab. Syst*. 1993, 20, 173.

53. Manne, R., Shen, H., Y. Liang, *Chemom. Intell. Lab. Syst*. 1999, 45, 171.

54. Shen, H.L., Manne, R., Q.S. Xu, D.Z. Chen, Y.Z. Liang, *Chemom. Intell. Lab. Syst*. 1999, 45, 323.

55. Kvalheim, O.M., Liang,Y.Z., *Anal. Chem*. 1992, 64, 936.

56. Liang, Y.Z., Kvalheim,O.M., Keller,H.R., Massart,D.L., P. Kiechle, F. Erni, *Anal. Chem*. 1992, 64, 946.

57. Jiang, J.H., Sasic, S. R., Yu, Y. Ozaki, *J. Chemom*. 2003,17, 186.

58. Malinowski, E.R., *J. Chemom.* 1996, 10, 273.

59. Keller,H.R., Massart,D.L., *Anal. Chim. Acta* 1991, 246, 379.

60. Keller, H.R., Massart, D.L., de Beer, J.O., *Anal. Chem*. 1993,65, 471.

61. de Juan, A., Tauler, R., *Journal of Chromatography* A, 2007, 1158, 184–195

62. Gilliard, J.A., Cumps, J.L., Tilquin, B.L., *Chemom. Intell. Lab. Syst.* 1993, 21, 235.

63. de Braekeleer, K., de Juan, A., Massart, D.L., *J. Chromatogr*. A, 1999,832, 67.

64. Archer, B. G., Hobbs, J. S., Pracrical resolution of overlapping diode array absorbance chromatogram profiles, poster presented at HPLC ,91 Conference held in Basel, Switzerland, June 2-7, 1991

65. Keller, H.R., Massart, D.L., *Anal. Chim. Acta* 1992,263, 21.

66. Keller, H.R., Massart, D.L., Y.Z. Liang, O.M. Kvalheim, *Anal. Chim. Acta* 1992,263, 29.

67. Keller,H.R., Massart,D., Kiechle, L, Erni,P. F., *Anal. Chim. Acta* 1992,256, 125.

68. Tauler, R., *Chemometr. Intell. Lab. Syst*. 1995,30, 133– 146.

69. Tauler, R., Smilde, A.K., Kowalski, B.J., *J. Chemom.* 1995, 9, 31– 58.

70. Tauler, R., *J. Chemom.* 2001, 15, 627–646.

71. Amrhein, M., Srinivasan, B., Bonvin, D., M.M. Schumacher, *Chemometr. Intell. Lab. Syst.* 1996, 33, 17– 33.

72. Saurina, J., Hernández-Cassou,S., Tauler, R., Izquierdo-Ridorsa, A., *J. Chemom.* 1998,12, 183–203.

73. Tauler, R., de Juan, A., in P. Gemperline (Ed.), Practical Guide to Chemometrics, CRC Press, Boca Raton, FL, 2006, p. 417.

74. de Juan, A., Tauler, R., *Anal. Chim. Acta* 2003, 500, 195.

75. de Juan, A., Tauler, R., *Crit. Rev. Anal. Chem*. 2006, 36, 163.

76. Vandeginste, B.G.M., Derks, W., Kateman, G., *Anal. Chim. Acta* 1985,173, 253.

77. Karjalainen, E.J. *Chemom. Intell. Lab. Syst*. 1989,7, 31.

78. van Zomeren, P.V., Darwinkel, H., Coenegracht, P.M.J., de Jong, G.J., *Anal. Chim. Acta* 2003, 487, 155.

79. Windig,W., Guilment,J., *Anal. Chem*. 1991,63, 1425.

80. Sánchez, F.C., Massart, D.L., *Anal. Chim. Acta* 1994, 298, 331.

81. Sánchez, F.C., Toft, J., van den Bogaert, B., Massart, D.L., *Anal. Chem*. 1996, 68, 79.

82. Grung, B., Nodland, E. G. M., Forland, *Journal of Chemical Education*, 2007, 84(7), 1193- 1200

83. Vosough, M., Mason, C., Tauler, R., Jalali-Heravi, M., Maeder, M., *J. Chemometrics* 2006, 20, 302–310

84. Hsieh, M.-D., Zellers, E. T. *Anal. Chem*. 2004, 76, 1885-1895.

85. Zellers, E. T., Batterman, S. A., Han, M., Patrash, S. J. *Anal. Chem*. 1995, 67, 1092-1106

86. Jin C., Kurzawski P., Hierlemann A., Zellers E. T. *Anal. Chem*. 2008, 80, 227-236

87. Park, J., Groves, W. A., Zellers, E. T. *Anal. Chem*. 1999, 71, 3877-3886

88. Zhong, Q.iongyan, Veeneman, R., William H. Steinecker, Chunrong Jia, Stuart A. Battermana and Edward T. Zellers, *J. Environ. Monit.*, 2007, 9, 440–448

89. Lu, C. J., Steinecker, W. H., Tian, W. C., Agah, M. J., Potkay,M. M., Oborny,C., J. Nichols, H. Chan, J. Driscoll, R. D. Sacks, S. W. Pang, K. D. Wise and E. T. Zellers, *Lab on a Chip*, 2005, 5, 1123–1131.

90. Rowe, M. P., Plass, K. E., Kim, K., C. Kurdak, E. T. Zellers and A. J. Matzger, *Chem. Mater*., 2004, 16, 3513–3517.

91. Steinecker, W.H., Rowe, W.P. and Zellers, E.T., *Analytical Chemistry*, 2007, 79(13), 1164-1172.

92. Gutierrez-Osuna, R. *IEEE Sensors Journal*, 2002, 2 (3), 189- 202

93. Kalivas, J. H., *Analytical Letters*, 2005, 38(14), 2259 – 2279

94. Giddings, J. C., Dynamics of Chromatography, vol. 1, edited by J. C. Giddings and R. A. Keller, Marcel Dekker, New York, 1965.

CHAPTER 6

Conclusions and Future Work

This dissertation has addressed a range of important issues related to the development and application of microsensor-array based microanalytical system to the determination of VOC contaminants in the occupational and ambient environment. The ability to detect trace level analyte at minimal cost and the selective detection of analyte in a complex matrix have made chemical sensor array an excellent VOC detector and explained its popularity in miniaturized VOC analyzers. [1] Coupling microsensor array detection with upstream chromatographic separation makes it possible to characterize complex VOC mixtures more reliably than a stand-alone sensor array detector. The multivariate nature of data generated from such hyphenated instrumentation demands the adaptation of sophisticated statistical methods to the task of data interpretation. Chemometrics is a subfield in chemistry where mathematics, statistics and modem computer technology are used to model and interpret multivariate data. Its basic application to the sensor-array based analytical system includes pattern recognition for vapor identification and multivariate calibration for vapor quantification. Extended disjoint principal component regression (EDPCR) model, the combination of principal component classification model with principal component regression model, has been extensively used with microsensor array for the purpose of VOC recognition and

quantification (see Chapter 1). The use of Monte Carlo simulation with EDPCR dispenses with laborious measurements and provides a valuable tool for evaluating the performance of a sensor array in analysis of individual vapors and vapor mixtures. This approach has been used in Chapter 3 and Chapter 4 for evaluation of the classification performance of multi-transducer arrays.

For the first time, the fidelity of response pattern to its reference pattern was studied by taking into consideration sensor response variability. Chapter 2 presented a decision rule for vapor identity assignment in the context of vapor analysis by sensor-array based GC. After separating a vapor mixture into pure components in the form of individual chromatographic peaks, the retention index of a vapor enables its tentative identification. To ensure unambiguous vapor identity assignment, the fidelity of the measured pattern to its reference pattern has to be evaluated. Variations on sample collection and analysis in field instrumentation cause distortions in response patterns and make it hard to judge whether the measured pattern is due solely to a single analyte. This problem is unique to GC-sensor array systems and has not been addressed before. By accounting for the variability of sensor responses during calibration, the variations in response pattern measured for a vapor can be quantified using the Mahalanobis distance which is the statistic distance from calibrated sample to their centroid. A residual distance can be defined. For a unknown sample if its Mahalanobis distance is below the threshold, its identity can be assigned to the vapor at a predefined confidence level, otherwise the identify cannot be assigned to the vapor for which the class model is built, and conclusion may be drawn that there might be more than one vapor or the instrument is out of calibration. The selection of the suitable number of principal components for describing

the pattern variations and establishing an optimal threshold is not an issue in this chapter because a boot-strap algorithm is employed to select the number of principal components that gives the minimal residual error of prediction of original calibrated data set. This decision rule has been applied successful to a surface acoustic wave sensor array and the stability of sensor responses, as well as the effects of environmental factors on sensor responses, has been evaluated.

This pattern fidelity testing method can not only be used for studying pattern distortion by environmental factors, but be applied to contaminated peaks. In future work it would be interesting to use this approach to quantify the threshold amount of impurities in an isolated peak beyond which the major peak is so contaminated that the measured response pattern can no longer be matched to its reference pattern. However, just like the multivariate curve resolution method introduced in Chapter 5, this pattern fidelity assessment method is insensitive to co-eluting analytes who have the same response pattern, since the measured relative response pattern of the major peak is in effect unchanged by increasing the concentration of the contaminant. Nevertheless, this Mahalanobis-distance based pattern fidelity method provides a basic uncertainty assessment tool for pattern recognition with GC-sensor array systems. It offers clue of abnormal measurement condition by monitoring variations in sensor responses, and helps with the decision-making with regards to whether or not the identity of an isolated peak can be unambiguously assigned through pattern matching.

The approach in Chapter 3 does not discriminate among the specific situations that cause pattern infidelity which may include abnormal measurement or peak contamination. Frequently, it is required to make a judgment whether or not an isolated

peak is contaminated and contains inner peak(s). Although it is possible to separate an embedded peak from the embedding peak via change of chromatographic conditions, peak purity can be effectively assessed by mathematical curve resolution. Peak purity assessment by Fixed-Sized Moving Window EFA is illustrated in Chapter 5 with an example. Furthermore a hybrid self-modeling curve resolution method that incorporates iterative alternating least square (ALS) algorithm into non-iterative evolving factor analysis (EFA) is described for its application to deconvolution of overlapped chromatographic peaks. Advantage of this hybrid algorithm over EFA is also illustrated, and the capacity limits of EFA-ALS in peak deconvolution as a function of chromatographic resolution *R*, response pattern similarity, signal-to-noise ratio, and quantitative composition have been explored. It is the first attempt in sensor array studies to use multivariate curve resolution for peak purity assessment and peak deconvolution. The ability to resolve overlapped peaks numerically is essential, considering the ubiquitous peak overlap in fast GC analysis of complex VOC mixtures.

The capacity of EFA-ALS presented in Chapter 5 has been studied based mostly on simulated chromatographic data. Further exploitation of multivariate curve resolution for mixture analysis entails the testing of EFA-ALS algorithm on chromatographic measurement data. Future work also includes the application of peak purity assessment, peak deconvolution, and pattern recognition in the context of retention time windows.

A seemingly separate issue is the selectivity of a sensor array that remains as the focus of chemical sensor research. In Chapter 3 a study of vapor recognition and quantification by polymer-coated multi-transducer (MT) arrays is described for maximizing sensor array selectivity. The primary data set consists of experimentally

derived sensitivities for 11 organic vapors obtained from 15 microsensors comprising five cantilever, capacitor, and calorimeter devices coated with five different sorptive-polymer films. These are used in Monte Carlo simulations coupled with principal component regression models to assess expected performance. Recognition rates for individual vapors and for vapor mixtures of up to four components are estimated for single-transducer (ST) arrays of up to five sensors and MT arrays of up to 15 sensors. Recognition rates are not significantly improved by including more than five sensors in an MT array for any specific analysis, regardless of difficulty. Optimal MT arrays consistently outperform optimal ST arrays of similar size, and with judiciously selected 5-sensor MT arrays one-third of all possible ternary vapor mixtures are reliably discriminated from their individual components and binary component mixtures, whereas none are reliably determined with any of the ST arrays. Quaternary mixtures could not be analyzed effectively with any of the arrays. A 'universal' MT array consisting of eight sensors is defined, which provides the best possible performance for all analytical scenarios. Accurate quantification is predicted for correctly identified vapors.

One shortcoming of the study presented in Chapter 3 is that effect of mixture concentrations is ignored by fixing all component concentrations to some constant values. This shortcoming is overcome in Chapter 4 where the discrimination of simple vapor mixtures from their components with polymer-coated multi-transducer (MT) arrays as a function of the absolute and relative concentrations of those components is explored. The data set consists of calibrated responses to 11 organic vapors from arrays of five or eight microsensors culled from a group of five cantilever, five capacitor, and five calorimeter transducers coated with one of five different sorptive-polymer films. Monte Carlo

methods are applied to simulate error-enhanced composite responses to all possible binary and ternary mixtures of the 11 vapors, and principal component regression models are established for estimating expected rates of recognition as a function of mixture composition. The limit of recognition (LOR), defined as the maximum recognizable mixture composition range, is used as the metric of performance. With the optimal 8-sensor MT array, 19 binary and three ternary mixtures could be discriminated from their components with < 5% error. The binary-mixture LORs are shown to decrease with increases in the baseline noise levels and random sensitivity variations of the sensors, as well as the similarity of the vapors. Importantly, most of the binary LOR contours are significantly asymmetric with respect to composition, and none of the mixtures could be recognized with < 5% error at component relative concentration ratios exceeding 20:1. Discrimination of ternary mixtures from their components and binary sub-component mixtures was possible only if the relative concentration ratio between any two of the components was <8:1. Although the binary LORs for the best 5-sensor single-transducer (ST) array were generally smaller than those of the best 5-sensor MT array, the differences are not significant and the ST-array LODs were about an order of magnitude lower. The implications of these results are considered in the context of using such as arrays as detectors in microanalytical systems.

These chapters serve the same goal: to allow chemical identification and quantification using sensor-array based GC analytical system. Linkage of the research presented in these chapters is obvious by referring to Figure 6-1. Given a chromatogram of a vapor mixture produced by a GC-sensor array system, the identity of each peak needs to be found by reference library search, and then the concentrations of each peak

needs to be quantified. Obviously the data analysis should be performed peak by peak in accordance with the sequential elution of mixture components. With each peak (or peak cluster in case of overlap) constituting a retention time window, the chromatogram is divided into a series retention time window that will be the units of chemometric computations. Depending on the efficiency of column separation, the window may contain an individual peak or a cluster of peaks that overlap each other. Since the retention time can used for qualitative analysis, the identity of the peak(s) in a retention time window can be screened by matching its retention index to a retention index library. Retention index is used because it is independent of chromatographic conditions. The candidates for vapor identity assignment can be narrowed to the vapors whose retention indices falls within the retention time window in which the peak is to be identified. Therefore the load of pattern library matching task is greatly relieved by focusing on the relatively few candidate vapors as a result of retention screening. For a single peak in the window, its purity can be assessed by finding if there is an inner peak embedded in it. Fixed-size moving window evolving factor analysis can be used for peak purity evaluation. If the peak is found contaminated, either a chromatographic condition adjustment or a numerical resolution may separate the contaminant from the peak. Numerical separation is especially useful when several peaks overlap together in a retention time window and hardware separation is hard to achieve. Evolving factor analysis can be thereby employed. Once the peaks are separated chromatographically or chemometrically, spectral library search will be performed to match the unknown to the reference patterns. The decision rule developed in Chapter 2 can be used for vapor identity assignment by reference library search. The merit of pattern fidelity testing relies

on its ability to quantify decision-making by taking into consideration of sensor response variations. Whereas it was developed for evaluation of pattern change caused by impurity or co-eluting unknown analyte, this pattern fidelity method can also be used for study of stability and reproducibility of sensor response under influence of environmental factors like humidity and temperature. By using this pattern fidelity testing, a quality control of sensor measurement can be established that is similar to the multivariate quality control chart used in industrial process. The fluctuation of response pattern under normal conditions can be quantified by using the method presented in Chapter 2. If an abnormal pattern change occurs, it suggests deviation of measurement from normal operations and serves as indicator of further remediation.

The capacity of chemometrics is not limited to the studies described in this dissertation. From data acquisition and sensor signal processing to array optimization and system evaluation, there exists a wide range of issues that can be addressed by chemometrics. Chemometrics has grown to a point that all sorts of algorithms are available for solving common chemical problems. The willingness to bring these chemometric methods to problem-solving and the awareness of the essence of math computing in exposure assessment becomes even more important when massive databases are constructed and digital scientific instruments capable of producing enormous multivariate data in a short amount of time are widely used, and the concomitant development of powerful computing hardware is never ending. A few thoughts specific to MicroGC are provided.

Chromatographic separation is the bottleneck of efficiency of mixture analysis. To ensure fast detection, a short column is required at the cost of decreased

chromatographic resolution. More efficient chromatographic technology like two-dimensional GC (2D GC or GC×GC) has emerged. A tandem of GC columns makes it possible to separate mixtures that cannot be done on a traditional capillary column. Modern fabrication technology is able to make the column very small and suitable to use in a microanalytical system. The addition of another separation dimension to MicroGC means three-way data produced for a sample run which demands more advanced chemometric methods to analyze these higher-dimensional data. Chemometric methods such as the generalized rank annihilation method (GRAM) have been used to extend the amount of retention information obtained from GC×GC instruments.[2] Three-way or higher-way data and their related chemometric methods (PARAFAC, GRAM, etc) have been proven to be useful in dealing with two-dimensional chromatographic data with a multivariate response (formatted into: retention time 1×retention time 2×response).[3] The application of chemometrics to illuminating analytical system with three or more information direction remains an intriguing research area.

As a last thought, the chemometric algorithms developed in this dissertation, no matter for multivariate calibration or self-modeling curve resolution, should be eventually programmed into intelligent user interface of MicroGC powered by a potent on-board microprocessor. Such application software, which should be developed by a chemometrician and a computer programmer, becomes an indispensable part of modern instrumentations. With minimal operator intervention, the software controls every aspect of an analysis from the instrument operation, data generation, detector calibration and library search, to output of sample identity and amounts of concentration. Quality controls and uncertainty assessment should also be embedded into the instrument to

realize intelligent instrumentation as well as automatic and accurate VOC analysis.

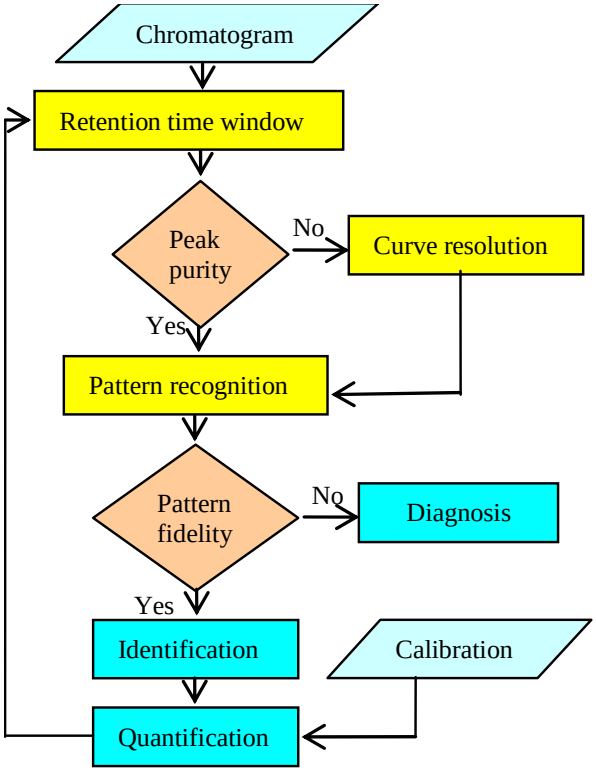

Figure 6-1: A chemometric view of VOC analysis by sensor-array based GC analytical system

Reference:

1. Röck, F., Barsan, N., Weimar, U., *Chem. Rev.* 2008, 108, 705-725.

2. Fraga, C. G., Prazen, B. J., Synovec, R. E., *Anal. Chem.* 2000, 72, 4154-4162

3. Porter, S.E.G., Stoll, D.R., Rutan, S.C., Carr, P.W., Cohen, J.D., *Anal. Chem.* 2006, 78, 5559.