Development of theory, instrumentation, and chemometric data analysis tools for comprehensive two- and three-dimensional gas chromatography

Timothy J. Trinklein

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2023

Reading Committee:

Robert E. Synovec, Chair

Dan Fu

Stefan Stoll

Program Authorized to Offer Degree:
Chemistry

University of Washington

Abstract

Development of theory, instrumentation, and chemometric data analysis tools for comprehensive two- and three-dimensional gas chromatography

Timothy J. Trinklein

Chair of the Supervisory Committee:
Title of Chair Robert E. Synovec
Department of Chair

Comprehensive two-dimensional gas chromatography coupled with time-of-flight mass spectrometry (GC×GC-TOFMS) is now recognized as a powerful analytical platform for the separation and analysis of complex mixtures containing volatile and semi-volatile components. However, two major impediments have limited widespread adoption of the technique. First, commercial instrumentation is cost-prohibitive and is not optimized for high peak capacity. Second, the technique generates large and complex data, requiring many man-hours of analysis by an expert user(s) to generate chemically meaningful information. This dissertation presents several developments in instrumentation, data analysis methods, and fundamental theory in order to address these challenges. First a new modulation method, termed dynamic pressure gradient modulation (DPGM), is described, and shown to produce high peak capacity separations in a short time frame using mostly off-the-shelf components. Next, the three-dimensional gas chromatograph - TOFMS (GC^3-TOFMS) is improved to provide total analyte transfer between all dimensions by employing DPGM as a modulator in series with cryogenic modulation. As a result of tandem modulation with total-transfer, exceptional signal enhancement from the first dimension to the third dimension was observed. To verify and study this observation, equations which predict signal enhancement for GC are derived and experimentally validated. In second part of this dissertation, the data analysis tool tile-based F-ratio analysis is investigated to provide suggestions on user-inputs and broaden the scope of the method. Using elemental bromine, olefins are derivatized in gasoline, followed by analysis of the original and brominated gasoline. Tile-based F-ratio analysis is used to quickly locate and identify the selectively derivatized analytes in the complex gasoline matrix. Finally, computer simulations are used to generate GC×GC-MS separations with realistic run-to-run retention time shifting using low-frequency "shift functions". The simulated data are used to evaluate the performance of tile-based F-ratio as a function of the amount of shifting along with chromatographic saturation and peak area *RSD*. The results of this study provide the user community with parameter recommendations (e.g., tile size) depending on the amount of shifting, saturation, and *RSD* observed.

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

It is difficult, if not impossible, to appropriately thank all of those who have played a role in my professional development. In attempting to do so, I will start at the University of Washington, and work back in time. Rob Synovec is sincerely thanked for his mentorship and enthusiastic support of this work. It goes without saying that this dissertation would not be possible without his guidance and input. I am grateful to my committee for their input and advice during my second-year, general, and final examinations. I also owe special thanks to my fellow graduate students, who have provided their support, advice, and friendship over the past five years. Of particular note are Sonia, Grant, and Caitlin. I thank my friends for necessary relief from graduate work: Edgar, Collin, Eric, Javi, and Fahad, to name just a few. I owe Kristina for her support and editorial comments to this thesis. Special thanks are due to my undergraduate research advisor, Joe Aldstadt. My first introduction to analytical chemistry was through a quantitative analysis course taught by Joe, who suggested I perform undergraduate research in his lab. Joe has continued to be a lifelong mentor and a friend. Lastly, I must thank my family. I have enjoyed spending time with my brother Mark and his family who live in nearby Olympia. I am grateful for the wisdom passed on from my grandparents, Rose, Ruth, and Mark. My parents, Mike and Lynne, have been very supportive of my personal and professional goals. I likely would not have pursued graduate (or undergraduate studies) without their support and the value of education and learning they instilled upon me from a young age.

DEDICATION

To my mother and father, who have given me the freedom to pursue my chosen craft.

Chapter 1

INTRODUCTION TO COMPREHENSIVE MULTIDIMENSIONAL GAS CHROMATOGRAPHY INSTRUMENTATION, PRINCIPLES, AND DATA ANALYSIS

1.1 Overview

The work in this dissertation covers new technological, theoretical, and computational developments in comprehensive multidimensional gas chromatography (GC). Pioneered by Liu and Phillips in 1991 [1], comprehensive two-dimensional (2D) gas chromatography (GC×GC) has become an indispensable technique for the separation and characterization of complex volatile and semi-volatile mixtures. The invention of GC×GC may very well be the most significant development in GC since the introduction (and theory of) of capillary GC columns by Marcel Golay [2]. Compared to one-dimensional gas chromatography (1D-GC), a well-designed GC×GC experiment can achieve an order-of-magnitude higher peak capacity [3] and increased detectability [4,5]. Further, in GC×GC separations the compound classes are highly ordered in 2D space, aiding interpretability. In GC×GC, fractions of effluent from the first dimension (^{1}D) column are continuously sampled and reinjected into the second dimension (^{2}D) column via a sampling and injection interface termed the modulator. Hyphenation with an information-rich, multivariate detector, such as time-of-flight mass spectrometry (GC×GC-TOFMS) or vacuum-ultraviolet spectroscopy (GC×GC-VUV) creates a powerful analytical tool well-suited to address challenging analyses in numerous application areas.

1.2 Components of a GC×GC instrument

A schematic of a GC×GC instrument is shown in Fig. 1.1. Key components of the instrument include the inlet, ^{1}D separation column, the modulator, the ^{2}D separation column, and the detector. The inlet is no different than that on a standard 1D-GC instrument, and any injection mode suitable for 1D-GC can be used for GC×GC. The ^{1}D column likewise has dimensions typical of a standard 1D-GC instrument (~10-60 m in length). The most critical component of a GC×GC instrument is the modulator, which periodically collects effluent from the ^{1}D column and injects it onto the ^{2}D column. The time between subsequent injections is termed the modulation period, denoted P_M. Many types of modulators exist, which will be discussed in detail subsequently, and the development of a new modulation method will be introduced in Chapter 2.

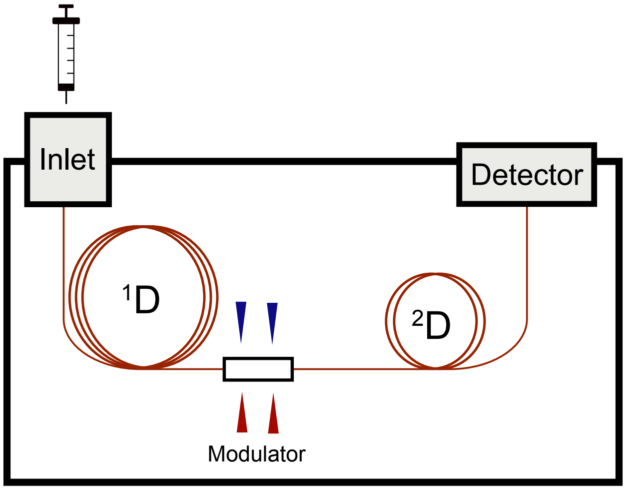


Figure 1.1: Schematic of a standard GC×GC instrument.

The ^{2}D column is short (~0.5-2 m in length) and is selected to have a stationary phase of different selectivity than the ^{1}D column such that peaks that coelute on ^{1}D can separate on ^{2}D due to different retention properties. Because the modulator generates narrow injection pulses and the ^{2}D column is short, peak widths exiting the ^{1}D column are on the order of tens to hundreds of milliseconds. Therefore, the detector in a GC×GC instrument must have a suitably high data acquisition frequency to reconstruct the narrow ^{2}D peaks. Two particularly useful detectors for GC×GC are the flame ionization detector (FID) and time-of-flight mass spectrometry (TOFMS), both capable of data acquisition frequencies suitable for narrow ^{2}D GC×GC peaks.

1.2.1 The modulator

1.2.1.1 A brief history of modulation

The modulator is the enabling technology of GC×GC—without it, the technique would not be possible. In fact, the invention of modulation for GC preceded GC×GC, and it was only later realized that modulators could be used to facilitate a comprehensive 2D separation. So, if not for GC×GC, what were modulators originally used for? The astute reader will recognize that the term modulation arises from signal processing and communications theory [1]. While conventional chromatographic separations are performed with one injection, a number of researchers in the 1970s and 1980s investigated a new technique termed multiplex chromatography [6], a method that simultaneously increases the signal-to-noise ratio (S/N) while allowing multiple chromatograms to be obtained nearly simultaneously. Multiplexing had long been used in optical spectroscopy, in particular for instruments with inherently poor S/N such as infrared spectroscopy [7] where all signals are acquired simultaneously rather than by using dispersing elements. Chromatographers wondered if these basic principle could be employed in chromatography by treating the separation occurring in the column as a chemical “signal”. Thus, an ideal modulator should raise the frequency of the chemical signal entering the detector. Low-frequency noise is a major contributor to the overall noise in chromatographic experiments, therefore increasing the frequency domain of the signal should improve the S/N. The modulation frequency can be generated using a pseudo-random number generator, or can occur at a fixed period.

Chemical modulation in a gas chromatographic column was first reported by James Lovelock [8]. Lovelock placed an electron capture detector (ECD) *in the middle* of a GC column and periodically switched the “detector” on and off, thus periodically destroying electronegative species. After demodulation, Lovelock observed a significantly improved S/N This early study demonstrated the principle of chemical modulation, although it had limited practical use [2]. Over a decade later, Phillips demonstrated a much more practical and broadly applicable chemical modulator, the thermal desorption modulator (TDM) [9]. The TDM operates by rapidly heating a section of the GC column, releasing any components retained by the stationary phase.

Phillips published several seminal papers applying the TDM to multiplex gas chromatography, such as its application to measuring methane in ambient air or trace organic compounds in water. and Then, Phillips had a striking realization: the TDM could operate as a sampling and injection apapratus to facilitate a comprehensive two-dimensional gas chromatographic separations [1]! And so, GC× was born.

---

[1] In comprehensive 2D liquid chromatography, the term “modulation” is used less frequently, and now only after the adoption of GC×GC. We can attribute this to historical reasons, as John Phillips developed GC×GC with a signal processing background, while Enri and Frei [59] and Bushey and Jorgenson [60] approached comprehensive separations with a more traditional chromatographic background. The terms “sampling” and “injection” were used insetad.

[2] James Lovelock! The brilliant English scientist invented the ECD, chemical modulation, and perhaps even the microwave! Lovelock built one of the first ever microwaves in early experiments attempting to determine if was feasible to resuscitate frozen rodents with microwave radiation [61]. No joke! Unfortunately, Lovelock didn’t realize that microwave radiation could be used to reheat frozen Lean Cuisine, and was never credited for the commercialization of the household appliance. Lovelock invented the ECD for GC, and then used it to discover the alarming abundance of chlorofluorocarbons in artic air. Lovelock is probably most famous for his development of the Gaia Hypothesis, which propose that everything living and non-living on earth operates as one in a synergistic and self-regulating complex system.

1.2.1.2 Thermal modulation

Today, modulators for GC×GC can be broadly classified into two groups: thermal modulators and flow modulators. The first thermal modulator was Phillips's TDM[3], but since then thermal modulation technology has significantly evolved. One of the first commercial thermal modulators was the "rotating sweeper" sold by Zoex Corporation,[4] whereby a segment of capillary column was slotted in-between a heater block, which rotated or "swept" around the column to periodically heat it [10,11], desorbing trapped compounds. This modulator enabled a number of early studies, but like Phillips early TDM, it was eventually realized that the rotating sweeper was not a robust solution, and frequently broke capillary columns. Further, at hot oven temperatures, modulators which only heated the column, with no cooling, were ineffective.

In 1997, Philip Marriott and his student Russell Kinghorn developed a cryogenic modulator initially used for 1D-GC, termed the longitudinally modulated cryogenic system (LMCS) [12]. This device operates by longitudinally moving a cryogenically-cooled coaxial cylinder up and down a section of capillary column, trapping and remobilizing solutes. While listening to talks about GC×GC at the International Symposium of Capillary Chromatography (ISCC), Marriott and Kinghorn realized their modulator could be applied for GC×GC. Soon, they reported the LMCS as the first cryogenic GC×GC modulator [13]. Because cooling is involved, the modulator was effective at modulating solutes at hot oven temperatures, extending the useful volatility range of GC×GC.

Still, a moving modulator system was not robust for long-term operation. In 2000, a year after John Phillip's untimely passing in 1999, Ed Ledford introduced a cryogenic jet thermal modulator [14]. This design was soon commercialized by LECO, a major manufacturer of GC×GC instrumentation. This thermal modulator operates by pulsing a jet of cryogenically cooled gas onto a portion of the GC capillary to trap solutes, and then pulsing a jet of hot air or nitrogen to rapidly remobilize the band. If two jets are used, the modulation process occurs in two-stages, where each band is trapped and injected twice, reducing breakthrough and injection bandwidth.[5] The jet-cooled modulator is the mainstay of LECO instruments, and the most popular thermal modulator today by far.

While cryogenic modulators such as the LMCS and the jet modulator are highly effective at trapping solutes across a broad volatility range, the primary drawback is the need for operators of GC×GC instruments to maintain a cryogen supply, imparting a significant operating cost. Further, the technical complexity of thermal modulators is high, incurring high capital costs. In the "early days", the cost of upgrading a 1D-GC system to a GC×GC system was unreasonable for many analysts, since the cost of the thermal modulators alone were nearly as much, if not more, than a 1D-GC instrument! As a result, recent research and development in thermal modulator technology (and modulation technology in general)

---

[3] The first TDM built by Phillips was a section of stainless steel connected to a packed GC column. Then, one cold winter morning, John Phillips noticed that his rearwindow defogger was not working—some of the resistive heating elements broken. To fix this issue, John went to automobile supply paint and bought conductive gold paint, and connected the damaged areas with paint. In one of those rare moments of scientific ingenuity, John realized he could apply the same conductive paint to a GC capillary to make a portion of the column exterior electrically conductive, and modulate the column temperature by computer-controlled resistive heating. The gold-painted TDM was eventually found not robust enough to be commercially viable, but without this development the TDM may not have entered mainstream interest, and perhaps the GC×GC story would be drastically different.

[4] Fun fact: the prototype rotating sweeper was built in the garage of Richard Gaines, an early adopter of GC×GC.

[5] The utility of two-stage thermal modulators was recognized early on by John Phillips. In later designs of his TDM, two sections of the GC capillary were coated in gold paint to enable two-stage modulation. Similarly, the LMCS operates in a two-stage fashion.

has shifted to developing systems that are cryogen-free with low capital cost. Perhaps the most popular of these is the "Thermal Independent Modulator (TIM)", also called the "Solid-State Modulator (SSM)" [15]. This modulator is placed external to the GC oven (hence "thermal independence") and uses thermoelectric cooling to trap solutes and micathermic heating for remobilization (hence "solid-state"). To effect modulation, a section of capillary column is physically moved back-and-forth across the heating and cooling elements. This design is certainly inspired by the LMCS, though in this case the *column* is moved rather than the modulator.[6] Despite thermal independence, a drawback of cryogen-free operation is the inability to trap species that are either highly volatile or non-volatile. Cryogenic thermal modulators typically cannot trap species less than C_4, and the TIM appears to trap species with a volatility range of C_6 to C_{24}. Nonetheless, cryogen-free thermal modulators have been successfully used in a wide rage of application areas [16,17].

In summary, thermal modulation is a highly effective approach and this form of modulation has seen intense research and development since Phillip's original TDM. Jet-cooled cryogenic modulation may very well be the most robust, reliable, and effective thermal modulator, despite the high operating and capital costs. Newer cryogen-free systems eliminate the costly use of cryogens and appear effective at modulating a sufficiently broad volatility range. A key operating principle of all thermal modulators is selective preconcentration of solutes over the carrier gas. The preconcentration effect is what produces intense peaks with a high frequency bandwidth and thus enhances the S/N. Due to preconcentration and remobilization, each injection pulse is Gauss-like. As we will see, this is not the case for all flow modulation approaches.

1.2.1.3 Flow modulators

At the same time that a number of researchers were developing various thermal modulation approaches, Bruckner et al. published a seminal paper in the history of GC×GC using a diaphragm valve to generate high-speed 2D separations [18]. Four ports of a six-port valve were used to periodically divert a portion of the effluent from the ^{1}D column to waste for a brief period of the P_{M}, and then quickly actuate to deliver a pulse of ^{1}D effluent onto ^{2}D. A 500 ms P_{M} was applied, markedly faster than that used in any thermal modulator.

However, while thermal modulator designs transferred 100% of the ^{1}D column effluent onto ^{2}D, the diaphragm valve in a four-port configuration discarded a majority of material to waste, and was unfortunately overlooked by the GC×GC community.[7] While Bruckner's report was the first use of a valve for modulation in GC×GC, it wasn't the first valve modulator used for gas chromatography. In the days of multiplex chromatography, several investigators were exploring a variety of valve-related means for multiplex GC, including fluidic logic gating [19] and early diaphragm valves [20].

In 2000 John Seeley, trained as a physical chemist, saw the benefits of the diaphragm valve system and improved it. Instead of using the valve to divert flow, a majority of the P_{M} was used to collect ^{1}D effluent in a sample loop. Then, the valve was actuated and a fast flow of auxiliary gas was used to inject the sample on the ^{2}D column and the ^{2}D separation was carried out. The duty cycle, or percentage of ^{1}D effluent transferred to the ^{2}D column, is determined by the fraction of the P_{M} that the valve is in the "load" position, assuming the sample loop does not overfill. Therefore, if the valve is collecting sample for 80% of the P_{M}, the duty cycle is 80%! However, this requires that the entire ^{2}D separation is accom-

---

[6] We earlier discussed that a major drawback of the rotating sweeper was the use of moving parts, hence the invention of the jet-cooled modulator. Some are skeptical that moving a capillary back-and-forth is a robust modulation approach, but for now, this modulator is enjoying widespread usage.

[7] Some even questioned if using a diaphragm valve truly generated a "comprehensive" separation. After calm discussion, it was decided that a 2D separation was comprehensive as long as a sufficient number of samples were taken from each primary peak, regardless of the fraction of the primary column effluent transferred to the second dimension.

plished in 20% of the modulation period, requiring a fast ^{2}D gas flow for a high-speed ^{2}D separation. The much greater ^{2}D flow relative to ^{1}D led Seeley to term the technique "differential flow modulation", which eventually led to the term "flow modulation" to be used to describe modulation techniques using one or more valves and fluidic connections to effect modulation. While Seeley's design greatly improved the duty cycle of valve-based modulation, the need for high flow increases theoretical plate height on ^{2}D and raised questions about compatibility with mass spectrometry. Sinha et al. showed that diaphragm valve-based modulation was compatible with mass spectrometry, albeit using a lower duty cycle of 10% (and thus a lower ^{2}D flow rate) [21]. This demonstrated the compromise between maintaining a high duty cycle and compatibility with mass spectrometry detection. Nonetheless, the GC×GC community soon realized that flow modulation was a simple and effective alternative to thermal methods, and several researchers began to develop and study flow modulation techniques as low-cost and simple alternatives to thermal modulation.

At this time, diaphragm valves were not capable of operating at temperatures above 175 °C, restricting their use inside the GC oven to relatively low-volatility samples. In Seeley's earlier report, the diaphragm valve was brought completely out of the oven, but this could cause solutes to condense in the valve. One solution reported by Sinha et al. was to face-mount thr diaphragm valve, such that only the ports were interior to the oven, enabling operation at oven temperatures up to 250 °C [22]. Seeley approached the issue from a different angle, developing flow modulators with 3-way solenoid valves exterior to the oven that controlled the loading and injection of effluent collected in accumulation loops [23]. Seeley's new designs were inspired by the Dean's switch, a device long used for heart-cutting "multidimensional" gas chromatography [24]. These designs eliminated the drawbacks of an internal diaphragm valve, but early modulator configurations were complicated and used multiple internal T-junctions and cross unions creating potential for leaks in the system. Ostensibly on the heels of Seeley, Aviv Amirav entered the GC×GC scene using Seeley's flow modulator design with supersonic molecular beam mass spectrometry detection, compatible with the high ^{2}D flow rates generated in differential flow modulation [25].

In 2006 Seeley reported a new design termed the "simple fluidic modulator" in *American Laboratory* using only two T-junctions internal to the GC, much simpler than his earlier modulators [26]. Amirav soon followed up, interfacing this modulation design with supersonic molecular beam mass spectrometry [27]. Diaphragm valve modulation also saw improvements, including a total-transfer design accomplished by blocking one port of a six-port valve [28], and connecting the ^{1}D column directly to the valve to enable long ^{2}D separation times [29]. The temperature limitations of diaphragm valves was realized to be the decomposition of the O-rings. The introduction of valves with perfluoroelastomer (Kalrez) O-rings now enables diaphragm valve-based modulation up to at least 325 °C [30,31].

Modulators based on the "simple fluidic modulator" and the reverse/fill flush mechanism are effective and robust approaches for flow modulation. Still, some drawbacks remain. First, in order to achieve maximum duty cycle, the size of the accumulation capillary must be optimized according to the ^{1}D column flow, loading time, and modulation period, or large amounts of effluent will be discarded to waste. Second, these modulators do not appear to be compatible with high P_M, a necessity for high peak capacity applications. Finally, there are a large number of fluidic connections, which increases the probability of leaks developing in the system. Therefore, there remained a need to develop a flow modulator that is fluidically simple, does not require manual adjustment for optimization, and can be applied for high-speed, high peak capacity applications. Ideally, this modulator should be compatible with mass spectrometry detection. To address these challenges, my research introduced a new flow modulation principle termed dynamic pressure gradient modulation. The modulator is fluidically simple, has a duty cycle of 100%, and can operate at P_M as short as 50 ms—faster than any other modulator reported.

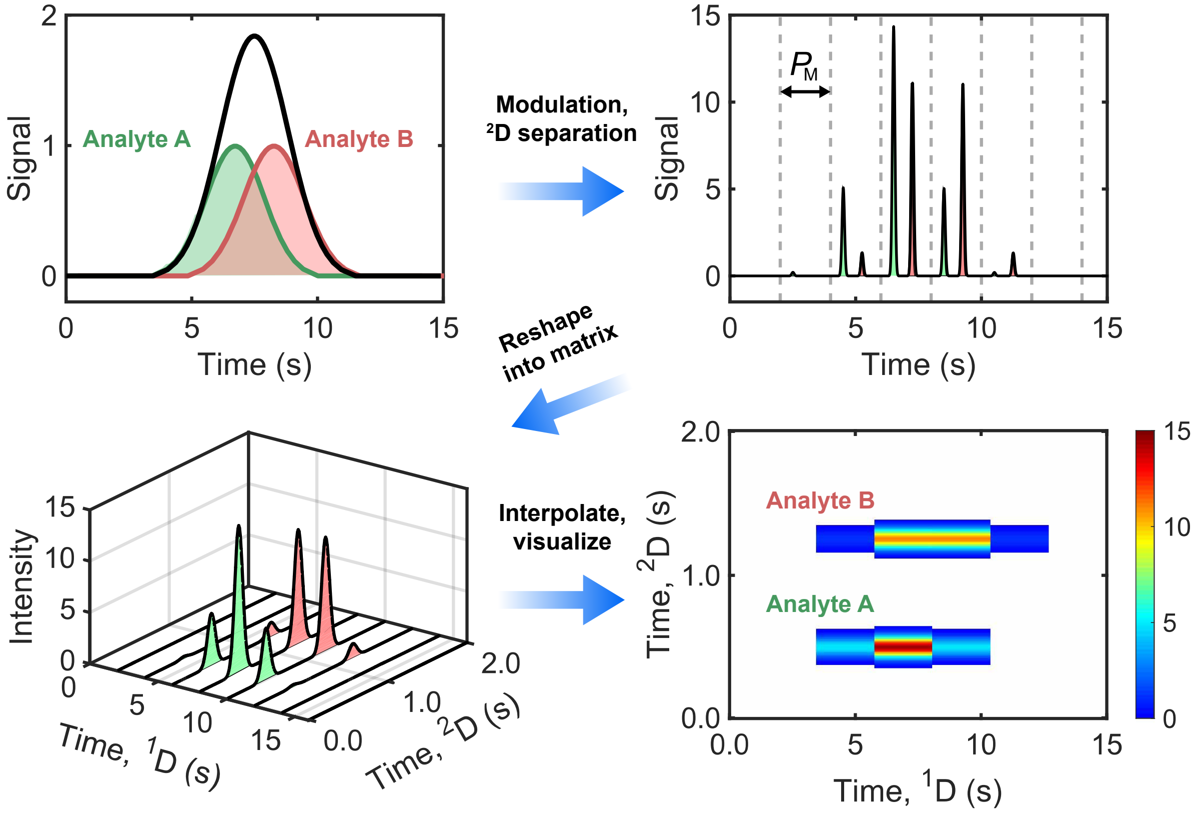


Figure 1.2: Process of constructing a GC×GC chromatogram

The design, optimization, and application of DPGM are discussed in Chapter 2. Then, it's utility as a modulator for comprehensive three-dimensional gas chromatography (GC^3) is shown in Chapter 3.

1.3 Data acquisition and visualization

Before the development of GC×GC and LC×LC, all two-dimensional separations were carried out *in space*. A common example is classical 2D thin layer chromatography (TLC). To perform 2D TLC, the ^{1}D separation is obtained by developing the TLC plate in one direction, rotating the plate 90°, and developing the separation with a different mobile phase. Clearly, this approach is not experimentally feasible using the instrument in Fig. 1.1. Instead, GC×GC separations are carried out *in time* and 2D chromatograms are generated computationally by reshaping the signal obtained (a 1D array) at the end of the ^{2}D column into a matrix. This process is illustrated in Fig. 1.2 for GC×GC with univariate detection (or only considering one channel of a multichannel detector). Let us consider two analytes, colored in green and red, who severely coelute on the ^{1}D column. In this example, the coelution is so severe that the signal at the detector would appear as a single peak. At each P_M, fractions of the ^{1}D effluent are collected and/or focused and reinjected onto the ^{2}D column. The signal detected at the end of the ^{2}D column, shown in Fig. 1.2B, is series of fast ^{2}D chromatograms of length P_M. If the stationary phase of the ^{2}D column has sufficiently different selectivity than the ^{1}D column, the analytes have a high likelihood to separate, as is the case in this example. To transform this signal into a two-dimensional image, the vector of data is reshaped into a matrix $I \times J$, where I is number of data points per P_{M} and J is the total number of modulations during the GC×GC separation [8]. Although GC×GC chromatograms shown in a three-dimensional view, such as in the bottom-left panel of Fig. 1.2 are occasionally shown in the literature, it is much more common to plot the chromatogram as an image where the color of each pixel denotes the signal intensity. An example is shown in the bottom-right panel of Fig. 1.2, where the image is generated by nearest-neighbor interpolation. The color of each pixel is then used to represent signal intensity. Most often, a linear color scale is used, but logarithmic color scales may be useful for visualizing samples that contain a dynamic range of component concentrations. In order to enhance the visibility of low signal components in a sample containing components of high signal, the color scale can also be truncated at some maximum intensity such that all components above the threshold intensity have the same color.

1.4 Figures-of-merit

1.4.1 Peak Capacity

There are several metrics which an analyst may use to quantify how well a particular GC×GC method performs. The most common of these is the peak capacity, n_{c}, which was originally defined by Giddings as a quantity "which approximates the maximum number of peaks to be separated on a given column" [34]. A particularly useful property of comprehensive multidimensional separations is that the overall $n_{\mathrm{c,2D}}$ peak capacity is the product of the peak capacities on each dimension [9]. Therefore, for GC×GC, the 2D peak capacity is given by

$$n_{\mathrm{c,2D}} = {}^{1}n_{\mathrm{c}} \times {}^{2}n_{\mathrm{c}} \tag{1.1}$$

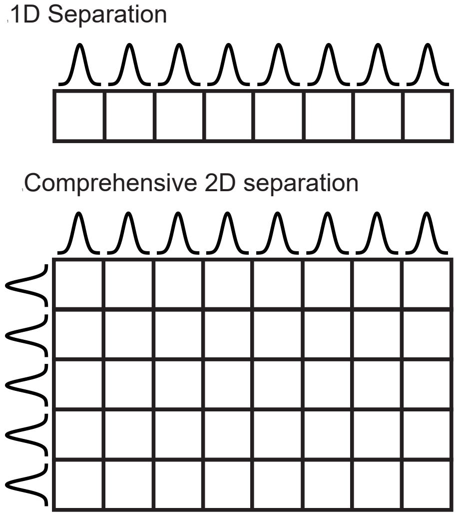


Figure 1.3: Illustration of the multiplicative property of peak capacity for comprehensive 2D separations.

where ${}^{1}n_{\mathrm{c}}$ and ${}^{2}n_{\mathrm{c}}$ are the first- and second- dimension peak capacities, respectively. This property is illustrated in Fig. 1.3 for one- and, two-dimensional separations, respectively. Eq. (1.1) will be referred to as the *ideal* 2D peak capacity; for reasons that we will show, the actual 2D peak capacity is always lower than that given by Eq (1.1).

Next, the general equations for ${}^{1}n_{\mathrm{c}}$ and ${}^{2}n_{\mathrm{c}}$ and their measurement will be derived. Consider a series of segments of time of infinitesimal length $d\mathrm{t}$, over a separation window of length t_{sep},

$$n_{\mathrm{c}} = \int_{0}^{t_{\mathrm{sep}}} \frac{1}{w_{\mathrm{b}}} d\mathrm{t} \tag{1.2}$$

If peak widths do not vary as a function of retention time, which is a reasonable approximation for temp-

---

[8] For GC×G with multichannel detection, each detector channel is reshaped to create a tensor $I \times J \times K$, where K is the number of detector channels (e.g., *m/z*).

[9] This expression is commonly attributed to Giddings. However, this property was actually first described by Karger, Snyder, and Horvath in *An Introduction To Separation Science* (1973), where the quantity was termed the "Fraction capacity", φ [62]. The authors stated that "The maximum fraction capacity of a separation scheme, $\varphi_{\max}$, is given as the product of values for the successive steps in the scheme". More interesting, the authors proposed a hypothetical comprehensive (offline) 3D separation!

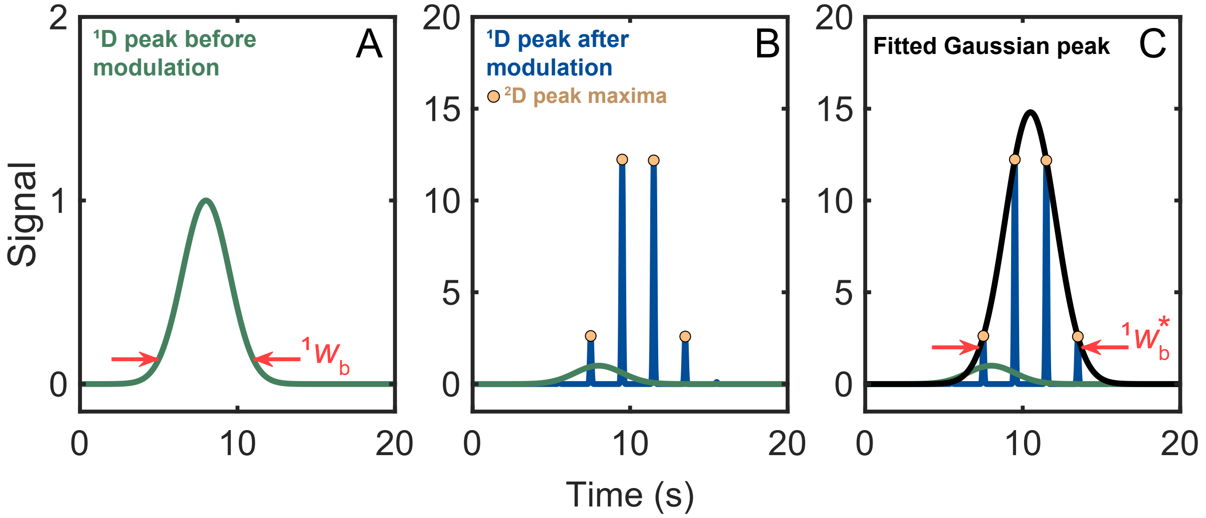


Figure 1.4: Illustration of sampling-induced broadening of a ^{1}D peak during modulation

erature programmed separations, Eq. (1.2) is given by

$$n_{\mathrm{c}} = \frac{t_{\mathrm{sep}}}{w_{\mathrm{b}}} \tag{1.3}$$

Eq. (1.2) is a good approximation for ${}^{1}n_{\mathrm{c}}$ in GC×GC separation because this dimension is usually temperature programmed. We can now explicitly write an equation for the first-dimension peak capacity as

$${}^{1}n_{\mathrm{c}} = \frac{{}^{1}t_{\mathrm{sep}}}{{}^{1}w_{\mathrm{b}}} \tag{1.4}$$

To estimate ${}^{1}n_{\mathrm{c}}$, it is common practice to measure ${}^{1}w_{\mathrm{b}}$ for several well-resolved analytes and average these values,

$${}^{1}n_{\mathrm{c}} = \frac{{}^{1}t_{\mathrm{sep}}}{N_{\mathrm{meas}}} \sum_{i=1}^{N_{\mathrm{meas}}} \frac{1}{{}^{1}w_{\mathrm{i}}} \tag{1.5}$$

where N_{meas} is the number of peaks measured. However, as shown in Fig. 1.2, the points defining the ^{1}D peak are spaced by P_{M}. Commonly, a ^{1}D peak in a GC×GC experiment is defined by only 2-4 points.

The low data density that defines the ^{1}D peak is a result of undersampling and broadens the effective ^{1}D peak width. This process is further illustrated in Fig. 1.4. The ^{1}D peak prior to modulation is shown in Fig. 1.4A and its width is well defined. After modulation and separation on ^{2}D, the ^{2}D peak is defined by the ^{2}D peak maxima (Figure 1.4B). To obtain an estimate of ${}^{1}w_{\mathrm{b}}$, the analyst can fit a Gaussian function to the ^{2}D peak maxima (Fig. 1.4B). However, due to the accumulation and resampling processes that occurs during modulation, the width of the fitted Gaussian function defining the ^{1}D peak is greater than the width of the ^{1}D peak prior to modulation. Therefore, we refer to the width of the ^{1}D peak after modulation, as obtained from fitting a Gaussian function, as the *effective* ^{1}D width, ${}^{1}w_{\mathrm{b}}^{*}$. Then, we can define the *effective* ^{1}D peak capacity as

$${}^{1}n_{\mathrm{c}}^{*} = \frac{{}^{1}t_{\mathrm{sep}}}{{}^{1}w_{\mathrm{b}}^{*}} \tag{1.6}$$

Because typical temperature programming rates (3-20 °C/min) are slow relative to the length of the ^{2}D separations (~500 ms to 4 s), each ^{2}D separation occurs under locally isothermal conditions. In this case, the assumption that ${}^{2}w_{\mathrm{b}}$ is invariant with time is no longer valid and w_{b} cannot be removed from the integrand in Eq (1.2). Fortunately, in fast capillary GC, ${}^{2}w_{\mathrm{b}}$ is generally observed to increase linearly with time,

$${}^{2}w_{\mathrm{b}} = \mathrm{a}\left({}^{2}t\right) +^{2} w_{\mathrm{o}} \tag{1.7}$$

where the slope is primarily governed by C-term broadening, and ${}^2w_\mathrm{o}$ is the width of the injection pulse before chromatographic broadening. Experimentally, an analyst can obtain the slope and intercept by linear regression of measured ${}^2w_\mathrm{b}$ for a sufficient number of analytes spanning ${}^2t_\mathrm{sep}$.

$$
\begin{aligned}
{}^2n_\mathrm{c} &= \int_0^{{}^2t_\mathrm{sep}} \frac{1}{\mathrm{a}\,({}^2t) +^2 w_\mathrm{o}} dt &(1.8)\\
&= \frac{1}{a} \ln\left(\mathrm{a}\left({}^2t\right) +^2 w_\mathrm{o}\right) &(1.9)
\end{aligned}
$$

To provide a very approximate estimation of ${}^2n_\mathrm{c}$, an analyst may also average the widths of ^{2}D peaks spanning ${}^2t_\mathrm{sep}$, akin to Eq (1.2), as

$$
{}^2n_\mathrm{c} = \frac{{}^2t_\mathrm{sep}}{N_\mathrm{meas}} \sum_{i=1}^{N_\mathrm{meas}} \frac{2}{{}^1w_\mathrm{i}} \tag{1.10}
$$

The accuracy of this approximation is best when the slope, a, in Eq (1.7) is small, which is true for (relatively) slower ^{2}D separations and when the influence of external broadening is high and does not vary with ${}^2t$. Considering Eq. (1.6) and (1.10), we can write the *effective* 2D peak capacity as

$$
n^*_\mathrm{c,2D} = {}^1n^*_\mathrm{c} \times {}^2n_\mathrm{c} \tag{1.11}
$$

where the superscript asterisk indicates the effective peak capacity.

1.4.2 Signal-to-noise enhancement

Another benefit of GC×GC is the potential for increasing the S/N relative to 1D-GC. As a result of modulation, the height of the tallest ^{2}D peak is substantially greater than the height of the original ^{1}D peak, were it detected at the exit of the ^{1}D column [10]. This results in a substantial S/N enhancement, and enables LODs much lower than that possible with an equivalent 1D-GC method. Both thermal modulation and flow modulation provide signal enhancement, albeit by different mechanisms. During thermal modulation, sample components are selectively preconcentrated over the carrier gas, resulting in signal enhancement due to preconcentration. In flow modulation, analytes are not preconcentrated; however, most flow modulation methods use a substantially higher flow rate on ^{2}D than on ^{1}D. As a result, the mass flux to the detector is increased and the signal is enhanced for any mass-sensitive detector, such as FID and electron ionization (EI)-MS[11]. Notably, it has been claimed that flow modulation does not provide a signal or S/N enhancement [35], but subsequent reports [36] and the results presented in this dissertation will demonstrate that this assumption is incorrect.

We will define the ratio of the tallest ^{2}D peak in a GC×GC separation to the height of the ^{1}D peak from which it originated as the *Signal Enhancement* (*SE*),

$$
SE = \frac{{}^1h}{{}^2h} \tag{1.12}
$$

where ${}^2h$ is the height of the tallest ${}^2D$ peak, and ${}^1h$ is the height of the corresponding 1D peak prior to modulation. If the magnitude of the noise is the same in a GC×GC experiment as in a 1D-GC experiment, one may conclude that the S/N enhancement scales linearly with the SE. However, this assumption overestimates the S/N enhancement if data are acquired at the same acquisition frequency. Let us assume that GC×GC data are obtained at a sufficient data acquisition frequency for ^{2}D peak sampling ($\geq$ 10 points/peak). Assuming a properly designed GC×GC experiment with three or more modulations per ^{2}D peak, many more points are acquired across the ^{1}D peak than necessary for accurate reconstruction of peak area and width. Therefore, the ^{1}D peak in an equivalent 1D-GC method could be collected at a lower data acquisition frequency, resulting in a higher S/N. Alternatively, if the 1D-GC peak is measured at the same data acquisition frequency as the GC×GC separation, the 1D-GC peak can be re-

---

[10] In rare cases where the modulator duty cycle is extremely small (< 10%), which occurs in some forms of flow diversion modulation, the height of the tallest ^{2}D peak may be lower than the ^{1}D peak.

[11] Some GC detectors, like thermal conductivity detection (TCD), are concentration sensitive; therefore, flow modulation would not provide signal enhancement (and by extension, S/N enhancement).

sampled (i.e., averaged) to increase the S/N without affecting peak shape or area. As a result, the true S/N enhancement is necessarily lower than the *SE*.

To measure the *SE* in a GC×GC experiment, an analyst first measures ${}^2h$ for a given analyte. To measure ${}^1h$, the analyst runs the GC×GC experiment with the modulator disabled, which provides a a 1D-GC separation. The assumption under this measurement is that the additional broadening of the ^{1}D peak during separation on the ^{2}D column is negligible.[12] This is a reasonable estimation for the column lengths used in a typical GC×GC experiment. Further, the analyst must select an analyte that is well-resolved on 1D ($R_\mathrm{s} \geq 1$) such that coelution does not contribute to the measurement of ${}^1h$.

In Chapter 3, we will derive and experimentally validate expressions for *SE* in comprehensive three-dimensional gas chromatography. To provide a foundation, we'll first derive expressions for *SE* in a GC×GC experiment based upon the fractional area sampled during each P_M. Then these expressions will be extended to model GC^3 separations. Finally, we will validate these expressions using both experimental and simulated GC^3 data.

1.4.3 Comprehensive three-dimensional gas chromatography

Despite the peak capacity provided by GC×GC, peak overlap remains likely for extremely complex mixtures. An intriguing technique to address this issue is comprehensive three-dimensional gas chromatography (GC^3). A GC^3 instrument is constructed by assembling three columns of sufficiently unique selectivity in series using two modulators (Fig. 1.5).

A few considerations in the design of a GC^3 instrument are noted. First, the ^{3}D column should have sufficiently unique selectivity relative to the ^{1}D and ^{2}D columns. One option that will be explored in this dissertation is the use of an ionic liquid stationary phase on ^{3}D. Second, recall that ^{2}D peaks in GC×GC are on the order of hundreds of milliseconds. Therefore, in order to adequately sample ^{2}D peaks in GC^3, the second modulator must be capable of operating at a P_M much faster than that used in conventional GC×GC. For ^{2}D peak widths on the order of 200-600 ms, a P_M of 50 to 250 ms would be necessary to meet the minimum sampling requirements. This further motivates the development of flow modulators that can operate at extremely fast P_M.

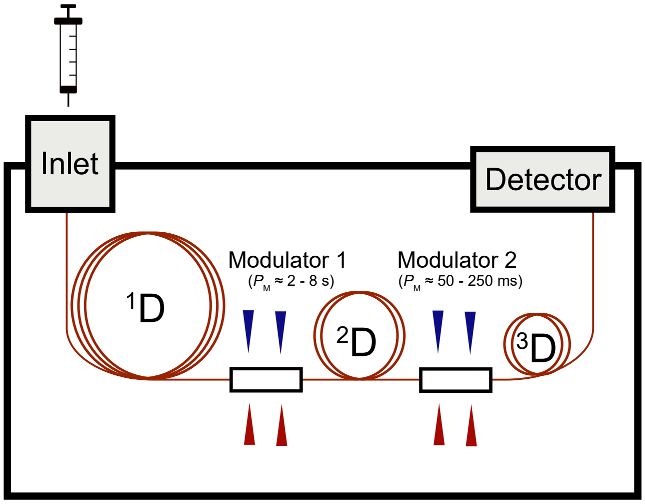


Figure 1.5: Schematic diagram of a GC^3 instrument consisting of three separation columns linked by two modulators.

The utility of GC^3 arises from the aforementioned multiplicative property of peak capacities. Thus, the ideal 3D peak capacity is given by

$$n_{\mathrm{c,3D}} = {}^1n_\mathrm{c} \times {}^2n_\mathrm{c} \times {}^2n_\mathrm{c} \tag{1.13}$$

as shown graphically in Fig. 1.6. Recall that in a typical GC×GC experiment, the ^{2}D peaks are sampled >10 times per peak such that there is no undersampling induced broadening of the ^{2}D peaks. However, in GC^3, the ^{2}D peaks are sampled by a secondary modulator, typically acquiring 3-4 samples per peak. This results in an undersampling contribution to ${}^2n_\mathrm{c}$. Therefore, the *effective* 3D peak capacity is given by

$$n_{\mathrm{c,3D}} = {}^1n^*_\mathrm{c} \times {}^2n^*_\mathrm{c} \times {}^2n_\mathrm{c} \tag{1.14}$$

where ${}^1n^*_\mathrm{c}$ and ${}^2n^*_\mathrm{c}$ denote that there is undersampling-induced broadening to both ^{1}D and

---

[12] Alternatively, the analyst could connect the 1D column directly to the detector.

^{2}D.

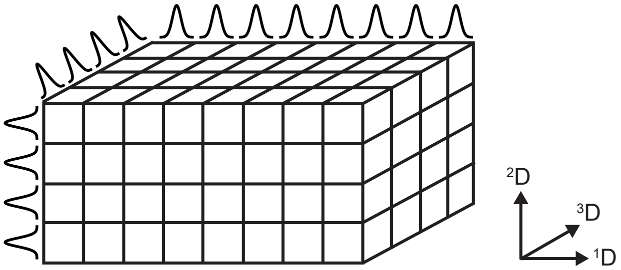


Figure 1.6: Illustration of the peak capacity of a three-dimensional separation with each peak volume represented by a cube.

Despite this additional broadening effect, the effective peak capacity is a well-designed GC^3 instrument is substantially higher than that in a GC×GC instrument, as will be demonstrated.

1.5 Data analysis methods

This section will walk the reader through the decisions an analyst must make in order to select an analytical method that is most appropriate for their analysis goal. The flow chart in Fig. 1.7 illustrates this process. A brief discussion and history of specific methods for different analysis types will be provided for context. Finally, we will explain in detail the data analysis method most pertinent to this thesis—tile-based *F*-ratio analysis.

The first question an analyst must ask themselves is whether or not the analytes of interest are known. As an example, a laboratory may want to determine the concentration of ten pesticides in human plasma. In this case, the identity of the analytes are known. This is termed a targeted analysis.

1.5.1 Targeted analysis

1.5.1.1 Traditional quantification

If a method can be developed to resolve the analytes (either chromatographically or using mass spectrometry) a "traditional" targeted analysis can be accomplished with well-established quantitative methods such as external calibration or the standard addition method. Thousands of analytical labs around the world run "traditional" targeted analyses with the goal of determining the concentrations of one or more compounds of known identity. Although GC×GC is certainly capable of performing traditional targeted quantitative analysis, and there are many methods reported in the literature, most traditional targeted analyses can be accomplished with less-expensive GC and GC-MS methods.

1.5.1.2 Chemometric resolution

If the analytes cannot be resolved, chemometric deconvolution methods can be used to obtain pure analyte peak profiles. Because these methods require the analyst to select a region of the chromatogram containing coeluting analytes, or know the identify and/or number of coeluting analytes, these methods can be categorized as targeted chemometric deconvolution techniques. The use of targeted chemometric deconvolution was fist applied to GC×GC in 1998 using the generalized rank annihilation method, GRAM [37].

Generalized rank annihilation method, GRAM [37]. GRAM natively operates on bilinear data, and had been previously applied to LC-UV and GC-MS data. In the context of GC×GC, a bilinear data structure requires that ^{2}D peaks do not shift appreciably along from one modulation to another and that ^{2}D peak shape is consistent. In this seminal report, it was demonstrated that the ^{2}D of a GC×GC instrument could be treated as a pseudo-"multichannel detector", which produced bilinear data, and thus the GRAM technique could be applied. Briefly, to perform GRAM, an analyst runs standards containing only the analytes of interest in addition to the sample data containing the analytes and interfering compounds. A singular value decomposition (SVD) is performed on the addition matrix, where these two runs are summed together. Then, the decomposed matrix is truncated to the chemical rank, which is the expected number of components in the sample. The use of both a standard and sample matrix eliminates "rotational ambiguity" in the solution. In short, GRAM is a direct eigenvalue decomposition method, and the amount of error in the solution can be readily ascertained.

The application of GRAM to GC×GC data demonstrated that bilinear data structure was advantageous in mathematically resolving analytes vs. 1D-GC data. Soon after, GC×GC with multivariate detection (e.g., GC×GC-TOFMS) became comm-

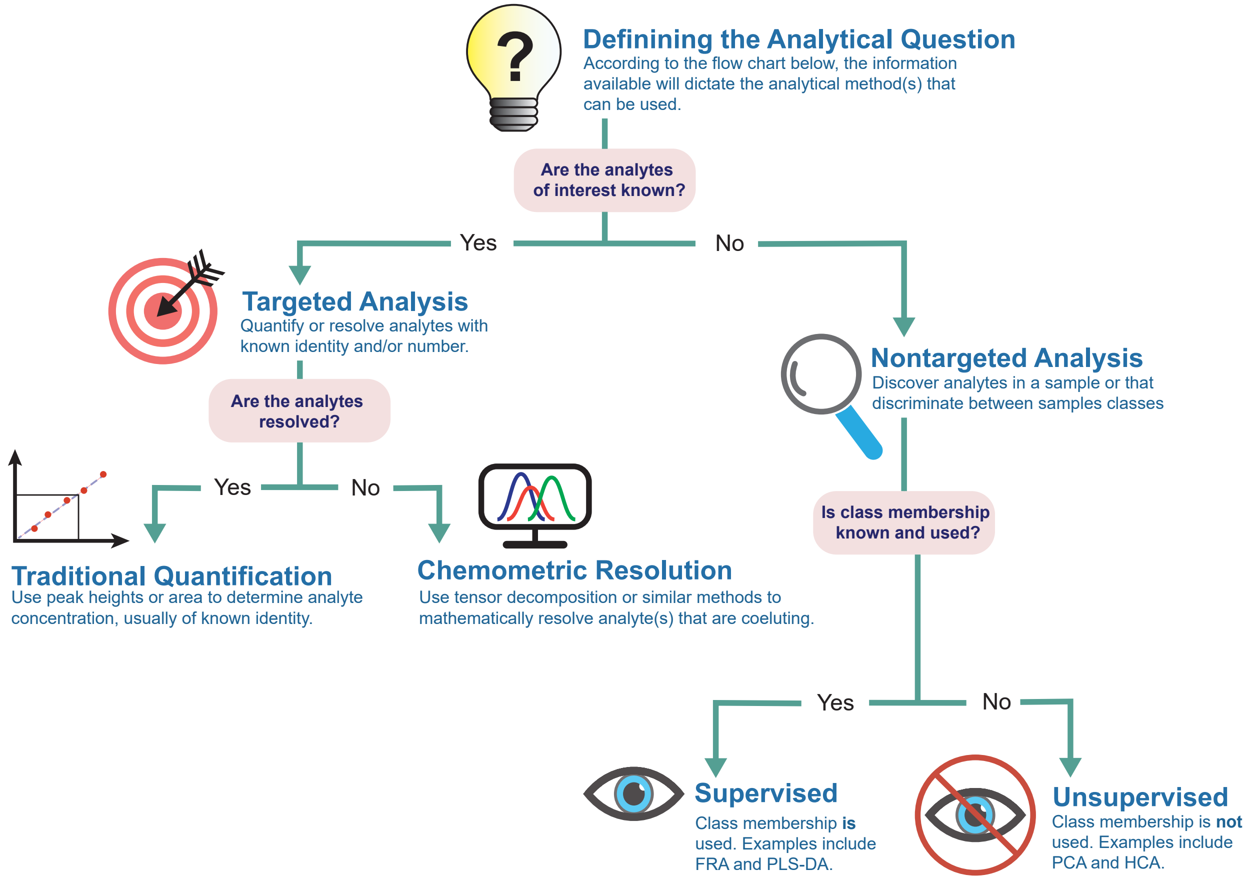


Figure 1.7: Flow chart of analytical methods depending upon the information, goals, and questions of the study.

ercially available and increasingly used. The addition of a third data dimension is even more advantageous—given appropriately designed experiments, GC×GC-MS generates trilinear data for each run. In the context of GC×GC-MS, a trilinear data structure further requires that the mass spectrum is consistent across a peak (i.e., does not "skew" [38]).

However, the use of an additional standard run in GRAM was considered burdensome, and so-called "iterative" tensor decomposition methods have since become the most widely used. The most common examples are multivariate curve resolution-alternative least squares (MCR-ALS) [39,40], which operates on bilinear data, and parallel factor analysis-alternative least squares (PARAFAC-ALS, or PARAFAC) [41,42], which operates on trilinear data. These methods use an initial "guess" of the pure chromatographic and spectral profiles, and use ALS to iteratively fit the model until convergence criteria are reached. Because the solution is obtained by iterative fitting, it is possible for the model to converge to a local minima, causing an erroneous solution. MCR-ALS also suffers from rotational ambiguity, where numerous mathematically sound solutions can be obtained. To some extent, this issue can be mitigated by applying constraints to the model, such as non-negativity, unimodality, and closure. Non-negativity is the most useful for GC×GC and GC×GC-MS, as chromatographic peaks and mass spectra are inherently non-negative. In principle, the issue of rotational ambiguity is eliminated with PARAFAC, but some variation in model results can occur if there is deviation from trilinearity [43].

1.5.2 Non-targeted analysis

In many studies, the identity and number of analytes of interest is not known. As an example, an analyst may seek to discover biomarkers indicative of a certain disease. These methods are termed non-targeted analyses (NTA), more broadly defined as studies with the goal to discover analytes that characterize a sample(s) with little or no *a priori* regarding analyte identity. Depending upon the information desired, and the capabilities of the analytical instruments and data analysis tools, NTA may not only include analyte discovery, but also their identification and quantification. NTA can be formally divide into unsupervised and supervised methods.[13] In unsupervised methods, samples are not assigned class membership (e.g., "infected subjects" vs. "controls"). Sample classes may not be known, may not be used, or may not exist.

1.5.2.1 Unsupervised analysis

The most fundamental and important of the unsupervised methods is principal component analysis (PCA) [44]. Briefly, PCA transforms high-dimensionality data into a lower dimensionality coordinate system where the majority of the variance in the data exists. Thus, PCA enables the analyst to visualize trends and make observations from large, multivariate data sets. One of the most useful outputs of PCA is the scores plot, which shows how each sample projects onto lower dimensionality variables called scores. Given proper preprocessing, samples that cluster together on a scores plot can be inferred to be chemically similar. If desired, an analyst could then classify different samples using other downstream chemometric tools, which may enable later use of a supervised analysis tool. PCA can also be used to determine if there are batch effects or other systematic error in their data; for example, if a series of quality control runs cluster close together, this provides the analyst with confidence that their data is reproducible. Another key output of PCA is the loadings plot, which shows how strongly each variable influenced a given principal component. If raw GC×GC are input into PCA, the loadings plot (upon reshaping) takes the appearance of a GC×GC chromatogram. The most intense peaks in a loadings plot are those that drive separation in the PC scores and can be further investigated as putative marker compounds [45,46]. Note that PCA operates on matrices. If multiple GC×GC chromatograms are analyzed in raw form, they must be unfolded and stacked row-wise to form a single matrix. Although GC×GC-MS chromatograms can be unfolded and stacked row-wise, this generally results in a matrix too large to perform PCA, so it is more common to apply PCA to only one *m/z* of a series of GC×GC-MS chromatograms [47]. Another approach is to apply feature selection or peak detection and only input peak areas into the PCA model, which reduces computational load. These issues will be discussed in more detail subsequently.

Another popular unsupervised method is hierarchical cluster analysis (HCA). The different variables in a dataset are clustered according with predominant ordering from top to bottom. The clustering accomplished by HCA is visualized via a dendrogram, a tree-like diagram which shows the multilevel hierarchy of the data. Several variations of HCA exist, depending on if the classification is built "bottom-up" or "top-down" and on the distance metric used to split and merge data.

1.5.2.2 Supervised analysis

Often, based upon the experimental design, samples can be organized by class membership. Examples can include disease-state vs. control samples in a bioanalytical study, different sources of coffee beans in a foodomics study, or different soil sample collection sites in an environmental study. When class

---

[13] In the lingo of machine learning, the same distinction is used with the terminology supervised and unsupervised *learning*.

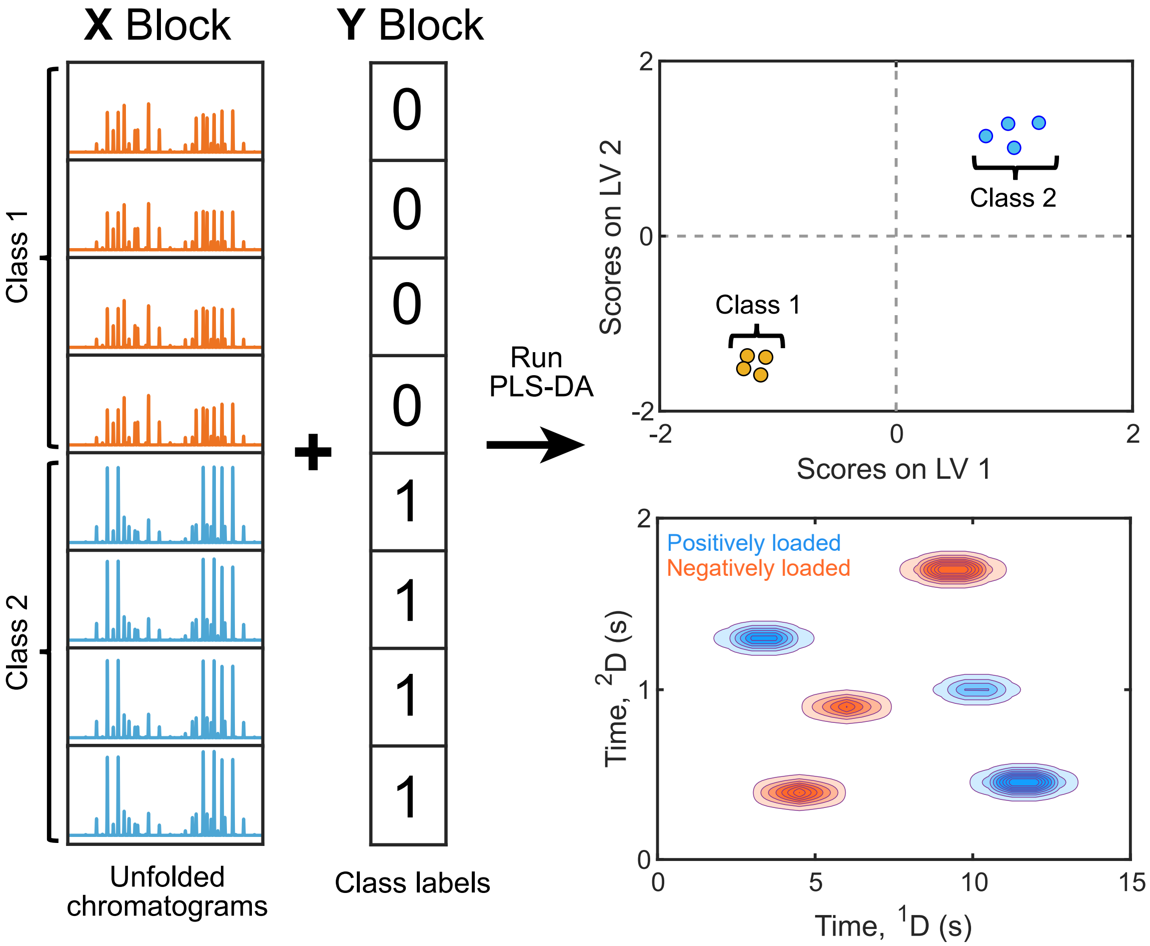


Figure 1.8: Illustration of performing PLS-DA.

membership exists and is used, powerful metrics can be used to discovery analytes which discriminate between different classes. Two of the most widely used supervised NTA methods for GC×GC data are partial least squares-discriminant analysis (PLS-DA) and *F*-ratio analysis (FRA). The optimization and application of (FRA) constituted a major part of my research, and will be discussed in detail in the next section. PLS-DA is a regression method which correlates GC×GC data in a matrix to an array containing categorical variables denoting class membership (e.g., 0 and 1 for a two-class model). The regression is performed such that the direction of the loading vectors, which are analogous to PCA, maximizes the covariance between the data matrix and the class labels. An illustration of PLS-DA regression is shown in Fig. 1.8.

A common method of data presentation is a PLS-DA scores plot, which has a similar appearance to the analogous PCA scores plot [14]. Given proper model building (i.e., no overfitting), clustering in the PLS-DA scores plot can enable the analyst to see the rel-

---

[14] In many instances, the overall appearance of a PLS-DA scores plot may suggest the existence better classification model than the underlying data should generate. Such model overfitting is particularly exacerbated for datasets containing many more variables than samples. In these instances, variables which do not correspond to meaningful chemical difference may be excellent sample classifiers solely due to random chance [63].

ative variance of each class. PLS-DA can be further used to classify new samples of unknown membership. In addition to classification, PLS-DA can be used to perform feature selection; if true chemical variance is driving class membership, the loadings plots generated by PLS-DA can be mapped back to the locations of class-distinguishing analytes. This information is contained in the loadings matrices, each of which can be refolded back to produce a plot visually similar to a GC×GC chromatogram. Regions of high intensity in the loading matrix may correspond to regions in the original GC×GC data that contain class-distinguishing analytes [48] [15]. Similar methodology can be used to identify analytes in PCA scores plots; however, PLS-DA leverages class membership information to better identify variables which distinguishes amongst sample classes. Finally, several statistical metrics can be used to perform feature selection using the PLS-DA outputs, such as Variable Importance in Projection (VIP). VIP can be computed for each variable using its loadings weight. As a simple rule of thumb, variables with VIP scores greater than one are retained as potential features [49].

Although PLS-DA is a powerful tool for classification and feature selection, chromatograms need to be well-aligned such that true chemical differences are identified as features rather than spurious results due to retention-time shifts. Additionally, model building in PLS-DA is computationally expensive and non-intuitive for the non-expert user. An increasingly useful supervised analysis tool, which is a key focus of my research, is *F*-ratio analysis (FRA).

1.5.2.3 F-ratio analysis

FRA is a multiple-hypothesis testing method based upon the applying one-way analysis of variance (ANOVA) to variables in GC×GC data. Developed by Ronald Fisher in 1921, ANOVA is used to determine the significance of differences between classes. Specifically, one-way ANOVA compares the variance *between* classes of samples to that *within* classes of samples, also referred to as treatment and error variance, respectively. The between class (BC) variance gives us information on how the mean of each class varies from the grand mean, defined in Eq. 1.14,

$$\sigma^2_{\mathrm{bc}} = \frac{1}{k-1} \sum \left(\overline{x}_{\mathrm{i}} - \overline{x}\right)^2 n_{\mathrm{i}} \qquad (1.15)$$

where k is the number of classes, n_{i} is the number of measurements in the i^{th} class, $\overline{x}_{\mathrm{i}}$ is the mean of the i^{th} class, and $\overline{x}$ is the grand mean. The within class variance indicates how much each measurement varies from the class mean, via Eq. (1.15),

$$\sigma^2_{\mathrm{wc}} = \frac{1}{N-1} \sum \sum \left(x_{\mathrm{ij}} - \overline{x}_{\mathrm{i}}\right)^2 \qquad (1.16)$$

where N is the total number of measurements, and x_{ij} is the j^{th} measurement of the i^{th} class. The *F*-ratio is then obtained by taking the ratio of these two quantities.

$$F - \mathrm{ratio} = \frac{\sigma^2_{\mathrm{bc}}}{\sigma^2_{\mathrm{wc}}} \qquad (1.17)$$

To perform FRA, software calculates *F*-ratios on variables from GC×GC data of two or more classes of data. Variables which have a high *F*-ratio are termed hits, and the output of an FRA is a "hit list", a table which rank orders all hits by their *F*-ratio, where hits with the highest *F*-ratios are most likely to correspond to class-distinguishing analytes. The analyst can then manually inspect the hit list, top-down, assigning tentative analyte identities to each hits based on their mass spectrum match value (MV) and ^{1}D and ^{2}D retention time information.

FRA can be performed on the pixel-level, the peak table level, or on "tile"-based level. In pixel-level analysis, *F*-ratios are calculated on every data point in a chromatogram. Pixel-level FRA was first reported in 2002 [50], and applied to discriminate between different blends of jet fuel by GC×GC-FID. Later, pixel-level FRA was further developed to GC×GC-TOFMS data [51,52]. Soon, LECO released commercial software for performing FRA as

---

[15] In order to verify that regions that are highly loaded correspond to meaningful chemical differences, the concentration change should be independently tested for statistical significance; for instance, by using a *t*-test at the desired confidence interval.

the Statistical Compare add-on to ChromaTOF, implemented on a peak-table level. At the peak table-level, an algorithm first finds all peaks, tentatively identifies them based upon MV and retention time information, and then calculates F-ratios on the integrated areas of matched analytes. Numerous studies have used Statistical Compare to performe peak table-based FRA [53–55]. However, there are shortcomings to both pixel-level and peak table-level FRA. At the pixel-level, FRA is highly susceptible to retention time variation, where "hits" appear due the variance caused by run-to-run shifting rather than true chemical differences between classes [56]. To some extent, this can be mitigated by applying alignment; however, alignment is computationally expensive in general, and alignment of GC×GC data specifically is less robust than alignment of 1D-GC data due to the poor data density on ^{1}D. Peak table FRA is also computationally expensive, as analytes are detected, deconvoluted, and tentatively identified on each sample prior to peak matching, integration, and the F-ratio calculation. To address the shortcomings associated with pixel-level and peak table-level FRA, tile-based FRA was introduced [56,57]. This method is illustrated schematically in Fig. 1.9. In this approach, data are binned into sections called "tiles", where the tile size is user-selected to be larger than the size of a typical 2D peak and its observed shifting. In this manner, the deleterious effects of retention time shifting are mitigated concurrent with a S/N enhancement. However, binning the data only once necessarily causes several peaks to be split across bins, impairing discoverability. Therefore, the tile-based approach bins chromatograms according to four spatially offset grids. This ensures that each peak is optimally captured in at least one grid (Fig. 1.9). As a result of peaks being split across bins and the use of four grids, any given analyte will necessarily appear more than once in the "initial" hit list. These entries are termed redundant hits. To remove redundant hits and keep only the optimally captured hit, a "pinning and clustering" routine is used. The algorithm locates the 2D maxima (i.e., "pins") for all the hits within a 2D box (the "cluster window") and keeps only the pin with the highest F-ratio. Then, the tile is recentered about the best pin and the F-ratio is recalculated. Although FRA bins data, after discovering hits the analyst can go back to the original pixel-level data contained within a tile, if desired, to investigate the high-fidelity data.

FRA on complex GC×GC datasets can generate hundreds of hits. This raises the following question—when does the analyst stop inspecting the hit list? That is, when does the analyst decide any hits past a certain F-ratio are not statistically significant? While the F-ratio does serve as an indication of the differences between classes it does not on its own imply significance. When only one variable is being tested for significance (i.e., a single F-test) one can consult widely available statistic tables to find p-values corresponding to different F-ratios based upon the number of samples to determine if the null hypothesis should be rejected or accepted. However, there are two problems to using this approach in FRA. First, when the F-test is performed on a large number of variables, as in FRA, the chance of spuriously rejecting the null hypothesis when it is true becomes large. For instance, selection of a p-value of 0.05 (95%) confidence is common. Then, if 100 F-tests are simultaneously performed, selection of $p = 0.05$ implies that 5 test will be false positives, which can be extremely problematic, depending on the study. This is termed the multiple hypothesis problem and the number of false positives expected to occur is termed the false discovery rate (FDR). To correct the multiple hypothesis problem, several corrections can be made to the p-values obtained, effectively reducing the p-value for the experimental F-distribution to reduce the FDR [58]. However, this can result in potentially significant hits being missed. The second issue with applying statistical tables in FRA is the assumption of a parametric test—in the context of an F-test, assumptions include that data are normally distributed[16] and that data are inde-

---

[16] It should be noted that the F-test is somewhat robust to non-normality.

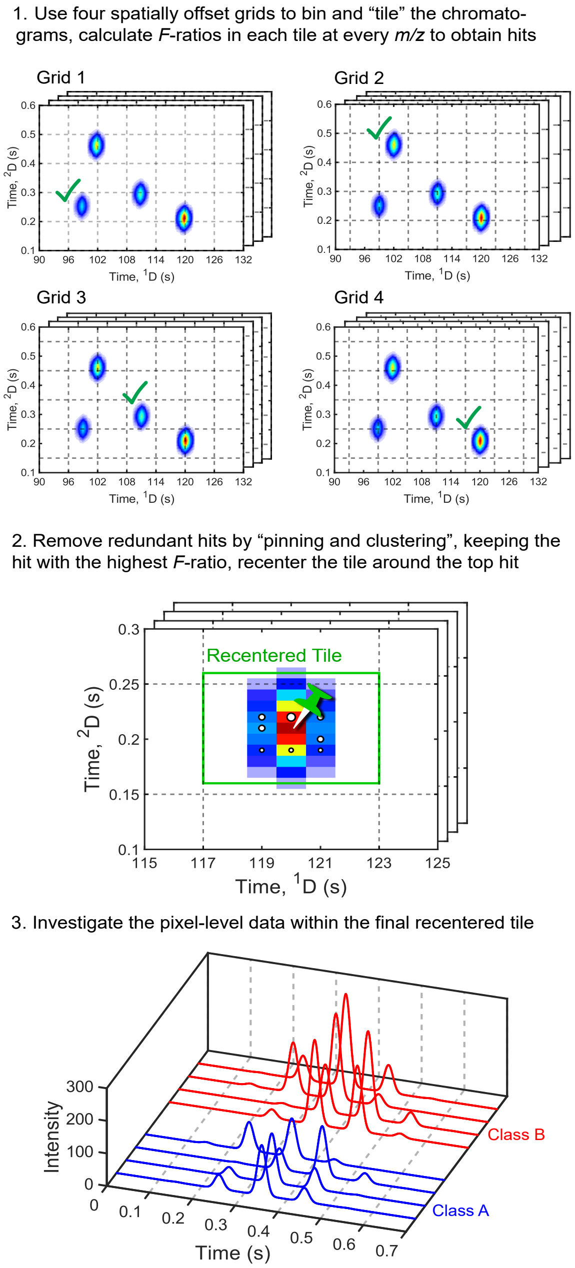


Figure 1.9: Illustration of performing FRA on a tile-based level.

pendent. In many biological studies, it is observed that data are not normally distributed. To address these two issues, an F-ratio cutoff can be obtained by performing a permutation-based test. In the context of FRA, this is termed a null distribution analysis [57]. Briefly, class labels are permuted and FRA is performed on each new "null class". From the resulting null F-distribution, confidence intervals can be derived to determine a suitable F-ratio cutoff in a manner that is both non-parametric and addresses the multiple hypothesis problem.

1.6 Overview of the following chapters

1.6.1 Dynamic pressure gradient modulation for comprehensive two-dimensional gas chromatography

The discovery, preliminary investigation, and demonstration of a novel form of differential flow modulation for comprehensive two-dimensional (2D) gas chromatography (GC×GC) is described. Commercially available components are used to apply a flow of carrier gas with a suitable applied auxiliary gas pressure (P_{aux}) to a T-junction joining the first (^{1}D) and second (^{2}D) dimension columns. The ^{1}D eluate is confined at the T-junction, and introduced for ^{2}D separation with a cyclic rhythm, dependent upon the relationship of the modulation period (P_{M}) to the pulse width (p_{w}), where p_{w} is defined as the time interval when the auxiliary gas flow at the T-junction is off. This flow modulation technique is referred to as "dynamic pressure gradient modulation" (DPGM) since a pressure gradient oscillates with the P_{M} along the ^{1}D and ^{2}D column ensemble providing temporary stop-flow conditions and fast ^{2}D flow rates, resulting in 100% duty cycle and full modulation. A 90-component test mixture was used to evaluate the technique with a p_{w} of 60 ms and a P_{M} of 750 ms. The resulting peaks were narrow, with $^2w_{\mathrm{b}}$ ranging from about 20-180 ms. With an average $^1w_{\mathrm{b}}$ of 3 s and a $^2n_{\mathrm{c}}$ of 10, a 2D peak capacity, $n_{\mathrm{c,2D}}$, for the 25 min separation was 5000. The signal enhancement (SE) is reported, defined as the peak height of the highest modulated ^{2}D peak divided by the unmodulated ^{1}D peak height ($SE = {}^2h/{}^1h$). The SE ranged from about 7-87, depending on the $^1w_{\mathrm{b}}$ and $^2w_{\mathrm{b}}$ for a given analyte. A diesel sample was analyzed to demonstrate performance with a complex sample. Based upon the average $^1w_{\mathrm{b}}$ of 5 s and an average $^2w_{\mathrm{b}}$ of 168 ms, a $n_{\mathrm{c,2D}}$ of 8640 was obtained for the 60 min diesel separation. Finally, the modulation principle was investigated as a function of P_{M}, p_{w}, and the volumetric flow rates, ${}^{1}F$ and ${}^{2}F$. The measured ${}^{2}w_{\mathrm{b}}$ correlate well with the theoretical ^{2}D injected width, given by ${}^{2}w_{\mathrm{inj}} = \left({}^{1}F/{}^{2}F\right) P_{\mathrm{M}}$, where the relevant ${}^{1}F$is the ^{1}D flow rate when no pressure is applied (during the p_{w} interval).

1.6.2 Total-transfer comprehensive three-dimensional gas chromatography with time-of-flight mass spectrometry

Although GC×GC is a powerful technique, component overlap remains likely for complex samples. An intriguing route to address this challenge is to utilize the additional peak capacity and chemical selectivity provided by comprehensive three-dimensional (3D) gas chromatography (GC^{3}), especially with time-of-flight mass spectrometry detection (GC^{3}-TOFMS). However, the GC^{3}-TOFMS instrumentation reported to date has employed one or both modulators with a duty cycle $< 100\%$, making the potential gain in detection sensitivity over GC×GC modest, or perhaps even worse. Herein, I describe instrumentation for GC^{3}-TOFMS in which both modulators provide total-transfer (100% duty cycle). Specifically, the instrument is based on the facile modification of a commercial thermally modulated comprehensive GC×GC-TOFMS platform for modulation from the ^{1}D column to the ^{2}D column, with recently described DPGM as the second modulator from the ^{2}D column to the ^{3}D column, which is a total-transfer flow modulation technique. Area measurements of ^{1}D peaks are compared to the sum of ^{3}D peak areas to validate the assumption that total-transfer from ^{1}D to ^{3}D is accomplished. Additionally, peak heights were amplified as high as a factor of 177 ($\bar{x} = 130$, $s = 47$) via comparison of ^{1}D peak heights to the maximum ^{3}D peak heights. Column selection is explored, with emphasis on the resulting peak width-at-base on each dimension and usage of 3D separation space as evaluation metrics. Using a nonpolar × polar × ionic liquid column combination, an effective peak capacity which considers modulation-induced broadening as high as 32,300 for select analytes was achieved ($\bar{x} = 19{,}900$, $s = 10{,}700$). The analytical benefits of employing three selective phases, mass spectrometry detection, and total-transfer modulation are explored with separations of a metabolomics-type sample, i.e., derivatized porcine serum, and a jet fuel spiked with various sulfur-containing compounds.

1.6.3 Determination of the signal-to-noise ratio enhancement in comprehensive three-dimensional gas chromatography

The extent to which comprehensive three-dimensional (3D) gas chromatography (GC^{3}) provides a signal enhancement (SE) and signal-to-noise ratio enhancement (S/N_{Rel}) relative to one-dimensional (1D)-GC has been explored. In the context of GC^{3}, the SE is defined as the ratio of the tallest ^{3}D peak height from the GC^{3} separation to the 1D peak height from the unmodulated 1D-GC separation. A model is proposed which allows the analyst to predict the theoretically attainable signal enhancement (SE_{T}) based upon peak width and sampling density inputs. The model is validated via comparison of the SE_{T} to the experimentally measured signal enhancement (SE_{M}) obtained using total-transfer GC^{3} (100% duty-cycle for both modulators) with time-of-flight mass spectrometry detection (TT-GC^{3}-TOFMS). Two experimental conditions were studied using the same GC^{3} column set, differing principally in the modulation period from the ^{1}D to ^{2}D columns: 4 s versus 8 s. Under the first set of conditions, the average SE_{M} was 97 (± 22), in excellent agreement with the SE_{T} of 97 (± 18). The second set of conditions improved the average SE_{M} to 181 (± 27), also in agreement with the average SE_{T} of 176 (± 26). The average S/N_{Rel} following correction for the mass spectrum acquisition frequency was 38.8 (± 11.2) and 59.0 (± 27.2), respectively, for the two sets of conditions. The enhancement in S/N is largely attributed to moving the signal to a higher frequency domain where the impact of "low frequency" noise is less detrimental. The findings here provide strong evidence that GC^{3} separations can provide enhanced detectability relative to 1D-GC and GC×GC separations.

1.6.4 Profiling olefins in gasoline by bromination using GC×GC-TOFMS followed by discovery-based comparative analysis

An analytical workflow for the analysis of olefins in gasoline that combines selective bromination and comprehensive two-dimensional (2D) gas chromatography time-of-flight mass spectrometry (GC×GC-TOFMS) with discovery-based analysis is reported. First, a standard mix containing n-alkanes, 1-alkenes, and aromatic species was brominated, and quantified using %Reacted as a metric for each compound class, defined as the difference in total peak area between the brominated and original samples normalized to the original sample. The average %Reacted (1 s.d.) were -1.45% (2.8%) for the alkanes, 99.5% (0.4%) for the alkenes, and 6.7% (11.6%) for the aromatics, demonstrating excellent selectivity towards the alkenes with only minor aromatic bromination. The bromination chemistry was then applied to gasoline, followed by GC×GC-TOFMS analysis of the original and brominated gasoline. This GC×GC-TOFMS data set was then submitted to the supervised discovery tool tile-based F-ratio analysis (FRA), which reduced the large data set to only the chromatographic regions which distinguish between the original and brominated gasoline samples. FRA discovered 314 hits, 56 of which were determined statistically significant using combinatorial null distribution analysis (CNDA), a permutation-based significance test. Since the brominated olefins elute in an uncrowded region of the 2D chromatogram and have no signal in the original sample, their discoverability was greatly increased relative to the original olefins. By combining the information gained from brominated olefin standards and the structured patterns of the GC×GC separations, the top hits were identified as the dibromoalkane products of mono-olefins, with five C5 mono-olefins identified on a species-level. The analytical workflow has broad implications for using selective reaction chemistries to facilitate supervised discovery by targeting desired compound classes.

1.6.5 Simulating comprehensive two-dimensional gas chromatography mass spectrometry data with realistic run-to-run shifting to evaluate the robustness of tile-based F-ratio analysis

Untargeted analysis of comprehensive two-dimensional (2D) gas chromatography time-of-flight mass spectrometry (GC×GC-TOFMS) data has the potential to be hindered by run-to-run retention time shifting. To address this challenge, tile-based Fisher ratio (F-ratio) analysis (FRA) has been developed, which utilizes a supervised, untargeted approach involving a chromatographic segmentation routine termed "tiling" combined with the ANOVA F-ratio statistic to discover class-distinguishing analytes while minimizing false positives arising from shifting. The tiling algorithm is designed to account for retention shifting in both separation dimensions. Although applications of FRA have been reported, there remains a need to thoroughly evaluate the robustness of FRA for different levels of run-to-run retention shifting in order to broaden the scope of its application. To this end, a novel method of simulating GC×GC-TOFMS chromatograms with realistic run-to-run shifting is presented by random generation of low-frequency "shift functions". The dimensionless retention-time precision, $\langle\delta\rangle$, which is four times the standard deviation in retention time normalized to the peak width-at-base is used as a key modeling variable along with the 2D chromatographic saturation, $\alpha_{\mathrm{e,2D}}$, and within-class relative standard deviation in peak area, RSD_{wc}. We demonstrate that all three of these variables operate together to impact true positive discovery. To quantify the "success" of true positive discovery, GC×GC-TOFMS datasets for various combinations of $\langle\delta\rangle$, $\alpha_{\mathrm{e,2D}}$, and RSD_{wc} were simulated and then analyzed by FRA using a wide range of relative tile areas (RTA), which is a dimensionless measure of tile size. Since each hit in the FRA hit list was known *a priori* as either a true or false positive based on the simulation inputs, receiver operating characteristic (ROC) curves were readily constructed. Then, the area under the ROC curve (AUROC) was used as a metric for discovery "success" for various combinations of the modeling variables. Based on the results of this study, recommendations for tile size selection and experimental design are provided, and further supported by comparison to previous tile-based FRA applications. For instance, values for $\langle\delta\rangle$, $\alpha_{\mathrm{e,2D}}$, and RSD_{wc} obtained from a GC×GC-TOFMS dataset of yeast metabolites suggested an optimum RTA of 6.25, corresponding closely to the RTA of 4.00 employed in the study, implying the simulation results obtained here can be generalized to real datasets.

1.7 References

(1) Z. Liu, J.B. Phillips, Comprehensive two-dimensional gas chromatography using an on-column thermal modulator interface, *J. Chromatogr. Sci.* **1991**, *29*, 227–231.

(2) M.J.E. Golay, Theory of chromatography in open and coated tubular columns with round and rectangular cross-sections, *Gas Chromatography,* **1958**, *2*, 36-55.

(3) M.S. Klee, J. Cochran, M. Merrick, L.M. Blumberg, Evaluation of conditions of comprehensive two-dimensional gas chromatography that yield a near-theoretical maximum in peak capacity gain, *J. Chromatogr. A* **2015**, *1383*, 151–159.

(4) A.L. Lee, K.D. Bartle, A.C. Lewis, A model of peak amplitude enhancement in orthogonal two-dimensional gas chromatography, *Anal. Chem.* **2001**, *73*, 1330–1335.

(5) A. Mostafa, T. Górecki, Sensitivity of comprehensive two-dimensional gas chromatography (GCXGC) versus one-dimensional gas chromatography (1D GC), *LC-GC Eur.* **2013**, *26*, 672–679.

(6) R. Annino, L.E. Bullock, Continuous chromatographic analysis using a pseudo random sample switching function, *Anal. Chem.* **1973**, *42*, 1221–1227.

(7) D.A. Skoog, F.J. Holler, T.A. Nieman, Principles of instrumental analysis, Sanders College Publishing, San Franciso, 1998.

(8) J.E. Lovelock, Solute switching and detection by synchronous demodulation in gas chromatography, *J. Chromatogr.* **1975***, 112*, 29–36.

(9) J.R. Valentin, G.C. Carfe, J.B. Phillips, A Non-mechanical chemical concentration Modulator for multiplex gas chromatography, *J. High Resolut. Chromatogr.* **1982**, *5*, 269–271.

(10) J. Blomberg, P.J. Schoenmakers, J. Beens, R. Tijssen, Comprehensive two-dimensional gas chromatography (GC×GC) and its applicability to the characterization of complex (petrochemical) mixtures, *J. High Resolut. Chromatogr.* **1997**, *20*, 539–544.

(11) H.J. De Geus, A. Schelvis, J. De Boer, U.A.Th. Brinkman, Comprehensive two-dimensional gas chromatography with a rotating thermal desorption modulator and independently temperature-programmable columns, *J. High Resolut. Chromatogr.* **2000,** *23*, 189-196.

(12) P.J. Marriott, R.M. Kinghorn, Longitudinally modulated cryogenic system. A generally applicable approach to solute trapping and mobilization in gas chromatography, *Anal. Chem.* **1997**, *69*, 2582–2588.

(13) R.M. Kinghorn, P.J. Marriott, P.A. Dawes, Design and implementation of comprehensive gas chromatography with cryogenic modulation, *J. High Resolut. Chromatogr.* **2000**, 23, 245-252.

(14) E.B. Ledford, C. Billesbach, Jet-cooled thermal modulator for comprehensive multidimensional gas chromatography, *J. High Resolut. Chromatogr.* **2000**, *23*, 202-204.

(15) J. Luong, X. Guan, S. Xu, R. Gras, R.A. Shellie, Thermal independent modulator for comprehensive two-dimensional gas chromatography, *Anal. Chem.* **2016**, *88*, 8428–8432.

(16) M. Zoccali, B. Giocastro, P.Q. Tranchida, L. Mondello, Use of a recently developed thermal modulator within the context of comprehensive two-dimensional gas chromatography combined with time-of-flight mass spectrometry: Gas flow optimization aspects, *J. Sep. Sci.* **2019**, *42*, 691–697.

(17) Z. An, H. Ren, M. Xue, X. Guan, J. Jiang, Comprehensive two-dimensional gas chromatography mass spectrometry with a solid-state thermal modulator for in-situ speciated measurement of organic aerosols, *J. Chromatogr.* A **2020**, *1625*, 461336.

(18) C.A. Bruckner, B.J. Prazen, R.E. Synovec, Comprehensive two-dimensional high-speed Gas chromatography with chemometric analysis, *Anal. Chem.* **1998**, *70*, 2796–2804.

(19) R.L. Wade, S.P. Cram, Fluidic logic sampling and injection system for gas chromatography, *Anal. Chem.* **1972**, *44*, 131-139.

(20) J.B. Phillips, M.F. Burke, Cross-correlation chromatography applied to gas-solid adsorption studies, *J. Chromatogr. Sci.* **1976**, *14*, 495–497.

(21) A.E. Sinha, B.J. Prazen, C.G. Fraga, R.E. Synovec, Valve-based comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometric detection: instrumentation and figures-of-merit, *J. Chromatogr. A* **2003**, *1019*, 79–87.

(22) A.E. Sinha, K.J. Johnson, B.J. Prazen, S.V. Lucas, C.G. Fraga, R.E. Synovec, Comprehensive two-dimensional gas chromatography of volatile and semi-volatile components using a diaphragm valve-based instrument, *J. Chromatogr. A* **2003**, *983*, 195–204.

(23) P.A. Bueno, J. V. Seeley, Flow-switching device for comprehensive two-dimensional gas chromatography, *J. Chromatogr. A* **2004**, *1027*, 3–10.

(24) D. Deans, A new technique for heart cutting in gas chromatography, *Chromatographia* 1968, 1, 18–22.

(25) M. Kochman, A. Gordin, T. Alon, A. Amirav, Flow modulation comprehensive two-dimensional gas chromatography-mass spectrometry with a supersonic molecular beam, *J. Chromatogr. A* **2006**, *1129*, 95–104.

(26) J. Seeley, N.J. Micyus, J.D. McCurry, S.K. Seeley, Comprehensive two-dimensional gas chromatography with a simple fluidic modulator, *Am. Lab.* **2006**, *39*, 24.

(27) M. Poliak, M. Kochman, A. Amirav, Pulsed flow modulation comprehensive two-dimensional gas chromatography, *J. Chromatogr. A* **2008**, *1186*, 189–195.

(28) R.E. Mohler, B.J. Prazen, R.E. Synovec, Total-transfer, valve-based comprehensive two-dimensional gas chromatography, *Anal. Chim. Acta.* **2006**, *555*, 68–74.

(29) H. Cai, S.D. Stearns, A comprehensive two-dimensional gas chromatography valve modulation method using hold-release primary column flow for long secondary separation time with 100% transfer, *J. Chromatogr. A* **2018**, *1569*, 200–211.

(30) C.E. Freye, L. Mu, R.E. Synovec, High temperature diaphragm valve-based comprehensive two-dimensional gas chromatography, *J. Chromatogr. A* **2016**, 1424, 127–133.

(31) C.E. Freye, R.E. Synovec, High temperaturee diaphragm valve-based comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry, *Talanta* **2016**, *161*, 675–680.

(32) F.A. Franchina, L.M. Dubois, J.F. Focant, In-depth cannabis multiclass metabolite profiling using sorptive extraction and multidimensional gas chromatography with low- and high-resolution mass spectrometry, *Anal. Chem.* **2020**, *92*, 10512–10520.

(33) J.F. Griffith, W.L. Winniford, K. Sun, R. Edam, J.C. Luong, A reversed-flow differential flow modulator for comprehensive two-dimensional gas chromatography, *J. Chromatogr. A* **2012**, *1226*, 116–123.

(34) J.C. Giddings, Maximum number of components resolvable by gel filtration and other elution chromatographic methods, *Anal. Chem.* **1967**, *39*, 1027–1028.

(35) P.M. Harvey, R.A. Shellie, Factors affecting peak shape in comprehensive two-dimensional gas chromatography with non-focusing modulation, *J. Chromatogr. A* **2011**, *1218*, 3153-3158.

(36) J. Krupčík, P. Májek, R. Gorovenko, I. Špánik, P. Sandra, D.W. Armstrong, On the determination of a detector response enhancement factor for flow modulated comprehensive two-dimensional gas chromatography, *J. Chromatogr. A* **2013**, *1286*, 235–240.

(37) C.A. Bruckner, B.J. Prazen, R.E. Synovec, Comprehensive two-dimensional high-speed gas chromatography with chemometricaAnalysis, *Anal. Chem.* **1998**, *70*, 2796–2804

(38) W.G. Pool, J.W. De Leeuw, B. Van De Graaf, A rapid routine to correct for skewing in gas chromatography/mass spectrometry, *J. Mass Spectrom.* **1996**, *31*, 213-215.

(39) Y. Izadmanesh, E. Garreta-Lara, J.B. Ghasemi, S. Lacorte, V. Matamoros, R. Tauler, Chemometric analysis of comprehensive two dimensional gas chromatography–mass spectrometry metabolomics data, *J. Chromatogr. A* **2017**, *1488*, 113–125.

(40) D.V. Gough, S. Schoneich, R.E. Synovec, Chemometric decomposition of comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry data employing partial modulation in the negative pulse mode, Talanta **2020**, *210*, 120670.

(41) R. Bro, PARAFAC. Tutorial and applications, Chemom. Intell. Lab. Syst. **1997**, *38*, 149–171.

(42) J.C. Hoggard, R.E. Synovec, Parallel factor analysis (PARAFAC) of target analytes in GC × GC-TOFMS data: Automated selection of a model with an appropriate number of factors, Anal. Chem. **2007**, *79*, 1611–1619.

(43) D.K. Pinkerton, B.A. Parsons, T.J. Anderson, R.E. Synovec, Trilinearity deviation ratio: A new metric for chemometric analysis of comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry data, *Anal. Chim. Acta.* **2015**, *871*, 66–76.

(44) R. Bro, A.K. Smilde, Principal component analysis, *Anal. Methods.* **2014**, *6*, 2812–2831.

(45) K.M. Pierce, J.L. Hope, J.C. Hoggard, R.E. Synovec, A principal component analysis based method to discover chemical differences in comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry (GC × GC-TOFMS) separations of metabolites in plant samples *Talanta* **2006**, *70*, 797–804.

(46) C. Monzón, S. Schöneich, R.E. Synovec, Non-targeted discovery of class-distinguishing metabolites in Argentinian pacu fish by comprehensive two-dimensional gas chromatography with principal component analysis, *Microchem. J.* **2021**, *164*, 106004.

(47) K.M. Pierce, J.L. Hope, J.C. Hoggard, R.E. Synovec, A principal component analysis based method to discover chemical differences in comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry (GC×GC-TOFMS) separations of metabolites in plant samples, *Talanta* **2006**, *70*, 797–804.

(48) L.W. Hantao, H.G. Aleme, M.M. Passador, E.L. Furtado, F.A. de L. Ribeiro, R.J. Poppi, F. Augusto, Determination of disease biomarkers in Eucalyptus by comprehensive two-dimensional gas chromatography and multivariate data analysis, *J. Chromatogr. A* **2013**, *1279*, 86–91.

(49) I.G. Chong, C.H. Jun, Performance of some variable selection methods when multicollinearity is present, *Chemom. Intell. Lab. Syst.* **2005**, *78*, 103–112.

(50) K.J. Johnson, R.E. Synovec, Pattern recognition of jet fuels: Comprehensive GC × GC with ANOVA-based feature selection and principal component analysis, *Chemom. Intell. Lab. Syst.*, **2002**, *60*, 225-237.

(51) K.M. Pierce, J.C. Hoggard, J.L. Hope, P.M. Rainey, A.N. Hoofnagle, R.M. Jack, B.W. Wright, R.E. Synovec, Fisher ratio method applied to third-order separation data to identify significant chemical components of metabolite extracts, *Anal. Chem.* 78 (2006) 5068–5075.

(52) R.E. Mohler, K.M. Dombek, J.C. Hoggard, K.M. Pierce, E.T. Young, R.E. Synovec, Comprehensive analysis of yeast metabolite GC×GC-TOFMS data: Combining discovery-mode and deconvolution chemometric software, *An-*

alyst **2007**, *132.* 756–767.

(53) H.D. Bean, J.E. Hill, J.M.D. Dimandja, Improving the quality of biomarker candidates in untargeted metabolomics via peak table-based alignment of comprehensive two-dimensional gas chromatography-mass spectrometry data, *J. Chromatogr. A* **2015**, *1394*, 111–117.

(54) S. Stadler, P.H. Stefanuto, M. Brokl, S.L. Forbes, J.F. Focant, Characterization of volatile organic compounds from human analogue decomposition using thermal desorption coupled to comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry, *Anal. Chem.* **2013**, *85*, 998–1005.

(55) M. Ueland, S. Harris, S.L. Forbes, Detecting volatile organic compounds to locate human remains in a simulated collapsed building, *Forensic Sci. Int.* **2021**, *323*, 110781.

(56) L.C. Marney, W. Christopher Siegler, B.A. Parsons, J.C. Hoggard, B.W. Wright, R.E. Synovec, Tile-based Fisher-ratio software for improved feature selection analysis of comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry data, *Talanta* **2013**, *115*, 887–895.

(57) B.A. Parsons, L.C. Marney, W.C. Siegler, J.C. Hoggard, B.W. Wright, R.E. Synovec, Tile-based Fisher ratio analysis of comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry (GC × GC-TOFMS) data using a null distribution approach, *Anal. Chem.* **2015**, *87,* 3812–3819.

(58) Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: A practical and powerful approach to multiple testing, *J. R. Stat. Soc. B.* **1995**, *57*, 289–300.

(59) F. Erni, R.W. Frei, Two-dimensional column liquid chromatographic technique for resolution of complex mixtures, *J. Chromatogr.* **1978**, *149*, 561–569.

(60) M.M. Bushey, J.W. Jorgenson, Automated instrumentation for comprehensive two-dimensional high-performance liquid chromatography/capillary zone electrophoresis, *Anal. Chem.* **1990**, *62*, 978–984.

(61) R.K. Andjus, J.E. Lovelock, Reanimation of rats from body temperatures between 0 and 1° C by microwave diathermy, *J. Physiol.* **1955**, *128*, 541–546.

(62) B.L. Karger, L.R. Snyder, C. Horvath, An introduction to separation science, Wily & Sons, New York, 1973.

(63) R.G. Brereton, G.R. Lloyd, Partial least squares discriminant analysis: Taking the magic away, *J. Chemom.* 28 **2014**, *28*, 213–225.

Chapter 2

DYNAMIC PRESSURE GRADIENT MODULATION FOR COMPREHENSIVE TWO-DIMENSIONAL GAS CHROMATOGRAPHY

This chapter was adapted from T.J. Trinklein, D.V. Gough, C.G. Warren, G.S. Ochoa, R.E. Synovec, *J. Chromatogr. A* **2020**, *1609*, 460488.

2.1 Introduction

Comprehensive two-dimensional gas chromatography (GC×GC), pioneered in 1991 by Liu and Phillips [1], has continued to evolve as an instrumental method of choice for the analysis of complex samples for volatile and semi-volatile analytes, with applications in metabolomics [2], forensic science [3], and the petroleum industry, among others [4,5]. With GC×GC, the first-dimension separation column (^{1}D) is coupled to the second-dimension separation column (^{2}D) by a modulator. GC×GC has been theoretically demonstrated to be superior to one-dimensional GC (1D-GC) by providing considerably more chemical selectivity concurrent with about a 10-fold enhancement in peak capacity [6]. If analytes are unresolved on the ^{1}D separation, they have another opportunity to be resolved when subjected to separation on the ^{2}D column with sufficient orthogonal selectivity. The modulator is the key to providing the superiority of GC×GC, by transferring eluate from the ^{1}D column to the ^{2}D column as quickly and efficiently as possible, essentially providing high-speed injections between the two serially connected columns.

Modulators can be broadly classified into two categories: thermal modulation [1,7–11], and flow modulation [12–20]. Based upon common usage, flow modulation is often referred to as either differential flow modulation [12–15], or flow diversion modulation [17–20]. With thermal modulation, temperature control is used to trap analytes from the ^{1}D eluate, and the trapped analytes are rapidly released for ^{2}D separation. Conversely, with flow modulation, gas flow is used to control and isolate portions of the ^{1}D eluite, with subsequent injection onto the ^{2}D column to provide modulation. Thermal modulation generally provides 100% duty cycle, i.e., a 100% transfer of ^{1}D eluate to the ^{2}D column. Depending upon their design, flow modulators provide a duty cycle $\leq$ 100%. A larger duty cycle generally translates into a higher detection signal-to-noise ratio (S/N). However, the S/N also depends upon the modulation period (P_M), how narrow the ^{2}D peaks are relative to the ^{1}D peaks, and detector performance.

The first reports of GC×GC used thermal modulation [1,21], which led to the use of cryogens [9], such as jet-cooled thermal modulation that remains popular to this day [10]. Additional developments in thermal modulation have been reported [22–24], with a recent focus on consumable free designs [25,26] due to the high operating costs associated with cryogens. Flow modulation was introduced as a simpler and less resource intensive alternative to thermal modulation. The first flow modulator used a diaphragm valve to rapidly divert small portions of the ^{1}D eluate to the ^{2}D column for separation [12]. However, this implementation suffered from a poor duty cycle. To improve the duty cycle, Seeley introduced differential flow modulation. In his initial report, a diaphragm valve was used to collect ^{1}D effluent in a sample loop for the majority of the P_M prior to injection and separation on ^{2}D with a fast flow rate, achieving duty cycles on the order of 80% [13]. However, diaphragm valve-based modulation remained relatively unpopular due to the restricted temperature range of early valve designs. Under high temperature conditions, the diaphragm valve O-rings decomposed. To overcome this limitation, "high-temperature" diaphragm valves have been introduced which use a perfluoroelastomer (Kalrez) O-ring [15,16]. Historically, implementation of flow modulation has been hampered by a trade-off between maintaining a high duty cycle and operating under flow rates compatible with mass spectrometry (MS) detection. Thus, there remains a need to continue to explore and develop advanced modulators for GC×GC, in order to provide improved performance, such as the ability to provide narrow ^{2}D peaks ($^2w_\mathrm{b} \geq 10$ ms), a wide P_M range (50 ms $\leq P_\mathrm{M} \leq 6$ s), and 100% duty cycle, concurrent with cryogen-free operation [27,28]. In this report, we describe a flow modulation technique that begins to address these issues using a two-position solenoid valve to control the gas flows on ^{1}D and ^{2}D.

The flow modulation technique reported herein uses a commercially available 2-position solenoid "pulse" valve to modulate, at a user-specified pulse width (p_w) and P_M, a flow of auxiliary carrier gas applied to a T-junction joining the ^{1}D and ^{2}D columns. The entire ^{1}D column eluate is transferred to the ^{2}D column, providing 100% duty cycle. The ^{1}D eluate is not isolated in a sample loop/compartment as in current implementations of differential flow modulation, and is not trapped as in thermal modulation. Rather, the ^{1}D eluite is temporarily stopped due to pressure of a suitably high auxiliary pressure under non-equilibrium pressure conditions. Then the valve is briefly closed, allowing flow to resume and providing injection of ^{1}D eluite onto ^{2}D. We refer to this flow modulation technique as "dynamic pressure gradient modulation" (DPGM) since a pressure gradient oscillates within the ^{1}D and ^{2}D column ensemble, in rhythm with the applied P_M and p_w, providing full modulation and a fast ^{2}D flow rate for analyte compression in time. Additionally, the pulse valve is located outside of the oven, so there are no temperature restrictions.

First, a 90-component test mixture was used to demonstrate and initially evaluate DPGM. Next, the signal enhancement (SE) was examined, defined as the peak height of the highest modulated ^{2}D peak divided by the unmodulated ^{1}D peak [30]. A diesel sample is then analyzed to demonstrate modulation performance with a complex sample. Finally, the modulation principle was investigated as a function of P_M, p_w, and the volumetric flow rates, ^1F and ^2F, on the ^{1}D and ^{2}D columns, respectively.

2.2 Principles and theory

In 2004 Cai and Stearns introduced partial modulation [31], a unique modulation technique wherein a rapid-actuating valve is used to introduce a pulse of carrier gas to a Y-junction, at a user-specified P_M, connecting the ^{1}D and ^{2}D columns. With partial modulation, only a portion of the ^{1}D eluate is modulated, while the remainder passes to the detector, with the ^{2}D peaks superimposed on a background of the 1D chromatogram. The resulting 1D and ^{2}D chromatographic information can be mathematically decoupled. In essence, partial modulation retains the integrity of the 1D separation at the cost of decreased ^{2}D detection sensitivity relative to if full modulation were applied. We have been developing a version of this instrument, based upon the use of a commercially available pulse valve that is interfaced to the ^{1}D and ^{2}D column ensemble of the GC×GC via a T-junction (Fig. 1) [27–29].

Partial modulation can be operated in two modes, either the positive pulse mode (PPM) or the negative pulse mode (NPM). In the positive pulse mode, the flow of auxiliary carrier gas from the pulse valve is turned off for nearly all of the P_M, and then is briefly turned on for a user-specified pulse width ($p_\mathrm{w,PPM}$). The brief pulse of auxiliary carrier gas creates a local analyte vacancy in the ^{1}D eluate as it enters the ^{2}D column. Through a series of data processing steps including differentiation, inversion, and baseline correction, the vacancy chromatograms from the positive pulse mode could be converted into "apparent" 2D chromatograms, which are cut and folded at each P_M into a plot which serves as a traditional GC×GC chromatogram [27,32].

In the negative pulse mode the flow of auxiliary carrier gas is applied for nearly all of the P_M, resulting in dilution of the analyte stream. The pulse valve is then very briefly turned off for a user-specified pulse width ($p_{\mathrm{w,NPM}}$), effectively providing an "injection" of undiluted ^{1}D eluate onto the ^{2}D column. The resultant data has the appearance of ^{2}D separations superimposed on top of the ^{1}D separation. We have shown that traditional 2D chromatograms can be obtained by applying a rolling ball minimum baseline correction (written in-house) [33] to remove the 1D separation, which can then be cut and folded into a GC×GC chromatogram. Recently, we demonstrated that partial modulation in the negative pulse mode can readily yield extremely narrow ^{2}D peak widths-at-base $^2w_{\mathrm{b}}$ (4σ) of about 10 ms with a P_{M} of 100 ms [29]. Since the pulse width in the positive pulse mode ($p_{\mathrm{w,PPM}}$) designates the time the auxiliary carrier gas is on, and the pulse width in the negative pulse mode is the time the auxiliary carrier gas is off ($p_{\mathrm{w,PPM}}$), the P_{M} is simply the sum of the two pulse widths. That is,

$$P_{\mathrm{M}} = p_{\mathrm{w,PPM}} + p_{\mathrm{w,NPM}} \tag{2.1}$$

Experimentally, it is convenient to think of the transition between the positive and negative pulse modes as being brought about by tuning the $p_{\mathrm{w,PPM}}$. Thus, we simply refer to the $p_{\mathrm{w,PPM}}$ as p_{w}, so at very low pulse widths ($p_{\mathrm{w}} \ll P_{\mathrm{M}}$) partial modulation in the negative pulse mode occurs, whereas partial modulation in the positive pulse mode occurs at very high pulse widths (p_{w} approaching P_{M}). Previously, it was thought that only these two modes of operation were obtainable by appropriate selection of p_{w}. In this report we describe that full modulation can also be obtained at intermediate p_{w}, between these two extremes of partial modulation. Achieving full modulation requires applying a suitable auxiliary pressure (P_{aux}) to the pulse valve (see Fig. 2.1) to provide temporary stop-flow conditions on the ^{1}D column. Indeed, use of a pneumatic valve with an applied P_{aux} to periodically stop ^{1}D flow was described by Górecki and co-workers [34]. In their study a constant stream of propane was injected by alternately switching the pneumatic valve on and off. At a sufficiently high P_{aux} the signal intensity briefly approached zero indicating temporary stop-flow conditions had been achieved. The stop-flow conditions remained for a period of time even after the auxiliary flow had been switched off until the pressure on the ^{1}D column became high enough to overcome the mid-point pressure [34].

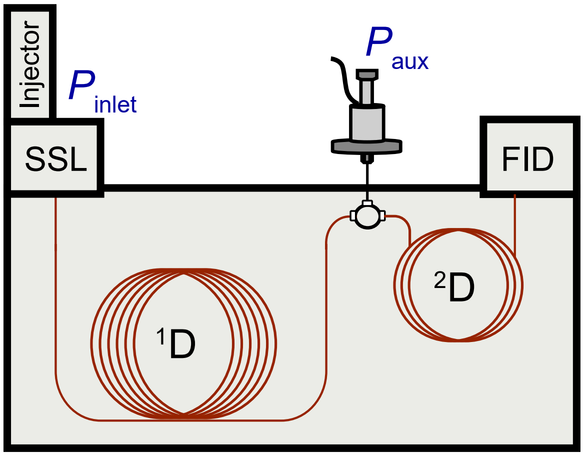


Figure 2.1: Diagram of a GC×GC instrument configured with DPGM.

Here, we describe how the same temporary stop-flow phenomenon can be utilized to perform fast differential flow modulation by rapidly actuating a pulse valve with an appropriate p_{w} and P_{aux}. For a given GC×GC column ensemble, an analyst makes an initial selection of P_{M} and p_{w}. An appropriate P_{aux} is chosen to provide the desired flow ratio ($^1F/^2F$), and subsequently tuned to obtain stop-flow conditions, i.e., baseline signal intensity between modulated ^{2}D peaks. If P_{aux} is not tuned to be suitably high, stop-flow conditions and full modulation will not occur for any p_{w} [34]. Finally, p_{w} can be adjusted both to ensure band-broadening contributions from ^{2}D injection are minimized, and as a means to provide full modulation. Similarly, stop-flow conditions will not be achieved for any P_{aux} if p_{w} is too short (positive pulse mode).

2.3 Experimental

All experiments were performed on an Agilent 7890 GC (Agilent Technologies, Palo Alto, CA, USA) with flame ionization detection (FID). The stock electrometer for the Agilent FID was replaced with an electrometer built in-house to allow data collection at 100 kHz. The data was then boxcar averaged to

2 kHz prior to further processing. The electrometer was interfaced to a data acquisition board and the data was collected using an in-house program written in LabVIEW (National Instruments, Austin, TX, USA). Data processing was performed in MATLAB R2018a (The Mathworks, Inc., Natick, MA, USA). Samples were injected using a 7683B autoinjector (Agilent Technologies, Palo Alto, CA, USA). Acetone was used as a wash solvent between each run. The carrier gas was ultrahigh purity hydrogen (Grade 5, 99.999%, Praxair, Seattle, WA, USA). The inlet and FID temperatures were set to 175 °C and 250 °C, respectively, for all experiments.

The ^{1}D and ^{2}D columns of the GC×GC instrument were connected via a microvolume T-union (Model MT.5CXS6, Valco Instruments Company Inc., Houston, TX, USA). The third port of the T-union was connected to a high speed pulse valve (Model 009-0347-900, Parker Hannifin, Hollis, NH, USA) using a 5 cm L × 125 μm i.d. stainless steel transfer line (Model T5C5D, Valco Instruments Company Inc., Houston, TX, USA) and a custom-built fitting (Fig. 2.1). The pulse valve was connected to a National Instruments data board (National Instruments, Austin, TX, USA) and controlled using the same LabVIEW program that collected data from the FID.

A 90-component test mixture containing a variety of compounds with a wide boiling points range (54-372 °C) was used to evaluate the instrumental platform (Table A1) [29]. A "normal configuration" column set was used. The ^{1}D column had a DB-5 stationary phase, equivalent to (5%-diphenyl)-dimethylpolysiloxane, with dimensions of 10.0 m × 180 μm d_c × 0.18 μm d_f. The ^{2}D column had a polar RTX-2330 stationary phase, equiavlent to 90% biscyanopropyl/10% phenylcyanopropyl polysiloxane, with dimensions of 1.0 m × 180 μm d_c × 0.10 μm d_f). A thinner film was used in the ^{2}D column in order to provide an optimal peak capacity with the ^{2}D separation [6] concurrent with preventing wraparound with a P_M = 750 ms [35,36]. For each run, 0.2 μ l of sample was injected at a 100:1 split ratio. The oven was held at 30 °C for 1 min and then ramped to 174 °C at 6 °C/min. The absolute head pressure at the inlet of the ^{1}D column, P_inlet (Fig. 2.1), was held at 128.9 kPa for 1 min, and then ramped to 211 kPa at 4.14 kPa/min, where it was held until the end of the separation. The auxiliary pressure, P_aux (Fig. 2.1) applied to the pulse valve was held at 133.8 kPa for 1 min, and subsequently ramped to 218.6 kPa at 4.14 kPa/min and held for 8 min. The applied P_aux produced a ^{2}D flow of 22.9 ml/min at the beginning of the separation. A wide range of p_w were examined: 0, 10, 20, 40, 60, 100, 200, 500, 725, 745, and 750 ms. The two extreme p_w produced no modulation, i.e., either p_w on with continuous auxiliary flow ($p_\mathrm{w} = 0$), or p_w off with no auxiliary flow ($p_\mathrm{w} = P_\mathrm{M} = 750$ ms).

A diesel sample sourced from a local gas station was analyzed to demonstrate the modulator performance with a more complex sample. A “reverse” column GC×GC configuration was implemented. The ^{1}D column had a polar Rxi-17Sil MS stationary phase, similar to (50%-phenyl)-methylpolysiloxane, with dimensions of 20.0 m × 180 μm d_c × 0.18 μm d_f. The ^{2}D column had a non-polar Rxi-1 stationary phase, equivalent to (100%methyl)-polysiloxane, with dimensions of 2.0 m × 180 μm d_c× 0.18 μm d_f. Three replicate diesel chromatograms were obtained by injecting a 0.2 μl sample with a 50:1 split ratio. A p_w of 150 ms was used. The oven was first held at 40 °C for 1 min, ramped to 300 °C at 5 °C/min, and held for 7 min at 300 °C. The absolute head pressure P_inlet was held at 211.7 kPa for 1 min and then ramped to 341 kPa at 2.2 kPa/min. TheP_aux was held initially at 197.9 kPa for 1 min and then ramped to 251.7 kPa at 1.72 kPa/min where it was held until the end of the separation.

An isothermal separation of acetone at 90 °C was performed to study the interdependence between, P_M, p_w, and the flow rates, ^1F and ^2F, on the ^{1}D and ^{2}D columns, respectively. A "normal configuration" column set was used. The ^{1}D column had a DB-5 stationary phase with dimensions of 10.0 m × 180 μm d_c × 0.18 μm d_f. The ^{2}D column had a polar RTX-2330 stationary phase, with dimensions of 2.0 m × 180 μm d_c × 0.10 μm d_f).Each sample injection was 0.2 μl with a 200:1 split ratio, and all chromatograms were acquired in triplicate. Acetone was found to be essentially unretained on both dimensions and served as a dead-time marker for ^{1}D and ^{2}D. The average linear flow velocities for ^{1}D and ^{2}D were obtained from the known column length and measured dead time. The average linear flow velocities were then converted to average volumetric flow rates $^1F_{\text{ave}}$ and $^2F_{\text{ave}}$ (ml/min) from the known cross-sectional area of the column. For comparison purposes, the ^1F when no pressure is applied (during the $p_{mathrmw}$ interval) was also calculated, referred to as $^1F_{\text{calc}}$, and determined to be 1.96 ml/min. To determine the optimum p_{w} for each P_{M}, chromatograms were acquired for at least six p_{w} for P_{M} of 100, 200, 300, 400, and 750 ms. The p_{w} range for each P_{M} was selected to include both partial modulation in the negative pulse mode and full modulation, such that the minimum p_{w} at which full modulation was achieved could be clearly identified. The P_{M} of 200 ms was the primary focus, with chromatograms obtained at p_{w} of 5, 7.5, 10, 15, 20, 30, 40, 50, 75, 100, and 125 ms.

2.4 Results and discussion

To explore the transitions between the three modulation modes the 90-component test mixture was examined using a broad range of p_{w} ranging from 0 ms to 750 ms with a $P_{\text{M}}$of 750 ms. Note that no modulation occurs either with P_{aux} on ($p_{\text{w}} = 0$), or with P_{aux} off ($p_{\text{w}} = P_{\text{M}}$). The first few minutes of the data for this test mixture are presented in Fig. 2.2. As a consequence of varying degrees of modulation the data are presented on different signal scales. When no pulsing is applied ($p_{\text{w}} = 0$ ms) a constant auxiliary gas flow is applied, an unmodulated, one-dimensional chromatogram results (Fig. 2.2A). Partial modulation in the negative pulse mode occurs for a p_{w} of about 10 ms to 40 ms, with data for a p_{w} = 10 ms presented (Fig. 2.2B). In the negative pulse mode, a portion of the ^{1}D eluate is separated on the ^{2}D column in a mobile phase consisting of both carrier gas and unresolved ^{1}D eluate [29]. The resulting $^2w_{\text{b}}$ are very narrow, and the ^{1}D and ^{2}D separations are readily uncoupled to produce 2D GC×GC plots. However, the sensitivity in this regime is lower since only a fraction of the ^{1}D peaks are modulated. Perhaps the most interesting mode of operation, and the focus of this report, occurs between a p_{w} of about 60 ms to 500 ms, with data for a $p_{\text{w}} = 60$ ms presented (Fig. 2.2C). In this time regime, full modulation was observed, wherein the entirety of the eluate from the ^{1}D separation is modulated to produce ^{2}D separations for each P_{M}. The $^2w_{\text{b}}$ are quite narrow, which provides a high detection sensitivity compared to the unmodulated ^{1}D peaks. Indeed, the signal scale is about 20-fold higher in Fig. 2.2C than in Fig. 2.2A (unmodulated) since full modulation significantly increased the signal intensity of the detected peaks.

At very high pulse widths (p_{w} approaching P_{M}), partial modulation in the positive pulse mode occurs, with data for a p_{w} = 745 ms presented (Fig. 2.2D). Based upon our recent studies, due to shortcomings in data processing requirements and compromised detection sensitivity, partial modulation in the positive pulse mode appears to be inferior to using partial modulation in the negative pulse mode. Full modulation should provide the highest detectability. Figure 2.3A compares chromatograms of the 90-component test mixture using full modulation to achieve the unfolded 1D array chromatogram (top, p_{w} = 60 ms), to the chromatogram with the auxiliary flow on but not pulsing (middle, p_{w} = 0 ms), to the chromatogram with P_{aux} off (bottom, p_{w} = 750 ms). The observed peaks are substantially higher with modulation. To see more detail of the unmodulated chromatograms, a zoom-in is provided in Fig. 2.3B. For the separation conditions applied, the overall appearance of the chromatograms with and without an applied P_{aux} is not dramatically different, and a separation is achieved in essentially the same time window. This demonstrates that the time required to achieve a temperature programmed separation using DPGM is, just like traditional 1D-GC, primarily governed by the boiling points of the analytes and the temperature program applied. Thus, despite the temporary stop-flow conditions, DPGM did not drastically increase in the ^{1}D dead time due to the addition of the auxiliary pressure at the T-

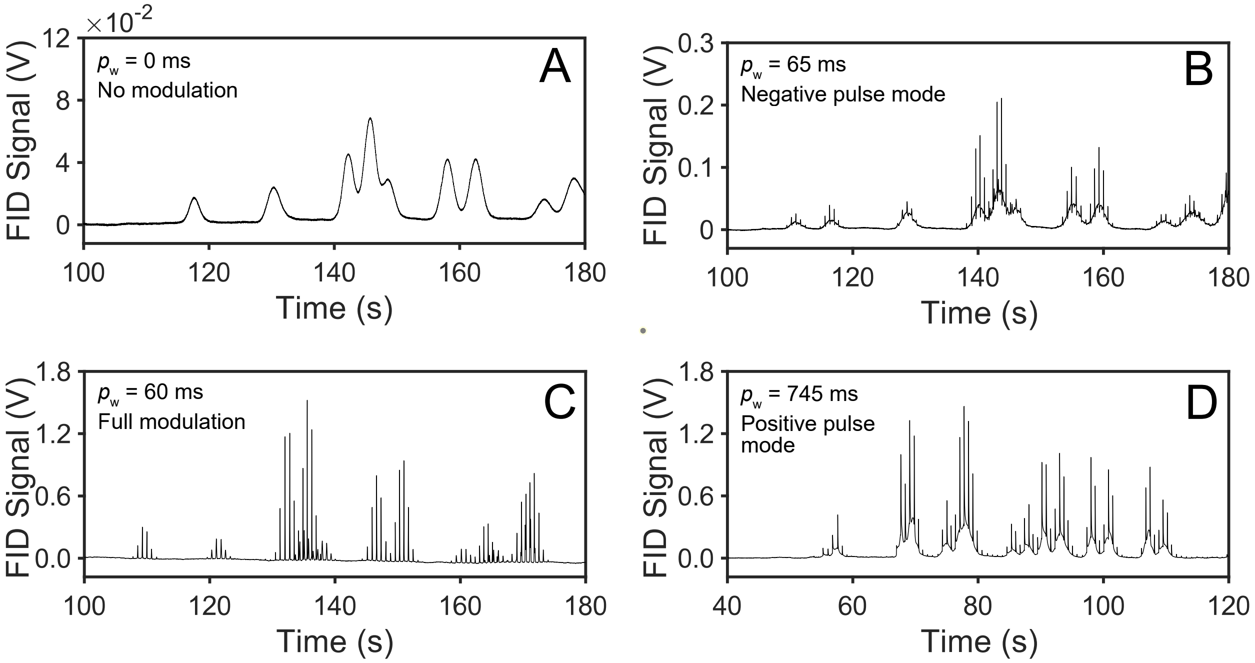


Figure 2.2: (A) One-dimensional chromatogram (p_w = 0 ms). (B) Partial modulation in the negative pulse mode (NPM) (C) Full modulation by DPGM (D) Partial modulation in the positive pulse mode (PPM).

union; this phenomenon is also observed in pressure tuning as introduced by Sacks and co-workers [37]. All else held constant, the p_w will affect analyte retention time (${}^1t_\mathrm{R}$) on the ^{1}D separation. When the eluate from the ^{1}D separation encounters the flow of auxiliary gas provided from the pulse valve at the T-union (Fig. 2.1), the ^{1}D separation is slowed. By definition, the p_w is the length of time that the pulse valve is closed (no auxiliary gas flow). Therefore, as p_w is increased, ^{1}D retention should decrease. To examine this principle, ${}^1t_\mathrm{R}$ was measured over a wide range of p_w for five representative analytes, as shown in Fig. 2.3C. The ${}^1t_\mathrm{R}$ obtained using modulation periods of 0, 745, and 750 ms are omitted due to the presence of overlap on the ^{1}D dimension. At very long p_w (positive pulse mode regime), the valve is open only briefly, and ${}^1t_\mathrm{R}$ is at a minimum for each analyte. Conversely, at very short p_w (negative pulse mode regime), the valve is open for the majority of the P_M, and the opposing gas flow from the pulse valve slows the migration of eluate along the ^{1}D column.

A maximum ${}^1t_\mathrm{R}$ is observed for the minimum p_w. Indeed, the ${}^1t_\mathrm{R}$ are slightly longer with the P_aux on all the time (p_w = 0 ms). The experimental conditions in the negative pulse mode regime are very similar to those applied with pressure tuning of column selectivity [37]. In fact, the only difference is that now the flow of auxiliary carrier gas is pulsed to modulate the eluate as it enters the ^{2}D column. To some extent, pressure tuning of the two separations appears to be occurring with pulse valve modulation. One might envisage that dynamic pressure gradient modulation, introduced herein, could provide a new approach for pressure tuning the two separation columns with GC×GC [38].

The most striking feature of the chromatograms in Figs. 2.3A and 2.3B is the marked increase in signal intensity observed in the modulated chromatogram relative to the unmodulated chromatograms, either with or without an auxiliary pressure applied. To quantify this observation, we report the signal enhancement (SE) as the ratio of peak heights,

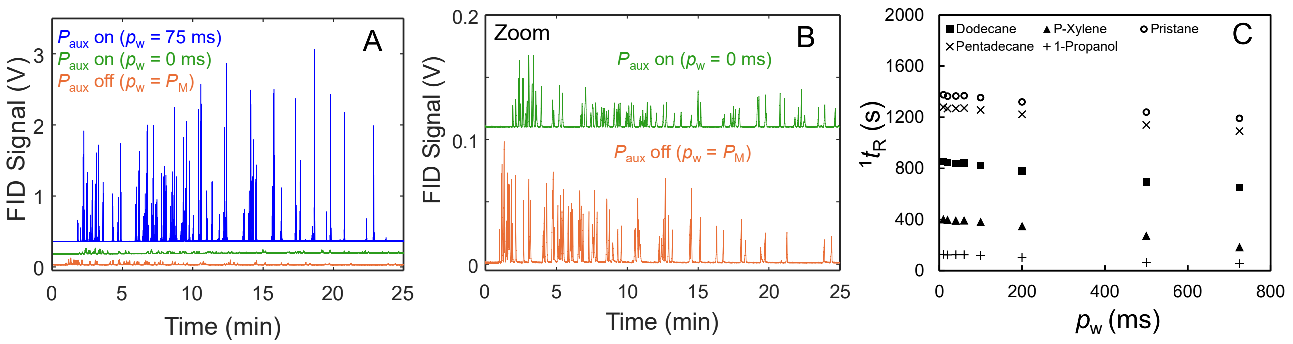


Figure 2.3: (A) Top: 1D chromatogram of the test mixture, using p_w = 60 ms. Middle: Unmodulated 1D chromatogram that results while a constant auxiliary gas is applied. Bottom: Unmodulated 1D chromatogram when no auxiliary gas is applied. (B) Zoom in view of the middle and bottom chromatograms from (A) for comparison. (C) Relationship of ^1t_R, vs. p_w , for five representative test analytes.

$$SE = \frac{^2h}{^1h} \qquad (2.2)$$

where ^2h is the height of the tallest modulated ^{2}D peak, and ^1h is the height of the unmodulated $\neg ^1$D peak from a chromatogram with the pulse valve applying a P_{aux} without actuation (p_w = 0 ms), shown in the top of Fig. 3B. The SE was calculated for pristane, a narrow ^{2}D peak (^2w_b = 22 ms), and for 1-propanol, a wide ^{2}D peak (^2w_b = 182 ms), to illustrate the SE range obtained.

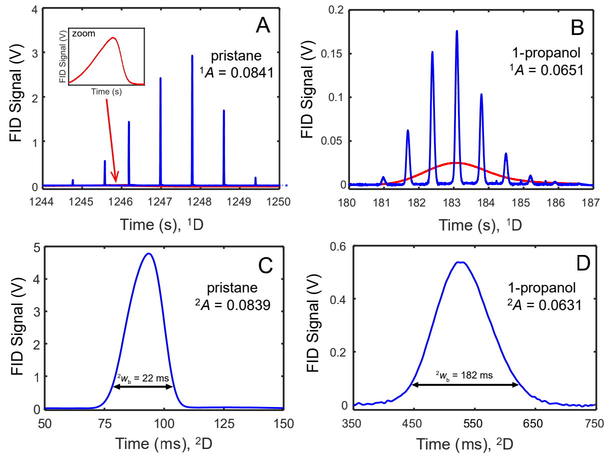


Figure 2.4: Illustration of signal enhancement in DPGM. (A) Unmodulated (red) and modulated (blue) signal for pristane. (B) Unmodulated and modulated signals ffor 1-propanol. (C) Summed ^{2}D peak for pristane. (D) Summed ^{2}D peak for 1-propanol.

The modulated and unmodulated peaks are shown in Figs. 2.4A and B for pristane and 1-propanol, respectively. For each analyte, the unmodulated ^{1}D peak is adjusted slightly in time so as to overlay with the tallest ^{2}D peak.

Applying Eq. (2.2), the SE = 87 for pristane, while SE = 7 for 1-propanol. Since the peak heights obtained by applying a constant pulse valve P_{aux} (p_w = 0) were about 2-fold shorter (due to the lower flow rate) than the peak heights obtained with the pulse valve off, as indicated in Fig. 2.3B, this should be taken into account when one considers the overall SE.

Next, the sum of the areas of the modulated ^{2}D peaks (Fig. 2.2A) were compared to the area from the unmodulated chromatogram (Fig. 2.1A) to demonstrate the conservation of mass obtained using the modulator (100% duty cycle). In order to minimize error, single ^{2}D peaks for pristane (Fig. 2.4C) and 1-propanol (Fig. 2.4D) were obtained by selecting a section of their 1D raw data array, cutting the array at the P_M to create a series of short arrays of length P_M, and adding each of the shorter arrays together. Then, the single ^{1}D and ^{2}D peak both required a single baseline correction prior to summing the baseline corrected signal to obtain accurate areas. The areas obtained for pristane were ^1A = 0.0841 and ^2A = 0.0839, while those for 1-propanol were ^1A = 0.0651 and ^2A = 0.0631. The good agreement between ^1A and ^2A for both analytes is additional,

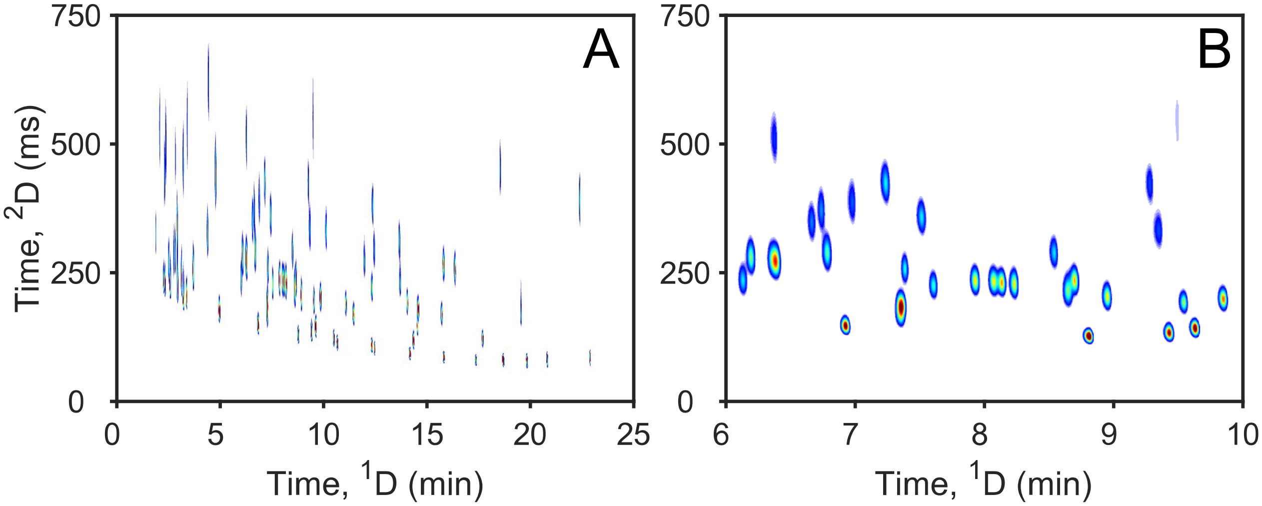


Figure 2.5: (A) GC×GC separation of the 90-component test mixture by DPGM. (B) Zoom-in of the section between 6 and 10 min.

strong evidence for achieving 100% duty cycle.

Using the full modulation, unfolded 1D chromatogram from Fig. 2.3A (top, $p_{\mathrm{w}} = 60$ ms, with $P_{\mathrm{M}}= 750$ ms), the 2D chromatogram of the 90-component test mixture was prepared and is shown in Fig. 2.5A. A zoom-in of the region between 6-10 min is provided in Fig. 2.5B, in order to highlight the excellent 2D separation performance. Since the ^{1}D separation was temperature programmed, the $^1w_{\mathrm{b}}$ were relatively constant with an average of about 3 s. Thus, based upon the separation run time of 25 min, the ^{1}D peak capacity is $^1n_{\mathrm{c}} = 500$ at a chromatographic resolution criterion of $R_{\mathrm{s}} = 1$. For the ^{2}D separations, the $^2w_{\mathrm{b}}$ ranged from about 20 ms to 180 ms, so the $^2n_{\mathrm{c}}$ peak capacity, calculated iteratively at $R_{\mathrm{s}}= 1$ was determined to be 10 [35]. Thus, the ideal 2D peak capacity given by $n_{\mathrm{c,2D}} =^1 n_{\mathrm{c}} \times^2 n_{\mathrm{c}}$ is 5000.

Next, a diesel was analyzed to demonstrate that DPGM produces 2D chromatograms with substantial chemical selectivity detail. A reverse column set was employed to enable better use of the 2D separation space. A longer ^{2}D column (2 m) was used to reduce the flow rate at the exit of the FID to 8 ml/min. Accordingly, a $P_{\mathrm{M}} = 2.0$ s was applied to further demonstrate the modulator performance and versatility. With $P_{\mathrm{M}} = 2$ s, a $p_{\mathrm{w}} = 150$ ms was required to achieve optimal full modulation (longer than the $p_{\mathrm{w}} = 60$ ms with a $P_{\mathrm{M}} = 750$ ms). The 2D chromatogram of the diesel and a zoom-in of the region between 20 and 25 min are given in Fig. 2.6B. Regions containing the alkanes, cycloalkanes, and aromatics are clearly visible in the highly structured chromatogram, which appears very similar to other analyses of diesel using reverse column formats with thermal modulation [39,40]. Figure 2.6C shows a three-dimensional view of the entire separation. The average $^2w_{\mathrm{b}}$ was 168 ms, implying a ^2nc of 12, while the average $^1n_{\mathrm{c}} \times^2 n_{\mathrm{c}}$, is an impressive 8640. To better understand the interdependence of P_{M}, p_{w}, and the flow rates ^1F and ^2F, an isothermal separation of a test mixture composed of acetone, 2-butanol, hexane, toluene, and octane was performed. The longer p_{w} required to provide full modulation during the diesel separation ($P_{\mathrm{M}} = 2$ s) prompted a study to determine the p_{w} necessary to achieve full modulation for a given P_{M}. To facilitate a consistent discussion, the optimum pulse width, $p_{\mathrm{w,opt}}$ is defined as the shortest p_{w} which produces full modulation. The $p_{\mathrm{w,opt}}$, for each P_{M} was determined by visually comparing a single modulation from a chromatogram

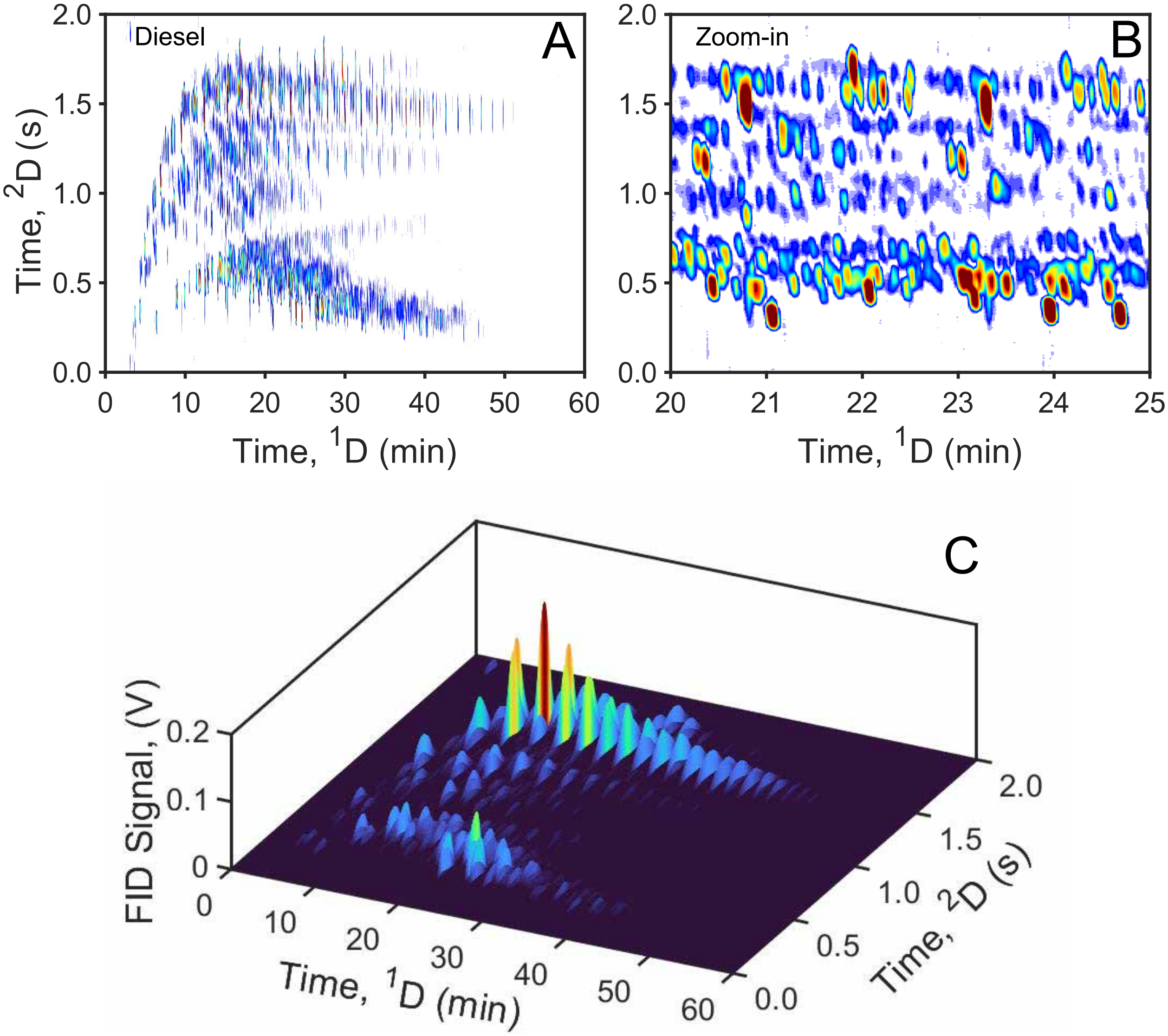


Figure 2.6: (A) GC×GC separation of diesel using DPGM with a reverse-column configuration. (B) Zoom-in of the section between 20 and 25 min. (C) Three-dimensional representation of the separation.

obtained at each p_{w}, normalizing the peak height, and overlaying all of the modulations at different p_{w}. This procedure is shown forP_{M}= 200 ms in Fig. 2.7A for acetone (unretained on ^{1}D and ^{2}D), where full modulation is achieved at $p_{\mathrm{w}} = 15$ ms. Data obtained for p_{w} between 30-50 ms are visually indistinguishable from $p_{\mathrm{w}} = 15$ ms (omitted for clarity).

Under differential flow conditions, the pulse width injected onto ^{2}D is given by

$$^2w_{\mathrm{inj}} = \frac{^1F}{^2F} P_{\mathrm{M}} \tag{2.3}$$

The flow rates $^1F_{\mathrm{ave}}$ and $^2F_{\mathrm{ave}}$ as a function of p_{w} are shown for $p_{\mathrm{w}} = 200$ ms in Fig. 2.7B. The $^2w_{\mathrm{inj}}$ were computed using Eq. (3) and plotted in Fig. 2.7C using both the $^1F_{\mathrm{ave}}$ which varies with p_{w}, and the $^1F_{\mathrm{calc}}$, assumed constant at 1.96 ml/min for all p_{w}. For comparison, the measured $^2w_{\mathrm{b}}$ at each p_{w} are

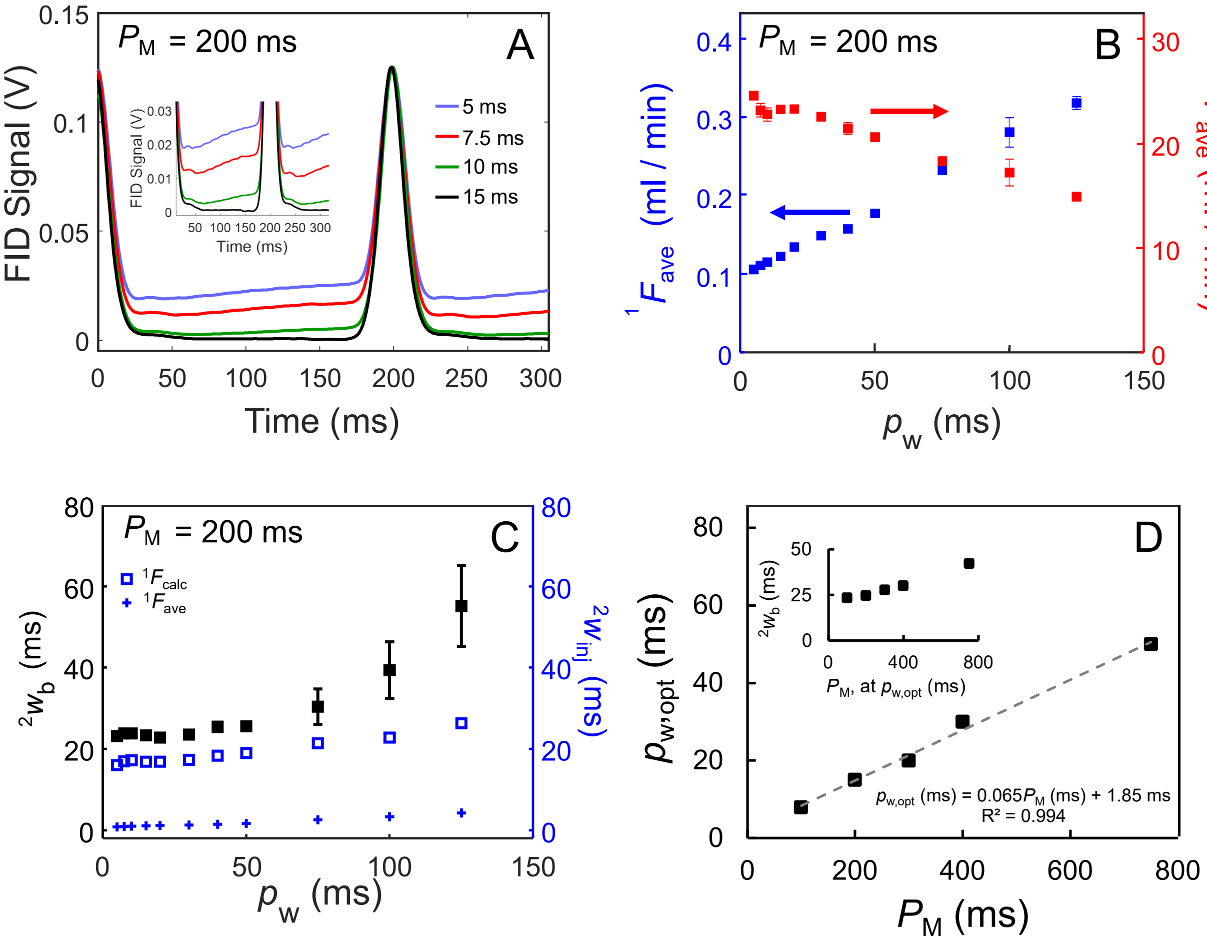


Figure 2.7: (A) Illustration of determining $p_{\mathrm{w,opt}}$ by increasing p_{w} until baseline between modulations is achieved. (B) $^1F_{\mathrm{ave}}$ and $^2F_{\mathrm{ave}}$as a function of p_{w} . (C) $^2w_{\mathrm{b}}$ and $^2w_{\mathrm{inj}}$ as a function of p_{w}. (D) Using the method in (A) $p_{\mathrm{w,opt}}$ is given for each P_{M}. The inset shows the $^2w_{\mathrm{b}}$ obtained using $p_{\mathrm{w,opt}}$ at each P_{M} studied.

also shown in Fig. 2.7C. Using $^1F_{\mathrm{ave}}$, the $^2w_{\mathrm{inj}}$ range from about 1 to 6 ms, markedly lower than the measured $^2w_{\mathrm{b}}$, which also includes on-column band broadening. However, the $^2w_{\mathrm{inj}}$ range determined using $^1F_{\mathrm{calc}}$ more closely follows the trend of $^2w_{\mathrm{b}}$, and moreover, the $^2w_{\mathrm{inj}}$ are reasonably less than $^2w_{\mathrm{b}}$ as well, suggesting that the operative “filling rate” for DPGM is governed by $^1F_{\mathrm{calc}}$ and not $^1F_{\mathrm{ave}}$. Relatively constant peak widths are also obtained (Fig. 2.7C) for a wide p_{w} range (15-50 ms), which is important for application. For $p_{\mathrm{w}} \geq 50$ ms, band broadening due to ^{2}D injection appears to become more of a significant contribution [42]. Finally, Fig. 2.7D displays a calibration of $p_{\mathrm{w,opt}}$ for each P_{M} obtained with the method shown in Fig. 7A. A larger p_{w} is required to achieve full modulation as P_{M} is increased, and as a result, the $^2w_{\mathrm{b}}$ obtained also increase, as shown in the inset of Fig. 2.7D. Based upon our experimental observations, and building upon the previous report by Górecki and co-workers [34], the following qualitative explanation of DPGM is provided. When the auxiliary pressure is off ($p_{\mathrm{w}} = 750$ ms), a “standard” 1D-GC separation is obtained, as in Fig. 2.3B (lower). Interestingly, when the auxiliary pressure is on, but not pulsed ($p_{\mathrm{w}} = 0$ ms), the flow rate is only slightly slowed and a different "standard" 1D-GC separation is obtained, as in Fig. 2.3B (upper). The chromatographic peak patterns in Fig. 2.3B (upper versus lower) differ due to the pressure tuning [37,38] that is being applied at the T-union. When full modulation is achieved, the two steady-state conditions observed in Fig. 2.3B (upper and lower) are never reached. Instead, a dynamic gas phase flow process is set up along the GC×GC column ensemble, functionally centered at the T-union and driven by the P_M and p_w. With full modulation as in Fig. 2.3A (upper), when auxiliary gas flow P_aux is on (during $P_\mathrm{M} - p_\mathrm{w}$), the flow of ^{1}D eluate is temporarily stopped from passing the T-union. Conversely, when the auxiliary gas flow is off (during p_w), the ^{1}D eluate is allowed to pass the T-union onto the ^{2}D column, i.e., producing a modulated injection. When the p_w has ended, and the P_aux is reapplied for P_M– p_w, the modulated injection onto ^{2}D is significantly compressed in time due to a fast ^{2}D flow rate, concurrent with the ^{1}D eluate being stopped from passing the T-union, as the cyclic modulation process continues. We hypothesize that during $P_\mathrm{M} - p_\mathrm{w}$, flow down the entire ^{1}D column may likely not stop, only ^{1}D flow past the T-union has stopped. Due to the release of ^{1}D eluate during p_w, the ^{1}D flow strives to back fill the previously released (modulated) ^{1}D eluate to prepare for the next modulation. In effect, DPGM bares some similarity to previous stopped flow designs [34,43]. It is important to note that in order to provide full modulation during $P_\mathrm{M} - p_\mathrm{w}$, the flow past the T-union only stops temporarily; if the P_M is too long relative to the p_w, the ^{1}D flow from P_inlet (Fig. 2.1) will build up at the T-union and "overflow", causing partial modulation in the negative pulse mode, eg., Fig. 2.2B. As p_w approaches P_M, the ^{1}D flow from P_inlet (Fig. 2.1) is not stopped by P_aux, which never reaches the equilibrium pressure due to the short opening period ($p_\mathrm{w} \ll P_\mathrm{M}$) causing partial modulation in the positive pulse mode, eg., Fig. 2.2D. In a sense, the partial modulation modes represent the cases where the two steady-state conditions in Fig. 2.3B are being subtly perturbed.

2.5 Conclusion

Development of new modulation technology has been an active research area since the invention of GC×GC. Flow modulation is an attractive low-cost option. The differential flow modulation technique presented herein, referred to as dynamic pressure gradient modulation (DPGM), provides full modulation with 100% duty cycle, a fast ^{2}D flow rate for analyte compression in time resulting in exceptional signal enhancement, and if needed, a short modulation period, $P_\mathrm{M} \geq 50$ ms [28]. The design is much simpler than current differential flow modulation approaches, utilizing only a 2-way solenoid valve and a T-union. Preliminary observations are that DPGM is relatively straight forward to implement. We suggest an analyst make an initial selection of$P_\mathrm{M}$and p_w, and then adjust P_aux to ensure temporary stop flow conditions (full modulation) are achieved. Then the analyst may finely tune p_w, if desired, to optimize the separation. For a given PM, a relatively wide range of p_w produces full modulation. As theP_M is increased, the required p_w must also be increased. Further work is necessary to investigate the full potential of this new modulation technique.

2.6 References

(1) Z. Liu, J.B. Phillips, Comprehensive two-dimensional gas chromatography using an on-column thermal modulator interface, *J. Chromatogr. Sci.* **1991**, *29,* 227–231.

(2) E.A. Higgins Keppler, C.L. Jenkins, T.J. Davis, H.D. Bean, Advances in the application of comprehensive two-dimensional gas chromatography in metabolomics, *TrAC - Trends Anal. Chem.* **2018**, *109*, 275–286.

(3) B. Gruber, B.A. Weggler, R. Jaramillo, K.A. Murrell, P.K. Piotrowski, F.L. Dorman, Comprehensive two-dimensional gas chromatography in forensic science: A critical review of recent trends, *TrAC - Trends Anal. Chem.* **2018**, *105*, 292–301.

(4) A.M. Muscalu, T. Górecki, Comprehensive two-dimensional gas chromatography in environmental analysis, *TrAC - Trends Anal. Chem.* **2018**, *106*, 225–245.

(5) S.E. Prebihalo, K.L. Berrier, C.E. Freye, H.D. Bahaghighat, N.R. Moore, D.K. Pinkerton, R.E. Synovec, Multidimensional gas chromatography: Advances in instrumentation, chemometrics, and applications, *Anal. Chem.* **2018**, *90*, 505–532.

(6) M.S. Klee, J. Cochran, M. Merrick, L.M. Blumberg, Evaluation of conditions of comprehensive two-dimensional gas chromatography that yield a near-theoretical maximum in

peak capacity gain, *J. Chromatogr. A* **2015**, *1383*, 151–159.

(7) H.-J. de Geus, J. de Boer, U.A.Th. Brinkman, Development of a thermal desorption modulator for gas chromatography, *J. Chromatogr. A* **1997**, *767*, 137–151.

(8) B. V. Burger, T. Snyman, W.J.G. Burger, W.F. van Rooyen, Thermal modulator array for analyte modulation and comprehensive two-dimensional gas chromatography, *J. Sep. Sci.* **2003**, *26*, 123–128.

(9) P.J. Marriott, R.M. Kinghorn, Longitudinally modulated cryogenic system. A generally applicable approach to solute trapping and mobilization in gas chromatography, *Anal. Chem.* **1997**,*69*, 2582–2588.

(10) E.B. Ledford, C. Billesbach, Jet-cooled thermal modulator for comprehensive multidimensional gas chromatography, *J. High Resolut. Chromatogr.* **2000**, *23*, 202–204.

(11) J. Harynuk, T. Górecki, New liquid nitrogen cryogenic modulator for comprehensive two-dimensional gas chromatography, *J. Chromatogr. A* **2003**, *1019*, 53–63.

(12) C.A. Bruckner, B.J. Prazen, R.E. Synovec, Comprehensive two-dimensional high-speed gas chromatography with chemometric analysis, *Anal. Chem.* **1998**, *70*, 2796–2804.

(13) J.V. Seeley, F. Kramp, C.J. Hicks, Comprehensive two-dimensional gas chromatography via differential flow modulation, *Anal. Chem.* **2002**, *72*, 4346–4352.

(14) A.E. Sinha, B.J. Prazen, C.G. Fraga, R.E. Synovec, Valve-based comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometric detection: instrumentation and figures-of-merit, *J. Chromatogr. A* **2003**, *1019*, 79–87.

(15) C.E. Freye, L. Mu, R.E. Synovec, High temperature diaphragm valve-based comprehensive two-dimensional gas chromatography, *J. Chromatogr. A* **2015**, *1424*, 127–133.

(16) C.E. Freye, R.E. Synovec, High temperature diaphragm valve-based comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry, *Talanta* **2016**, *161*, 675–680.

(17) J.V. Seeley, Recent advances in flow-controlled multidimensional gas chromatography, *J. Chromatogr. A* **2012**, *1255* 24–37.

(18) C. Duhamel, P. Cardinael, V. Peulon-Agasse, R. Firor, L. Pascaud, G. Semard-Jousset, P. Giusti, V. Livadaris, Comparison of cryogenic and differential flow (forward and reverse fill/flush) modulators and applications to the analysis of heavy petroleum cuts by high-temperature comprehensive gas chromatography, *J. Chromatogr. A*, **2015**, *1387*, 95-103.

(19) J.V. Seeley, N.J. Micyus, S. V. Bandurski, S.K. Seeley, J.D. McCurry, Microfluidic deans switch for comprehensive two-dimensional gas chromatography, *Anal. Chem.* **2007**, *79*, 1840–1847.

(20) J.V. Seeley, N.E. Schimmel, S.K. Seeley, The multi-mode modulator: A versatile fluidic device for two-dimensional gas chromatography, *J. Chromatogr. A* **2018**, *1536*, 6–15.

(21) J.B. Phillips, E.B. Ledford, Thermal modulation: A chemical instrumentation component of potential value in improving portability, *F. Anal. Chem. Technol.* bf1996 *1*, 23–29.

(22) M. Edwards, A. Mostafa, T. Górecki, Modulation in comprehensive two-dimensional gas chromatography: 20 years of innovation, *Anal. Bioanal. Chem.* **2011**, *401*, 2335–2349.

(23) P.Q. Tranchida, G. Purcaro, P. Dugo, L. Mondello, G. Purcaro, Modulators for comprehensive two-dimensional gas chromatography, *TrAC - Trends Anal. Chem.* **2011**, *30*, 1437–1461.

(24) J. Beens, M. Adahchour, R.J.J. Vreuls, K. Van Altena, U.A.Th. Brinkman, Simple, non-moving modulation interface for comprehensive two-dimensional gas chromatography, *J. Chromatogr. A* **2001**, *919*, 127–132.

(25) A.M. Muscalu, M. Edwards, T. Górecki, E.J. Reiner, Evaluation of a single-stage consumable-free modulator for comprehensive two-dimensional gas chromatography: Analysis of polychlorinated biphenyls, organochlorine pesticides and chlorobenzenes, *J. Chromatogr. A* **2015**, *1391*, 93–101.

(26) J. Luong, X. Guan, S. Xu, R. Gras, R.A. Shellie, Thermal independent modulator for comprehensive two-dimensional gas chromatography, *Anal. Chem.* **2016**, *88*, 8428–8432.

(27) C.E. Freye, H.D. Bahaghighat, R.E. Synovec, Comprehensive two-dimensional gas chromatography using partial modulation via a pulsed flow valve with a short modulation period, *Talanta*, **2018**, *177*, 142–149.

(28) H.D. Bahaghighat, C.E. Freye, D.V. Gough, R.E. Synovec, Comprehensive two-dimensional gas chromatography and time-of-flight mass spectrometry detection with a 50 ms modulation period, *J. Chromatogr. A* **2018**, *1583*, 117–123.

(29) D.V. Gough, D.H. Song, S. Schöneich, S.E. Prebihalo, R.E. Synovec, Development of ultrafast separations using negative pulse partial modulation to enable new directions in gas chromatography, *Anal. Chem.* **2019**, *91*, 7328–7335.

(30) J. Krupčík, P. Májek, R. Gorovenko, I. Špánik, P. Sandra, D.W. Armstrong, On the determination of a detector response enhancement factor for flow modulated comprehensive two-dimensional gas chromatography, *J. Chromatogr. A* **2013**, *1286*, 235–240.

(31) H. Cai, S.D. Stearns, Partial modulation method via pulsed flow modulator for comprehensive two-dimensional gas chromatography, *Anal. Chem.* **2004**, *76*, 6064–6076.

(32) H.D. Bahaghighat, C.E. Freye, D. V. Gough, P.E. Sudol, R.E. Synovec, Ultrafast separations via pulse flow valve modulation to enable high peak capacity multidimensional gas chromatography, *J. Chromatogr. A* **2018**, *1573*, 115–124.

(33) B. Kehimkar, J.C. Hoggard, L.C. Marney, M.C. Billingsley, C.G. Fraga, T.J. Bruno, R.E. Synovec, Correlation of rocket propulsion fuel properties with chemical composition using comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry followed by partial least squares regression analysis, *J. Chromatogr. A* **2014**, *1327*, 132–140.

(34) N. Oldridge, O. Panic, T. Górecki, Stop-flow comprehensive two-dimensional gas chromatography with pneumatic

switching, *J. Sep. Sci.* **2008**, *31*, 3375–3384.

(35) B.A. Parsons, D.K. Pinkerton, R.E. Synovec, Implications of phase ratio for maximizing peak capacity in comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry, *J. Chromatogr. A* **2018**, *1536*, 16–26.

(36) D.V. Gough, H.D. Bahaghighat, R.E. Synovec, Column selection approach to achieve a high peak capacity in comprehensive three-dimensional gas chromatography, *Talanta* **2019**, *195*, 822–829.

(37) M. Akard, R.D. Sacks, Pressure-tunable selectivity for high-speed gas chromatography, *Anal. Chem.* **1994**, *66*, 3036–3041.

(38) K.M. Sharif, C. Kulsing, P.J. Marriott, Pressure tuning of first dimension columns in comprehensive two-dimensional gas chromatography, *Anal. Chem.* **2016**, *88*, 9087–9094.

(39) B.C. Reaser, B.W. Wright, R.E. Synovec, Using receiver operating characteristic curves to optimize discovery-based software with comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry, *Anal. Chem.* **2017**, *89*, 3606–3612.

(40) M. Adahchour, J. Beens, R.J.J. Vreuls, A.M. Batenburg, U.A.Th. Brinkman, Comprehensive two-dimensional gas chromatography of complex samples by using a “reversed-type” column combination: Application to food analysis, *J. Chromatogr. A* **2004**, *1054*, 47–55.

(41) R.W. LaClair, P.A. Bueno, J.V. Seeley, A systematic analysis of a flow-switching modulator for comprehensive two-dimensional gas chromatography, *J. Sep. Sci.* **2004**, *27*, 389–396.

(42) G.M. Gross, B.J. Prazen, J.W. Grate, R.E. Synovec, High-speed gas chromatography using synchronized dual-valve injection, *Anal. Chem.* **2004**, *76*, 3517–3524.

(43) R.E. Mohler, B.J. Prazen, R.E. Synovec, Total-transfer, valve-based comprehensive two-dimensional gas chromatography, *Anal. Chim. Acta.* **2006**, *555*, 68–74.

2

Chapter 3

TOTAL-TRANSFER COMPREHENSIVE THREE-DIMENSIONAL GAS CHROMATOGRAPHY WITH TIME-OF-FLIGHT MASS SPECTROMETRY

This chapter was adapted from T.J. Trinklein, J.Jiang, R.E. Synovec, *J. Chromatogr. A* **2020**, 1634, 461654.
.

3.1 Introduction

Over the past three decades, comprehensive two-dimensional gas chromatography (GC×GC) has advanced from demonstration to commercialization with broad implementation [1–4]. Today, GC×GC has a demonstrated utility in a diverse range of applications, such as metabolomics [5], forensics [6,7], and environmental samples [8,9]. New technology continues to be developed for GC×GC, including both instrumentation [10–13] and data analysis tools [14–17]. Motivations for employing the technique include increased peak capacity versus one-dimensional GC [18], highly selective and ordered separations, and potential for increased detectability.

Despite the excellent performance of GC×GC, there remains a high probability of component overlap for complex samples [19,20]. One intriguing route to address this issue is comprehensivd three-dimensional gas chromatography, or GC^3 [21–30]. GC^3 is operationally like GC×GC, with the extension that the ^{2}D column is interfaced via a second modulator to a ^{3}D column of sufficiently orthogonal selectivity. A primary motivation for employing GC^3 separation techniques derives from the multiplicative property of peak capacities. Envisage a GC^3 system which maintains a reasonably attainable ^{1}D×^{2}D peak capacity of 4000 to 7000 [18] by adding only a modest ^{3}D peak capacity (^3n_c) of 5; the total ideal peak capacity, which is the product of the peak capacities on each dimension, would be an astounding 20,000 to 35,000. Although the potential analytical benefits warrant more investigation, comprehensive 3D separations remain rare, with a few exceptions [31,32]. Notably, several 3D (but not comprehensive) GC platforms have been described such as "smart" 3D-GC [33] and sequential hybrid 3D-GC [34]. These innovative systems are highly selective and provide a powerful approach to targeted analysis, but do not comprehensively sample both ^{1}D and ^{2}D and therefore do not provide a significant peak capacity advantage over GC×GC.

Certainly, the high chemical selectivity and peak capacity provided via GC^3 is best utilized in combination with an information-rich detector, such as time-of-flight mass spectrometry (TOFMS). A GC^3-TOFMS instrument was introduced in 2017, however; the duty cycle was low and the peak capacity was not substantially higher than optimized G×GC systems [24]. Recent instrumental developments should enable more robust and powerful GC^3-TOFMS platforms to be developed, as will be demonstrated here. The first GC^3-TOFMS instrument utilized a diaphragm valve to modulate between ^{1}D and ^{2}D (^{1}M) and a thermal modulator (cryogenic quad-jet) to modulate from ^{2}D to ^{3}D (^{2}M). While sufficient for a proof-of-concept study, the diaphragm valve only transferred a small portion (~15%) of the ^{1}D effluent to the subsequent dimensions, impairing detectability relative to if total-transfer modulation were applied. Moreover, flow modulation is better suited as ^{2}M with the thermal modulator used as ^{1}M, due to the generally higher speeds at which flow modulators can operate [22,35], which is important to adequately sample narrow ^{2}D peaks. Recently, we introduced a GC^3-TOFMS platform which used "partial" modulation for ^{2}M, wherein only a portion of the ^{2}D effluent is modulated [28]. The instrument produced an exceptional 3D peak capacity of 31,500; however, the use of partial modulation sacrificed detectability (^{2}M duty cycle 10%) and required additional data processing steps in order to visualize the data in a common format and to perform quantification [35].

To address these issues, we describe a total-transfer GC^3-TOFMS system (100% duty cycle for ^{1}M and ^{2}M), based upon the facile modification of a commercial thermally modulated GC×GC-TOFMS instrument. The stock thermal modulator generates standard ^{1}D×^{2}D separations, with the ^{2}D peaks modulated via differential flow modulation to produce the ^{3}D separations. Specifically, a variant of differential flow modulation termed dynamic pressure gradient modulation (DPGM) is employed [36-38].

DPGM operates by pulsing an auxiliary pressure (P_{aux}) to a junction separatin two separation columns. When the off time of the valve (p_{w}) is optimized for a given modulation period (P_{M}), a dynamic pressure gradient is setup along the column ensemble which temporarily stops flow just prior to the junction while providing fast flow along the separation leading to the detector. A major benefit of DPGM is the ability to operate reliably at extremely fast modulation periods ($P_{\mathrm{M}} \geq 50$ ms), enabling an adequate sampling density for modulated ^{2}D peaks in GC^3. Moreover, DPGM remains highly effective at MS-compatible flow rates of 4-6 ml/min, enabling direct connection to the TOFMS without flow splitting [37,38]. Importantly, DPGM provides total-transfer (100% duty cycle); thus, its hyphenation with thermal modulation in a GC^3-TOFMS instrument provides total-transfer between all dimensions.

To validate the assumption that total-transfer is obtained, area measurements of analytes from unmodulated chromatograms (where modulators ^{1}M and ^{2}M are both turned off) are compared to the sum of the areas of all ^{3}D peaks for a given analyte from a 3D chromatogram (where both modulators are enabled). Separations of a test mixture are compared in terms of peak widths and the usage of 3D space for two column sets to provide insight into tuning and optimizing the column selection for GC^3. Finally, the utility of GC^3-TOFMS for complex sample analysis is investigated via separations of derivatized porcine serum, and a jet fuel spiked with sulfur containing compounds.

3.2 Experimental

The total-transfer GC^3-TOFMS instrument was based upon a modified Pegasus 4D GC × GC-TOFMS (LECO Corporation, St. Joseph, MI) with an integrated Agilent 6890 gas chromatograph (Agilent Technologies, Santa Clara, CA, U.S.A.) and a 4D thermal modulator upgrade (LECO Corporation, St. Joseph, MI, USA). The column sets and various experimental parameters are given in Table 3.1. The ^{1}D and ^{2}D columns were interfaced via the thermal modulator and an Ultimate Union (Agilent Technologies, Santa Clara, CA, U.S.A.). The end of the ^{2}D column was connected via a microvolume T-union (Model MT.5CXS6, Valco Instruments Company Inc., Houston, TX, USA) to the ^{3}D column as part of the DPGM assembly [38]. A schematic of the instrument is given in Fig. 3.1A. A high-speed 2-way solenoid "pulse"nvalve (Model 009–0347–900, Parker Hannifin, Hollis, NH) was interfaced by an in-house fabricated stainless-steel plate and connected to the T-union via stainless steel connection tubing (5 cm L × of 127 µm i.d., 800 µm o.d.). A schematic of the exterior of the pulse valve and custom interface is shown in Fig. 2.1B. The opening to the pulse valve was reamed to 800 µm to provide a secure fitting with the connection tubing. A circular stainless-steel plate was machined to a 1/32" union and screwed onto the pulse valve base. The connection tubing was held in place using a 1/32" screw and stainless-steel ferrules ferrules (Valco Instruments, ZF.5S6-10). A

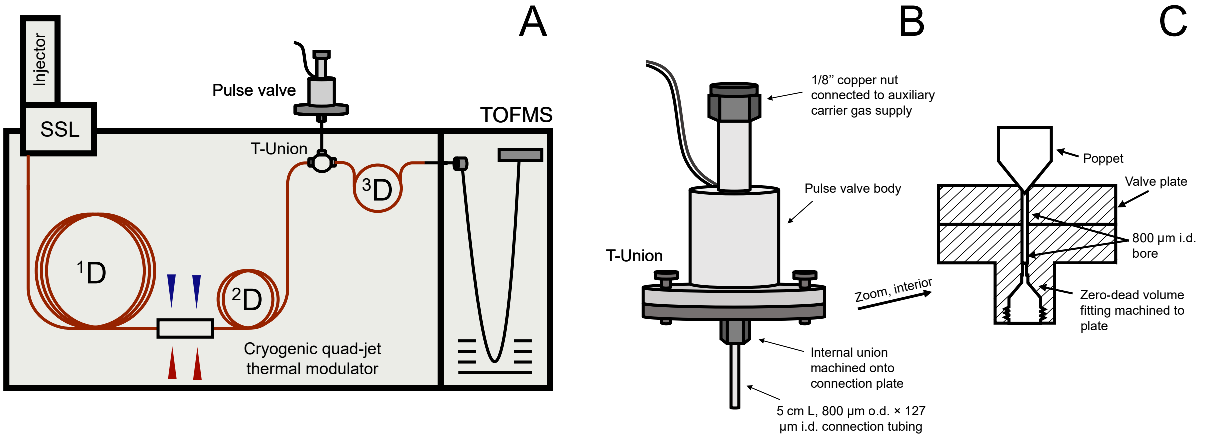


Figure 3.1: (A) Schematic diagram of the GC^3-TOFMS instrument. (B) Components of the DPGM assembly. (C) Cross-section view of the pulse valve and interface.

cross section of the flow modulator assembly is given in Fig. 3.1C. The pulse valve was powered via an in-house made solenoid driver and controlled using a LabVIEW program synchronized with GC injection and thermal modulation using the remote-out port of the Agilent 6890. Actuation of the pulse valve was delayed for 15 s after injection to ensure that the injection process was not affected by the pressure fluctuations resulting from applying DPGM.

The GC^3-TOFMS instrument was evaluated with a 90-component test mixture containing a moderate degree of sample dimensionality (Table A1). Each component was diluted to one part-per-thousand by volume (ppth) in methanol. Aporcine serum sample (Sigma Aldrich, Milwaukee WI) was derivatized by trimethylsilation and methoximation according to as described elsewhere [39]. A jet fuel was spiked with sulfur-containing species with each spiked compound at 200 ppm in the final mixture. For all experiments, 0.2 µl of liquid mixture was injected at split ratio of 40:1. The "first" modulation period, $^1P_\mathrm{M}$ was 4 s the "second" was of 250 ms. Run-specific method parameters for all sample analyses are provided in Table 3.1. The auxiliary pressure programs applied to the pulse valve (P_aux) were chosen to provide a flow rate exiting the ^{3}D column of ~6 ml/min throughout the course of the run. To accomplish this the P_aux was manually varied until the vacuum pressure in the flight tube reached the calibration point for ~6 ml/min. Despite the relatively high ^{3}D flow rate, no overheating of the turbomolecular pumps or any discernable effect on the mass spectral quality were observed, with this flow rate successfully employed previously with GC×GC-TOFMS with DPGM [39]. The initial and final temperature and P_aux set points, along with the corresponding ramp rates, are provided in Table 3.1. Note that the flow rates on ^{1}D, ^{2}D, and ^{3}D were essentially equivalent for the two column set combinations for the test mixture, along with an identical temperature program, to facilitate comparison. Although the P_inlet program was set to constant flow mode at 0.5-0.8 ml/min for the three conditions, the application of DPGM slows the flow along ^{1}D, therefore we report the average linear flow velocity in Table 3.1 via measurement of the dead time and known column length. The TOFMS collected a mass range of 33 to 256 m/z at 200 Hz. This mass range was selected to provide adequate identification of analytes in the test mixture (confirmed with retention time matching) as well as enough selective m/z for visualization purposes in the real samples, without introducing more noise to the TIC. An ionization energy of 70 eV was used with a detector

Condition(s) studied	**Column Set 1**	**Column Set 2**	**Column Set 3**
^{1}D Column	60 m DB5 250 µm d_c × 0.25 µm d_f	60 m DB5 250 µm d_c × 0.25 µm d_f	30 m RTX-17 250 µm d_c × 0.25 µm d_f
^{2}D Column	3.5 m DB-Wax 100 µm d_c × 0.1 µm d_f	4.5 m DB-Wax 100 µm d_c × 0.1 µm d_f	4.5 m Rxi-1 100 µm d_c × 0.1 µm d_f
^{3}D Column	1 m Rtx-200 180 µm d_c × 0.18 µm d_f	2.5 m IL-111 250 µm d_c × 0.25 µm d_f	2.5 m IL-111 250 µm d_c × 0.25 µm d_f
Temperature program	30 °C for 5 min, ramp at 4 °C/min to 280 °C for 10 min	30 °C for 5 min, ramp at 4 °C/min to 280 °C for 10 min	30 °C for 5 min, ramp at 4 °C/min to 240 °C for 15 min
Inlet flow setting	0.8 ml/min	0.7 ml/min	0.5 ml/min
Average ^{1}D Linear Flow Velocity	0.080 m/s	0.075 m/s	0.052 m/s

Table 3.1: Experimental parameters for the three column sets used.

voltage of 1562 V. The resulting data were imported into MATLAB 2019a (The Mathworks, Inc., Natick, MA, U.S.A.) using an in-house written data converter for all subsequent analyses. All chromatograms were baseline-centered around zero before curve fitting or area measurements.

3.3 Results and discussion

The GC^3-TOFMS instrument was initially assessed using Column Set 1, a nonpolar × polar × midpolar combination, which provided ample usage of 3D space in the first GC^3-TOFMS report [24]. Since the instrument produces a fourth-order data tensor pervreplicate, simultaneous visualization of all chromatographic dimensions and the mass spectral dimension is not possible. Therefore, the set of images in Fig. 3.2 is provided, in which each subfigure shows a different "view" of the separation. Starting with Fig. 3.2A, the matrix dimensions corresponding to ^{2}D, ^{3}D and the *m/z* dimension are summed to create a reconstructed 1D total ion current (TIC) chromatogram, which is a 1D array. This view illustrates the resolution achieved if the sample were analyzed using only the first chromatographic dimension and univariate detection operating at a data collection rate of one point every 4 s (i.e., $^1P_{\mathrm{M}}$, the modulation period from ^{1}D to ^{2}D). The inset, showing representative analyte pristane, displays the appearance of an individual peak in this reconstructed chromatogram.

In Fig. 3.2B an isosurface TIC chromatogram is shown, in which data points at a user-selected intensity are connected into peak volumes. This view allows the analyst to view the distribution of compounds within the 3D space but sacrifices intensity and *m/z* information. To complement the information provided in the isosurface plot, an analyst can sum up any one chromatographic dimension to view a two-dimensional chromatographic plane. For instance, the ^{3}D dimension has been summed to produce the ^{1}D×^{2}D chromatogram in Fig. 2.2C, which is akin to a to a traditional GC×GC chromatogram. Ten analytes were selected which span a broad range of chemical functionalities and will be used subsequently to discuss figures-of-merit (Table 3.2). Similarly, the ^{1}D dimension can be summed up to view the ^{2}D×^{3}D separation. However, the ^{2}D×^{3}D chromatogram has limited peak capacity relative to the

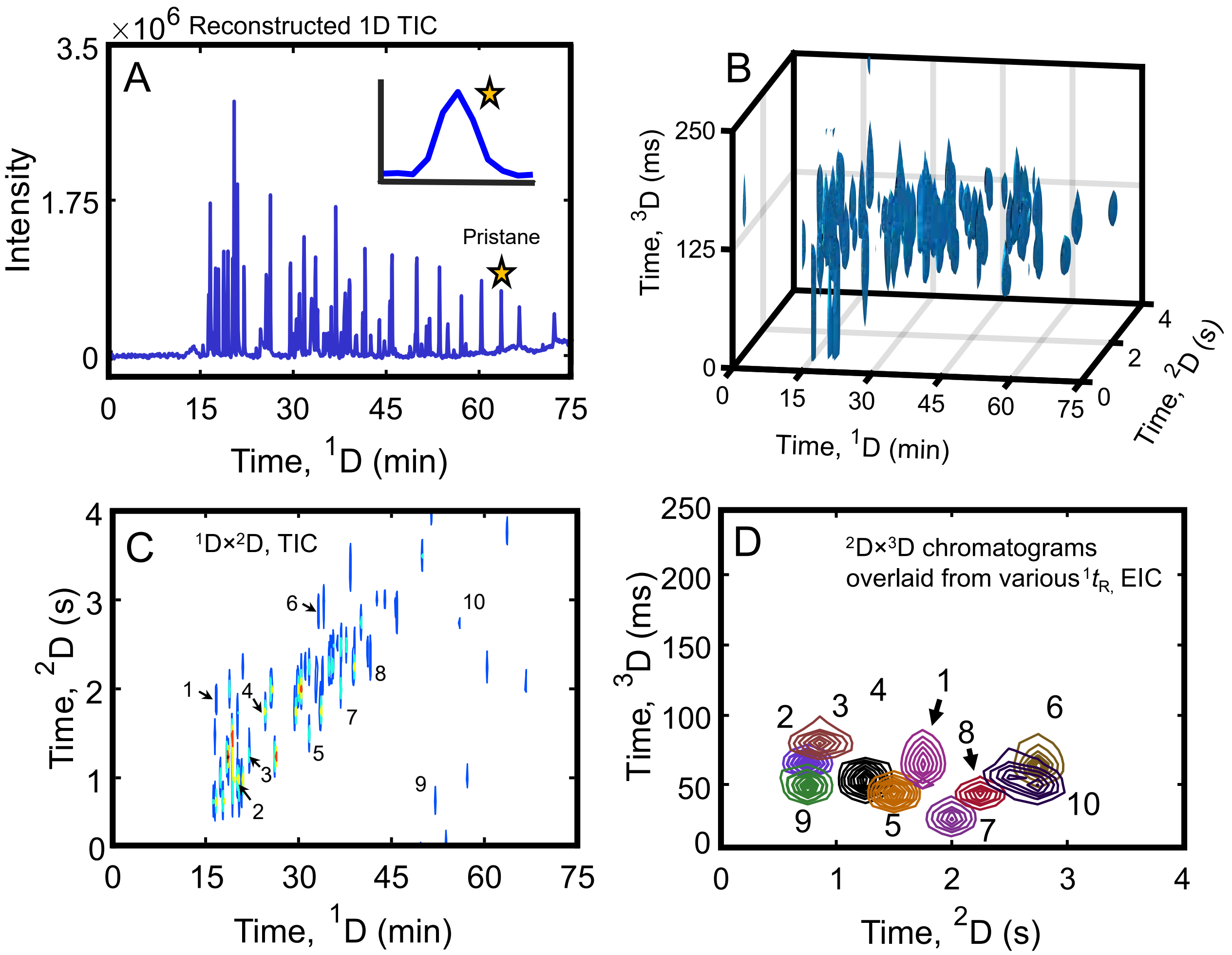


Figure 3.2: (A) Reconstructed 1D TIC chromatogram of the 90-component test mixture using Column Set 1. The starred peak and inset provide a close-up view of pristane. (B) TIC isosurface 3D chromatogram. (C) View of the ^{1}D $\times^2$ D separation plane. ^{a}Due to low signal intensity in the TIC, analyte 10 is not visible. (D). Series of overlays of ^{2}D $\times$ ^{3}D EIC "slices" containing the same 10 analytes.

^{1}D$\times^2$D or ^{1}D$\times^3$D planes and will contain numerous overlapping peaks if viewed over the entire separation time. To better visualize the ^{2}D$\times^3$D plane, it is more effective to view specific section(s) along ^{1}D. For example, one can view several ^{2}D$\times^3$D "slices" (i.e., ultra-fast GC×GC chromatograms) taken from various, not necessarily sequential, modulations from ^{1}D. Figure 3.2D was created by excising and overlaying various ^{2}D$\times^3$D slices containing the 10 analytes indicated in Fig. 2.2C at various ^{1}D retention times ($^1t_{\mathrm{R}}$) and using selective m/z. From Figs. 2.2C and 2.2D, it is apparent that column set 1, a nonpolar × polar × midpolar configuration, does not effectively utilize the 3D separation space. We hypothesize this is a result of the fast flow along ^{3}D as a consequence of employing differential flow modulation as ^{2}M. The suboptimal usage of separation space will be subsequently addressed with a better column set. Nonetheless, the separation in Fig. 3.2 is well suited to illustrate fundamental principles regarding the total-transfer GC^3-TOFMS instrument, such as conservation of area.

To validate the assumption that total-transfer was obtained between all three separation dimensions, area measurements were performed on extrac-

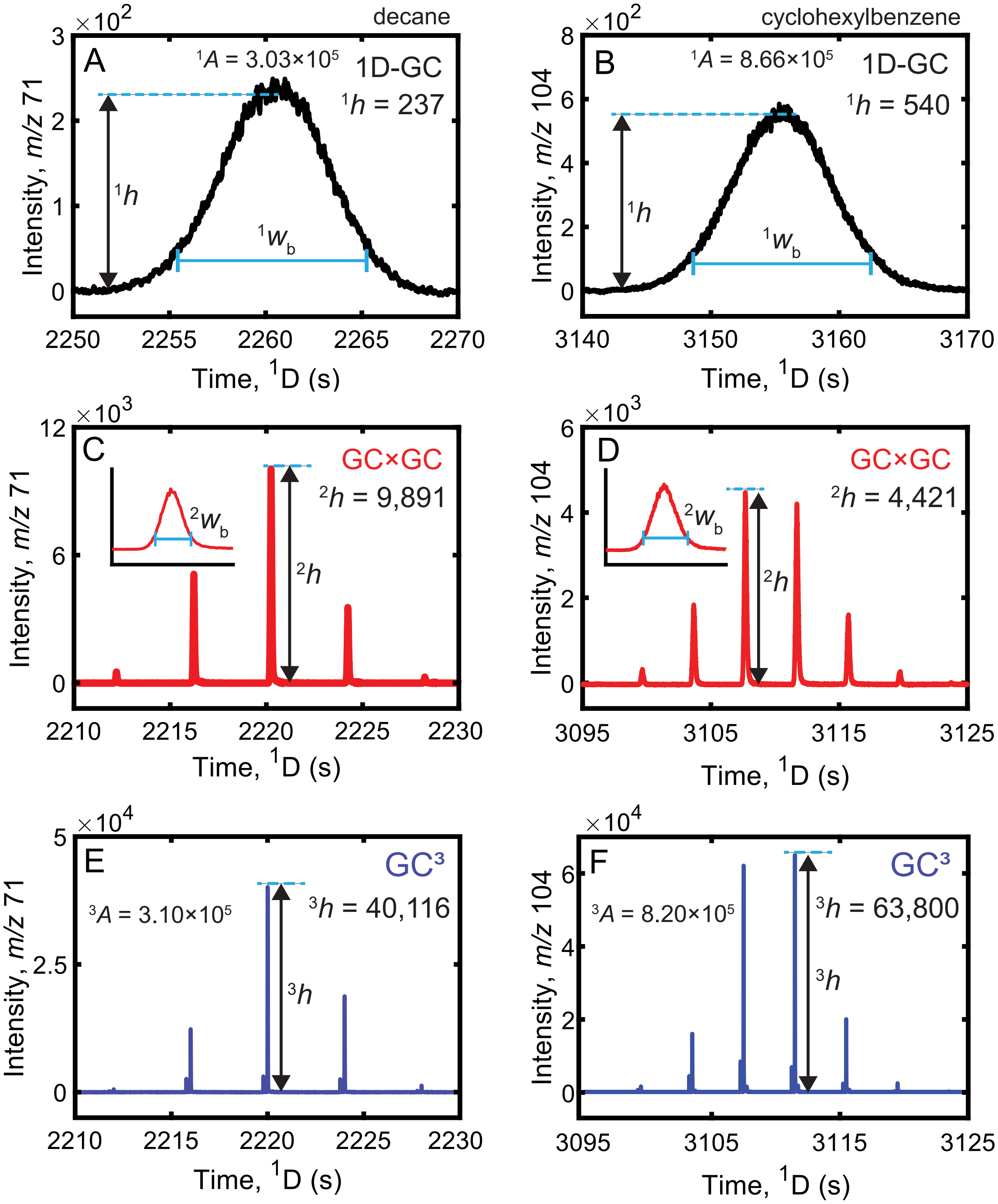


Figure 3.3: (A) 1D-GC data for decane, wherein neither modulator is enabled. Measurements of ^1h and $^1w_\mathrm{b}$ are illustrated. (B) 1D-GC data for cyclohexylbenzene. (C) Unfolded GC×GC data for decane, where thermal modulation is enabled but DPGM is not. Measurement of ^2h is indicated. The insert demonstrates measurement of $^2w_\mathrm{b}$ from the tallest ^{2}D peak. (D) Unfolded GC×GC for cyclohexylbenzene with ^2h and $^2w_\mathrm{b}$ similarly indicated. (E) Unfolded GC^3 data for decane, wherein both modulators are enabled, with measurement of ^3h from the tallest 3D peak indicated. (F) Unfolded GC^3 data for cyclohexylbenzene.

ed ion chromatograms (EICs) by summing all of the ^{3}D peaks for a given analyte (^3A), and comparing ^3A to the ^{1}D peak area (^1A) peak, obtained from unmodulated chromatograms wherein modulators ^{1}M and ^{2}M are both turned off, for the 10 analytes indicated in Fig. 3.2C-D. To illustrate the relevant area measurements, the raw data for two representative analytes with varying degrees of modulation are provided in Fig. 2.3 for decane and cyclohexylbenzene. Note that the P_{aux} program in all runs was the same to ensure the flow rate on each dimension was not significantly different. Fig. 3.3A-B shows raw ^{1}D data, wherein neither modulator is enabled for decane and cyclohexylbenzene. Measurements of ^1A were obtained by numerical integration via the trapezoid rule of the ^{1}D data following baseline correction. Fig. 3.3C-D displays the ^{1}D×^{2}D data for decane and cyclohexylbenzene wherein thermal modulation (^{1}M) is enabled but DPGM (^{2}M) is not. (^{2}M) is not. The sum of areas of the ^{2}D peaks, ^2A, must be equivalent to ^1A, and in turn the sum of the ^{3}D peaks, ^3A, must be equal ^2A, if the GC^3 separations provide 100% duty cycle for ^{1}M and ^{2}M. Thus, we simply need to determine if ^1A is equal to ^3A, to demonstrate 100% duty cycle. To do so, Fig. 3.3E-F displays the raw (unfolded) GC^3 data wherein modulators ^{1}M and ^{2}M are both enabled. To minimize the potential for baseline error, measurement of ^3A is obtained by folding the GC^3 data section along $^2P_{\text{M}}$ and adding all the modulations together to create a single ^{3}D peak. Next, the summed ^{3}D peak is baseline corrected and integrated via the trapezoid rule. Then, the %Difference between ^1A and ^3A is calculated as

$$\%\text{Difference} = \frac{^3A -^1 A}{^1A} \times 100 \tag{3.1}$$

If total-transfer is achieved, the %Difference should equal zero within reasonable experimental error. Percent difference values are given in

	Analyte	m/z	$^1t_{\text{R}}$ **(min)**	^1A	^3A	**%Difference**	^3_1SE
1	2-butanol	59	16.6	–[a]	1.75×10^4	–[a]	157
2	2,2,4-trimethylpentane	57	20.2	2.21×10^6	2.08×10^6	-11%	152
3	3,4-dimethylhexene	39	21.7	6.89×10^5	6.79×10^5	-1%	–[a]
4	toluene	91	24.5	–[a]	1.23×10^6	–[a]	–[a]
5	nonane	57	31.5	1.10×10^6	1.15×10^6	7%	154
6	anisole	108	33.0	1.09×10^6	1.16×10^6	7%	65
7	decane	71	36.7	3.03×10^5	3.10×10^5	2%	168
8	undecane	57	41.4	1.23×10^6	9.80×10^6	-4%	177
9	cyclohexylbenzene	104	51.8	8.66×10^5	8.20×10^5	-5%	118
10	diphenylmethane	167	55.8	–[a]	1.16×10^6	–[a]	–[a]
	Average					-2.31% ± 7.60%	130 ± 47

Table 3.2: Chromatographic peak area, %Difference, peak height, and *SE* measurements for 10 compounds representing a variety of chemical functionality in the 90-component test mixture separation using column set 1. –[a] Not measured due to coelution

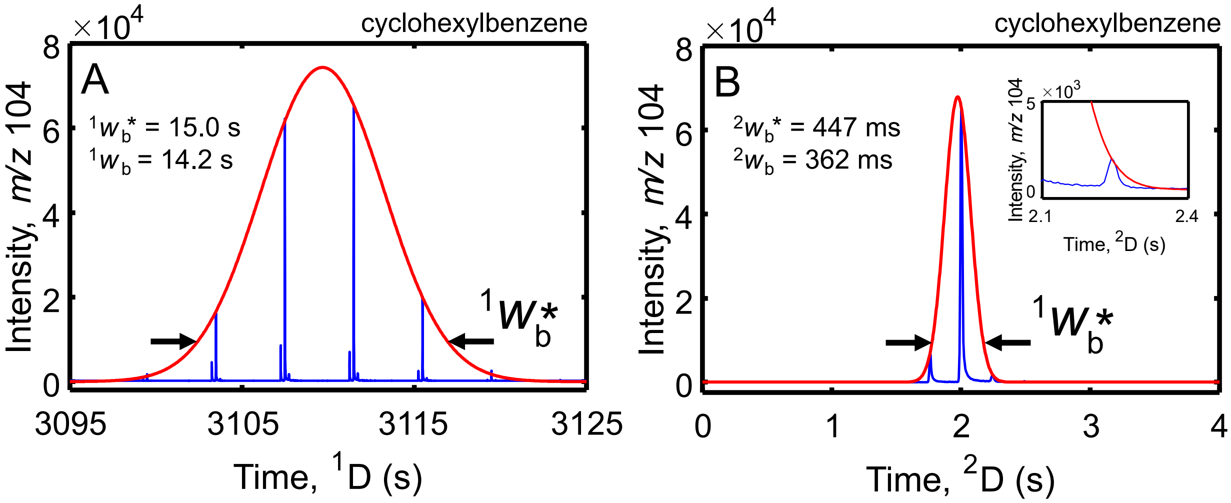


Figure 3.4: Illustration of measurement of $^1w_b^*$ and $^2w_b^*$. (A) Gaussian curve fitting over the unfolded GC^3 data for cyclohexylbenzene, with $^1w_b^*$ obtained at 13.5% of the Gaussian peak height (±2 s.d.). (B) Gaussian curve fit to the middle-right set of ^{3}D peaks which comprise a ^{2}D peak to obtain $^2w_b^*$. The $^1w_b^*$ and $^2w_b^*$ are comparedto ^1w_b and ^2w_b obtained from the unmodulated chromatograms, as shown in Fig. 3.3.

Table 3.2 for duplicate measurements at each degree of modulation. An average %Difference of -2.31% ± 7.60% (1 s.d) was obtained, corroborating the assumption of total-transfer with adequate precision and without significant bias. Note that due to coelution on ^{1}D for the 1D separation for 3 of the 10 analytes, each overlapping with different interfering analytes, the %Difference is not given for these 3 analytes: 2-butanol, toluene and diphenylmethane. One of the most striking features of Fig. 3.3 is the successive gain in signal enhancement obtained by enabling successive stages of modulation. Indeed, the intensity scale in Fig. 3.3E-F for GC^3 data is about two orders of magnitude higher than that in Fig. 3.3A-B for ^{1}D data. Signal enhancement is an established benefit of modulation, since a higher signal amplitude enhancement results in a higher signal-to-noise ratio (S/N), and has been treated both theoretically and experimentally for 1D-GC and GC×GC [40-43]. However, the effect of tandem modulation on signal enhancement, as in GC^3, has not been explored and will be introduced herein. First, we define the signal enhancement ${}^x_iSE$ as the ratio of the tallest peak in the x$^{\text{th}}$ dimension originating from the tallest peak in the i$^{\text{th}}$ dimension. Thus, ${}^2_1SE$ is the ratio of a tallest modulated ^{2}D peak (^2h) to the ^{1}D peak height (^1h). Likewise, ${}^3_2SE$ is the ratio of the tallest modulated ^{3}D peak (^3h) to the tallest ^{2}D peak height (^2h). Importantly, the product of successive amplitude enhancements ${}^2_1SE$ and ${}^3_2SE$ yields the overall 3D signal enhancement, ${}^3_1SE$,

$$ {}^3_1SE = \frac{^2h}{^1h}\frac{^3h}{^2h} = \frac{^3h}{^1h} \tag{3.2} $$

This principle is illustrated for decane and cyclohexylbenzene in Fig. 3.3. The ${}^2_1SE$ for decane is $^2h/^1h = 42$, and the ${}^3_2SE$ is $^3h/^2h = 4.0$, so via Eq. (3.2) the ${}^3_1SE$ is 168. Relative to decane, cyclohexylbenzene has a lower ${}^2_1SE$ of $^2h/^1h = 8.2$, but a higher $^3h/^2h = {}^3_2SE$ of 14.4. Overall, cyclohexylbenzene still attained an excellent ${}^3_1SE$ of 118. The height measurements and ${}^3_1SEs$ for the 10 analytes shown in Fig. 3.2C-D are provided in Table 3.2, with ${}^3_1SEs$ ranging from 67 – 177, with an average of 130 ± 47. As a result of the multiplicative property the ${}^3_1SE$s are markedly higher than ${}^2_1SE$s obtained with GC×GC [43]. Note that modulation phasing plays an important role in the observed ${}^3_1SE$. This is apparent in Fig. 3.3E-F, with decane modulated in-phase between ^{1}D and ^{2}D and cyclohexylbenzene modulated out-of-phase between ^{1}D and ^{2}D, and decane obtaining the higher ${}^3_1SE$ (168 versus 118). Although we limited the discussion of *SE* to illustrative examples, further investigation beyond the scope of this study is warranted, on the sampling phase aspect for example, to better understand how to maximize the *SE*.

We next turn our attention to the measurements of peak widths in order to facilitate a discussion of instrument performance, including modulator-induced band broadening. Using the 1D-GC and GC×GC chromatograms (i.e., as in Fig. 3.3A-D), measurements of the true ^{1}D and ^{2}D peak widths at ±2 standard deviations (s.d.) were obtained before modulation, designated as $^1w_\mathrm{b}$ and $^2w_\mathrm{b}$. The average $^1w_\mathrm{b}$ was 12.90 s ± 3.46 s, while $^2w_\mathrm{b}$ was 324 ms ± 116 ms. The higher range in $^2w_\mathrm{b}$ is due to the pseudo-isothermal nature of the ^{2}D separations. Due to the exemplary performance of the DPGM as ^{2}M, an average $^3w_\mathrm{b}$ of 28 ms ± 3 ms was obtained. However, in GC^3 the observed ^{1}D and ^{2}D peak widths are reconstructed from the ^{2}D and ^{3}D separations. This leads to undersampling, and hence is a source of peak broadening. Therefore, the more releveant measurements are the *effective* peak widths on ^{1}D and ^{2}D, which account for sampling-induced broadening, designated as $^1w_\mathrm{b}{}^*$ and $^2w_\mathrm{b}{}^*$. A comparison of the true and effective peak widths allows the analyst to assess the extent to which modulator-induced band broadening is diminishing the performance of the system.

The effective peak widths were obtained by fitting Gaussian profiles to the tops of the modulated peaks, as illustrated in Fig. 3.4 for cyclohexylbenzene. In Fig. 3.4A, $^1w_{*\mathrm{b}}$ is obtained as ±2 s.d. of a Gaussian curve fit to the tops of the ^{3}D peaks. In Fig. 3.4B, a set of ^{3}D peaks from Fig. 3.4A comprising one ^{2}D peak is highlighted to illustrate the measurement of $^2w_\mathrm{b}^*$. Given an average $^1w_\mathrm{b}^*$ of 14.40 s and an average $^1w_\mathrm{b}$ of 12.90 s, the ^{1}D peaks were broadened on by 11.6% due to sampling-induced broadening. Similarly, the average $^2w_{*\mathrm{b}}$ of 425 ms and $^2w_\mathrm{b}$ of 339 ms indicates a a 25.7% increase in width due to sampling of the ^{2}D peaks. The higher average increase of $^2w_{*\mathrm{b}}$ from $^2w_\mathrm{b}$ (compared to the increase in $^1w_\mathrm{b}^*$ from $^1w_\mathrm{b}$) results from the fact that ^{1}D was sampled at a higher frequency than ^{2}D ($^1\rho >^2 \rho$) [44], visually evident in Fig. 3.4, where the sampling density (ρ) is defined as $\rho =^1 w_\mathrm{b}/P_\mathrm{M}$.

While the test mixture separation employing Column Set 1 used to prepare Figs. 3.2-4 was sufficient for demonstrating proof-of-concept of total-transfer and illustrating *SE* for GC^3-TOFMS separations, the fractional coverage of the separation space was low. As a result, the peak capacity was poor despite the narrow peak widths obtained. In order to increase the fractional coverage, a different column set was installed. Although a very similar column configuration to Column Set 1 was used in the first report of GC^3-TOFMS with adequate fractional coverage [24], the instrument used thermal modulation for ^{2}M. Since a differential flow modulator was used herein for ^{2}M, we hypothesize that the faster flow rate along ^{3}D (~6 ml/min) resulted in poor usage of 3D space due to limited retention of analytes on ^{3}D.

Gas chromatography, in contrast to liquid chromatography, has fewer commercial stationary phase and retention mechanism options that can be exploited. Nonetheless, experimental and theoretical studies have demonstrated highly orthogonal 3D separations are attainable using "standard" GC stationary phases. Aside from the column configuration used in our 2017 report [24], upon which Column Set 1 in the present study is based upon, several other combinations have been evaluated. For instance, a recent thermodynamics-guided study suggested excellent use of 3D space could be achieved using a Rtx-5 column on ^{1}D (5% diphenyl-dimethylpolysiloxane), an Rtx-200 column on ^{2}D (trifluoropropylmethylpolysiloxane), and an Rxi-17 Sil MS on ^{3}D (50% diphenyl-50%dimethylpolysiloxane) [45]. In our initial GC^3-FID report, a highly orthogonal 3D separation was achieved using a DB-5 column on ^{1}D, an Rtx-200 on

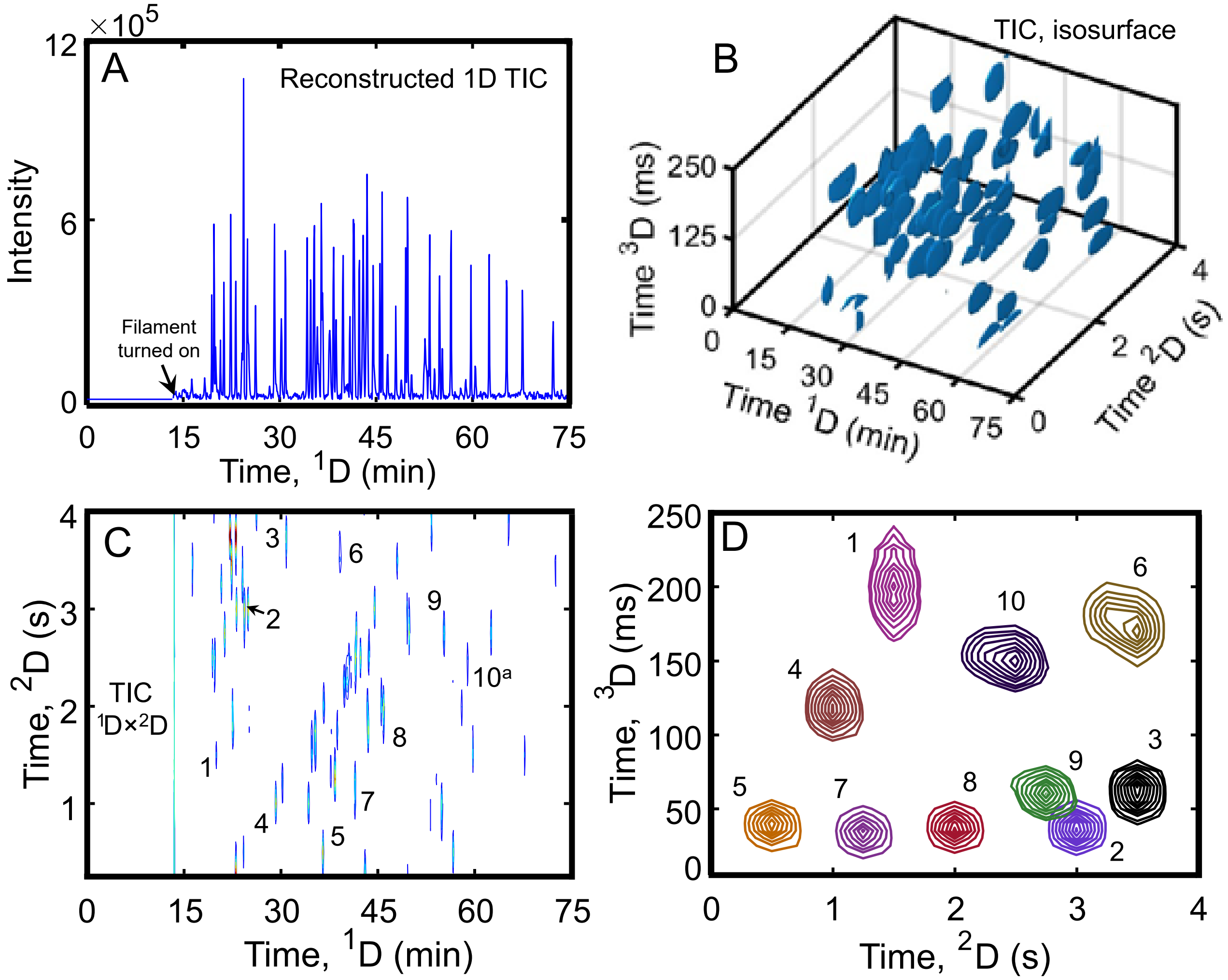


Figure 3.5: Reconstructed 1D TIC chromatogram of the 90-component test mixture using Column Set 2. (B) TIC isosurface 3D chromatogram, highlighting improved usage of separation space relative to Fig. 3.2. (C) View of the ^{1}D×^{2}D separation plane obtained by summing ^{3}D and the m/z dimensions. The same 10 analytes indicated in Fig. 3.2C are labeled. (D). Series of overlays of ^{2}D×^{3}D "slices" with the 10 model analytes labeled, demonstrating improved usage of the ^{2}D×^{3}D space relative to Column Set 2, as in Fig. 3.2D.

^{2}D, and a DB-Wax ^{3}D [24]. Additionally, optimal use of the 3D space with any column combination requires an adequate k' range is obtained on the ^{3}D separation. To this end, we elected to use a longer ^{3}D column (2.5 m versus 1 m with Column Set 1) with a more retentive ionic liquid phase, resulting in Column Set 2 (Table 3.1). Previously, ionic liquid columns have been employed as the ^{2}D column in GC^3 with excellent results [23,29] as well as in sequential hybrid 3D-GC [34] though their usage as the ^{3}D column in GC^3 has not yet been explored. Additionally, a longer ^{2}D column of the same phase as with Column Set 1 was employed with Column Set 2 to increase use of the ^{1}D×^{2}D separation space (Table 3.1). Using Column Set 2, the complete set of images in Fig. 3.5 were prepared in the same manner as those in Fig. 3.2 with Column Set 1 to facilitate performance comparison of the two column configur-

	Analyte	m/z	$^1t_\mathrm{R}$ (min)	$^1w_\mathrm{b}^*$ (s)	$^2w_\mathrm{b}^*$ (ms)	$^3w_\mathrm{b}$ (ms)	$n^*_\mathrm{c,3D}$
1	2-butanol	59	19.9	12.44	565	75	6,800
2	2,2,4-trimethylpentane	57	24.3	11.79	293	36	29,000
3	3,4-dimethylhexene	39	26.1	11.68	308	37	27,000
4	toluene	91	29.0	13.84	294	46	19,200
5	nonane	57	36.3	11.4	368	31	27,300
6	anisole	108	37.5	15.2	659	52	6,900
7	decane	71	41.3	11.94	283	33	32,300
8	undecane	57	45.8	11.92	295	33	31,000
9	cyclohexylbenzene	104	55.2	13.56	565	38	12,400
10	diphenylmethane	167	58.8	16.9	706	44	6,900
	Average ± 1 s.d.	-	-	13.07 ± 1.80	434 ± 170	-2.31% ± 7.60%	19,900 ± 10,700

Table 3.3: Peak width and peak capacity measurements for the same compounds in Table 3.2, using Column Set 2.

ations. Note that the separation parameters were essentially the same between both column sets, with the P_aux program adjusted slightly to ensure the same flow conditions were obtained on ^{2}D and ^{3}D. Figure 3.5A displays the 1D separation of the test mixture by summing the ^{2}D and ^{3}D chromatographic dimensions and m/z dimension onto ^{1}D. The improvement using Column Set 2 is clear in the isosurface TIC chromatogram in Fig. 3.5B, with compounds well distributed about the 3D space. In Fig. 3.5C, markedly better usage of the ^{1}D×^{2}D space is apparent, likely as a result of the longer ^{2}D column, increasing analyte retention times on ^{2}D. Figure 3.5D is prepared with the same 10 analytes as in Fig. 3.2D, illustrating a marked improvement in the distribution of the analytes within the ^{2}D×^{3}D space.

Peak measurements of the 10 analytes depicted in Fig. 3.5C-D are given in Table 3.3. It was observed that $^1w_\mathrm{b}^*$ and $^2w_\mathrm{b}{}^*$ remain effectively unchanged, though there is an increase in the magnitude and spread of $^3w_\mathrm{b}$ due to the longer and more retentive ^{3}D column, yielding a greater ^3k' range.

Given the fractional coverage, we report the peak capacity of this separation as a benchmark to compare against previous implementations of GC^3. The *ideal* peak capacity of a comprehensive 3D separation is explained in Chapter 1.1 as

$$n_\mathrm{c,3D} = {}^1n_\mathrm{c} \times {}^2n_\mathrm{c} \times {}^3n_\mathrm{c} \tag{3.3}$$

where the peak capacity for a given dimension is the separation time divided by the peak width-at-base (±2 s.d.) at $R_\mathrm{S} = 1$,

$$n_\mathrm{c} = \frac{t_\mathrm{sep}}{w_\mathrm{b}} \tag{3.4}$$

where $^1w_\mathrm{b}$ and $^2w_\mathrm{b}$ are obtained from the ^{1}D and ^{1}D×^{2}D chromatograms, as depicted in Fig. 3.3A-D, or extrapolated from w_b^* via previously described theory [44]. In practice however, both the ^{1}D and ^{2}D peaks are broadened by the modulator, and the more relevant *effective* peak capacity that reflects the undersampling is given by

$$n^*_\mathrm{c,3D} = {}^1n^*_\mathrm{c} \times {}^2n^*_\mathrm{c} \times {}^3n_\mathrm{c} \tag{3.5}$$

where $^1n^*_\mathrm{c}$ and $^2n^*_\mathrm{c}$ are calculated using the

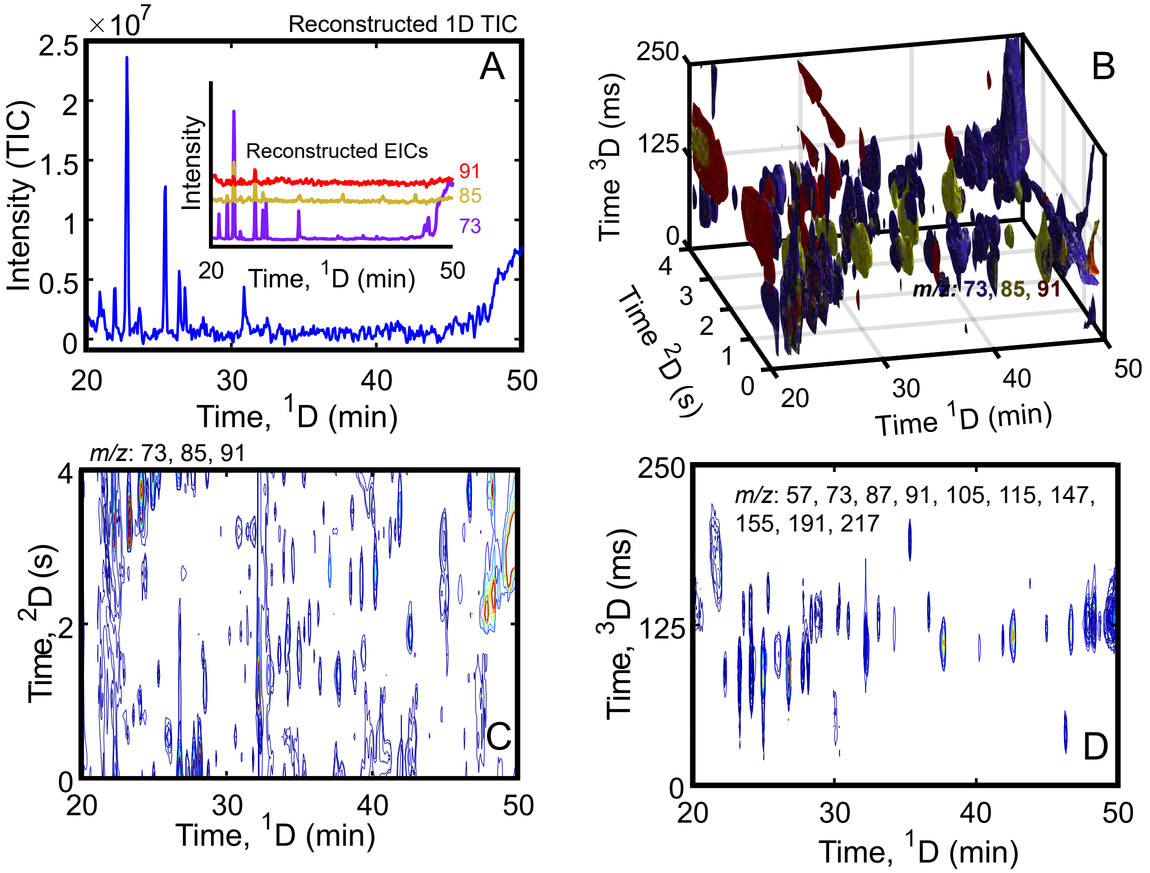


Figure 3.6: (A) Reconstructed 1D TIC of derivatized porcine serum, obtained with Column Set 2, as used in Fig. 3.5. The inset shows reconstructed EICs at *m/z* 73, 85, and 91. The intensity of *m/z* 85 and 91 have been increased ten-fold that of *m/z* 73 and offset, for clarity. (B) Isosurface analytical ion chromatogram of derivatized porcine serum, with *m/z* 73 in blue, 85 in gold, and 91 in red. (C) ^{1}D×^{2}D separation plane of derivatized porcine serum with *m/z* 73, 85, and 91 summed. (D) View of the ^{1}D×^{3}D separation plane with selective *m/z* used to show usage of the separation ^{1}D×^{3}D space.

effective peak widths measured from modulated chromatograms, as depicted in Fig. 3.4,

$$n_c^* = \frac{t_{\text{sep}}}{w_b^*} \tag{3.6}$$

The effective 3D peak capacity, $^1n_c^*$ via Eq. (3.5), calculated for the 10 representative analytes ranged from 6,800 – 32,300 ($\bar{x}$ = 19,900 ± 10,700). Since in prior studies we reported the ideal 3D peak capacity ($n_{\text{c,3D}}$) via Eq. (3.3), we also report the ideal peak capacity for comparison. The ideal and effective peak capacities are related by the broadening factor, β,

$$n_c^* = \frac{n_c}{\langle\beta\rangle} \tag{3.7}$$

Given that $\langle{}^1\beta\rangle$ was 1.12 and $\langle{}^2\beta\rangle$ was 1.26, $n_{\text{c,3D}}$, is ~28,000. This greatly exceeds the peak capacity of the first GC^3-TOFMS instrument which yielded a $n_{\text{c,3D}}$ of 7000. Importantly, we have demonstrated that this instrument design provides tandem total-transfer modulation while previous systems employed one or more modulators with a duty cycle less than 100%.

To demonstrate applicability of the total-transfer GC^3-TOFMS instrument, two complex samples were analyzed: a metabolomics-type sample, i.e., deriva-

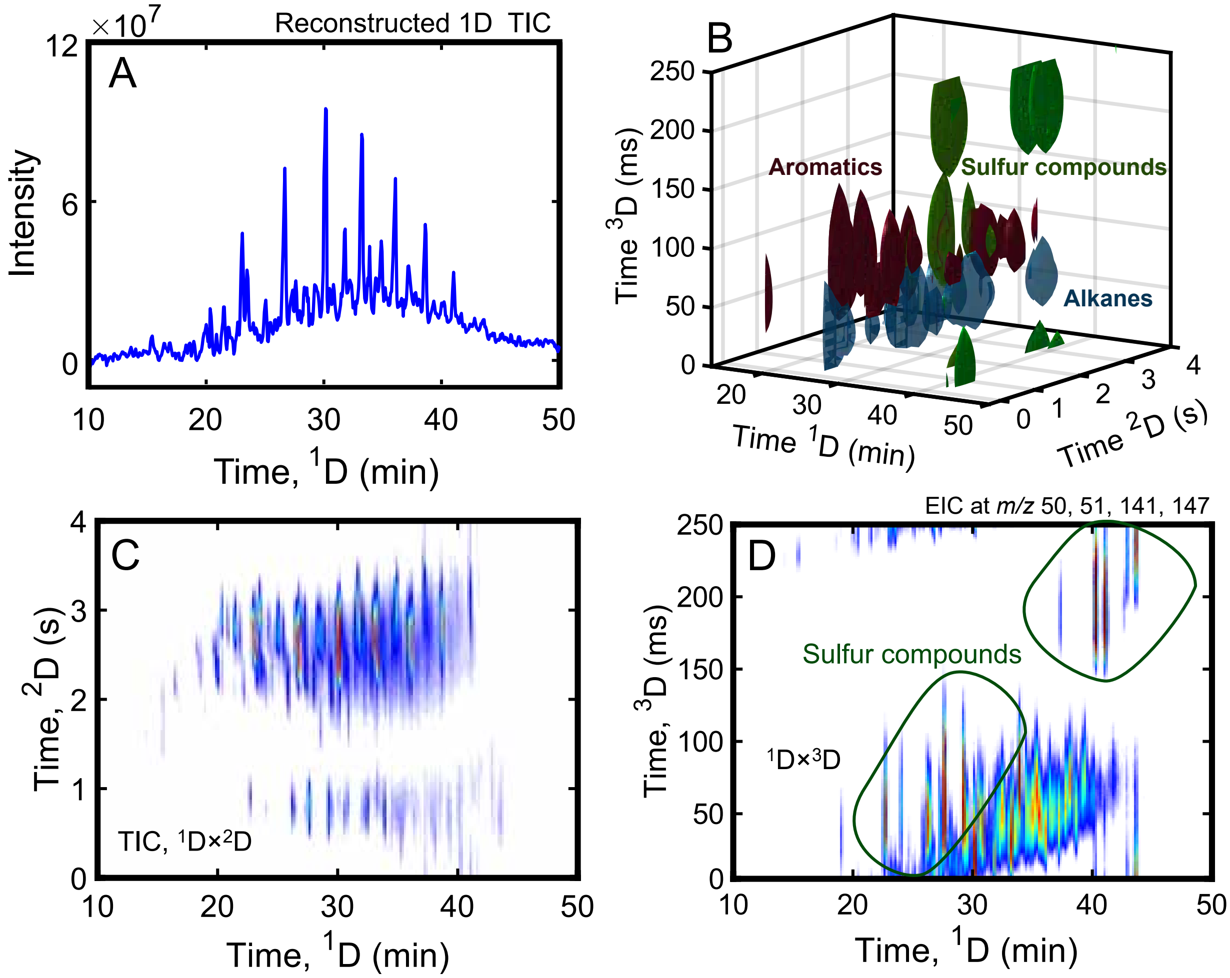


Figure 3.7: Output from GC^3-TOFMS analysis of jet fuel spiked with sulfur-containing compounds listed in Table A2. (A) Reconstructed 1D TIC chromatogram. (B) Isosurface EIC, with alkanes in blue (m/z 57), aromatics in red (m/z 91), and sulfur compounds in green (m/z 50, 51). (C) Traditional 2D TIC chromatogram of the ^{1}D×^{2}D separation plane. (D) View of the ^{1}D×^{3}D separation plane, using selective m/z 50, 51, 141, and 147 to illustrate separation of sulfur species relative to the bulk mixture.

porcine serum, and a jet fuel spiked with sulfur-containing compounds. For the tized porcine serum, Column Set 2 was applied with the same conditions as in Fig. 3.5. The reconstructed TIC in Fig. 3.6A is prepared in same manner as in Fig. 3.2A. The inset of reconstructed EICs (m/z 73, 85, and 91) were prepared similarly by indexing the indicated m/z and summing up ^{2}D and ^{3}D. The inset emphasizes that selective m/z are necessary to visualize this sample, unlike the test mixture separations, since there are few peaks visible in the TIC. Then, the same m/z used to construct the EICs were used to construct the 3D isosurface EIC in Fig. 3.5B. Here, isosurfaces at different m/z are overlaid in different colors for visualization. Owing largely to the additional chromatographic dimensions, an abundance of additional peaks are visible. Summation of the ^{3}D dimension at these m/z results in the ^{1}D×^{2}D EIC in Fig. 3.6C. Excellent usage of the ^{1}D×^{2}D space is obtained, with wraparound intentionally permitted to maximize the peak capacity in this plane. A ^{1}D×^{3}D EIC is prepared in Fig. 3.6D with ^{2}D summed and m/z 57, 73, 87, 91, 105, 115, 147, 155, 191, and 217 added together. Fig. 3.6D demonstrates that the ^{3}D column provides selectivity beyond the "standard" ^{1}D×^{2}D separation, with several analytes levitated away from the the band containing most of the components in this separation window.

Next, total-transfer GC^3-TOFMS was investigated to "filter" sulfur-containing compounds from a jet fuel sample. For this purpose, a third column set was employed (Table 3.1), using an ionic liquid phase (IL-59) on ^{3}D, which has selectivity for organosulfur compounds [47]. Fig. 3.7A depicts the reconstructed 1D TIC, processed in the same manner as in Fig. 3.2A, highlighting the need for additional separation dimensions. Selective m/z are used to produce the 3D isosurface EIC in Fig. 3.7B. Compound class distinction is apparent in the 3D space. Here, alkanes (m/z 57) are depicted in blue, aromatics (m/z 91) in red, and sulfur-containing compounds (m/z 50, 51) in green. As before, the ^{1}D×^{2}D TIC is provided in Fig. 3.7C. Since the ^{1}D column is more polar than ^{2}D, a "reverse type" 2D separation is obtained, with aromatics less retained on ^{2}D than saturated compounds. A ^{2}D×^{3}D EIC is provided in Fig. 3.7D, with m/z 50, 51, 141, and 147 employed to display most of the sulfur-containing compounds in the fuel. The sulfur compounds are highlighted, and adequately separated away from the bulk fuel sample, which elutes together on ^{3}D. As applied herein, a selective ionic liquid column on ^{3}D was successful in selectively separating sulfur-containing compounds from a complex hydrocarbon mixture, providing a highly ordered 3D separation.

3.4 Conclusion

Since the first reports of GC^3 [21–23], both modulator technology and data processing tools have undergone significant advances. Principally among these is the development of fast, total-transfer modulators, such as DPGM. Herein, DPGM is added as a facile modification to a commercial thermally modulated GC×GC-TOFMS platform, providing a total-transfer GC^3-TOFMS system. The new design enables impressive signal enhancement relative to unmodulated 1D chromatograms. Using a relatively long ^{3}D column with a retentive phase improves the usage of the overall 3D space when employing DPGM as ^{2}M. Using column set 2, an improved peak capacity relative to previous GC^3 reports was achieved with a range of 6,800 − 32,300 ($\bar{x} = 19{,}900 \pm 10{,}700$). Future studies will focus on leveraging the 4D data structure to address emerging challenges in chemical analysis of complex samples.

3.5 References

(1) Z. Liu, J.B. Phillips, Comprehensive two-dimensional gas chromatography using an on-column thermal modulator interface, *J. Chromatogr. Sci.* **1991**, 29 227– 231.

(2) M.S.S. Amaral, Y. Nolvachai, P.J. Marriott, Comprehensive two-dimensional gas chromatography advances in technology and applications: Biennial update, *Anal. Chem.* **2020**, *92*, 85–104.

(3) S.E. Prebihalo, K.L. Berrier, C.E. Freye, H.D. Bahaghighat, N.R. Moore, D.K. Pinkerton, R.E. Synovec, Multidimensional gas chromatography: Advances in instrumentation, chemometrics, and applications, *Anal. Chem.* **2018**, *90*, 505–532.

(4) J.V. Seeley, S.K. Seeley, Multidimensional gas chromatography: Fundamental advances and new applications, *Anal. Chem.* **2013**, *85*, 557–578

(5) F.A. Franchina, G. Purcaro, A. Burklund, M. Beccaria, J.E. Hill, Evaluation of different adsorbent materials for the untargeted and targeted bacterial VOC analysis using GC×GC-MS, *Anal. Chim. Acta* **2019**, *1066*, 146–153.

(6) L.M. Dubois, P.H. Stefanuto, L. Heudt, J.F. Focant, K.A. Perrault, Characterizing decomposition odor from soil and adipocere samples at a death scene using HS-SPME-GC×GC-HRTOFMS, *Forensic Chem.* **2018**, *8*, 11–20.

(7) K.A. Perrault, K.D. Nizio, S.L. Forbes, A comparison of one-dimensional and comprehensive two-dimensional gas chromatography for decomposition odour profiling using inter-year replicate field trials, *Chromatographia* **2015**, *78*, 1057–1070.

(8) P. Korytár, P.E.G. Leonards, J. de Boer, U.A.Th. Brinkman, Group separation of organohalogenated compounds by means of comprehensive two-dimensional gas chromatography, *J. Chromatogr. A* **2005**, *1086*, 29–44.

(9) A.M. Muscalu, T. Górecki, Comprehensive two-dimensional gas chromatography in environmental analysis, *TrAC - Trends Anal. Chem.* **2018**, *106*, 225–245.

(10) X. Guan, J. Luong, Z. Yu, H. Jiang, Quasi-stop-flow modulation strategy for comprehensive two-dimensional gas chromatography, *Anal. Chem.* **2020**, *92*, 6251-6256.

(11) T. Gröger, B. Gruber, D. Harrison, M. Saraji-Bozorgzad, M. Mthembu, A. Sutherland, R. Zimmermann, A vacuum ultraviolet absorption array spectrometer as a selective detector

for comprehensive two-dimensional gas chromatography: concept and first results, *Anal. Chem.* **2016**, *88*, 3031–3039.

(12) J. Luong, X. Guan, S. Xu, R. Gras, R.A. Shellie, Thermal independent modulator for comprehensive two-dimensional gas chromatography, *Anal. Chem.* **2016**, *88*, 8428–8432.

(13) F.A. Franchina, M. Maimone, P.Q. Tranchida, L. Mondello, Flow modulation comprehensive two-dimensional gas chromatography-mass spectrometry using 4 mL min-1 gas flows, *J. Chromatogr. A*, **2016**, *1441*, 134–139.

(14) L.A. Adutwum, J.J. Harynuk, Unique ion filter: A data reduction tool for GC/MS data preprocessing prior to chemometric analysis, *Anal. Chem.* **2014**, *86*, 7726–7733.

(15) B.A. Weggler, B. Gruber, F.L. Dorman, Rapid screening of complex matrices: utilizing Kendrick mass defect to enhance knowledge-based group type evaluation of multidimensional gas chromatography-high-resolution time-of-flight mass spectrometry data, *Anal. Chem.* **2019**, *91*, 10949–10954.

(16) K.L. Berrier, C.E. Freye, M.C. Billingsley, R.E. Synovec, Predictive modeling of aerospace fuel properties using comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry and partial least squares analysis, *Energy Fuels* **2020**, *34*, 4084-4094.

(17) C.N. Cain, S. Schöneich, R.E. Synovec, Development of an enhanced total ion current chromatogram algorithm to improve untargeted peak detection, *Anal. Chem.* **2020**, *92*, 11365-11373.

(18) M.S. Klee, J. Cochran, M. Merrick, L.M. Blumberg, Evaluation of conditions of comprehensive two-dimensional gas chromatography that yield a near-theoretical maximum in peak capacity gain, *J. Chromatogr. A* **2015,** *1383*, 151–159.

(19) J.M. Davis, Statistical-overlap theory for elliptical zones of high aspect ratio in comprehensive two-dimensional separations, *J. Sep. Sci.* **2005**, *28*, 347–359.

(20) J.M. Davis, Statistical theory of spot overlap in two-dimensional separations, *Anal. Chem.* **1991**, *63*, 2141–2152.

(21) E.B. Ledford, C.A. Billesbach, GC3: Comprehensive three-dimensional gas chromatography, J. *High Resolut. Chromatgr.* **2000,** *23,* 205–207.

(22) N.E. Watson, W.C. Siegler, J.C. Hoggard, R.E. Synovec, Comprehensive three-dimensional gas chromatography with parallel factor analysis, *Anal. Chem.* **2007**, *79*, 8270–8280.

(23) W.C. Siegler, J.A. Crank, D.W. Armstrong, R.E. Synovec, Increasing selectivity in comprehensive three-dimensional gas chromatography via an ionic liquid stationary phase column in one dimension, *J. Chromatogr. A* **2010**, *1217*, 3144-3149.

(24) N.E. Watson, H.D. Bahaghighat, K. Cui, R.E. Synovec, Comprehensive three-dimensional gas chromatography with time-of-flight mass spectrometry, *Anal. Chem.* **2017**, *89*, 1793–1800.

(25) N.E. Watson, S.E. Prebihalo, R.E. Synovec, Targeted analyte deconvolution and identification by four-way parallel factor analysis using three-dimensional gas chromatography with mass spectrometry data, *Anal. Chim. Acta* **2017,** *983*, 67–75.

(26) N.E. Watson, H.D. Bahaghighat, K. Cui, R.E. Synovec, Comprehensive three-dimensional gas chromatography with time-of-flight mass spectrometry, *Anal. Chem.* **2017**, *89*, 1793–1800.

(27) D.V. Gough, H.D. Bahaghighat, R.E. Synovec, Column selection approach to achieve a high peak capacity in comprehensive three-dimensional gas chromatography, *Talanta* **2019,** *195*, 822–829.

(28) D.V. Gough, D.H. Song, S. Schöneich, S.E. Prebihalo, R.E. Synovec, Development of ultrafast separations using negative pulse partial modulation to enable new directions in gas chromatography, *Anal. Chem.* **2019**, *91*, 7328-7335.

(29) V.H.C. Ferreira, L.W. Hantao, R.J. Poppi, Consumable-free comprehensive three-dimensional gas chromatography and PARAFAC for determination of allergens in perfumes, Chromatographia **2020**, *83*, 581-592.

(30) T.J. Trinklein, S.E. Prebihalo, C.G. Warren, G.S. Ochoa, R.E. Synovec, Discovery-based analysis and quantification for comprehensive three-dimensional gas chromatography flame ionization detection data, *J. Chromatogr. A* **2020**, *1623*, 461190.

(31) B. Wouters, E. Davydova, S. Wouters, G. Vivó-Truyols, P.J. Schoenmakers, S. Eeltink, Towards ultra-high peak capacities and peak-production rates using spatial three-dimensional liquid chromatography, *Lab Chip* **2015**, *15*, 4415–4422.

(32) P. Venter, M. Muller, J. Vestner, M.A. Stander, A.G.J. Tredoux, H. Pasch, A. de Villiers, Comprehensive three-dimensional LC × LC × ion mobility spectrometry separation combined with high-resolution MS for the analysis of complex samples, *Anal. Chem.*, **2018**, *90* 11643–11650.

(33) D. Chen, J.H. Seo, J. Liu, K. Kurabayashi, X. Fan, Smart three-dimensional gas chromatography, *Anal. Chem.* **2013**, *85*, 6871–6875.

(34) D.D. Yan, Y.F. Wong, S.P. Whittock, A. Koutoulis, R.A. Shellie, P.J. Marriott, Sequential hybrid three-dimensional gas chromatography with accurate mass spectrometry: a novel tool for high-resolution characterization of multicomponent samples, *Anal. Chem.* **2018**, *90*, 5264–5271.

(35) D.V. Gough, S. Schneich, R.E. Synovec, Chemometric decomposition of comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry data employing partial modulation in the negative pulse mode, *Talanta* **2020**, *210*, 120670.

(36) T.J. Trinklein, D. V. Gough, C.G. Warren, G.S. Ochoa, R.E. Synovec, Dynamic pressure gradient modulation for comprehensive two-dimensional gas chromatography, *J. Chromatogr. A* **2020**, *1609*, 460488.

(37) S. Schöneich, D. V Gough, T.J. Trinklein, R.E. Synovec, Dynamic pressure gradient modulation for comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry detection, *J. Chromatogr. A* **2020**, *1620*,

460982.

(38) S. Schöneich, T.J. Trinklein, C.G. Warren, R.E. Synovec, A systematic investigation of comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry with dynamic pressure gradient modulation for high peak capacity separations, *Anal. Chim. Acta* **2020**, *1134*, 115-124.

(39) A.C. Beckstrom, E.M. Humston, L.R. Snyder, R.E. Synovec, S.E. Juul, Application of comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry method to identify potential biomarkers of perinatal asphyxia in a non-human primate model, *J. Chromatogr. A* **2011**, *1218*, 1899–1906.

(40) H.-J. de Geus, J. de Boer, J.B. Phillips, E.B. Ledford Jr., U.A.Th. Brinkman, Increased signal amplitude due to mass conservation in a thermal desorption modulator, *J. High Resolut. Chromatogr.* **1998**, *21*, 411-413.

(41) R.M. Kinghorn, P.J. Marriott, Enhancement of signal-to-noise ratios in capillary gas chromatography by using a longitudinally modulated cryogenic system, *J. High Resolut. Chromatogr.* **1998**, *21*, 32-38.

(42) Z. Liu, J.B. Phillips, Sensitivity and detection limit enhancement of gas chromatographic detection by thermal modulation, *J. Microcol. Sep.* **1994**, *6*, 229-235.

(43) A.L. Lee, K.D. Bartle, A.C. Lewis, A model of peak amplitude enhancement in orthogonal two-dimensional gas chromatography, *Anal. Chem.* **2001**, *73*, 1330–1335.

(44) D.K. Pinkerton, B.A. Parsons, R.E. Synovec, Method to determine the true modulation ratio for comprehensive two-dimensional gas chromatography, *J. Chromatogr. A* **2016**, *1476*, 114–123.

(45) K.A.J.M., Stevenson, L.M. Blumberg, J.J. Harynuk, Thermodynamics-based retention maps to guide column choices for comprehensive multi-dimensional gas chromatography, *Anal. Chim. Acta* **2019**, *1086*, 133-141.

(46) S.C. Rutan, J.M. Davis, P.W. Carr, Fractional coverage metrics based on ecological home range for calculation of the effective peak capacity in comprehensive two-dimensional separations, *J. Chromatogr. A* **2012**, *1255*, 267–276.

(47) F. Cappelli Fontanive, É.A. Souza-Silva, J. Macedo da Silva, E. Bastos Caramão, C. Alcaraz Zini, Characterization of sulfur and nitrogen compounds in Brazilian petroleum derivatives using ionic liquid capillary columns in comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometric detection, *J. Chromatogr. A* **2016**, *1461*, 131–143.

Chapter 4

DETERMINATION OF THE SIGNAL-TO-NOISE RATIO ENHANCEMENT IN COMPREHENSIVE THREE-DIMENSIONAL GAS CHROMATOGRAPHY

This chapter was adapted from T.J. Trinklein, C.G. Warren, R.E. Synovec, *Anal. Chem.* **2021**, *93*, 8526-8535.

4.1 Introduction

Comprehensive two-dimensional gas chromatography (GC×GC) is becoming a mature technique for the analysis of volatile and semi-volatile compounds [1–4]. While new instrumental developments are being reported [5], the bulk of recent research has shifted to developing data analysis tools [2,6] and exploring new applications and sample preparation methods [7]. Three key benefits motivate the usage of GC×GC over 1D-GC: ~10 fold higher peak capacity [8], chemical class structured separations, and increased detectability from modulation [9–18]. Despite the excellent separation performance provided by GC×GC, many samples are of such complexity that the peak capacity is insufficient to resolve the analytes of interest [19,20].

One intriguing solution is the addition of a third dimension (^{3}D) to a GC×GC system, i.e., comprehensive three-dimensional (3D) gas chromatography (GC^3) [21–30]. configured by sampling the ^{2}D peaks by a second modulator interfaced to a ^{3}D column of unique selectivity. The potential benefits of GC^3 are well supported by theory [19,31], and other comprehensive 3D platforms, such as LC×LC×CE [32], spatial 3D-LC [33,34], and LC×LC×IMS [35] have also been investigated. Recent GC^3 reports have investigated the attainable peak capacity and improved selectivity afforded by the technique, [25–26,29–30], However, the potential for GC^3 (and other 3D platforms) to provide signal enhancement (*SE*), and hence, signal-to-noise ratio (*S/N*) enhancement vis-á-vis GC×GC and 1D-GC has not been explored.

In Chapter 3, we introduced a GC^3 instrument design coupled with time-of-flight mass spectrometry where both modulators provided total-transfer, i.e., 100% duty-cycle (TT-GC^3-TOFMS) [30]. Notably, a substantial *SE* was observed, defined as the ratio of the tallest ^{3}D peak height relative to the unmodulated 1D peak height; the largest *SE* observed was 175 (=130 ± 47). Since thermal and total-transfer differential flow modulation were employed in series, they both enhance the signal. Thermal modulation selectively preconcentrates the eluting analyte over the carrier gas, whereas differential flow modulation increases analyte flux to the detector, which enhances signal for mass-sensitive detectors (e.g., flame ionization detection and electron ionization mass spectrometry) as a result of the increased flow rate on the terminal dimension [14,16,18,36,37]. While the *SE*s were encouraging, no theoretical framework yet exists to support this observation, nor is it clear if the practically relevant metric, i.e., the *S/N*, is truly enhanced. The objective of the present report is to address these questions.

To this end, we propose a model of *SE* for GC^3, extending a model developed by Lee et al. for GC×GC [12]. While their approach laid a strong foundation for examining *SE* in comprehensive multi-dimensional separations, we found quantitative discrepancies when applying their modeling expressions. To address these shortcomings, we rederive expressions for *SE* based upon concepts outlined in their report and then extend the model to GC^3. The proposed theoretical model is validated using both experimentally obtained and simulated GC^3 data. We then investigate whether the *S/N* is truly enhanced for GC^3 separations relative to 1D-

GC.

4.2 Principles and theory

Formally, the signal enhancement ${}^{i}_{x}SE$ of a given analyte is defined as the ratio of the tallest peak in the x^{th} dimension originating from the tallest peak in the i^{th} dimension, which is independent of sampling phase. Thus, ${}^{2}_{1}SE$ is the ratio of the tallest ^{2}D peak (${}^{2}h$) to the ^{1}D peak height (${}^{1}h$),

$$ {}^{2}_{1}SE = \frac{{}^{2}h}{{}^{1}h} \tag{4.1} $$

Similarly, ${}^{3}_{2}SE$ is the ratio of the tallest ^{3}D peak (${}^{3}h$) that originates from the tallest ^{2}D peak. Note that the product of ${}^{2}_{1}SE$ and ${}^{3}_{2}SE$ yields ${}^{3}_{1}SE$, which is the ratio of the tallest ^{3}D peak to the ^{1}D peak height,

$$ {}^{3}_{1}SE = \frac{{}^{2}h}{{}^{1}h} \times \frac{{}^{3}h}{{}^{2}h} = \frac{{}^{3}h}{{}^{1}h} \tag{4.2} $$

However, in practice, the relevant metric is the relative signal-to-noise ratio enhancement from ^{1}D to ^{3}D, ${}^{3}_{1}S/N_{\text{Rel}}$, since the S/N of a method ultimately determines the limit of detection (LOD). The S/N on the k^{th} dimension is given by ${}^{\text{k}}h/{}^{\text{k}}s_{\text{N}}$, where ${}^{\text{k}}s_{\text{N}}$ is one standard deviation of the baseline noise measured herein on an event window of the 6σ peak width of the analyte on the k^{th} dimension. However, in order to make an objective comparison ${}^{1}s_{\text{N}}$ must be resampled to contain the same number of data points (i.e., spectra) as ${}^{3}s_{\text{N}}$, thus denoted ${}^{1}s_{\text{N,RS}}$. Then ${}^{3}_{1}S/N_{\text{Rel}}$ is given by

$$ {}^{3}_{1}S/N_{\text{Rel}} = {}^{3}_{1}SE \frac{{}^{1}s_{\text{N,RS}}}{{}^{3}s_{\text{N}}} \tag{4.3} $$

Equation (4.3) highlights that ${}^{3}_{1}S/N_{\text{Rel}}$ is inextricably linked to ${}^{3}_{1}SE$. Therefore, the initial focus of this report is to derive expressions for ${}^{3}_{1}SE$ and their validation by experimental and simulated data. In subsequent sections, the experimentally obtained ${}^{3}_{1}S/N_{\text{Rel}}$ are investigated.

Herein, Gaussian peaks of the general form

$$ C(t) = \frac{4A}{w_{\text{b}}\sqrt{2\pi}} \exp\left(\frac{-8(t - t_{\text{R}})^2}{w_{\text{b}}^2}\right) \tag{4.4} $$

are assumed, where A is the total analyte peak area and w_{b} is the 4 peak width. The ^{1}D peak height is obtained by evaluating Eq. (4.4) at $t_{\text{R}} = t =^{1} t_{\text{R}}$,

$$ {}^{1}h = \frac{4 \cdot {}^{1}A}{{}^{1}w_{\text{b}}\sqrt{2\pi}} \tag{4.5} $$

where ${}^{1}A$ is taken to be unit area for the analyte peak eluting from ^{1}D. Upon modulation, ${}^{1}A$ is sampled with a modulation period of ${}^{1}P_{\text{M}}$, with each fraction transferred onto ^{2}D, derived here for the GC^{3} instrument that provides 100% duty-cycle. The largest fraction sampled from the ^{1}D peak becomes the tallest ^{2}D peak and the area is designated ${}^{2}a$. The tallest ^{2}D peak height is then given by

$$ {}^{2}h = \frac{4 \cdot {}^{2}a}{{}^{2}w_{\text{b}}\sqrt{2\pi}} \tag{4.6} $$

This process is shown in Fig. 4.1A and 4.1B for the case of in-phase modulation from ^{1}D to ^{2}D. The phase of a set of peaks sampled from the k^{th} dimension is defined as the difference in position of the peak maximum (${}^{\text{k}}t_{\text{R}}$) relative to the center of the nearest sampling event (denoted ${}^{\text{k}}t_{\text{mod}}$) and normalized by P_{M} [38]

$$ {}^{k}\theta = \frac{\left|{}^{k}t_{\text{mod}} - {}^{k}t_{\text{R}}\right|}{{}^{k}P_{\text{M}}} \tag{4.7} $$

where the absolute value is taken because the symmetry of the Gaussian peak model produces equivalent results for positive and negative phases. We designate in-phase modulation of the k^{th} dimension with ${}^{k}\theta_{\text{i},}$ and out-of-phase modulation with ${}^{k}\theta_{\text{o}.}$ By this definition, the possible values of ${}^{k}\theta$ range from 0.0 to 0.5 with ${}^{k}\theta_{\text{i}} = 0.0$ and ${}^{k}\theta_{\text{o}} = +0.5$.

For instructive purposes, and for comparison to prior work, we focus on the four "ideal" GC^{3} sampling cases, where ^{1}D and ^{2}D are sampled only at θ_{i} and θ_{o}. These principles are introduced by first deriving one of these cases, in-phase by in-phase modulation (${}^{1}\theta_{\text{i}} = {}^{2}\theta_{\text{i}} = 0.0$), shown in Fig. 4.1. However, the experimentally observed ${}^{k}\theta$ will vary between ${}^{k}\theta_{\text{i}}$ and ${}^{k}\theta_{\text{o}}$ due to the randomness of analyte retention times in a fixed modulation periods. Therefore, following this derivation of the in-phase by in-phase

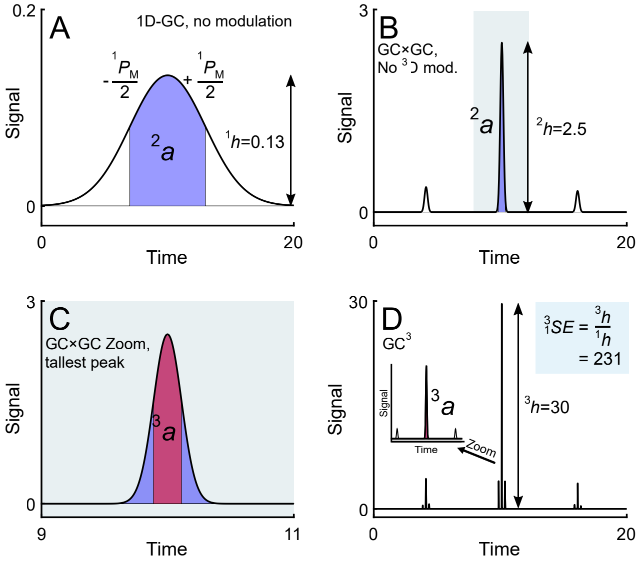


Figure 4.1: Simulated illustration of in-phase by in-phase modulation. (A) A ^{1}D peak before modulation. (B) Unfolded GC×GC peaks after in-phase modulation. (C) Zoom-in of the tallest ^{2}D peak from (B). (C) Unfolded GC^3 peaks after in-phase by in-phase modulation.

modulation case, we also derive a general ^3_1SE expression which can be evaluated at any possible combination of $^1\theta$ and $^2\theta$.

In an effectively identical process to what occurs in modulation from ^{1}D to ^{2}D, a fraction of the ^{2}D peak area is modulated and submitted to ^{3}D separation. The tallest ^{3}D peak, which originates from the tallest ^{2}D peak, is given by

$$^3h = \frac{4 \cdot {}^3a}{^3w_\mathrm{b}\sqrt{2\pi}} \tag{4.8}$$

where ^3a is the total area of the tallest ^{3}D peak and is equivalent to the sampled area originating from the tallest ^{2}D peak. This process is shown graphically for completely in-phase sampling between ^{2}D and ^{3}D ($^2\theta_\mathrm{i}$) in Fig. 4.1C-D where the red area illustrates ^3a. Substitution of Eqs (4.5) and (4.8) into Eq (4.2) yields the following general expression for ^3_1SE

$$^3_1SE = \frac{^1w_\mathrm{b}{}^3a}{^3w_\mathrm{b}} \tag{4.9}$$

Note that as shown in Fig. 4.1, ^3a is a fraction of ^2a. Thus, we begin by deriving ^2a in terms of the error function and use similar principles to arrive at ^3a.

The area under points a and b of a general probability density function, φ, can be obtained by evaluating the cumulative distribution function (CDF), Φ, at both points,

$$\int_a^b = \varphi(\tau)\mathrm{d}\tau = \Phi(b) - \Phi(a) \tag{4.10}$$

where the reduced time, τ, is $\tau = (t - t_\mathrm{R})/\sigma$. The CDF for a standard normal probability density function is given by

$$\Phi(\tau) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\tau} \exp\left(\frac{-u^2}{2}\right) \mathrm{d}u \tag{4.11}$$

where u is a dummy variable. In the case of in-phase modulation of ^{1}D, the area sampled during modulation is bound by -$^1P_\mathrm{M}/2$ and +$^1P_\mathrm{M}/2$ (Fig. 4.1A). Therefore, ^2a is obtained as follows,

$$^2a_{^1\theta_\mathrm{i}} = {}^1A \int_{-^1P_\mathrm{M}/2}^{+^1P_\mathrm{M}/2} {}^1C(t)\mathrm{d}t \tag{4.12}$$

$$= {}^1A \left[\Phi\left(\frac{+^1P_\mathrm{M}}{2}\right) - \Phi\left(\frac{-^1P_\mathrm{M}}{2}\right)\right] \tag{4.13}$$

We will express Φ in terms of the closely related error function in order to facilitate comparison with signal enhancement expressions reported previously for GC×GC [12]. By definition, the error function is

$$\mathrm{erf}(\tau) = \frac{2}{\sqrt{2\pi}} \int_0^{\tau} \exp\left(-z^2\right) \mathrm{d}z \tag{4.14}$$

where z is a unique dummy variable, applied so that Eq. (4.14) can be rewritten in terms of the variable u in Eq. (4.11). In order to express the fractional peak area with the error function, the error function is expressed in terms of Φ. Therefore, we rewrite z^2 = $u^2/2$ which implies that $z = u/2$ (because z cannot be negative) and $\mathrm{d}z = \mathrm{d}u/2$. The relationship $z = u/2$ also causes the upper endpoint of $z =$ to become $u = 2$. Substitution of these relationships into Eq. (4.14) yields

$$\mathrm{erf}(\tau) = \frac{2}{\sqrt{2\pi}} \int_{0}^{\tau\sqrt{2}} \exp\left(\frac{-u^2}{2}\right) \mathrm{d}u \tag{4.15}$$

Since the CDF is defined with a lower limit of $-\infty$, the error function must be rewritten with lower limits of $-\infty$

$$\mathrm{erf}(\tau) = \frac{2}{\sqrt{2\pi}} \left(\int_{-\infty}^{\tau\sqrt{2}} \exp\frac{-u^2}{2} \mathrm{d}u - \int_{-\infty}^{0} \exp\frac{-u^2}{2} \mathrm{d}u \right) \tag{4.16}$$

Note that now both integrals take the form of Φ. Thus, Eq. (4.16) can be expressed as

$$\mathrm{erf}(\tau) = 2\left(\Phi\left(\tau\sqrt{2}\right) - \Phi\left(0\right)\right) \tag{4.17}$$

Note that $\Phi(o) = 1/2$ due to symmetry, therefore rearranging for $\Phi(\tau)$ yields

$$\Phi(\tau) = \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{\tau}{\sqrt{2}}\right)\right) \tag{4.18}$$

For a chromatographic peak of retention time t_R and standard deviation σ Eq. (4.18) can be expanded in terms of τ,

$$\Phi(t) = \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{t - t_\mathrm{R}}{\sigma\sqrt{2}}\right)\right) \tag{4.19}$$

We let the ^{1}D peak have $t_\mathrm{R} = {}^1t_\mathrm{R} = 0$ and $\sigma = {}^1\sigma$. Substitution of Eq. (4.19) into Eq. (4.10) and integrating over the ^{1}D peak, ${}^1C(t)$ gives an expression for the sampled area

$$
\begin{aligned}
{}^2a_{{}^1\theta_\mathrm{i}} &= {}^1A \int_{-{}^1P_\mathrm{M}/2}^{+{}^1P_\mathrm{M}/2} {}^1C(t)\ \mathrm{d}t & (4.20)\\
&= {}^1A\left[\mathrm{erf}\left(\frac{{}^1P_\mathrm{M}}{2\cdot{}^1\sigma\sqrt{2}}\right)\right] & (4.21)
\end{aligned}
$$

Equation (4.21) is not equivalent to the formula reported by Lee et al. for GC×GC, and subsequently reproduced [12–14]. An equivalent expression to Eq. (4.21) was reported in the context of LC×LC, although no derivation was provided [39].

Recall that Eq. (4.21) is derived based upon total-transfer (100% duty cycle). If the duty cycle is < 100%, Eq(4.21) needs only to be multiplied by the fraction that is sampled (i.e., multiplied by 0.5 for 1:1 split modulation).

For a ^{1}D peak modulated at ${}^1P_\mathrm{M}$, ${}^1w_\mathrm{b} = 4\cdot{}^1\sigma$ and the sampling density (i.e., modulation ratio) is ${}^k\rho = {}^kw_\mathrm{b}/{}^kP_\mathrm{M}$. Then, taking ${}^1A$ to have unit area we have

$${}^2a_{{}^1\theta_\mathrm{i}} = \mathrm{erf}\left(\frac{\sqrt{2}}{{}^1\rho}\right) \tag{4.22}$$

The ^{3}D peak area for in-phase modulation can be obtained similarly be integrating over teh tallest ^{2}D peak concentration profile, ${}^2C(t)$, between the limits of $-{}^1P_\mathrm{M}/2$ and $+{}^1P_\mathrm{M}/2$, and is scaled by ${}^2a$. Applying the definition of Eqs. (4.10) and (4.19) yields

$$
\begin{aligned}
{}^3a_{{}^1\theta_\mathrm{i},{}^2\theta_\mathrm{i}} &= {}^2a \int_{-{}^2P_\mathrm{M}/2}^{+{}^2P_\mathrm{M}/2} {}^2C(t)\ \mathrm{d}t & (4.23)\\
&= \mathrm{erf}\left(\frac{\sqrt{2}}{{}^2\rho}\right) & (4.24)
\end{aligned}
$$

If the ^{1}D and ^{2}D peaks are both sampled in-phase, the tallest ^{3}D peak area is obtained by substitution of Eq (4.22) into Eq (4.24)

$${}^3a_{{}^1\theta_\mathrm{i},{}^2\theta_\mathrm{i}} = \mathrm{erf}\left(\frac{\sqrt{2}}{{}^1\rho}\right)\mathrm{erf}\left(\frac{\sqrt{2}}{{}^2\rho}\right) \tag{4.25}$$

Then, ${}^3_1SE_{{}^1\theta_\mathrm{i},{}^2\theta_\mathrm{i}}$ is obtained by substituting Eq (4.25) into Eq (4.9)

$${}^3_1SE_{{}^1\theta_\mathrm{i},{}^2\theta_\mathrm{i}} = \frac{{}^1w_\mathrm{b}}{{}^3w_\mathrm{b}}\,\mathrm{erf}\left(\frac{\sqrt{2}}{{}^1\rho}\right)\mathrm{erf}\left(\frac{\sqrt{2}}{{}^2\rho}\right) \tag{4.26}$$

Further, it can be shown that the ${}^3_1SE$ for any possible phase combination is given generally by the following expression (see Appendix B for a complete derivation),

$$
\begin{aligned}
{}^{3}_{1}SE = \frac{{}^{1}w_{\mathrm{b}}}{{}^{3}w_{\mathrm{b}}} \Bigg[& \frac{\operatorname{erf}}{2}\left(\frac{-\sqrt{2}(2\cdot {}^{1}\theta - 1)}{{}^{1}\rho}\right) \\
& - \frac{\operatorname{erf}}{2}\left(\frac{-\sqrt{2}(2\cdot {}^{1}\theta + 1)}{{}^{1}\rho}\right) \\
\times \Bigg[\frac{\operatorname{erf}}{2}\left(\frac{-\sqrt{2}(2\cdot {}^{2}\theta - 1)}{{}^{2}\rho}\right) & - \frac{\operatorname{erf}}{2}\left(\frac{-\sqrt{2}(2\cdot {}^{2}\theta + 1)}{{}^{2}\rho}\right)\Bigg]
\end{aligned}
\tag{4.27}
$$

which can be used to arrive at the four "ideal" phase cases, where we restrict $^{1}\theta$ and $^{2}\theta$ to only θ_{i} and θ_{o}: both ^{1}D and ^{2}D sampled in-phase ($^{1}\theta_{\mathrm{i}}$, $^{2}\theta_{\mathrm{i}}$), both ^{1}D and ^{2}D sampled out-of-phase ($^{1}\theta_{\mathrm{o}}$, $^{2}\theta_{\mathrm{o}}$), ^{1}D sampled in-phase while ^{2}D is sampled out-of-phase ($^{1}\theta_{\mathrm{i}}$, $^{2}\theta_{\mathrm{o}}$), and ^{1}D sampled out-of-phase with ^{2}D sampled in-phase ($^{1}\theta_{\mathrm{o}}$, $^{2}\theta_{\mathrm{i}}$) (Figure B3). Substitution of θ_{i} and θ_{o} into Eq (4.27) elds the four ${}^{3}_{1}SE$ cases, given in Table 4.1.

Phase case	**Expression for ${}^{3}_{1}SE$**
$^{1}\theta_{\mathrm{i}}, {}^{2}\theta_{\mathrm{i}}$	$\frac{{}^{1}w_{\mathrm{b}}}{{}^{3}w_{\mathrm{b}}} \operatorname{erf}\left(\frac{\sqrt{2}}{{}^{1}\rho}\right) \operatorname{erf}\left(\frac{\sqrt{2}}{{}^{2}\rho}\right)$
$^{1}\theta_{\mathrm{o}}, {}^{2}\theta_{\mathrm{o}}$	$\frac{{}^{1}w_{\mathrm{b}}}{4\cdot{}^{3}w_{\mathrm{b}}} \operatorname{erf}\left(\frac{2\sqrt{2}}{{}^{1}\rho}\right) \operatorname{erf}\left(\frac{2\sqrt{2}}{{}^{2}\rho}\right)$
$^{1}\theta_{\mathrm{i}}, {}^{2}\theta_{\mathrm{o}}$	$\frac{{}^{1}w_{\mathrm{b}}}{2\cdot{}^{3}w_{\mathrm{b}}} \operatorname{erf}\left(\frac{\sqrt{2}}{{}^{1}\rho}\right) \operatorname{erf}\left(\frac{2\sqrt{2}}{{}^{2}\rho}\right)$
$^{1}\theta_{\mathrm{o}}, {}^{2}\theta_{\mathrm{i}}$	$\frac{{}^{1}w_{\mathrm{b}}}{2\cdot{}^{3}w_{\mathrm{b}}} \operatorname{erf}\left(\frac{2\sqrt{2}}{{}^{1}\rho}\right) \operatorname{erf}\left(\frac{\sqrt{2}}{{}^{2}\rho}\right)$

Table 4.1: Expressions of ${}^{3}_{1}SE$ for the four ideal sampling cases

Consideration of the theoretically attainable range of ${}^{3}_{1}SE$ under readily obtainable GC^{3} sampling densities and peak widths is instructive. For this purpose, ${}^{3}_{1}SE$ were calculated using the four expressions in Table 4.1 as a function of $^{1}\rho$ and $^{2}\rho$ with a fixed $^{1}w_{\mathrm{b}} = 12$ s and $^{3}w_{\mathrm{b}} = 30$ ms (Fig. 4.2), which are readily obtainable for GC^{3}-TOFMS.[30] Two important features are noted. First, ${}^{3}_{1}SE$ increases with decreasing ^{1} and ^{2} for all phase cases. Typically, practitioners of GC×GC (and GC^{3}) seek a sampling density on the order of $^{1}\rho$ and $^{2}\rho$ 1.5-3 [40–42] to maximize the ^{2}D peak capacity, $^{2}n_{\mathrm{c}}$, while not so low as to markedly broaden ^{1}D peaks [43] or incur quantitative inaccuracy for low-duty cycle modulators [38,42]. Second, the lowest ${}^{3}_{1}SE$ are obtained for the $^{1}\theta_{\mathrm{o}}, {}^{2}\theta_{\mathrm{o}}$ case, while the highest are obtained for the $^{1}\theta_{\mathrm{i}}, {}^{2}\theta_{\mathrm{i}}$ case, reaching well over 200. Therefore, we anticipate that a majority of the ${}^{3}_{1}SE$s in a real or simulated experiment to lie randomly between these two boundary conditions. Although the theoretical investigation via Fig, 4.2 is promising for obtaining high ${}^{3}_{1}SE$ with TT-GC^{3}-TOFMS, experimental validation, as presented herein, is warranted.

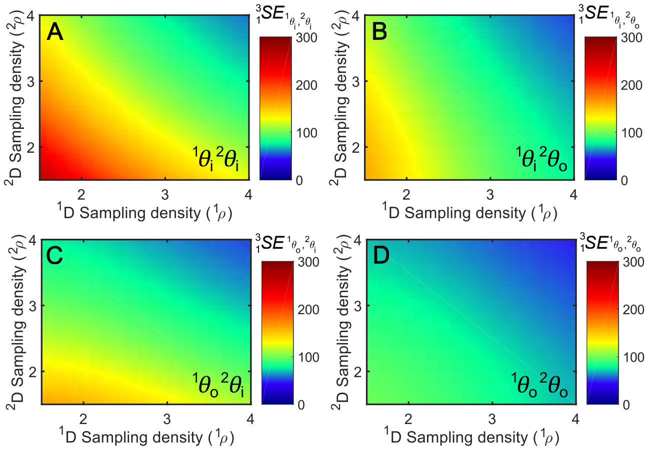


Figure 4.2: Effect of sampling densities, $^{1}\rho$ and $^{2}\rho$ on the theoretically attainable ${}^{3}_{1}SE$. (A) In-phase by in-phase case. (B) In-phase by out-of-phase case. (C) Out-of-phase by in-phase case. (D) Out-of-phase by out-of-phase case.

4.3 Experimental

The TT-GC^{3}-TOFMS instrument (Figure 3.1) is described in detail Chapter 3 [30]. Briefly, the instrument is based upon a modified Pegasus 4D GC×GC TOFMS (LECO Corporation, St. Joseph, MI) utilizing an Agilent 6890 gas chromatograph. Modulation from ^{1}D to ^{2}D was accomplished by the stock cryogenic quad-jet thermal modulator. Dynamic pressure gradient modulation (DPGM), a recently described total-transfer flow modulation method (Chapter 2), was used to perform modulation from ^{2}D to ^{3}D [36,37,44]. Mass spectra were acquired from m/z 33-256 at an electron ionization energy of -70 eV and an acquisition frequency of 500 Hz.

The expressions for ${}^{3}_{1}SE$ in Table 4.1 were validated by the analysis of a 100-component test mixture using two sets of experimental conditions by the analysis of a 100-component test mixture . The two sets of experimental conditions studied used the same GC^3 column set, differing principally in the modulation period $^1P_\mathrm{M}$ from ^{1}D to ^{2}D. Conditions 1, which used a $^1P_\mathrm{M}$ of 4 s and $^2P_\mathrm{M}$ of 250 ms, were chosen primarily in order to compare to previous work [30]. Conditions 2 used a $^1P_\mathrm{M}$ of 8 s and $^2P_\mathrm{M}$ of 250 ms.

For both sets of conditions, the measured ${}^3_1SE$ (${}^3_1SE_\mathrm{M}$) for the same twenty analytes were obtained, via Eq (4.2), following measurements of ^1h and ^3h from 1D-GC and GC^3 chromatograms, respectively. 1D-GC, GC×GC, and GC^3 data were all obtained on the TT-GC^3-TOFMS instrument by enabling successive stages of modulation: 1D-GC data were obtained with no modulators enabled, GC×GC data were obtained by enabling only the thermal modulator, and GC^3 data were obtained with both modulators enabled. The %Difference in area between the GC^3 and 1D-GC chromatograms for each of the analytes was determined using selective *m/z* extracted ion chromatograms, yielding an average %Difference of -3.6% ± 9.1%.

The ${}^3_1SE_\mathrm{M}$ were compared to the theoretical ${}^3_1SE$) (${}^3_1SE_\mathrm{T}$)), calculated by substitution of analyte-specific measurements into Eq. (4.27). To obtain $^1\theta$ and $^2\theta$, Gaussian curves were fitted to the envelope of ^{2}D peaks and the tallest set of ^{3}D peaks, respectively. Then, $^1\theta$ and $^2\theta$ were calculated as

$$^k\theta = \frac{|^kt_\mathrm{R,fit} - {}^{k+1}t_\mathrm{R,fit}|}{^kP_\mathrm{M}} \tag{4.28}$$

where $^\mathrm{k}t_\mathrm{R,fit}$, is the retention time of the fitted Gaussian curve, $^\mathrm{k+1}t_\mathrm{R,max}$ is the retention time of the tallest peak on the $k{+}1^\mathrm{th}$ dimension.

The true, unmodulated peak widths-at-base, $^1w_\mathrm{b}$, $^2w_\mathrm{b}$ and $^3w_\mathrm{b}$, which are the relevant widths in Table 4.1, were obtained from 1D-GC, GC×GC, and GC^3 chromatograms, respectively, by measuring the peak width-at-half-height and multiplying by 1.7. The sampling densities, $^1\rho$ and $^2\rho$ were calculated as $^1w_\mathrm{b}/^1P_\mathrm{M}$ and $^2w_\mathrm{b}/^2P_\mathrm{M}$, respectively.

To calculate the effective peak capacity ($n^*_\mathrm{c,3D}$), peak widths which include sampling-induced broadening, $^1w^*_\mathrm{b}$ and $^2w^*_\mathrm{b}$, were obtained from the GC^3 data. The $^1w^*_\mathrm{b}$ were obtained by fitting Gaussian curves to reconstructed 1D data by summation of the ^{2}D and ^{3}D chromatographic dimensions of the GC^3 data. Likewise, to obtain $^2w^*_\mathrm{b}$, Gaussian curves were fitted to reconstructed GC×GC data following summation of the ^{3}D chromatographic dimension.

The ${}^3_1SE_\mathrm{M}$ obtained from the experimental data were corroborated by simulated data. Two sets of 1D-GC and GC^3 data were simulated based on the experimental results obtained under Conditions 1 and 2. Specifically, $^1w_\mathrm{b}$ and $^2w_\mathrm{b}$ were kept constant at the experimental average, and $^3w_\mathrm{b}$ was varied in accordance with the experimentally observed range for each of the two conditions. Retention times were assigned randomly from the uniform distribution on ^{1}D and ^{2}D to ensure that the entire range of possible phasing cases would be obtained. Gaussian noise was added to both the 1D-GC and GC^3 data which was comparable to that observed in the experimental GC^3 data (*vide infra*). Five hundred single component peaks (SCPs) were simulated for each of the two conditions studied, and ${}^3_1SE_\mathrm{M}$ were obtained algorithmically. For each SCP, $^1\theta$ and $^2\theta$ were measured by fitting Gaussian curves and application of Eq. (4.28). ${}^3_1SE_\mathrm{T}$ was calculated for each SCP, using the measured $^1\theta$ and $^2\theta$ and the $^1w_\mathrm{b}$, $^2w_\mathrm{b}$ and $^3w_\mathrm{b}$ as input into the simulation, via Eq. (4.27). Thus, ${}^3_1SE_\mathrm{M}$ and ${}^3_1SE_\mathrm{T}$ were obtained for both experimental and simulated data. All data analysis and simulations were performed in MATLAB 2019a (The Mathworks, Inc., Natick, MA, U.S.A.). The only preprocessing applied to the data was baseline centering around zero.

4.4 Results and discussion

4.4.1 Validation of the proposed model

We begin by investigating how well the derived expressions explain the ${}^3_1SE$ observed experimentally under the two sets of GC^3 operating conditions (Table B1). Conditions 1 were similar to those in our recent report [30], optimized with consideration of the peak capacity, and produced the chromatographic data in Fig. 4.3. By summing away the ^{3}D chromatographic dimension of the GC^3 data (Fig. 4.3A), the result appears like a traditional GC×GC-TIC chromatogram. Note that wraparound was intentionally permitted in order to maximize the usage of the separation space in these dimensions. Alternatively, by reshaping the GC^3 data into a 3D tensor, each peak can be represented as an ellipsoid, as shown in the isosurface chromatogram of the section at ${}^1t_{\mathrm{sep}} = 3550$ sto 3710 s in Fig. 4.3B. The raw data as in Fig. 4.3, facilitated a comparison of theoretically predicted ${}^3_1SE$s (${}^3_1SE_{\mathrm{T}}$) to those measured from experimental data (${}^3_1SE_{\mathrm{M}}$).

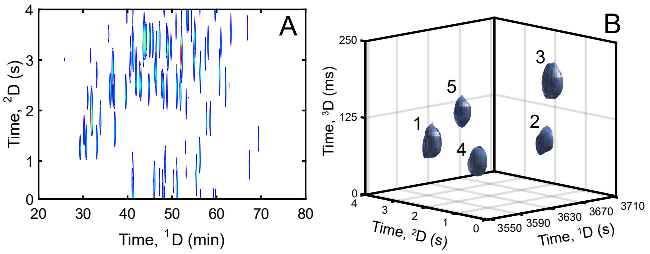


Figure 4.3: (A) Summation of the GC^3-total ion count (TIC) data along 3D produces a plot that appears like a GC×GC-TIC chromatogram. (B) Isosurface 3D TIC chromatogram of a section containing five analytes ordered by ${}^1t_{\mathrm{R}}$: (1) dodecanoic acid, (2) diphenylmethane, (3) dibenzofuran, (4) hexadecane, and (5) diphenylsulfide.

The validation study is introduced first in the context of Conditions 1. Unfolded data are given in the left panel of Fig. 4.4 and successive modulation stages for each analyte are enabled column-wise. For discussion purposes, four of the twenty analytes were selected from each of the four cases in Table 4.1 whose measured phase was close to the anticipated ideal ${}^1\theta$ and ${}^2\theta$ (≤ 0.15 absolute error, see bottom row for exact phase measurements). Note that only the GC^3 data are used to measure ${}^1\theta$ and ${}^2\theta$, because retention time shifting from enabling successive stages of modulation may cause the phasing in the GC×GC data to be different than that of the GC^3 data. The utility of the 1D-GC and GC×GC data is to demonstrate the signal enhancement obtained from successive modulation, to obtain an accurate measurement of ${}^1h$ from the ^{1}D peak, and to obtain unmodulated peak widths. Although there is minor tailing in the GC×GC data, the ${}^2w_{\mathrm{b}}$ varied at most by 4.3% from ${}^2w_{\mathrm{b}}$ for ideal Gaussian peaks of equivalent area. The insets in the bottom row (GC^3 data) display a zoom-in of the tallest set of ^{3}D peaks to better highlight the ${}^2\theta$.

Overall, average unmodulated peak widths, ${}^3_1SE_{\mathrm{M}}$, and ${}^3_1SE_{\mathrm{T}}$ were obtained for twenty analytes (including the four shown in the left panel of Fig. 4.4) and are given in Table 4.2. The ${}^3_1SE_{\mathrm{T}}$ are calculated on a per-analyte basis using the peak widths measured in the unmodulated chromatograms and ${}^1\theta$ and ${}^2\theta$ obtained by curve fitting and application of Eq. (4.28). The results are summarized in Table 4.2; overall for Conditions 1, the average ${}^3_1SE_{\mathrm{M}}$ was 97 ± 22, in excellent agreement with the ${}^3_1SE_{\mathrm{T}}$ of 97 ± 18.

Using $t_{\mathrm{sep}} = 50$ min, an average ideal 3D peak capacity, $n_{\mathrm{c,3D}}$, of 11,470 (± 3,600) was obtained (Table 4.2). Here, the peak capacity is calculated as

$$n_{\mathrm{c,3D}} = {}^1n_{\mathrm{c}} \times {}^2n_{\mathrm{c}} \times {}^3n_{\mathrm{c}} \tag{4.29}$$

Because the peak widths for a given dimension did not vary significantly, ${}^{\mathrm{k}}n_{\mathrm{c}}$ is estimated as ${}^{\mathrm{k}}t_{\mathrm{sep}}/{}^{\mathrm{k}}w_{\mathrm{b}}$. Reporting $n_{\mathrm{c,3D}}$ facilitates comparison with prior work, where often only the ideal peak capacity was reported [24,27,28,30]. However, a more useful metric is the *effective* 3D peak capacity, $n^*_{\mathrm{c,3D}}$, which incorporates correction for the sampling-induced broadening of ${}^1w_{\mathrm{b}}$ and ${}^2w_{\mathrm{b}}$, designated as ${}^1w^*_{\mathrm{b}}$ and ${}^2w^*_{\mathrm{b}}$. Thus, $n^*_{\mathrm{c,3D}}$ is calculated as

$$n^*_{\mathrm{c,3D}} = {}^1n^*_{\mathrm{c}} \times {}^2n^*_{\mathrm{c}} \times {}^3n_{\mathrm{c}} \tag{4.30}$$

where ${}^1n^*_{\mathrm{c}}$ and ${}^2n^*_{\mathrm{c}}$ are estimated using ${}^1w^*_{\mathrm{b}}$ and ${}^2w^*_{\mathrm{b}}$, respectively. The average ${}^1w^*_{\mathrm{b}}$ and ${}^2w^*_{\mathrm{b}}$ were 17.2 s (± 2.9 s) and 550 ms (± 80 ms) respectively; therefore, the average $n^*_{\mathrm{c,3D}}$ was 9,070 ± 2,660 (Table 4.2). This corresponds to a ~26% loss in peak capacity from undersampling, an improvement over prior work, where the loss was ~41% [30]. In summary, the results demonstrate that GC^3 exceptional signal enhancement of ^{1}D peaks can be obtained concurrently with high peak capacity separations for the resolution of complex mixtures.

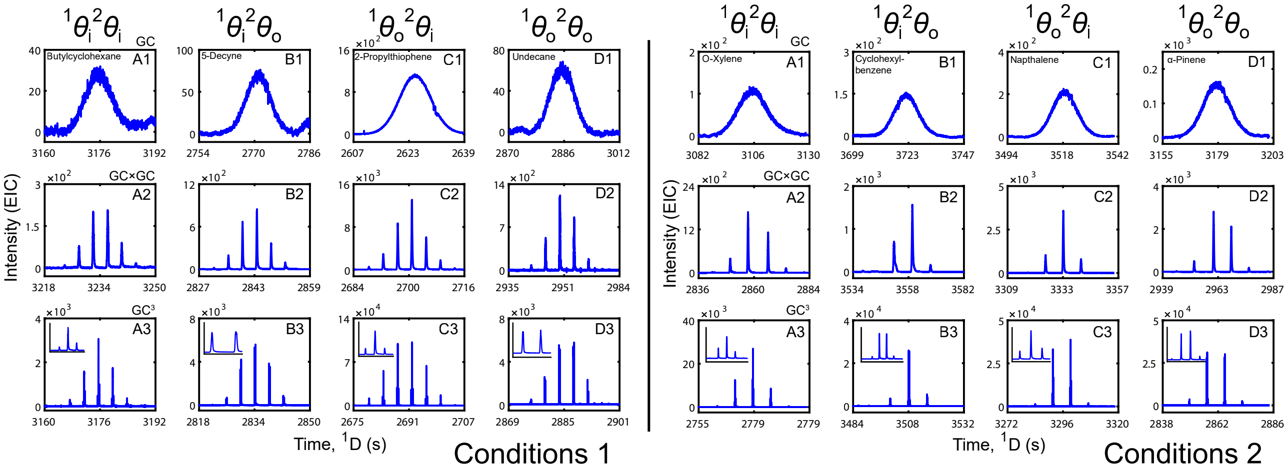


Figure 4.4: Unfolded 1D-GC, GC×GC, and GC^3 data for four representative analytes of each phasing case studied under Conditions 1 (left) and Conditions 2 (right). Note that only the GC^3 data are used to compute the phases, and an insert of the most intense set of 3D peaks is provided to illustrate $^2\theta$.

Next, the results obtained under Conditions 2 for the same twenty analytes are discussed and compared with Conditions 1. Based on the equations in Table 4.1, increasing $^1w_{\rm b}$ and decreasing $3w_{\rm b}$ should increase the $^3_1SE_{\rm M}$. Therefore, the auxiliary pressure ($P_{\rm aux}$) applied to the DPGM assembly was the key factor changed under Conditions 2 to decrease $^3w_{\rm b}$. By increasing $P_{\rm aux}$, $^3w_{\rm b}$ decreases due to the increased flow ratio ($^3F/^2F$), which governs the injected pulse width under differential flow modulation conditions [46]. Data for four analytes of each phasing case are given in the right panel of Fig. 4.4. Note that although the same twenty analytes were studied, the four analytes included in Fig. 4.4 were necessarily different in order to provide one analyte of each phase case. The average $^1w_{\rm b}$ and $^2w_{\rm b}$ along with $n_{\rm c,3D}$ for the 20 analytes studied under Conditions 2 are also given in Table 4.2. Notably, the $^3_1SE_{\rm M}$ with Conditions 2 were substantially higher than for Conditions 1 with an average value of 181 ± 27 and were still in excellent agreement with the average $^3_1SE_{\rm T}$ of 176 ± 26. However, although $^1w_{\rm b}$ and $^2w_{\rm b}$ were broader with Conditions 2, the longer $^2P_{\rm M}$ significantly increased $^2n_{\rm c}$, and along with the lower $^3w_{\rm b}$ resulted in a higher $n_{\rm c,3D}$ and $n^*_{\rm c,3D}$ (Table 4.2). These assessments are reinforced in a comparison of Conditions 1 and 2 for the $^3_1SE_{\rm M}$ and $^3_1SE_{\rm T}$, for both the theoretical boundary conditions and simulated data, as shown in Fig. 4.5 and discussed next.

We now investigate the overall agreement between the $^3_1SE_{\rm M}$ and $^3_1SE_T$ of both experimental data (both Conditions) and simulated data. The utility of the simulated data is to investigate the accuracy of the theory using a large sample size and without experimental error. The data were simulated to correspond to the experimental conditions. That is, a set of 500 SCPs were simulated using the average $^1w_{\rm b}$, $^2w_{\rm b}$, and range of $^3w_{\rm b}$ obtained under Conditions 1 (Table 4.2), and likewise 500 SCPs were simulated to match the peak widths obtained under Conditions 2. Like the experimental data, the phase of the simulated data was measured by a fitting Gaussian function and application of Eq(4.28). All 1040 $^3_1SE_{\rm M}$ (500 simulated SCPs and 20 experimental analytes from each set of conditions) are plotted against their $^3_1SE_{\rm T}$, over a 1:1 line representing ideal agreement, in Fig. 4.5A.

To assess the accuracy of the theory, linear regression models were fitted to the experimental data and simulated data. The regression model for Conditions 1 and 2 for the simulated data was $^3_1SE_{\rm M} =$

	${}^1w_{\mathrm{b}}$ (s)	${}^2w_{\mathrm{b}}$ (ms)	${}^3w_{\mathrm{b}}$ (ms)	${}^3_1SE_{\mathrm{M}}$	${}^3_1SE_{\mathrm{T}}$	$n^*_{\mathrm{c,3D}}$	${}^3_1S/N_{\mathrm{Rel}}$
Conditions 1	15.9 (3.0)	480 (86)	37 (5)	97 (22)	97(18)	11,470 (3600)	38.8 (11.2)
Conditions 2	23.4 (4.5)	711 (99)	29 (3)	181 (27)	176 (26)	13,060 (3530)	59.0 (27.2)

Table 4.2: Average (1 s.d) peak widths, ${}^3_1SE_{\mathrm{M}}$, ${}^3_1SE_{\mathrm{T}}$, $n^*_{\mathrm{c,3D}}$, S/N_{Rel} for the GC^3 data analyzed by two sets of conditions.

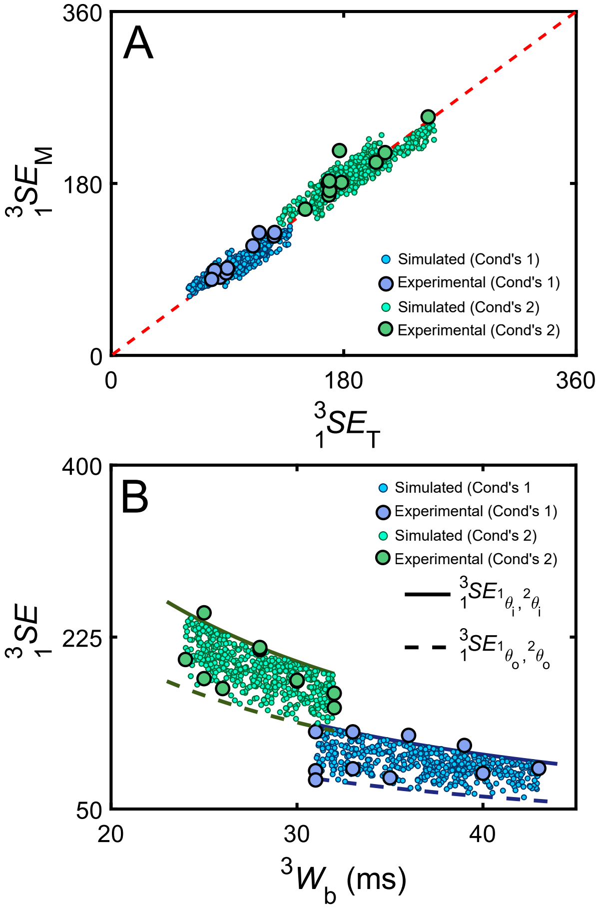


Figure 4.5: (A) Comparison of measured ${}^3_1SE$ ${}^3_1(SE_{\mathrm{M}})$ to theoretical ${}^3_1SE$ (${}^3_1SE_{\mathrm{T}}$) via the Eqs. in Table 4.1 using two sets of experimental and simulated data. The red dashed line indicates ideal agreement with a slope of 1 and a y-intercept of 0. (B) Data in (A) are plotted vs ${}^3w_{\mathrm{b}}$. The solid and dashed lines represent the maximum and minimum predicted ${}^3_1SE$ for the average experimental conditions.

$1.00{}^3_1SE_T + 0.78$ and for the experimental data over Conditions 1 and 2 was ${}^1_3SE_{\mathrm{M}}$= $1.04{}^1_3SE_{\mathrm{T}}$ - 3.04. To verify that the slopes of these two models were not statistically different from the 1:1 line, a *t*-test was performed on the regression coefficients. For the simulated data, the *t*-test yielded $p = 0.277$ ($p > 0.05$); therefore, we conclude that the theory is accurate. A *t*-test comparing the experimental data to the 1:1 line also confirmed there was no statistical difference ($p = 0.114$). We note that in the prior report [12] no validation of the theory was performed.

Although there is some scatter about the 1:1 line for the experimental data, the overall agreement between ${}^3_1SE_{\mathrm{M}}$ and ${}^3_1SE_{\mathrm{T}}$ is satisfactory. The minor discrepancies may arise from intra-analyte variation in area and/or the difference between the analyte profile had it been detected at the end of the ^{2}D column versus the Gaussian model. Nonetheless, the model is advantageous in that the analyst can reliably predict the ${}^3_1SE$ based on readily measurable chromatographic parameters. Moreover, the theoretical treatment of signal enhancement should in principle enable the analyst to readily formalize method optimization.

In Fig. 4.5B, the ${}^3_1SE_{\mathrm{M}}$ are plotted against ${}^3w_{\mathrm{b}}$. The solid line and dashed line for each condition represent the theoretical ${}^3_1SE_{{}^1\theta_{\mathrm{i}},{}^2\theta_{\mathrm{i}}}$ and ${}^3_1SE_{{}^1\theta_{\mathrm{o}},{}^2\theta_{\mathrm{o}}}$. For a given set of w_{b} and ρ, the maximum ${}^3_1SE$ is obtained with ${}^3_1SE_{{}^1\theta_{\mathrm{i}},{}^2\theta_{\mathrm{i}}}$and the minimum with ${}^3_1SE_{{}^1\theta_{\mathrm{o}},{}^2\theta_{\mathrm{o}}}$. In accordance with the theory, all of the simulated data fall within these two boundaries. The experimental data shows good agreement, although some points fall just outside the boundaries. This does not necessarily disagree with the theory, since the lines are drawn with the average ${}^1w_{\mathrm{b}}$ and ${}^2w_{\mathrm{b}}$, and a particular analyte may have broader and/or narrower peak widths relative to the average, causing it to shift above or below the boundaries. Our interpretation of Fig. 4.5B is that ${}^3w_{\mathrm{b}}$ appears to be an extremely important factor in optimizing ${}^3_1SE$. Likewise, minimizing ${}^3w_{\mathrm{b}}$ is also critical for maximizing peak capacity. Therefore, an analyst seeking to improve ${}^{3}_{1}SE$ and/or $n_{\mathrm{c,3D}}$ should always employ experimental conditions and instrumentation that minimize ${}^{3}w_{\mathrm{b}}$.

4.4.2 Signal-to-noise considerations

Although the present study demonstrates GC^3 can provide exceptional signal enhancement, a more useful metric in practice is ${}^{3}_{1}S/N_{\mathrm{Rel}}$, given by Eq. (4.3). We define ${}^{1}s_{\mathrm{N}}$ as one standard deviation of the baseline noise of a ^{1}D separation and ${}^{3}s_{\mathrm{N}}$ as one standard deviation of the baseline noise of a GC^3 separation. Here, ${}^{1}s_{\mathrm{N}}$ and ${}^{3}s_{\mathrm{N}}$ are measured over an event width equal to the 6σ peak width on the 1D-GC and GC^3 runs, respectively, which is considered the minimum time segment required for baseline correction and subsequent quantification. The noise sections were measured for each replicate of both the 1D-GC and GC^3 data. The noise sections were selected within a maximum tolerance of two min from the retention time of the ^{1}D or ^{3}D peak. Herein, a 500 Hz spectrum acquisition frequency was used throughout, which is adequate for sampling 20-40 ms ^{3}D peaks, but far higher than required to sample 10-30 s wide ^{1}D peaks. Therefore, the data-dense 1D-GC chromatograms can tolerate averaging and/or digital filtering to improve the S/N without compromising the underlying signal. To make an objective comparison, the ^{1}D noise section for each analyte was resampled (i.e., boxcar averaged) such that ${}^{1}s_{\mathrm{N}}$ was composed of the same number of data points (spectra) as ${}^{1}3s_{\mathrm{N}}$ for a given analyte, denoted ${}^{1}s_{\mathrm{N,RS}}$. Ideally, one could collect 1D-GC data using a lower acquisition frequency; however, the lowest acquisition frequency available with the TOFMS detector used in this study required averaging in order to make the comparison.

The process of calculating ${}^{3}_{1}S/N_{\mathrm{Rel}}$ is illustrated for 2-methylthiophene (Conditions 2) in Fig. 4.6. First, a section of noise equal in length to the 6σ ^{1}D peak width (46.6 s or 23325 points) is located near the peak in the raw ^{1}D data (Fig. 4.6A). The calculation of ${}^{1}s_{\mathrm{N}}$ over this interval is 3.74 but should be improved by resampling. The data has been resampled in Fig. 4.6B, yielding ${}^{1}s_{\mathrm{N,RS}} = 1.69$ measured over the same interval. In Fig. 4.6C, measurement of ${}^{3}s_{\mathrm{N}}$ is performed on the raw ^{3}D data for a section of noise equal in length to the 6 ^{3}D peak width (36 ms or 18 points). Given a ${}^{3}_{1}SE_{\mathrm{M}}$ of 202, the measurements in Fig. 4.6B-C yield a ${}^{3}_{1}S/N_{\mathrm{Rel}}$ of 111 via Eq. (4.3).

Using this method, the average ${}^{3}_{1}S/N_{\mathrm{Rel}}$ for all analytes under Conditions 1 and 2 is given Table 4.2. Overall, the average ${}^{3}_{1}S/N_{\mathrm{Rel}}$ were 38.8 ($\pm$11.2) and 59.0 ($\pm$ 27.2) for Conditions 1 and 2, respectively, which is appreciably higher than typically obtained for GC×GC, with reported values ranging from ~5-10 [12,17], although methods for calculating ${}^{3}_{1}S/N_{\mathrm{Rel}}$ vary. It should be noted that the 1D-GC separation was not fully optimized if the sole purpose was to optimize 1D-GC, and that ${}^{1}w_{\mathrm{b}}$ of ~ 6 s are attainable using the same ^{1}D column dimensions and carrier gas, following extensive optimization. Thus, we anticipate that ${}^{3}_{1}SE_{\mathrm{M}}$ for GC^3 with an optimized 1D-GC separation to be ~2.7-fold (Conditions 1) and ~3.9-fold (Conditions 2) lower than obtained herein, concurrent with lower ${}^{3}_{1}S/N_{\mathrm{Rel}}$, depending on the characteristics of the noise. However, GC×GC and GC^3 (as well as 1D-GC) are frequently performed with a less-than-optimal 1D separation. The study herein is intended to provide results that are indicative of what is likely to be achieved in typical practice of GC×GC and GC^3.

One might expect that ${}^{3}_{1}S/N_{\mathrm{Rel}}$ could be appreciably lower if the 1D baseline noise were random, as ${}^{1}s_{\mathrm{N}}$ of white noise should improve by a factor of $n^{1/2}$ by averaging, where n is the number of data points averaged. This suggests that ${}^{3}_{1}S/N_{\mathrm{Rel}}$ would be ~30-fold lower than the ${}^{3}_{1}SE_{\mathrm{M}}$. However, chromatographic systems are often dominated by low-frequency noise [9,47]. The presence of low-frequency noise can be inferred from visual inspection of both Fig. 4.6A and Fig. 4.6B. We observed low-frequency noise for every analyte studied. The power spectra of the noise section and the raw ^{1}D peak of 2-methylthiophene (Fig. 4.6) are shown in Fig. B3. Both the noise section and ^{1}D peak occupy the same frequency domain of about -0.05 to 0.05 Hz. Therefore, smoothing or filtering methods applied in the

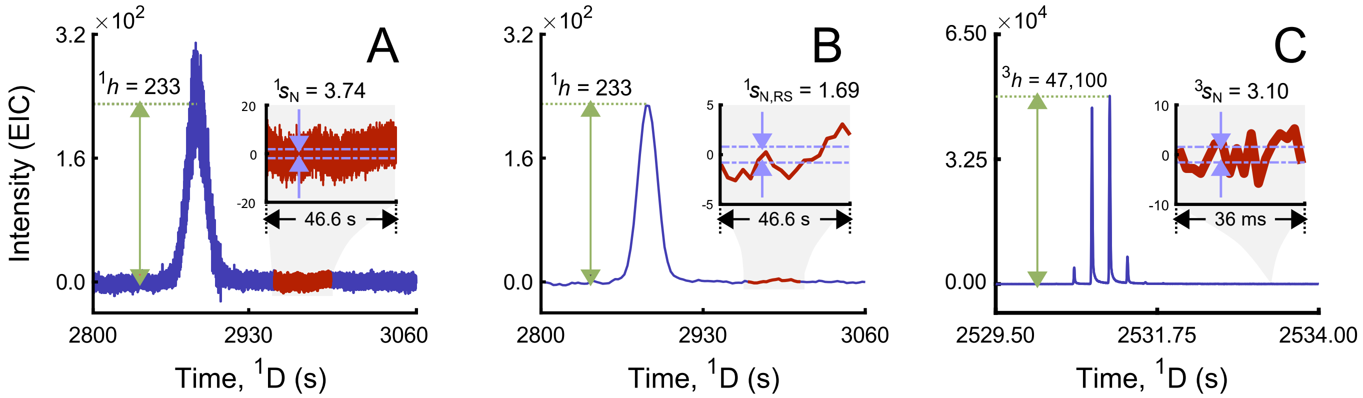


Figure 4.6: (A) Raw 1D-GC data for 2-methylthiophene (Conditions 2), illustrating the measurement of ^1h and $^1s_\mathrm{N}$. (B) 1D-GC data after resampling. The same noise region is used to measure $^1s_\mathrm{N,RS}$. (C) Raw GC^3 data illustrating measurement of ^3h and $^3s_\mathrm{N}$.

frequency domain are unlikely to further enhance the S/N without compromising the underlying signal. In fact, one impetus for the development of thermal modulation, which subsequently led to the invention of GC×GC, was to attenuate low-frequency noise by shifting the signal to a higher frequency domain [9], which is further enhanced here by tandem modulation. Although noise is discussed here in the context of TT-GC^3-TOFMS, noise in other chromatographic systems, such as GC with flame ionization detection (GC-FID) [9,48] and LC with absorbance detection [47,49], has been observed to be limited by low-frequency noise.

In addition to instrumental noise, chemical noise can hinder analyte detectability. Column bleed and residual solvent are well-known examples of chemical noise, and manifest in GC×GC chromatograms as streaks with decreasing retention as the oven temperature increases. Further, significant chemical noise arises from matrix components, including those which are at or below the LOD [50,51]. Chemical noise can be largely mitigated by employing high peak capacity separations, because the analyte(s) of interest are more likely to be separated from matrix interferences. The enhanced peak capacity of GC×GC reduces the effects of chemical noise relative to 1D-GC, and we anticipate this benefit is further enhanced by the additional peak capacity of GC^3.

4.5 Conclusion

To date, interest in comprehensive 3D separations has focused on obtaining a high peak capacity and/or unique selectivity [24,28,29], Herein, we provide evidence that these systems can also enhance signal and the S/N. The enhancement is attributed not only to spatial and temporal peak compression, but also to shifting the signal to a higher frequency domain where low-frequency noise is less detrimental.

While the theory presented herein is discussed and evaluated in the context of the TT-GC^3-TOFMS instrument, the principles can be easily extended to other 3D platforms with modulator duty cycles 100%, provided that the separations are carried out *in-time*. The model could also be applied to a select subset of peaks in systems which are not truly comprehensive, such as "selective comprehensive" multidimensional LC [52] and sequential-hybrid 3D GC [53]. It is our hope that further developing the theory of 3D separations will encourage others to explore this exciting area.

4.6 References

(1) Z. Liu, J.B. Phillips, Comprehensive two-dimensional gas chromatography using an on-column thermal modulator interface, *J. Chromatogr. Sci.* **1991**, 29 227– 231.

(2) Zanella, D.; Focant, J.; Franchina, F. A. 30th Anniversary of comprehensive two-dimensional gas chromatography: Latest advances. *Anal. Sci. Adv.* **2021**, 1-12.

(3) Amaral, M. S. S.; Nolvachai, Y.; Marriott, P. J. Comprehensive two-dimensional gas chromatography advances in

technology and applications: Biennial update. *Anal. Chem.* **2020**, *92*, 85–104.

(4) Prebihalo, S. E.; Berrier, K. L.; Freye, C. E.; Bahaghighat, H. D.; Moore, N. R.; Pinkerton, D. K.; Synovec, R. E. Multidimensional gas chromatography: Advances in instrumentation, chemometrics, and applications. *Anal. Chem.* **2018**, *90*, 505–532.

(5) X. Guan, J. Luong, Z. Yu, H. Jiang, Quasi-stop-flow modulation strategy for comprehensive two-dimensional gas chromatography, *Anal. Chem.* **2020**, *92*, 6251-6256.

(6) C.N. Cain, S. Schöneich, R.E. Synovec, Development of an enhanced total ion current chromatogram algorithm to improve untargeted peak detection, *Anal. Chem.* **2020**, *92*, 11365-11373.

(7) Franchina, F. A.; Dubois, L. M.; Focant, J. F. In-Depth dannabis multiclass metabolite profiling using sorptive extraction and multidimensional gas chromatography with low- and high-resolution mass spectrometry. *Anal. Chem.* **2020**, *92*, 10512–10520.

(8) M.S. Klee, J. Cochran, M. Merrick, L.M. Blumberg, Evaluation of conditions of comprehensive two-dimensional gas chromatography that yield a near-theoretical maximum in peak capacity gain, *J. Chromatogr. A* **2015,** *1383*, 151–159.

(9) Z. Liu, J.B. Phillips, Sensitivity and detection limit enhancement of gas chromatographic detection by thermal modulation, *J. Microcol. Sep.* **1994**, *6*, 229-235.

(10) H.-J. de Geus, J. de Boer, J.B. Phillips, E.B. Ledford Jr., U.A.Th. Brinkman, Increased signal amplitude due to mass conservation in a thermal desorption modulator, *J. High Resolut. Chromatogr.* **1998**, *21*, 411-413.

(11) Kinghorn, R. M.; Marriott, P. J. Enhancement of signal-to-noise ratios in capillary gas chromatography by using a longitudinally modulated cryogenic system. *J. High Resolut. Chromatogr.* **1998**, *21*, 32-38.

(12) Lee, A. L.; Bartle, K. D.; Lewis, A. C. A model of peak amplitude enhancement in orthogonal two-dimensional gas chromatography. *Anal. Chem.* **2001**, *73*, 1330-1335.

(13) Hamilton, J. F.; Lewis, A. C.; Bartle, K. D. Peak Amplitude and Resolution in Comprehensive Gas Chromatography Using Valve Modulation. *J. Sep. Sci.* **2003**, *26*, 578-584.

(14) Krupčík, J.; Májek, P.; Gorovenko, R.; Špánik, I.; Sandra, P.; Armstrong, D. W. On the Determination of a Detector Response Enhancement Factor for Flow Modulated Comprehensive Two-Dimensional Gas Chromatography. *J. Chromatogr. A* **2013**, *1286*, 235-240.

(15) Zoccali, M.; Giocastro, B.; Tranchida, P. Q.; Mondello, L. Use of a Recently Developed Thermal Modulator within the Context of Comprehensive Two-Dimensional Gas Chromatography Combined with Time-of-Flight Mass Spectrometry: Gas Flow Optimization Aspects. *J. Sep. Sci.* **2019**, *42*, 691-697.

m(16) Hayward, T.; Gras, R.; Luong, J. Flow-Modulated Targeted Signal Enhancement for Volatile Organic Compounds. *J. Sep. Sci.* **2016**, *39*, 2284-2291.

(17) Mostafa, A.; Górecki, T. Sensitivity of comprehensive two-dimensional gas chromatography (GCXGC) versus one-dimensional gas chromatography (1D GC). *LC-GC Eur.* **2013**, *26*, 672- 676.

(18) K.J. Jobst, J.V. Seeley, E.J. Reiner, L. Mullin, A. Ladak, A, Enhancing the sensitivity of atmospheric pressure ionization mass spectrometry using flow modulated gas chromatography, *LC-GC North Am.* **2018**, *14*, 23-28.

(19) J.M. Davis, Statistical theory of spot overlap for n-dimensional separations. *Anal. Chem.* **2014**, *65*, 2014–2023.

(20) Davis, J. M. Statistical-Overlap Theory for Elliptical Zones of High Aspect Ratio in Comprehensive Two-Dimensional Separations. *J. Sep. Sci.* **2005**, *28*, 347–359.

(21) E.B. Ledford, C.A. Billesbach, GC3: Comprehensive three-dimensional gas chromatography, J. *High Resolut. Chromatgr.* **2000,** *23*, 205-207.

(22) N.E. Watson, W.C. Siegler, J.C. Hoggard, R.E. Synovec, Comprehensive three-dimensional gas chromatography with parallel factor analysis, *Anal. Chem.* **2007**, *79*, 8270–8280.

(23) W.C. Siegler, J.A. Crank, D.W. Armstrong, R.E. Synovec, Increasing selectivity in comprehensive three-dimensional gas chromatography via an ionic liquid stationary phase column in one dimension, *J. Chromatogr. A* **2010**, *1217*, 3144-3149.

(24) N.E. Watson, H.D. Bahaghighat, K. Cui, R.E. Synovec, Comprehensive three-dimensional gas chromatography with time-of-flight mass spectrometry, *Anal. Chem.* **2017**, *89*, 1793–1800.

(25) Bahaghighat, H. D.; Freye, C. E.; Gough, D. V.; Sudol, P. E.; Synovec, R. E. Ultrafast Separations via Pulse Flow Valve Modulation to Enable High Peak Capacity Multidimensional Gas Chromatography. *J. Chromatogr. A* **2018**, *1573*, 115-124.

(26) D.V. Gough, H.D. Bahaghighat, R.E. Synovec, Column selection approach to achieve a high peak capacity in comprehensive three-dimensional gas chromatography, *Talanta* **2019**. *195*, 822-829.

(27) D.V. Gough, D.H. Song, S. Schöneich, S.E. Prebihalo, R.E. Synovec, Development of ultrafast separations using negative pulse partial modulation to enable new directions in gas chromatography, *Anal. Chem.* **2019**, *91*, 7328-7335.

(28) V.H.C. Ferreira, L.W. Hantao, R.J. Poppi, Consumable-free comprehensive three-dimensional gas chromatography and PARAFAC for determination of allergens in perfumes, Chromatographia **2020**, *83*, 581-592.

(29) T.J. Trinklein, S.E. Prebihalo, C.G. Warren, G.S. Ochoa, R.E. Synovec, Discovery-based analysis and quantification for comprehensive three-dimensional gas chromatography flame ionization detection data, *J. Chromatogr. A* **2020**, *1623*, 461190.

(30) Trinklein, T. J.; Schöneich, S.; Sudol, P. E.; Warren, C. G.; Gough, D. V.; Synovec, R. E. Total-Transfer Comprehensive Three-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometry. *J. Chromatogr. A* **2020**, *1634*,

461654.

(31) Guiochon, G.; Beaver, I. A.; Gonnord, M. F.; Siolffi, A. M.; Zakaria, M. Theoretical Investigation of the Potentialities of the Use of a Multidimensional Column in Chromatography. *J. Chromatogr. A* **1983**, *255*, 415-437.

(32) Moore, A. W.; Jorgenson, J. W. Comprehensive Three-Dimensional Separation of Peptides Using Size Exclusion Chromatography/ Reversed Phase Liquid Chromatography/Optically Gated Capillary Zone Electrophoresis. *Anal. Chem.* **1995**, *67*, 3456-3463.

(33) Wouters, B.; Davydova, E.; Wouters, S.; Vivo-Truyols, G.; Schoenmakers, P. J.; Eeltink, S. Towards Ultra-High Peak Capacities and Peak-Production Rates Using Spatial Three-Dimensional Liquid Chromatography. *Lab Chip* **2015**, *15*, 4415-4422.

(34) Themelis, T.; Amini, A.; De Vos, J.; Eeltink, S. Towards Spatial Comprehensive Three-Dimensional Liquid Chromatography: A Tutorial Review. *Anal. Chim. Acta* **2021**, *1148*, 238157.

(35) Venter, P.; Muller, M.; Vestner, J.; Stander, M. A.; Tredoux, A. G. J.; Pasch, H.; De Villiers, A. Comprehensive Three-Dimensional LC×LC×Ion Mobility Spectrometry Separation Combined with High-Resolution MS for the Analysis of Complex Samples. *Anal. Chem.* **2018**, *90*, 11643-11650.

(36) Trinklein, T. J.; Gough, D. V.; Warren, C. G.; Ochoa, G. S.; Synovec, R. E. Dynamic Pressure Gradient Modulation for Comprehensive Two-Dimensional Gas Chromatography. *J. Chromatogr. A* **2020**, *1609*, 460488.

(37) Schöneich, S.; Trinklein, T. J.; Warren, C. G.; Synovec, R. E. A Systematic Investigation of Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry with Dynamic Pressure Gradient Modulation for High Peak Capacity Separations. *Anal. Chim. Acta* **2020**, *1134*, 115-124.

(38) Seeley, J.V. Theoretical Study of Incomplete Sampling of the First Dimension in Comprehensive Two-Dimensional Chromatography. *J. Chromatogr. A* **2002**, *962*, 21-27.

(39) Horváth, K.; Fairchild, J. N.; Guiochon, G. Detection Issues in Two-Dimensional on-Line Chromatography. *J. Chromatogr. A* **2009**, *1216*, 7785-7792.

(40) Khummueng, W.; Harynuk, J.; Marriott, P.J. Modulation Ratio in Comprehensive Two-Dimensional Gas Chromatography. *Anal. Chem.* **2006**, *78*, 4578-4587.

(41) Marriott, P. J.; Mühlen, C.v. The Modulation Ratio in Comprehensive Two-Dimensional Gas Chromatography: A Review of Fundamental and Practical Considerations. *Sci. Chromatogr.* **2016**, *8*, 7-23.

(42) Siegler, W. C.; Fitz, B. D.; Hoggard, J. C.; Synovec, R. E. Experimental Study of the Quantitative Precision for Valve-Based Comprehensive Two-Dimensional Gas Chromatography. *Anal. Chem.* **2011**, *83*, 5190-5196.

(43) Davis, J. M.; Stoll, D. R.; Carr, P. W. Effect of First-Dimension Undersampling on Effective Peak Capacity in Comprehensive Two-Dimensional Separations. *Anal. Chem.* **2008**, *80*, 461-473.

(44) Schöneich, S.; Gough, D. V.; Trinklein, T. J.; Synovec, R. E. Dynamic Pressure Gradient Modulation for Comprehensive Two-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometry Detection. *J. Chromatogr. A* **2020**, *1620*.

(45) Pinkerton, D. K.; Parsons, B. A.; Synovec, R. E. Method to Determine the True Modulation Ratio for Comprehensive Two-Dimensional Gas Chromatography. *J. Chromatogr. A* **2016**, *1476*, 114-123.

(46) LaClair, R. W.; Bueno, P. A.; Seeley, J. V. A Systematic Analysis of a Flow-Switching Modulator for Comprehensive Two-Dimensional Gas Chromatography. *J. Sep. Sci.* **2004**, *27*, 389-396.

(47) Renn, C. N.; Synovec, R. E. Improved Detection in Superspeed Size-Exclusion Chromatography by Reducing On-Column Dilution and Detector Noise. *Anal. Chem.* **1988**, *60*, 1829-1832.

(48) Smit, H. C.; Walg, H. L. Base-Line Noise and Detection Limits in Signal-Integrating Analytical Methods. Application to Chromatography. *Chromatographia* **1975**, *8*, 311-323.

(49) Hayashi, Y.; Matsuda, R.; Poe, R. B. Prediction of Precision from Signal and Noise Measurement in Liquid Chromatography: Limit of Detection. *Chromatographia* **1997**, *41*, 66-74.

(50) Enke, C. G.; Nagels, L. J. Undetected Components in Natural Mixtures: How Many? What Concentrations? Do They Account for Chemical Noise? What Is Needed to Detect Them? *Anal. Chem.* **2011**, *83*, 2539-2546.

(51) N.M. Devitt, J.M. Davis, M.P. Schure, Estimation of low-level components lost through chromatographic separations with finite detection limits. *J. Chromatogr. A* **2020**, *162*, 461266.

(52) Groskreutz, S. R.; Swenson, M. M.; Secor, L. B.; Stoll, D. R. Selective Comprehensive Multi-Dimensional Separation for Resolution Enhancement in High Performance Liquid Chromatography. Part I: Principles and Instrumentation. *J. Chromatogr. A* **2012**, *1228*, 31-40.

(53) D.D. Yan, Y.F. Wong, S.P. Whittock, A. Koutoulis, R.A. Shellie, P.J. Marriott, Sequential hybrid three-dimensional gas chromatography with accurate mass spectrometry: a novel tool for high-resolution characterization of multicomponent samples, *Anal. Chem.* **2018**, *90*, 5264–5271.

Chapter 5

PROFILING OLEFINS IN GASOLINE BY BROMINATION USING GC×GC-TOFMS FOLLOWED BY DISCOVERY-BASED COMPARATIVE ANALYSIS

This chapter was adapated from T.J. Trinklein, J.Jiang, R.E. Synovec, *Anal. Chem.* **2022**, *94*, 9407–9414.

5.1 Introduction

The formation of complex, macromolecular compounds, termed "gums," can severely compromise the performance of gasoline [1,2]. Adverse effects associated with gum formation include clogged filters and fuel lines, increased fuel consumption, engine stalling, knocking, and increased exhaust gas emission [3,4]. Of the hydrocarbon components in gasoline, olefins are the most reactive and are a major contributor to gum formation [1,2,5]. Although complete mechanism(s) for gum formation are not fully known, there is consensus that free radical chain reactions, initiated by peroxide radical addition to the double bond(s) [3,4,6] subsequently lead to numerous routes of gum formation. Therefore, analytical tools are needed to determine the olefin content of gasoline to aid production and quality control. Currently, several standardized methods exist for the estimation of olefin content in gasoline and similar engine fuels. These include several American Society for Testing and Materials (ASTM) methods such as fluorescence indicator absorption (ASTM D1319) [7], heart-cutting gas chromatography (GC) (ASTM D6296) [8], and supercritical-fluid chromatography (ASTM 6550) [9], among others. However, these methods provide only the total olefin content, and do not provide the analyst with information as to the total number of olefin species present or their carbon number. Numerous methods have been reported for olefin analysis, including infrared spectrometry [10], supercritical fluid chromatography [11], and NMR [12]. Hyphenated GC-spectroscopic techniques, such as GC with vacuum ultraviolet detection (GC-VUV) [13] and GC-IR [14] are promising for olefin profiling due to the selectivity in both the ultraviolet and infrared regions for double bonds. Recently, GC stationary phases containing silver ions to selectively separate olefins has been developed [15,16]. Although these phases are not yet commercially available, they appear to be very promising for application to complex petrochemical samples. One intriguing method for estimating olefin content is the measurement of the bromine number, which is the number of grams of bromine which will add to 100 g of a given sample [17]. The method uses the well-characterized reaction where bromine (Br_2) adds across the double bond of olefins to produce vicinal dibromoalkanes [18,19]. (Note that we shall adopt IUPAC nomenclature when necessary to discuss reactions, mechanistic aspects, and analyte names, and use the term olefin to refer more generally to any molecule containing a double-bond, per common usage in the petrochemical field.) The general principles of the Br_2 method date back to at least 1881 [20], and were further developed into a standardized method (via electrometric titration) in the 1940's [17]. The method remains an indispensable metric for estimating olefinic content today (e.g., ASTM D1159 [21] and ISO 3839). Due to the presence of a bromonium ion intermediate, the bromine atoms add via anti-addition and generate an enantiomeric pair for any alkene which is asymmetric about the double bond; symmetric molecules will yield two identical products. The "textbook mechanism" for alkene bromination is provided in the appendix as Fig. C1.

Much like other standard methods, calculation of bromine number only estimates total olefin content and does not provide detailed information about the olefinic profile. Subsequent analysis of the brominated products by a separation method, such as one-dimensional (1D)-GC, facilitates a more precise analysis of the olefin content [22,23]. However, 1D-GC is inherently limited in its peak capacity, thus the total number of mixture components that can be resolved within a reasonable analysis time is also limited [24,25]. Comprehensive two-dimensional (2D) GC, or GC×GC [26–29], provides an enhancement in peak capacity of about an order-of-magnitude [30,31] increased detectability [32,33], and additional selectivity with compound class-ordered separations. Coupling GC×GC with time-of-flight mass spectrometry (GC×GC-TOFMS) creates a powerful instrument, well-suited to the characterization of petroleum-based fuels [34–39]. Unsurprisingly, several researchers have investigated the utility of GC×GC to analyze olefins in various mixtures, including Fischer-Tropsch products [40], commercial dodecenes [41,42], diesel [43], and unsaturated fatty acids [15].

In this report, a method for profiling olefins by combining bromination with GC×GC-TOFMS analysis and comparative analysis is presented. The method uses tile-based F-ratio analysis (FRA) [44–46] to identify the differences between the original gasoline and the brominated gasoline. In this manner, not only the brominated sample, but the combination of the original and brominated samples is leveraged to facilitate the discovery of olefins in gasoline. In principle, both the locations where the brominated products "appear", and the locations where olefins "disappear" should be discoverable. However, we hypothesize that due to the retention properties GC×GC column set applied, the brominated olefins will elute in a less crowded region of the 2D chromatographic space, and their discovery by FRA will be easier than discovery of the original olefins which coelute with other constituents in the gasoline. In order to determine the reaction selectivity, we begin method development by brominating a test mixture containing a homologous series of n-alkanes, 1-alkenes (α-mono-olefins), monoaromatic compounds and naphthalene. Next, the method is applied to profiling olefins in a sample of a gasoline using FRA. Using retention information and an inhouse library of mass spectra obtained from bromination of both the 1-alkenes and a series of C5 olefins, several olefins in the original gasoline could be identified.

5.2 Experimental

5.2.1 Sample preparation

In order to evaluate the reaction chemistry, a test mixture containing a homologous series of n-alkanes, 1-alkenes (α-mono-olefins), and a mixture of monoaromatic compounds (toluene, o-xylene, propylbenzene, 4-ethyltoluene, butylbenzene,) and naphthalene (Sigma-Aldrich, Milwaukee, WI, USA, 95% purity) was prepared at a nominal concentration of 200 ppm (m/v) in HPLC grade n-pentane (Table C1).

A few comments on the selection of the bromination solvent are noted. The halogenation of alkenes is often performed with carbon tetrachloride (CCl_4) or dichloromethane (CH_2Cl_2) as the solvent [18,19], e.g., CCl_4 has been used where bromination was coupled with 1D-GC [22]. However, any non-nucleophilic solvent is suitable, whereas performing the reaction with a nucleophilic solvent such as water or methanol forms a mixture of products that would complicate a discovery-based analysis [18]. Therefore, pentane, a non-nucleophilic solvent, was selected as a greener alternative to CCl_4 and CH_2Cl_2.

A gasoline sample was collected from a fueling station in Seattle, WA, USA. The initial gasoline sample was prepared at 5890 ppm (m/v) in HPLC-grade pentane with 420 ppm of 1-chlorohexane as an internal standard. Three separate sample aliquots were brominated, producing three reaction replicates, and each brominated sample was subsequently diluted again 1:10 in HPLC-grade pentane. Duplicate chromatograms of each brominated sample were acquired for a total of six chromatograms in the brominated sample class which were paired with six chromatograms of the original gasoline sample.

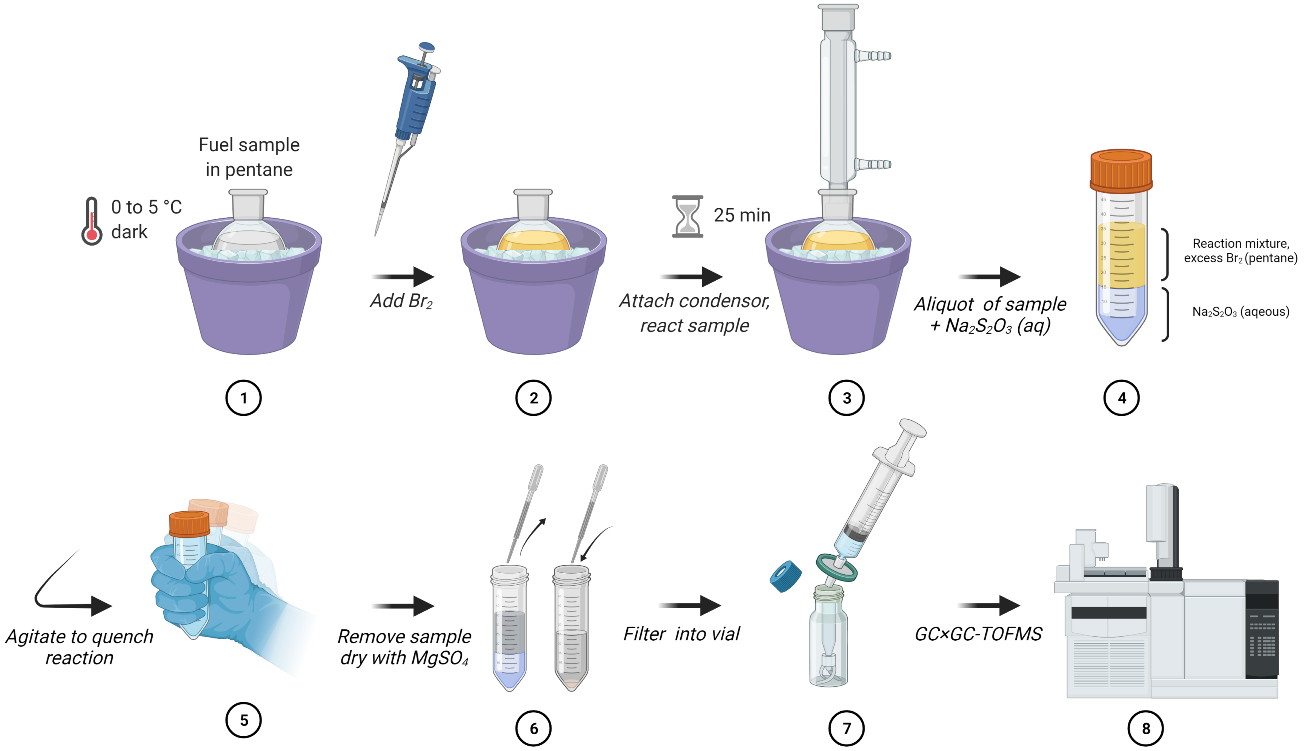


Figure 5.1: Scheme for the bromination and analysis of gasoline.

To aid in the identification of olefins in the gasoline, standards of C5 olefins 2-pentene, 2-methyl-1-butene, 2-methyl-2-butene, 3-methyl-2-butene, and cyclopentene (95% purity, Sigma Aldrich) were prepared at 200 ppm (nominally, m/v) in *n*-pentane and were brominated.

5.2.2 Bromination method

The steps of the bromination method are illustrated in Fig. 5.1. First, a 10.00 ml aliquot of the sample was added to a round-bottom flask covered in aluminum foil and submerged in an ice-water bath. Once the sample was between 0 °C and 5 °C, a 50 µl aliquot of ice-cold Br_2 (Sigma-Aldrich) was added to the flask, immediately turning the reaction mixture to a brownish red color. A Liebig condenser was attached, circulating with cold water, and the reaction was allowed to proceed for 25 min under moderate mixing. Upon completion, a 5.00 ml aliquot of sample was quickly removed and shaken in a flask containing 10.00 ml of ice-cold aqueous saturated sodium thiosulfate ($Na_2S_2O_3$), turning the dark solution (indicating excess Br_2) clear. The colorless solution was dried by the addition of solid anhydrous $MgSO_3$, filtered through a 0.45 µm PTFE syringe filter, and analyzed by GC×GC-TOFMS. To ensure that the sample solvent did not contain any olefins, a method blank was performed by brominating a 10.00 ml aliquot of pentane as the sample. Only trace amounts of brominated pentane were detected (Fig. C2 and C3), verifying that no detectable olefins were present in the solvent.

5.2.3 GC×GC-TOFMS method

Samples were analyzed using a Pegasus BT 4D GC×GC-TOFMS (LECO Corporation, St. Joseph, MI, USA), incorporating an Agilent 7890B gas chromatograph with a quad-jet thermal modulator, a

2D column oven, and a split/spitless injector (maintained at 250 °C). Liquid samples (0.5 µl) were injected at a split ratio of 100:1 using an L-PAL3 automatic sampler (LECO). The ^{1}D column had a semi-polar Rxi-17 phase (50% diphenyl / 50% dimethylpolysiloxane) with dimensions of 29 m × 280 µm d_c × 0.25 µm d_f, while the ^{2}D column had a non-polar DB1 phase (100% dimethylpolysiloxane) with dimensions of 2 m × 180 µm d_c × 0.18 µm d_f. The columns were connected using a SilTite™ union (Fisher Scientific, Fair Lawn, NJ, USA) with the 2D column installed in the ^{2}D oven. The 1D oven temperature was maintained at 40 °C for 1 min, then ramped to 200 °C at 5 °C /min where it was held for 1 min. The 2D oven followed the same program with a +20 °C offset. The modulation period (P_M) was 3 s, including a hot pulse of 0.90 s and a cool time of 0.60 s. Helium (Grade 5, Praxair, Seattle, WA) was the carrier gas at a constant flow rate of 2.00 ml/min. A mass range of 40-400 m/z at unit-resolution was monitored at 100 Hz using electron ionization (EI) with an 85 s acquisition delay to avoid damaging the detector from the solvent front. The transfer line and ion source temperatures were maintained at 285 °C and 225 °C, respectively. Data were acquired using ChromaTOF (Version 5.40) and processed as netCDF files for downstream data analysis.

5.2.4 Data analysis

Data were imported into MATLAB 2021a for all analyses and preprocessed by normalizing each chromatogram to the sum of the TIC (for the test mixture) or the internal standard (for the original and brominated gasoline) and were baseline corrected by subtraction of blank chromatograms (solvent injection only). All chromatograms are plotted following this baseline correction procedure. Comparative analysis of original and brominated gasoline samples was performed using an in-house version of the tile-based F-ratio software [44], functionally equivalent to the commercially available package ChromaTOF Tile (LECO). However, in order to use a novel downstream data analysis method, herein described as "hit list splitting" (*vide infra*), the in-house software was used. A tile size of 6 modulations (18 s) on ^{1}D by 50 spectra (500 ms) on ^{2}D was used, with a cluster window of 4 modulations on ^{1}D and 30 spectra on ^{2}D. A S/N filter of 10 was used based on previous guidelines [47]. The F-ratio for each hit was computed on the recentered tile following the "pinning and clustering" routine [44] and hits were ranked by the top F-ratio m/z tile [47]. "Hit list splitting" was introduced to discriminate between analytes discovered due to disappearance in the original gasoline sample and those discovered due to appearance in the brominated gasoline sample by computing the relative signal in each sample class for each hit using a pure m/z, previously found to strongly correlate to the top F-ratio m/z [48]. Hits having appreciable signal in the original sample but negligible signal in the brominated sample ($S_\mathrm{original} > 0, S_\mathrm{brominated} \approx 0$) were relegated in one hit list, while hits having appreciable signal in the brominated sample but negligible signal in the original sample ($S_\mathrm{brominated} > 0, S_\mathrm{original} \approx 0$) were placed into a second hit list.

5.3 Results and discussion

5.3.1 Method development by test mixture bromination

We began method development by brominating a test mixture containing a homologous series of n-alkanes, 1-alkenes, and several aromatic compounds (Table C1). A GC×GC TIC chromatogram of the original test mixture is shown in Fig. 5.2A. A compound class-ordered separation is obtained, with the retention order on 2D in agreement with the anticipated result for a "reverse" column GC×GC configuration (more polar ^{1}D × less polar ^{2}D). Compounds 1-pentene, 1-hexene, and hexane co-elute with the solvent (pentane), during which time an acquisition delay is employed; therefore, they are not detected in the original test mixture. A chromatogram of the brominated test mixture is given in Fig. 5.2B. The 1-alkenes are no longer visible and are replaced by a band containing a homologous series of 1,2-dibromoalkanes. Although the reaction chemistry produces an enantiomeric pair of vicinal dibromides for each 1-alkene reacted, only one product peak is produced due to co-elution of the two enantiomers

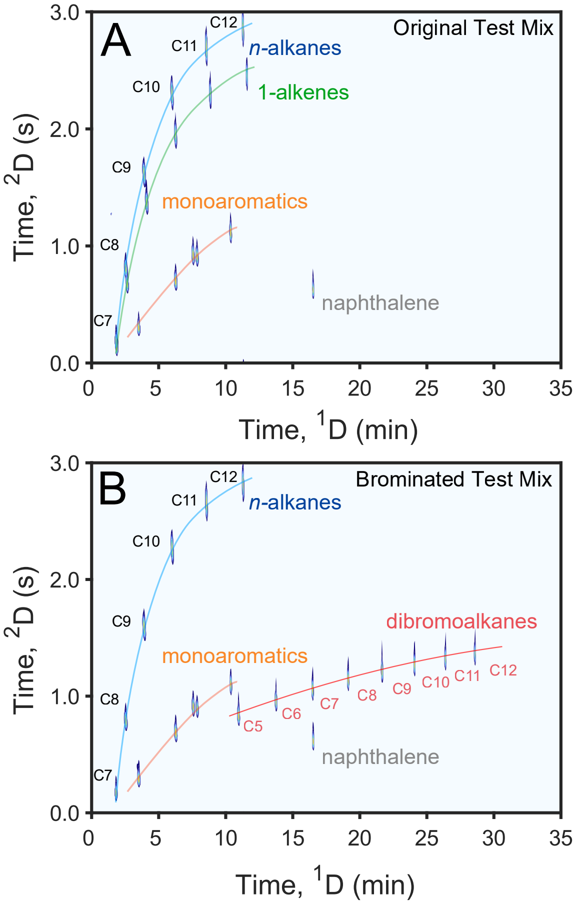


Figure 5.2: (A) GC×GC-TOFMS TIC chromatogram of the original test mixture. Alkenes ≤ C6 are not observed due to coelution withthe solvent. (B) GC×GC-TOFMS TIC chromatogram of the brominated test mixture. The band consisting of the 1-alkenes is no longer visible, but has been replaced by a new band consisting of the homologous series of the 1,2-dibromoalkane products.

on the achiral column set used (Rxi-17 × DB1), simplifying chromatographic interpretation. Notably, while 1-pentene and 1-hexene were not detected in the original test mixture due to elution during the solvent front, their products 1,2-dibromopentane and 1,2-dibromohexane elute later on ^{1}D due to their lower volatility from bromine addition and are well resolved. The entire band of dibromoalkanes elutes in a relatively uncrowded 2D chromatographic region. The n-alkanes and aromatics visually appear the same in the brominated test mixture relative to the original test mixture, suggesting these species were unaffected by the bromination reaction. However, to determine if the 1-alkenes quantitatively reacted and if any side reactions occurred for the other compound groups, the %Reacted for each component was calculated

$$\%\text{Reacted} = \frac{A_{\text{original}} - A_{\text{brominated}}}{A_{\text{original}}} \tag{5.1}$$

where A_{original} and $A_{\text{brominated}}$ are the sum of all the ^{2}D peak signal areas of a given analyte in the original and brominated test mixtures, respectively, measured on extracted ion chromatograms (EICs, n=3, with m/z used given in Table C1). No appreciable reaction was observed for the n-alkanes ($\bar{x}$ = -1.45% ± 2.8%), while the 1-alkenes reacted to virtually 100% conversion ($\bar{x}$ = 99.5% ± 0.4%). Minor bromination of the aromatic species was observed ($\bar{x}$ = 6.7% ± 11.6%), however no side-products had substantial signal in the brominated data to be identified.

Reaction of aromatic species requires the formation of bromine radicals by homolysis [19]. The bond dissociation energy for the homolysis of Br_2 is relatively low with ΔG = 192 kJ mol^{-1}; blue light has an associated energy of approximately 293 kJ mol^{-1} [19]. Since performing the reaction in complete darkness is not practical for routine use, some bromination of aromatic species by radical reactions cannot be entirely avoided. Nonetheless, the test mixture bromination demonstrated excellent selectivity towards 1-alkenes versus typical native fuel compounds, with relatively minor aromatic bromination observed.

5.3.2 Qualitative profiling olefins in gasoline

Next, the Br_2 method was applied to profile olefins in a locally sourced gasoline sample. GC×GC chromatograms of the gasoline before and after bromination are given in Fig. 5.3A and 5.3B, respectively. The chromatograms are plotted as EICs where the signal at several m/z is summed together to facilitate visualization of the alkanes, olefins, and brominated products. The TIC chromatograms are provided in Fig. C4. In the brominated gasoline sample chromatogram, a band of peaks corresponding to the dibromoalkane products is readily identified.

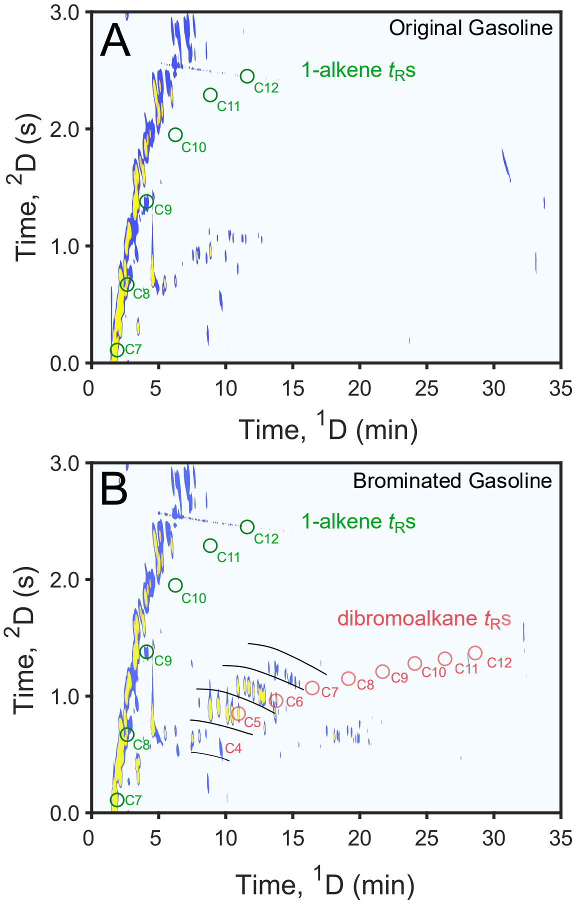


Figure 5.3: (A) GC×GC-TOFMS chromatogram (EIC at m/z 41, 43, 55, and 57) of the original gasoline sample. The 2D retention locations of 1-alkenes, obtained from the standard mixture, are indicated. (B) GC×GC-TOFMS chromatogram (EIC at m/z 41, 43, 55, and 57) of the brominated gasoline sample. The 2D retention locations of the 1,2-dibromoalkanes, obtained from the brominated test mixture, are indicated.

Within this band, several slanting sub-bands are observed, corresponding to an isomeric series of dibromoalkanes at a given carbon number. Occurrence of such structured sub-band patterns is an asset for GC×GC analyses, and has been described as the "roof-tile" effect [49]. Using the retention information provided by the test mixture comparison (Fig. 5.3A,B), lines are drawn on 5.3B to delineate each isomeric series by its carbon number. By qualitative assessment, 5.3B suggests that the olefin profile includes numerous C5 and C6 mono-olefins, with some contribution of C4 and C7 olefins to the overall olefinic profile. Mono-olefins ≥ C8 do not appear present in this gasoline sample. However, to facilitate a more comprehensive characterization of the olefinic profile, tile-based FRA was performed, as is discussed next.

5.3.3 Profiling olefins in gasoline by tile-based FRA

A thorough tile-based FRA comparison of the two sample classes comprised of the GC×GC-TOFMS data set for the original and brominated gasoline samples was performed. Briefly, the software integrates the signal within rectangular "tiles" of the chromatograms which are user-selected to be larger than the 2D peak size and the observed run-to-run shifting thus nullifying the deleterious effects of shifting without requiring explicit alignment. Then, tiles at every m/z are compared between chromatograms using the one-way analysis of variance (ANOVA) F-ratio, the ratio of the between-class variance over the sum of the within-class variances [44]. The initial output is a "hit list", a table which contains the chromatographic locations (^1t_R, ^2t_R) of all analyte features that were discovered, rank-ordered by their F-ratio. The analyst can then mine the hit list to catalogue and identify class-distinguishing analytes, which are most likely to be higher up on the hit list.

Tile-based FRA of the GC×GC-TOFMS data set for the original and brominated gasoline samples generated 314 total hits, of which 56 were determined to be statistically significant via combinatorial null distribution analysis (CNDA). Detailed information describing CNDA is provided in prior reports [44–46]. Briefly, CNDA is a permutation-based test which enables the analyst to empirically determine an F-ratio threshold (F_{crit}) below which hits are likely to not be significant [44–46]. First, class labels are repeatedly permuted (Fig. D6A) and tile-based FRA is performed on each of these new "null classes". The resulting null F-distributions (Fig. D6B) are converted to null probabilities (Fig. D6C), defined equivalently to the false discovery rate (FDR), the expected ratio of false positives to true positives. An F_{crit} of 99% was used herein at a null probability of 5% indicating that with 99% confidence an FDR of 5% will be achieved in the true class comparison (Fig. D6D).

The F-ratio distribution of the true class comparison for all the hits is shown in Figure 5.4A, with the vertical line indicating the significance threshold based upon performing CNDA. As a result of using bromination chemistry, the discoverable olefins were pinpointed in the original gasoline sample class ($S_{\text{original}} > 0$, $S_{\text{brominated}} \approx 0$). The same, if not more, of the olefins were found as dibromoalkanes with appreciable signal only in the brominated gasoline sample class ($S_{\text{brominated}} > 0$, $S_{\text{original}} \approx 0$). To highlight the difference in the relative "discoverability" of these two categories of analyte features, a hit list splitting routine was employed, whereby each analyte feature was assigned to one of the two forementioned cases.

Figure 5.4B shows hits with appreciable signal in the original gasoline sample that disappeared upon bromination. The first hit discovered in this class is hit 18. These hits should correspond to olefins which were removed from the gasoline by bromination. In contrast, Fig, 5.4C shows hits with appreciable signal in the brominated gasoline sample and should correspond to the dibromoalkane products of the olefins. The top 17 hits were all discovered due to signal in the brominated gasoline sample, supporting the notion that they are more "discoverable" than those found in the original gasoline sample. This is largely due to the appearance of the brominated compounds in a relatively unoccupied region of the chromatographic 2D space. In contrast, the unreacted olefins in the original gasoline sample are poorly resolved from branched and cyclic alkanes.

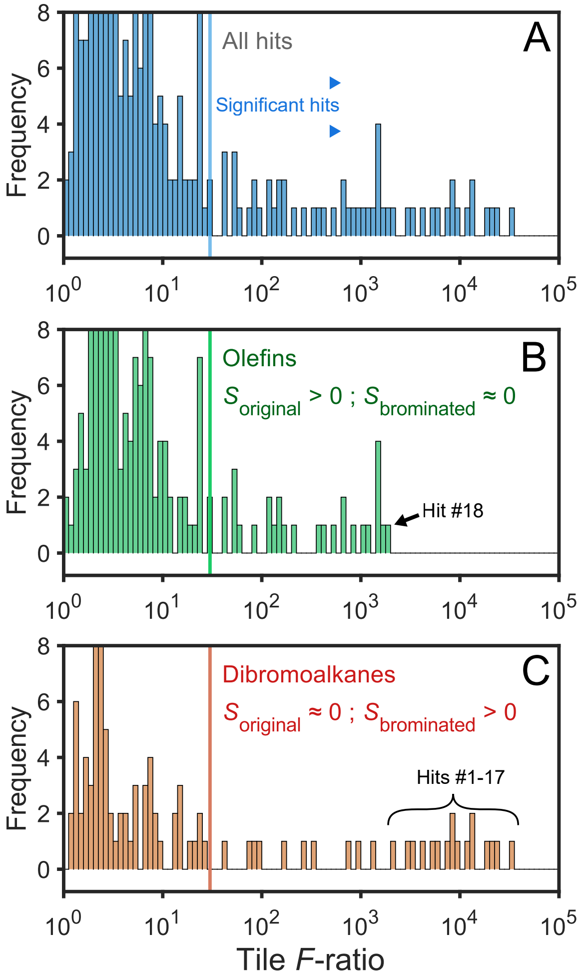


Figure 5.4: (A) F-ratio distribution for all hits in the tile-based FRA output. The vertical line indicates the significance threshold from CNDA. All hits to the right of the line are considered statistically significant. (B) F-ratio distribution for all hits discovered directly due to the olefins ($S_{\text{original}} > 0$ and $S_{\text{brominated}} \approx 0$), constructed by hit list splitting. (C) F-ratio distribution for all hits discovered due to the brominated olefins ($S_{\text{brominated}} > 0$ and $S_{\text{original}} \approx 0$), constructed by hit list splitting.

Next, the discovered brominated hits are investigated in more detail (Fig. 5.5). A region of the brominated gasoline chromatogram (from 5.3B) containing a majority of the significant hits (44/56) is shown in Fig. 5.5A. Each significant hit is plotted as a circle (i.e., a "hit marker") linearly scaled by the F-ratio magnitude. This 2D elution region envelopes the brominated compounds. Compound identification of the top discovered hits was attempted. Working from the assumption that each sub-band corresponds to an isomeric peak series, each peak in the C5 sub-band could correspond to one or more of a total of six possible brominated products of the acyclic C5 olefins, so identification was demonstrated for the C5 olefins. Figure C7 shows all possible linear and branched C5 alkenes and their predicted products, along with the boiling points (bp) of each compound. Commonly, tentative identifications in non-targeted analysis are made by matching the mass spectrum for a given hit to a mass spectrum library

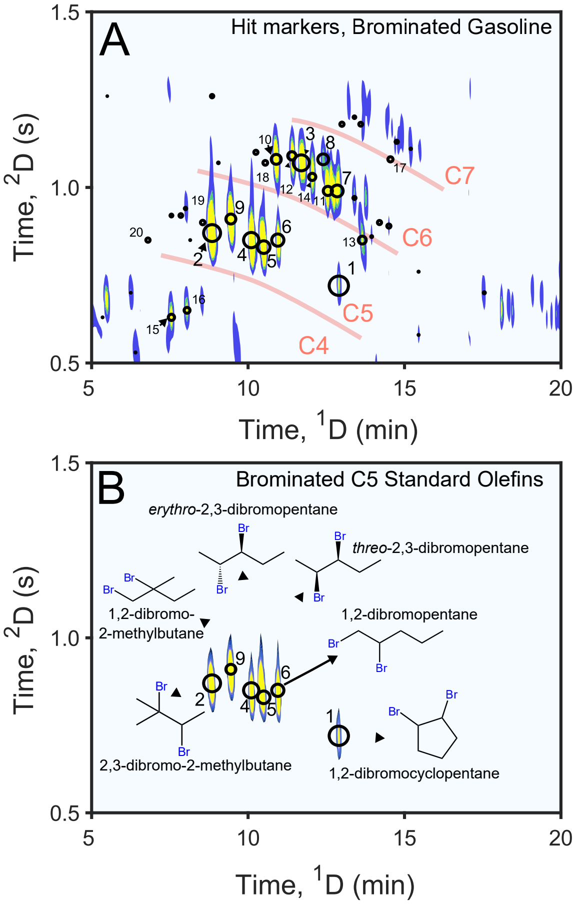


Figure 5.5: (A) Region of the brominated gasoline sample chromatogram (from Fig. 4.3B) containing 44 of the 56 significant hits. Hits are indicated by circles scaled by the magnitude of their F-ratio with the top 20 numbered. (B) Brominated C5 standard olefin mixture used to identify several analytes in the hit list, which can be seen in (A). Data are plotted as an EIC at m/z 41, 43, 55, and 57 in both plots.

(e.g., NIST), aided by retention index (RI) information when available [46]. However, of the six possible C5 dibromoalkane products shown in Fig. C7, only three are present in the NIST 2020 library (Table C2), and fewer are present in older libraries , making identifications based on matching to existing libraries impractical. Therefore, to identify peaks in the C5 sub-band, a standard olefin mixture was sequentially prepared and reacted to generate a brominated standard mix. The standard olefin mixture contained all acyclic C5 olefins except 3-methylbutene (which is gaseous room temperature) and also included cyclopentene (which is also absent from the NIST 2020 library), for six total compounds.

A GC×GC chromatogram of the brominated C5 standard olefin mixture is given in Fig. 5.5B. Using the retention and mass spectrum information provided by the standards, the identities of five hits were obtained as indicated in Fig. 5.4B. The hit markers and numbers from the closest matching hits are overlaid to demonstrate the excellent agreement in 2D retention times. To further confirm analyte identities, class comparison-enabled mass spectrum purification (CCE- MSP), a recently described mass spectrum decomposition technique [50], was used to obtain the mass spectra of the top hits. These mass spectra were matched against those of the brominated standards. Match values (MVs) for all standards were $\geq$ 995. Coupled with 2D retention time matching, the mass spectrum similarity confirms of the identities of each brominated C5 olefin (Table C3). Particularly interesting in 4B is the baseline separation (2D resolution, $R_{s,2D} > 1.5$) for the diastereomers erythro- and threo-2,3-dibromopentane, which are the products of trans-pentene and cis-pentene, respectively. The GC separation of trans- and cis-pentene is very challenging due to their virtually identical enthalpies of vaporization, with a boiling point of 36.3 for the trans isomer and 36.9 °C for the cis isomer [51,52]. In contrast, the boiling points of erythro- and threo-2,3-dibromopentane are 91 °C and 94 °C (50 mm Hg), respectively, as a result of their conformational differences (Fig. C8), enabling an easier separation [53,54]. The excellent separation of the brominated cis- and trans- light olefins demonstrates another benefit of the present methodology. In summary, a hit list containing the top 10 F-ratio hits is provided in Table 5.1. Because the top 10 hits were discovered due to substantial signal in the brominated class, the name of the discovered brominated compound is provided along with the name of the alkene that produced the discovered compound based on the known reaction chemistry. Six of the top 10 hits were compounds included in the C5 standard mix and were definitively identified as described. To support the identifications, Table C3 provides a MV for each dis-

Hit	Discovered Compound	Corresponding Alkene
1	1,2-dibromocyclopentane	cyclopentene
2	2,3-dibromo-2-methylbutane	2-methyl-2-butene
3	dibromohexane isomer	hexene isomer
4	*erythro*-2,3-dibromopentane	*trans*-2-pentene
5	*threo*-2,3-dibromopentane	*cis*-2-pentene
6	1,2-dibromopentane	1-pentene
7	dibromohexane isomer	hexene isomer
8	dibromohexane isomer	hexene isomer
9	1,2-dibromo-2-methylbutane	2-methyl-1-butene
10	dibromohexane isomer	hexene isomer

Table 5.1: Hit list of the top ten discovered analytes via F-ratio analysis.

covered compound along with retention time information for the standards and hits. This method of identifying the C5 olefins on a species-level could also be performed to identify the C6 olefins on a species level. However, confirmation that the isomeric peak series corresponds to a band of olefins at a given number enables assignment of hits in each band to an olefin at a given carbon number. The remainder of the hits in Table 5.1 eluted in the C6 isomeric series (Fig, 4.5B), and were therefore tentatively identified as a dibromohexane isomers, suggesting a series of hexene isomers in the original sample.

5.4 Conclusion

Comparative analysis of GC×GC-TOFMS data is a rapidly growing area of interest. The primary novelty of this work is use of selective reaction chemistry to target a specific compound class combined with a supervised, non-targeted workflow that provides a comprehensive examination of the targeted compound class. The non-targeted aspect of the workflow provides comprehensive discovery, which is juxtaposed with selective reaction chemistry so only the selected compound class is thoroughly analyzed. In doing so, the analysis workflow was streamlined to focus only on the 2D regions of the GC×GC-TOFMS chromatograms where products appeared, or reactants disappeared. The dibromoalkane products generated in this study eluted in a region of chromatographic 2D space less crowded than that occupied by the native olefins, enhancing discoverability relative to if only the disappearance of the olefins were used. Further, the product peaks had no signal in the original class, generating extremely high F-ratios, even for relatively low signal peaks. Implementation areas of future interest include automating the reaction chemistry, such as by flow-injection analysis, and multivariate optimization of reaction conditions. In this report, Br_2 was used in minimal amounts ($\leq 50\,\mu L$); however, the potential safety issue associated with handling Br_2 could be greatly mitigated by generating Br_2 *in situ* with a solid brominating agent, such as pyridinium bromide tribromide. Although the focus of this report was the characterization of olefins by bromination in gasoline, numerous reaction chemistries could be exploited to target a given compound class of interest in a variety of sample matrices. The development of analogous methods to target different compound classes is an on-going area of research.

5.5 References

(1) G. De La Puente; U. Sedran, Formation of gum precursors in FCC napthas. *Energy Fuels* **2004**, *18*, 460–464.

(2) F. Pradelle; S.L. Braga; A.R.F.A. Martins, F. Turkovics; R.N.C. Pradelle, Gum formation in gasoline and its blends: A Review, *Energy Fuels* **2015**, *29*, 7753–7770.

(3) Pereira, R. C. C.; Pasa, V. M. D. Effect of Mono-olefins and diolefins on the stability of automotive gasoline. *Fuel* **2006**, *85*, 1860–1865.

(4) Epping, R.; Kerkering, S.; Andersson, J. T. Influence of different compound classes on the formation of sediments in fossil fuels during aging. *Energy Fuels* **2014**, *28*, 5649–5656.

(5) Nagpal, J. M.; Joshi, G. C.; Singh, J.; Rastogi, S. N. Gum Forming Olefinic Precursors in Motor Gasoline, a Model Compound Study. *Fuel Sci. Technol. Int.* **1994**, *12*, 873–894.

(6) Pullen, J.; Saeed, K. An Overview of biodiesel oxidation stability. *Renew. Sustain. Energy Rev.* **2012**, *16*, 5924–5950.

(7) ASTM D1319, Standard Test Method for Hydrocarbon Types in Liquid Petroleum Products by Fluorescent Indicator Adsorption; *ASTM International: West Conshohocken*, PA, **2015**. www.astm.org

(8) ASTM D6296, Standard Test Method for Total Olefins in Spark-Ignition Engine Fuels by Multidimensional Gas Chromatography; *ASTM International: West Conshohocken*, PA, **2015**. www.astm.org

(9) ASTM D655, Standard Test Method for Determination of Olefin Content of Gasolines by Supercritical-Fluid Chromatography; *ASTM International: West Conshohocken*, PA, **2015**. www.astm.org

(10) Johnston, R. W.B.; Appleby, W. G.; Baker, O. Determination of Olefins in Gasoline. *Anal. Chem.* **1947**, *20*, 805–812.

(11) Squicciarini, M. P. Paraffin, olefin, naphthene, and aromatic determination of gasoline and JP-4 jet fuel with supercritical fluid chromatography. *J. Chromatogr. Sci.* **1996**, *34*, 7–12.

(12) Mondal, S.; Chattopadhayay, K.; Srivastav, B.; Garg, K.; Kumar, R.; Chopra, A.; Christopher, J.; Kapur, G. S. Dynamic approach for the estimation of olefins in cracked fuel range products of variable nature and composition by 1H NMR spectroscopy. *Energy Fuels* **2019**, *33*, 1114–1122.

(13) Walsh, P.; Garbalena, M.; Schug, K. A. Rapid analysis and time interval deconvolution for comprehensive fuel compound group classification and speciation using gas chromatography-vacuum ultraviolet spectroscopy. *Anal. Chem.* **2016**, *88*, 11130–11138.

(14) Nolvachai, Y.; Salzmann, S.; Zavahir, J. S.; Doetzer, R.; Steiner, S.; Kulsing, C.; Marriott, P. J. Structure elucidation using gas chromatography-infrared spectroscopy/mass spectrometry supported by quantum chemical IR spectrum simulations. *Anal. Chem.* **2021**, *93*, 15508–15516.

(15) Souza, I. D.; Nan, H.; Queiroz, M. E. C.; Anderson, J. L. Tunable Silver-containing stationary phases for multidimensional Gas chromatography. *Anal. Chem.* **2019**, *91*, 4969–4974.

(16) Eor, P.; Anderson, J. L. Using a Chromatographic pseudophase model to elucidate the mechanism of olefin separation by silver(I) ions in ionic liquids. *Anal. Chem.* **2021**, *93*, 13284–13292.

(17) Johnson, H. L.; Clark, R. A. Procedure for determination of the bromine number of olefinic hydrocarbons. *Anal. Chem.* **1947**, *19*, 869–872.

(18) Morrison, R. T.; Boyd, R. N. Organic Chemistry, 6th ed.; Prentice Hall, Inc.: Upper Saddle River, NJ, **1992**

(19) Clayden, J.; Greeves, N.; Warren, S. Organic Chemistry, 2nd ed.; Oxford University Press: New York, NY, **2012**

(20) Allen, A.A. On the Relative Proportions of Olefins in Shale and Petroleum Products. *Analyst* **1881**, *6*, 177b – 180.

(21) ASTM D1159, Determination of Bromine Number by Automated Titration; *ASTM International: West Conshohocken*, PA, **2015**. www.astm.org

(22) Hardas, N. R.; Adam, R.; Uden, P. C. Alkene determination by bromination and gas chromatography with element-selective atomic plasma spectroscopic detection. *J. Chromatogr. A* **1999**, *844*, 249–258.

(23) Trigg, D.; Simmonds, P.; Nickless, G. Gas chromatographic determination of volatile alkenes by on-column bromination and electron-capture detection. *J. Chromatogr. A* **1995**, *690*, 197–206.

(24) Davis, J. M.; Giddings, J. C. Statistical theory of component overlap in multicomponent chromatograms. *Anal. Chem.* **1983**, *55*, 418–424.

(25) Blumberg, L. M.; David, F.; Klee, M. S.; Sandra, P. Comparison of one-dimensional and comprehensive two-dimensional separations by gas chromatography. *J. Chromatogr. A* **2008**, *1188*, 2–16.

(26) Liu, Z.; Phillips, J. B. Comprehensive two-dimensional gas chromatography using an on-Column thermal modulator interface. *J. Chromatogr. Sci.* **1991**, *29*, 227–231.

(27) Hyötyläinen, T.; Kallio, M.; Hartonen, K.; Jussila, M.; Palonen, S.; Riekkola, M. L. Modulator design for comprehensive two-dimensional gas chromatography: Quantitative analysis of polyaromatic hydrocarbons and polychlorinated Biphenyls. *Anal. Chem.* **2002**, *74*, 4441–4446.

(28) Amaral, M. S. S.; Nolvachai, Y.; Marriott, P. J. Comprehensive two-dimensional gas chromatography advances in technology and applications: Biennial update. *Anal. Chem.* **2020**, *92*, 85–104.

(29) Zanella, D.; Focant, J.; Franchina, F. A. 30th Anniversary of comprehensive two-dimensional gas chromatography: Latest advances. *Anal. Sci. Adv.* **2021**, *2*, 213–224.

(30) Klee, M. S.; Cochran, J.; Merrick, M.; Blumberg, L. M. Evaluation of conditions of comprehensive two-dimensional gas chromatography that yield a near-theoretical maximum in peak capacity gain. *J. Chromatogr. A* **2015**, *1383*, 151–159.

(31) Nizio, K. D.; Cochran, J. W.; Forbes, S. L. Achieving a near-theoretical maximum in peak capacity gain for the forensic analysis of ignitable liquids using GC×GC-TOFMS. *Separations* **2016**, *3*, 1–17.

(32) Lee, A. L.; Bartle, K. D.; Lewis, A. C. A model of peak amplitude enhancement in orthogonal two-dimensional gas chromatography. *Anal. Chem.* **2001**, *73*, 1330–1335.

(33) Mostafa, A.; Górecki, T. Sensitivity of comprehensive two-dimensional gas chromatography (GCXGC) versus one-dimensional gas chromatography (1D GC). *LC-GC Eur.* **2013**, *26*, 672–679.

(34) Nelson, R. K.; Kile, B. M.; Plata, D. L.; Sylva, S. P.; Xu, L.; Reddy, C. M.; Gaines, R. B.; Frysinger, G. S.; Reichenbach, S. E. Tracking the weathering of an oil spill with comprehensive two-dimensional gas chromatography. *Environ. Forensics* **2006**, *7*, 33–44.

(35) Von Mühlen, C.; De Oliveira, E. C.; Zini, C. A.; Caramão, E. B.; Marriott, P. J. Characterization of nitrogen-containing compounds in heavy gas oil petroleum fractions using comprehensive two-dimensional gas chromatography coupled to time-of-flight mass spectrometry. *Energy Fuels* **2010**, *24*, 3572–3580.

(36) Pollo, B. J.; Alexandrino, G. L.; Augusto, F.; Hantao, L. W. The impact of comprehensive two-dimensional gas chromatography on oil gas analysis: Recent advances and applications in petroleum industry. *TrAC - Trends Anal. Chem.* **2018**, *105*, 202–217.

(37) Giri, A.; Coutriade, M.; Racaud, A.; Stefanuto, P. H.; Okuda, K.; Dane, J.; Cody, R. B.; Focant, J. F. Compositional elucidation of heavy petroleum base oil by GC × GC-EI/PI/CI/FI-TOFMS. *J Mass Spectrom.* **2019**, *54*, 148–157.

(38) Vozka, P.; Modereger, B. A.; Park, A. C.; Zhang, W. T. J.; Trice, R. W.; Kenttämaa, H. I.; Kilaz, G. Jet fuel density via GC×GC-FID. *Fuel* **2019**, *235*, 1052–1060.

(39) Boswell, H. A.; Edwards, M.; Górecki, T. Comparison of thermal and flow-based modulation in comprehensive two-dimensional gas chromatography—time-of-flight mass spectrometry (GC × GC-TOFMS) for the analysis of base oils. *Separations* **2020**, *7*, 1–17.

(40) van der Westhuizen, R.; Potgieter, H.; Prinsloo, N.; de Villiers, A.; Sandra, P. Fractionation by liquid chromatography combined with comprehensive two-dimensional gas chromatography-mass spectrometry for analysis of cyclics in oligomerisation products of fischer-tropsch derived light alkenes. *J. Chromatogr. A* **2011**, *1218*, 3173–3179

(41) Zou, Y.; Stefanuto, P. H.; Maimone, M.; Janssen, M.; Focant, J. F. Unraveling the complex olefin isomer mixture using two-dimensional gas chromatography-photoionization-time of flight mass spectrometry. *J. Chromatogr. A* **2021**, *1645*, 462103.

(42) Zou, Y.; Stefanuto, P.H.; Maimone, M.; Janssen, M.; Focant, J.F. Insights into dodecenes produced from olefin oligomerization based on two-dimensional gas chromatography-photoionization-time of flight mass spectrometry and multivariate statistics. *ACS Omega* **2021**, *46*, 30971-30982.

(43) Wang, F. C. Y. Comprehensive two-dimensional gas chromatography hyphenated with a vacuum ultraviolet spectrometer to analyze diesel-A three-dimensional separation (GC×GC×VUV) approach. *Energy Fuels* **2020**, *34*, 8012–8017.

(44) Parsons, B. A.; Marney, L. C.; Siegler, W. C.; Hoggard, J. C.; Wright, B. W.; Synovec, R. E. Tile-based fisher ratio analysis of comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry (GC × GC-TOFMS) data using a null distribution approach. *Anal. Chem.* **2015**, *87*, 3812–3819.

(45) Parsons, B. A.; Pinkerton, D. K.; Wright, B. W.; Synovec, R. E. Chemical characterization of the acid alteration of diesel fuel: Non-targeted analysis by two-dimensional gas chromatography coupled with time-of-flight mass spectrometry with tile-based fisher ratio and combinatorial threshold determination. *J. Chromatogr. A* **2016**, *1440*, 179–190.

(46) Watson, N. E.; Parsons, B. A.; Synovec, R. E. Performance evaluation of tile-based fisher ratio analysis using a benchmark yeast metabolome dataset. *J. Chromatogr. A* **2016**, *1459*, 101–111.

(47) Sudol, P. E.; Ochoa, G. S.; Synovec, R. E. Investigation of the limit of discovery Using tile-based fisher ratio analysis with comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry. *J. Chromatogr. A* **2021**, 1644.

(48) Ochoa, G. S.; Prebihalo, S. E.; Reaser, B. C.; Marney, L. C.; Synovec, R. E. Statistical inference of mass channel purity from fisher ratio analysis using comprehensive two-dimensional gas chromatography with time of flight mass spectrometry data. *J. Chromatogr. A* **2020**, *1627*, 461401.

(49) Beens, J.; Blomberg, J.; Schoenmakers, P. J. Proper tuning of comprehensive two-dimensional gas chromatography (GC×GC) to optimize the separation of complex oil fractions. *J. High Resolut. Chromatogr.* **2000**, *23*, 182–188.

(50) Ochoa, G. S.; Sudol, P. E.; Trinklein, T. J.; Synovec, R. E. Class comparison enabled mass spectrum purification for comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry. *Talanta* **2022**, *236*, 122844.

(51) CRC Handbook of Chemistry and Physics, 88th ed.; Lide, D. R., Ed.; CRC Press: Boca Raton, FL, **2007**.

(52) Scott, D. W.; Waddington, G. Vapor pressure of *cis*-2-pentene, *trans*-2-pentene and 3-methyl-1-butene. *J. Am. Chem. Soc.* **1950**, *72*, 4310–4311.

(53) Lucas, H. J.; Schlatter, M. J.; Jones, R. C. The isomeric 2,3-epoxypentanes and 2-pentenes. The Extent to which mixtures of diastereomers are formed in reactions of some pentane compounds. *J. Am. Chem. Soc.* **1940**, *63*, 22–28.

(54) Bayer, F. L.; Goodfey, P. G.; Gordon, M. Rapid gas chromatographic separation of diastereomeric dihalo-butanes, pentanes, and hexanes. *J. Chromatogr. Sci.* **1973**, *11*, 443–446.

Chapter 6

SIMULATING COMPREHENSIVE TWO-DIMENSIONAL GAS CHROMATOGRAPHY MASS SPECTROMETRY DATA WITH REALISTIC RUN-TO-RUN SHIFTING TO EVALUATE THE ROBUSTNESS OF TILE-BASED FISHER RATIO ANALYSIS

This chapter was reproduced from T.J. Trinklein and R.E. Synovec, *J. Chromatogr. A* **2022**, *1677*, 463321.

6.1 Introduction

Comprehensive two-dimensional gas chromatography coupled to time-of-flight mass spectrometry (GC×GC-TOFMS) has evolved into an instrumental method of choice for the analysis of sufficiently volatile complex mixtures [1-4]. Compared with optimized one-dimensional (1D) GC, GC×GC can provide roughly a ten-fold enhancement in peak capacity [5,6]. This makes GC×GC-TOFMS an indispensable tool for untargeted, discovery-based analyses, where the analytes of interest are not necessarily known *a priori*. Gleaning useful information from the large data sets generated by GC×GC-TOFMS, still an arduous task, is becoming easier, more rapid and reliable due to advances in commercial software (e.g., ChromaTOF, ChromSpace, and GC Image, among others) in concert with the ongoing academic development of data processing tools [7–11]. Recently, untargeted GC×GC-TOFMS workflows have been developed in various application areas, including metabolomics [12,13], food science [14,15], forensic science [16], environmental samples [17], and petroleomics [18,19].

When sample classes are known based upon the experimental design (e.g., "infected" vs. "control"), an untargeted analysis is more precisely described as a *supervised* untargeted method. One such method, Fisher ratio analysis (FRA) uses the one-way analysis of variance (ANOVA) F-ratio statistic, which is the ratio of the between-class variance to the pooled within-class variance, calculated for each chromatographic peak "feature." Based upon the F-ratio calculation, the FRA output is a "hit list" of compounds, rank-ordered by their F-ratio. Features with high F-ratios can be further scrutinized as potential discriminatory compounds, greatly reducing the amount of data necessary to analyze. A feature is a variable flagged by FRA as being potentially class-discriminatory, and features can correspond to pixels, tiles, or tabulated peak areas, as will be discussed next. Note that features are not the initially analytes themselves, but variables which were flagged for further investigation.

FRA was first applied to GC×GC-TOFMS on a pixel-level, however, even minor run-to-run misalignment generated numerous false positives that were often near the top of the hit list [20]. Such false positives were due to retention time shifting rather than meaningful chemical differences between classes [21]. To some extent pixel-based FRA can be applied following alignment of the raw data. However, the resulting 2D separation alignment is computationally expensive and challenging due to the low data density resulting from undersampling the first dimension (^{1}D) separation, and may not accurately preserve peak areas and mass spectra [21,22]. Another approach to perform FRA is based upon F-ratio calculations using peak table data [12,23], which was formerly commercialized as the Statistical Compare feature of ChromaTOF [24]. Peak table-based FRA has been widely used in a variety of applications [23,25–27], often combined with other data analysis tools. However, this method requires sample identification prior to peak table construction and peak matching (via mass spectrum match criteria), making this approach time-consuming and necessitating extensive optimization [12].

To mitigate the shortcomings associated with pixel- and peak table-based FRA, a novel form of FRA termed tile-based FRA was introduced [21,28–30]. Briefly, the tiling step of the software segments each GC×GC-TOFMS chromatogram into four spatially offset grids of segments, termed tiles, and the signal within each tile is summed as a function of mass channel (*m/z*) and compared across sample classes via the *F*-ratio calculation. The size of the tile is a user-selected parameter that should encompass both the width of a typical 2D chromatographic peak as well as the observed run-to-run shifting in both separation dimensions, thus mitigating the occurrences of false positives due to run-to-run shifting. Because tile-based FRA does not require explicit alignment or component identification for peak-table matching and the tiling algorithm is computationally expedient, a significant advantage over the aforementioned approaches is achieved. Recently, tile-based FRA software has been released in the commercial software packages ChromaTOF Tile [31] and ChromCompare+ [32]. A detailed explanation of the algorithmic steps can be found in previous work [21,28].

As usage of tile-based FRA increases, a more thorough investigation of the software is warranted in order for the user community to better understand its strengths and limitations, thus broadening the scope of its application. Successful implementation of FRA relies upon proper tile size selection, which we hypothesize to be a function of separation parameters which are characteristic of different sample types and experimental design, namely, the variable amounts of run-to-run shifting, the chromatographic saturation, and the amount of within-class sample variance. A primary objective of the present report is to provide a detailed and thorough investigation of the interplay between these variables as they relate to the success of tile-based FRA. In order to systematically model the effects of changing these variables, we focus on the evaluation of simulated GC×GC-TOFMS data sets. Note that a comparison to ChromaTOF Tile [31] and ChromCompare+ [32] is beyond the scope of this study because simulated GC×GC-TOFMS data sets are used herein. Therefore, the present study relies upon using the in-house version of tile-based FRA which performs most similar to ChromaTOF Tile [21,28]. There are a few reported examples of simulated GC×GC data in the literature [7,33]; however, simulation complexity varies considerably depending on the needs of the particular study. Of importance to the present study is the simulation of realistic retention-time shifting. Further, evaluation of a realistic amount of detector noise in simulations is lacking in the literature. Therefore, a secondary aim was to develop a simulation workflow that provides confidence that the results are generalizable to real GC×GC-TOFMS data sets.

To quantify the success of tile-based FRA for each set of simulated chromatographic conditions and for different tile sizes, receiver operating characteristic (ROC) curves are employed [30]. The area under the ROC curve, or AUROC, defines the probability that a randomly selected hit would be correctly ranked as either a true positive or false positive. An AUROC of 1 is an ideal test, indicating all true positives were located at the top of the hit list, and the AUROC decreases as the ranking of true positives decreases and when fewer than the total number of simulated true positives (known *a priori*) are discovered. Ultimately, we envisage the results of this study will provide the user community with insight into how to best select their experimental design and tile-based FRA parameters (namely, tile size selection) in order to glean the most information from the untargeted analysis of the data set at hand.

6.1.1 Computational procedures

All simulations are based on a 6 sample versus 6 sample experimental design, i.e., two sample classes, each composed of six sample chromatograms. For this purpose, the term simulated class-based experiment (SCBE) is used to refer to a given simulated set of 12 sample chromatograms. We also adopt the terminology “single-component peak” (SCP) [34-36],

Parameter	**Condition(s) Studied**
^{1}D Separation time, $^{1}t_{\mathrm{sep}}$	600 s
Modulation period, P_{M}	2 s
^{1}D peak width, $^{1}w_{\mathrm{b}}$	6 s
^{2}D peak width, $^{2}w_{\mathrm{b}}$	200 ms
Ideal 2D peak capacity, $n_{\mathrm{c,3D}}$	1000 in 600 s
Acquisition frequency, f	100 Hz
Effective 2D saturation, $\alpha_{\mathrm{e,2D}}$	0.1, 0.5, 2.5
Number of components, m	100, 500, 2500
Number SCPs upregulated per SCBE	20
Detector noise per m/z	4000 (1 s.d.)
Retention time precision, $\langle\delta_{\mathrm{r}}\rangle$	0.0, 0.5, 1.0, 2.0
Within-class variance, RSD_{wc}	0.5-2% or 5-50%
Tile sizes (pixels) / RTAs	20, 80, 320, 500, 720, 2000 / 0.25, 1.00, 2.25, 4.00, 6.25, 9.00, 20.00

Table 6.1: GC×GC-TOFMS simulation parameters and conditions studied.

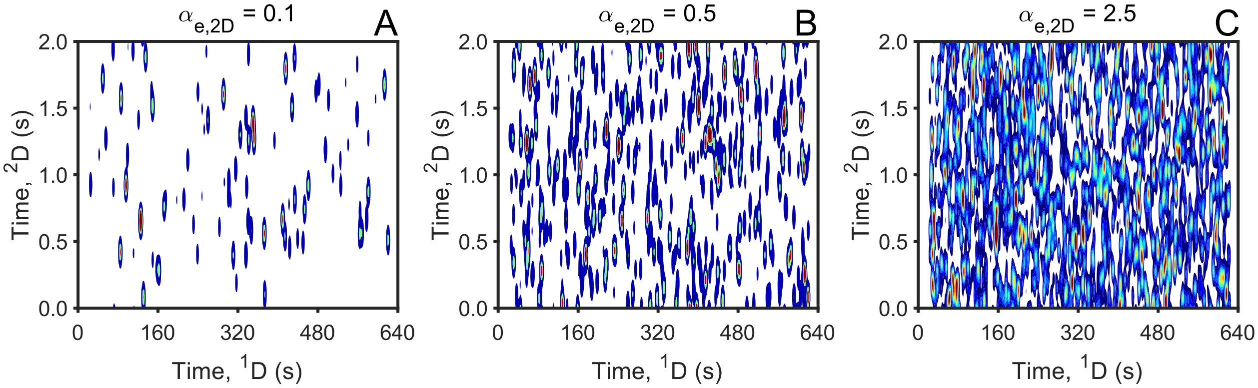


Figure 6.1: Example simulated chromatograms of the three saturation levels studied. (A) $\alpha_{\mathrm{e,2D}}$ = 0.1, (B) $\alpha_{\mathrm{e,2D}}$ = 0.5, (C) $\alpha_{\mathrm{e,2D}}$ = 2.5. Each chromatogram is plotted in a time window of 640 s, although $^{1}t_{\mathrm{sep}}$ is limited to a 600 s interval (20 s to 620 s) for unshifted chromatograms.

where an SCP is the chromatographic signal produced by a single constituent, while more broadly a "peak" is a chromatographic signal with a maximum that can be either one resolved SCP or multiple unresolved SCPs that overlap to produce one maximum [35]. In each SCBE, twenty SCPs are nominally upregulated by a factor of two. These twenty SCPs define the analytes, whereas the remaining compounds which do not change between classes are considered interference SCPs. Note that a feature is a variable discovered by FRA which could potentially correspond to a discriminatory compound(s). After finding features, an analyst can identify and quantify the analyte(s) that are contained in the tile features using the original high-fidelity data. Table 6.1 lists the simulation and tile-based FRA parameters studied. The methodology behind the selection of these key parameters is discussed in the following sections, while the computational equipment and simulation coding method is discussed in the Supplementary Material.

6.1.2 Peak capacity

The traditional definition of peak capacity for a comprehensive 2D separation is given by [37,38]

$$n_{\mathrm{c,2D}} = {}^{1}n_{\mathrm{c}} \times {}^{2}n_{\mathrm{c}} \tag{6.1}$$

where ${}^{1}n_{\mathrm{c}}$ and ${}^{2}n_{\mathrm{c}}$ are the peak capacities on the first and second dimension, respectively. If the entire separation space is occupied (i.e., the separation is fully orthogonal) [39–41], and peak widths are constant, at unit resolution, $R_{\mathrm{s}} = 1$, Eq. (6.1) can be expressed as

$$n_{\mathrm{c,2D}} \frac{{}^{1}t_{\mathrm{sep}}}{{}^{1}w_{\mathrm{b}}} \times \frac{{}^{2}t_{\mathrm{sep}}}{{}^{2}w_{\mathrm{b}}} \frac{{}^{1}t_{\mathrm{sep}}}{{}^{1}w_{\mathrm{b}}} \times \frac{P_{\mathrm{b}}}{{}^{2}w_{\mathrm{b}}} \tag{6.2}$$

We endeavored to simulate GC×GC-TOFMS chromatograms with a reasonably attainable peak capacity. A peak capacity of ~6000 can be readily achieved in a 1 hr run with GC×GC [5,42]; therefore, a $n_{\mathrm{c,2D}}$ of 1000 was selected for 10 min simulated separations, These 10 min simulated chromatograms can be considered to be a portion of a hypothetical 1 hr separation with a peak capacity of 6000. Ten minute separations, rather than 60 min separations, were selected for computational efficiency. To achieve $n_{\mathrm{c,2D}}$ = 1000 via Eq. (6.2), the ^{1}D peak width, ^{1}w+b, (4σ) was 6 s, sampled with a modulation period, P_{M}, of 2 s with a ^{2}D peak width, ${}^{2}w_{\mathrm{b}}$, of 200 ms (Table 6.1). Note that further corrections can be made to Eq. (2) to account for broadening of the ^{1}D separation due to undersampling [43,44]. However, conditions were selected such that undersampling-induced ^{1}D broadening was relatively minor, thus, peak capacity as defined in Eq. (6.1) was deemed a suitable metric for this work. Finally, mass spectra were obtained at an acquisition frequency, *f*, of 100 Hz (Table 6.1).

6.1.3 Chromatographic saturation

The “crowdedness” or saturation of a 2D separation can be defined by the effective 2D saturation, $\alpha_{\mathrm{e,2D}}$ [45],

$$\alpha_{\mathrm{e,2D}} = \frac{m}{n_{\mathrm{c,2D}}} \tag{6.3}$$

where m is the total number of detectable mixture components (all SCPs comprised of analytes and interferences). Three levels of $\alpha_{\mathrm{e,2D}}$ are examined while holding $n_{\mathrm{c,2D}}$ constant at 1000 by letting m = 100, 500, or 2500, generating $\alpha_{\mathrm{e,2D}}$ of 0.1, 0.5, and 2.5, respectively, to represent samples of varying complexity. Note that the number of analytes is held constant at 20 for all levels of $\alpha_{\mathrm{e,2D}}$. Examples of simulated chromatograms at each $\alpha_{\mathrm{e,2D}}$ are given in Fig. 6.1. Note that retention times for the SCPs were permitted to occur on the separation boundaries, such that some peaks were “split” across modulations. In some instances when significant ^{2}D shifting occurred, SCPs wrapped around completely (${}^{2}t_{\mathrm{R}} > 2$ s). The occurrence of wraparound peaks is similar to the description of “unbounded” separations in SOT literature [45]. The instance of peaks split across modulations is well handled by the tile-based FRA software and is illustrated and further explained in Fig. D1.

Although determination of $\alpha_{\mathrm{e,2D}}$ for different sample types and GC×GC conditions has not been evaluated in the literature, the following examples are suggested. At $\alpha_{\mathrm{e,2D}} = 0.1$ the peak capacity is 10-fold higher than the number of components, which may correspond to extracts of food and fragrance samples [46]. An $\alpha_{\mathrm{e,2D}} = 0.5$ is representative of GC×GC applications involving many metabolomic [29] and environmental samples. Although few (if any) complete separations likely yield $\alpha_{\mathrm{e,2D}} = 2.5$, such a high saturation level may occur locally, that is, within a portion of the 2D separation space, such as in the paraffin- and naphthene-containing regions of fuel separations. Moreover, $\alpha_{\mathrm{e,2D}} = 2.5$ was employed as an extreme condition to provide a rigorous evaluation of the tile-based FRA software.

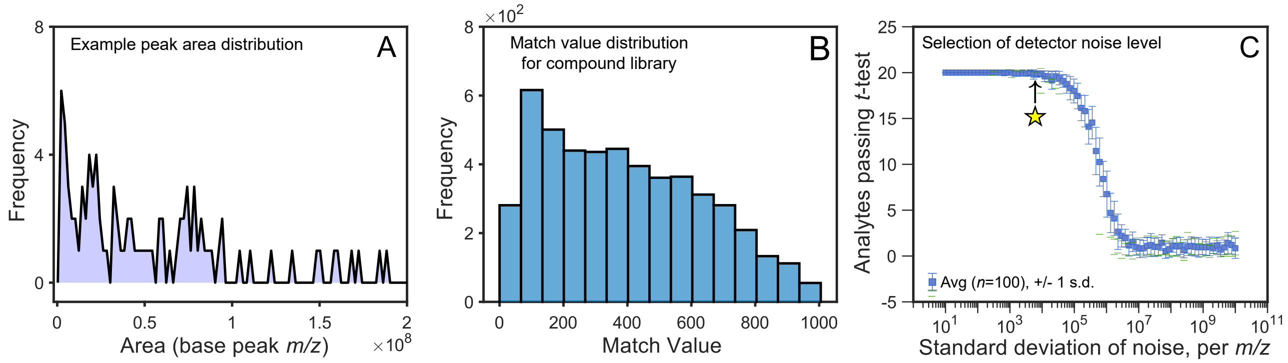


Figure 6.2: (A) Representative area distribution, generated using Eq. (6.4). (B) Match value distribution of 100 compounds in the mass spectrun library. (C) Plot of the number of discoverable analytes (via Student's *t*-test) as a function of the amount of detector noise (per m/z, $s_{\mathrm{Det,m/z}}$), introduced into the simulation. The starred arrow indicates the selected noise level of $s_{\mathrm{Det,m/z}}$ = 4000.

6.1.4 *Concentration distribution and dynamic range*

In real samples the distribution of component concentrations is currently understood to follow a right-skewed probability distribution, with the exponential, Weibull, and log-normal distributions proposed in the literature [36,47,48]. Herein, sample components are assumed to follow an exponential distribution of concentrations (and thus peak areas). For each simulation, an array of m SCP areas are generated by drawing m random variates from an exponential distribution. However, by the nature of drawing SCP areas randomly from an exponential distribution, the effective linear dynamic range (LDR) in each simulated chromatogram can be highly variable between simulations. Therefore, the LDR was prohibited from exceeding 10^3, a typical LDR for a TOFMS detector with 70 eV electron ionization [49,50], by constraining the upper end of the exponential function as

$$f(x;\mu) = \begin{cases} \frac{1}{\mu}\exp\left(-\frac{x}{\mu}\right) & 40^5 \le x \le 40^8 \\ 0 & x < 40^5 \text{ or } x > 40^8 \end{cases} \tag{6.4}$$

where μ is the distribution mean and x is the base m/z peak area. Note that a LDR of 10^3 was selected conservatively as a readily achievable dynamic range. The LDR experimentally achieved is typically constrained by detector saturation, which is dependent on the ionization method (e.g., hard vs. soft) [51] and the digitizing sample rate [52]. Some instrument manufacturers report an LDR as high as 10^5 [53]. Use of an area distribution in this manner ensures that all m components will be detectable, up to some amount of detector noise (*vide infra*). An example of an area distribution generated from Eq. (6.4) is given in Fig. 6.2A. For each SCBE, a new set of m SCP areas was drawn from the distribution in Eq. (6.4).

6.1.5 *Analyte/interference component library*

Electron ionization (EI) mass spectra (70 eV, unit mass resolution, 40 to 240 m/z) of 100 compounds were obtained from the NIST11 library (Table D1). EI-MS spectra were then randomly assigned to each SCP. For simulations containing more SCPs than spectra in the library, the same spectrum was assigned to more than one SCP for the interference components. However, the spectra assigned to the twenty SCP analytes upregulated per chromatogram were not repeated within a SCBE. Fig. 6.2B displays a histogram for all possible pairwise match values (normalized cosine similarity [54]) of the 100 mass spectra. The distribution shows a broad range of spectrum similarity, indicating the compounds in the library will generate realistic sample types for the simulated chromatograms.

6.1.6 *Within-class variance*

The success of an untargeted analysis is adversely impacted by the amount of background variance (overall noise in the data set), which can be expressed as [55],

$$s_{\mathrm{bg}}^2 = s_{\mathrm{NatVar}}^2 + s_{\mathrm{SP}}^2 + s_{\mathrm{NatVar}}^2 + s_{\mathrm{Inj}}^2 + s_{\mathrm{Det}}^2 \quad (6.5)$$

where the subscripts indicate three contributions to s_{bg}^2: natural sample variance, s_{NatVar}^2, such as biological variance, sample preparation s_{SP}^2, injection, s_{inj}^2, and detection noise variance, s_{Det}^2. Although detection variance contributes to the s_{bg}^2, in terms of the simulated chromatograms we consider its contribution to s_{bg}^2 separately from the other three contributions (*vide infra*).

Here, regarding the contributions of natural sample variance, sample preparation variance, and injection variance, the overall background variance s_{bg}^2 is expressed in terms of the within-class relative standard deviation in peak area, RSD_{wc}. Two broad categories of experiments are considered, the more challenging case in which s_{NatVar}^2 and s_{SP}^2 are the principal sources of variance (high RSD_{wc} case), and the less challenging case where s_{Inj}^2 is the principal source of variance (low RSD_{wc} case). In the former case, s_{NatVar}^2, including biological variation (e.g., culture-to-culture variation [20]), and s_{SP}^2, such as solvent-based and sorptive extraction methods [46], contribute to the large amount of s_{bg}^2 observed in environmental and metabolomics-based investigations [55]. Supported by prior observation [29,56], the high variance case was varied from 5-50% RSD_{wc}, assigned randomly for each SCP from a uniform distribution. For the low variance "injection-limited" case, RSD_{wc} was varied from 0.5 to 2% (Table 6.1).

6.2 Detector noise

Gaussian-distributed detector noise was added to each simulated chromatogram independently at every *m/z*. In order to generate simulations that were both realistic and analytically challenging, we endeavored to ensure that all m components were both detectable and discoverable. To do so, the lower concentration components were on the "cusp" of discoverability, where the level of simulated detector noise produced a sufficiently challenging experimental design. Our goal was to create a realistic and challenging scenario to facilitate a thorough evaluation of the tile-based FRA software. To accomplish this, a series of SCBEs containing twenty analytes were generated with increasing levels of detector noise added to each *m/z* spaced logarithmically over the interval of 10^1 to 10^{10}. Here, we more precisely write s_{Det} from Eq. (6.5) as $s_{\mathrm{Det,m/z}}$ to emphasize that this noise level corresponds to the amount added to each *m/z*. Distinct from all other SCBEs discussed later in this report, the twenty SCPs were assigned fixed retention times and were constrained to be well resolved from one another ($R_{\mathrm{S,2D}} > 4$ for all pairwise combinations). At this stage no interfering compounds were included, such that the effect of noise could be studied independent of peak overlap. All other conditions were kept constant: peak areas were drawn from the distribution in Eq. (6.4), mass spectra were assigned randomly from the library (Table D1), and no RSD_{wc} was added.

To determine the fraction of "discoverable" analytes, tile-based FRA was performed, and the number of analytes discovered was reported as the number with a significant concentration difference via Student's *t*-test. The results are shown in Fig. 6.2C. At $s_{\mathrm{Det,m/z}} = 4000$, the number of passing analytes is 19.96 ± 0.22, while further increasing $s_{\mathrm{Det,m/z}}$ rapidly decreases the discoverable analytes. Therefore, $s_{\mathrm{Det,m/z}} = 4000$ (starred, Fig. 2C) was selected as optimal. To support the selection of $s_{\mathrm{Det,m/z}}$, Fig. 6.3A and 6.3B show the unfolded (vectorized) tiles containing the highest concentration and lowest concentration SCP analytes from a representative simulated chromatogram at $s_{\mathrm{Det,m/z}} = 4000$. Figures 6.3C and 6.3D show the same unfolded tiles, plotted as extracted ion chromatograms (EICs) at the *m/z* with the highest *F*-ratio for the SCBE that includes the chromatograms in Fig. 3A and B. All twelve runs are overlaid. Note that the high concentration SCP analyte (Fig. 3A) has numerous *m/z* sufficiently above the detector noise, indicating facile discovery and li-

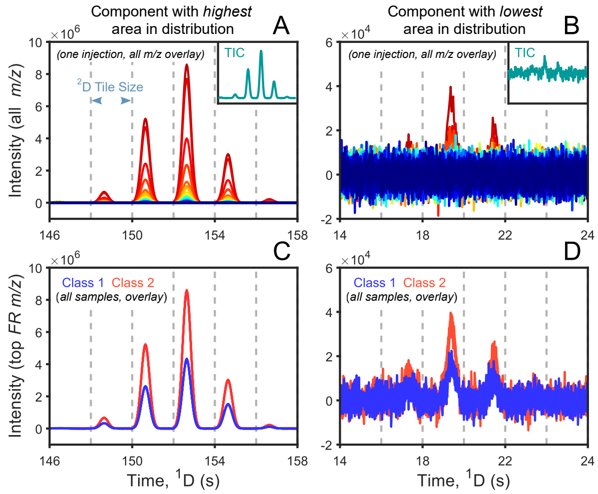


Figure 6.3: (A) Analyte in a SCBE simulated with an area drawn from the upper end of the area distribution in Fig. 6.2A (one replicate, overlay of all m/z). The inset shows the appearance of the analyte in the TIC. (B) Equivalent type of plot as in (A), now for an analyte with the smallest area via the distribution in Fig. 6.2A. (C) Overlay of all runs for the higher concentration analyte, plotted as EICs at the top F-ratio m/z illustrating facile discovery of a high concentration analyte. Although this is an overlay of 12 runs, because $RSD_{\mathrm{wc}} = 0$, the traces closely overlap. (D) Equivalent type of plot as in (C), now for the analyte with the lowest area from Fig. 6.2B, illustrating successful discovery of the lowest concentration analyte despite the small number of discernable m/z. Again, only the top F-ratio m/z is plotted as an overlay for all 12 runs.

kely identification by library matching. In contrast, the low concentration SCP analyte (Fig. 6.3B) only has ~3 m/z that are easily distinguishable from noise, suggesting successful identification is less likely. However, there is a discernable concentration difference at the highest F-ratio m/z in Fig. 6.3D, and the analyte passed the t-test. Therefore, while all SCPs simulated are *detectable* and *discoverable*, they are not necessarily *identifiable* from mass spectrum information [57] but are identified from the *a priori* information provided by the simulation.

6.2.1 Simulation of run-to-run retention time shifting

Run-to-run retention shifting is the last parameter modeled and is the primary variable of interest, with the goal being to simulate realistically shifted data that spans the range that would be encountered experimentally, from no run-to-run shifting to substantial shifting that significantly challenges the tile-based FRA software performance. The actual behavior of run-to-run shifting can be assessed by inspection of retention-time “shift functions”, defined here as a function which traces the shift, as a function of run-time, between a given sample chromatogram and a “target” chromatogram. The target chromatogram can be defined a number of ways such as simply selecting one of the sample runs or by using the mean retention times of a subset of components. For alignment purposes, the target is usually taken to be the most representative chromatogram from the sample set. However, selection of the appropriate target is less critical for the present application, as will be demonstrated.

Highly informative examples of shift functions were generated by Krebs et al. in the context of developing a retention time alignment algorithm for 1D GC-MS data [58]. To accomplish this, the authors used a peak finding script to locate several high S/N peaks (therein referred to as “landmark” peaks) that were present in all chromatograms. Then the retention time of each landmark peak in a given run was compared to that of a target chromatogram. Because misalignments were not simple linear shifts, the authors used cubic spline interpolation to generate shift functions. The shift functions were low frequency waveforms containing a small number (*ca.* 2-3) of maxima and minima.

While the shift functions reported by Krebs et al. were instructive for understanding run-to-run shifting in 1D-GC-MS data, investigating shift functions in experimentally obtained GC×GC-TOFMS data is necessary to extend the principles to the ^{2}D dimension. To this end, Fig. 6.4 shows individual analyte shifts (dots) and shift functions (lines) on ^{1}D (Fig. 6.4A) and ^{2}D (Fig. 6.4B) generated using sixteen analytes from a yeast data set previously analyzed by GC×GC-TOFMS with thermal modulation [20] (Fig. D2-D3). This data set was collected over a period of approximately two months and displayed a marked amount of shifting as a result (Fig. D4). Consistent with the Krebs study, as previously ob-

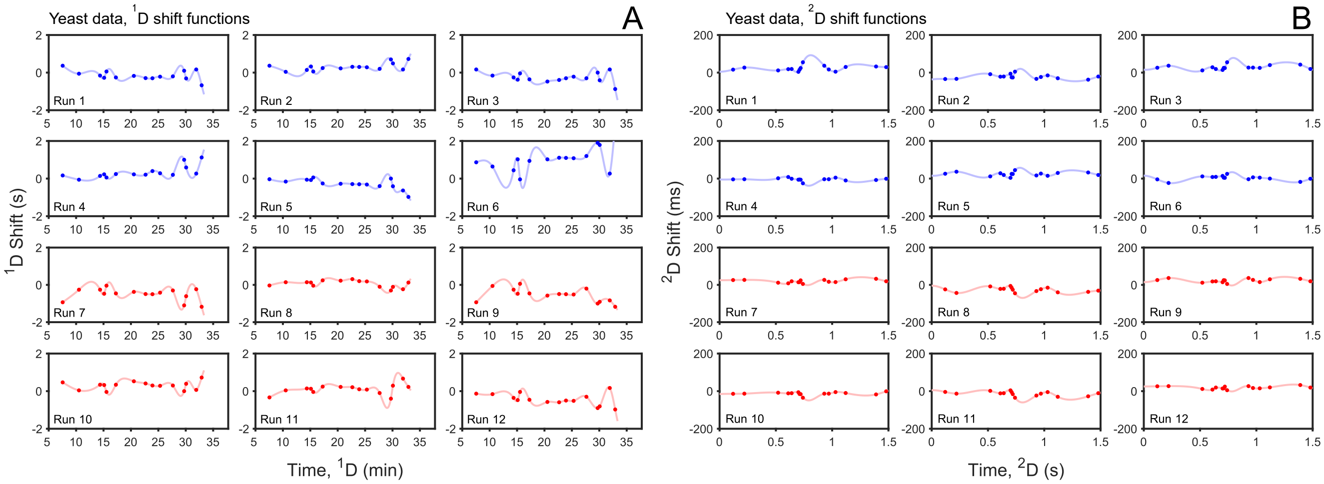


Figure 6.4: (A) Experimental ^{1}D shifts (dots) and shift functions (lines) obtained from a previously analyzed yeast data set [20]. Each point is the difference in retention time for each analyte from a given run to a target set of retention times, which are taken here to be the mean retention times for all twelve runs. Shift functions are obtained by cubic spline interpolation. (B) Equivalent plot for experimental ^{2}D shifts.

served [58] simple polynomials did not fit the data, rather cubic spline interpolation was needed to generate adequate shift functions. Although Fig. 6.4 shows shift functions using mean retention times to create the target, the same conclusions can be drawn observing shift functions with arbitrary target selection, e.g., using Run 1 as the target. Inspection of the 1D GC-MS shift functions in [58] and both sets of shift functions (^{1}D and ^{2}D) from the yeast data set (Fig. 6.4) leads to several important observations. First, the appearance of shift functions between different chromatograms, even within a single experiment, is highly variable. Second, one or a small number of maxima and minima in the shift function occur, wherein peaks shift towards or away from one another between runs. Therefore, peak retention order inversions between runs should occur only at very few locations at the maxima and minima of the shift functions as result of minor, low-frequency changes in temperature and flow in one separation relative to another. Finally, we note that the appearance of the ^{1}D and ^{2}D shift functions are strikingly similar, suggesting that a common methodology can be used to simulate shifting on both dimensions.

With the aforementioned qualities in the run-to-run shifting in mind, the following method was developed to generate realistically shifted GC×GC chromatograms: First, a “base set” of uniformly random retention times is generated (Fig. 6.5A). Here, the base set is the initial set of points which are then shifted to created twelve shifted runs, serving as the target for the purpose of visualizing run-to-run shifting (via shift functions). Next, a ^{1}D shift function is modeled as a low-frequency waveform by applying a low-pass filter to white Gaussian noise (Fig. 6.5B). The low-pass filter was selected such that the number of maxima and minima in the simulated shift functions was approximately the same as the shift functions constructed from real data sets [20,58]. Using low-frequency noise to generate shift functions is logical because the origin of shifting in real data is the result of random sources of low-frequency noise, such as slow changes in temperature and flow rate. The new, shifted ^{1}D retention time of each SCP is obtained by mapping its location onto the shift function, as illustrated in Fig. 4B, where each dot indicates the point of a given SCP on the shift function. For each run, a new shift function is generated, and a new set of shifted ^{1}D retention times are obtained (Fig. 6.5C). Note that the mapped retention times of each SCP (i.e., dots) are omitted for clarity in Fig.

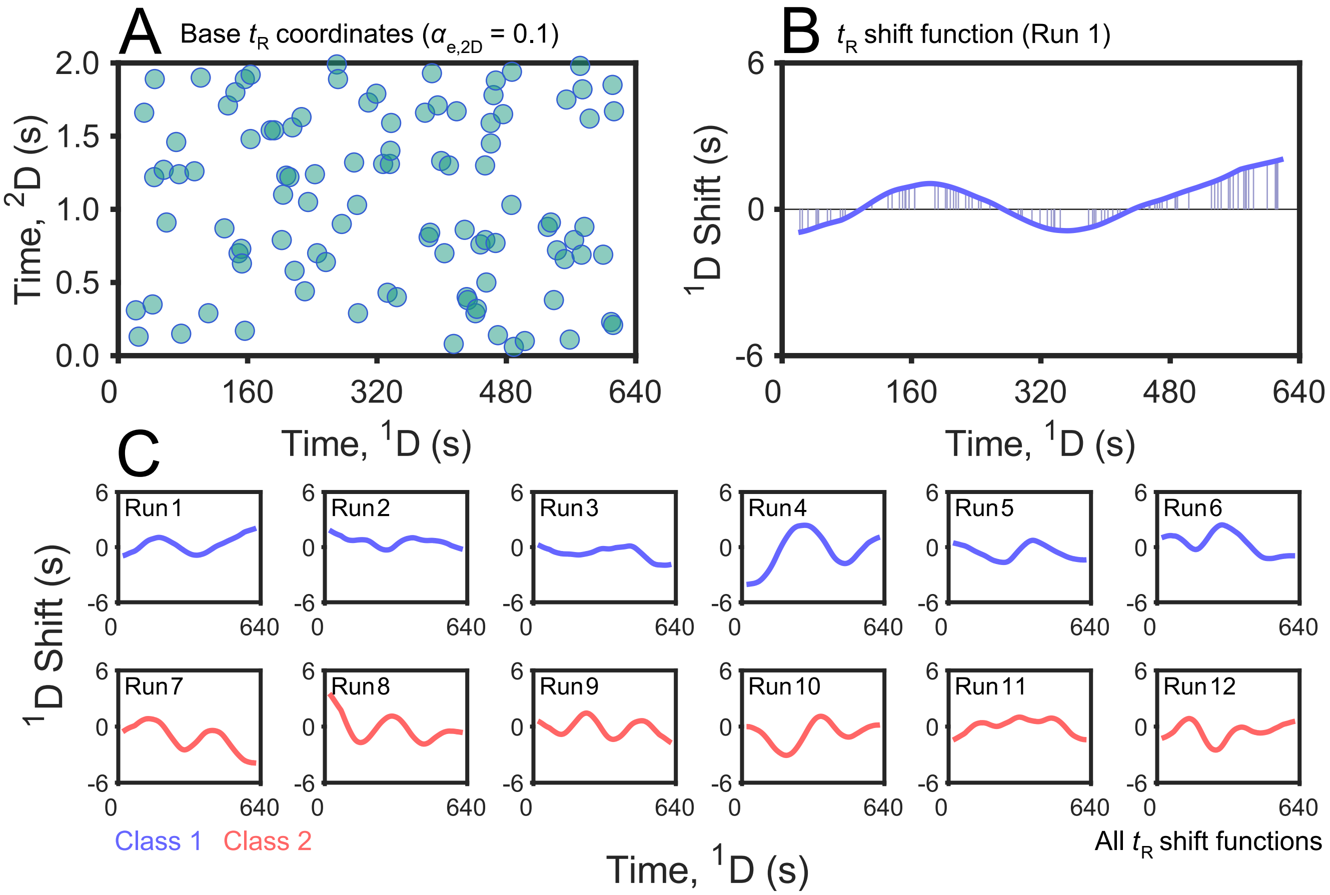


Figure 6.5: Illustration of the method to generate realistic retention time shifts., illustrated here for the ^{1}D separations. (A) First, a "base set" of uniform randomly (i.e., Poisson) distributed retention time coordinates is generated. (B) Next, a shift function is generated, which determines the direction and magnitude of an analyte's shift as a function of its retention time, shown here for Run 1 on the primary separation axis. (C) Examples of all twelve ^{1}D shift functions generated for one SCBE, including that in (B). Using the shift functions, retention time coordinates for the shifted chromatograms are computed. Dots are omitted for clarity. Equivalent methodology is used to shift SCPs on ^{2}D. Conditions: $\alpha_{e,2D} = 0.1$, $\langle\delta_r\rangle = 1.0$.

6.5C. Because it was observed that ^{2}D shift functions exhibit similar behavior to ^{1}D shift functions (Fig. 4B), the same steps are used to generate shifted ^{2}D retention times. We model several levels of retention-time shifting by varying the dimensionless retention time precision, δ_r[59], defined herein as

$$\delta_r = \frac{s_t}{\sigma_b} = \frac{4s_t}{w_b} \tag{6.6}$$

where s_t is the standard deviation of a given SCP retention time between runs. Interpretation of δ_r is simply the ratio of the retention time "distribution" of a given SCP to its width "distribution". For instance, an analyte with $\delta_r = 1.0$ has a distribution of retention times of about one peak-width-at-base, whereas an analyte with $\delta_r = 2.0$ has a shift distribution twice the width of the peak. Computationally, δ_r is modified by varying the amplitude of the shift function. For simplicity, we focus on separations with the same relative amount of shifting on both dimensions, *viz.* δ_r on ^{1}D ($^1\delta_r$) is simulated to be equal to δ_r on ^{2}D ($^2\delta_r$). Results of the shifting method can be visualized via an example simulation (Fig. 6.6A), which shows an overlay of the retention time coordinates for all components in a SCBE

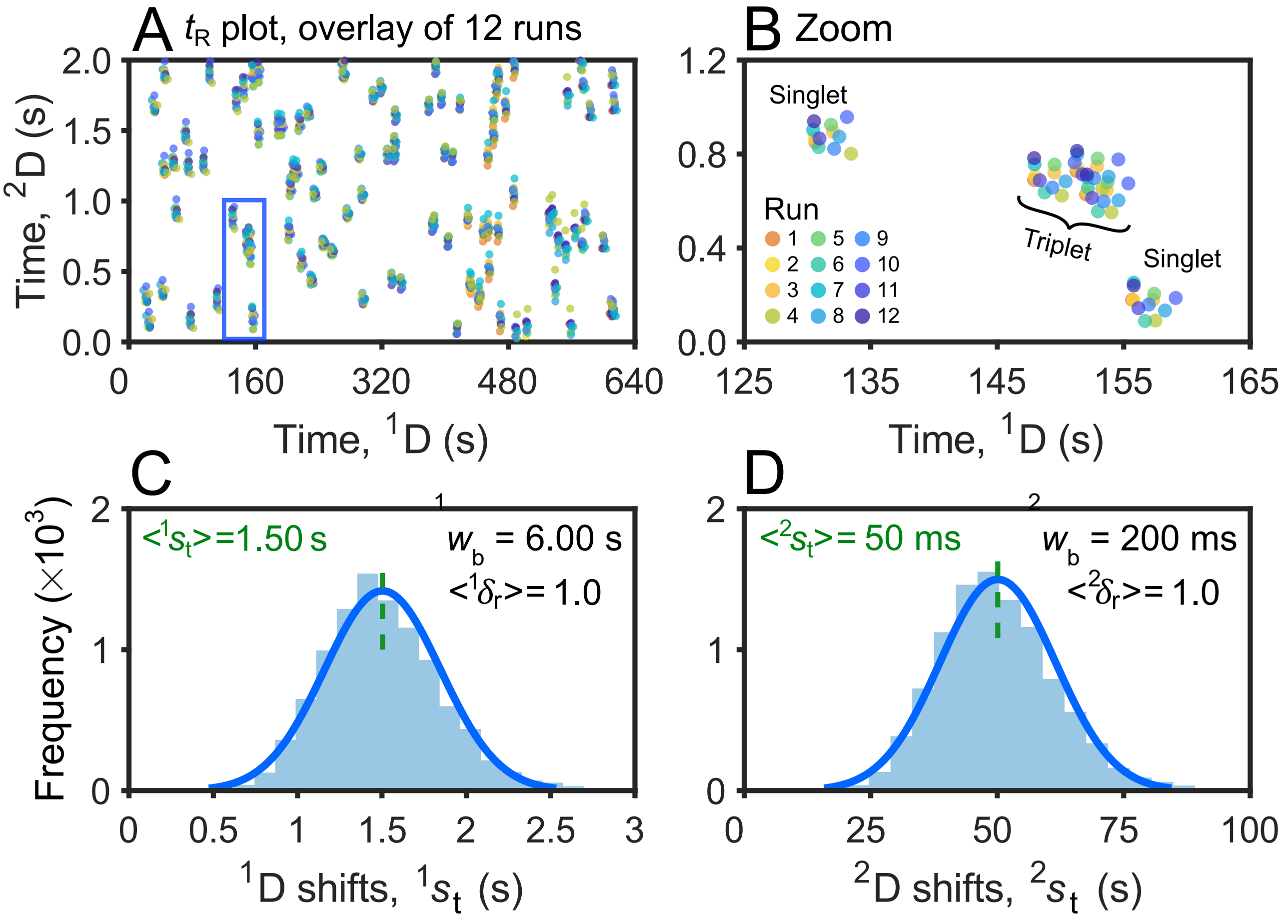


Figure 6.6: (A) Overlay of the retention time coordinates for all twelve runs of the SCBE created in Fig. 6.5. (B) Zoom-in of overlaid retention time coordinates for four components, including a co-elution of three components (triplet) and two pure SCPs (singlets). (C) The standard deviation in ^{1}D retention time, ^1s_t , is computed for every component from 100 SCBE's and plotted as a histogram. (D) Equivalent histogram for retention time shifts on ^{2}D.

simulated with $\alpha_{e,2D} = 0.1$ and a nominal $\delta_r = 1.0$. These conditions were selected for illustration since visualization of shifting at a higher $\alpha_{e,2D}$ is obfuscated by the high extent of peak overlap, while lower levels of δ_r make visualization of shifting difficult across the entire separation space.

Here, 17 of the 20 analytes were pure peaks (singlets), and only three were overlapped with interference SCPs, which is typical for this saturation level. A section of Fig. 6.6A (blue box) containing five SCPs is shown in Fig. 6.6B. Here, two SCP singlets are present along with three SCPs overlapping as a triplet, illustrating the interplay between δ_r and $\alpha_{e,2D}$.

In real chromatograms δ_r is not the same for all components, and the same is true for the SCPs in the simulated chromatograms or SCBEs herein. However, given a sufficient number of SCBEs, the distribution of s_t, and thus δ_r, converges to a normal distribution. Then, the expected, or mean retention time precision is the mean of the distribution of δ_r's and is useful as a modeling variable,

$$\langle \delta_r \rangle = \frac{\langle s_t \rangle}{w_b} \tag{6.7}$$

This is illustrated in Fig. 6.6C and 6.6D, which show distributions of the ^{1}D and ^{2}D shifts at $<\langle\delta_\mathrm{r}\rangle>$ = 1.0. Each point in the distribution is calculated for all runs of a given SCP, and the distributions include 100 SCBEs each containing 100 SCPs.

Four levels of $\langle\delta_\mathrm{r}\rangle$ were examined in this study: 0.0, 0.5, 1.0 and 2.0 (Table 6.1). This $\langle\delta_\mathrm{r}\rangle$ range spans the amount of shifting anticipated in a shorter time-course application such as a data set collected in a single day to a few days ($\langle\delta_\mathrm{r}\rangle$ = 0.0 to 0.5) up to extreme levels of shifting that may occur during a long-term data collection campaign such as a month or more ($\langle\delta_\mathrm{r}\rangle$ = 1.0 to 2.0).

6.2.2 Tile-based FRA

For each combination of modeling parameters, tile-based FRA was performed. Seven different tile sizes (i.e., tile areas) were tested, ranging from tiles smaller than the size of a SCP to tile sizes many times the size of a SCP. The ratio of the tile width to height (i.e., the aspect ratio) was 1 modulation by 5 spectra for all tile sizes studied, enabling a direct comparison of different tile sizes. The tile areas studied spanned an order of magnitude, from 2 modulations by 10 spectra up to 20 modulations by 100 spectra. To make the results generalizable to SCPs of any size, the dimensionless relative tile area (*RTA*) is used, defined as the ratio of the tile area in pixels to the SCP area in pixels,

$$RTA = \frac{\text{Tile area}}{\text{SCP Area}} \tag{6.8}$$

The tile area and SCP area in pixels are given by

$$\text{Tile Area} = {}^1\text{D modulations} \times {}^2\text{D spectra} \tag{6.9}$$

$$\text{SCP area} = n_\mathrm{M} \times f \times {}^2w_\mathrm{b} \tag{6.10}$$

where n_M is the number of detectable ^{2}D peaks, $^2w_\mathrm{b}$ is the ^{2}D peak width-at-base, and f is the mass spectrometer acquisition frequency. Note that n_M is not necessarily equal to the modulation ratio, M_R = $^1w_\mathrm{b}/P_\mathrm{M}$, which for a typically sampled ^{1}D peak is $M_\mathrm{R} \approx$ 2-4. Rather, n_M varies depending on the modulation phase and the S/N [60]. In the present study, n_M = 4 for a ^{1}D peak modulated completely out-of-phase, given $^1w_\mathrm{b}$ = 6 s and P_M = 2 s. We elected to use n_M for an out-of-phase peak because this corresponds to the broadest effective ^{1}D width, $^1w_\mathrm{b}^*$ [44], and therefore the greatest effective SCP size. Then, given $^2w_\mathrm{b}$ = 200 ms and f = 100 Hz the SCP area in the present study is ~80 pixels via Eq. (6.10). Given the SCP size of 80 pixels, the *RTA* can now be calculated for each tile area (Table 6.1). For example, an SCP of 80 pixels in a tile area of 500 pixels (5 modulations by 10 spectra) yields an RTA = 6.25 via Eq. (6.8). This specific example is illustrated in Fig. D5. Of course, SCPs have a 2D profile originating from peaks in each chromatographic dimension that are Gaussian-like, whereas the tiles are rectangular. Nonetheless, the use of *RTA* enables the results of the study to be generalizable to studies with different peak widths than those employed herein. Both the absolute tile areas in pixels and *RTA* used are given in Table 6.1. Grids made at each *RTA* studied are shown in Fig. D6.

6.3 Results and discussion

6.3.1 Method tutorial

We begin with a brief tutorial of how tile-based FRA is performed for each set of conditions, and how the success of each set of conditions is quantified using ROC curves. This workflow is presented in Fig. 6.7. For a given combination of $\alpha_\mathrm{e,2D}$, $\langle\delta_\mathrm{R}\rangle$, and RSD_wc, a SCBE is indexed and input into the tile-based FRA software (Fig. 6.7A). The methodology is illustrated with a typical SCBE simulated under conditions of $\langle\delta_\mathrm{R}\rangle$ = 2.0, $\alpha_\mathrm{e,2D}$ = 0.1, and RSD_wc = 0.5 to 2%. These conditions illustrate successful mitigation of significant run-to-run shifting with minimal chromatographic interference and without high within-class variance. The output is a hit list, which provides the locations of each feature rank-ordered by their F-ratio (Fig. 6.7B). Next, a ROC curve is constructed and the AUROC is calculated. In or-

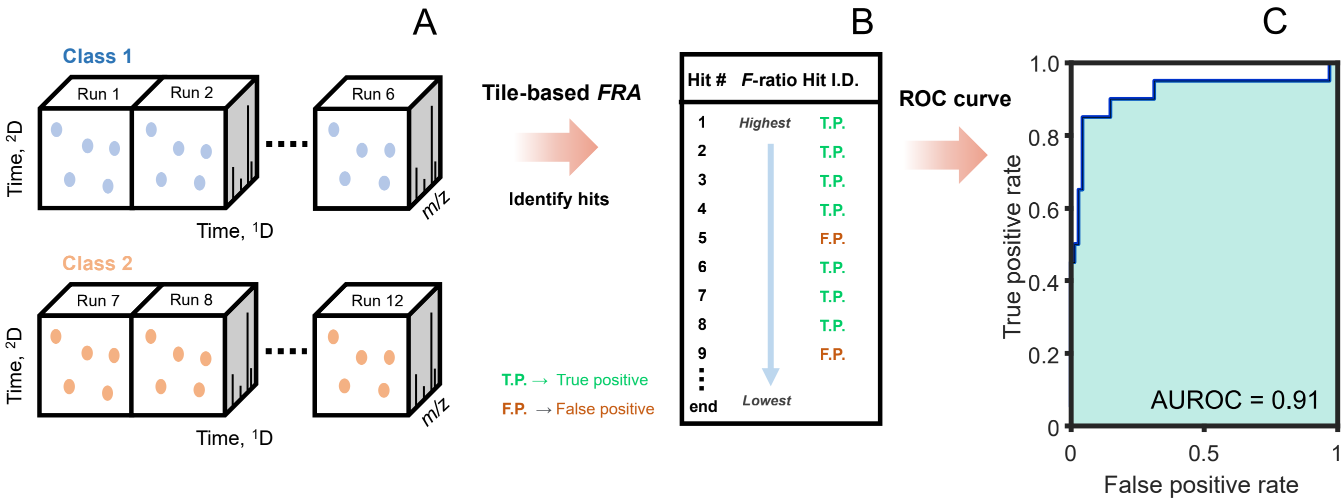


Figure 6.7: Methodology to quantify the "success" of tile-based FRA under each condition ($\langle\delta_{\mathrm{r}}\rangle, \alpha_{\mathrm{e,2D}}, RSD_{\mathrm{wc}}$) , illustrated for a simulation at $\langle\delta_{\mathrm{r}}\rangle = 2.0$, $\alpha_{\mathrm{e,2D}} = 0.1$, and $RSD_{\mathrm{wc}} = 0.5$ to 2%. (A) A SCBE for a given set of conditions is indexed and submitted to tile-based FRA. (B) A hit list is constructed, and each hit is identified as a true positive or false positive based on the known simulated retention time. (C) A ROC curve is generated using the classifications made in (B) and the AUROC is computed. This process is repeated for all forty SCBEs under a given set of conditions to calculate the mean AUROC and repeated for all combinations of conditions.

der to build a ROC curve, each analyte feature on the hit list must be classified as a true positive or a false positive. A flow chart of this hit classification process is given in Fig. D7 and is accomplished by iterating through each hit and determining if the known simulated retention time of any analytes were located within the boundaries of the tile corresponding to the given hit. If an analyte was within the tile, the hit is classified as a true positive and the identity of the analyte is recorded. If no analyte was within the tile, the hit is classified as a false positive. This process repeats iteratively until all true positives are discovered, or until the 20th false positive is reached. We emphasize that the cutoff of the 20th false positive was selected as a reasonable cut-off point to enable a comparison of different combinations of simulation conditions and that an analyst investigating real data might be inclined to exhaustively mine the hit list, well past the 20th false positive. Therefore, the AUROC numbers here are intended for comparison of different method conditions and that more true positives may be discovered if exhaustive examination of each hit list was performed. After hit classification has been accomplished, a ROC curve can be constructed (Fig. 6.7C). In the present context, a ROC curve provides the running sum of true positives discovered, relative to the total number of simulated analytes (i.e., discoverable true positives) versus the running sum of false positives [30]. The AUROC gives the probability that a randomly selected hit would be correctly ranked as either a true positive or false positive. If all of the twenty simulated true positives are at the top of the hit list, the AUROC is 1, and the AUROC decreases as the ranking of true positives in the hit list decreases. If all twenty true positives are discovered but interspersed with false positives (and before the 20th false positive is registered) the true positive rate will reach 1.0 when the last true positive is found, but the AUROC will be less than 1.0.

Note that other metrics could be used as a measure of discovery success. For instance, one could count the number of true positives discovered up to some specified hit number or F-ratio threshold, and then use the percentage of true positives discovered as the metric. However, the information in the AU-

ROC incorporates not only the number of true positives found, but also encompasses their ranking in the hit list, and was therefore utilized in the current study.

In an optimized scenario, the tile should encompass the analyte width on both dimensions as well as its observed run-to-run shifting. To demonstrate optimal capture of an analyte by tile-based FRA, Fig. 6.8 illustrates the tiling of hit #1, N-ethylbenzamine. Here, the tile size is 10 modulations by 50 spectra ($RTA = 6.25$). Figure 6.8A shows the shows a section of a 2D chromatogram of one run, plotted as a pixel-level image where the dashed lines indicate the tile boundaries. Figure 6.8B shows an overlay of contour plots for all twelve runs of the SCBE. In this view, the tile size optimally captures all of the observed shifting. Figure 6.8C and 6.8B show the reconstructed ^{1}D and summed ^{2}D profiles, respectively, illustrating optimal analyte peak capture.

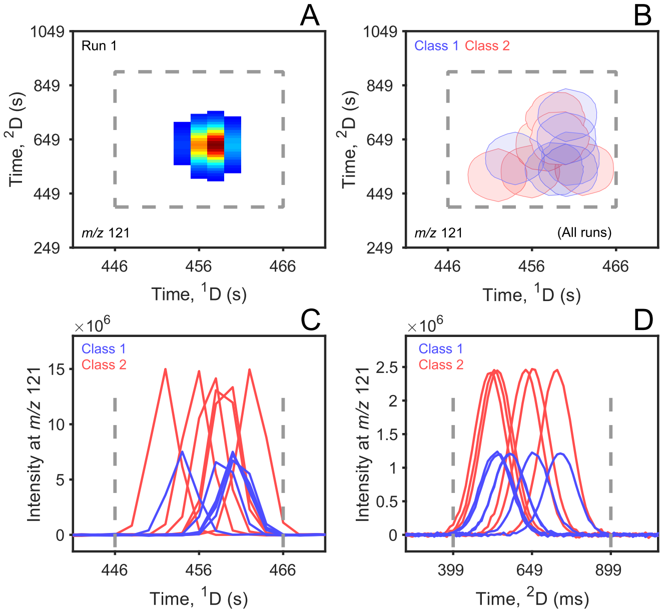


Figure 6.8: Four-panel expansion a true positive (N-ethylbenzamine, hit 1) from a chromatogram simulated under conditions of $\langle \delta_{\mathrm{r}} \rangle = 2.0$, $alpha_{\mathrm{e,2D}} = 0.1$, and $RSD_{\mathrm{wc}} = 0.5$ to 2%. (A) Pixel-level section of the 2D chromatogram for one replicate. (B) Contour overlay of all runs to illustrate shifting across both separation dimensions. (C) Reconstructed 1D peak profiles obtained by summing the total signal in each modulation. (D) Summed ^{2}D profiles obtained by summing together all of the ^{2}D peaks.

Although N-ethylbenzamine was optimally captured, we hypothesize that increasing $\alpha_{\mathrm{e,2D}}$ and RSD_{wc} will hinder analyte discoverability, acting in concert with $\langle \delta_{\mathrm{r}} \rangle$, as is discussed next.

6.3.2 Effect of saturation, shifting, and RSD_{wc} on tile size selection

The interplay of $\alpha_{\mathrm{e,2D}}$, $\langle \delta_{\mathrm{r}} \rangle$, and RSD_{wc} on analyte discoverability is summarized in Fig. 9. Here, the mean AUROCs from forty SCBEs are plotted as a function of tile size, i.e., as a function of RTA in Eq. (6.8), with each trace corresponding to a different level of $\langle \delta_{\mathrm{r}} \rangle$. Note that variation in the mean AUROC for each set of simulated conditions is omitted for clarity but will be provided in a subsequent figure. Figure 6.9 A-C displays the results for $RSD_{\mathrm{wc}} = 0.5\%$ to 2% with an increasing $\alpha_{\mathrm{e,2D}}$ left to right, while Fig. 6.9 D-F similarly shows the results for $RSD_{\mathrm{wc}} = 5\%$ to 50%. A few general observations are noted. First, as $\alpha_{\mathrm{e,2D}}$ increases, the effects of run-to-run shifting become more detrimental, with the difference in AUROC between each successive $\langle \delta_{\mathrm{r}} \rangle$ curves decreasing more rapidly. Likewise, as RSD_{wc} increases (Fig. 6.9 A-C versus Fig. 6.9 D-F), the same trend is observed. Note also that the drawbacks of using a tile size smaller than the optimum are much worse than using a tile size larger than the optimum for all but completely unshifted data. When the tile size is too small, false positives due to shifting are not sufficiently mitigated, negatively affecting the AUROC due to a high number of false positives interspersed near the top of the hit list. In contrast, when the tile size is too large, run-to-run shifting is mitigated, but the likelihood of capturing interferences within the tile increases. Interferences, especially those with much higher concentrations than the analyte(s) of interest, can diminish the contribution of the analyte to a given tile's overall F-ratio, thus impairing discovery.

A valuable distillation of the information from Fig. 6.9 is the optimum relative tile area (RTA_{opt}), determined for each of the sets of conditions. The RTA_{opt} for a given combination of $\alpha_{\mathrm{e,2D}}$, $\langle \delta_{\mathrm{r}} \rangle$, and RSD_{wc} conditions corresponds simply to the curve

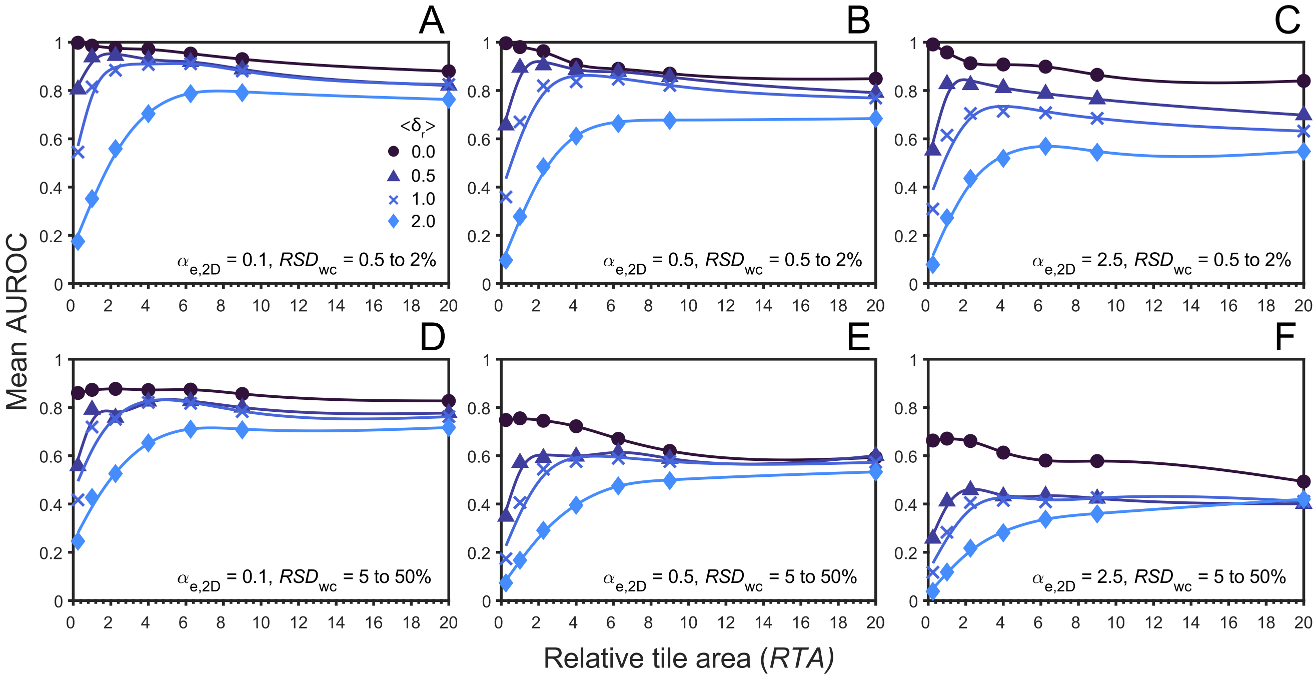


Figure 6.9: Summary of the interplay of $\alpha_{e,2D}$, $\langle\delta_r\rangle$, and RSD_{wc} , on the AUROC, as a function of RTA. Each trace corresponds to different amount of shifting, $\langle\delta_r\rangle$. Titles indicate the $\alpha_{e,2D}$ and RSDwc for each subplot. Abscissa: Relative tile area, RTA, via Eq. (6.8). Ordinate: Mean area under the ROC curve (Mean AUROC), average of 40 SCBEs.

maximum, i.e., the highest mean AUROC, which equates to optimal discoverability of analytes that are changing in concentration at a statistically significant level between sample classes. For example, consider the analyte illustrated in Fig. 6.8, with $\langle\delta_r\rangle$ = 2.0 and RSD_{wc} = 0.5% to 2%. If this analyte were in a separation where $\alpha_{e,2D} = 0.1$, the RTA_{opt} would be 6.25 and the mean AUROC would be 0.8 (Fig. 6.9A). By doing this exercise for all combinations of $\alpha_{e,2D}$, $\langle\delta_r\rangle$, and RSD_{wc}, a summary of the RTA_{opt} obtained from the results in Fig. 6.9 are plotted in Fig. 6.10 as a function of $\langle\delta_r\rangle$. Logically, there is a strong dependence of RTA_{opt} on $\langle\delta_r\rangle$, as larger tiles are necessary to capture and account for larger amounts of shifting. There is also a smaller, yet notable dependence of RTA_{opt} on RSD_{wc}. At each level of $\langle\delta_r\rangle$ the RTA_{opt} is higher for RSD_{wc} = 5% to 50% (shown in red) than for RSD_{wc} = 0.5% to 2% (shown in blue). Finally, there appears to be a weak dependence of RTA_{opt} on $\alpha_{e,2D}$ at a given $\langle\delta_r\rangle$, whereby RTA_{wc} does increase slightly with $\alpha_{e,2D}$ in some instances. Based on these observations, a larger RTA may be necessary when there is significant RSD_{wc} and/or run-to-run shifting. However, when there is minor shifting and/or low RSD_{wc}, the user may benefit from the selection of a relatively smaller RTA.

In order to gain additional insight, the maximum

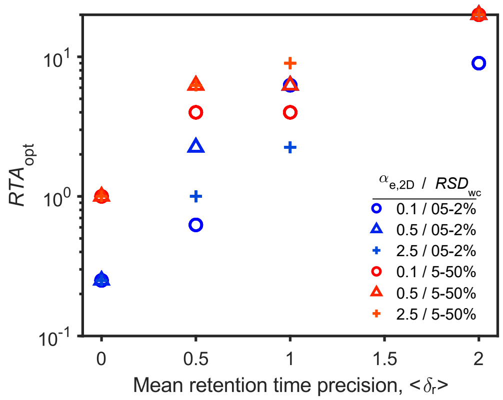


Figure 6.10: Optimum relative tile area (RTA_{opt}) as a function of mean retention time precision, $\langle\delta_r\rangle$. The RTA_{opt} are obtained from the maxima of each trace in Fig. 9. A larger tile size (i.e., RTA) is needed as both $\langle\delta_r\rangle$ and RSD_{wc} increase.

AUROC (from Fig. 6.9) are plotted at RTA_{opt} in Fig. 6.11. This plot illustrates the optimized results for each of the sets of simulated conditions. The error bars in the simulations are provided, which were initially omitted in Fig. 6.9 for clarity. A few key observations are noted. As anticipated, the AUROC for RSD_{wc} = 0.5% to 2% is higher than the RSD_{wc} = 5% to 50% for a given set of conditions. When there is little to minor shifting ($\langle\delta_{\mathrm{r}}\rangle = 0.0$ to 0.5), discoverability is high even at high saturation, e.g., the AUROC is 0.8 for $\alpha_{\mathrm{e,2D}} = 0.1$, 0.5 and 2.5. Conversely, when there is high RSD_{wc} and high shifting, discoverability is impaired, eg., the AUROC is 0.6 for RSD_{wc} = 5 to 50% and $\alpha_{\mathrm{e,2D}} = 0.5$ and 2.5.

It is insightful to relate the observations from the present study to previous work to then provide practical guidance. In a recent study, Sudol et al. used tile-based FRA to characterize the flavor-profile of Sicilian wines [31]. Typical peak widths were ${}^1w_{\mathrm{b}}$ = 3.6 s, ${}^2w_{\mathrm{b}}$ = 240 ms, obtained with a P_{M} = 700 ms and an acquisition frequency of 150 Hz, yielding an SCP size in pixels of 185 pixels via Eq. (10). The tile size was 10 modulations on ^{1}D by 45 spectra, for a tile area of 450 pixels via Eq. (9). Therefore, using Eq. (8) the RTA was 2.43. By visual inspection, the $\alpha_{\mathrm{e,2D}}$ appears closest to $\alpha_{\mathrm{e,2D}} = 0.1$ of the three levels investigated in this study. (Although $\alpha_{\mathrm{e,2D}}$ was not in the scope of the report by Sudol et al. a peak-finding script could be applied to estimate m and thus $\alpha_{\mathrm{e,2D}}$ [61]). Moderate shifting was observed, corresponding to $\delta_{\mathrm{r}} \approx 0.4$. The RSD_{wc} ranged from ~5 to 50%, with a median RSD_{wc} of 12% (likely due to the use of SPME, resulting in s_{SP}^2 contributing significantly). Based on these observations, the conditions were

determined to be closest to those in Fig. 6.9D, triangle trace ($\alpha_{\mathrm{e,2D}} = 0.1$, RSD_{wc} = 5 to 50 %, $\langle\delta_{\mathrm{r}}\rangle$ = 0.5), which yielded an RTA_{opt} of 4.00, very close to the RTA experimentally selected for the wine study. Importantly, the selection of an RTA in the range of 2.25 to 6.25 would have yield a nearly identical level of discoverability. In summary, this study was a moderately challenging case, with minor shifting and RSD_{wc}, but low $\alpha_{\mathrm{e,2D}}$, enabling a high level of discovery.

The yeast metabolome data set [20], used in Fig. 6.4 to generate shift functions provides a more challenging analysis with a higher $\alpha_{\mathrm{e,2D}}$, $\langle\delta_{\mathrm{r}}\rangle$, and RSD_{wc} [29]. We estimate an $\alpha_{\mathrm{e,2D}}$ of about 0.5 due to the complexity of the yeast metabolome, despite a high

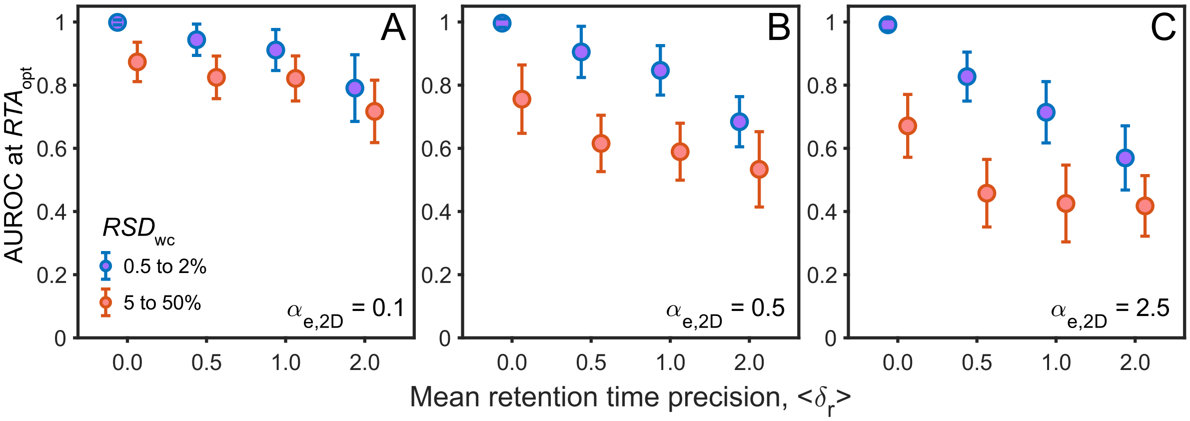


Figure 6.11: Summary plot of the optimized AUROC for each combination of $\alpha_{\mathrm{e,2D}}$ and $\langle\delta_{\mathrm{r}}\rangle$ from Fig. 6.9. Each marker distills the best result from the seven RTA tested per combination of conditions. The mean AUROC is plotted at its RTA_{opt} , which are shown in Fig. 6.10. Although omitted for clarity in Fig. 6.9, error bars are shown here at one standard deviation

$n_{\mathrm{c,2D}} = 6400$. Average peak widths of ${}^{1}w_{\mathrm{b}} = 4.5$ s and ${}^{2}w_{\mathrm{b}} = 80$ ms were observed using an acquisition frequency of 100 Hz and $P_{\mathrm{M}} = 1.5$, yielding an SCP area of 32 pixels via Eq. (6.10). The data were collected over a two month-long campaign, resulting in significant run-to-run shifting. Using the same sixteen analytes used earlier to generate shift functions an average ${}^{1}\delta_{\mathrm{r}}$ of 0.57 and ${}^{2}\delta_{\mathrm{r}}$ of 1.18 were obtained, though ${}^{1}\delta_{\mathrm{r}}$ and ${}^{2}\delta_{\mathrm{r}}$ were highly variable (Table D2). Therefore, a relatively large tile size was used: 6 s on ^{1}D, and 30 spectra on ^{2}D, giving an RTA of 5.62. Significant RSD_{wc} (10 to 88%) [20] was observed due to the combination of s^2_{NatVar} due to biological variation in yeast cultures and the s^2_{SP} due to derivatization. Given the $\alpha_{\mathrm{e,2D}} \approx 0.1$ to 0.5, $\delta_{\mathrm{r}} \approx 0.5$ to 1.0, and high RSD_{wc}, this analysis likely lies between Figs. 6.9D and 6.9E (triangle and "x" traces) with an $RTA_{\mathrm{opt}} \approx 6.25$. This is slightly larger than the applied RTA of 4.00; however, the shifting on both dimensions was highly variable between analytes with ${}^{2}\delta_{\mathrm{r}}$ as high as 2.60 (Table D2). The take-home message is as follows. First, when data is collected over a long campaign, the likelihood of higher run-to-run shifting increases, and a larger tile size is likely necessary to optimize true positive discovery. Also, the effects of "noise" in the data (i.e., peak overlap, shifting, RSD_{wc}) can be mitigated but not completely avoided. As a result, the inferred AUROC from the Watson et al. study [29] is ~0.6 to ~0.8 versus that of the Sudol et al. study [31] at >0.8 where the data was collected in a shorter campaign and with less within-class variance.

We acknowledge that the random attribution of mass spectra to SCPs does not account for the likelihood that closely eluting compounds may exhibit similar fragmentation patterns in real data sets and hence impact discovery. However, since tile-based FRA is performed on every *m/z*, discovery of a discriminatory compound is possible even when a structurally similar compound is overlapping as long as there is sufficient partial selectivity in at least one *m/z*. Thus, we anticipate the current study was not highly impacted by the simulation methodology employed. However, simulating structurally ordered GC×GC separations may be a logical extension to the present work. Since a mass spectrum library of real compounds containing a broad distribution of match values (Fig. 6.2B) was used, samples with low sample dimensionality [62], such as fuels, would be anticipated to have a tighter mass spectrum distribution, which may make discovery more challenging in structurally ordered GC×GC separations.

Beyond using tile-based FRA to mitigate retention time shifting, another approach worthy of investigation is the use of a simple linear shift of the vectorized 1D data to provide an initial "coarse" retention-time correction, which may improve the discovery ability of subsequent tiling with FRA. The linear shift required could be obtained by monitoring a prominent sample peak or by using quality control (QC) standards injected intermediately during the experiment campaign. However, due to the non-linear behavior of retention-time shifting, such coarse corrections are anticipated only to improve the results from tiling and are not suggested as an alternate solution.

6.4 Conclusion

Successful untargeted analysis of GC×GC-TOFMS data sets requires addressing the challenge of run-to-run retention shifting. Although tile-based FRA is a computationally fast and robust data analysis tool that optimizes true positive discoverability while mitigating false positive occurrences due to run-to-run shifting, getting the most out of the software requires knowledge of the chromatographic saturation, $\alpha_{\mathrm{e,2D}}$, nominal retention shifting, $\langle\delta_{\mathrm{r}}\rangle$, and observed within-class signal variability RSD_{wc}, in order to select the optimum tile size, RTA_{opt}. Based on the results presented herein, the following recommendations and observations can be made: (1) for data with any appreciable amount of shifting, choosing a tile size that is too large is better than one with that is too small, (2) in cases where there is high RSD_{wc} and/or $\langle\delta_{\mathrm{r}}\rangle$, the selection of a relatively larger tile size is beneficial, and (3) tile-based FRA can facilitate successful discovery in situations of high chromatographic complexity (high $\alpha_{\mathrm{e,2D}}$), and high levels of run-to-run shifting as long as only two out of these three variables are significant, i.e., $\alpha_{\mathrm{e,2D}}$, RSD_{wc}, and/or $\langle\delta_{\mathrm{r}}\rangle$. Overall, the results in this report are promising for untargeted analysis of GC×GC-TOFMS data using tile-based FRA for a wide range of experimental conditions and should provide the user community with insight into how best to apply this data analysis tool. Furthermore, of significant importance to this study was the simulation of GC×GC-TOFMS data sets containing realistic run-to-run retention shifting. This was accomplished by using low-frequency waveforms to model the "shift-function" between chromatograms. Although the focus of this report was on the evaluation of tile-based FRA preprocessing, the data simulation approach with its shifting methodology introduced herein could be used to evaluate the robustness to shifting of other data analysis/chemometric tools for comprehensive two-dimensional separations data.

6.5 References

(1) Z. Liu, J.B. Phillips, Comprehensive two-dimensional gas chromatography using an on-column thermal modulator interface, *J. Chromatogr. Sci.* **1991**, *29*, 227-231.

(2) J.V. Seeley, S.K. Seeley, Multidimensional gas chromatography: Fundamental advances and new applications, *Anal. Chem.* **2013**, *85*, 557–578.

(3) M.S.S. Amaral, Y. Nolvachai, P.J. Marriott, Comprehensive two-dimensional gas chromatography advances in technology and applications: biennial update, *Anal. Chem.* **2020**, *92*, 85–104.

(4) D. Zanella, J. Focant, F.A. Franchina, 30th Anniversary of comprehensive two-dimensional gas chromatography: Latest advances, *Anal. Sci. Adv.* **2021**, *2*, 213–224.

(5) M.S. Klee, J. Cochran, M. Merrick, L.M. Blumberg, Evaluation of conditions of comprehensive two-dimensional gas chromatography that yield a near-theoretical maximum in peak capacity gain, *J. Chromatogr. A* **2015**, *1383*, 151-159.

(6) K.D. Nizio, J.W. Cochran, S.L. Forbes, Achieving a near-theoretical maximum in peak capacity gain for the forensic analysis of ignitable liquids using GC×GC-TOFMS, *Separations* **2016**, *3*, 1-17.

(7) C.N. Cain, S. Schöneich, R.E. Synovec, Development of an enhanced total ion current chromatogram algorithm to improve untargeted peak detection, *Anal. Chem.* **202**, *92*, 11365-11373.

(8) B. Li, S.E. Reichenbach, Q. Tao, R. Zhu, A streak detection approach for comprehensive two-dimensional gas chromatography based on image analysis, *Neural Comput. Appl.* **2020**, *32*, 649-663.

(9) M.J. Wilde, B. Zhao, R.L. Cordell, W. Ibrahim, A. Singapuri, N.J. Greening, C.E. Brightling, S. Siddiqui, P.S. Monks, R.C. Free, Automating and extending comprehensive two-dimensional gas chromatography data processing by interfacing open-source and commercial software, *Anal. Chem.* **2020**, *92*, 13953–13960.

(10) L.A. Adutwum, J.K. Kwao, J.J. Harynuk, Unique ion filter–A data reduction tool for chemometric analysis of raw comprehensive two-dimensional gas chromatography-mass spectrometry data, *J. Sep. Sci.* **2021**, *44*, 2773-2784.

(11) B.A. Weggler, L.M. Dubois, N. Gawlitta, T. Gröger, J. Moncur, L. Mondello, S.E. Reichenbach, P. Tranchida, Z. Zhao, R. Zimmermann, M. Zoccali, J.F. Focant, A unique data analysis framework and open source benchmark data set for the analysis of comprehensive two-dimensional gas chromatography software, *J. Chromatogr. A* **2021**, *1635*, 461721.

(12) H.D. Bean, J.E. Hill, J.M.D. Dimandja, Improving the quality of biomarker candidates in untargeted metabolomics via peak table-based alignment of comprehensive two-dimensional gas chromatography-mass spectrometry data, *J. Chromatogr. A* **2015**, *1394*, 111-117.

(13) E.A. Higgins Keppler, C.L. Jenkins, T.J. Davis, H.D. Bean, Advances in the application of comprehensive two-dimensional gas chromatography in metabolomics, *TrAC - Trends Anal. Chem.* **2018**, *109*, 275-286.

(14) P.H. Stefanuto, K.A. Perrault, L.M. Dubois, B. L'Homme, C. Allen, C. Loughnane, N. Ochiai, J.F. Focant, Advanced method optimization for volatile aroma profiling of beer using two-dimensional gas chromatography time-of-flight mass spectrometry, *J. Chromatogr. A* **2017**, *1507*, 45-52.

(15) F. Stilo, C. Bicchi, A. Robbat, S.E. Reichenbach, C. Cordero, Untargeted approaches in food-omics: The potential of comprehensive two-dimensional gas chromatography/mass spectrometry, *TrAC - Trends Anal. Chem.* **2021** *135*, 116162.

(16) B. Gruber, B.A. Weggler, R. Jaramillo, K.A. Murrell, P.K. Piotrowski, F.L. Dorman, Comprehensive two-dimensional gas chromatography in forensic science: A critical review of recent trends, *TrAC - Trends Anal. Chem.* **2018**, *105*, 292-301.

(17) A.M. Muscalu, T. Górecki, Comprehensive two-dimensional gas chromatography in environmental analysis, *TrAC - Trends Anal. Chem.* **2018**, *106*, 225-245.

(18) A. Barcaru, G. Vivó-Truyols, Use of Bayesian statistics for pairwise comparison of megavariate data sets: Extracting meaningful differences between GC×GC-MS chromatograms using Jensen-Shannon divergence, *Anal. Chem.* **2016**, *88*, 2096–2104.

(19) B.J. Pollo, G.L. Alexandrino, F. Augusto, L.W. Hantao, The impact of comprehensive two-dimensional gas chromatography on oil & gas analysis: Recent advances and applications in petroleum industry, *TrAC - Trends Anal. Chem.* **2018**, *105*, 202-217.

(20) R.E. Mohler, K.M. Dombek, J.C. Hoggard, K.M. Pierce, E.T. Young, R.E. Synovec, Comprehensive analy-

sis of yeast metabolite GC×GC-TOFMS data: Combining discovery-mode and deconvolution chemometric software, *Analyst* **2007**, *132*, 756–767.

(21) L.C. Marney, W. Christopher Siegler, B.A. Parsons, J.C. Hoggard, B.W. Wright, R.E. Synovec, Tile-based Fisher-ratio software for improved feature selection analysis of comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry data, Talanta 115 (2013) 887–895.

(22) Y. Zushi, J. Gros, Q. Tao, S.E. Reichenbach, S. Hashimoto, J.S. Arey, Pixel-by-pixel correction of retention time shifts in chromatograms from comprehensive two-dimensional gas chromatography coupled to high resolution time-of-flight mass spectrometry, *J. Chromatogr. A* **2017**, *1508*, 121-129.

(23) S. Stadler, P.H. Stefanuto, M. Brokl, S.L. Forbes, J.F. Focant, Characterization of volatile organic compounds from human analogue decomposition using thermal desorption coupled to comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry, *Anal. Chem.* **2013**, *85*, 998-1005.

(24) J. Heim, Using the Statistical Compare and Fisher ratio ChromaTOF features to define variance prior to multivariate analysis in the small metabolite profile of diabetic versus non-diabetic urine by GC×GC-TOFMS, *Sep. Sci. Appl. Notes.* **2009**, 1-8.

(25) C.A. Manzano, N.G. Dodder, E. Hoh, R. Morales, Patterns of personal exposure to urban pollutants using personal passive samplers and GC×GC/ToFMS, *Environ. Sci. Technol.* **2018**, *53*, 614-624.

(26) M. Ueland, S. Harris, S.L. Forbes, Detecting volatile organic compounds to locate human remains in a simulated collapsed building, *Forensic Sci. Int.* **2021**, *323*, 110781.

(27) P. Teehan, M.K. Schall, V.S. Blazer, F.L. Dorman, Targeted and non-targeted analysis of young-of-year smallmouth bass using comprehensive two-dimensional gas chromatography coupled with time-of-flight mass spectrometry, *Sci. Total Environ.* **2022**, *806*, 150378.

(28) B.A. Parsons, L.C. Marney, W.C. Siegler, J.C. Hoggard, B.W. Wright, R.E. Synovec, Tile-based Fisher ratio analysis of comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry (GC×GC-TOFMS) data using a null distribution approach, *Anal. Chem.* **2015**, *87*, 3812-3819.

(29) N.E. Watson, B.A. Parsons, R.E. Synovec, Performance evaluation of tile-based Fisher Ratio analysis using a benchmark yeast metabolome data set, *J. Chromatogr. A* **2016**, *1459*, 101–111.

(30) B.C. Reaser, B.W. Wright, R.E. Synovec, Using receiver operating characteristic curves to optimize discovery-based software with comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry, *Anal. Chem.* **2017**, *89*, 3606-3612.

(31) P.E. Sudol, M. Galletta, P.Q. Tranchida, M. Zoccali, L. Mondello, R.E. Synovec, Untargeted profiling and differentiation of geographical variants of wine samples using headspace solid-phase microextraction flow-modulated comprehensive two-dimensional gas chromatography with the support of tile-based Fisher ratio analysis, *J. Chromatogr. A* **2022**, *1662*, 462735.

(32) K. Murtada, D. Bowman, M. Edwards, J. Pawliszyn, Thin-film microextraction combined with comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry screening for presence of multiclass organic pollutants in drinking water samples, *Talanta* **2022**, *242*, 123301.

(33) A.P. de la Mata, J.J. Harynuk, Limits of detection and quantification in comprehensive multidimensional separations. 1. A theoretical look, *Anal. Chem.* **2012**, *84*, 6646-6653.

(34) J.M. Davis, J.C. Giddings, Statistical method for estimation of number of components from single complex chromatograms: Theory, computer-based testing, and analysis of errors, *Anal. Chem.* **1985**, *57*, 2168-2177.

(35) J.M. Davis, C. Samuel, The need for two-dimensional gas chromatography: Extent of overlap in one-dimensional gas chromatograms, *J. High Resolut. Chromatogr.* **2000**, *23*, 235-244.

(36) N.M. Devitt, J.M. Davis, M.R. Schure, Estimation of low-level components lost through chromatographic separations with finite detection limits, *J. Chromatogr. A* **2020**, *1626*, 461266.

(37) B.L. Karger, L.R. Snyder, C. Horvath, An Introduction to Separation Science, John Wiley & Sons, New York, 1973.

(38) J.C. Giddings, Two-dimensional separations: Concept and promise, *Anal. Chem.* **1984**, *56*, 1258A-1270A.

(39) C.J. Venkatramani, J. Xu, J.B. Phillips, Separation orthogonality in temperature-programmed comprehensive two-dimensional gas chromatography, *Anal. Chem.* **1996**, *68*, 1486-1492.

(40) D. Ryan, P. Morrison, P.J. Marriott, Orthogonality considerations in comprehensive two-dimensional gas chromatography, *J. Chromatogr. A* **2005**, *1071*, 47-3.

(41) S.C. Rutan, J.M. Davis, P.W. Carr, Fractional coverage metrics based on ecological home range for calculation of the effective peak capacity in comprehensive two-dimensional separations, *J. Chromatogr. A* **2012**, *1255*, 267-276.

(42) C.J. Venkatramani, J.B. Phillips, Comprehensive two-dimensional gas chromatography applied to the analysis of complex mixtures, *J. Microcol. Sep.* **1993**, *5*, 511-516.

(43) R.E. Murphy, M.R. Schure, J.P. Foley, Effect of sampling rate on resolution in comprehensive two-dimensional liquid chromatography, *Anal. Chem.*, *1998*, *80*, 1585-1594.

(44) J.M. Davis, D.R. Stoll, P.W. Carr, Effect of first-dimension undersampling on effective peak capacity in comprehensive two-dimensional separations, *Anal. Chem.* **2008**, *80*, 461-73.

(45) J.M. Davis, P.W. Carr, Effective saturation: A more informative metric for comparing peak separation in one- and two-dimensional separations, *Anal. Chem.* **2009**, *81*, 1198-1207.

(46) F.A. Franchina, L.M. Dubois, J.F. Focant, In-depth cannabis multiclass metabolite profiling using sorptive extraction and multidimensional gas chromatography with low- and high-resolution mass spectrometry, *Anal. Chem.* **2020**, *92*, 10512-10520.

(47) L.J. Nagels, W.L. Cretan, P.M. Vanpeperstraete, Determination limits and distribution function of ultraviolet absorbing substances in liquid chromatographic analysis of plant extracts, *Anal. Chem.* **1983**, *55*, 216-220.

(48) C.G. Enke, L.J. Nagels, Undetected components in natural mixtures: How many? What concentrations? Do they account for chemical noise? What is needed to detect them?, *Anal. Chem.* **2011**, *83*, 2539-2546.

(49) R.J.J. Vreuls, J. Dallüge, U.A.Th. Brinkman, Gas chromatography-time-of-flight mass spectrometry for sensitive determination of organic microcontaminants, *J. Microcol. Sep.* **1999**, *11*, 663-375.

(50) L.N. Williamson, M.G. Bartlett, L.N. Williamson, M.G. Bartlett, Quantitative gas chromatography/time-of-flight mass spectrometry: A review, *Biomed. Chromatogr.* **2007**, *21*, 664-669.

(51) C. Cordero, J. Kiefl, S.E. Reichenbach, C. Bicchi, Characterization of odorant patterns by comprehensive two-dimensional gas chromatography: A challenge in omic studies, *TrAC - Trends Anal. Chem.* **2019**, *113*, 364-78.

(52) H. Liu, L. Lam, L. Yan, B. Chi, P.K. Dasgupta, Expanding the linear dynamic range for quantitative liquid chromatography-high resolution mass spectrometry utilizing natural isotopologue signals, *Anal. Chim. Acta.* **2014**, *850*, 65-70.

(53) Pegasus® GC-HRT+, https://www.leco.com/product/pegasus-gc-hrt (accessed 29.06.22)

(54) S.E. Stein, D.R. Scott, Optimization and testing of mass spectral library search algorithms for compound identification, *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 859-866.

(55) C.N. Cain, P.E. Sudol, K.L. Berrier, R.E. Synovec, Development of variance rank initiated-unsupervised sample indexing for gas chromatography-mass spectrometry analysis, *Talanta* **2021**, *233*, 122495.

(56) C. Monzón, S. Schöneich, R.E. Synovec, Non-targeted discovery of class-distinguishing metabolites in Argentinian pacu fish by comprehensive two-dimensional gas chromatography with principal component analysis, *Microchem. J.* **2021**, *164*, 106004.

(57) P.E. Sudol, G.S. Ochoa, R.E. Synovec, Investigation of the limit of discovery using tile-based Fisher ratio analysis with comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry, *J. Chromatogr. A* **2021**, *1644*, 462092.

(58) M.D. Krebs, R.D. Tingley, J.E. Zeskind, M.E. Holmboe, J.M. Kang, C.E. Davis, Alignment of gas chromatography-mass spectrometry data by landmark selection from complex chemical mixtures, *Chemom. Intell. Lab. Syst.* **2006**, *81*, 74-81.

(59) T.J. Bahowick, R.E. Synovec, Sequential chromatogram ratio technique: Evaluation of the effects of retention time precision, adsorption isotherm linearity, and detector linearity on qualitative and quantitative analysis, *Anal. Chem.* **1992** 64 (1992) 489–496.

(60) O. Amador-Muñoz, P.J. Marriott, Quantification in comprehensive two-dimensional gas chromatography and a model of quantification based on selected summed modulated peaks, *J. Chromatogr. A* **2008**, *1184*, 323-340.

(61) K. Rowe, D. Bowlin, M. Zou, J.M. Davis, Application of 2-D statistical theory of overlap to three separation types: 2-D thinlayer chromatography, 2-D gas chromatography, and liquid chromatography/capillary electrophoresis, *Anal. Chem.* **1995**, *67*, 2994-3003.

(62) J.C. Giddings, Sample dimensionality: A predictor of order-disorder in component peak distribution in multidimensional separation, *J. Chromatogr. A* **1995**, *703*, 3-15.

APPENDIX A

This Appendix is reproduced in part from the Electronic Supplementary Content of T.J. Trinklein, S. Schoneich, P.E. Sudol, C.G. Warren, D.V. Gough, R.E. Synovec, Total-transfer comprehensive three-dimensional gas chromatography with time-of-flight mass spectrometry, *J Chromatogr. A* **2020**, *1634*, 461654.

Alkanes	bp (°C)
hexane	69
heptane	98
heptane	126
nonane	151
decane	174
undecane	196
dodecane	216
tridecane	235
tetradecane	254
pentadecane	271
Ketones	**bp (°C)**
2-butanone	80
2-pentanone	102
3-hexanone	128
2-heptanone	152
3-heptanone	141
3-octanone	167
2-decanone	209
2-undecanone	231
2-dodecanone	231
Halogenated Alkanes	**bp** (°C)
1-bromohexane	155
1-bromoheptane	179
1-bromooctane	200
1-chlorobutane	78
1,1,1-chloroethane	74
1,2-dichloroethane	84
carbon tetrachloride	77
1,5-dichloropentane	179
1-chlorohexane	135
Esters	**bp (°C)**
ethyl formate	54
methyl decanoate	224
methyl caprylate	193
methyl salicylate	223
ethy salicylate	223
methyl caproate	151
diethyl phthalate	299
Cyclic alkanes	**bp** (°C)
methylcyclopentane	72
cycohexane	81
butylcyclohexane	181
bicyclohexyl	227
1-undecene	194
Alcohols	**bp (°C)**
1-propanol	97
2-butanol	177
1-pentanol	137
1-decanol	229
1-tetradecanol	289
1-octadecanol	210
hexyl alchohol	160
2-heptanol	160
1-nonanol	215
1-eicosanol	372
benzyl alcohol	205
2-ethyl-1-hexanol	185
α-terpineol	219
Alkenes	**bp** (°C)
1-hexene	63
cyclohexene	83
dodecene	214
1-undecene	194
Aromatics	**bp (°C)**
benzene	80
toluene	111
3-ethyltoluene	161
mesitylene	165
ethylbenzene	136
butylbenzene	183
isobutylbenzene	170
tert-butylbenzene	167
propylbenzene	159
1-ethylnaphtalene	260
bromobenzene	155
cyclohexylbenzene	239
diphenylmethane	265
p-xylene	189
o-xylene	145
m-xylene	123
1,2,3-trimethylbenzene	169
anisole	1254
dibutyl phthalate	340
Alkynes	**bp (°C)**
5-decyne	177
1-hexyne	71
1-heptyne	109
1-nonyne	151

Table A1: List of analytes in the 90-component test mixture and their boiling points (°C).

APPENDIX B

This Appendix is reproduced in part from the Supporting Information of T.J. Trinklein, C.G. Warren, R.E. Synovec, Determination of the signal-to-noise ratio enhancement in comprehensive three-dimensional gas chromatography, *Anal. Chem.* **2021**, *93*, 8526-8535.

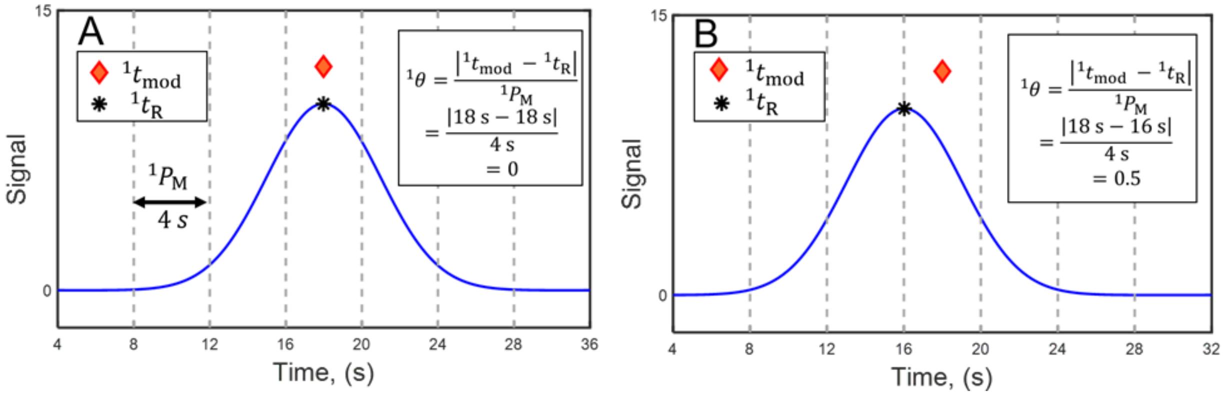


Figure B1: Illustration of sample phase in the context of Eq (4.7). (A) Illustration of in-phase sampling of a 1D peak (k = ki = 0.0), with relevant variables from Eq (4.7) labeled ($^k\theta = |^k t_{\text{mod}} - {}^k t_{\text{R}}|/{}^k P_{\text{M}}$). (B) Illustration of out-of-phase modulation of a ^{1}D peak ($^k\theta = {}^k\theta_{\text{o}} = 0.5$). Note that the sampling events remain constant, but the 1D peak has been repositioned by $\Delta^1 t_{\text{R}} = {}^1P_M/2 = 2$ s.

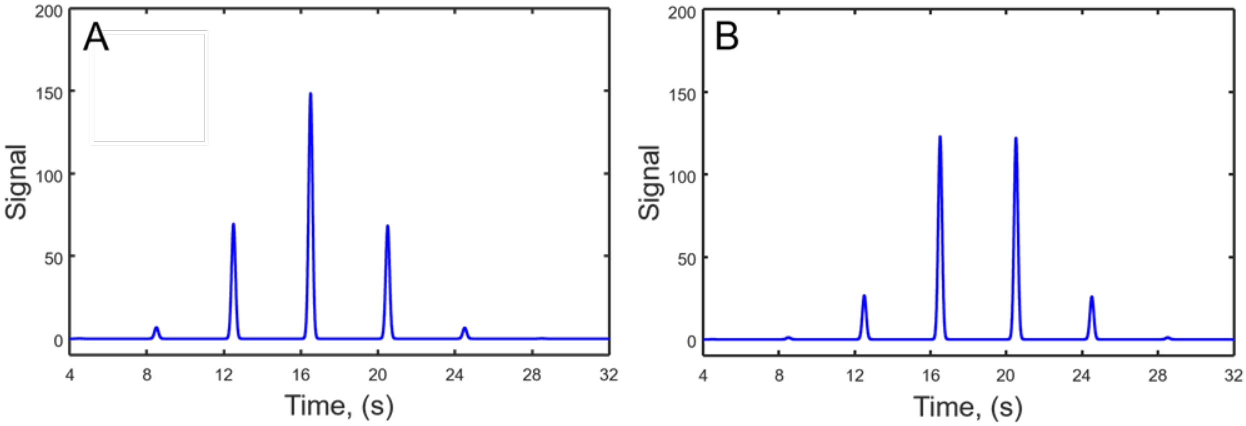
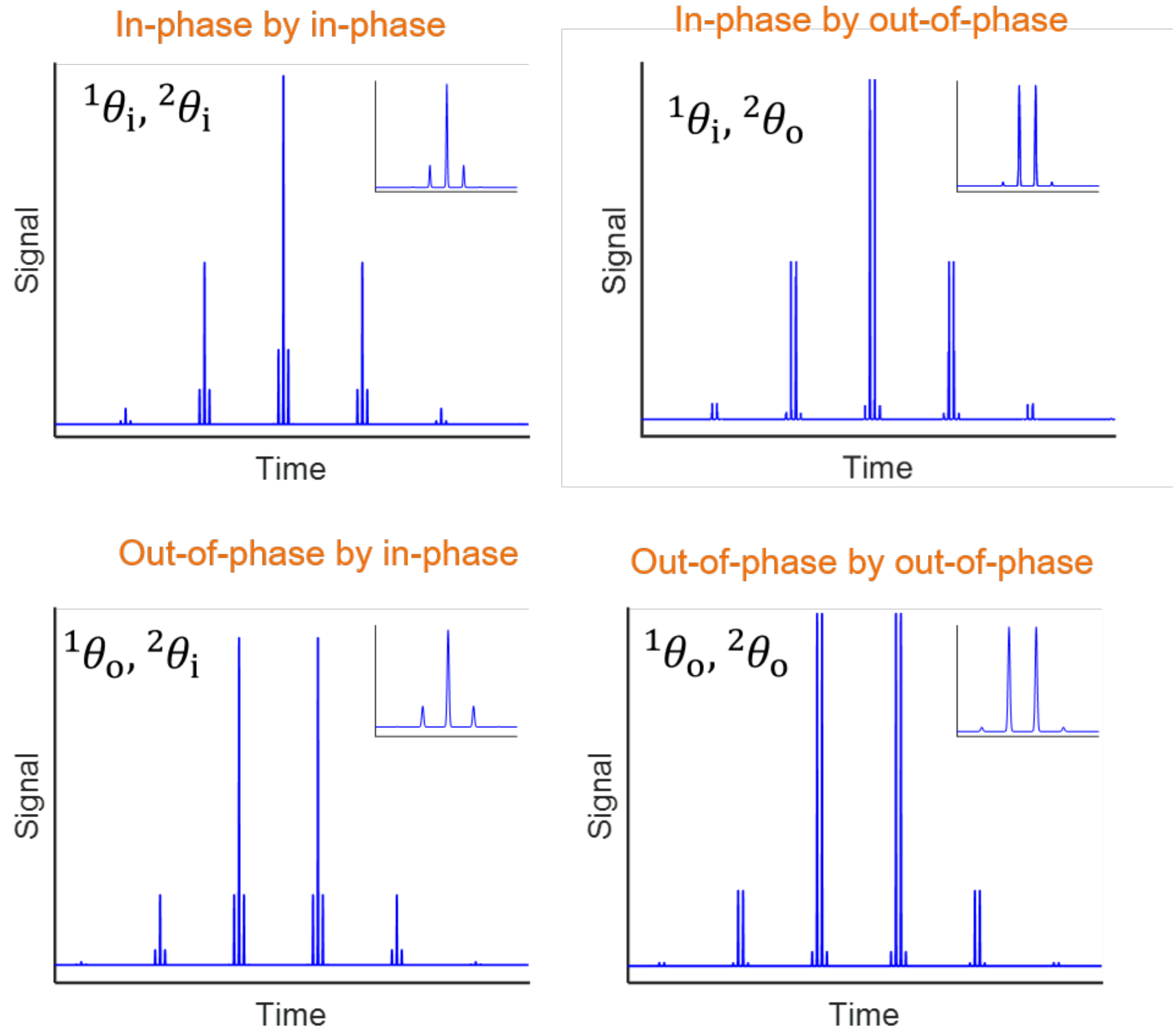


Figure B2: Peak patterns for in-phase and out-of-phase sampled peaks. (A) Modulated ^{2}D peaks following in-phase sampling of the 1D peak in Figure B1A. (B) Modulated ^{2}D peaks following out-of-phase sampling of the ^{1}D peak in Figure B1B.


In-phase by in-phase

In-phase by out-of-phase

Out-of-phase by in-phase

Out-of-phase by out-of-phase

Figure B3: Illustration of the four ideal GC^3 phase cases from simulated data. The insets provide a close-up view of the tallest set of ^{3}D peaks, which make up the tallest ^{2}D peak.

6.6 Derivation of the general ${}^3_1SE$ expression for any phase combination

We now consider the general phase case. The modulation phase Eq. (4.7)is given by

$$ {}^k\theta = \frac{|{}^kt_{\mathrm{mod}} - {}^kt_{\mathrm{R}}|}{{}^kP_{\mathrm{M}}} \tag{B11} $$

To obtain ${}^2a$, we integrate over the ^{1}D peak, Eq. (4.10), where the sampled fraction is bound by the arbitrary limits ${}^1t_1$ and ${}^1t_2$.

$$ {}^2a = \int_{{}^1t_1}^{{}^1t_2} = {}^1C(t)\mathrm{d}t = \Phi({}^1t_2) - \Phi({}^1t_1) \tag{B12} $$

In order to derive an expression for any ${}^1\theta$, the integration limits must be written in terms of ${}^1\theta$. Because ${}^1\theta$ is referenced to the middle of the closest modulation, and normalized by ${}^1P_{\mathrm{M}}$ via Eq. (B1), the lower integration bound can be written as

$$ {}^1t_1 = {}^1t_{\mathrm{R}} - {}^1P_{\mathrm{M}}\left({}^1\theta + 0.5\right) \tag{B13} $$

Each modulation samples the preceding dimension peak for a duration of ${}^1P_{\mathrm{M}}$, thus, the upper integration bound is one ${}^1P_{\mathrm{M}}$ after the lower bound,

$$ {}^1t_2 = {}^1t_{\mathrm{R}} - {}^1P_{\mathrm{M}}\left({}^1\theta + 0.5\right) + {}^1P_{\mathrm{M}} \tag{B14} $$

The retention time of a peak does not affect its area, thus, we let ${}^1t_{\mathrm{R}} = 0$. Applying Eq. (4.10) and simplifying yields

$$ {}^2a = \int_{-{}^1P_{\mathrm{M}}/2(2\cdot{}^1\theta+1)}^{-{}^1P_{\mathrm{M}}/2\left(2\cdot{}^1\theta-1\right)} {}^1C(t)\mathrm{d}t \tag{B15} $$

$$ \Phi\left(\frac{{}^1P_{\mathrm{M}}(2\cdot{}^1\theta-1)}{2}\right) - \Phi\left(\frac{{}^1P_{\mathrm{M}}(2\cdot{}^1\theta+1)}{2}\right) \tag{B16} $$

Next, the expression is written in terms of the error function, via Eq. 4.18

$$ {}^2a = \frac{\mathrm{erf}}{2}\left(\frac{{}^1P_{\mathrm{M}}(2\cdot{}^1\theta-1)}{2\cdot{}^1\sigma\sqrt{2}}\right) - \frac{\mathrm{erf}}{2}\left(\frac{{}^1P_{\mathrm{M}}(2\cdot{}^1\theta+1)}{2\cdot{}^1\sigma\sqrt{2}}\right) \tag{B17} $$

For a 1D peak modulated at ${}^1P_{\mathrm{M}}$, ${}^1w_{\mathrm{b}} = 4\cdot{}^1\sigma$ and the sampling density (i.e., modulation ratio) is ${}^1\rho = {}^1w_{\mathrm{b}}/{}^1P_{\mathrm{M}}$, giving

$$ {}^2a = \frac{\mathrm{erf}}{2}\left(\frac{-\sqrt{2}(2\cdot{}^1\theta-1)}{{}^1\rho}\right) - \frac{\mathrm{erf}}{2}\left(\frac{-\sqrt{2}(2\cdot{}^1\theta+1)}{{}^1\rho}\right) \tag{B18} $$

For modulation which is completely in-phase, or completely out-of-phase, this expression simplifies considerably. Recall that in-phase modulation has k = 0.0, and noting that $\mathrm{erf}(-x) = -\mathrm{erf}(x)$, we obtain

$$ {}^2a_{{}^2\theta_{\mathrm{i}}} = \mathrm{erf}\left(\frac{\sqrt{2}}{{}^1\rho}\right) \tag{B19} $$

Which is equivalent to Eq. (B22). Similarly, ${}^{3}a$ is obtained by the same principles, integrating over ${}^{2}C(t)$

$$\begin{aligned}
{}^{3}a &= {}^{2}a \int_{-{}^{2}P_{\mathrm{M}}/2(2\cdot{}^{2}\theta+1)}^{-{}^{2}P_{\mathrm{M}}/2\left(2\cdot{}^{2}\theta-1\right)} {}^{2}C(t)\mathrm{d}t \qquad &\text{(B20)}\\
&= {}^{2}a\left[\frac{\mathrm{erf}}{2}\left(\frac{-\sqrt{2}(2\cdot{}^{2}\theta-1)}{{}^{2}\rho}\right)-\frac{\mathrm{erf}}{2}\left(\frac{-\sqrt{2}(2\cdot{}^{2}\theta+1)}{{}^{2}\rho}\right)\right] \qquad &\text{(B21)}\\
& &\text{(B22)}
\end{aligned}$$

Then, ${}_{1}^{3}SE$ for any combination of ${}^{1}\theta$ and ${}^{2}\theta$ is obtained by substitution of Eq. B8 and B11 into Eq. (B18)

$$\begin{aligned}
{}_{1}^{3}SE = \frac{{}^{1}w_{\mathrm{b}}}{{}^{3}w_{\mathrm{b}}}&\left[\frac{\mathrm{erf}}{2}\left(\frac{-\sqrt{2}(2\cdot{}^{1}\theta-1}{{}^{1}\rho}\right)-\frac{\mathrm{erf}}{2}\left(\frac{-\sqrt{2}(2\cdot{}^{1}\theta+1}{{}^{1}\rho}\right)\right]\\
&\times\left[\frac{\mathrm{erf}}{2}\left(\frac{-\sqrt{2}(2\cdot{}^{2}\theta-1}{{}^{2}\rho}\right)-\frac{\mathrm{erf}}{2}\left(\frac{-\sqrt{2}(2\cdot{}^{2}\theta+1}{{}^{2}\rho}\right)\right] \qquad \text{(B23)}
\end{aligned}$$

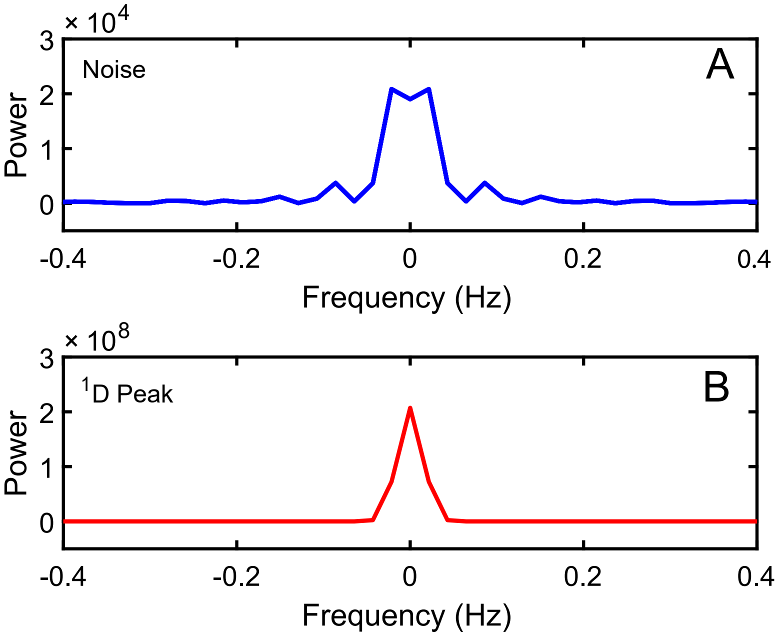


Figure B4: Power spectrum of the baseline noise section and 1D peak for 2-methylthiophene (Fig. 4.7A). (A) The power spectrum for the raw noise (shown in Fig. 4.7A) is compared with (B) the power spectrum of the 1D peak of 2-methylthiophene (with 10^4-fold larger y-axis). The bandwidth of the noise and raw 1D peak are roughly equivalent in frequency space.

APPENDIX C

This Appendix is reproduced in part from the Supporting Information of T.J. Trinklein, J. Jiang, R.E. Synovec, Profiling olefins in gasoline by bromination Using GC×GC-TOFMS followed by discovery-based comparative analysis, *Anal. Chem.* **2022**, *94*, 9407-9414.

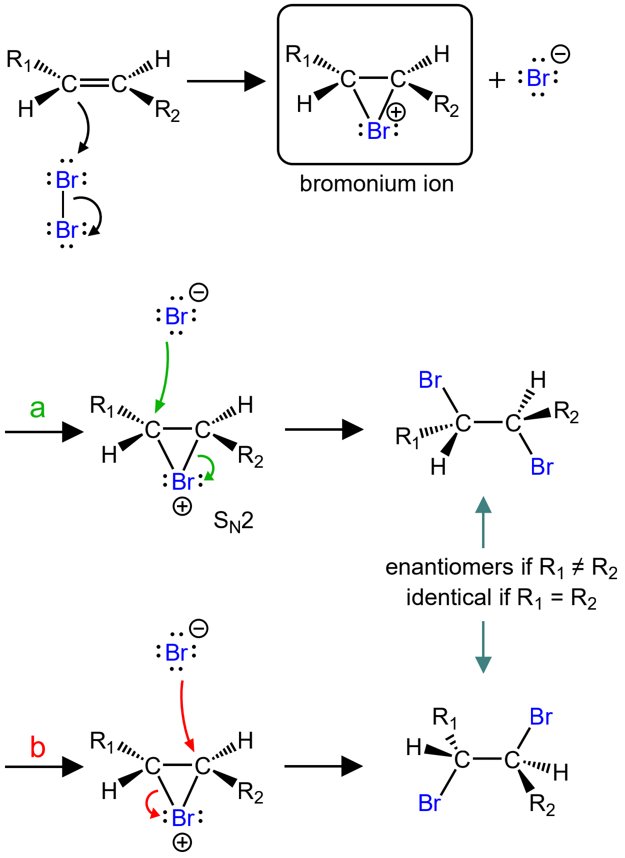


Figure C1: Alkene bromination reaction mechanism. When the Br-Br bond approaches the bond of an alkene, the Br-Br bond becomes polarized. Then, the p electrons of the alkene π bond attack the partial positive Br atom forming a new σ bond. This leads to formation of the cyclic bromonium ion intermediate. The bromonium ion is unstable due to high ring strain, making it susceptible to nucleophilic attack from the free Br^- ion. Because attack occurs from backside of the bromonium ring (via an $\mathrm{S_N}2$ mechanism), anti-addition occurs. However, the free Br^- can attack either carbon (paths a and b) which leads to two final products. For any asymmetric molecule ($R_1 \neq R_2$), two enantiomers are produced, otherwise the products are identical (meso compounds).

Analyte	**Concentration (ppm)**	^1t_R **(min)**	^2t_R **(sec)**	**Quant** m/z	**%Reacted (1.s.d.)**
Alkanes					
hexane	236	-[a]	-[a]	-[a]	-[a]
hexane	266	1.77	0.18	57	-1.5 (5.4)
octane	290	2.52	0.79	57	-3.4 (3.5)
nonane	276	3.87	1.59	57	-5.3 (3.6)
decane	324	5.97	2.27	57	-1.9 (3.6)
undecane	248	11.27	15.52	57	1.1 (3.8)
Average	-	-	-	-	-1.45 (2.8)
1-alkenes					
1-pentene	227	-[a]	-[a]	-[a]	-[a]
1-hexene	267	-[a]	-[a]	-[a]	-[a]
1-heptene	263	1.82	2.07	57	99.1 (0.2)
1-octene	284	2.62	4.77	57	99.5 (1.0)
1-nonene	284	4.07	8.07	57	99.4 (0.6)
1-decene	276	6.22	11.12	57	100 (0.1)
1-dodecene	286	11.52	13.67	57	99.8 (0.5
Average	-	-	-	-	99.5 (0.5)
Aromatics					
toluene	281	3.47	0.29	91	-4.3 (2.9
o-xylene	337	6.22	0.69	91	-0.1 (3.1)
propylbenzene	255	7.52	0.93	91	10.3 (3.5)
4-ethyltoluene	278	7.82	0.91	105	23.8 (3.0)
butylbenzene	316	10.32	1.12	91	15.3 (4.5)
naphthalene	238	16.47	0.62	128	-4.6 (2.1)
Average	-	-	-	-	6.7 (11.6)

Table C1: List of analytes in the test mixture with their actual concentration, retention times and %Reacted via Eq. (5.1). The "Quant m/z" column denotes the extracted ion m/z used to obtain peak signal areas. -[a]Compound not detected due to co-elution with the solvent front.

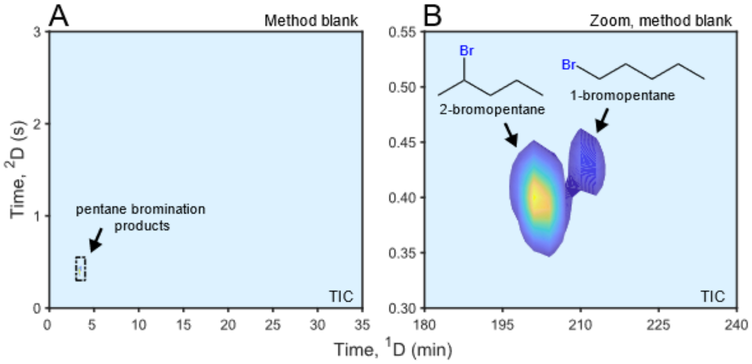


Figure C2: (A) Method blank obtained by reacting the sample solvent (HPLC grade pentane) with Br2 in the same manner as the test mixture and gasoline sample. (B) Zoom-in view. Only two low-intensity peaks were detected; the color scale (z-axis) is 100-fold lower than that in Figure 2 in the main paper. The more intense peak eluting first on 1D is tentatively identified as 2-bromopentane (Match Value: 863 to NIST). The second and less intense peak is tentatively identified as 1-bromopentane (Match Value: 850 to NIST). The higher ratio of 2-bromopentane to 1-bromopentane agrees with the proportions expected from radical-induced substitution of pentane (Figure C3), which is an anticipated minor side reaction of the method. The third anticipated product, 3-bromopentane (which is produced in the lowest proportion), was not observed but may have co-eluted with 2-bromopentane.

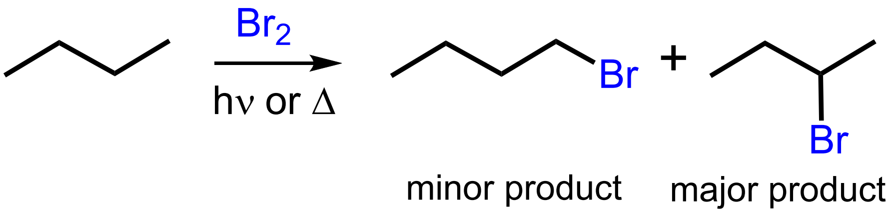


Figure C3: Reaction scheme for the bromination of alkanes, illustrated for butane. Bromination of alkanes is a free radical reaction, which would require significant Br2 homolysis for the present conditions (>192 kJ mol^{-}1). However, the reaction proceeds extremely slowly, which accounts for observation of only trace brominated pentane. The product proportions follow Markovnikov addition behavior where the major product is the most substituted species. The number of possible products increases as the length of the alkane chain increases.

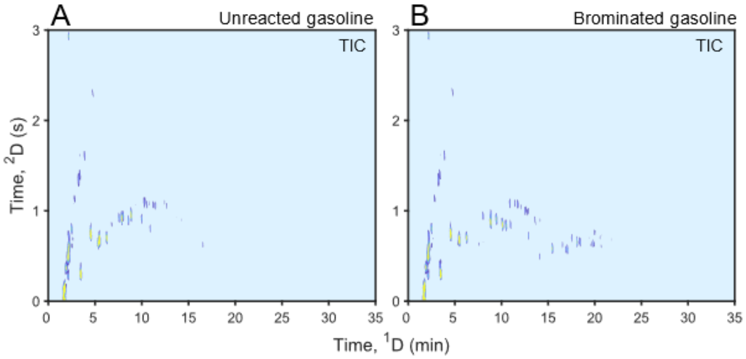


Figure C4: TIC chromatograms of the original and brominated gasoline. (A) Unreacted gasoline. (B) Brominated gasoline.

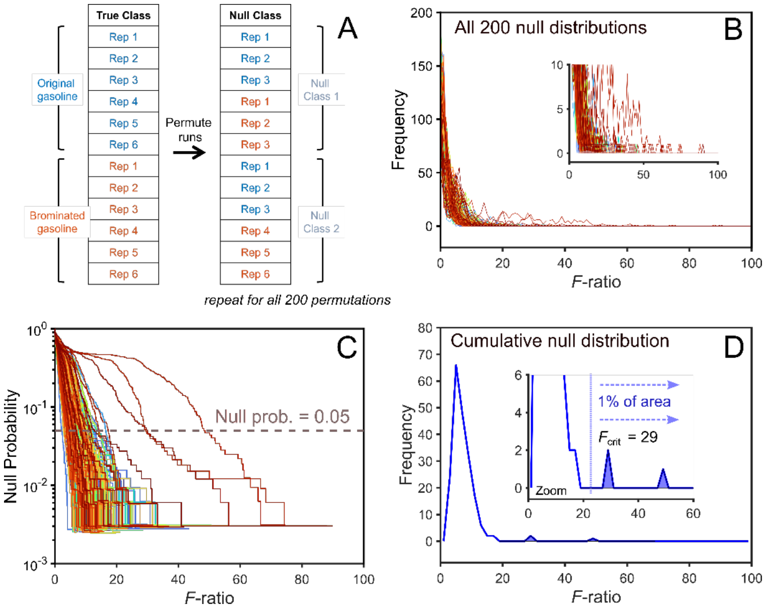


Figure C5: Illustration of combinatorial null distribution analysis (CNDA). (A) First, class labels are permuted to create "null classes". This is repeated for all possible permutations, which is 200 for a 6-sample versus 6 sample class comparison. Each null comparison is an estimation of the amount of non-class distinguishing variance due to random and systematic measurement errors. (B) Next, each null class is submitted to tile-based FRA and a hit list is generated. From the F-ratio of each hit in the hitlist a "null F-distribution" is created. (C) Each null distribution is transformed to a null probability curve, obtained by integration followed by determining the number of features above a given F-ratio value divided by the total features. A cumulative F-distribution can be constructed by selecting a single F-ratio from each null probability curve. The null probability thus acts as a surrogate for the commonly used false discovery rate (FDR) which is the probability of considering a false positive significant (type I error). However, the null probability output of CNDA is more accurate than, for instance, the venerable Benjamini-Hochberg correction which only considers Type I error as a result of the multiple hypothesis problem and does not consider the systematic and random variability of the dataset at hand. Here, the null probability was selected at 0.05 (5%) but could be made more stringent or liberal by shifting the null probability up or down. (D) Plotting all the F-ratios along the horizontal null probability line generates a final "cumulative" null F-distribution, where the analyst can visualize the range of F-ratios that could be selected as thresholds (i.e., F_{crit}). In the cumulative null F-distribution integration of 99% of the curve area provides 99% confidence that the null probability (i.e., FDR) of 5% will be achieved. For the present study, this value is $F_{\mathrm{crit}} = 29.9$.

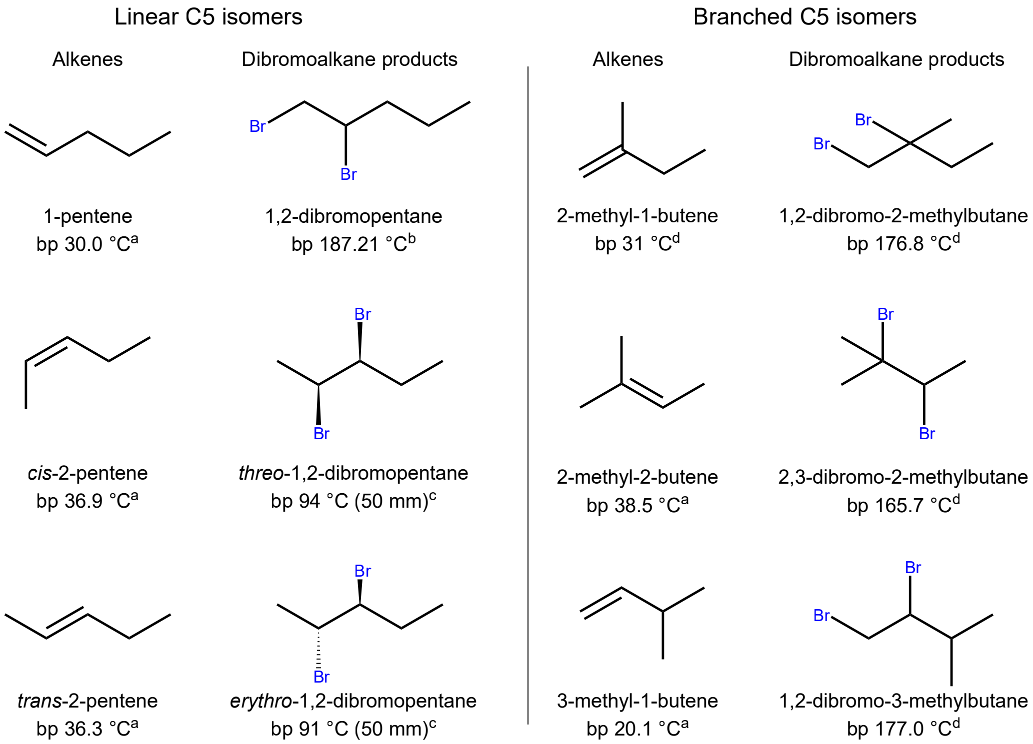


Figure C6: All possible linear and branched C5 alkene isomers and their expected reaction products. Boiling points (bp) were obtained from the superscripted references.

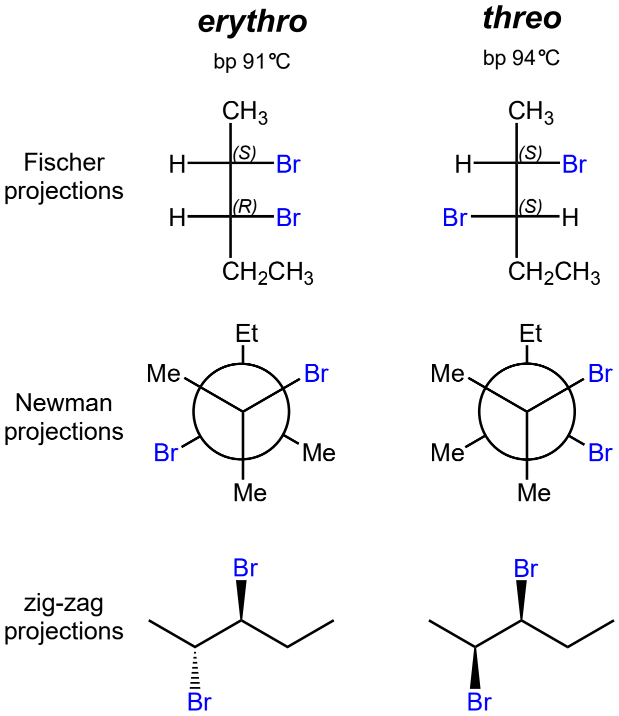


Figure C7: Fischer, Newman, and zig-zag projections of *erythro*- and *threo*-2,3-dibromopentane. Note that the *erythro* isomer encompasses the (S)(R) and (R)(S) enantiomers and the *threo* isomer includes both the (S)(S) and (R)(R) enantiomers. Boiling points (bp) are at 50 mm Hg.

Hit	Discovered compound	Match Value	^1t_R (s) standard/hit	^2t_R (s) standard/hit
1	1,2-dibromocylopentane	995 ± 1	861/861	0.72/0.72
2	2,3-dibromo-2-methylbutane	996 ± 2	615/618	0.87/0.87
4	*erythro*-2,3-dibromopentane	997 ± 0	693/693	0.84/0.85
5	*threo*-2,3-dibromopentane	997 ± 0	711/717	0.87/0.86
6	1,2-dibromopentane	996 ± 1	744/744	0.85/0.85
9	1,2-dibromo-2-methylbutane	998 ± 1	654/654	0.90/0.91

Table C2: Additional information for definitively identified hits. Match values are obtained by matching the mass spectrum of the brominated standard to the discovered hit (n=3). Columns four and five give comparisons of the retention times between the brominated standard compound and the given hit.

APPENDIX D

This Appendix is reproduced in part from the Electronic Supplementary Content of T.J. Trinklein and R.E. Synovec, Simulating comprehensive two-dimensional gas chromatography mass spectrometry data with realistic run-to-run shifting to evaluate the robustness of tile-based Fisher ratio analysis, J Chromatogr. A 2022, 1677, 463321.

D7 Computational equipment

Simulations were performed on a mid-level personal computer consisting of an Intel 6 Core i7-8700K processor (3.70 GHz), a 500 GB solid state hard drive, and 64 GB dual-channel DDR4 RAM. All simulated chromatograms were stored on an 8 TB optical drive. Simulations and tile-based FRA were performed in MATLAB 2021a (MathWorks, Inc., Natick, MA).

D8 Simulation methodology

First, a vector of chromatographic space was defined, of 640 s length with a 0.01 sampling density, corresponding to a 100 Hz collection frequency. For each SCP simulated, a 1D Gaussian peak was uniform randomly distributed in a 620 s separation window ($t_{\mathrm{sep}} = 20$ s to 620 s). To generate ^{2}D peaks, the area captured in each simulated modulation was calculated. The time at the beginning of the n^{th} sampling event is given by

$$^1t_{\mathrm{n}} = \left(n \times {}^1P_{\mathrm{M}}\right) + {}^1t_{\mathrm{st}} \tag{C24}$$

where n is the modulation number and $^1t_{\mathrm{st}}$ is the start time of the modulator. The ^{2}D peak area in the n^{th} modulation is represented by

$$^2a_{\mathrm{n}} = \int_{t_{\mathrm{n}-1}}^{t_{\mathrm{n}}} {}^1C(t)\mathrm{d}t = \int_{t_{\mathrm{n}-1}}^{t_{\mathrm{n}}} \frac{4 \cdot A_{\mathrm{SCP}}}{{}^1w_{\mathrm{b}}\sqrt{2\pi}} \exp\left(\frac{8(t - t_{\mathrm{R}})^2}{{}^1w_{\mathrm{b}}^2}\mathrm{d}t\right) \tag{C25}$$

where A_{SCP} is the SCP's 1D area, which is a random variate drawn from an exponential distribution (Fig. 6.2A). Equation (D2) is computed numerically via the trapezoid rule for each modulation, generating a vector of areas

$$^2\boldsymbol{A} = \{{}^2A_1, {}^2A_2 \ldots {}^2A_{\mathrm{n}}\} \tag{C26}$$

Then, an n-term Gaussian function is generated where the area of the n^{th} 2D peak is the n^{th} entry from $^2\boldsymbol{A}$

$$^2C_{\mathrm{n}}(t) = \frac{4 \cdot {}^2A_{\mathrm{SCP}}}{{}^2w_{\mathrm{b}}\sqrt{2\pi}} \exp\left(\frac{8({}^2t - {}^2t_{\mathrm{R}})^2}{{}^2w_{\mathrm{b}}^2}\mathrm{d}t\right) \tag{C27}$$

Note that this method preserves the area of the ^{1}D peak, as in total-transfer modulation, viz.

$$\sum_{i}^{i=n} {}^2C_{\mathrm{n}}(t) = {}^1C(t) \tag{C28}$$

Once the peaks were modeled, a randomly selected mass spectrum was multiplied element-wise (i.e., outer product) across each peak to form a series of Gaussian peaks at each m/z channel. This process is repeated for all m components, and the noiseless SCPs are summed together to generate a chromatogram. Finally, white Gaussian noise is added to each m/z channel.

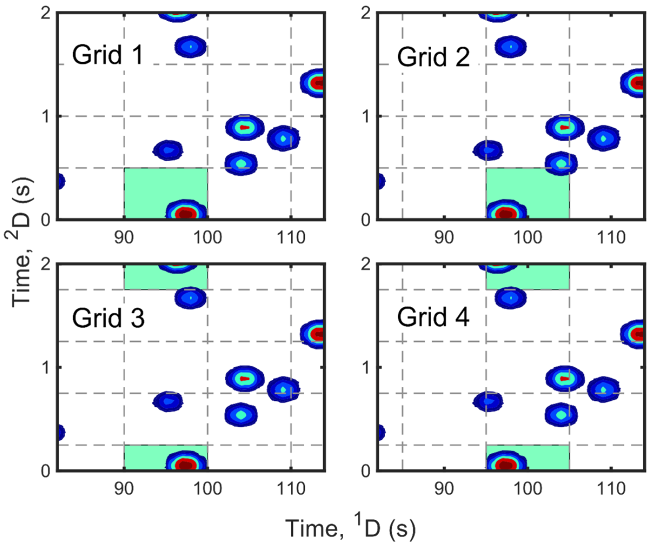
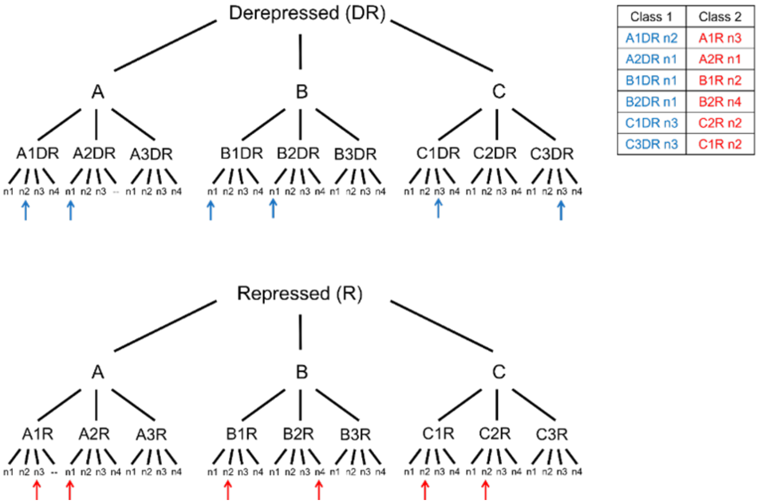


Figure D1: Illustration of how tile-based FRA addresses wraparound peaks. The tiling procedure is distinct from binning in that four overlapping grids are used by shifting grids on 1D and/or 2D by half of a tile width. In this example, one analyte peak is wrapping around and "split" between modulations. The tile closest to this analyte is highlighted in green. In Grid 1, which is anchored in the origin of the 2D chromatogram (no tiles shifted), the tile does not fully capture the analyte. The same is true for Grid 2, which is only shifted along 1D. However, the tiles in Grids 3 and 4 are shifted along 2D. As a result, the tile itself "wraps around" such that the tile in Grids 3 and 4 integrates signal at both the bottom and top of the chromatogram. In this example, Grid 3 captures all of the analyte signal and is the optimum tile.

D9 Yeast dataset experimental details

To generate shift functions from a "real" dataset, the retention times of sixteen analytes from a previously obtained GC×GC -TOFMS dataset of respiring, i.e., glucose DeRepressed (DR) and fermenting, i.e., glucose Repressed yeast (R) were used. More experimental details can be found in the original reports [1-2]. The sample preparation workflow is outlined in Fig. D2. Briefly, cells were cultured in a medium containing either 5% glucose, promoting fermentation (Repressed, R) or 3% ethanol and 0.05% glucose to promote respiration (Derepressed, DR). Three cultures of R and DR cells were grown to obtain a measure of biological variability, denoted A, B and C. Then, from each culture, three extracts were obtained, denoted (e.g., for culture A) A1R, A2R, A3R for the R cell cultures and A1DR, A2DR, and A3DR for the DR cell cultures. Four replicate GC×GC-TOFMS runs were acquired for each sample (denoted, e.g., A1R n1, A1R n2, etc.) except A1R and A2DR which were obtained in triplicate resulting in a total of seventy chromatograms. In order to create a 6 versus 6 sample class comparison, a managed random selection was used [3]. The selection was constrained to contain two runs from each culture, one from each extract. The twelve samples randomly selected are shown in Fig D2.

(1) R.E. Mohler et al., *Anal. Chem.* **2006**, *78*, 2700-2709.
(2) R.E. Mohler, et al., *Analyst* **2007**, *132*, 756–767.
(3) N.E. Watson et al., *J. Chromatogr. A* **2006**, *1459*, 101-111.



Class 1	Class 2
A1DR n2	A1R n3
A2DR n1	A2R n1
B1DR n1	B1R n2
B2DR n1	B2R n4
C1DR n3	C2R n2
C3DR n3	C1R n2

Figure D2: Experimental scheme and nomenclature of the yeast dataset samples. The table shows the twelve samples that were selected for analysis.

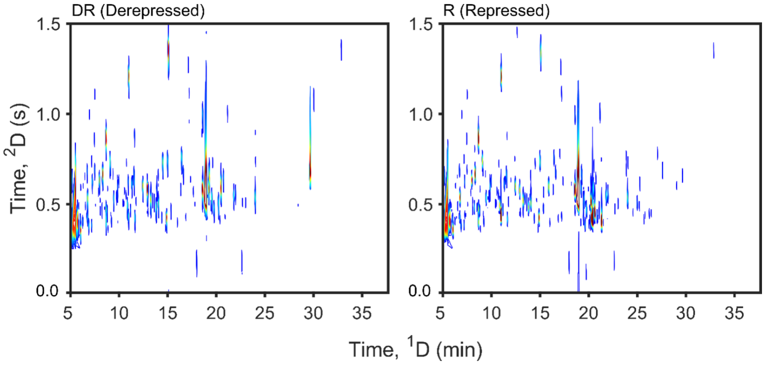


Figure D3: GC×GC chromatograms of the two classes from the yeast data set. Average GC×GC chromatograms (m/z 73) of the (DR) derepressed and (R) repressed yeast samples. A high level of chromatographic complexity ($\alpha_{\mathrm{e,2D}} \approx 0.1$ to 0.5) is obtained.

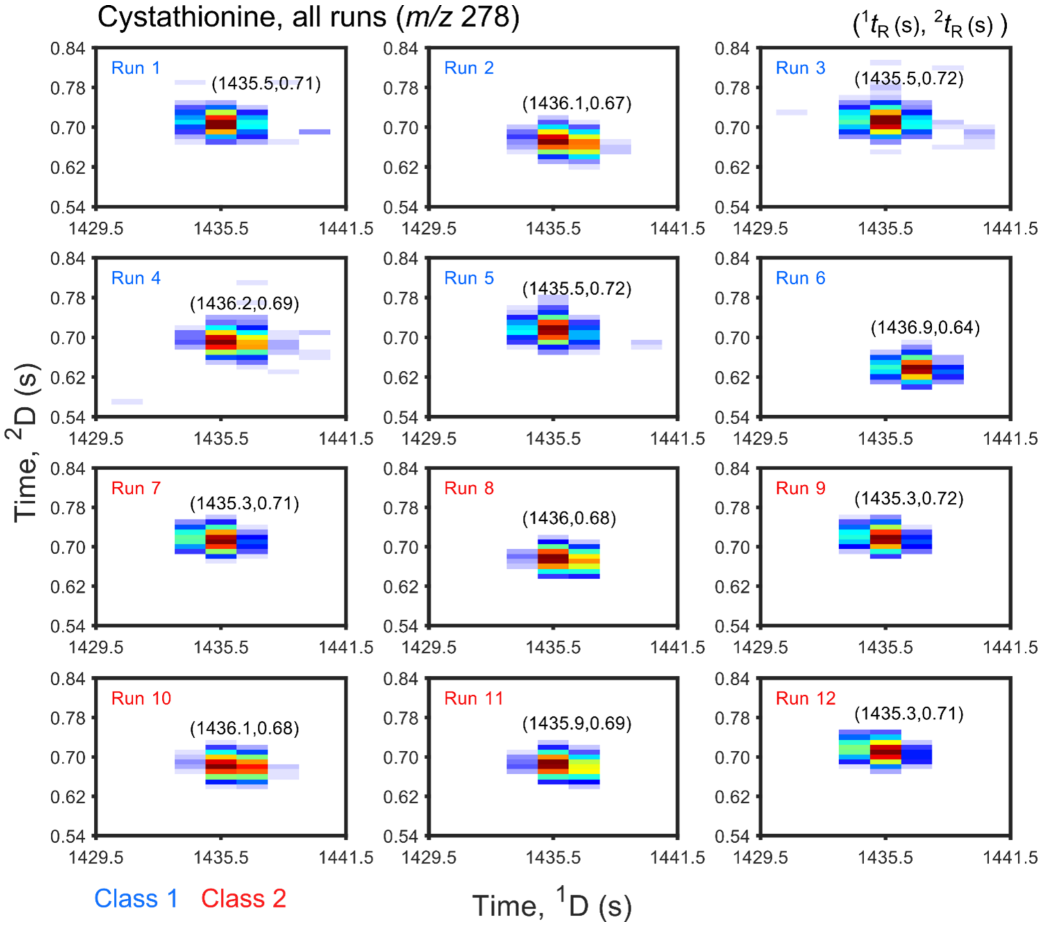


Figure D4: Illustration of 2D shifting for cystathionine from the yeast dataset. The $1t_{\mathrm{R}}$ and $2t_{\mathrm{R}}$ are measured for each run. First, a peak finder is used to find the rough location of the peak maximum on both dimensions. Then, to obtain a more accurate measure of $^1t_{\mathrm{R}}$, a Gaussian profile is fit to the envelope of the peaks in each dimension and obtained as the maximum of the fitted Gaussian function. The $^2t_{\mathrm{R}}$ is obtained as the average of the 2D peak maxima. The entire set of retention times from the twelve runs is used to compute the average $^1t_{\mathrm{R}}$ and $^2t_{\mathrm{R}}$ along with $^1\delta_{\mathrm{r}}$ and $^2\delta_{\mathrm{r}}$, via Eq. (6.6), using an average $^1w_{\mathrm{b}} = 4.5$ s and $^2w_{\mathrm{b}} = 80$ ms as given in Table D2.

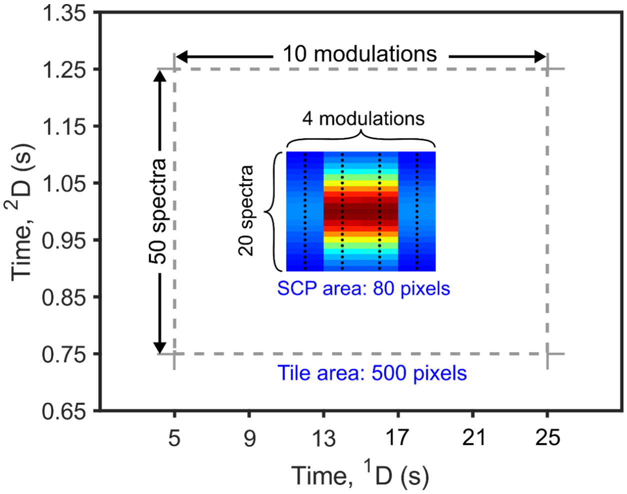


Figure D5: Illustration of SCP area, Tile area, and RTA. An SCP sampled out-of-phase (80 pixels) is illustrated within a tile area of 500 pixels (10 modulations by 50 spectra). This yields an RTA of 6.25 via Eq. (6.8).

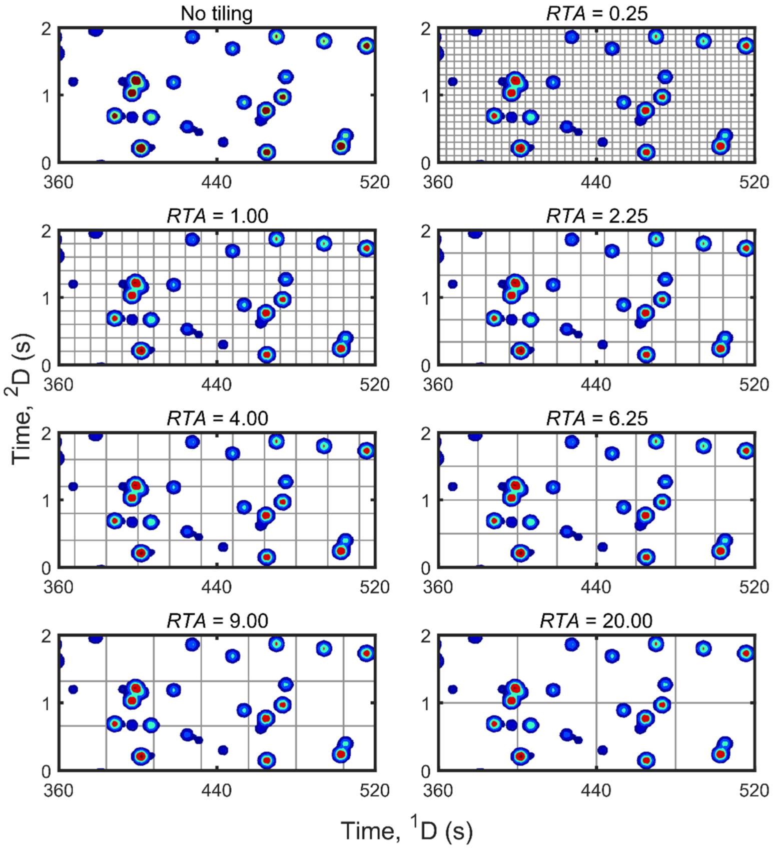


Figure D6: Tile grids for the seven tile sizes (RTA) tested. The grids are drawn over a representative separation at $\alpha_{\mathrm{e,2D}} = 0.1$. For $RTA = 2.25$ and 9.00, the number of ^{2}D spectra does not divide evenly into the P_{M}. For these RTA, the tile-based FRA software linearly interpolates the number of spectra (i.e., data points) per each P_{M} to a number of spectra that divide evenly into the 2D tile dimension.

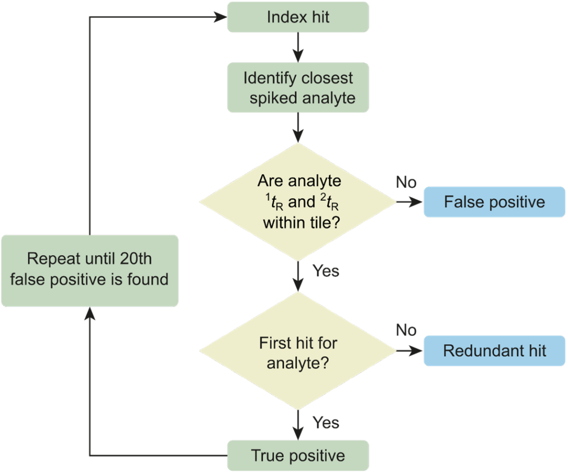


Figure D7: Flowchart of the method to classify hits. The flow chart illustrates automated hit list evaluation for tile-based FRA. First, a given hit was matched to the spiked analyte closest to the hit tile. If the analyte's retention time was within the tile area, the hit was considered a potential true positive. If not, the hit was recorded as a false positive. If the hit was a potential true positive, the hit was checked for prior appearance on the hit list, in which case the hit was recorded as a "redundant hit" and is removed from the hit list. If the hit had not appeared prior, the hit was recorded as a true positive. This process was iterated until all of the known analytes were discovered. If fewer than the known number of analytes was discovered, the process was iterated until the 20$^{\text{th}}$ false positive.