

**Automated Control and Data Analysis for
Multicellular Multidetector Microscale Gas Chromatographs**

by

Qu Xu

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Engineering
(Manufacturing)
in the University of Michigan
2024

Doctoral Committee:

Professor Yogesh B. Gianchandani, Co-Chair
Dr. Yutao Qin, Co-Chair
Professor Judy Jin
Professor Khalil Najafi
Professor Robert Bordley

Qu Xu

quxu@umich.edu

ORCID iD: 0009-0006-6851-3293

© Qu Xu 2023

To my loving family

Acknowledgements

I would like to express my deepest gratitude to my research advisor, Professor Gianchandani, who gave me the opportunity for me to become a DEng student under his guidance. His leadership, research integrity, vision, enthusiasm, and advice have all been invaluable throughout my DEng studies. I would also like to thank my mentor, Dr. Yutao Qin who has provided training, inspiration, and advice. His patience, guidance, diligence, and mentorship have all been indispensable throughout my graduate studies. He has always aided me through the most difficult parts of my degree. I would like to thank my committee members, Professor Judy Jin, Professor Khalil Najafi and Professor Robert Bordley for their valuable feedback, insight, and guidance.

I would like to thank everyone who has worked with me and aided me greatly in my research. Dr. Xiangyu (Daniel) Zhao provided the PCB designs for my control software interfaces; peak detection algorithm, the output of which is used in my recognition algorithm; and for collaborating in software development. Declan Winship for testing my software and providing feedback. Dr. Weilin Liao for designing the microfluidic chip which my software provides control for. The master's and undergraduate students who aided my research Yingkun and Tao. I would also like to thank other research group members and office mates, including Alex, Zeyu, Hanfei, Adrienne, Partha, Sajal, Neeharika, Behnoush, Ester, and Farzad.

I would also like to thank the ISD department, Prof. Judy Jin, Prof. Diann Brei, and Professor Mihaela Banu who gave me the opportunity to become a DEng student in ISD, especially during the difficult COVID-19 pandemic.

Finally, I would like to express my sincerest gratitude to my family. My Girlfriend Esther Friedrich who has offered immense emotional support in the most difficult moments of my DEng study. My parents, who have loved and supported me through my entire life. My grandpa, who inspired me to become an engineer.

Table of Contents

Acknowledgements	iii
List of Tables	vii
List of Figures	viii
List of Appendices	xi
Abstract	xii
Chapter 1 Introduction	1
1.1 Gas Phase Analysis and MPCA	1
1.2 Electronic Interfaces and Control for Microscale Gas Chromatography	5
1.3 Recognition of Chemicals	9
1.4 Motivation for This Work	13
1.5 Organization of This Thesis	15
References	17
Chapter 2 Control Software Design for a Multisensing Multicellular Microscale Gas Chromatography System	21
2.1 Introduction	22
2.2 Design and Implementation	25
2.2.1 MPCA System Overview	25
2.2.2 Electronic Interfaces	27
2.2.3 Control Software Implementation	34
2.3 Experimental Validation	54
2.4 Discussion and Conclusions	61

References.....	63
Chapter 3 A Rule-Based Automated Chemical Recognition Algorithm for a Multi-Cell Multi-Detector Micro Gas Chromatograph.....	66
3.1 Introduction.....	67
3.2 Operating Principle and Method	70
3.2.1 Overall Algorithm flow.....	70
3.2.2 Chemical Library and recognition parameters	72
3.2.3 Recognition by retention time and cell number	75
3.2.4 Recognition by DRP	78
3.2.5 Total score for recognition	84
3.2.6 Estimated concentration.	85
3.2.7 Using a reference chemical	85
3.3 Results and Discussion	86
3.4 Conclusions.....	99
References.....	101
Chapter 4 Conclusions and Future Work	104
4.1 Conclusions.....	104
4.2 Future Work	109
Appendix A: Other Publications and Manuscripts	112
A.1 Highly Integrated μ GC Based on a Multisensing Progressive Cellular Architecture with a Valveless Sample Inlet.....	112
A2. An Enhanced-performance Multisensing Progressive Cellular μ GC: Design Advances and Blind Test Results.....	112
References.....	113

List of Tables

Table 2.1: The IC model and conversion time used for each type of readout.....	39
Table 2.2: Normal mode and low power mode power consumption of the RPi (Model 3B+ in this work).....	46
Table 2.3 Actual time interval between measurements (TIBM) of all the readout component types (measured at the peak software workload step, <i>i.e.</i> , the Cell2 separation step).	59
Table 2.4 Noise data of the detectors.....	60
Table 3.1: Example of the basic chemical library that contains the chemical properties and MPCA system response characteristics.....	73
Table 3.2: Example of the expanded chemical library incorporating additional parameters for use by the algorithm.	74
Table 3.3: Recognition results for the first example. The labels on the left indicate comparison to the ground truth, in which the bold italic text indicates chemicals that are in the library. This section describes three representative examples of the chemical recognition results. In each example, the recognition results are compared against the true list of chemicals, <i>i.e.</i> , the ground truth, on a peak-by-peak basis.	89
Table 3.4: Recognition results for the second example. The labels on the left indicate comparison to the ground truth, in which the bold italic text indicates chemicals that are in the library.	92
Table 3.5: Recognition results of the third example. The labels on the left indicate comparison to the ground truth. The columns in light blue background are results with the use of o-xylene as the reference chemical. For this example, other columns resulting from the use of the reference chemical are identical as those without using a reference chemical and therefore not repeated here.	97
Table 3.6: Confusion matrix based on the recognition results.	98

List of Figures

Figure 1.1: Diagram of a basic gas chromatograph.	2
Figure 1.2: Diagram of MPCA	4
Figure 1.3: Conceptual diagram of SPI communication layout with multiple peripherals [Gru23].	6
Figure 1.4: Illustration of I ² C communication layout with multiple peripherals.	7
Figure 1.5: Illustration of an artificial neural network.	11
Figure 1.6: Diagram of the architecture of an expert system	12
Figure 2.1: MPCA fluidic architecture and photo of the chip.	26
Figure 2.2: Block diagram of the electronic interface between components.	28
Figure 2.3: Circuit schematics of μ GC system hardware. (a) PWM heater control circuit. (b) Analog voltage heater control circuit. (c) Thermistor readout circuit. (d) Sampling pump control circuit. (e) Separation pump control circuit. (f) Pressure sensor readout circuit. (g) Valve control circuit. (h) Capacitive detector readout circuit. (i) Photoionization detector readout circuit.	32
Figure 2.4: Illustration of connections between multiple CapDets and CDCs for the cases of (a) fewest CDC units needed, and (b) this work with a level of design redundancy.	33
Figure 2.5: Schematic illustration of the power management circuit.	34
Figure 2.6: Suitable use cases for multithreading and multiprocessing.	36
Figure 2.7: Flow charts of the main thread Th0.	37
Figure 2.8: Flow charts of the temperature control thread.	38
Figure 2.9: Flow charts of the capacitive detector readout thread.	39
Figure 2.10: Flow charts of the valve control thread.	40
Figure 2.11: Flow charts of the pump control thread.	41
Figure 2.12: Flow charts of the AiPD readout thread.	42

Figure 2.13: Graphical representation of the main UI window with real time plotting of data. ...	47
Figure 2.14: Graphical representation of UI for customizing the operation method. The entries in blue are representative values filled by a user for the Cell2 separation step.....	48
Figure 2.15: Main page of the UI. The obsolete or unrelated contents are covered by black boxes.....	50
Figure 2.16: Flowchart of user interface operations.	52
Figure 2.17: Method configuration page for editing the operation parameters of a run step. (The obsolete or unrelated contents are covered by black boxes)	53
Figure 2.18: Main window during operation with real-time plotting of the readouts.	53
Figure 2.19: Typical temperature and flow control results in the MPCA system during (a) the Cell2 separation step and (b) the Cell3 separation step.	56
Figure 2.20: Chromatogram of a mixture containing cyclohexane, pentanal, heptane, pinacolyl alcohol, toluene, methyl isobutyl ketone, butyl acetate, ethylbenzene, m-xylene, o-xylene, 1-chloroheptane, nitrobenzene, nonane, mesitylene, decane, 2-nonanone, undecane and dodecane. Currently, the chromatograms are only presented for Cell2 and Cell3, as Cell1 is still under hardware development.	58
Figure 2.21: The typical experimentally measured time interval between measurement (TIBM) for each type of readout over time (measured during the MPCA Cell2 separation step). The number of data points in each plot is greater than 1200.....	59
Figure 3.1: Overall chemical recognition algorithm flow for each peak in a chromatogram.	72
Figure 3.2: Likelihood scoring function for retention time.....	76
Figure 3.3: Experimental data points on retention times and peak heights and fitted curves for DMMP in Cell2 and DEMP and DIMP in Cell3.	78
Figure 3.4: Flow diagram of DRP match scoring when two detector peak heights are below their peak height thresholds. $H_{th,i}$ is the peak height threshold for detector i , $H_{d,i}$ is the detected peak height of the detector i , $H_{p,i,j}$ is the peak height of detector i projected from detector j , and S_{ij} is the score that corresponds to the response ratio between detector i and detector j	82
Figure 3.5: Flow diagram of DRP scoring when one detector peak height is below its peak height thresholds. $H_{th,i}$ is the peak height threshold for detector i , $H_{d,i}$ is the detected peak height of the detector i , $H_{p,i,j}$ is the peak height of detector i projected from detector j , and S_{ij} is the score that corresponds to the response ratio between detector i and detector j	83
Figure 3.6: Raw chromatograms and detected peaks of the first example. For the peaks with known identities, the identities are labeled. Known peaks present in the library are labeled in bold italic.	88

Figure 3.7: Raw chromatograms and detected peaks of the second example. For the peaks with known identities, the identities are labeled. Known peaks present in the library are labeled in bold italic.91

Figure 3.8: Raw chromatograms and detected peaks of the third example. For the peaks with known identities, the identities are labeled. Known peaks present in the library are labeled in bold italic.95

List of Appendices

Appendix A: Other Publications.....	112
-------------------------------------	-----

Abstract

Industrial manufacturing processes often release volatile organic compounds (VOCs). VOCs can be harmful to humans and the environment but can also serve as markers for process control. The timely detection and recognition of VOCs are, therefore, important. Gas chromatography (GC) systems are the current gold standard for VOC detection, but are limited by their high bulk, cost, and power consumption. Microscale gas chromatographs (μ GCs) are miniaturized GC instruments that can be distributed and deployed more easily, thus show greater promise for automated *in situ* analysis. As μ GCs become increasingly sophisticated, their controls and automated recognition become increasingly complex as well. Current μ GC systems often require trained personnel to operate. Automation of controls and VOC recognition is, therefore, necessary to achieve fully autonomous and automated *in situ* analysis.

This work investigates the challenges of automating the controls, and recognition of VOCs in highly complex μ GC systems. The representative μ GC system is based on a multicellular progressive cellular architecture that uses multiple cells to broaden the range of analytes. The multiple detectors within each cell further enhance chemical recognition. The control software manages data acquisition of the μ GC system in a time sensitive manner, while operating multiple control loops and error conditions. The multithreaded control software enables concurrent control of heaters, pumps, and valves, while also gathering data from thermistors, pressure sensors, capacitive detectors, and photoionization detectors. A graphical user interface (UI), implemented on a laptop computer, provides remote control and real time data visualization. In experimental evaluations, the control software provided successful automation of all the components, including 8 sets of thermistors and heaters for temperature feedback loops, 2 sets of pressure sensors and pumps that form pressure head feedback loops, 6

capacitive detectors, 3 photoionization detectors, 6 valves, and a fixed-flow gas pump. A typical run analyzing 18 chemicals is presented. Despite the use of a non-real-time operating system, the standard deviations of the control loop timings were <0.5% of the intended time interval between measurements. The control software successfully supported >1000 μ GC runs that analyzed various mixtures.

A chemical recognition algorithm can be a valuable part of any autonomous μ GC system. For a multi-detector μ GC system, the chemical analysis must account for the retention time of each chemical analyte as well as the relative response of each detector to each analyte, i.e., the detector response pattern (DRP). This work reports a rule-based automated chemical recognition algorithm for a multi-cell, multi-detector μ GC system. The algorithm applies rules based on expert knowledge to compare the detected peaks. Consequently, this algorithm only requires a small amount of calibration data but not extensive training data. Additionally, the algorithm provides special subroutines for chromatogram peaks with a small signal-to-noise ratio and for asymmetrical peaks that may result from surface adsorptive analytes. The algorithm was verified by multiple experimental tests. Each test included chromatograms with 21–31 peaks for each detector. The true positive rate was 96.3%, the true negative rate was 94.1%, the false positive rate was 5.9%, and the false negative rate was 3.7%. The results demonstrated that the algorithm could support μ GC systems for automated chemical screening and early warning applications.

Chapter 1 Introduction

Across many manufacturing sectors, gases are generated. The detection of these gases is important for personal, environmental, and industrial health monitoring. The instruments for gas phase analysis, and data analysis are therefore important. Gas chromatography is a commonly used technique for gas phase analysis. Advancements in microelectromechanical systems enable microscale gas chromatographs (μ GCs) that have become increasingly complex. This work investigates the automation and data analysis of highly complex microscale gas chromatographs using a representative example as a case study. This chapter describes the context and thesis goals. It is divided into five sections. Section 1 introduces gas phase analysis and the representative microsystem. Section 2 introduces the electronic interfaces and control for this representative μ GC. Section 3 introduces the commonly used data analysis techniques for μ GCs. Section 4 discusses the motivation of this work. Section 5 discusses the organization of the thesis.

1.1 Gas Phase Analysis and MPCA

Gases are often generated in many manufacturing industries and sectors. A notable group released are volatile organic compounds (VOCs) [Zhe09, Wan23, Cao16]. At high enough concentrations, certain VOCs can cause a host of health-related issues, ranging from irritation to organ damage and cancer [EPA23]. VOCs are not only damaging to humans but can also cause adverse ecological effects [Dav21]. VOCs can also serve as markers for industrial processes [Kan23, Dre19, Cal23]. The detection of VOCs is therefore important. However, traditional

benchtop GCs are often expensive, bulky, and require training to operate, making them suitable for laboratory analysis. As VOCs can travel great distances [Far05] and are hazardous, mobile, autonomous, and low-cost alternatives are needed for widespread testing.

The current gold standard for VOC monitoring is the use of a benchtop gas chromatograph (GC) systems, as approved by government agencies such as the EPA [EPA99]. Benchtop GCs are often operated in a laboratory setting. In a benchtop GC, the gas sample is first collected in the field. The gas sample is injected into the inlet of the GC. The gas sample then flows through the separation column and is separated based on the volatility of the chemicals [Reg18]. Temperature of the columns play a key role in the separation quality and repeatability [McN19a], as does the flow rate [McN19b]. A downstream detector detects the eluting peaks. The temporal delay, or the retention time of the peaks is a key feature of the chemicals that can be used for qualitative analysis of the sample. The peak area is a key feature for quantitative analysis of the sample.

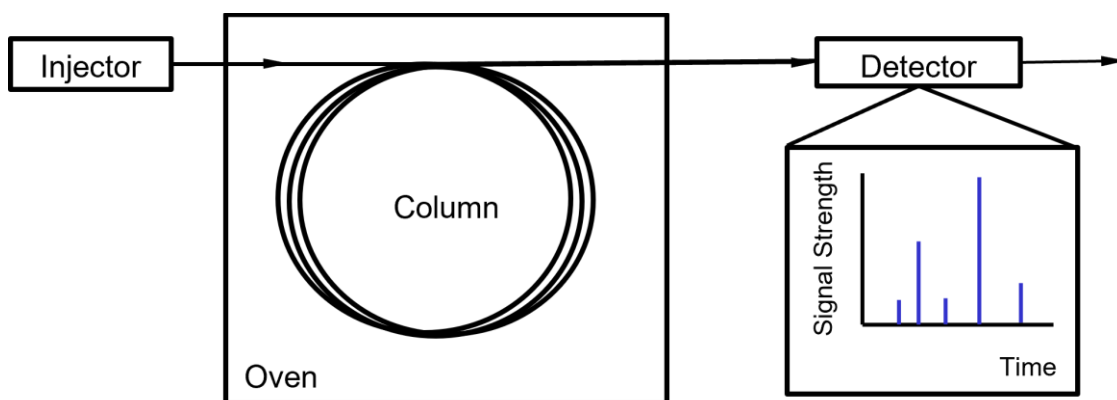


Figure 1.1: Diagram of a basic gas chromatograph.

Although retention time is useful for the qualitative analysis of chemicals, it is not always sufficient to speciate chemicals as some chemicals have indistinguishable retention times [Gro04]. Additional techniques are often used in conjunction with GC to improve differentiation of chemicals. A common technique used with GC is mass spectroscopy (MS) [Spa11]. MS first ionizes the samples, and the different composition of chemicals results in different mass and charge of the ions. The different mass to charge ratios of the ions can then be used to differentiate the samples [Bra23]. Another technique is to add another GC column to the existing GC (GCxGC) [Ama19]. In GCxGC, a modulator takes small samples from the first column periodically and injects the small samples into another column. Two-dimensional separation can be achieved by this method. Using multiple detectors [Dąb18] that sense different properties of the chemicals is another technique to improve differentiability.

Microelectromechanical systems (MEMS) technology has enabled the miniaturization of the aforementioned gas phase analysis instruments. However, the increased mobility of these miniaturized instruments often comes at the tradeoff of performance. Ever sophisticated μ GCs have been developed to increase performance, and the progress of the μ GC hardware has been well documented. In contrast, reports on the control software, an area vital to the automation of μ GC for autonomous *in situ* analysis, have been rare.

Researchers at the Center for Wireless Integrated MicroSensing and Systems (WIMS²) have developed μ GCs with ever increasing performance [Qin16, Wan19, Lia23, Zho16, Hua22]. One of such designs is the highly complex μ GC based on a Multisensing Progressive Cellular Architecture (MPCA) [Lia23]. The MPCA features three separate cells, each containing a preconcentrator, a separation column, two capacitive detectors (CapDetA and CapDetB), and one photoionization detector (AiPD). Preconcentrators are used to decrease the detection limit for μ GC. The MPCA also incorporates a mini motor driven pump, two piezoelectric pumps, two

differential pressure sensors, and six valves for sample flow control and eight thermistor heater pairs for temperature control. The microfluidic components and the electronics are on separate circuit boards, so temperature of the microfluidic components has minimal effect on the electronics.

The first step to operating MPCA is usually the sampling step. In this step, the valves are arranged so the sample gas is carried into the preconcentrators. The more upstream the preconcentrator, the less volatile the chemical it absorbs. The sample is pre-separated based on the preconception. Then, for each cell, the valves are arranged so the sample absorbed in the preconcentrator can only flow to the column and detector of this cell. The preconcentrator is heated to release the sample. The sample is then separated in the column, where the coating of the column, the flow in the column and the temperature of the column all play key roles in the separation. For a given column, the more repeatable the control of flow and temperature, the more repeatable the separation. Finally, the separated sample passes the three detectors. The ensemble of the chromatograms of the three detectors are used for chemical recognition.

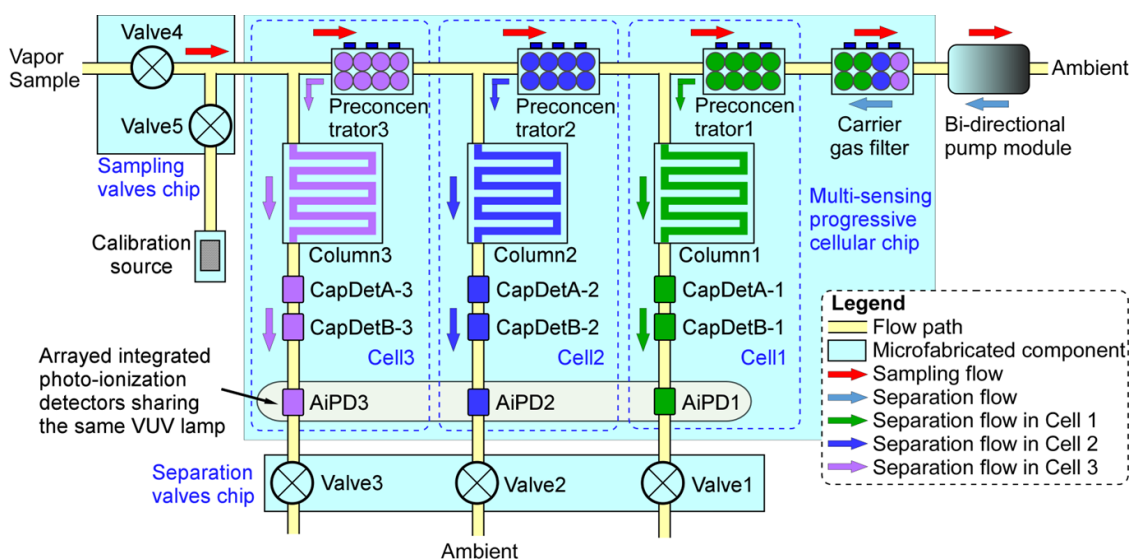


Figure 1.2: Diagram of MPCA

1.2 Electronic Interfaces and Control for Microscale Gas Chromatography

The various components of a μ GC are often connected to various electronic interfaces, mounted on an electronics board. A control unit usually schedules data acquisition and component control. The control unit is often a microcontroller [Gar15], a single board embedded computer [Wan19], or a data acquisition card (DAQ) [Zho16]. Traditionally, microprocessors and microcontrollers differ in the microprocessors are more suitable for computational needs and a microcontroller focuses on embedded uses and integration with peripherals [Ban98]. Nowadays, the distinction between the two terms is increasingly blurred with some embedded computers also supporting a wide array of peripherals available at low cost on the market. However, in general, microprocessors tend to possess stronger computational power and have on board data storage, while microcontrollers tend to focus on specific applications and need external data storage. Microprocessors also tend to use more energy than microcontrollers. Traditionally, DAQ cards needed an external computer, though more recent standalone DAQs incorporate a microcontroller as well. However, DAQ cards are often proprietary and less cost effective and are not designed for more intensive computational needs.

Some control units have built in analog to digital converters to read voltages directly [Mbe23]. However, the resolution and signal range of these units are predetermined and may not be suitable for the signals from a μ GC. Analog to digital converters (ADC) and capacitance to digital converters are often used to tailor the measurement resolution and range. The converted data needs to be read out from a register to the controller unit.

The most common means of communication between a converter and a control unit are serial peripheral interface (SPI) protocol, Inter-Integrated Circuit (I^2C) protocol and universal asynchronous receiver/transmitter (UART) protocol. SPI is a four-wire communication protocol [Gru23]. A clock signal (SCK) is sent from the controller to the peripheral (here the ADC) on one

wire. The peripheral in controller out (PICO) wire is dedicated to sending the controller command to the peripheral. The peripheral out controller wire in is dedicated to sending the peripheral signal to the controller. The chip select wire sends the signal from the controller to the peripheral indicating if the peripheral should be active for communication. The SPI protocol is stable but requires increasing more wire for chip select as the number of peripherals grows, making it difficult to scale as control units have a limited number of control pins.

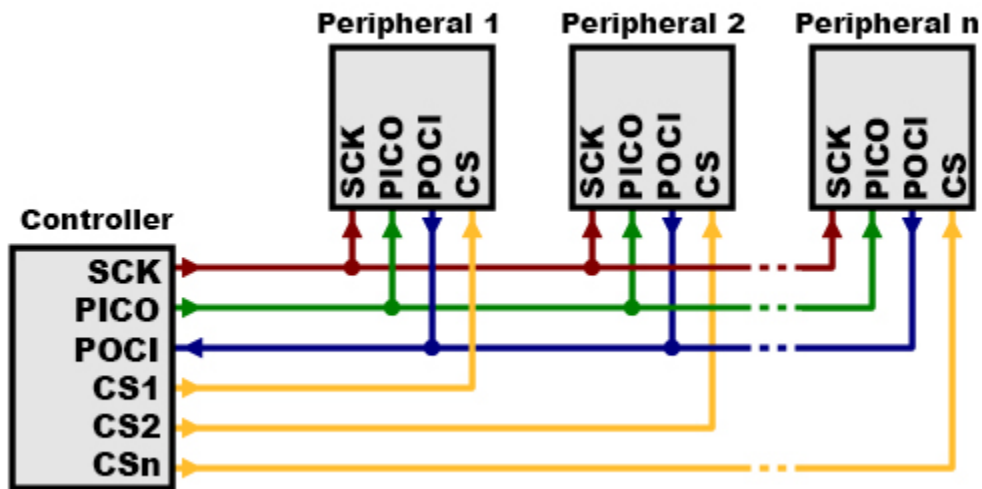


Figure 1.3: Conceptual diagram of SPI communication layout with multiple peripherals [Gru23].

I²C utilizes only two wires, serial clock (SCL) and serial data (SDA) [Val15]. The SCL wire provides the clock signal for both controller and peripheral. To select a specific peripheral to read from, the controller sends a data address to all peripherals on SDA wire. The peripheral with the correct address will respond and data communications between the controller and the peripheral may take place on the SDA wire. Due to the number of shared lines, I²C may be less stable than SPI, but is much more scalable. The practical limit to scalability is the number of addresses available for use. UART is an universal two-wire communication protocol but cannot be scaled up for a controller to read from multiple peripherals [Pen23].

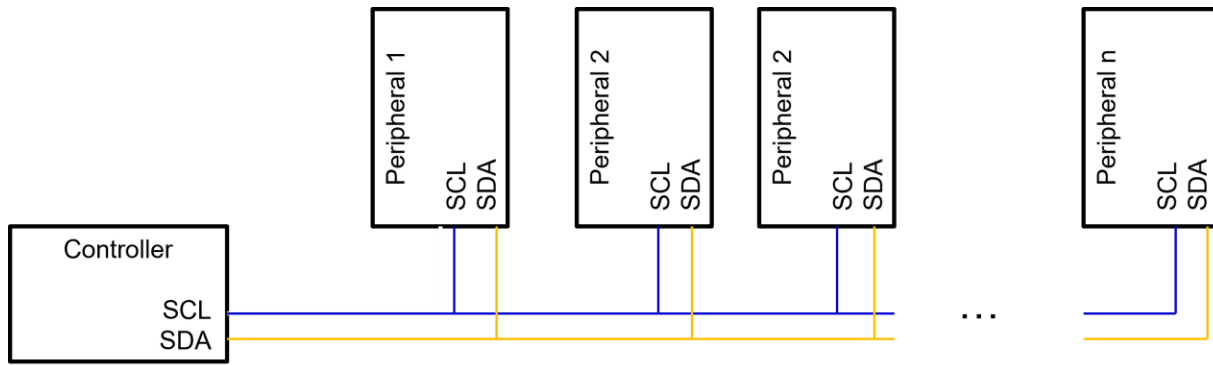


Figure 1.4: Illustration of I²C communication layout with multiple peripherals.

To actuate the components on the electronics board, the controller generally has two options, control using a General-Purpose Input Output (GPIO) pin and to control an external device using a standard communication protocol. GPIO pins generally could supply a binary signal of either 0 or 1. A pulsed signal can be generated with GPIO pins. Pulse Width Modulation (PWM) controls the frequency and duty signal of the pulses. Through the control of the duty cycle, the average signal strength can be controlled as well. There are two types of PWM, software PWM, and hardware PWM [Pig23]. Software PWM uses software timing to keep track of the pulse frequency while hardware PWM uses a hardware clock to track the pulse frequency. As such, hardware PWM offers far finer control of frequency and can be utilized for controls that require much higher frequencies than software PWM. Due to the switching nature of PWM, flicker noise is often generated and can affect detector readout. A circuit that controls signals using a digital potentiometer can be used in lieu of PWM to reduce flicker noise. However, the digital potentiometer must use one of the standard protocols, adding complexity to the lines of communication.

Proportional, integral, differential controllers (three term controllers) are used as a general purpose closed loop controller. As the control loop is not based on an existing physical system, the control parameters must be fine tuned using trial and error. The discrete time control loop can be expressed as:

$$C = C_{\max} (P \cdot \Delta x_{curr} + I \cdot \Delta x_{accu} + D(\Delta x_{curr} - \Delta x_{prev})) \quad \dots\dots(1)$$

where C is the control signal strength; C_{\max} is the maximum available signal strength; P , I , D are the proportional, integral, and derivative coefficients, respectively; Δx_{curr} is the difference between the target value of x and the measured value of x , both at the current (*i.e.*, most recent) measurement time; Δx_{accu} is the accumulated difference between the target value of x and the measured value of x since the beginning of each operation step; Δx_{prev} is the difference between the target value of x and measured value of x , both at the preceding (*i.e.*, second most recent) measurement time. As can be seen in the equation, the time between each measurement is not considered and should stay consistent for predictable control loop results.

In general, controller units utilize either a real time operating system (RTOS) or a general-purpose operating system (GPOS). RTOS has the advantage of being able to complete tasks strictly following a schedule [Ham14]. As such, it is best for time sensitive tasks. However, not all platform support RTOS. GPOS, in contrast, does not promise strict deadlines but are generally done on a best effort basis. As such, GPOS tend to be less viable for time sensitive tasks.

There are two main ways to maintain concurrent control of multiple tasks, multiprocessing, and multithreading. Multiprocessing takes advantage of the multiple cores of a processor and completes different tasks on each core to achieve true parallelism. However, multiprocessing can only take place on controllers with multiple cores. As such, multiprocessing is more viable to increasing computing power by parallel computing. Multithreading, however, is a kind of context switching. It takes utilizing the idle time of each task to complete parts of other tasks. Therefore, multithreading achieves concurrency of tasks but does not provide benefits to increasing computing power.

1.3 Recognition of Chemicals

In a multidetector μ GC, each chemical produces their characteristic retention time and detector signals. The relative strength of the signals as a ratio form detector response pattern (DRP). DRP has been used as a secondary metric to recognize chemicals in support of retention time. In gas chromatograph analysis, principal component analysis (PrCoAn) is often used [Skr07, Wan19, Hov89] heuristically to aid in recognition using DRP along with other statistical methods. PrCoAn does not provide recognition; it serves to increase data separation while reducing data dimensions [Jol11]. PrCoAn rotates data so that the variance is maximized on the new axes. The rotation is a linear transformation. The eigenvalues from the rotation explain the amount of variance in the new axis dimension. Often, only the axes with high eigenvalues are retained and are deemed to be the principal component, thus achieving maximizing data separation while reducing data dimension. The PrCoAn process can be mathematically expressed as

$$AX = \lambda X \quad \dots\dots\dots(2)$$

where A is the n dimension by n dimension covariance matrix, X is matrix of eigenvectors and λ is the eigen values of the eigenvectors. The user can then choose to remove eigenvectors from the matrix of eigenvectors based on their eigenvalue. The rotated and dimension reduced data is obtained by multiplying the original data by the matrix containing the selected eigenvectors.

Numerous classifications methods can be used to recognize data, sometimes after the use of PrCoAn. Linear discriminate analysis (LDA) [Skr07], is similar to PrCoAn but is used to maximize separation between known classes while reducing dimensions. This means LDA is a supervised learning method, where the classification of the data points must be known. As the classifications are known, LDA first calculates the eigenvectors and eigenvalues of the pooled

within group covariance instead of the covariance matrix. The eigenvector provides the rotation and the eigenvalue indicates the amount of classification strength the new axis provides.

Another measure of separation of data is Fisher's Ratio (F-Ratio), which scores the discrimination power of two classes. The F-Ratio is expressed as:

$$F_{Ratio} = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2} \dots\dots\dots(3)$$

where μ_1 and μ_2 are the means of the the first and second class, and σ_1^2 and σ_2^2 are the variances of the first and second class.

Machine learning methods like artificial neural networks (ANN) are yet another tool to help in the recognition of chemicals. ANN aims to train a model inspired by the neural connectivity of biological brains. A typical ANN has an input layer, hidden layer(s) and an output layer. Each hidden layer consists of interconnected neurons [Jia96, Zur92] The hidden layer takes the inputs and transforms it to a different value using an activation function. Sometimes multiple hidden layers are used. The only limitation to activation functions is it should be nonconstant and bounded [Hor91]. Machine learning techniques, however, often require massive amounts of training data [Alw18].

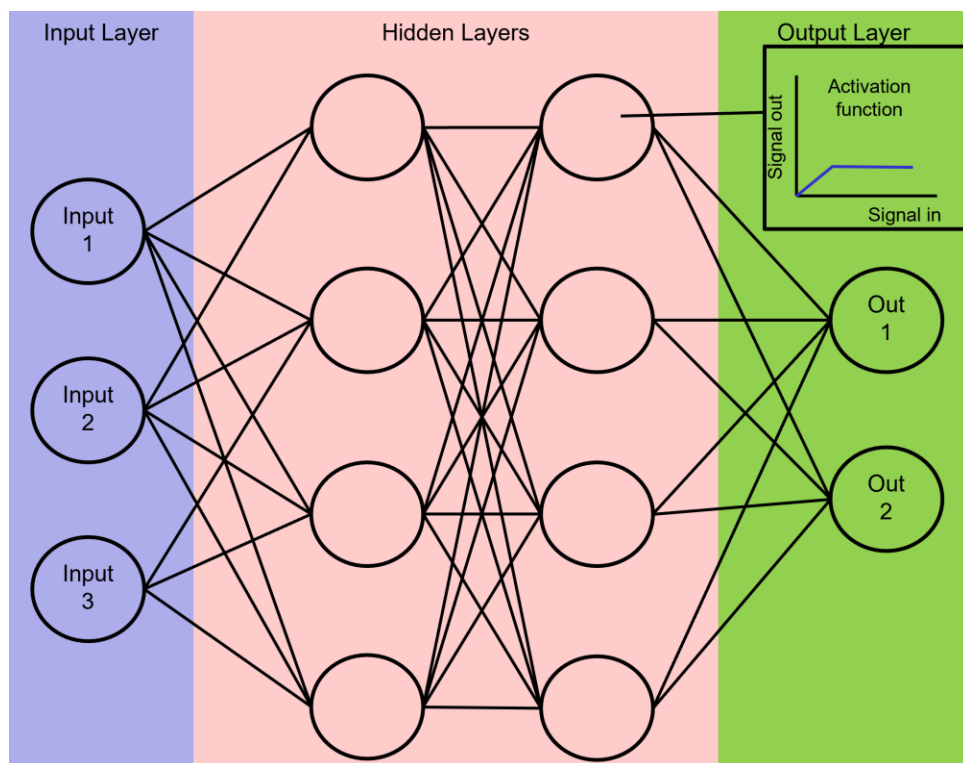


Figure 1.5: Illustration of an artificial neural network.

Though powerful, the training data sets required from statistical and machine learning techniques place a heavy burden on the characterization of a μ GC. Additionally, statistical and machine learning techniques often do not take into account of expert knowledge of the DRP, which can be useful for recognition. For example, the response of special chemicals may produce outlier data. Outlier data is useful for human based chemical recognition but its inclusion in PrCoAn can significantly alter the results [Neu19].

An expert system (sometimes called inference engine) transforms expert knowledge into logical expressions and outputs results based on expert knowledge [Buc82]. For recognition of chemicals, an expert system takes in features used in recognition, applies recognition rules, and infers the most likely chemicals. Rules based on expert knowledge can be constructed for recognition without deference to statistics. Expert systems usually have two modes of reasoning, forward

changing and backwards chaining [Yan19]. In forward chaining , the rules are directly applied to the inputs and an outcome is presented. In backwards chaining, an outcome is first proposed, and the rules try to determine if the inputs are suitable for the outcome.

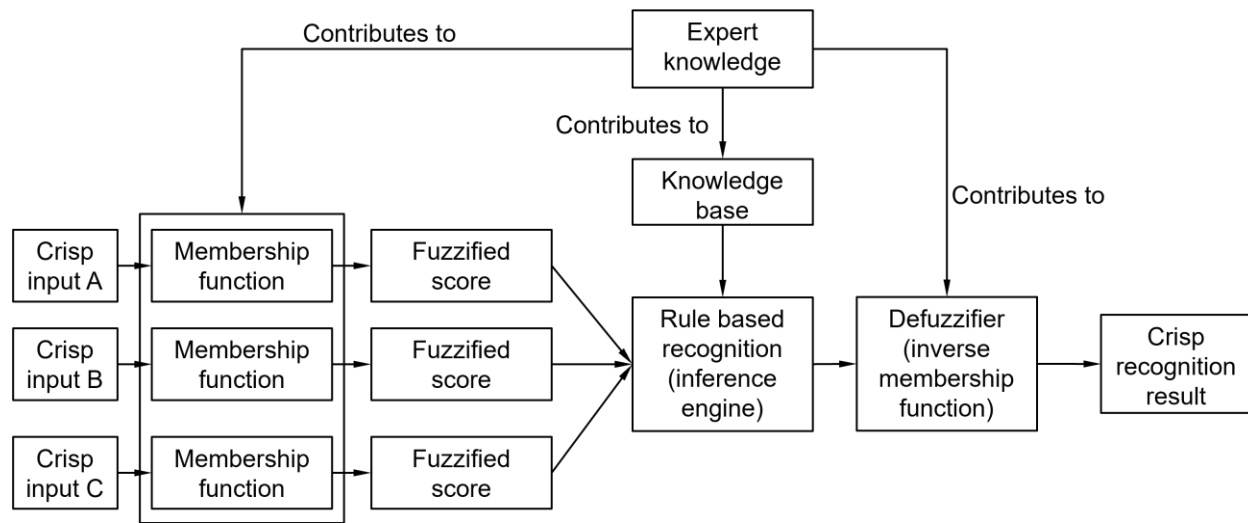


Figure 1.6: Diagram of the architecture of an expert system

As an expert system is a representation of logical thinking, inputs often need to take on values conducive to logical representation (i.e., if A is true and B is true, then C). However, real life values are rarely binary, and the values of input features often deviate from the ideal values. Therefore, a method of measuring the degree of truth is needed. Fuzzy logic uses a membership function to determine the degree of truth of inputs [Pec21]. A mathematical expression for the rules can then determine the degree of likelihood of the outcome (i.e., if A is 80% true and B is 90% true, then the outcome 72% likely C). Membership functions can take any form if it is bounded. Expert knowledge can be used to establish a membership function as well.

Compared to statistical methods, rule based expert systems require a minimal set of training data and special rules for special case recognition can be readily added while maintaining the integrity of the main recognition process. The results from an expert system are

also traceable, showing the cause of success or failure. However, the accuracy of recognition then depends on the accuracy of expert knowledge.

1.4 Motivation for This Work

VOCs can be harmful to humans and the environment, especially in high concentrations. VOCs are generally encountered by people in everyday life but are more prevalent in industrial settings [Har15]. Various industries, including oil refining, printing, energy generation, garbage collection, chemical processing, food processing and many other manufacturing processes emit VOCs [Zhe09, Wan23, Cao16]. Guidelines and regulations for VOC exposure have been put forward by various government agencies such as Occupational Safety and Health Administration (OSHA) and National Institute for Occupational Safety and Health (NIOSH) [Wan19]. Though potentially hazardous to humans, VOCs are also a viable marker for process control and process health [Kan23, Dre19]. Due to these risks and potential uses, the identification and quantification of VOCs are therefore important for ensuring the health of humans and the environment, and for use in industry. The current gold standard for VOC detection is benchtop GCs. However, benchtop GCs are often bulky, high in cost, high in energy consumption, high in heat generation, and often require trained personnel to operate. These downsides make them suitable for laboratory analysis but unsuitable for automated or autonomous *in situ* analysis of VOCs.

Microelectromechanical systems have enabled the miniaturization of many instruments, including GC. The first publication of the concept of a microscale gas chromatograph (μ GC) was in 1972 [Car72], and the first successful manufacturing of μ GC was reported in 1979 [Ter79]. The small form factor, lower cost, lower power consumption, lower heat generation, and often faster analysis speed make μ GC a suitable candidate for a fully automated *in situ* test

system. A fully automated μ GC system includes the fluidics hardware, the electronics, the control software, and chemical recognition software.

Although the developments of fluidic hardware of μ GC systems are well documented, the design of control software is scarcely documented. The recognition of chemicals from chromatograms are either accomplished by human experts [Gar11], by various statistical methods [Wan19, Hov89, Gar11], machine learning techniques [Mat20] or library look up methods [Ste99]. However, most techniques used are some distance away from being automated, often require extensive characterization of the μ GC and do not consider expert knowledge of the system.

This thesis aims to advance the level of automation of μ GC systems in both controls and chemical recognition on a representative highly complex μ GC based on a Multisensing Progressive Cellular Architecture [Lia23] (in this work MPCA refers to the μ GC). The MPCA features various and numerous transducers that have different control mechanisms and often must be controlled concurrently, sometimes with closed loop controls. The control software supports concurrent controls. The acquisition is designed to have high data rates and consistent mean time between measurements. Fail safes are implemented to cover possible failure modes. A graphical user interface (UI) is designed for remote control and monitoring of the MPCA. The control software is validated by successful test runs that generate viable data. The data is then fed to a peak detection algorithm (not within the scope of this work), which provides the needed data for the chemical recognition algorithm. The chemical recognition algorithm implements fuzzy logic, and rules based on expert knowledge to recognize chemicals. The recognition algorithm is validated by the successful recognition of chemicals. The control software and

recognition algorithm serve to increase the level of automation of μ GC. Within the objective to increase automation, there are six goals:

1. Investigate challenges to automated controls of a highly complex μ GC. Highly complex μ GCs incorporate different and various electronics that often must be controlled concurrently. Electronics may need heterogenous controls.
2. Design the architecture of the control software and implement it for representative system. Demonstrate a pathway for controlling complex μ GCs.
3. Validate the control software by inspecting the chromatograph results and closed loop control results.
4. Investigate the challenges of automating chemical recognition. Heuristically used methods do not take full advantage of the emerging trends in hardware, and alternative approaches need to be investigated.
5. Implement the automated recognition algorithm. Special rules to improve recognition must be implemented.
6. Validate the chemical recognition algorithm with test sets and explain the recognition results.

1.5 Organization of This Thesis

This thesis consists of four chapters and one appendix. The first chapter introduced the principles of GC, MPCA operations, the challenges of controls of MPCA, the recognition of chemicals in μ GCs, and the importance of automation for μ GCs.

Chapter 2 investigates the challenges of automated control software of MPCA and provides a design for the challenges. The MPCA features three cells, with each cell

incorporating a preconcentrator, a separation column, two capacitive detectors (CapDetA and CapDetB), and one photoionization detector (AiPD). The MPCA also incorporates a mini motor driven pump, two piezoelectric pumps, two differential pressure sensors, and six valves for sample flow control and eight thermistor heater pairs for temperature control. The control software provides solutions for concurrent controls of heterogenous components, heterogenous control mechanism, heterogenous control loops, fail safes, and addresses challenges in timing while also providing data acquisition. A UI for remote control is also implemented. This work has been published [Xu23a].

Chapter 3 investigates the challenges of automated chemical recognition for MPCA. Traditionally used techniques do not take advantage of expert knowledge of MPCA and may require large training sets. The chemical recognition algorithm uses fuzzy logic, and expert knowledge to create a rule-based recognition algorithm with a minimal training set. The orthogonality of the detectors, expert knowledge of the retention time and detector response patterns (DRPs), expert knowledge of special chemical were leveraged to construct the rules of recognition. Additional rules for using a reference chemical and for treating small signals are also included. The rule-based recognition algorithm achieved success with minimal training data. This work has been published [Xu23b].

Chapter 4 concludes the current work and discusses the possible improvements to the control software and chemical recognition algorithm.

Appendix A includes published journals and manuscripts in preparation that uses the contents of this work to achieve automated controls.

References

- [Alw18] A. Alwosheel, S. van Cranenburgh, C.G. Chorus, C.G. Is your dataset big enough? Sample size requirements when using artificial neural networks for discrete choice analysis. *Choice Model*, vol 28, pp167–182, 2018.
- [Ama19] M.S. Amaral, Y. Nolvachai, and P. J. Marriott, “Comprehensive two-dimensional gas chromatography advances in technology and applications: Biennial update,” *Analytical Chemistry*, vol. 92, no. 1, pp. 85–104, 2019.
- [Ban98] R. Bannatyne and G. Viot, "Introduction to microcontrollers. I," Wescon/98. Conference Proceedings (Cat. No.98CH36265), Anaheim, CA, USA, 1998, pp. 350-360
- [Bra23] S.E. Bramer, *An Introduction to Mass Spectrometry*. LibreText.
- [Buc82] B. G. Buchanan and R. O. Duda, *Principles of Rule-Based Expert Systems*. Stanford, CA: Stanford University. Computer Science Department, 1982.
- [Cal23] A. Calabres, P. Battstoni, S. Ceylan, L. Zeni, A. Capo, A. Varriale, S. D’Auria, M. Staiano, “An impedimetric biosensor for detection of volatile organic compounds in food,” *Biosensors*, vol. 13, no. 3, pp. 34.
- [Cao16] X.L. Cao , M. Sparling, R. Dabeka, “Occurrence of 13 volatile organic compounds in foods from the Canadian total diet study”. *Food Additives & Contaminants: Part A*,. 2016;33(2):373-82.
- [Car72] G.C. Carle, R.W. Donaldson, S.C. Terry, K.D. Wise, “Microminiature gas chromatograph,” *NASA Tech Briefs*, Sept. 1972.
- [Dąb18] Ł. Dąbrowski, “Multidetector systems in gas chromatography,” *TrAC Trends in Analytical Chemistry*, vol. 102, pp. 185–193, 2018.
- [Dav21] E. David and V.-C. Niculescu, “Volatile Organic Compounds (vocs) as environmental pollutants: Occurrence and mitigation using nanomaterials,” *International Journal of Environmental Research and Public Health*, vol. 18, no. 24, pp. 13147, 2021.
- [Dre19] J. M. Dreimann, E. Kohls, H. F. Warmeling, M. Stein, L. F. Guo, M. Garland, T. N. Dinh, and A. J. Vorholt, “In situ infrared spectroscopy as a tool for monitoring molecular catalyst for hydroformylation in continuous processes,” *ACS Catalysis*, vol. 9, no. 5, pp. 4308–4319, 2019.
- [EPA99] EPA, “Air method, toxic organics-15 (TO-15): Compendium of methods for the determination of toxic organic compounds in ambient air, second edition: determination of volatile organic compounds (VOCs) in air collected in specially-prepared canisters and analyzed by gas chromatography/mass spectrometry (GC/MS),” *EPA 625/R96/010b*, 1999.

[EPA23] EPA, “Volatile organic compounds’ impact on indoor air quality,” Volatile Organic Compounds’ Impact on Indoor Air Quality, <https://www.epa.gov/indoor-air-quality-iaq/volatile-organic-compounds-impact-indoor-air-quality> (accessed Nov. 20, 2023).

[Far05] O. Faroon., J. Taylor, N. Roney, M.E Fransen, S. Bogaczyk , G, Diamond. “Toxicological Profile for Carbon Tetrachloride. Department of Health and Human Services”, Public Health Service Agency for Toxic Substances and Disease Registry; Atlanta, GA, USA: 2005.

[Gar11] E. García-Berríos, T. Gao, J.C. Theriot, M.D. Woodka, B.S. Brunshwig, N.S. Lewis, “Response and discrimination performance of arrays of organothiol-capped au nanoparticle chemiresistive vapor sensors,” *The Journal of Physical Chemistry C*, vol. 115, no. 14, pp. 6208–6217, 2011.

[Gro04] R. L. Grob and E. F. Barry, *Modern Practice of Gas Chromatography*. Hoboken, NJ: Wiley-Interscience, 2004.

[Gru23] M. Grusin, “Serial peripheral interface (SPI),” Serial Peripheral Interface (SPI) - SparkFun Learn, <https://learn.sparkfun.com/tutorials/serial-peripheral-interface-spi/all> (accessed Nov. 20, 2023).

[Ham14] P. Hambarde, R. Varma, and S. Jha, “The survey of Real Time Operating System: RTOS,” *International Conference on Electronic Systems, Signal Processing and Computing Technologies*, 2014.

[Hov89] H.T. Hove, O. Grahl-Nielsen, A. Rogstad, “Assay for dinoflagellate toxins in mussels by gas chromatography and Principal Components Analysis”, *Anal. Chim. Acta* 1989 pp 222, 35–42.

[Hor91] K. Hornik, “Approximation capabilities of multilayer Feedforward Networks,” *Neural Networks*, vol. 4, no. 2, pp. 251–257, 1991.

[Hua22] X. Huang, M.W Li, W. Zhang, X. Huang, A.D. Sivakumar, R. Sharma, X. Fan, “Portable comprehensive two-dimensional micro-gas chromatography using an integrated flow-restricted pneumatic modulator”, *Microsystems & Nanoengineering*, vol 8, no. 1, 2022.

[Jia96] A.K. Jain, J. Mao K.M. Mohiuddin, Artificial neural networks- a tutorial, *Computer*, 29(3) (1996) 31-44.

[Jol11] I. T. Jolliffe, *Principal Component Analysis*. New York: Springer, 2011.

[Kan23] W. Kang et al., “Advanced sensing of volatile organic compounds in the fermentation of kombucha tea extract enabled by nano-colorimetric sensor array based on density functional theory,” *Food Chemistry*, vol. 405, pp. 134193, 2023.

[Lia23] W. Liao, D. Winship, L. Lara-Ibeas, X. Zhao, Q. Xu, H. Lu, T. Qian, R. Gordenker, Y. Qin, Y.B. Gianchandani, “Highly Integrated μ GC Based on a Multi-sensing Progressive Cellular Architecture with a Valveless Sample Inlet”, *Analytical Chemistry*, vol 95, no. 4, pp. 2157–2167, 2023.

[Mat20] D.D. Matyushin, A.Y. Sholokhova, A.K. Buryak, “Deep Learning Driven GC- MS Library Search and its application for Metabolomics,” *Analytical Chemistry*, vol. 92, no. 17, pp. 11818–11825, 2020.

[Mbe23] “Mbed LPC1768,” Mbed, <https://os.mbed.com/platforms/mbed-LPC1768/> (accessed Nov. 20, 2023).

[McN19a] H.M. McNair, J. M. Miller, and N. H. Snow, “Chapter 6 Basic Gas Chromatography, Third Edition Chapter 6 TEMPERATURE PROGRAMMING,” in *Basic gas chromatography*, Hoboken , N.J: Wiley, 2019.

[McN19b] H.M. McNair, J. M. Miller, and N. H. Snow, “Chapter 6 Basic Gas Chromatography, Third Edition Chapter 3 INSTRUMENT OVERVIEW,” in *Basic gas chromatography*, Hoboken , N.J: Wiley, 2019.

[Neu19] S. Neumayer, M. Nimmer, S. Setzer, and G. Steidl, “On the robust PCA and Weiszfeld’s algorithm,” *Applied Mathematics and Optimization*, vol. 82, no. 3, pp. 1017–1048, 2019.

[Pen23] E. Pena and M. G. Legaspi, “UART: A hardware communication protocol understanding universal asynchronous receiver/transmitter,” *Analog Devices*, <https://www.analog.com/en/analog-dialogue/articles/uart-a-hardware-communication-protocol.html> (accessed Nov. 20, 2023).

[Pig23] pigpio Library, “The Pigpio Library,” <http://abyz.me.uk/rpi/pigpio/> (accessed Jul. 13, 2023).

[Pec21] J. K. Peckol, *Introduction to fuzzy logic*, 2021.

[Qin16] Y. Qin and Y.B. Gianchandani, “A fully electronic microfabricated gas chromatograph with complementary capacitive detectors for indoor pollutants,” *Microsystems & Nanoengineering*, vol. 2, no. 1, 2016.

[Skr07] V. L. Skrobot, E. V. Castro, R. C. Pereira, V. M. Pasa, and I. C. Fortes, “Use of principal component analysis (PCA) and linear discriminant analysis (LDA) in gas chromatographic (GC) data in the investigation of gasoline adulteration,” *Energy & Fuels*, vol. 21, no. 6, pp. 3394–3400, 2007.

[Ste99] S.E. Stein, “An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data,” *Journal of the American Society for Mass Spectrometry*, vol. 10, no. 8, pp. 770–781, 1999.

- [Spa11] O.D. Sparkman, Z. Penton, and F. G. Kitson, *Gas Chromatography and Mass Spectrometry: A Practical Guide*. Burlington, MA: Academic Press, 2011.
- [Ter79] S.C. Terry, J.H. Jerman, J.B. Angell, “A gas chromatographic air analyzer fabricated on a silicon wafer,” *IEEE Transactions on Electron Devices*, vol. 26, no. 12, pp. 1880-1886, 1979.
- [Val15] J. Valdez and J. Becker, “Understanding the I2C bus ,” Understanding the I2C Bus, <https://www.ti.com/lit/an/slva704/slva704.pdf> (accessed Nov. 20, 2023).
- [Wan19] J. Wang, N. Nuñoovero, R. Nidetz, S. J. Peterson, B. M. Brookover, W. H. Steinecker, and E. T. Zellers, “Belt-mounted micro-gas-chromatograph prototype for determining personal exposures to volatile-organic-compound mixture components,” *Analytical Chemistry*, vol. 91, no. 7, pp. 4747–4754, 2019.
- [Wan23] S. Wang, J. Zhang, Y. Zhang, L. Wang, Z. Sun, and H. Wang, “Review on Source Profiles of Volatile Organic Compounds (VOCs) in Typical Industries in China,” *Atmosphere*, vol. 14, no. 5, p. 878, 2023.
- [Xu23a] Q. Xu, Y. Zhao, Y. Qin, Y.B. Gianchandani, “Control Software Design for a Multisensing Multicellular Microscale Gas Chromatography System,” *Micromachines*, vol. 15, no.1, p 95, 2023.
- [Xu23b] Q. Xu, Y. Qin, Y.B. Gianchandani, “A Rule-Based Automated Chemical Recognition Algorithm for a Multi-Cell Multi-Detector Micro Gas Chromatograph,” *Separations*, vol 10, no.11, p 555, 2023
- [Yan19] J. Yanase and E. Triantaphyllou, “A systematic survey of computer-aided diagnosis in medicine: Past and present developments,” *Expert Systems with Applications*, vol. 138, p. 112821, 2019.
- [Zhe09] J. Zheng, M. Shao, W. Che, L. Zhang, L. Zhong, Y. Zhang, and D. Streets, “Speciated VOC Emission Inventory and Spatial Patterns of Ozone Formation Potential in the Pearl River Delta, China”, *Environmental Science & Technology*, 2009.
- [Zho16] M. Zhou, J. Lee, H. Zhu, R. Nidetz, K. Kurabayashi, X. Fan, “A fully automated portable gas chromatography system for sensitive and rapid quantification of volatile organic compounds in water,” *RSC Advances*, vol. 6, no. 55, pp. 49416–49424, 2016.
- [Zur92] J. Zurada, *Introduction to artificial neural systems*, West Publishing Co., St. Paul, MN, USA, 1992.

Chapter 2 Control Software Design for a Multisensing Multicellular Microscale Gas Chromatography System¹

Microscale gas chromatography (μ GC) systems are miniaturized instruments that typically incorporate one or more microfabricated fluidic elements; such systems are generally well-suited for automated sampling and analysis of gas phase chemicals. Advanced μ GC systems may incorporate more than 15 elements and operate these elements in different coordinated sequences to execute complex operations. In particular, the control software must manage sampling and analysis operations of the μ GC system² in a time sensitive manner; while operating multiple control loops it must also manage error conditions, data acquisition, and user interactions when necessary. To address these challenges, this work presents a multithreaded control software for a representative μ GC system¹ based on a progressive cellular architecture that uses multiple μ GC cells to efficiently broaden the range of chemical analytes as well as multiple detectors within each cell to enhance chemical recognition. Implemented in Python software on an embedded single-board computer, the control software enables concurrent control of heaters, pumps, and valves, while also gathering data from thermistors, pressure sensors, capacitive detectors, and photoionization detectors. A graphical user interface (UI), which operates on a laptop, provides remote control and real time data visualization. In experimental evaluations, the control software provided successful operation and readout of all the

¹ This chapter is reproduced from a journal manuscript that has been published [Xu23a].

² PCB design was performed by Dr. Xiangyu Zhao, Mr. Declan Winship and Mr. Robert Gordenker. Section 2.2.2 was written in collaboration with Dr. Xiangyu Zhao. The microfluidics components were designed by and fabricated by Dr. Weilin Liao. An early prototype of the software was developed by Mr. Ali Dowair and Mr. Chandler Creech.

components, including 8 sets of thermistors and heaters that form temperature feedback loops, 2 sets of pressure sensors and tunable gas pumps that form pressure head feedback loops, 6 capacitive detectors, 3 photoionization detectors, 6 valves, and an additional fixed-flow gas pump. A typical run analyzing 18 chemicals is presented. Despite the use of an operating system that does not guarantee real-time operation, the relative standard deviations of the control loop timings were <0.5%. The control software successfully supported >1000 μ GC runs that analyzed various chemical mixtures.

2.1 Introduction

Microscale gas chromatography (μ GC) systems have shown promise for use in *in situ* detection and measurement of volatile organic compounds (VOCs) since their introduction in the 1970s [Car72, Ter79]. A μ GC system incorporates a number of fluidic components for gas-phase flow management. The major fluidic components in μ GC systems include preconcentrators, separation columns, detectors, pumps, and valves [Reg18]. Typical operation involves two steps that constitute an analytical run: first, a sampling step is performed, in which the VOC molecules are drawn into the system using a sampling pump and adsorbed in the preconcentrator; subsequently, the VOC molecules are thermally desorbed and transported through the separation column by a carrier gas flow, during which the VOC molecules are spatiotemporally separated. Further downstream, the detector responds to the eluting analytes and forms a chromatogram, in which the location and magnitude of each peak indicate the species and quantity of a VOC, respectively.

The performance of μ GC systems can be enhanced using complex architectures with a larger number of fluidic components. Some μ GC systems incorporate multiple complementary detectors to enhance chemical recognition. These detectors are based either on a single sensing

principle (*e.g.*, capacitive detection [Qin16] or chemi-resistive detection [Wan19], or resonance-based detection [6]) or on a mix of sensing principles with higher orthogonality [Lia23]. A 2D μ GC system typically incorporates two columns of complementary separation mechanisms to enhance the separation; typically, it requires either a thermal modulator [Col15] or a flow modulator [Hua22]. The recently reported progressive cellular architecture [Lia23, Lia21] incorporates multiple sets of preconcentrators and columns, with each set tailored for a different range of the VOC volatility.

In a μ GC system, the diverse fluidic components all require specialized controls. For example, the preconcentrators, separation columns, and detectors require separate temperature controls, as they are operated at different temperatures and with different heating rates. The heating control is usually provided by either pulse width modulation (PWM) of a fixed voltage or by controlling an analog voltage. This control is preferably implemented in a servo-controlled (closed-loop) manner using temperature measurements from either thermistors or thermocouples connected to analog-to-digital converters (ADCs) [Gar15, McC17]. Capacitive detectors are typically read out using capacitance-to-digital converters (CDCs) [Qin16, Lia23, Lia21, Pat03]; photoionization detectors are read out amplifiers and ADCs [Lia23, Li21]; resonance-based detectors are read out by phase-locked loop circuits [Li10] and ADCs. The valves used in μ GC systems are typically active valves – either latching or non-latching. Latching valves are actuated using a short voltage pulse, the polarity of which controls whether the valve is opened or closed. In contrast, non-latching valves need sustained voltage levels, but with an additional complexity favored for certain models in which the sustained voltage should be lower than the initial actuating voltage. The commonly used miniature piezoelectric or motor-driven diaphragm

pumps are controlled by an on/off power and sometimes in combination with a tunable actuation frequency.

Within an analytical run, all the μ GC components must be continuously and concurrently controlled. One of the challenges is that the data rates of different elements can vary widely. For example, the data rates of detectors vary based on the sensing principles and the readout electronics; higher data rates are generally favored. Other components may differ even more widely in data rates. The control software must accommodate the heterogeneous timing of individual components, with some components operating in closed-loop control.

The control software must also implement fail-safe protocols to handle hardware error cases, such as accidental overheating and thermal runaway. Otherwise, these error cases may cause irreversible damage to the μ GC system.

To achieve standalone operation, all these hardware controls must be provided by an embedded computer (EC). Conversely, to facilitate user interaction, a user interface is often preferred on an external computer that can communicate with the EC.

This work presents advances in control software using a representative μ GC system that is based on a multisensing progressive cellular architecture (MPCA) [Lia23]. This MPCA system is an example of a highly complex μ GC system with a large number and variety of components that require concurrent software control. The MPCA system incorporates eight elements that require individual temperature control, six detectors that require capacitive readout, three detectors that require voltage readout, six latching valves, and three pumps with different control requirements. This work investigates a multithreaded software architecture for the concurrent control of the MPCA system, with specific considerations regarding thread implementation, thread timing, temperature and flow control, communications between the EC and electronics,

data saving and storage, low-power operation, and the user interface (UI) that communicates with the EC.

2.2 Design and Implementation

2.2.1 MPCA System Overview

The MPCA system incorporates a monolithic microfabricated MPCA chip in addition to multiple miniature valves, pumps, and pressure sensors. The MPCA chip integrates three μ GC cells, each containing a preconcentrator, a separation column, and three complementary detectors (Figure 2.1). The three cells are tailored to different ranges of analyte volatility, primarily by altering the sorbent materials in the preconcentrators and adjusting the stationary phase thicknesses in the separation columns. The three detectors within each cell are based on different sensing principles, providing complementary responses that enhance chemical recognition. Two of the detectors are capacitive detectors (CapDets), that use coplanar interdigitated thin-film metal electrodes by with a sensing polymer of controlled thickness. The absorption of incoming analytes causes changes in both the thickness and dielectric constant of this layer, leading to changes in the detector capacitance. The first capacitive detector (CapDetA) primarily responds to polymer swelling, which produces only positive changes to the capacitance for all analytes. In contrast, the response of the second capacitive detector (CapDetB) is dominated by the change in dielectric constant, which may produce positive or negative changes to the capacitance depending on the dielectric constant of the analytes relative to that of the polymer itself. The third detector is an arrayed integrated photoionization detector (AiPD) which uses biased coplanar electrodes to draw a current from the photoionization of the analyte molecules. The photoionization is caused by ultraviolet radiation. The AiPDs across the three cells share a vacuum ultraviolet (VUV) lamp. Additionally, the MPCA chip incorporates a carrier gas filter,

which dehumidifies ambient air to supply the carrier gas. For each preconcentrator, each separation column, the carrier gas filter, and the ensemble of all the detectors, a pair of on-chip heater and thermistor are incorporated to support closed-loop temperature control.

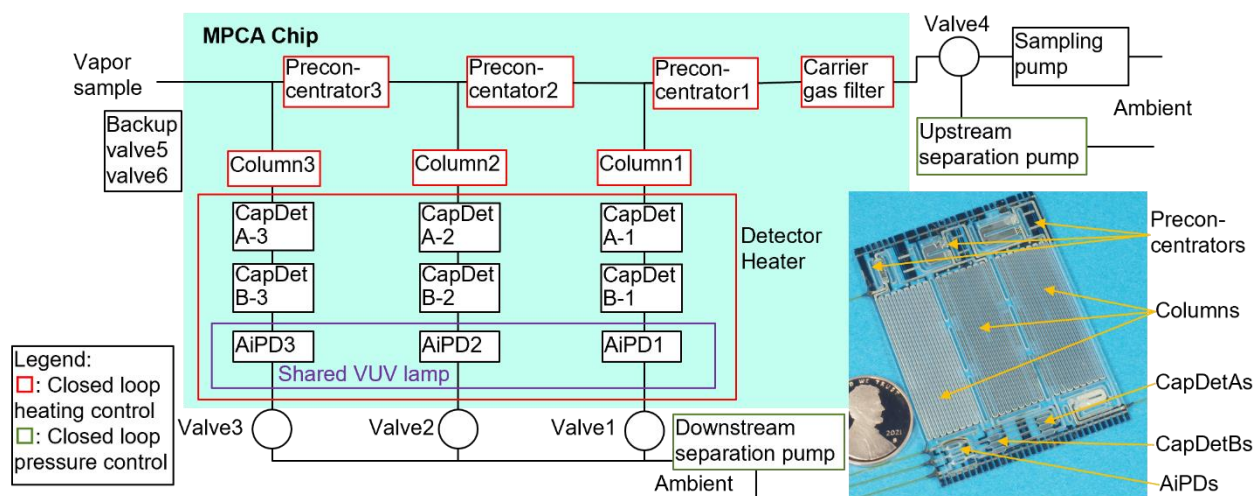


Figure 2.1: MPCA fluidic architecture and photo of the chip.

The MPCA system operation includes a sampling step, three sequential separation steps (one for each cell), and multiple purging steps. During the sampling step, the flow is generated using a sampling pump and routed using the valves. The analyte samples are transported through the preconcentrators sequentially, first through the one with the least adsorptive sorbent and then through those with progressively more adsorptive sorbents. The analytes are adsorbed by the preconcentrators, with the lower-volatility chemicals adsorbed by the upstream preconcentrator, and the more volatile chemicals breaking through the upstream preconcentrator and being adsorbed by downstream preconcentrators. During the separation step for each cell, the corresponding preconcentrator is heated to desorb the chemicals, which are then carried forward by the carrier gas flow and separated by the corresponding column and subsequently detected by

the detectors. The carrier gas flow during separation is generated using both an upstream separation pump and a downstream separation pump, both of which may be used in conjunction with a pressure sensor to support closed-loop flow control. Prior to the next sampling step, purging steps may be used to reduce humidity and chemical carryover; these purging steps involve combinations of component heating and carrier gas flow. To support the MPCA system operation and continuing development, the control software must provide users with enough flexibility to modify operation sequence and parameters.

2.2.2 Electronic Interfaces

The EC serves as both the controller for the MPCA system electronics and the interface with the laptop UI (Figure 2.2). The EC may be embodied as either a microcontroller [Gar15] or a single-board computer [Wan19]. Microcontrollers are highly compact and inexpensive, but their limitations in computational power, visualization, and storage resources increase the burden on software development. In contrast, single-board computers provide a favorable balance between device size, cost, computational power, and ease of software development. The control and readout functions in this work can be supported by either a microcontroller or a single-board computer. However, from the perspective of a fully autonomous system for continuous, near-real-time *in situ* analysis, the post-processing of raw data requires extensive computation. [Xu23b, Zha19]. An alternative to the EC is a data acquisition (DAQ) device [Zho16]; it can be easily programmed and is attractive for proof of concept. However, without a built-in controller, the control software for a DAQ device typically resides on an external computer, which does not always facilitate standalone operation. This work uses a Raspberry Pi (RPI) series single-board computer [Ras23] as the EC. The RPI series includes various models with different computational power, power consumption, and sizes. Most models have the same pinout and the

same operating system, which facilitates upgrades to newer models. For this work, the control software is built in the RPi Model 3B+ [Ras23], which was the latest model available at the start of this work.

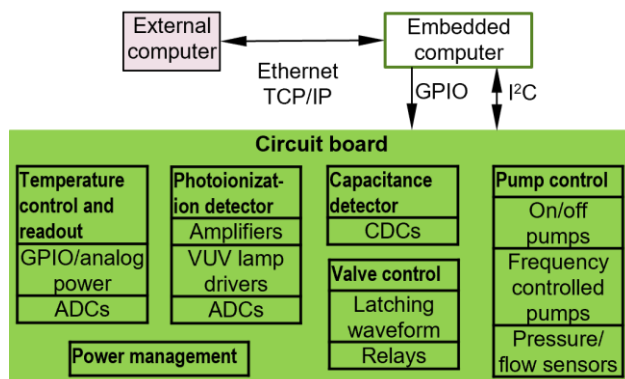


Figure 2.2: Block diagram of the electronic interface between components.

The electronic interfaces used within the MPCA system are selected for their compatibility with the general-purpose input/output (GPIO) pins and/or the inter-integrated circuit (I²C), which are readily available on the EC. A GPIO pin can be easily programmed to provide simple on/off control or to provide square waveforms with controllable frequency and duty cycles. An I²C bus, operating with just two wires, may be used to communicate with tens of integrated circuit (IC) chips. This capability makes the I²C more suitable for the MPCA system than other protocols, such as the serial peripheral interface (SPI) and the universal asynchronous receiver/transmitter (UART).

The component heating is controlled by pulse width modulation (PWM) for certain elements, and by analog voltage control for other elements. The PWM-based circuit uses a relay (#CS326, Coto Technology, North Kingstown, RI, USA) controlled by a PWM signal from the EC (Figure 2.3a). By varying the duty cycle of a 24 V square wave, the PWM signal controls the average output voltage using just one GPIO pin, while eliminating additional ICs. However, for

certain components that are sensitive to electrical interference, an analog voltage control circuit is desired. This is implemented using a digital potentiometer (# MCP45HVX1, Microchip Technology, Chandler, AZ, USA) that controls the output voltage of a buck converter (# TPS54061DRBR, Texas Instruments, Dallas, TX, USA) (Figure 2.3b). In the MPCA system, the heating of the preconcentrators, columns, and carrier gas filter uses PWM-based circuit, whereas the heating of the ensemble of the detectors uses the analog voltage control circuit.

For temperature measurements, each on-chip thermistor is read out using a voltage divider coupled to an ADC (#ADS1115, Texas Instruments, Dallas, TX, USA) (Figure 2.3c). This ADC provides 16-bit resolution at a data rate of 8-860 Hz. It contains a converter core that is shared by four input channels in a time-multiplexed manner. Therefore, the eight on-chip temperature channels can be read out using two chips of this ADC, maintaining the same resolution but at $\frac{1}{4}$ of the data rate compared to an individual channel. These are necessary for the closed-loop temperature control of the MPCA system.

The sampling pump is a mini motor-driven pump (#NMP03KPDC-L, KNF Neuberger Inc., Trenton, NJ, USA) that is controlled using a relay (Figure 2.3d). The separation flow requires more accurate and adjustable control. The two separation pumps are piezoelectric micropumps (#mp6-air, Servoflo Corporation, Lexington, MA, USA), each driven using an original equipment manufacturer (OEM) driver circuit. The driver circuit receives on/off control from a GPIO pin; it also receives a square wave with a variable frequency of 0–1000 Hz to control the flow rate of the separation pump (Figure 2.3e). To implement closed-loop control of the separation flow rate, two differential pressure sensors (#MPX5010DP, NXP Semiconductors, Eindhoven, The Netherlands) are used to measure upstream and downstream pressure heads generated by the two separation pumps; these two pressure sensors are directly read out using a

high-resolution ADC (#LTC2493, Linear Technology, Milpitas, CA, USA) (Figure 2.3f), which is rated with a 24-bit resolution at a data rate of 7.5 Hz. It contains a converter core shared by four time-multiplexed input channels. Therefore, the two pressure sensors can be read out using one chip of this ADC, maintaining the same resolution but at half of the data rate (i.e., 3.75 Hz) compared to an individual channel.

Six latching solenoid valves (#LHLA1231211H, The Lee Company, Westbrook, CT, USA) are used to route gas flow within the system. A single-pole double-throw relay (#1462042-1, TE Connectivity, Bevaix, Switzerland) is used to select the voltage level between that required for latching or unlatching a valve. Additionally, multiple relays (#CS326, Coto Technology, East Greenwich, RI, USA) are used to separately control the actuation of individual valves (Figure 2.3g).

Each of the two CapDet elements in the active cell is read out using a separate high-resolution capacitance-to-digital converter (CDC) (#AD7746ARUZ, Analog Devices Inc., Wilmington, MA, USA) (Fig. 3h); the CDC is rated for a root mean square (RMS) input-equivalent noise level of 4 aF and a peak-to-peak noise of 27 aF, at a data rate of 9.1 Hz. A limitation of this CDC is its single fixed I²C address, which allows only one CDC chip on each I²C bus. This limitation is overcome by using an I²C address translator (#LTC4316, Linear Technology, Milpitas, CA, USA) to assign different I²C addresses to the additional CDC chips. All the CDC chips are set in continuous conversion mode, operating at the slowest data rate (9.1 Hz) to achieve the highest available resolution. The excitation signal is set at 32 kHz frequency and ± 1.65 V peak-to-peak amplitude. The built-in capacitance offset is set at 13.344 pF, which provides an effective measurement range of 9.248 pF to 17.440 pF.

The CDC used in this work (#AD7746ARUZ) provides two time-multiplexed input channels; in principle, the 6 CapDets across the three μ GC cells can be covered by three chips of this CDC (Fig. 2.4a). Because of the sequential (*i.e.*, non-concurrent) operation of the separation steps in the three cells, the two CapDets in each cell are connected to two input channels that belong to two different chips of the CDC, hence avoiding time multiplexing in each CDC chip and allowing each CapDet to be read out at the maximum available rate for a given resolution. In this work, four CDC units are used to provide a level of design redundancy (Fig. 2.4b).

For the AiPDs, the shared VUV lamp is controlled using on/off power. The current generated by each AiPD is amplified and converted into a voltage signal using a low noise transimpedance amplifier circuit. This voltage is read out using a high-resolution ADC (#LTC2493, Linear Technology, Milpitas, CA, USA), which is the same model as that for the pressure sensors (Fig. 3i). Because of the sequential (*i.e.*, non-concurrent) operation of the separation steps in the three cells, the three AiPDs can be read out using three channels of this ADC chip without time multiplexing. Therefore, each AiPD can be read out at the nominal data rate of 7.5 Hz (which is twice as fast as for the pressure sensors).

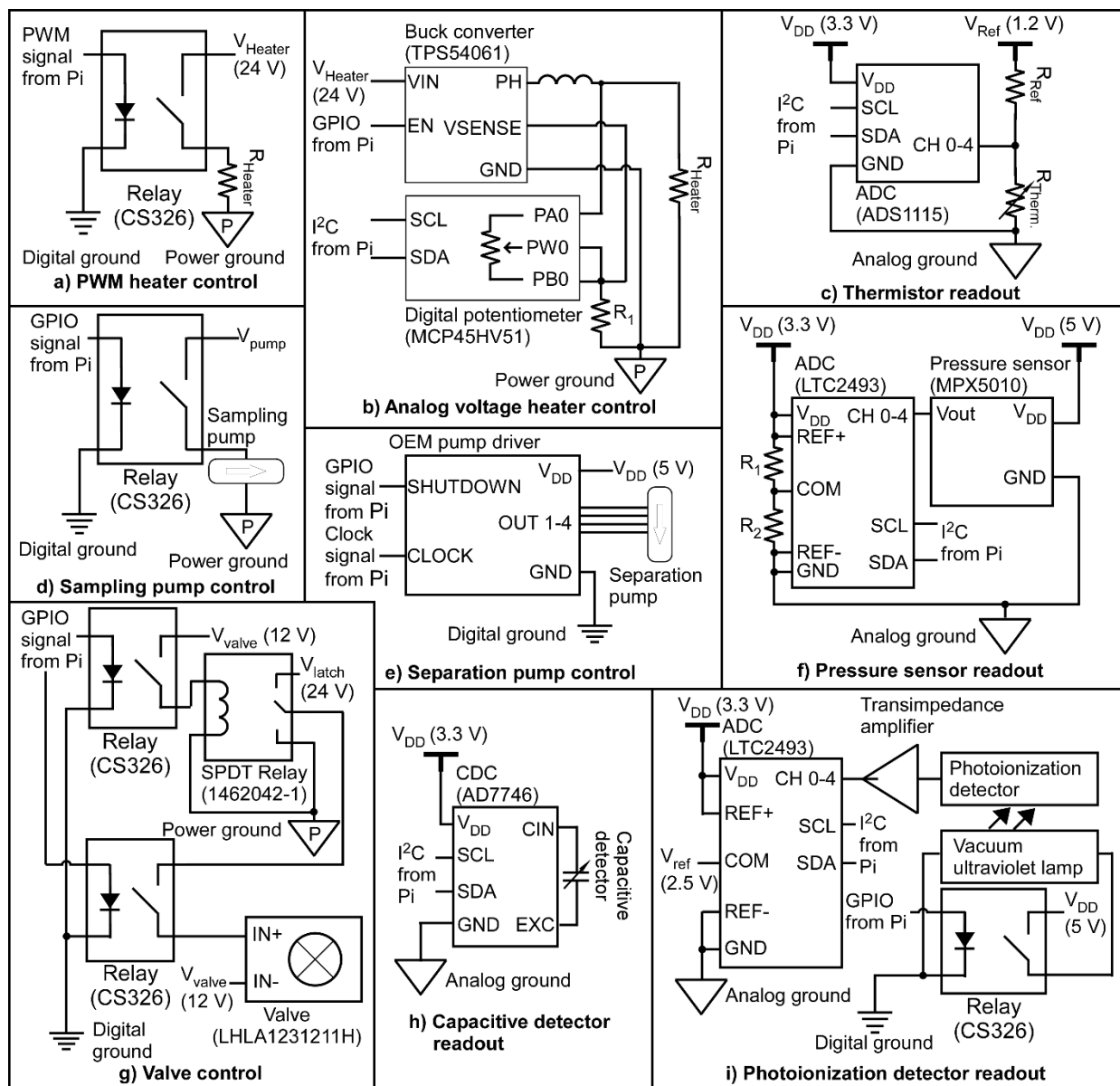


Figure 2.3: Circuit schematics of μ GC system hardware. (a) PWM heater control circuit. (b) Analog voltage heater control circuit. (c) Thermistor readout circuit. (d) Sampling pump control circuit. (e) Separation pump control circuit. (f) Pressure sensor readout circuit. (g) Valve control circuit. (h) Capacitive detector readout circuit. (i) Photoionization detector readout circuit.

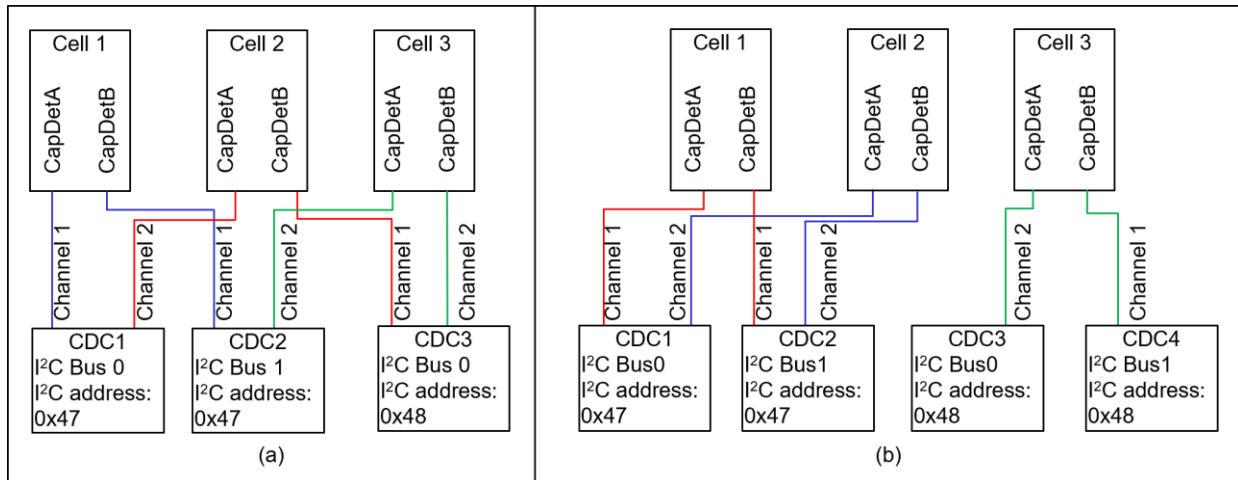


Figure 2.4: Illustration of connections between multiple CapDets and CDCs for the cases of (a) fewest CDC units needed, and (b) this work with a level of design redundancy.

To power the MPCA system, both 24 V and 12 V supplies are needed, each of which can be sourced from either a wall plug power supply or a battery pack (Fig. 4). The 12 V supply is dedicated to valve actuation. In contrast, the 24 V supply is used directly for component heaters (Fig. 3a,b) and for valve actuation (Fig. 3g). Additionally, the 24 V supply is used to provide 2 V for sampling pump actuation using a DC-DC converter (LMR14010A, Texas Instruments, Dallas, TX, USA). The 24 V supply is used to provide a 5 V supply, i.e., $V_{\text{Digital}} (5 \text{ V})$, using another DC-DC converter (R-745.0, RECOM Power, Gmunden, Austria). The $V_{\text{Digital}} (5 \text{ V})$ is used to provide a 3.3 V supply, i.e., $V_{\text{Digital}} (3.3 \text{ V})$, using a voltage regulator (NCP51460, Onsemi, Scottsdale, AZ, USA). These digital voltage supplies are used for digital electronics. Additionally, the $V_{\text{Digital}} (5 \text{ V})$ is treated using a Pi filter to provide another low-noise 5 V supply, i.e., $V_{\text{Analog}} (5 \text{ V})$, which is used to provide another low-noise 3.3 V supply, i.e., $V_{\text{Analog}} (3.3 \text{ V})$. These analog voltage supplies supply the data converters. The $V_{\text{Analog}} (5 \text{ V})$ is also converted to 1.2 V and 2.5 V reference voltages for the ADC chips using two different power regulators (LM4041, Texas Instruments, Dallas, TX, USA and LT6657, Analog Devices Inc., Wilmington, MA, USA, respectively).

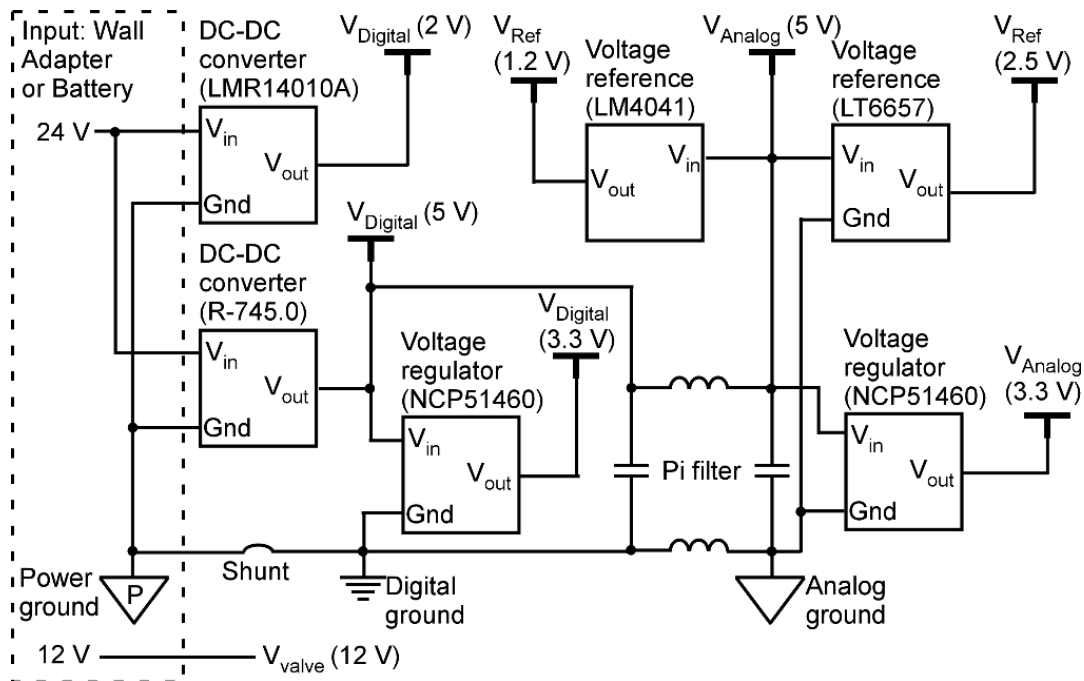


Figure 2.5: Schematic illustration of the power management circuit.

2.2.3 Control Software Implementation

In the MPCA system, the control software executed by the EC is programmed in Python 3.7.3. As the default language for the RPi, Python is compatible with a wide range of built-in and online open-source libraries, easing control software development. The main libraries that were used for this work include the open-source *pigpio* library [Pig23] for GPIO pin control, the open-source *SMBus* library [SMB23] for I²C communication, the Python built-in *multithreading* library for achieving multithreading, and the Python built-in *JSON* library for handling operation method files in JavaScript Object Notation (JSON) format. An operation method file contains all the necessary information for automating an analytical run. Each analytical run can be divided into up to eight run steps (such as a sampling step, a separation step, or a purging step). The operation methods file contains parameters for thermal, fluidic, detector, and timing controls of

each run step. The libraries for interfacing with the electronics, namely *SMBus* and *pigpio*, are open-source libraries with core functions implemented in C programming language [Fig23, SMb23], which provides the necessary performance for the control software. The specific challenges and solutions are described in the following sub-sections.

2.2.3.1 Multithreading Architecture

In each run step of the MPCA system operation, multiple fluidic components must be controlled concurrently. There are two mainstream approaches to providing concurrency: multiprocessing and multithreading (Fig. 2.6). Multiprocessing provides true parallelism [Pyt23a], where multiple tasks are executed simultaneously using multiple central processing unit (CPU) cores. It is most suitable for tasks that require intensive and parallel CPU usage. In contrast, multithreading achieves concurrency through context switching [Pyt23b], where only one task is active at a time while the others are waiting. Multithreading is better suited for the MPCA system because all the readout ICs require relatively long data conversion intervals (of 15.6–146.9 ms) between measurements, increasing the computational power does not benefit the performance. In addition, multithreading is compatible with single-core CPUs that are commonly used in low-power single-core microprocessors (*e.g.*, the RPi Model Zero [Ada23]) that are favored for ultra-portable systems. Therefore, for this work, multithreading was chosen to provide concurrency.

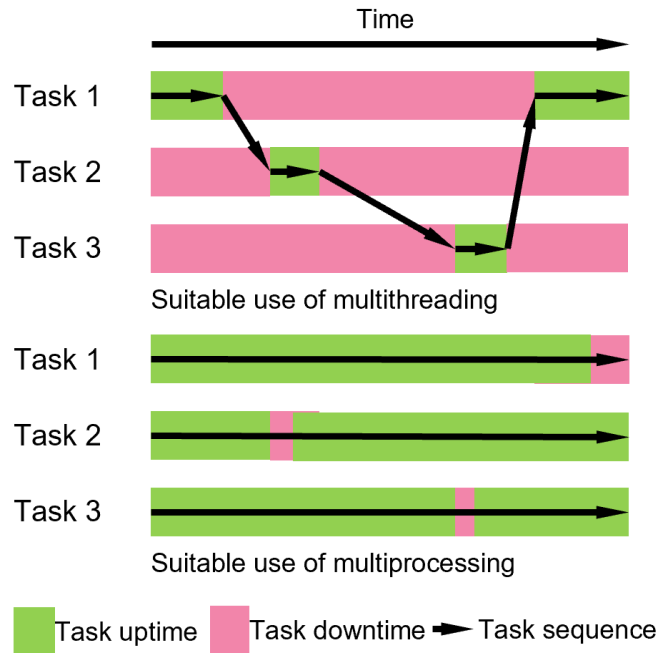


Figure 2.6: Suitable use cases for multithreading and multiprocessing.

To align multithreading with the MPCA system hardware, the control software is divided into six threads: the main thread (Th0), the temperature control thread (Th1), the capacitive detector readout thread (Th2), the valve control thread (Th3), the pump control thread (Th4), and the AiPD readout thread (Th5). Each thread operates using its own dedicated library that is specifically designed to interface with and control the corresponding electronics. The main thread manages the other threads, handles communication with the UI, monitors overall progress, saves readout data, and provides centralized control to start or stop all the threads as needed (Fig. 2.7).

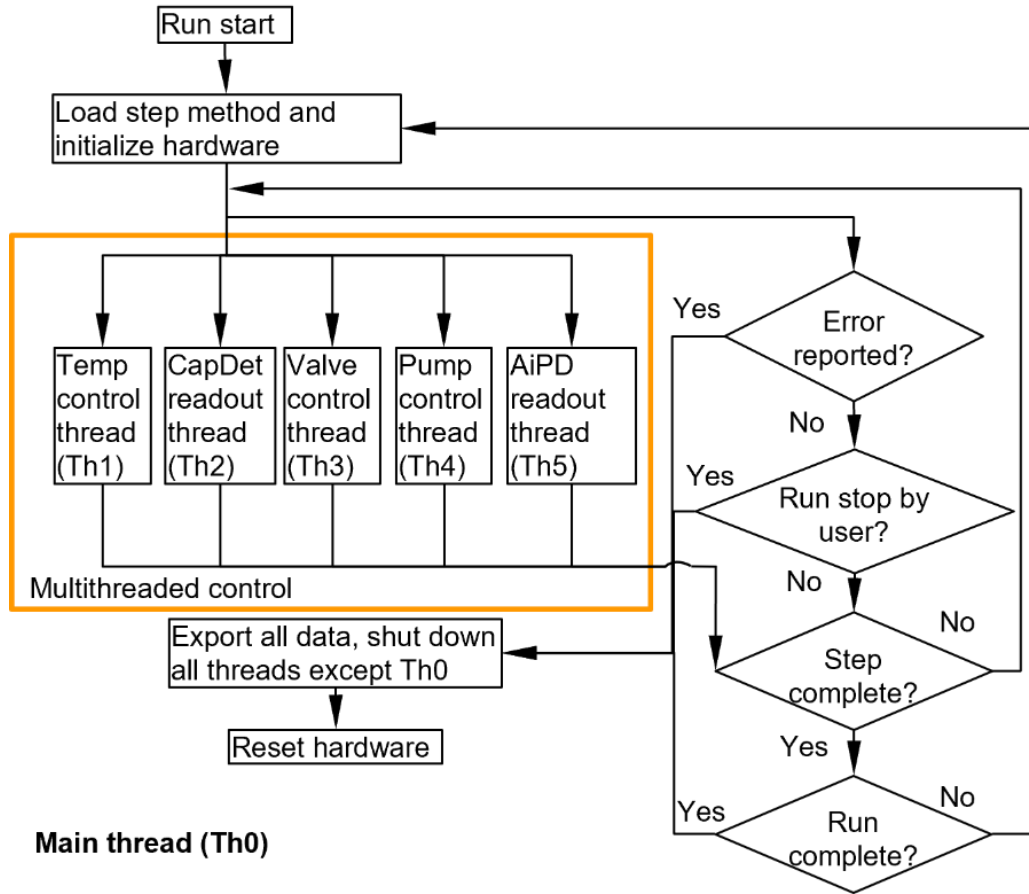


Figure 2.7: Flow charts of the main thread Th0.

2.2.3.2 . Concurrent Threads Th1-Th5

The Th1 thread manages the closed-loop controls of all the eight temperature channels that are read using two ADC chips (Fig. 2.8). In a control loop, it first initiates data conversion on one channel of each ADC chip and then enters sleep mode, awaiting the data conversion while allowing other threads to progress. Once the prescribed data conversion interval has elapsed, the Th1 thread reads the output data. If a channel is designated to perform active heating, its temperature data is used to compute the parameter that controls the heating power (as described in Section 2.3.4). The process is repeated for the four channels of each ADC chip, and the total duration of the four channel process is recorded as the current loop execution time. The

Th1 thread then re-enters sleep mode for a duration equal to the difference between the predetermined loop cycle time (which is 100.0 ms) and the execution time (which is typically between 62.5 ms and 100.0 ms). The Th1 thread also monitors any anomalous temperature readings, which indicate system malfunction that may lead to hardware damage if left uncorrected. In this process, for each temperature channel, if anomalous temperature readings occur consecutively over an excessive count (here set at three consecutive readings), this thread terminates all the heating and temperature readout actions; it also alerts the Th0 thread, which then terminates all the threads.

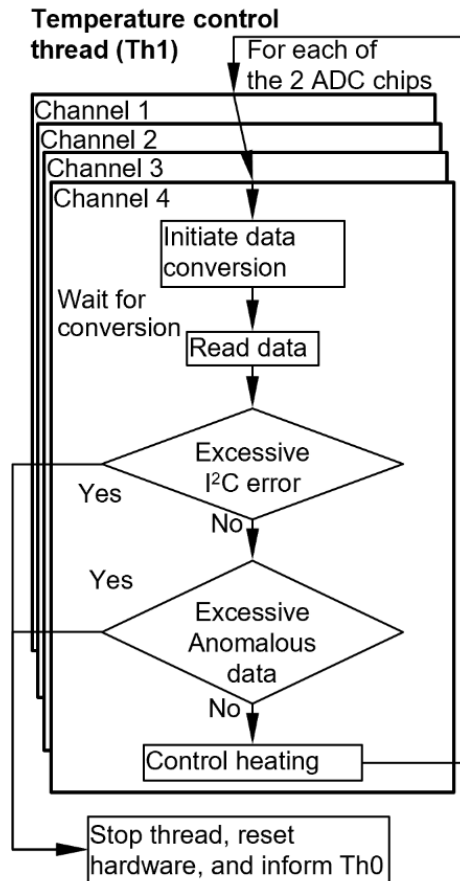


Figure 2.8: Flow charts of the temperature control thread.

The Th2 thread manages the readout of the two CapDet elements in each cell. The two CDC chips for CapDet readout are programmed to operate in the continuous data conversion mode, which automatically initiates a new data conversion immediately after the preceding one completes (Fig. 2.9). After initiation, each CDC chip is read at a fixed interval, i.e., cycle time, of 110.0 ms, which is slightly longer than the conversion interval of 109.6 ms (Table 1).

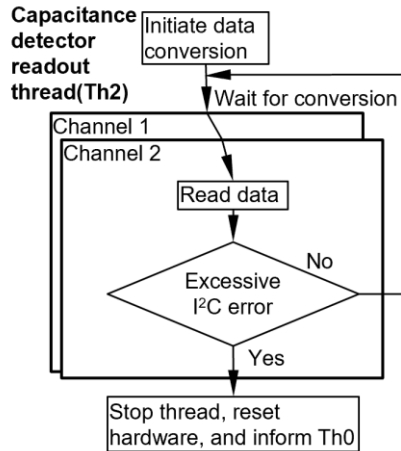


Figure 2.9: Flow charts of the capacitive detector readout thread.

The Th3 thread controls the 6 latching valves (Fig. 2.10). The valves are actuated sequentially using voltage pulses of 50.0 ms duration. The cycle time of Th3 is set at one second in order to reduce the computational burden (Table 2.1).

Table 2.1: The IC model and conversion time used for each type of readout.

Readout type	IC model	OCCT (ms)	Time multiplexed?	ECCT (ms)	PLCT (ms)
Temperature	ADS1115	15.6	Yes (4 channels)	62.4	100.0
CapDet	AD7746	109.6	No	109.6	110.0
Pressure head	LTC2493	146.9	Yes (2 channels)	293.8	400.0
AiPD	LTC2493	146.9	No	146.9	200.0

Abbreviations. OCCT: one channel conversion time; ECCT: effective channel conversion time (depending on whether multiple channels are time-multiplexed); PLCT: programmed loop cycle time.

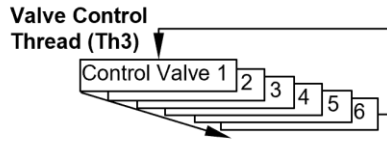


Figure 2.10: Flow charts of the valve control thread.

The Th4 thread controls the three pumps (Fig. 2.11). For the sampling pump, the thread sends an on/off signal to start or stop the sampling pump. Then, for the separation pumps, it initiates data conversion for pressure head readout on one of the two ADC channels, and then enters sleep mode. After the prescribed data conversion interval has elapsed, the Th4 thread reads the output pressure head from this channel and computes the actuation frequency required for its corresponding separation pump control (as described in Section 2.3.4). This process is then repeated for the other ADC channel. To accommodate the total required conversion time of both channels, which is 293.8 ms, the loop cycle time of the thread is set at 400.0 ms (Table 2.1). Such a much longer loop cycle time leaves enough margin for any algorithmic processing that may potentially be required.

The Th5 thread controls the AiPD power supply and readout of the AiPD (Fig. 2.12). In its loop, it first sends an on/off signal to the VUV lamp. Next, it initiates data conversion on the ADC, and then enters sleep mode. After the prescribed data conversion interval, the Th5 thread reads the output data. To accommodate the data conversion time of 146.9 ms, the cycle time of the thread is set at 200.0 ms (Table 2.1). Although this cycle time can be reduced, it is set at 200 ms for consistency with the Th4 thread, where the same ADC model is used but with two time-multiplexed input channels.

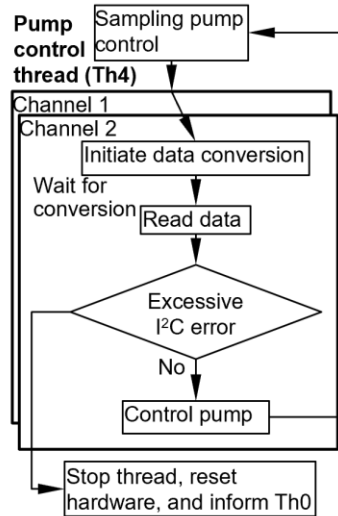


Figure 2.11: Flow charts of the pump control thread.

Within each run step in the operation method, the Th2–Th5 threads are programmed to be active only when their specific functions are required. Otherwise, these threads enter sleep mode. For example, during the sampling step where users typically do not need detector readouts, both the Th2 and Th5 threads enter sleep mode. When in sleep mode, a thread performs no action other than checking periodically – once per cycle time – to determine whether it needs to resume activity. Sleep mode conserves computational power and energy expenditure.

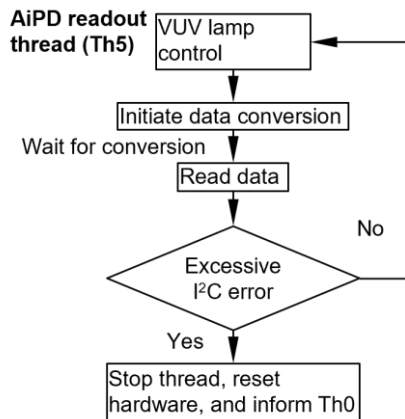


Figure 2.12: Flow charts of the AiPD readout thread.

2.2.3.3 Thread Timing

The Raspberry Pi OS, which is the main operating system on RPi, is not a real time operating system (RTOS). Consequently, programmed events are not always executed at precisely scheduled time points [Ham14]. The timing errors can accumulate and consequently affect the accuracy of the chromatograms. To address this, two solutions are implemented. Firstly, the loop cycle time in each thread is set to be sufficiently longer than the time required by the hardware (Table 1), as described above. Secondly, all readout data are timestamped relative to the start of each run step. This allows back-end chromatogram processing (although not performed for this work) to interpolate the data, thereby compensating for any accumulated timing error, and obtaining time-corrected chromatograms.

2.2.3.4 Control of Temperature and Flow

In this work, all the temperature and pressure head channels are closed-loop controlled to follow user-defined transient temperature and pressure head profiles. After reading out the temperature or pressure head data, the corresponding output control is computed using a

proportional-integral-derivative (PID) algorithm. For a parameter x under control, a unified transfer function of the control is as follows:

$$C = C_{\max} (P \cdot \Delta x_n + I \cdot \Delta x_{\Sigma} + D(\Delta x_n - \Delta x_{n-1})) \quad \dots\dots(2.1)$$

Where C is the control signal strength; C_{\max} is the maximum available signal strength; P, I, D are the proportional, integral, and derivative coefficients, respectively; Δx_n is the difference between the target value of x and the measured value of x , both at the current (i.e., most recent) measurement time, n . The measurement time is discretized in 0.1 s intervals; Δx_{Σ} is the accumulated difference between the target value of x and the measured value of x since the beginning of each run step; Δx_{n-1} is the difference between the target value of x and measured value of x , both at the preceding (i.e., second most recent) measurement time. In temperature control, x represents the temperature. Further, for the PWM-based control, C represents the output duty cycle with a maximum available value (i.e., C_{\max}) at 100%. For the analog voltage control via a potentiometer, C represents the output resistance, whereas C_{\max} represents the maximum available output resistance of the potentiometer. In pressure head control, which effectively controls the flow rate, C represents the actuation frequency for a separation pump, C_{\max} represents the maximum available actuation frequency, whereas x represents the pressure head.

The square waves required for PWM and for pump actuation may be generated using either hardware PWM or software PWM to provide the timing pulses. Hardware PWM supports precise timing control over a large range of frequency (up to 30 MHz) [Zho16]. In contrast, software PWM is less precise in timing, because its pulses are generated using a software clock. Additionally, it can only support a lower frequency range (up to 40 kHz) [Zho16]. Despite the superiority of the hardware PWM, on the RPi it is limited to only 4 GPIO pins, which are

insufficient for all the channels in the MPCA system. Considering the relatively large thermal time constants in the heating of the MPCA components (typically >0.5 s), all the temperature control channels are set to use software PWM. This software PWM operates at a frequency of 50 Hz with an 8-bit duty cycle resolution. Conversely, for the two separation pumps that require 0–1000 Hz frequency control, hardware PWM is selected. It provides a square wave at 50% duty cycle with a frequency control resolution down to a millionth of the set frequency. In this work, both types of PWM are managed by the *pigpio* library.

2.2.3.5 Communication between the EC and the electronics

As I²C bus is shared by the EC and multiple electronic components, the communication on this bus may be affected by power glitches, noise, and interference. However, I²C communications use a handshake protocol, so the *Smbus* used in this work can detect a failed handshake and report an error. If left unaddressed, accumulated I²C errors can prematurely end a thread without implementing proper hardware controls. This is especially risky during component heating, which may eventually damage the hardware irreversibly. Nevertheless, tolerating a limited number of I²C errors can sometimes be beneficial, especially at early stages of hardware development. Therefore, in the threads that use I²C, namely Th1, Th2, Th4, Th5, an error handling protocol is implemented as follows. Each thread continuously monitors its cumulative I²C errors. If the accumulated number of I²C errors in a thread exceeds a pre-determined threshold (here set at 3), the I²C communication is deemed unreliable. As a result, this thread terminates its associated hardware operation and alerts the main thread, which then shuts down all the other threads and saves the data.

2.2.3.6 Data Saving and Storage

The control software is designed to automatically export all raw data, including timestamps and values of the readout data points, into a comma-separated value (CSV) file at the end of each run step. CSV files are compatible with a wide range of commonly used software, such as Microsoft Excel, MATLAB, and Python. After an analytical run, the raw data files that are generated can be readily used for automated peak detection [Zha19] and subsequently for automated chemical detection [Xu23b]. Additionally, after each run, the control software automatically saves as a metadata file the operation method file that contains the entire set of operation parameters used for the run (which is further described in Section 2.3.8). This file also serves as a record of the exact operating parameters used for the run, which is important for systems where the operating parameters are frequently adjusted by users. For clear and convenient record-keeping, all data and metadata files of a particular run are saved in a dedicated folder that is uniquely named with the hardware serial number and the date and time of the run.

2.2.3.7 Low Power Mode

For applications with constrained power sources, *e.g.*, small-size batteries or solar panels, μ GC systems must operate within a limited power budget. Therefore, this work also investigates software approaches to reduce the power consumption of the MPCA system while maintaining essential hardware control capabilities. Table 202 shows various options for reducing power consumption by turning off certain RPi functions. Among these, the most significant power saving is achieved by turning off the Ethernet and USB functions. By implementing a low power mode with all these options active, a total reduction in power consumption of 1.6–1.8 W is achievable.

Table 2.2: Normal mode and low power mode power consumption of the RPi (Model 3B+ in this work)

Functionalities	Normal mode	Low power mode savings
Bluetooth + Wifi	On	Off, 0.2 W reduction
Soundboard	On	Off, 0.2 W reduction
Ethernet + USB	On	Off, 1.2 W reduction
HDMI	On	Off, 0.15 W reduction
Energy usage (idle state)	2.8-2.9 W	1.6-1.8 W reduction 1.1-1.2W total consumption

In the low power mode, the MPCA system must be able to perform the analytical runs without Ethernet communication to the UI. Upon starting, the control software executes a script to terminate Ethernet and USB functions, then operates the MPCA system based on the previously transferred operation method file. After the prescribed runs are complete, the control software executes another script to restore Ethernet and USB functions, allowing users to configure the system for the next series of runs. Turning off the other RPi functions in the low power mode does not affect the system function in any way.

2.2.3.8 UI Design and Communications to the EC

The user interface (UI) of the MPCA system allows users to configure the operation method and to start and stop the system operation; it also provides a real-time display of the data. The UI operates in a Windows (Microsoft Corporation, Redmond, WA, USA) environment using the Windows Presentation Foundation (WPF) [Mic23], which is a proven framework with strong open-access library support. The UI is programmed in *C#* and *Extensible Application Markup Language (xaml)*, which are the primary languages for WPF. In the UI, the main window contains plots for the temperature profiles, pressure head profiles, and the chromatograms. The plotting of real time readout data (Fig. 2.12) is Implemented using LiveCharts [Liv23]. To

manage resource demands of plotting, each data stream is accompanied by a checkbox, which allows users to select whether to plot the data stream. After each run step, the plots are automatically cleared and prepared for the next run step.

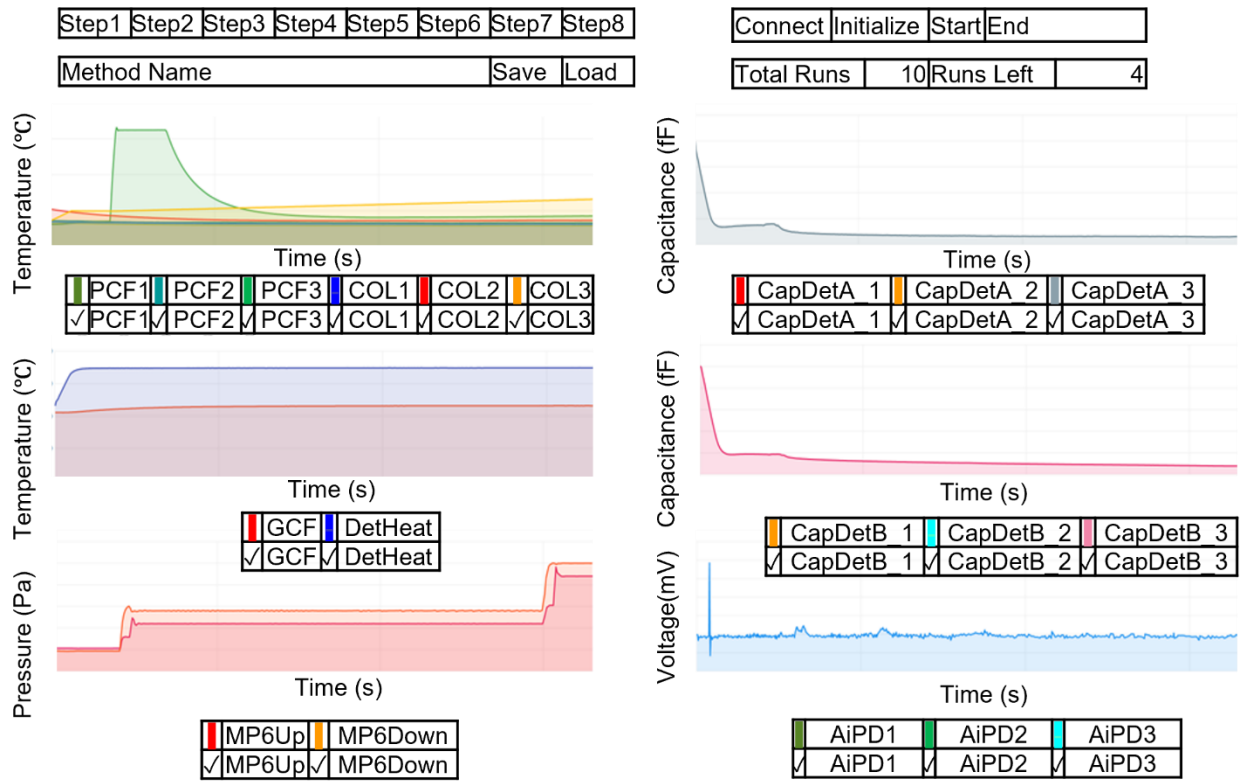


Figure 2.13: Graphical representation of the main UI window with real time plotting of data. The method configuration page is implemented for the user to customize the operation parameters of each run step (Fig. 2.14). This page is implemented to pop up when the user clicks the corresponding button of a step from the main UI window. This page also enables the user to save the current method and load it for future use.

	Ramp Start Time (s)	Ramp End Time (s)	Heating End Time (s)	Initial Temp. (°C)	Ramp Target Temp. (°C)	Enabled
Preconcentrator1	0	0	0	0	0	False
Preconcentrator2	15	20	35	20	165	True
Preconcentrator3	0	0	0	0	0	False
Column1	0	0	0	0	0	False
Column2	20	398	500	30	70	True
Column3	0	0	0	0	0	False
Carrier Gas Filter	0	0	0	0	0	False
Detector Heater	1	5	500	20	40	True
Test Chip Heater	0	0	0	0	0	False

	Open Time	Close Time	Enable
Valve1	-1	-1	False
Valve2	1	498	True
Valve3	-1	-1	False
Valve4	-1	-1	False
Valve5	-1	-1	False
Valve6	-1	-1	False

	Start Time (s)	End Time (s)	Enable
AiPD1	0	0	False
AiPD2	0	500	True
AiPD3	0	0	False
AiPD Lamp	0	500	True

Sampling pump	Start	Stop	Duty Cycle	Enable
	-1	-1	-1	False

Upstream separation pump	Start time (s)	20	260	-1	-1	Enable	True
	End time	-1	498	-1	-1	Closed Loop	True
	Freq/Press	500	1700	-1	-1		
Downstream separation pump	Start time (s)	20	260	-1	-1	Enable	True
	End time	260	498	-1	-1	Closed Loop	True
	Freq/Press	900	2000	-1	-1		

Total Time (s)	500
----------------	-----

Enable Step	True
-------------	------

	Start Time	End Time	Enable
CapDetA_1	0	0	False
CapDetA_2	0	500	True
CapDetA_3	0	0	False
CapDetB_1	0	0	False
CapDetB_2	0	500	True
CapDetB_3	0	0	False

Figure 2.14: Graphical representation of UI for customizing the operation method. The entries in blue are representative values filled by a user for the Cell2 separation step.

The operation parameters of all the run steps set up in the UI are arranged into an operation method file in the JSON format [JSO23], which is both machine-readable and comprehensible to users. The widely used open-source *Newtonsoft* library is used to generate this file from the parameters entered into the UI. For the MPCA system, this file is typically larger than the TCP/IP receive buffer in the EC. Therefore, for transmitting it to the EC, this file is first segmented into smaller packets, which are then individually sent to the EC and finally reassembled into the full operation method file. The EC then reads the operation method file and prepares for Initiating the MPCA system operation. During automated runs, the EC sequentially executes each run step based on the operation method file; in this process, the system operation

is controlled by the EC and does not rely on the UI. The operation method file is also saved by the EC as the metadata for the corresponding MPCA system run.

In the normal operation mode, as opposed to low power mode, real time data transfer between the EC and the UI is performed. This includes sending the readout data from the RPi to the UI and sending the user commands from the UI to the EC. For the readout data, the UI issues requests for the latest real time data by sending a simple string to the EC. To avoid overburdening the UI, the frequency of these requests is programmed to vary inversely with the number of data streams already plotted. The EC then replies with the readout data in a string, which is parsed by the UI. The user commands include those to initialize the TCP/IP connection, to start the runs, and to stop the runs. These commands are sent to the EC in the form of a simple string immediately after being issued from the UI .

2.2.3.9 Detailed UI Implementation

The graphical user interface (UI) provides the following functions: 1) facilitating the user in setting up run parameters; 2) saving and loading run parameters, 3) real-time plotting of data, and 4) issuing start and stop commands. In the main page (Fig. 2.15), the eight step configuration buttons “STEP1”-“STEP8” in the upper left area provides entries to the method configuration pages of up to eight individual steps that a single run may consist of. A textbox underneath these eight buttons is used to input the operation method file name. The “Save” and “Load” buttons are used to save the currently set run parameters into the operation method file and to load a previously configured operation method file.

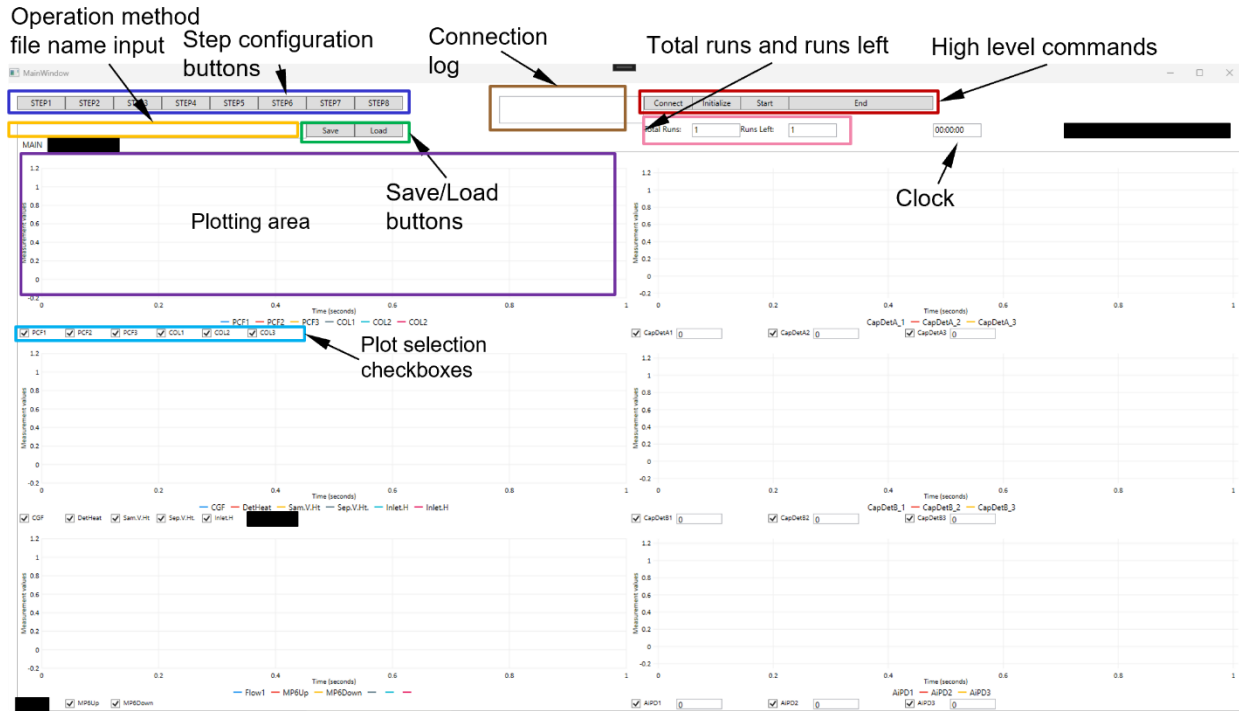


Figure 2.15: Main page of the UI. The obsolete or unrelated contents are covered by black boxes.

The four buttons in the upper center area of the UI main page (Fig. 2.15) are used to issue high-level commands to the embedded computer (EC). The “Connect” button is used to establish a TCP/IP connection with the EC. The result of the connection attempt (*i.e.*, successful or unsuccessful connection) is displayed in the textbox on the left side of the button. After the connection is established, pressing the “Initialize” button sends the run parameters (in the form of the operation method file) to the EC, which then automatically loads the run parameters for all the steps. Next, the “Start” button can be pressed to execute the run process, which is completely automated and controlled by the EC. The UI does not control the run automation process. During a run, if at any point the user needs to stop the run, the user can press the “End” button. Once the EC receives the “End” command, the EC terminates communication with the UI and saves all data.

If the EC encounters any unresolvable error, a message will be sent from the EC to the UI. A pop-up window will appear on the UI, notifying the user of the error. The EC saves all existing readout data, shuts down the MPCA hardware and stops communicating with the UI.

Underneath these four high-level control buttons is a “Total Runs” textbox, which is used to configure how many runs should be repeated. The neighboring “Runs Left” textbox is used to display how many runs are remaining. On the right side of the “Runs left” textbox is a clock showing the elapsed time since the beginning of the runs. Further on the right is a “Low Power Mode” checkbox, which is used to enable system operation in the low power mode.

The six plots that constitute the majority of the UI main page are used to show the real-time readout data. The upper two plots in the left column are for temperature readouts; the bottom plot in the left column is for pressure readouts; the upper two plots in the right column are for capacitive detector readouts; the bottom plot in the right column is for AiPD readouts. Underneath each plot are checkboxes to select or unselect which individual component(s) to plot. The checkboxes can be selected or unselected at any time during the run.

To set up a typical set of runs (Fig. 2.16), the user first starts the EC software and then the UI software. If a previously configured operation method needs to be modified, the user can click any of 8 step configuration buttons, which opens the step configuration window for the selected step (Fig. 2.17). After the user edits the parameters, the user can save the step parameters by clicking the “Save” button in this window, which then automatically closes and returns the user to the main window. After all steps are configured and the total number of repeated runs entered, the user can click the “Connect” button to connect with the EC, then click the “Initialize” button to load all the parameters, and finally click the “Start” button to start the automated control process on the EC side. The UI communicates regularly with the EC to fetch

the most recent data. The step configuration buttons change to green color when the corresponding steps are completed or ongoing in a run and reset to grey after a run (Fig. 2.18).

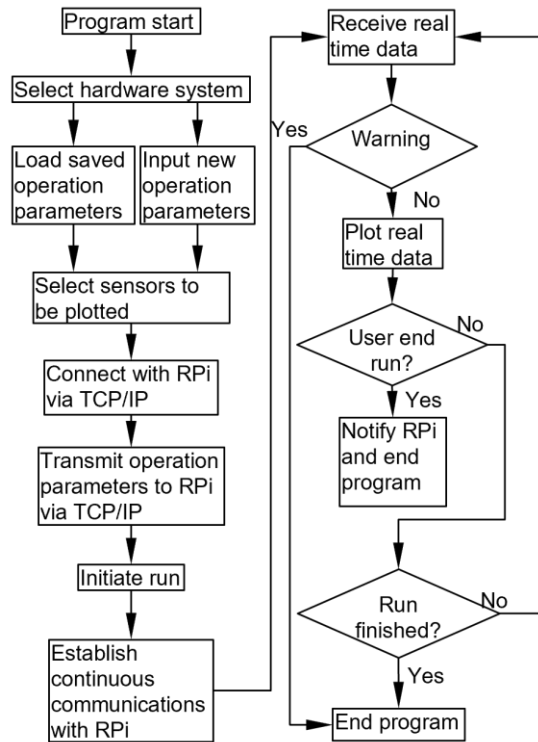


Figure 2.16: Flowchart of user interface operations.

MPCA Step 1

C/T	Ramp Start Time	Ramp End Time	Heating End Time	Initial Temperature	Ramp Target Temperature	Enable
PCF1	0	0	0	0	0	<input type="checkbox"/>
PCF2	0	0	0	0	0	<input type="checkbox"/>
PCF3	0	0	0	0	0	<input type="checkbox"/>
COL1	0	0	0	0	0	<input type="checkbox"/>
COL2	0	0	0	0	0	<input type="checkbox"/>
COL3	0	0	0	0	0	<input type="checkbox"/>
CarrierGasFilter	0	0	0	0	0	<input type="checkbox"/>
DetectorHeater	0	0	0	0	0	<input type="checkbox"/>

Valve Open Time	Valve Close Time	Enable	
Valve1	-1	-1	<input type="checkbox"/>
Valve2	-1	-1	<input type="checkbox"/>
Valve3	-1	-1	<input type="checkbox"/>
Valve4	-1	-1	<input type="checkbox"/>
Valve5	-1	-1	<input type="checkbox"/>
Valve6	-1	-1	<input type="checkbox"/>

AIPD Start Time	AIPD Stop Time	Enable	
AIPD1	0	0	<input type="checkbox"/>
AIPD2	0	0	<input type="checkbox"/>
AIPD3	0	0	<input type="checkbox"/>

Lamp Start	Lamp End	Enable	
AIPD Lamp	0	0	<input type="checkbox"/>

Capacitor Meter Start Time	Capacitor Meter End Time	Enable	
CapDetA_1	0	0	<input type="checkbox"/>
CapDetA_2	0	0	<input type="checkbox"/>
CapDetA_3	0	0	<input type="checkbox"/>
CapDetB_1	0	0	<input type="checkbox"/>
CapDetB_2	0	0	<input type="checkbox"/>
CapDetB_3	0	0	<input type="checkbox"/>

Pump Start Time	Pump End Time	Duty Cycle	Enable	
SampPump	-1	-1	0	<input type="checkbox"/>

MP6Up Start time	MP6Up End time	MP6Up Freq/Pressure	MP6Down Start time	MP6Down End time	MP6Down Freq/Pressure	Enable	CloseLoop
-1	-1	-1	-1	-1	-1	<input type="checkbox"/>	<input type="checkbox"/>
-1	-1	-1	-1	-1	-1	<input type="checkbox"/>	<input type="checkbox"/>

Total Time: 0 Enable Step

Confirm Save

Figure 2.17: Method configuration page for editing the operation parameters of a run step. (The obsolete or unrelated contents are covered by black boxes)

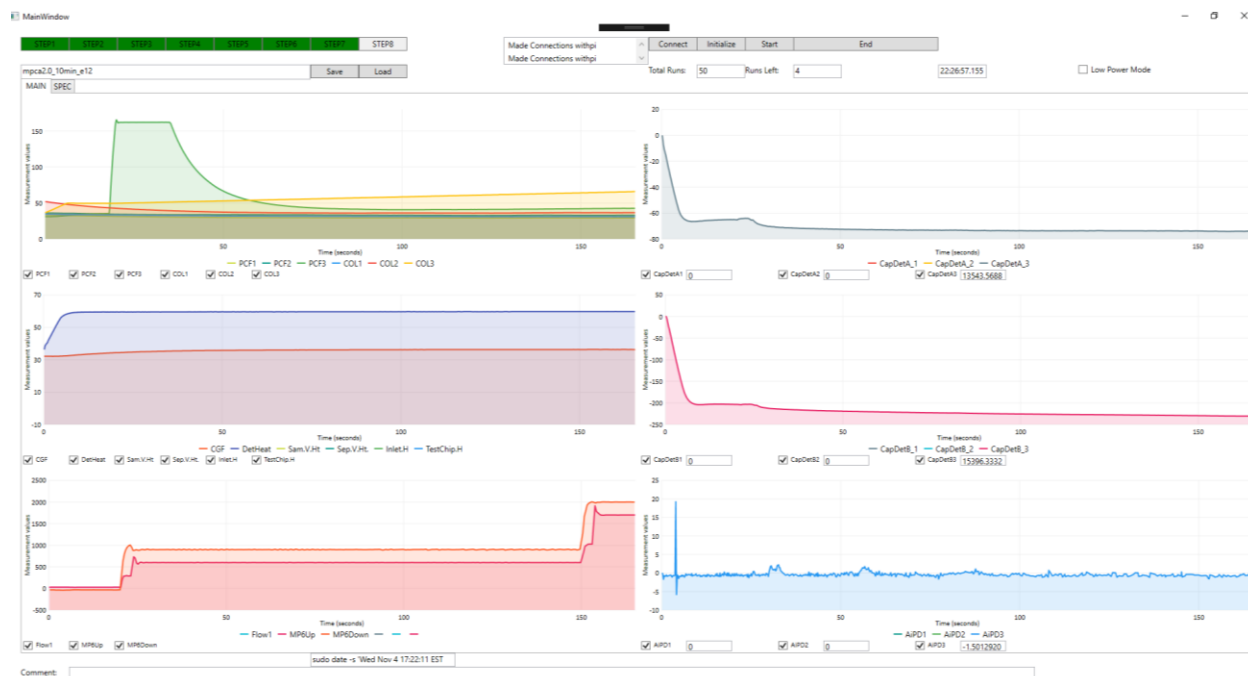


Figure 2.18: Main window during operation with real-time plotting of the readouts.

2.3 Experimental Validation

The control software has been deployed in multiple generations and multiple units of the MPCA system [Lia23] and has supported >1000 analytical runs, which have analyzed >40 different VOCs in various mixtures. Additional results of chemical analyses conducted using this control software have been previously reported [Lia23]. The details of the system setup for experimental evaluation are described in Supporting Information Section S3. The performance of the control and readout can be assessed by the readout data from a typical run, as described below.

The efficacy of the closed-loop control was confirmed by verifying the measured temperature and pressure head profiles during the separation steps. For example, for the separation step performed in Cell2, Preconcentrator2 was set to ramp from 20 °C to 165 °C from 15 s to 20 s, to maintain a steady temperature at 165 °C from 20 to 35 s, and then to cool down naturally; Column2 was set to ramp from 20 °C to 70 °C from 20 s to 398 s and to maintain at 70 °C till 500 s; the detectors were set to heat up to 40 °C by 20 s and to maintain this temperature till 500 s. In a representative set of results, all these components adhered closely to their programmed temperature profiles (Fig. 14a). Preconcentrator2 reached 93% of the target temperature at 20 s and remained within $\pm 1\%$ of the target temperature during the maintaining period. Column2 tightly followed its target temperature profile, showing an average deviation of only 0.38 °C. The region of the detectors also tightly followed its target temperature profile, showing an average deviation of only 0.02 °C.

A parasitic temperature rise was observed in the unheated components, including Preconcentrator1, Preconcentrator3, Column1, Column3, and the carrier gas filter. This phenomenon was anticipated, because both the heated and unheated components were

monolithically integrated in close proximity on the same MPCA chip, which had incomplete thermal isolation. Nevertheless, this parasitic temperature rise did not adversely affect the system operation.

For the separation step performed in MPCA Cell2, each flow control loop – both the upstream and downstream – was programmed for a two-stage flow rate pattern: an initial modest flow rate, followed by a stronger flow. The initial flow was used to provide an amenable flow rate for separation, whereas the stronger flow was used to expedite the elution of the low-volatility chemicals in the cell. For example, the upstream loop was set to provide a pressure head at 500 Pa during 20–260 s and 1700 Pa during 260–500 s. From the pressure readout, the upstream pressure head stabilized by 30 s into a range of 485–513 Pa, which was within $\pm 3.0\%$ of the target pressure. Following the scheduled increase at 260 s, the upstream pressure head stabilized by 275 s within a range of 1704–1696 Pa, i.e., within $\pm 0.3\%$ of the target pressure. Similar control performance was also obtained for the downstream loop (Fig. 2.19a).

For the separation step in Cell3, the operation parameters were set up in a similar manner as those for Cell2, albeit with slightly different values. Both heating and pressure head controls in Cell3 provided similar performance as Cell2 (Fig. 2.19b). The unheated components also showed unremarkable temperature responses. For example, Column2 started at an elevated temperature because it was heated in the preceding separation step in Cell2; as it was not actively heated during the current separation step in Cell3, it gradually cooled down during this step.

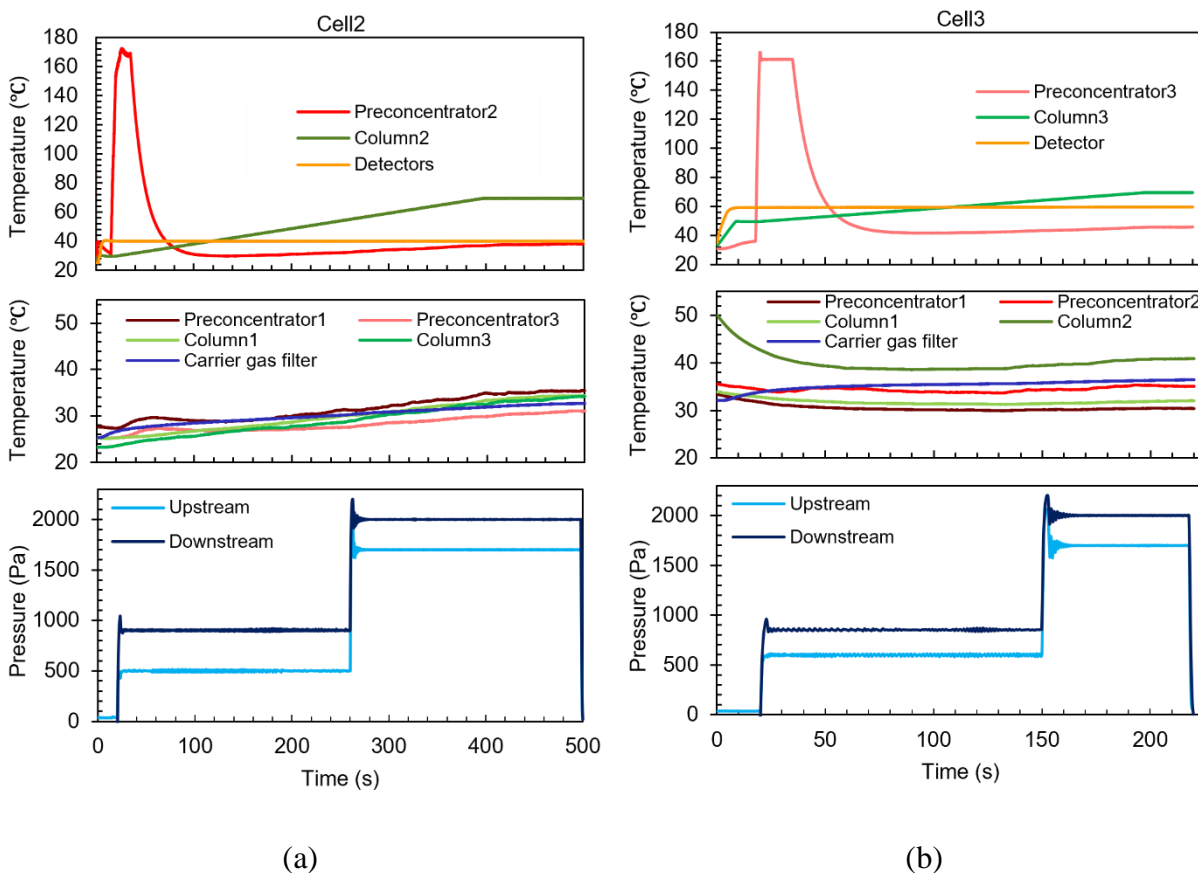


Figure 2.19: Typical temperature and flow control results in the MPCA system during (a) the Cell2 separation step and (b) the Cell3 separation step.

The effective control of the MPCA system was further validated by the chromatograms generated in analytical runs. In one such run, the MPCA system was tested with a mixture that included 18 detectable analyte chemicals, including cyclohexane, pentanal, heptane, pinacolyl alcohol, methyl isobutyl ketone (MIBK), toluene, butyl acetate, ethylbenzene, m-xylene, o-xylene, 1-chloropheptane, nitrobenzene, mesitylene, decane, 2-nonanone, undecane and dodecane. All the chemicals were prepared at a concentration of 200 parts per billion (ppb), except for o-xylene, which was present at an unspecified concentration. The sampling time was set by the user to be 10 min.

After the run, the exported data successfully provided an ensemble of chromatograms (Fig. 2.20). The analyte chemicals presented various degrees of separation, such as pairs with complete baseline separation (*e.g.*, between 1-chloroheptane and nitrobenzene), partial separation (*e.g.*, between ethylbenzene and m-xylene), and complete overlap (*e.g.*, between pinacolyl alcohol and MIBK). It was evident that the chromatograms obtained from the three different types of detectors were complementary in nature, even within each cell. For example, in Cell3 nonane produced relatively small positive CapDetA and AiPD peaks and a relatively large negative CapDetB peak, whereas ethylbenzene produced a relatively small positive CapDetA peaks, almost no CapDetB peak, and a relatively large AiPD peak.

Most peaks showed peak widths >10 s, which were much larger than the data acquisition rate of any detector. Therefore, each peak was well profiled by at least 50 data points. Even for those peaks that were partially separated within a narrow time window (*e.g.*, the three peaks in Cell3 from 17 s to 20 s), their profiles were well represented by >5 data points per peak. The available time resolution provided accurate and undistorted acquisition of the peak profiles, allowing recognition and quantification of the chemicals.

To examine the data acquisition rate more closely, readout data streams in the Cell2 separation step are used here as a case study. In this step, the data stream of each readout transducer type (*i.e.*, temperature, CapDet, pressure head, or AiPD) contained >1200 data points. For each readout type, the actual time interval between measurements (TIBM) was obtained as the difference in the recorded timestamps between every two consecutive data points (Fig. 2.21). Across all the readout types, the standard deviation of TIBM was $<0.5\%$ of the mean TIBM value, demonstrating timing stability (Table 2.3). This low variation in TIBM is desired, as it contributes to repeatable control of temperature and flow from run to run, which in turn

contributes to repeatable system results. Additionally, the mean TIBM values for all readout types closely matched the intended cycle times, with <0.1% discrepancy in the temperature readout, <0.5% discrepancy in the CapDet readout, and <1.2% discrepancy in the pressure head and AiPD readouts. These discrepancies are further discussed in Section 4.

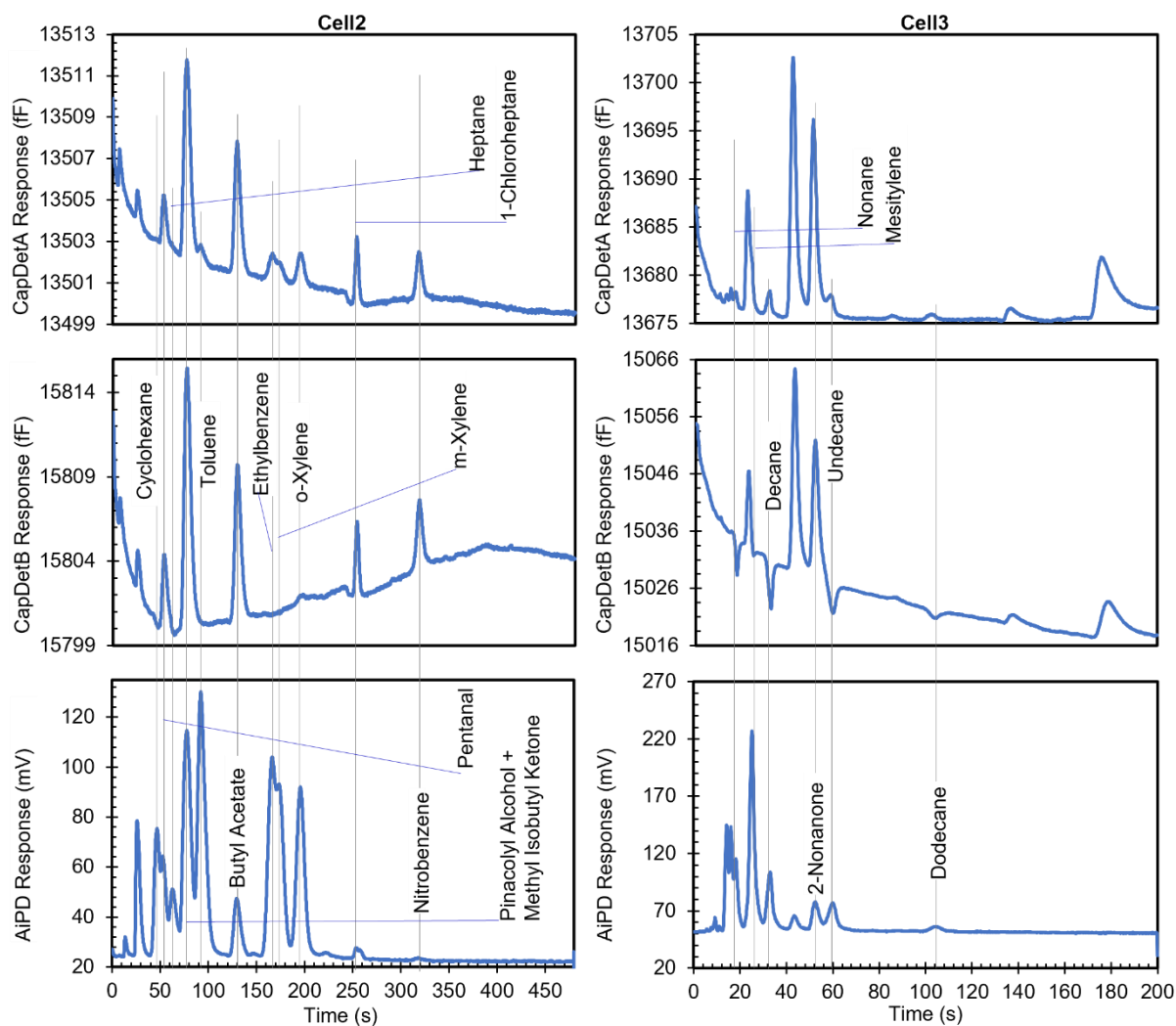


Figure 2.20: Chromatogram of a mixture containing cyclohexane, pentanal, heptane, pinacolyl alcohol, toluene, methyl isobutyl ketone, butyl acetate, ethylbenzene, m-xylene, o-xylene, 1-chloroheptane, nitrobenzene, nonane, mesitylene, decane, 2-nonanone, undecane and dodecane. Currently, the chromatograms are only presented for Cell2 and Cell3, as Cell1 is still under hardware development.

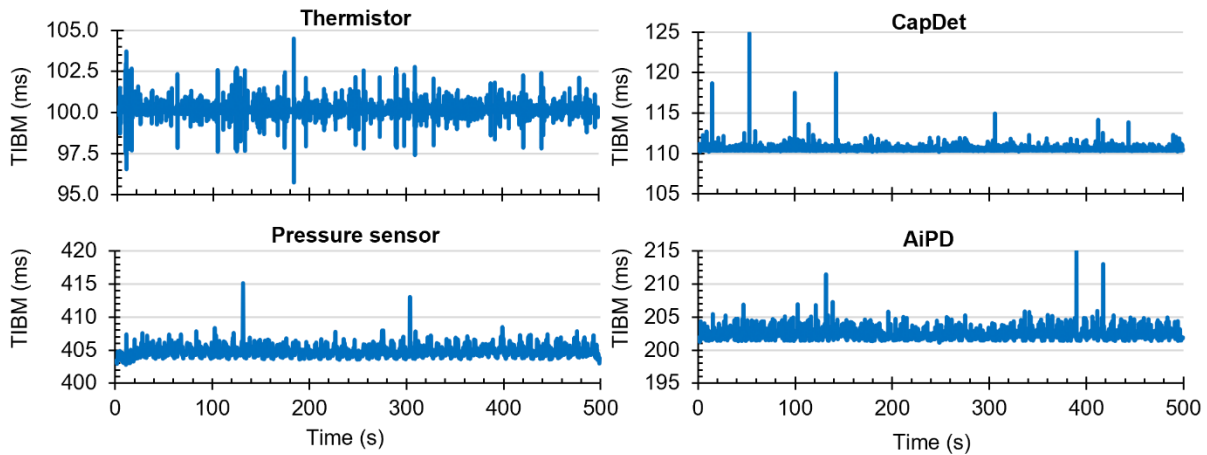


Figure 2.21: The typical experimentally measured time interval between measurement (TIBM) for each type of readout over time (measured during the MPCA Cell2 separation step). The number of data points in each plot is greater than 1200.

Table 2.3 Actual time interval between measurements (TIBM) of all the readout component types (measured at the peak software workload step, *i.e.*, the Cell2 separation step).

Readout type	Intended loop cycle (ms)	Actual mean TIBM (ms)	TIBM standard deviation (ms)
Temperature	100.0	100.1	0.4
CapDet	110.0	110.5	0.4
Pressure head	400.0	404.7	1.0
AiPD	200.0	202.3	0.9

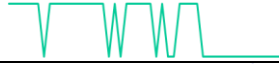



The noise level of each readout was also experimentally measured and compared to the rated noise in the converter datasheet (Table 4). The measured RMS noise values for the temperature, CapDet, pressure head, and AiPD readouts were 32.0 μV , 53.0 aF, 66.9 μV and 107.1 μV , respectively.

The measured noise of the temperature readout was comparable to that rated in the datasheet. According to the datasheet, the voltage resolution that corresponds to a least significant bit (LSB) of the ADC, under the settings in this work, is 62.5 μV , which aligns with the rated RMS noise. As shown by the measured noise waveform (Table 4), the measured

voltage of the temperature readout typically toggled between two discretized values with an increment that corresponded to an LSB of the ADC. This observation not only confirmed the datasheet ratings but also indicated that the temperature readout resolution was limited by the LSB or the quantization error of the ADC.

In contrast, the CapDets measured much higher noise than that rated in the CDC datasheet. Similarly, the pressure head and AiPD readouts (both using the same ADC model) measured noise levels that significantly exceeded that rated in the datasheet (Table 2.4). These observations indicated that the resolutions of these readouts were not limited by the quantization errors of the ADC and CDC models. Instead, they appeared to be limited by the overall noise of the system, which possibly resulted from various sources such as the noise in the reference voltage, noise in the power supply, interference, and the amplifier. Additionally, the elevated noise observed in the CDC readings was likely caused by the large baseline capacitances of the CapDets.

Table 2.4 Noise data of the detectors

Readout type	Noise (RMS)		Typical noise waveform (over 5 s duration)
	Measured	Datasheet	
Temperature	32.0 μV (0.04 $^{\circ}\text{C}$)	62.5 μV^*	 64.1 μV (0.85 $^{\circ}\text{C}$)
CapDet	53.0 aF	4.0 aF	 219.8 aF
Pressure head	66.9 μV (0.15 Pa)	0.6 μV^*	 249.3 μV (0.55 Pa)
AiPD	107.1 μV	0.6 μV^*	 400.1 μV

* Per datasheet, this number is constant over a range of data rates and corresponds to the voltage for the least significant bit

2.4 Discussion and Conclusions

Beyond the demonstrated functionalities, the control software may be improved in several ways. Firstly, ringing behaviors observed in the pressure head profiles likely resulted from the relatively long loop cycles (0.4 s) used for controlling the separation pumps. This problem can be mitigated by using an ADC model with a faster data rate.

Secondly, the recorded loop cycle times for pressure sensor and AiPD readouts were slower than the targets by 1% for all the data points. This was most likely because the control software assumed the computation times of these threads were negligible in the pressure sensor control thread (Th4) and the AiPD readout thread (Th5), whereas in fact the computation times of these threads were not negligible and cause some delay in these threads. In contrast, this issue was less pronounced in the temperature control thread, likely because of the implementation of a variable cycle wait time based on the measured computation time. In the capacitance readout thread Th2, the behavior of slower loop cycle time was also less significant, because the CDC chips operated in the continuous conversion mode, requiring only two I²C read operations per loop. For the pressure sensor control thread (Th4) and the AiPD readout thread (Th5), the approach of using variable cycle wait time can be implemented in the future.

In conclusion, the experimental results demonstrate that the control software reported in this work is successful in using multithreading to provide concurrent control of all aspects of the μ GC system operation, including temperature control, flow control, capacitive detector readout, AiPD readout, and user interface. The efficacy of the control software is reflected by the generated temperature profiles, pressure head profiles, and chromatograms, all aligning with the user-defined operation method and expectations. Despite the use of a non-real-time operating system, the variations in TIBM are small and show no adverse impact on system control. The

raw chromatogram data acquired can be readily processed first using an automated peak detection algorithm [Zha23] and then using an automated chemical recognition algorithm [Xu23b], culminating in a fully automated VOC analyzer.

With the demonstrated functionality for the MPCA system, the control software reported in this work shows great potential for applications in other μ GC systems. Its modular thread design can be easily expanded or adjusted to support different hardware configurations. In the future, real time data transfer between the RPi and the UI can be implemented wirelessly, *e.g.*, via Bluetooth or Wi-Fi. Additionally, the control software can be tested and assessed on a lower-power, single-core processor (*e.g.*, the RPi Zero) to further reduce power consumption.

References

- [Ada23] Lady Ada, “Introducing the Raspberry Pi Zero”, <https://cdn-learn.adafruit.com/downloads/pdf/introducing-the-raspberry-pi-zero.pdf> (accessed Jul 21, 2023).
- [Car72] G.C. Carle, R.W. Donaldson, S. C. Terry, K. D Wise, *Microminiature gas chromatograph*; NASA Tech Briefs, 1972.
- [Col15] W.R. Collin, A. Bondy, D. Paul, K. Kurabayashi, and E.T. Zellers, “ $\mu\text{GC} \times \mu\text{GC}$: Comprehensive Two-dimensional gas chromatographic separations with microfabricated components,” *Analytical Chemistry*, vol. 87, no. 3, pp. 1630–1637, 2015.
- [Gar15] A. Garg, M. Akbar, E. Vejerano, S. Narayanan, L. Nazhandali, L.C. Marr, M. Agah, “Zebra GC: A mini gas chromatography system for Trace-level determination of Hazardous Air Pollutants,” *Sensors and Actuators B: Chemical*, vol. 212, pp. 145–154, 2015.
- [Ham14] P. Hambarde, R. Varma, and S. Jha, “The survey of Real Time Operating System: RTOS,” 2014 *International Conference on Electronic Systems, Signal Processing and Computing Technologies*, 2014.
- [Hua21] X. Huang, M.W. Li, W. Zang, X. Huang, A.D. Sivakumar, R. Sharma, X. Fan, “Portable comprehensive two-dimensional micro-gas chromatography using an integrated flow-restricted pneumatic modulator,” *Microsystems and Nanoengineering*, vol. 8, no. 1, 2022.
- [JSO23] JSON, “Introducing json,” <https://www.json.org/json-en.html> (accessed Jul. 13, 2023).
- [Li10] M. Li, E.B. Myers, H.X. Tang, S.J. Aldridge, H.C. McCaig, J.J. Whiting, R.J. Simonson, N.S. Lewis, M.L. Roukes, “Nanoelectromechanical resonator arrays for ultrafast, gas-phase chromatographic chemical analysis,” *Nano Letters*, vol. 10, no. 10, pp. 3899–3903, 2010.
- [Li 21] M.W Li, A. Ghosh, A. Venkatasubramanian, R. Sharma, X. Huang, X. Fan “High-sensitivity micro-gas chromatograph–photoionization detector for trace vapor detection,” *ACS Sensors*, vol. 6, no. 6, pp. 2348–2355, 2021.
- [Lia21] W. Liao X. Zhao, H. Lu, T. Byambadorj, Y. Qin, Y.B. Gianchandani, “Progressive cellular architecture in microscale gas chromatography for broad chemical analyses,” *Sensors*, vol. 21, no. 9, p. 3089, 2021.
- [Lia23] W. Liao, D. Winship, I. Lara-Ibeas, X. Zhao, Q. Xu, H. Lu, T. Qian, R. Gordenker, Y. Qin, Y.B. Gianchandani, “Highly Integrated μGC Based on a Multisensing Progressive Cellular Architecture with a Valveless Sample Inlet,” *Analytical Chemistry*, vol. 95, no. 4, pp. 2157–2167, 2023.
- [Liv23] LiveCharts, “LiveCharts Simple, flexible, powerful and open source data visualization for .Net”, <https://v0.lvcharts.com/> (accessed Jul. 11, 2023).

- [Mic23] Microsoft Learn, “Windows Presentation Foundation for .NET 7 documentation,” Windows Presentation Foundation for .NET 7 documentation, <https://learn.microsoft.com/en-us/dotnet/desktop/wpf/?view=netdesktop-7.0> (accessed Jul. 11, 2023).
- [McC17] M.M. McCartney, Y. Zrodnikov, A.G. Fung, M.K. LeVasseur, J.M. Pedersen, K.O. Zamuruyev, A.A. Aksenov, N.J. Kenyon, C.E. Davis., “An easy to manufacture micro gas preconcentrator for chemical sensing applications,” *ACS Sensors*, vol. 2, no. 8, pp. 1167–1174, 2017.
- [Pat03] S. Patel, T. Mlsna, B. Fruhberger, E. Klaassen, S. Cemalovic, D. Baselt, “Chemical capacitive microsensors for volatile organic compound detection”, *Sensors and Actuators B: Chemical*, vol 96, no. 3, pp 541–553, December 2003
- [Pig23] pigpio Library, “The Pigpio Library,” <http://abyz.me.uk/rpi/pigpio/> (accessed Jul. 13, 2023).
- [Pyt23a] Python documentation, “Multiprocessing - process-based parallelism,” <https://docs.python.org/3/library/multiprocessing.html?highlight=multiprocessing#module-multiprocessing> (accessed Jul. 11, 2023).
- [Pyt23b] Python documentation, “Threading - thread-based parallelism,” <https://docs.python.org/3/library/threading.html> (accessed Jul. 11, 2023).
- [Qin16] Y. Qin, Y.B. Gianchandani, “A fully electronic microfabricated gas chromatograph with complementary capacitive detectors for indoor pollutants,” *Microsystems & Nanoengineering*, vol. 2, no. 1, 2016.
- [Ras23] Raspberry Pi Foundation, “Raspberry Pi Model 3B+”, <https://static.raspberrypi.org/files/product-briefs/Raspberry-Pi-Model-Bplus-Product-Brief.pdf> (accessed Jul 20, 2023)
- [Reg18] B. P. Regmi and M. Agah, “Micro Gas Chromatography: An overview of critical components and their integration,” *Analytical Chemistry*, vol. 90, no. 22, pp. 13133–13150, 2018.
- [SMB23] SMBus Protocol Summary, <kernel.org/doc/Documentation/i2c/smbus-protocol> (accessed Sept. 06, 2023)
- [Ter79] S. C. Terry, J. H. Jerman, and J. B. Angell, “A gas chromatographic air analyzer fabricated on a silicon wafer,” *IEEE Transactions on Electron Devices*, vol. 26, no. 12, pp. 1880–1886, 1979.
- [Wan19] J. Wang, N. Nuño, R. Nidetz, S.J. Peterson, B.M. Brookover, W.H. Steinecker, E.T. Zellers, “Belt-mounted micro-gas-chromatograph prototype for determining personal exposures to volatile-organic-compound mixture components,” *Analytical Chemistry*, vol. 91, no. 7, pp. 4747–4754, 2019.

[Xu23a] Q. Xu, Y. Zhao, Y. Qin, Y.B. Gianchandani, “Control Software Design for a Multisensing Multicellular Microscale Gas Chromatography System”, *Micromachines*, vol.15, no.1, p 95, 2023.

[Xu23b] Q. Xu, Y. Qin, Y.B. Gianchandani, “A Rule-based Automated Chemical Recognition Algorithm for a Multi-cell Multi-detector Micro Gas Chromatograph”, *Separations*, vol 10, no. 11, pp 555, 2023.

[Zha19] X. Zhao, Y. Qin, and Y. B. Gianchandani, "Automatic Peak Integration and Baseline Correction for Micro-scale Gas Chromatographs Using Continuous Wavelet Transform," *IEEE SENSORS*, Montreal, QC, Canada, 2019, pp. 1-4.

[Zho16] M. Zhou, J. Lee, H. Zhu, R. Nidetz, K. Kurabayashi, X. Fan, “A fully automated portable gas chromatography system for sensitive and rapid quantification of volatile organic compounds in water,” *RSC Advances*, vol. 6, no. 55, pp. 49416–49424, 2016.

Chapter 3 A Rule-Based Automated Chemical Recognition Algorithm for a Multi-Cell Multi-Detector Micro Gas Chromatograph¹

A chemical recognition algorithm is an integral part of any autonomous microscale gas chromatography (μ GC) system for automated chemical analysis. For a multi-detector μ GC system, the chemical analysis must account for the retention time of each chemical analyte as well as the relative response of each detector to each analyte, i.e., the detector response pattern (DRP). In contrast to the common approaches of heuristically using principal component analysis and machine learning, this chapter reports a rule-based automated chemical recognition algorithm for a multi-cell, multi-detector μ GC system, in which the DRP is related to theoretical principles; consequently, this algorithm only requires a small amount of calibration data but not extensive training data. For processing both the retention time and the raw DRP, the algorithm applies rules based on expert knowledge² to compare the detected peaks; these rules are located in a customized software library. Additionally, the algorithm provides special handling for chromatogram peaks with a small signal-to-noise ratio. It also provides separate special handling for asymmetrical peaks that may result from surface adsorptive analytes. This work also describes an experimental evaluation in which the algorithm used the relative response of two complementary types of capacitive detectors as well as a photoionization detector that were incorporated into the μ GC system of interest. In these tests, which were performed on

¹ This chapter is reproduced from a journal manuscript that has been published [Xu23]

² Expert knowledge for recognition was provided by Dr. Yutao Qin and Dr. Weilin Liao. The input for this work (the peak information) are provided by a peak detection algorithm developed by Dr. Xiangyu Zhao, Et al.

chromatograms with 21–31 peaks for each detector, the true positive rate was 96.3%, the true negative rate was 94.1%, the false positive rate was 5.9%, and the false negative rate was 3.7%.

The results demonstrated that the algorithm can support μ GC systems for automated chemical screening and early warning applications.

3.1 Introduction

With the research and development dating back to the 1970s [Car72, Ter79], microscale gas chromatography (μ GC) systems are now achieving commercial significance and becoming increasingly promising for *in situ* measurements of volatile organic compounds (VOCs). A μ GC system typically incorporates pump(s), valve(s), separation column(s), preconcentrator(s), and detector(s) [Reg18]. There are two steps necessary to generate a chromatogram: the sampling and separation steps. During the sampling step, analytes are adsorbed by the preconcentrator(s). During the subsequent separation step, the analytes are desorbed and injected into the separation column(s), where the analytes are separated based on volatility as they pass through the separation column(s). A chromatogram is produced by the output of each detector that is located downstream of a column. The temporal delay of each analyte peak is called the elution (or retention) time, and is a characteristic to that analyte for the particular separation column and test conditions. To enhance the differentiation of analytes with similar retention times, some μ GC systems incorporate multiple complementary detectors [Jin09, Li10, Qin16, Hu18, Wang19].

Our group has recently reported a multi-cell, multi-detector μ GC system based on a multi-sensing progressive cellular architecture (MPCA) [Lia23]. This MPCA system monolithically integrates three cells, each incorporating a preconcentrator and a separation column that are tailored for a specific volatility range of analytes. Downstream of each separation column in an MPCA cell are located three detectors in series: two capacitive detectors

with different polymer coating thicknesses (CapDetA and CapDetB) and an arrayed integrated photoionization detector (AiPD). By their inherent nature, these detectors provide complementary responses to chemical analytes, leading to well differentiated responses.

An analytical run in an MPCA system includes a single (collective) sampling step followed by three sequential separation sub-steps – one for each cell. In the sampling step, the sampled vapor passes the three preconcentrators that are ordered by the sorbent adsorptivity, from lowest to highest. As such, the least adsorptive preconcentrator, which is located upstream, traps the least volatile chemicals, whereas the downstream, more adsorptive preconcentrators trap more volatile chemicals. In the subsequent separation sub-steps, one for each cell, the corresponding preconcentrator injects the adsorbed chemicals into the separation column within its cell, with the separation column being tailored for the volatility range of that cell. The output of the three detectors in each cell is tracked to create the overall chromatogram.

The essential information that emerges from an analytical run is the retention time of each analyte and the relative intensity of the response of each detector in each cell to each analyte that passes across it. The relative response of each detector to each analyte is defined as the detector response pattern (DRP) in this work. This chapter focuses on an automated algorithm that recognizes chemicals based on the detected peaks in an MPCA μ GC system using a combination of retention time and the DRP; the findings can be applied more generally to other μ GC systems.

The DRP is a valuable metric for chemical recognition [Kim11]. Various tools for DRP recognition have been reported. One of the most common statistical tools is the principal component analysis (PCA) [Wan18, Hov89], which is often used to reduce data dimensions while maximally preserving information [Jol16]. For chemical recognition from the DRP, the

PCA result is often combined with visual inspection [Gar11], Fisher ratio (F-ratio) analysis [Gar11], or other tools [Par00, Par99, Zel95] to cluster the data, so that statistical boundaries for different chemicals may be established. Another common statistical tool is library lookup [Ste99] combined with machine learning techniques [Sho22, Bac21, Sen06, Mat20], which is also used to cluster the data for recognition.

Although statistical tools have achieved various degrees of success, they tend to have some limitations. First, most statistical tools neglect expert knowledge about the DRP that is intrinsically determined by the sensing principles of the detectors. Consequently, when processed by the statistical tools, certain DRP features that benefit the use of expert knowledge may be reduced or eliminated. Second, these tools typically require a substantial amount of data for training, especially for machine learning techniques like neural networks [Alw18, Bel13]. This requirement increases the burden on performing extensive experimental characterization, even though the experimental efforts can sometimes be alleviated by Monte Carlo simulation. Efforts to fully automate the recognition have been reported [Bac21, Wil20, Beh11] but have not been targeted towards multi-detector μ GC systems.

To effectively treat the DRP with expert knowledge and relatively small amount of experimental characterization, a rule-based algorithm can be used for the raw DRP without statistical processing. The rules can be constructed by setting up an acceptable range of values for each feature of interest to account for uncertainty in the feature, essentially forming a fuzzy logic system [Baf84, Ott86, Lov94, rea07]. In fuzzy logic, a membership function assigns to each object a score of membership ranging between 0 and 1, where a score closer to 1 indicates that the object is more likely a member of a certain class [Zad65]. The membership function can be determined by statistical data, *e.g.*, derived from the probability density function of a feature

of the object. Alternatively, the membership function can be a mathematical expression of expert knowledge [Men19]. Triangular, trapezoidal, Gaussian, rectangular, and piecewise functions are commonly used for membership functions [Men19].

This chapter reports an automated rule-based chemical recognition algorithm for the MPCA system, in which expert knowledge is applied to construct a set of fuzzy logic rules. After the peaks are detected from a raw chromatogram, this algorithm uses the peak information (including the retention time, asymmetry, and peak heights) as the input. The recognition uses a chemical library that is constructed from a small number of experiments and applies the rules on the retention time and DRP parameters. Additionally, special rules are implemented for handling surface adsorptive chemicals and peaks with low signal-to-noise ratios. Finally, for each peak, the algorithm provides a list of possible candidates with likelihood scores as the output.

3.2 Operating Principle and Method

3.2.1 Overall Algorithm flow

For ease of discussion, the temporal response of each detector in the MPCA system during the analysis of an analytical run is defined as a sub-chromatogram. For the whole MPCA system with two effective cells and three detectors in each cell; the ensemble of the 6 sub-chromatograms forms a chromatogram. In the following description, a peak in the chromatogram refers to the responses from all three detectors at a matched retention time.

The raw chromatograms must first be processed with a peak detection algorithm. For the MPCA system, the information provided by the peak detection algorithm includes, for each peak, the cell number, retention time, peak asymmetry, and peak heights from the three detectors. For each set of peaks corresponding to an analyte, the chemical recognition algorithm (Fig. 3.1) compares the experimentally observed retention time to values in the chemical library and selects

matching candidates. Among these candidates, the algorithm further compares the DRP of this set of peaks against those in the chemical library. Both comparisons use custom-defined recognition rules, which are combined to provide an overall likelihood score.

In this process, multiple special handling cases may be triggered. First, a known reference chemical may be introduced by the user into the sample and used as a reference for the retention time. In this case, the algorithm first finds the peak for the reference chemical based on the chemical library, then converts the retention times of all other peaks to retention times relative to the reference chemical) and uses these relative values for recognition. Second, for surface absorptive analytes such as the phosphonate esters that usually exhibit asymmetric peaks, the retention times may vary with the injected masses in the retention peaks. For these certain analytes, the retention times are not independently incorporated in the library. Instead, the relationship between retention time and peak height must be pre-determined experimentally and this height-dependent retention time must be incorporated into the library. Third, when analyte is such that the responses of the detectors are highly orthogonal, some detectors may show strong responses, whereas others may show very weak or zero responses. Such cases may result in small detector signals, which require special handling that overrides the result of direct DRP matching.

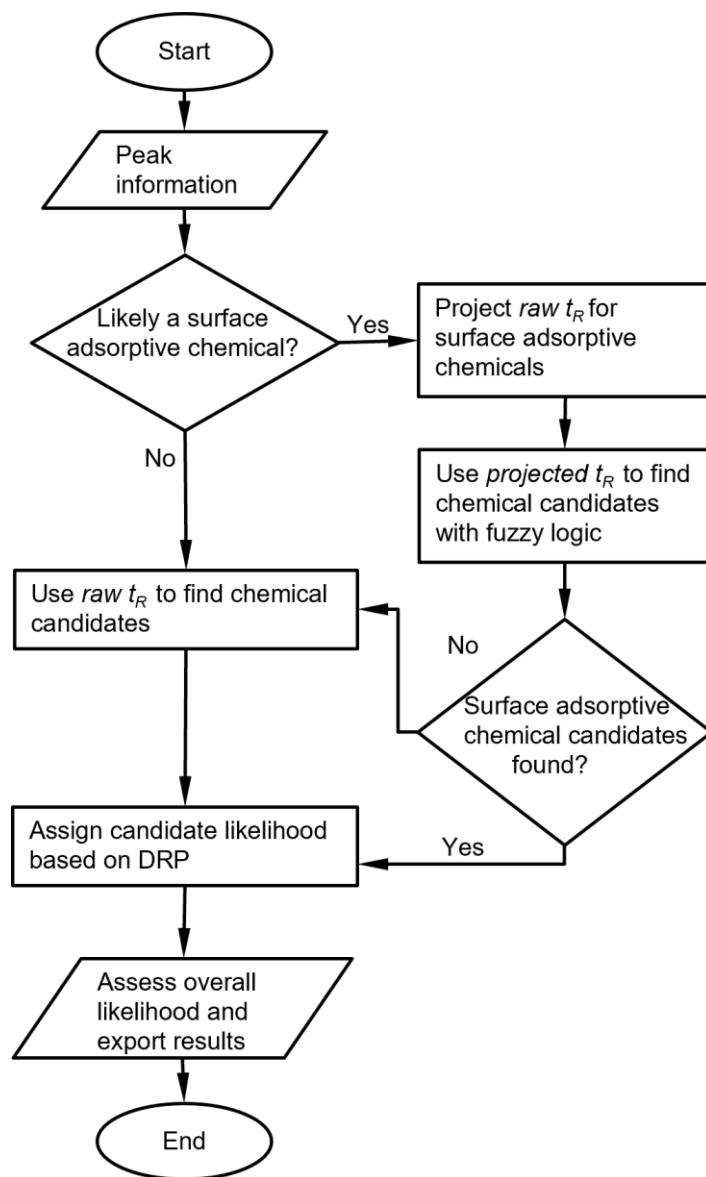


Figure 3.1: Overall chemical recognition algorithm flow for each peak in a chromatogram.

3.2.2 Chemical Library and recognition parameters

The chemical library stores all the parameters of the expected system response to each chemical and of the recognition criteria. The library is constructed as a tabulated file using Microsoft Excel; all the parameters can be edited readily. The chemical library incorporates a basic library, an expanded library, and other algorithmic parameters. The basic library contains chemical properties and system response parameters of all target chemicals (Table 3.1). The

chemical properties, obtained from online databases, include molecular weight, Kovats retention index (RI), dielectric constant, and ionization potentials. These chemical properties are retained for background, but not for recognition by the algorithm. The system response parameters include nominal values of the retention time, DRP, the primary cell used for recognition, detector sensitivities, and chemical peak asymmetry. These parameters are obtained from a minimum set of experimental calibration runs, where the system is operated to sample and analyze custom-prepared chemical standards that contain well-separable chemicals. Such calibration runs are normally performed on any system and may include experimentally obtained values combined with theoretically interpolated values. These parameters are used for recognition by the algorithm.

Table 3.1: Example of the basic chemical library that contains the chemical properties and MPCA system response characteristics.

Chemical Analyte	Physical properties				Primary Cell	Retention Time (s)	Sensitivity			Nominal response ratio			Asym
	Mol Weight (g/mol)	Kovats RI	Dielectric Constant	Ionization Potential (eV)			CapDetA (fF/ppb/min)	CapDetB (fF/ppb/min)	AiPD (mV/ppb/min)	CapDetA/AiPD (fF/mV)	CapDetB/CapDetA (fF/fF)	CapDetB/AiPD (fF/mV)	
Benzene	78.1	654	2.3	9.2	2	43.5	1.91×10^{-4}	4.43×10^{-5}	8.45×10^{-2}	2.26×10^{-3}	2.32×10^{-1}	5.25×10^{-4}	0
2,3-Butanediol	90.1	753	21.6	Unknown	2	123.6	9.01×10^{-3}	1.47×10^{-2}	3.75×10^{-2}	2.40×10^{-1}	1.63×10^0	3.92×10^{-1}	0
Butyl Acetate	116.2	796	5.1	10.0	2	129.4	2.63×10^{-3}	3.47×10^{-3}	1.55×10^{-2}	1.70×10^{-1}	1.32×10^0	2.24×10^{-1}	0
Carbon Tetrachloride	153.8	658	2.2	11.5	2	45.1	1.35×10^{-4}	-1.00×10^{-4}	0	infinity	-7.63×10^{-1}	infinity	0
Cyclohexane	84.2	662	2.0	9.9	2	46.5	1.00×10^{-4}	-2.64×10^{-4}	4.78×10^{-2}	2.09×10^{-3}	-2.64×10^0	-5.52×10^{-3}	0
Decane	142.3	1000	2.0	9.7	3	33.5	5.60×10^{-3}	-1.08×10^{-2}	7.82×10^{-2}	7.16×10^{-2}	-1.93×10^0	-1.38×10^{-1}	0
DEMP	152.1	975	13.4	Unknown	3	42.6	3.98×10^{-2}	5.62×10^{-2}	1.35×10^{-2}	2.95×10^0	1.41×10^0	4.16×10^0	1
DIMP	180.2	1073	7.7	Unknown	3	61.1	5.51×10^{-2}	6.45×10^{-2}	3.13×10^{-2}	1.76×10^0	1.17×10^0	2.06×10^0	1
DMMP	124.1	840	20.3	10.0	2	230.5	1.55×10^{-2}	3.25×10^{-2}	1.03×10^{-2}	1.50×10^0	2.10×10^0	3.16×10^0	1
o-Xylene	106.2	881	2.6	8.6	2	196.5	6.51×10^{-4}	2.93×10^{-4}	3.39×10^{-2}	1.92×10^{-2}	4.49×10^{-1}	8.63×10^{-3}	0

The expanded library includes windows for the retention time and the DRP. Here, each window is a range defined by a lower bound value and an upper bound value. A parameter of an experimentally observed peak is considered a match to the corresponding parameter in the library if the experimental value is between the upper and lower bounds. The expanded library

also includes parameters for the relationship between retention time and peak height of the surface adsorptive chemicals, which are dependent on the peak heights (Table 3.2).

Other parameters include identifying whether a reference chemical has been used, the name of the reference chemical, the limits of detection of the detectors, the sampling time of the run, the retention time windows (as fractions of the nominal values) for the surface adsorptive chemicals, and peak asymmetry threshold for treatment as surface adsorptive chemicals. The purpose of these parameters are described in the following sub-sections.

The library established for this work incorporates a total of 28 chemicals. Among these, three are surface adsorptive chemicals that were characterized together (*i.e.*, in mixtures) by 10 experimental runs that covered a practical range of concentrations. The other 25 chemicals were characterized either as neat chemicals or in mixtures of 2-6 chemicals, in a total of 7 experimental runs. Each run was performed at a single concentration that provided enough signal-to-noise ratios for most of the detector responses.

Table 3.2: Example of the expanded chemical library incorporating additional parameters for use by the algorithm.

Chemical	Cell	Retention time likelihood window				DRP likelihood window						Surface adsorptive chemical fitting parameters				
		$t_{R,HC,L}$ (s)	$t_{R,HC,U}$ (s)	$t_{R,MC,L}$ (s)	$t_{R,MC,U}$ (s)	BA_L (fF/fF)	BA_U (fF/fF)	BD_L (fF/mV)	BD_U (fF/mV)	AD_L (fF/mV)	AD_U (fF/mV)	p_1	p_2	p_3	p_4	p_5
Benzene	2	40.9	46.1	39.1	47.8	-1	1	-9.45×10^{-4}	-1.05×10^{-4}	4.52×10^{-4}	4.07×10^{-3}					
2,3-Butanediol	2	116.2	131.0	111.2	136.0	1	3	7.84×10^{-2}	7.06×10^{-1}	4.81×10^{-2}	4.32×10^{-1}					
Butyl Acetate	2	121.6	137.1	116.4	142.3	1	3	4.48×10^{-2}	4.03×10^{-1}	3.39×10^{-2}	3.06×10^{-2}					
Carbon Tetrachloride	2	42.4	47.8	40.6	49.6	-1	1	-infinity	-1.00×10^0	1.00×10^0	infinity					
Cyclohexane	2	43.7	49.3	41.9	51.2	-infinity	-1	-9.94×10^{-3}	-1.10×10^{-3}	2.09×10^{-4}	3.77×10^{-3}					
Decane	3	31.4	35.5	30.1	36.8	-infinity	-1	-2.49×10^{-1}	-2.76×10^{-2}	1.44×10^{-2}	1.30×10^{-1}					
DEMP	3	0	0	0	0	1	3	8.33×10^{-1}	7.49×10^0	5.89×10^{-1}	5.30×10^0	47.69	0.25	18.92	0.01	29.00
DIMP	3	0	0	0	0	1	3	4.12×10^{-1}	3.71×10^0	3.52×10^{-1}	3.17×10^0	22.95	0.06	14.73	0.38	53.00
DMMP	2	0	0	0	0	1	3	6.13×10^{-1}	5.68×10^0	3.03×10^{-1}	2.73×10^0	46.85	0.59	246.30	0.01	0.01
o-Xylene	2	184.7	208.3	176.9	216.2	-1	1	1.55×10^{-2}	1.73×10^{-3}	3.84×10^{-3}	3.46×10^{-2}					
o-Xylene	3	14.9	16.9	14.3	17.5	-1	1	1.55×10^{-2}	1.73×10^{-3}	2.52×10^{-3}	2.27×10^{-2}					

Cell indicates the cell that provides the peak to the chemical.

$t_{R,HC,L}$ and $t_{R,HC,U}$ indicate the lower bound and upper bound of the retention time high-confidence window, respectively

$t_{R,MC,L}$ and $t_{R,MC,U}$ indicate the lower bound and upper bound of the retention time medium-confidence window, respectively

BA_L and BA_U indicate the lower and upper bound of the CapDetB/CapDetA window, respectively

BD_L and BD_U indicate the lower and upper bound of the CapDetB/AiPD window, respectively

AD_L and AD_U indicate the lower and upper bound of the CapDetA/AiPD window, respectively

3.2.3 Recognition by retention time and cell number

As noted previously, the first step to recognizing a peak is to find chemicals in the library with matching retention times. The repeatability in the retention time is determined by the repeatability in the column flow rate and temperature, which may drift in field environments. Surface-adsorptive chemicals eluting a μ GC column without sufficient deactivation may cause asymmetric peaks, for which the retention times may vary with concentration. Therefore, an appropriate window must be selected for the retention time based on the knowledge of the μ GC system hardware.

The algorithm leverages the characteristics of the microsystem to which it is applied. Because the MPCA system incorporates multiple cells, the chemicals in the library are first pre-filtered by the cell which detected the peak. Within the MPCA system architecture, some chemicals may have responses in only one cell, whereas others may have responses in two cells. In the latter case, usually one cell is superior to the other cell for recognition, because of better peak separation or a taller peak. Hence, the superior cell is defined in the library as the primary cell for this chemical and used for recognition and quantification; the response of the other cell is excluded from the library, except for special cases (*e.g.*, a reference chemical that has responses in two cells and can serve both the cells).

Next, the algorithm checks the peak asymmetry. A peak with significant tailing (*i.e.*, with a tail portion that is much longer than the duration of the rising edge) is typically caused by a surface adsorptive chemical, which requires pre-treatment with an adjustment of the retention time as discussed below. For a symmetrical peak (*i.e.*, without significant tailing), the algorithm checks if its retention time falls within the retention time windows of each chemical in the library. For the retention time, the recognition involves the use of a high-confidence window

and a medium-confidence window (which encloses the high-confidence window). If the retention time of the peak is located within the high-confidence window of a chemical, the chemical is considered a candidate and assigned a retention time likelihood score (S_{tR}) of 1. If the peak is located outside the high-confidence window but within the medium-confidence window of a chemical, the chemical is also considered a potential candidate but is assigned a lower S_{tR} of 0.5. The combination of these windows forms a piecewise membership function (Fig. 3.2). By default, the bounds of the high-confidence window and medium-confidence window are empirically set at $\pm 6\%$ and $\pm 10\%$ around the nominal retention times, respectively.

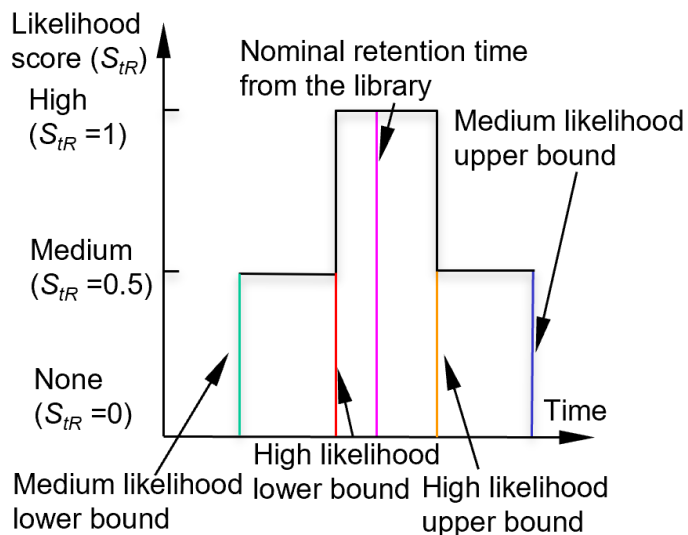


Figure 3.2: Likelihood scoring function for retention time.

For the special case of surface adsorptive chemicals, *e.g.*, phosphonate esters such as dimethyl methylphosphonate (DMMP), diethyl methylphosphonate (DEMP) and diisopropyl methylphosphonate (DIMP), the surface adsorption along the fluidic paths causes not only peak tailing but also retention time variation with concentration. For recognition, the latter factor must be specially treated. The peak tailing is a good indicator to trigger this special treatment. However, in a μ GC system, peak tailing may also result from imperfections, such as slow

preconcentrator desorption or leakage. To reduce the number of special treatments required, a second indicator can be added based on the fact that the surface adsorptive chemicals typically have high polarity, which produces positive responses in both the capacitive detectors [Lia23]. Therefore, the algorithm implements a rule such that, if the asymmetry of a peak exceeds a certain threshold defined in the library, and both the CapDetA and CapDetB peaks are positive, this peak is considered likely a surface adsorptive chemical. Empirically, this threshold value is set to 3.

The next step is to determine whether this peak can be recognized as a surface adsorptive chemical in the library. For each of the applicable surface adsorptive chemicals in the library (*i.e.*, in the same cell as this peak), its projected retention time (t_{Rp}) can be calculated based on a pre-characterized function of the actual peak height (H), which is empirically formulated as:

$$t_{Rp} = p_1 e^{-p_2 H} + p_3 e^{-p_4 H} + p_5 \dots\dots\dots(3.1)$$

where p_1 , p_2 , p_3 , p_4 , and p_5 are fitting parameters obtained from multiple prior experimental characterization runs over a concentration range of interest (Fig. 3.3). Considering the relatively large capacitive detector responses to the surface adsorptive chemicals, the peak height for this function is obtained from a capacitive detector (in this case CapDetA). In prior experimental characterization, the fitting parameters for all the tested phosphonate esters provided R^2 values ≥ 0.99 (Fig. 3). Next, for each surface adsorptive chemical in the library, the bounds of the high-confidence window and medium-confidence window are empirically set at $\pm 10\%$ and $\pm 20\%$ around the t_{Rp} , respectively. The algorithm then searches for the surface adsorptive chemical candidates that have retention time windows enclosing the retention time of the detected peak. If no such candidate is found, the algorithm falls back to treat the peak as a normal peak.

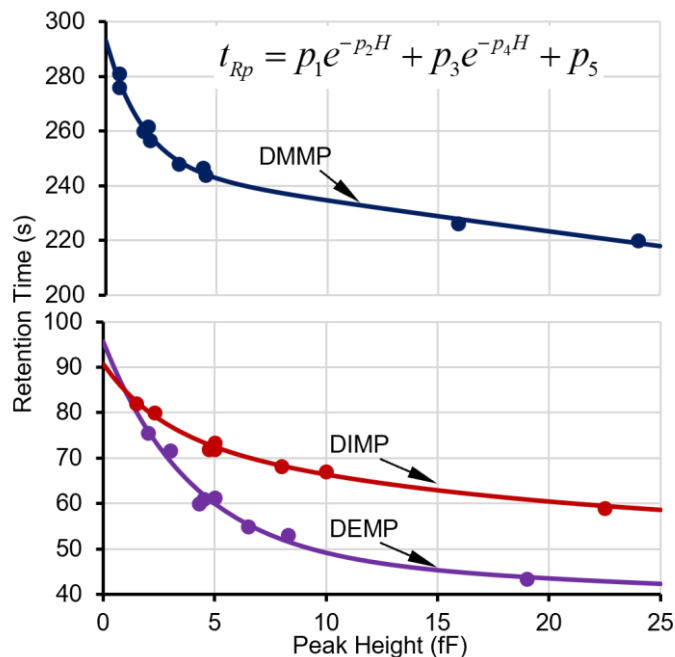


Figure 3.3: Experimental data points on retention times and peak heights and fitted curves for DMMP in Cell2 and DEMP and DIMP in Cell3.

3.2.4 Recognition by DRP

After the selection of chemical candidates by retention time, leveraging the multi-detector benefit offered by the MPCA system, the DRP is further used for chemical recognition. In the MPCA system, the DRP consists of the three peak height response ratios representing ratios of the three detectors, *i.e.*, CapDetB/CapDetA, CapDetA/AiPD and CapDetB/AiPD. The nominal values of these response ratios are stored in the library. In the general case, a library lookup process is applied to each response ratio. If a response ratio of a detected peak falls within the corresponding window of a chemical candidate in the library, this response ratio is considered a match to the library and is assigned a likelihood score of 1, otherwise it is assigned a likelihood score of 0. As a result, the membership function for each individual response ratio is essentially a rectangular function. The DRP match scores are denoted as S_{BA} (for CapDetB/CapDetA), S_{AD} (for CapDetA/AiPD) and S_{BD} (for CapDetB/AiPD).

The CapDetB/CapDetA response ratio is a reliable indicator of the chemical polarity. CapDetA is designed to always provide a positive response, which is dominated by swelling of the detector coating (*i.e.*, polydimethylsiloxane) upon chemical absorption. In contrast, CapDetB is designed to provide either a positive response or a negative response, depending on the dielectric constant difference between the chemical and the detector coating [Lia23]. The CapDetB/CapDetA response ratio is most useful for differentiating among the following three chemical categories: 1) highly non-polar chemicals such as alkanes, for which this ratio is typically < -1 ; 2) less non-polar chemicals such as aromatic hydrocarbons, for which this ratio typically is between -1 and 1 ; and 3) polar chemicals, for which this ratio is typically between 1 and 3 . However, within each chemical category, this ratio is less reliable in differentiating individual chemicals [Lia23]. Therefore, in the expanded library, the windows for this ratio are coarsely set at these ranges, rather than fixed percentages around the nominal values. For certain chemicals whose CapDetB/CapDetA response ratios are experimentally verified to deviate from the ranges above, the windows can be adjusted accordingly in the library. For example, 1-octanol is a polar chemical but has been experimentally verified to have a nominal CapDetB/CapDetA response ratio of 0.89 ; its CapDetB/CapDetA response window is adjusted to be from 0.45 to 1.35 in the library.

In contrast, the AiPD response is dominated by the ionization potential of the chemical, which is minimally or not correlated with the capacitive detector responses. In principle, relatively narrow windows can be used for the CapDetA/AiPD and CapDetB/AiPD response ratios. However, in practice, one must consider the inaccuracies in the peak height values provided by the peak detection algorithm, which can be caused by the slightly nonlinear AiPD

response to concentration [Lia23], baseline drift, noise, asymmetry, and co-elution (*i.e.*, overlap of peaks).

In practice, for a combination of detectors with a high level of orthogonality (*e.g.*, the 3 types of detectors in the MPCA system), it is not uncommon for one or more detectors to provide nearly zero peak heights. This scenario can cause significant variability in the calculation of the response ratios in the detected peaks. To address this scenario, a subroutine is implemented for comparing the response ratios of the detected peaks to the chemical library. In this subroutine, for each detector, a peak height threshold (H_{th}) is set at 6 times of the detector noise (represented by the standard deviation σ), below which the response is considered potentially compromised. As a result, the H_{th} values for both the CapDetA and CapDetB are set at 0.24 fF, and that for the AiPD is set at 0.36 mV. Corresponding to the number of peaks with heights below H_{th} , the following three cases are considered: 1) All three peaks with heights below H_{th} ; 2) two peaks with heights below H_{th} , 3) only one peak with a height below H_{th} . In the first case, for a chemical peak, if all the three detectors provide peak height magnitudes below the thresholds, this peak is considered not recognizable by response ratio, and all the three DRP match scores are set to zero.

The second is that in which two detectors provide peak height magnitudes that are below their corresponding thresholds, whereas the third detector provides a peak height magnitude that is above its thresholds. The DRP match score for the response ratio between the first two detectors is assigned zero. For each of the other two DRP scores, a projected peak height is calculated as the product of the third detector (above H_{th}) and the library value of the response ratio of respective first or second detector to the third detector. It is possible that the projected peak height magnitude of the first or second detector is also below the respective H_{th} . In such a

case, the small magnitude of the detected peak height is confirmed, and subsequently an exception is created where the DRP match score for the response ratio between this detector and the third detector is assigned to be 1, indicating a match of this response ratio to the chemical candidate. However, if the projected peak height magnitude of a detector is above H_{th} , the small magnitude of the detected peak height cannot be confirmed. In such a case, the exception above is not created, and the DRP match score for the peak height ratio between this detector and the third detector is determined by the library lookup result (Fig. 3.4). As an example, assume that both the detected CapDetA and CapDetB peak height magnitudes are below their H_{th} whereas the detected AiPD peak height magnitude is above its H_{th} . In this case, S_{BA} is automatically assigned a zero. The projected CapDetA and CapDetB peak height magnitudes are calculated based on the detected AiPD peak height. If the projected CapDetA peak height is also below its corresponding H_{th} , the S_{AD} is assigned a 1. Otherwise, the S_{AD} is dependent on the library lookup result. The same judgement procedure is applied to the CapDetB and the S_{BD} .

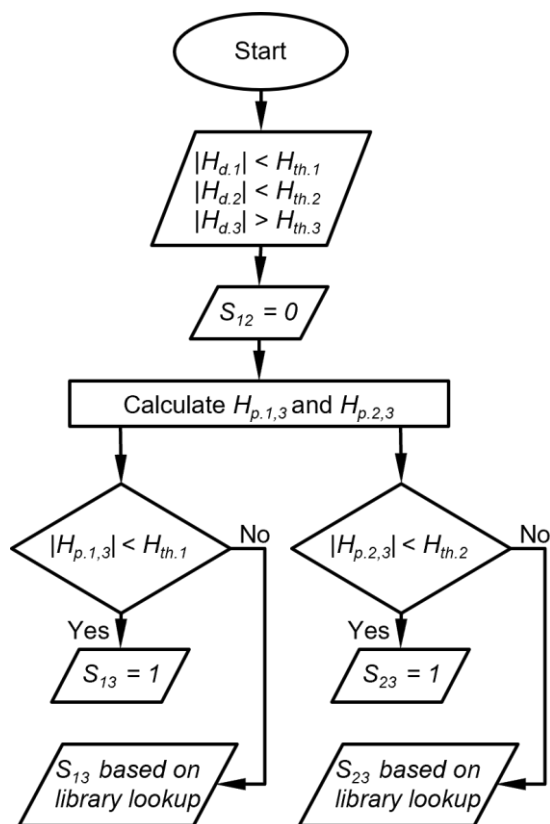


Figure 3.4: Flow diagram of DRP match scoring when two detector peak heights are below their peak height thresholds. $H_{th,i}$ is the peak height threshold for detector i , $H_{d,i}$ is the detected peak height of the detector i , $H_{p,i,j}$ is the peak height of detector i projected from detector j , and S_{ij} is the score that corresponds to the response ratio between detector i and detector j .

The third case is that in which one detector provides a peak height magnitude below its H_{th} , whereas the other two detectors provide peak height magnitudes above their H_{th} (Fig. 3.4). In this case, the DRP match score for the peak height ratio between the latter two detectors (S_{23}) is determined by the library lookup result. Then, two projected peak heights are calculated for the former detector, each based on the detected peak height of one of the latter detectors. If a projected peak height magnitude is below its H_{th} , the DRP match score for the response ratio between the former detector and the corresponding latter detector is assigned to be 1. If a projected peak height magnitude is above its H_{th} , the DRP match score for the peak height ratio between the former detector and the corresponding latter detector is determined by the library

lookup result (Fig. 3.5). As an example, suppose the CapDetA peak height magnitude is below its H_{th} whereas both the CapDetB and AiPD peak height magnitudes are above their H_{th} , the S_{BD} is determined by the library lookup result. Two projections of the CapDetA peak height are calculated from the detected CapDetB and AiPD peak heights. If the CapDetA peak height magnitude that is projected from CapDetB is below the H_{th} of CapDetA, S_{BA} is assigned 1. Otherwise, the S_{BA} is determined by whether the library lookup result for the CapDetA/CapDetB response ratio. The same judgement procedure is applied to the S_{AD} .

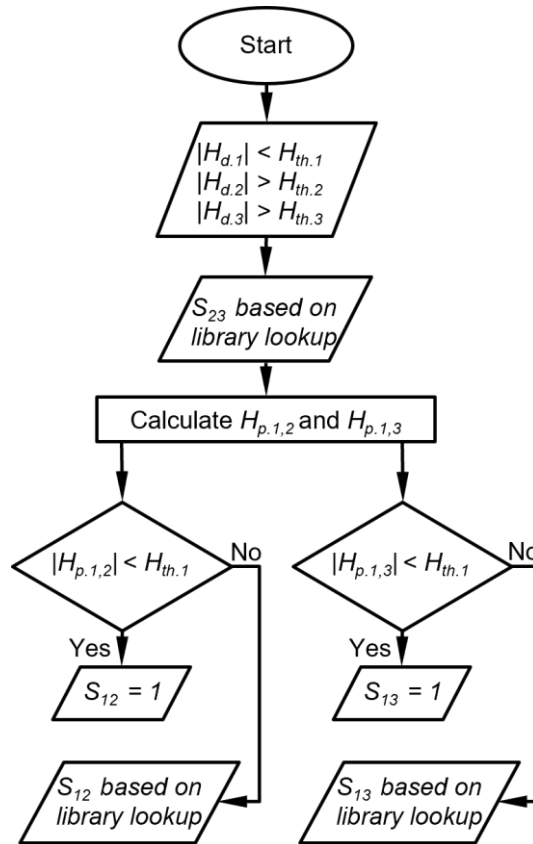


Figure 3.5: Flow diagram of DRP scoring when one detector peak height is below its peak height thresholds. $H_{th,i}$ is the peak height threshold for detector i , $H_{d,i}$ is the detected peak height of the detector i , $H_{p,i,j}$ is the peak height of detector i projected from detector j , and S_{ij} is the score that corresponds to the response ratio between detector i and detector j .

3.2.5 Total score for recognition

To quantify the overall likelihood of a chemical recognition result, a total score (S_{Total}) is calculated for each chemical candidate of a peak. The S_{Total} is defined based on the individual retention time and DRP match scores:

$$S_{Total} = S_{tR} \sum_{i=1}^3 w_i S_i \dots\dots\dots(3.2)$$

where, S_i is the i^{th} DRP match score, w_i is the corresponding weight assigned by the user, and S_{tR} is the previously defined retention time likelihood score. The three DRP match scores are combined by summing rather than multiplication; a single DRP match score of 0 does not eliminate a possible chemical candidate. In contrast, the S_{tR} and the sum of the DRP match score are multiplied (rather than being added); because a 0 in either of these two values is a good indicator of mismatch between a peak and a listed chemical in the library. By default, the w_i values are all set to be 1/3. For a chemical candidate, as S_{tR} can be either 1 or 0.5 and S_i can be either 1 or 0, the possible values of S_{Total} include 0, 0.17, 0.33, 0.5, 0.67, and 1. For those users who prefer only binary results (*i.e.*, presence or absence of target chemicals), the S_{Total} criterion for positive recognition (*i.e.*, presence of a target chemical) is set to be ≥ 0.67 , *i.e.*, with $S_{tR} = 1$ and at least two S_i values to be 1. For other users who prefer more granularity in the results, the S_{Total} value can provide additional insights. For example, a S_{Total} of 1 represents higher confidence in the recognition than a S_{Total} of 0.67; whereas a S_{Total} of 0.5 may be set as an indicator for further investigation. Note that the weights of the DRP and the criterion for positive recognition can be changed by the user, if needed.

3.2.6 Estimated concentration.

After chemical recognition, the algorithm provides an estimate of the chemical concentration, even though the quantification is not the focus of this work. To minimize the impact of quantification error from the detector noise, the algorithm selects in the primary cell the detector that provides the largest signal-to-noise ratio to compute the concentration associated with a peak. The concentration C is calculated by:

$$C = \frac{H_i}{t_{\text{sampling}} \alpha_i} \dots\dots\dots(3.3)$$

where H_i is the peak height for the detector with the largest signal-to-noise ratio, α_i is the sensitivity of the corresponding detector, and t_{sampling} is the sampling time.

3.2.7 Using a reference chemical

If the user specifies a reference chemical to be used for recognition of other chemicals, the algorithm first verifies the presence of the reference chemical among all the detected peaks in both cells. This verification checks the retention time and the DRP between the detected peaks and the reference chemical information in the library and, using a similar process as described above. If the reference chemical is recognized in a cell, for all the other peaks in this cell, the algorithm computes the relative retention times ($t_{R,r}$) as the ratios between the raw retention times and that of the reference chemical. All the retention time bounds are also ratioed accordingly. Then the algorithm searches for chemical candidates, scores them as described above, and computes the relative concentrations (C_r) as a ratio to that of the reference chemical.

3.3 Results and Discussion

The chemical recognition algorithm was assessed using the peak detection results of the MPCA system chromatograms as the input data. In this work, the peak detection results were provided by a wavelet-based algorithm [Zha23] that was further adapted for the MPCA system chromatograms. This peak detection algorithm provides a set of processed chromatograms for user visualization and provides tabulated peak information of all the detected peaks to be used for chemical recognition. The tabulated peak information includes the peak number, retention time, peak asymmetry, CapDetA peak height, CapDetB peak height, and AiPD peak height. To facilitate user inspection, the tabulated peak information is repeated in the output results table of the chemical recognition algorithm (*e.g.*, Table 3), in the columns “*Chemical Number*”, “*Retention time (s)*”, “*Asymmetry*”, “*CapDetA (fF)*”, “*CapDetB (fF)*”, and “*AiPD (mV)*”, respectively. The output results table also includes the recognition scores S_{iR} , S_{BA} , S_{AD} , S_{BD} , and S_{Total} in the corresponding columns. The names and estimated concentrations of the recognized chemicals are reported in the columns “*Chemical Name*” and “*C (ppb)*”, respectively. For recognition results obtained with reference chemicals, the estimated concentration is labeled “ C_r (*ppb*)” instead. Each recognized chemical contains all the information above in a row, which is uniquely numbered in the column “*Chemical Number*” using the format of “ $x.y.(z)$ ”, where x indicates which cell the peak is from, y indicates which peak in the cell it refers to, and z is the chemical candidate number. For example, the first possible chemical candidate for the third peak detected in Cell2 is numbered as “2.3.(1)”. In this work, only the Cell2 and Cell3 data are used for recognition; Cell1 is in continuing development and will be included in the future [Lia23].

The first example is a relatively simple chromatogram, which contains 8 detected peaks in Cell2 and 13 detected peaks in Cell3. The chromatogram shows all the peaks with known

identities and whether they are in the library (Fig. 3.6). Among the other peaks with unknown identities, Peaks 3.4, 3.7, 3.9, 3.12, and part of Peak 2.4 resulted from system outgassing, as verified by separate chromatograms of blank runs (i.e., with zero-grade air as the samples), which also contained these peaks. The remaining unknown peaks likely resulted from trace impurities in the sample. After the algorithm performed chemical recognition, the output (Table 3.3) showed that all the chemicals that were within the library were correctly recognized, with a total score of 0.67 or higher. The peak for dodecane, which was not in the library, was correctly treated as unknown and not falsely recognized as another chemical in the library.

Using Peak 2.2 as an example to navigate the recognition algorithm, it was recognized as follows. In the first step, the algorithm decided that this peak was unlikely to be a surface adsorptive chemical and hence no special treatment was needed. Despite a peak asymmetry of 3.96 (which exceeded the threshold value of 3 for being considered to have significant tailing), this decision was made because neither peak height from CapDetA and CapDetB was a positive value, indicating that this peak was unlikely to be from a polar chemical. In the second step, based on the cell number and the retention time (29.2 s), both hexane and ethyl acetate were found as possible candidates. In the library, hexane has a nominal retention time at 29.2 s, a high-confidence window of 27.5-31.1 s, and a medium-confidence window of 26.3-32.1 s. Therefore, the S_{IR} for hexane was 1. In the library, ethyl acetate has a nominal retention time at 31.1 s, a high-confidence window of 29.2-33.0 s, and a medium-confidence window of 28.0-34.2 s. Therefore, the S_{IR} for ethyl acetate was also 1.

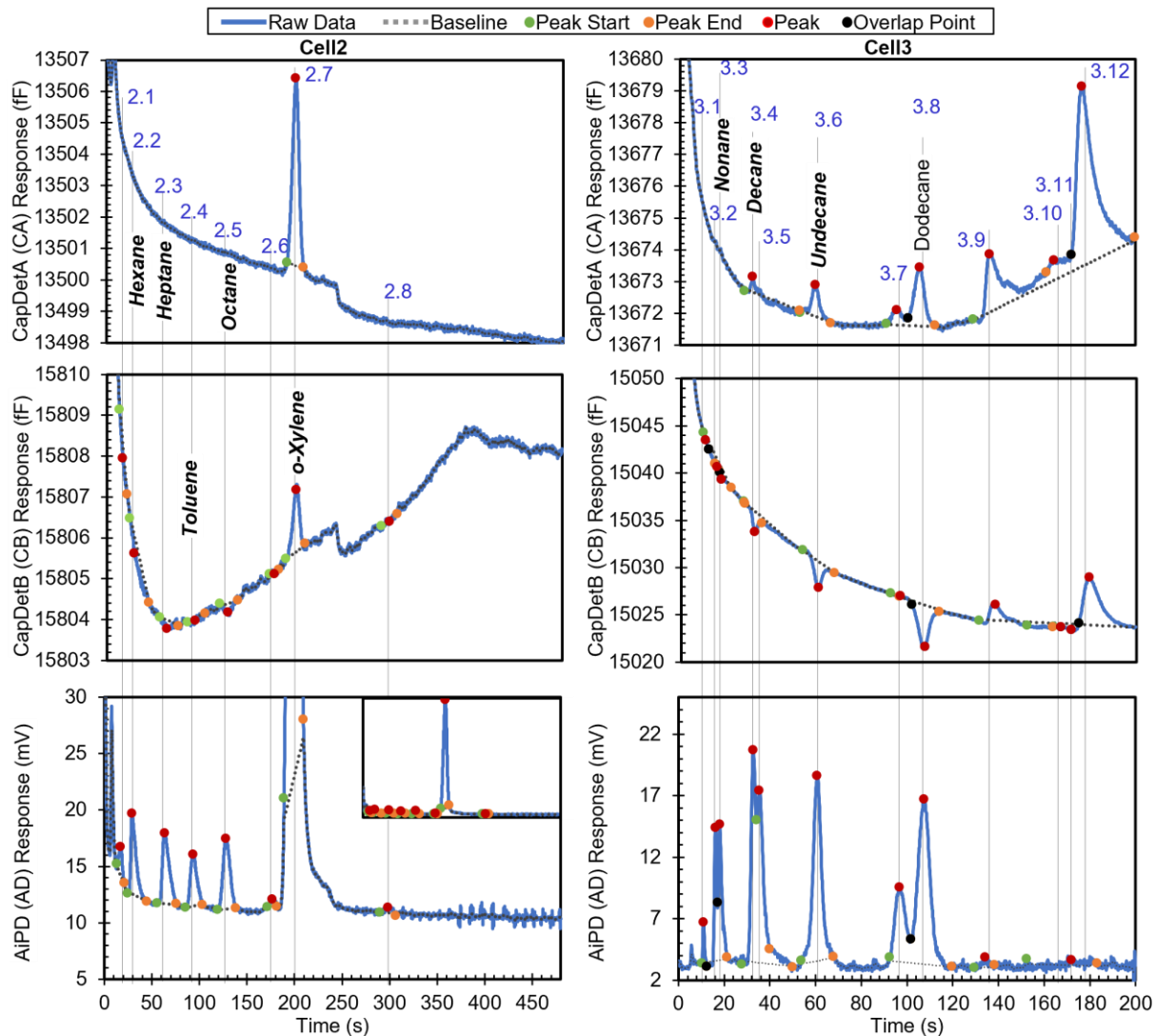


Figure 3.6: Raw chromatograms and detected peaks of the first example. For the peaks with known identities, the identities are labeled. Known peaks present in the library are labeled in bold italic.

Table 3.3: Recognition results for the first example. The labels on the left indicate comparison to the ground truth, in which the bold italic text indicates chemicals that are in the library. This section describes three representative examples of the chemical recognition results. In each example, the recognition results are compared against the true list of chemicals, *i.e.*, the ground truth, on a peak-by-peak basis.

Ground Truth	Output from chemical recognition algorithm												
	Chemical Number	Chemical Name	t_R (s)	Asym	CapDetA (fF)	CapDetB (fF)	AI PD (mV)	S _{tr}	S _{BA}	S _{AD}	S _{BD}	S _{total}	C (ppb)
Unknown	✓ 2.1.(1)	Unknown#1	17.1	1.14	0.00	-0.32	2.50	0	0	0	0	0	
Hexane	✓ 2.2.(1)	Hexane	29.2	3.96	0.00	-0.41	7.03	1	1	1	1	1	70.26
	✓ 2.2.(2)	Ethyl Acetate	29.2	3.96	0.00	-0.41	7.03	1	0	0	0	0	
Heptane	✓ 2.3.(1)	Heptane	63.5	2.51	0.00	-0.18	6.23	1	0	1	1	0.67	66.21
	✓ 2.3.(2)	1-Nitropropane	63.5	2.51	0.00	-0.18	6.23	1	0	0	0	0	
Toluene	✓ 2.4.(1)	Toluene	93.3	1.69	0.00	-0.05	4.63	1	0	1	1	0.67	5.05
Octane	✓ 2.5.(1)	Octane	128.1	1.35	0.00	-0.24	6.20	1	1	1	1	1	114.48
	✓ 2.5.(2)	2,3-Butanediol	128.1	1.35	0.00	-0.24	6.20	1	1	0	0	0.33	
	✓ 2.5.(3)	Butyl Acetate	128.1	1.35	0.00	-0.24	6.20	1	1	0	0	0.33	
Unknown	✓ 2.6.(1)	Ethylbenzene	176.8	0.96	0.00	0.01	0.65	0.50	1	1	1	0.33	
o-Xylene	✓ 2.7.(1)	o-Xylene	199.9	0.83	5.86	1.55	204.29	1	1	1	1	1	602.61
Unknown	✓ 2.8.(1)	Unknown#2	298.5	0.92	0.00	-0.02	0.53	0	0	0	0	0	
Unknown	✓ 3.1.(1)	Unknown#3	11	1.80	0.00	-0.48	3.23	0	0	0	0	0	
Unknown	✓ 3.2.(1)	o-Xylene	16.2	1.25	0.00	-0.77	10.66	1	0	1	0	0.33	
Nonane	✓ 3.3.(1)	Nonane	18.1	3.71	0.00	-1.20	10.98	1	1	1	1	1	18.26
Decane	✓ 3.4.(1)	Decane	32.7	2.77	0.48	-2.18	16.67	1	1	1	1	1	21.32
	✓ 3.4.(2)	Limonene	32.7	2.77	0.48	-2.18	16.67	0.50	1	0	0	0.17	
Unknown	✓ 3.5.(1)	Limonene	35.5	2.83	0.00	0.00	13.21	1	0	1	0	0.33	
Unknown	✓ 3.5.(2)	Decane	35.5	2.83	0.00	0.00	13.21	1	0	0	0	0	
Undecane	✓ 3.6.(1)	Undecane	60.8	1.14	0.97	-2.80	14.96	1	1	1	1	1	26.42
Unknown	✓ 3.7.(1)	Unknown#4	96.8	0.93	0.49	0.10	6.10	0	0	0	0	0	
Dodecane	✓ 3.8.(1)	Unknown#5	107.6	0.94	1.82	-4.33	13.34	0	0	0	0	0	
Unknown	✓ 3.9.(1)	Unknown#6	136.2	6.83	1.79	1.74	0.35	0	0	0	0	0	
Unknown	✓ 3.10.(1)	Unknown#7	167.3	2.19	0.63	-0.31	0.00	0	0	0	0	0	
Unknown	✓ 3.11.(1)	Unknown#8	171.9	5.65	0.00	-0.58	0.45	0	0	0	0	0	
Unknown	✓ 3.12.(1)	Unknown#9	179.8	3.57	5.61	4.97	0.00	0	0	0	0	0	

Legend:

- ✓ Correct recognition of a peak as a chemical in the library
- ✓ Correct recognition for true unknowns and chemicals not in the library

The third step used the DRP information. For Peak 2.2, the detected CapDetA peak height magnitude was below its H_{th} ; in fact, the peak detection algorithm did not detect a CapDetA peak for it. Therefore, Peak 2.2 was treated as the third case in the DRP subroutine.

Assuming that this peak was generated by hexane, the CapDetA peak height was projected from both the detected peak heights from CapDetB and AiPD, based on the nominal DRP for hexane. The former projected values were 0.06 fF and 0.03 fF, respectively; both were smaller than their corresponding H_{th} values. Therefore, both S_{BA} and S_{AD} were assigned to be 1. The CapDetB/AiPD peak height ratio was -5.68×10^{-2} , which was within the corresponding window in the library (-6.09×10^{-2} to -6.76×10^{-3}), therefore S_{BD} was assigned 1. As a result, hexane received a S_{Total} of 1, indicating that this peak was likely generated by hexane. Subsequently, the concentration was calculated for hexane using Eq. (3); based on the peak height of the AiPD, which provided the strongest signal-to-noise ratio among the three detectors, the concentration was estimated to be 70.26 ppb. Next, assuming that this peak was generated by ethyl acetate, the CapDetA peak height was projected. However, neither the projected value was below the corresponding H_{th} . In the subsequent comparison of the peak height ratios against the library, none of the ratios formed a match. Therefore, ethyl acetate received a 0 in every DRP match score and consequently a S_{Total} of 0, indicating that this peak was unlikely to have been generated by ethyl acetate.

The second example is a more complex chromatogram, with 31 peaks. The complexity was manifested in the extent of partially and even fully coeluting peaks, for which the peak detection algorithm may produce inaccurate peak information, particularly on the peak heights and hence the DRP. The chromatogram in Fig. 3.7 shows all the peaks with known identities and indicates whether or not they are in the chemical library. Other peaks correspond to unknown chemicals that resulted from system outgassing or sample impurities. Except methyl isobutyl ketone (MIBK), m-xylene, 1-chloroheptane, nitrobenzene, and dodecane, the other known chemicals were in the library. After the algorithm performed chemical recognition, the

output (Table 3.4) showed that most of the chemicals that were within the library were correctly recognized, with a total score of 0.67 or higher. The peaks for m-xylene, 1-chloroheptane, nitrobenzene, and dodecane, which were not in the library, were correctly treated as either unknowns or a chemical with $S_{Total} < 0.67$, *i.e.*, they were not falsely recognized as another chemical in the library.

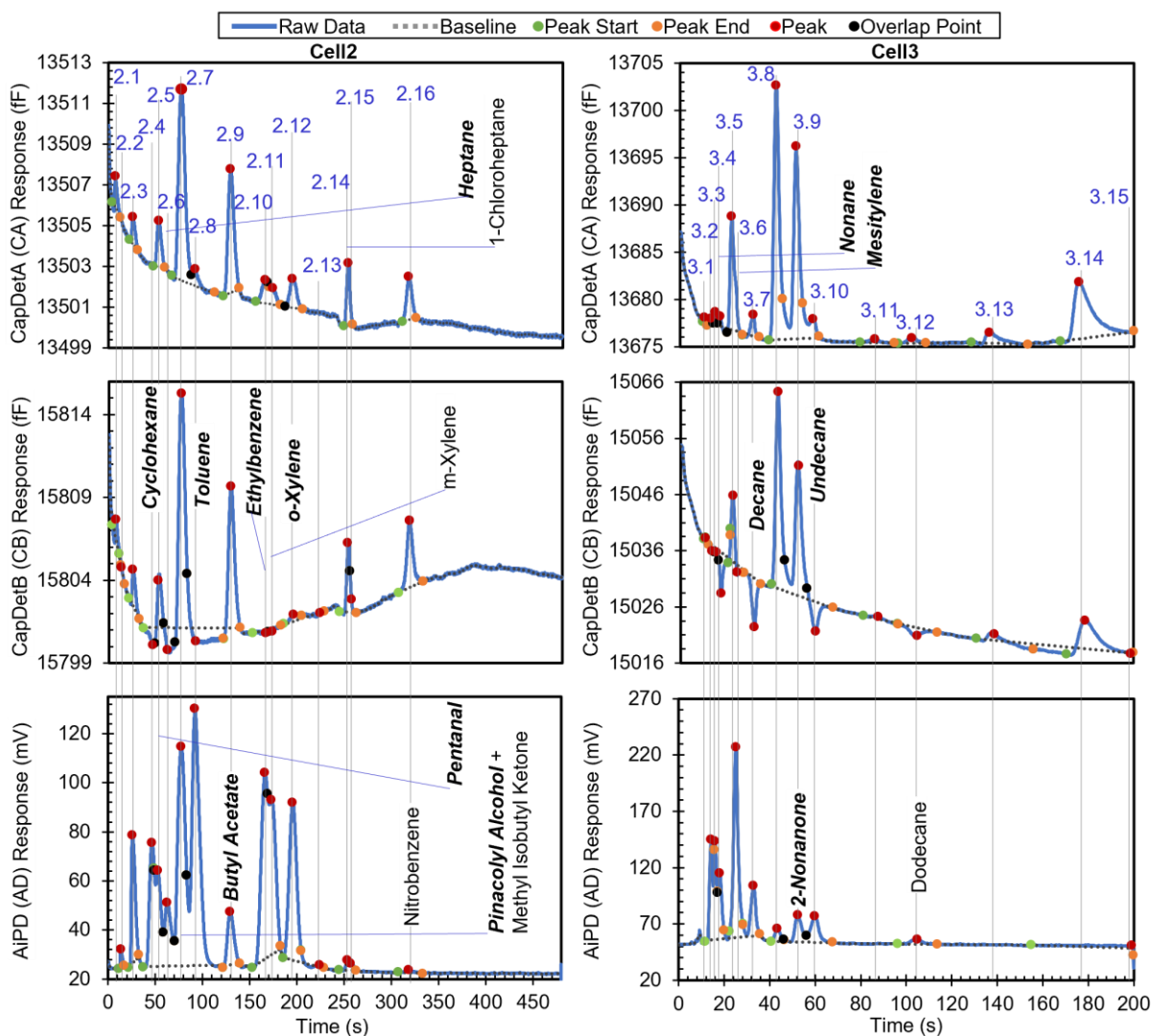


Figure 3.7: Raw chromatograms and detected peaks of the second example. For the peaks with known identities, the identities are labeled. Known peaks present in the library are labeled in bold italic.

Table 3.4: Recognition results for the second example. The labels on the left indicate comparison to the ground truth, in which the bold italic text indicates chemicals that are in the library.

Ground Truth	Output from chemical recognition algorithm												
	Chemical Number	Chemical Name	t _r (s)	Asym	CapDetA (fF)	CapDetB (fF)	AiPD (mV)	S _{tr}	S _{BA}	S _{AD}	S _{BD}	S _{total}	C (ppb)
Unknown	✓ 2.1.(1)	Unknown#1	7.8	3.50	1.53	1.19	0.00	0	0	0	0	0	
	✓ 2.2.(1)	Unknown#2	13.7	3.27	0.00	-0.05	7.51	0	0	0	0	0	
	✓ 2.3.(1)	Unknown#3	26.1	2.94	1.32	2.18	52.08	0	0	0	0	0	
Cyclohexane	✓ 2.4.(1)	Cyclohexane	46.8	0.48	0.00	0.00	50.22	1	0	1	1	0.67	105.07
	✓ 2.4.(2)	Benzene	46.8	0.48	0.00	0.00	50.22	0.5	0	1	1	0.33	
	✓ 2.4.(3)	Carbon Tetrachloride	46.8	0.48	0.00	0.00	50.22	1	0	0	0	0	
	✓ 2.4.(4)	2-Pentanone	46.8	0.48	0.00	0.00	50.22	0.5	0	0	0	0	
Pentanal	✓ 2.5.(1)	Pentanal	52.7	1.69	2.21	3.77	38.90	1	1	1	1	1	233.61
	✓ 2.5.(2)	2-Pentanone	52.7	1.69	2.21	3.77	38.90	0.5	1	1	1	0.5	
	✓ 2.5.(3)	Isooctane	52.7	1.69	2.21	3.77	38.90	0.5	0	1	0	0.17	
Heptane	✓ 2.6.(1)	Heptane	62.8	1.90	0.00	-0.53	25.76	1	1	1	1	1	273.72
	✓ 2.6.(2)	Isooctane	62.8	1.90	0.00	-0.53	25.76	0.5	1	1	1	0.5	
	✓ 2.6.(3)	1-Nitropropane	62.8	1.90	0.00	-0.53	25.76	1	0	0	0	0	
Pinacolyl alcohol + MIBK	✗ 2.7.(1)	Pinacolyl Alcohol	77.8	1.52	9.38	14.86	88.95	1	1	0	0	0.33	
Toluene	✓ 2.8.(1)	Toluene	92.2	2.09	0.74	0.00	104.29	1	1	1	1	1	113.73
Butyl Acetate	✗ 2.9.(1)	2,3-Butanediol	129.6	1.40	6.05	8.93	21.50	1	1	1	1	1	57.32
	✓ 2.9.(2)	Butyl Acetate	129.6	1.40	6.05	8.93	21.50	1	1	1	0	0.67	138.69
	✓ 2.9.(3)	Octane	129.6	1.40	6.05	8.93	21.50	1	0	1	0	0.33	
Ethylbenzene	✓ 2.10.(1)	Ethylbenzene	166.3	0.61	1.20	-0.17	75.60	1	1	1	1	1	258.91
m-Xylene	✓ 2.11.(1)	Ethylbenzene	173.2	63.00	0.78	-0.18	63.12	0.5	1	1	1	0.5	
o-Xylene	✓ 2.12.(1)	o-Xylene	195.7	1.41	1.42	0.42	64.00	1	1	1	1	1	188.78
Unknown	✓ 2.13.(1)	Unknown#4	224.1	0.98	0.00	-0.06	0.61	0	0	0	0	0	
1-Chloroheptane	✓ 2.14.(1)	Unknown#5	253.5	1.22	3.09	4.12	4.07	0	0	0	0	0	
Unknown	✓ 2.15.(1)	Unknown#6	257.2	24.50	0.00	0.83	3.15	0	0	0	0	0	
Nitrobenzene	✓ 2.16.(1)	Unknown#7	318.6	1.27	2.11	4.02	1.11	0	0	0	0	0	
Unknown	✓ 3.1.(1)	Unknown#8	11.3	1.67	0.43	0.39	0.00	0	0	0	0	0	
Unknown	✓ 3.2.(1)	Unknown#9	14.2	2.40	0.54	-0.99	87.00	0	0	0	0	0	
o-Xylene	✓ 3.3.(1)	o-Xylene	16.0	1.83	1.29	-0.57	84.28	1	1	1	0	0.67	
Nonane	✓ 3.4.(1)	Nonane	18.0	0.90	1.08	-7.37	55.10	1	1	1	0	0.67	91.67
Unknown	✓ 3.5.(1)	Mesitylene	23.5	1.00	11.89	11.75	0.00	1	1	0	0	0.33	
Mesitylene	✓ 3.6.(1)	Mesitylene	25.1	15.00	5.36	-1.35	166.32	1	1	0	1	0.67	109.42
Decane	✓ 3.7.(1)	Decane	32.9	0.83	2.22	-8.55	44.08	1	1	1	1	1	56.36
	✓ 3.7.(2)	Limonene	32.9	0.83	2.22	-8.55	44.08	0.5	0	1	0	0.17	
Unknown	✓ 3.8.(1)	Unknown#10	43.4	1.50	26.67	34.71	11.39	0	0	0	0	0	
2-Nonanone	✓ 3.9.(1)	2-Nonanone	52.4	0.74	19.97	23.01	23.75	1	1	1	1	1	100.9
	✗ 3.9.(2)	1-Octanol	52.4	0.74	19.97	23.01	23.75	1	1	1	0	0.67	360.02
Undecane	✓ 3.10.(1)	Undecane	59.9	1.21	1.95	-5.22	23.15	1	1	1	1	1	40.88
Unknown	✓ 3.11.(1)	Unknown#11	87.5	1.04	0.36	0.42	0.00	0	0	0	0	0	
Dodecane	✓ 3.12.(1)	Unknown#12	104.8	1.15	0.54	-1.41	4.07	0	0	0	0	0	
Unknown	✓ 3.13.(1)	Unknown#13	138.9	4.23	1.17	1.37	0.00	0	0	0	0	0	
Unknown	✓ 3.14.(1)	Unknown#14	179.0	3.87	5.98	5.52	0.00	0	0	0	0	0	
Unknown	✓ 3.15.(1)	Unknown#15	199.1	0.02	0.00	-0.05	2.10	0	0	0	0	0	

Legend:

- ✓ Correct recognition of a peak as a chemical in the library
- ✗ False positive
- ✓ Correct recognition for true unknowns and chemicals not in the library
- ✗ False negative

However, there were some cases of misrecognition. Peak 2.7, which resulted from full coelution of pinacolyl alcohol (in the library) and MIBK (not in the library), was not recognized correctly. Based on the retention time, pinacolyl alcohol was correctly found as the possible candidate with a S_{IR} of 1. In the subsequent recognition based on the DRP, all the three detector peak height magnitudes were above their corresponding H_{th} values, so all DRP ratios were directly compared to the windows in the library. Only the CapDetB/CapDetA response (1.59×10^0) was within the corresponding window (1-3), whereas CapDetB/AiPD (1.67×10^{-1}) was not within the corresponding window (9.81×10^{-3} to 8.83×10^{-2}), and CapDetA/AiPD (1.05×10^{-2}) was also not within the corresponding window (7.60×10^{-3} to 6.84×10^{-2}). Therefore, S_{BA} was assigned 1, whereas S_{BD} and S_{AD} were assigned 0, producing a S_{Total} of only 0.33. This false negative recognition resulted from the coelution of MIBK, which altered the DRP. In the future, this problem can be addressed by incorporating MIBK into the library and implementing the algorithm to handle coeluting chemicals.

For Peak 2.9, while butyl acetate was correctly reported with $S_{Total} = 0.67$, 2,3-butanediol was reported with $S_{Total} = 1$. This was because both 2,3-butanediol and butyl acetate had relatively similar nominal retention times (123.6 s vs 129.3 s), so their high-confidence retention time windows both covered the retention time of Peak 2.9 (at 129.6 s); additionally, they had relatively similar DRPs. Evidently, the retention time of Peak 2.9 was much closer to that of butyl acetate than 2,3-butanediol. If the retention time windows can be narrowed to, for example, within $\pm 3\%$ of the nominal retention time, the high-confidence window of 2,3-butanediol becomes 119.9-127.3s, whereas that of butyl acetate becomes 125.5-133.3s. In this scenario, the retention time of Peak 2.9 only falls within the high-confidence window of butyl acetate.

For the same reason as Peak 2,9, both 1-octanol and 2-nonanone were reported for Peak 3.9, although only the former was truly in the sample. However, the nominal retention times of 2-nonanone (51.7 s) and 1-octanol (51.8 s) are almost identical. Therefore, for these two chemicals, simply narrowing the retention time window cannot improve recognition. Differentiation of these two chemicals may require narrower DRP windows.

The third example is a chromatogram with phosphonate esters (DMMP, DEMP, and DIMP), which are surface adsorptive chemicals. The chromatogram contains 21 peaks. The chromatogram in Fig. 3.8 shows all the peaks with known identities and whether they are in the chemical library. All the known chemicals were in the library. After the algorithm performed chemical recognition, the output (Table 3.5) showed that all the chemicals that were within the library were correctly recognized, with a total score of 0.67 or higher. Peak 2.2 resulted from system outgassing, but the algorithm recognized it as toluene; a possible solution to addressing this problem is presented later.

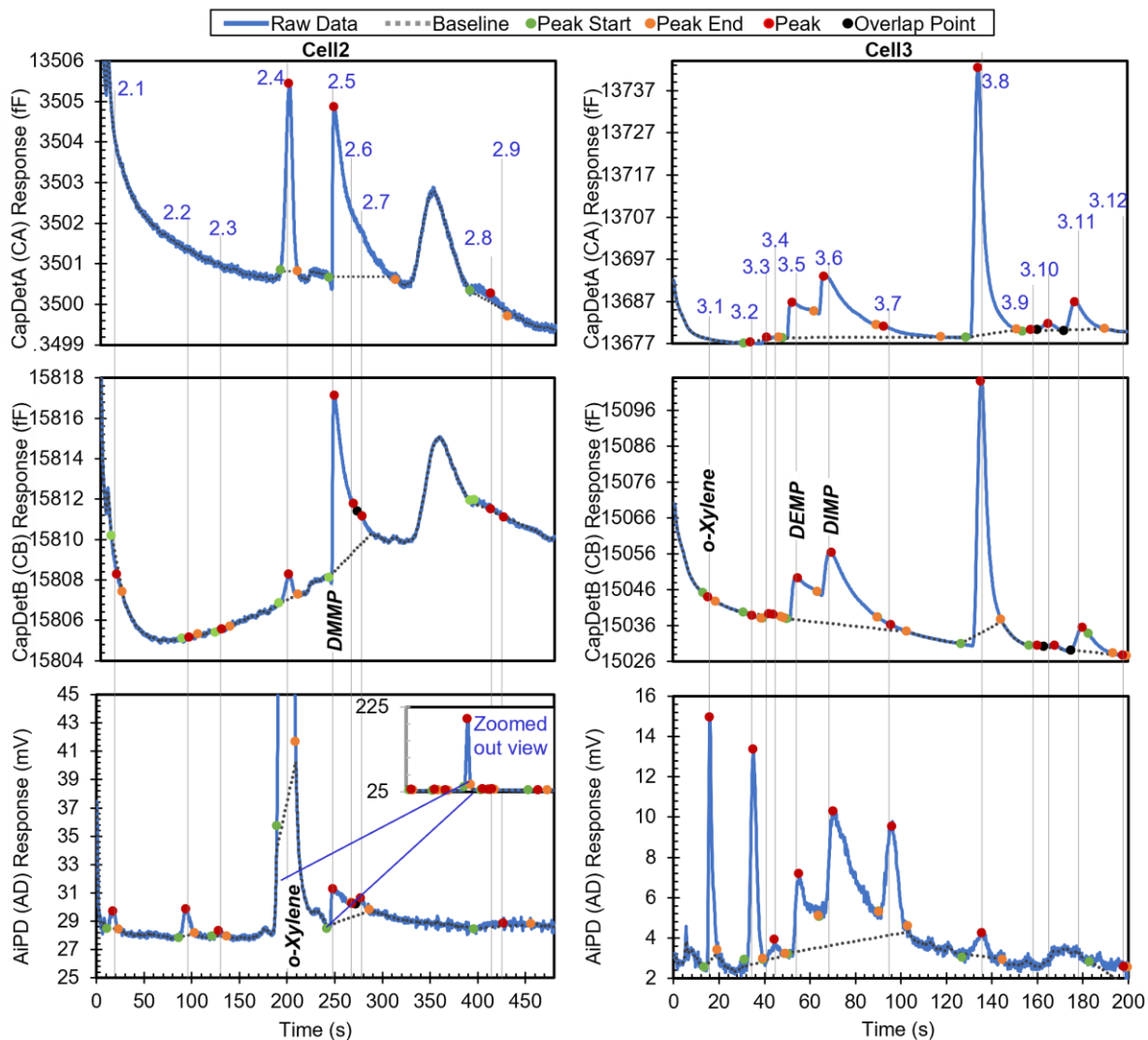


Figure 3.8: Raw chromatograms and detected peaks of the third example. For the peaks with known identities, the identities are labeled. Known peaks present in the library are labeled in bold italic.

When the user specified the use of *o*-xylene as the reference chemical, the phosphonate esters were also correctly recognized and quantified in relative quantities (Table 3.5). In this process, the algorithm first searched for *o*-xylene in both cells and correctly recognized Peaks 2.4 and 3.1 as *o*-xylene. Next, retention times of *o*-xylene in these two cells were set as reference retention times, with 199.7 s for Cell2 and 15.9 s for Cell3. Take Peak 2.5 as an example. It had a retention time of 248.1 s in Cell2 and therefore a relative retention time of 1.2 (*i.e.*, 248.1 s / 199.7 s). In the first step, based on the peak asymmetry (24.5) and the CapDetA and CapDetB

peak heights (both being positive), the peak underwent special treatment. In Cell2, the only surface adsorptive chemical in the library was DMMP. Using Eq. (1) and the detected CapDetA peak height, the projected retention time for DMMP was 245.3 s. With a nominal retention time of 196.5 s in the library for o-xylene in Cell2, the projected nominal relative retention time of DMMP was therefore 1.2, which well matched the relative retention time of Peak 2.5. Therefore, S_{iR} was assigned 1. Subsequently, the recognition based on the DRP was performed as previously discussed. Although most peaks that resulted from system outgassing were correctly treated as unknowns, some may have affected the recognition results. For instance, Peak 2.2 in the third example was incorrectly reported as toluene, whereas Peak 2.4 in the first example (which constituted both system outgassing and toluene from the sample) was possibly overestimated in concentration. One solution to address these problems is to incorporate the features of the outgassing peak into the chemical library, including not only the retention times and the DRPs but also the expected peak heights. As a result, these peaks can be correctly recognized as outgassing chemicals and separated from true unknown chemicals in the sample. Furthermore, if the outgassing peaks are well characterized, their contributions to the peak heights of the actual analyte chemicals can be subtracted before quantifying the analyte chemicals.

Table 3.5: Recognition results of the third example. The labels on the left indicate comparison to the ground truth. The columns in light blue background are results with the use of o-xylene as the reference chemical. For this example, other columns resulting from the use of the reference chemical are identical as those without using a reference chemical and therefore not repeated here.

Ground Truth	Output from chemical recognition algorithm															
	Chemical Number	Chemical Name	t _R (s)	Asym	CapDetA (fF)	CapDetB (fF)	AiPD (mV)	S _{rR}	S _{BA}	S _{AD}	S _{BD}	S _{total}	C (ppb)	t _{R,r}	C _r	
Unknown	✓	2.1.(1)	UnknownChem#1	17.6	0.81	0	-0.51	1.3	0	0	0	0		0.1		
Unknown	✗	2.2.(1)	Toluene	94.1	1.44	0	0.03	1.90	1	0	1	1	0.67	2.07	0.5	0.004
Unknown	✓	2.3.(1)	2,3-Butanediol	128.4	1.55	0	0.05	0.36	1	0	0	0		0.6		
Unknown	✓	2.3.(2)	Butyl Acetate	128.4	1.55	0	0.05	0.36	1	0	0	0		0.6		
Unknown	✓	2.3.(3)	Octane	128.4	1.55	0	0.05	0.36	1	0	0	0		0.6		
o-Xylene	✓	2.4.(1)	o-Xylene	199.7	0.93	4.58	1.23	159.35	1	1	1	1	470.06	1.0	1	
DMMP	✓	2.5.(1)	DMMP	248.1	24.50	4.17	8.65	2.44	1	1	1	1	26.61	1.2	0.057	
Unknown	✓	2.6.(1)	UnknownChem#2	268.1	35.00	0	2.33	0.99	0	0	0	0		1.3		
Unknown	✓	2.7.(1)	UnknownChem#3	277.2	22.52	0	1.22	1.18	0	0	0	0		1.4		
Unknown	✓	2.8.(1)	UnknownChem#4	413.0	0.74	0.21	0.09	0	0	0	0	0		2.1		
Unknown	✓	2.9.(1)	UnknownChem#5	427.1	0.96	0	-0.01	0.35	0	0	0	0		2.1		
o-Xylene	✓	3.1.(1)	o-Xylene	15.9	3.38	0	-0.15	11.99	1	0	1	1	0.67	1.0		
Unknown	✓	3.2.(1)	Decane	35.1	0.04	-0.18	-0.63	10.59	1	0	0	1	0.33	2.2		
Unknown	✓	3.2.(2)	Limonene	35.1	0.04	-0.18	-0.63	10.59	1	0	0	0		2.2		
Unknown	✓	3.3.(1)	UnknownChem#6	42.6	4.00	0.37	0.50	0	0	0	0	0		2.7		
Unknown	✓	3.4.(1)	UnknownChem#7	44.5	1.38	0	0.54	0.78	0	0	0	0		2.8		
DEMP	✓	3.5.(1)	DEMP	55.2	4.57	8.32	11.68	3.64	1	1	1	1	20.79	3.5	0.044	
Unknown	✓	3.5.(2)	2-Nonanone	55.2	4.57	8.32	11.68	3.64	0.5	1	0	0	0.17	3.5		
Unknown	✓	3.5.(3)	1-Octanol	55.2	4.57	8.32	11.68	3.64	0.5	0	0	0		3.5		
DIMP	✓	3.6.(1)	DIMP	70.2	10.59	14.46	19.70	6.60	1	1	1	1	30.54	4.4	0.065	
Unknown	✓	3.7.(1)	DIMP	96.1	96.00	2.48	1.42	5.40	0.5	0	1	0	0.17	6.0		
Dodecane	✓	3.8.(1)	UnknownChem#8	135.7	3.16	62.76	69.25	1.18	0	0	0	0		8.5		
Unknown	✓	3.9.(1)	UnknownChem#9	160.6	1.05	0.32	0.20	0	0	0	0	0		10.0		
Unknown	✓	3.10.(1)	UnknownChem#10	168.3	1.13	1.45	0.77	0	0	0	0	0		11.0		
Unknown	✓	3.11.(1)	UnknownChem#11	180.4	2.73	6.59	6.52	0	0	0	0	0		11.0		
Unknown	✓	3.12.(1)	UnknownChem#12	198.1	0.04	0	-0.04	0.94	0	0	0	0		13.0		

Legend:

- ✓ Correct recognition of a peak as a chemical in the library
- ✗ False positive
- ✓ Correct recognition for true unknowns and chemicals not in the library

Other false positive cases resulted from chemicals with similar parameters, such as Peak 2.9 (butyl acetate and 2,3-butanediol) and Peak 3.9 (2-nonanone and 1-octanol) in the second example. Although further narrowing the retention time and DRP windows may enhance differentiation between these chemicals, it requires more experimental characterization of the system hardware to provide more statistics on the variations of these parameters. Additionally, as the AiPD exhibits non-linear response to chemical concentrations [Lia23], it requires

additional considerations. Regardless, for an expert user who recognizes the MPCA system as a chemical screening and early warning system rather than a fingerprint-type of identification system, the current style of reporting multiple possibilities for each peak may be considered a useful feature.

To quantitatively summarize on the recognition accuracy, the recognition results were categorized to form a confusion matrix (Table 3.6). For each peak, if its recognition result included the correct chemical identity, it was considered a true positive ; if a peak from any chemical outside the library was falsely recognized as a chemical in the library, it was considered a false positive; if its recognition result included both a correct chemical identity and an incorrect chemical identity, it was considered both a true positive and a false positive (*e.g.*, Peak 2.9 in the second example). Additionally, if a peak from any chemical outside the library was reported as unknown, it was considered as a true negative; if a peak was from any chemical in the library but was reported as unknown, it was considered a false negative. Overall, the chemical recognition algorithm in this work provided a true positive rate of 96.3%, a true negative rate of 94.1%. a false positive rate of 5.9%, and a false negative rate of 3.7%.

Table 3.6: Confusion matrix based on the recognition results.

True positive 96.3%	False negative 3.7%
False positive 5.9%	True negative 94.1%

Compared to the statistical methods, the advantages of this work are summarized as follows:

- A viable library can be constructed using small data sets (7 total for non-special case chemicals) and expert knowledge.

- The causal relationship between the recognition score, the DRP, and expert knowledge increases the traceability of the algorithm.
- Special rules for small signals and for surface absorptive chemicals enhance the recognition for complex samples.

3.4 Conclusions

Overall, the algorithm reported in this work provided chemical recognition for a μ GC system with three complementary types of detectors and achieved its intended performance. Recognition rules for the retention time and the detector response pattern were developed based on both the physical attributes and expert understanding of the hardware. Despite the use of relatively coarse windows for individual parameters, the combined use of these recognition rules enabled chemical recognition in complex chromatograms, overcoming the uncertainty of peak information resulting from low signal-to-noise ratios, asymmetry, and overlapping peaks. Additionally, customized special treatments were implemented to further address these problems. The recognition capabilities of the algorithm were illustrated in three examples, with all the results well expected and explainable. Such a chemical recognition algorithm requires only a small amount of experimental data, because typically one experimental run is enough to establish the nominal responses for multiple chemicals. Therefore, this algorithm shows the prospects of reducing the burden on system calibration while providing satisfactory results to expert users.

Although this work adopts fuzzy logic, the membership functions applied to the individual parameters are relatively simple, *i.e.*, with step functions providing binary results to a small number of discrete windows. With more experimental characterization and better understanding of the parameter variations, probability density functions can be established and

used as the membership functions, thus providing more granularity to the individual recognition scores.

To further improve the algorithm, future work may include the following. 1) Characterize the outgassing peaks and add them into the library. 2) Conduct more experimental characterization on the chemical response variations, which can provide a better quantitative basis for setting the membership functions and likelihood windows. 3) Implement an algorithm to handle coeluting chemicals within a peak. 4) Expand the library to include more chemicals.

References

- [Car72] G.C. Carle R.W. Donaldson, S.C. Terry, K.D. Wise *Microminiature gas chromatograph*; NASA Tech Briefs 1972.
- [Alw18] A. Alwosheel, S. van Cranenburgh, C.G. Chorus, “Is your dataset big enough? sample size requirements when using artificial neural networks for discrete choice analysis,” *Journal of Choice Modelling*, vol. 28, pp. 167–182, 2018.
- [Bac21] G. Baccolo, B. Quintanilla-Casas, S. Vichi, D. Augustijn, R. Bro, “From untargeted chemical profiling to peak tables – a fully automated AI driven approach to untargeted GC-MS,” *Trends in Analytical Chemistry*, vol. 145, p. 116451, 2021.
- [Beh11] V. Behrends, G. D. Tredwell, J.G. Bundy, “A software complement to AMDIS for processing GC-MS Metabolomic Data,” *Analytical Biochemistry*, vol. 415, no. 2, pp. 206–208, 2011.
- [Baf84] T. Blaffert, “Computer-assisted multicomponent spectral analysis with fuzzy data sets,” *Analytica Chimica Acta*, vol. 161, pp. 135–148, 1984.
- [Bel13] C. Beleites, U. Neugebauer, T. Bocklitz, C. Krafft, J. Popp, “Sample size planning for classification models,” *Analytica Chimica Acta*, vol. 760, pp. 25–33, 2013.
- [Gar11] E. García-Berríos, T. Gao, J.C. Theriot, M.D. Woodka, B.S. Brunshwig, N.S. Lewis, “Response and discrimination performance of arrays of organothiol-capped au nanoparticle chemiresistive vapor sensors,” *The Journal of Physical Chemistry C*, vol. 115, no. 14, pp. 6208–6217, 2011.
- [Hov89] H.T. Hove, O. Grahl-Nielsen, A. Rogstad, “Assay for dinoflagellate toxins in mussels by gas chromatography and Principal Components Analysis,” *Analytica Chimica Acta*, vol. 222, no. 1, pp. 35–42, 1989.
- [Hu18] J. Hu, H. Qu, Y. Chang, W. Pang, Q. Zhang, J. Liu, X. Duan, “Miniaturized polymer coated film bulk acoustic wave resonator sensor array for quantitative gas chromatographic analysis,” *Sensors and Actuators B: Chemical*, vol. 274, pp. 419–426, 2018.
- [Jin09] C. Jin, E.T. Zellers, “Chemometric analysis of gas chromatographic peaks measured with a microsensor array: Methodology and performance assessment,” *Sensors and Actuators B: Chemical*, vol. 139, no. 2, pp. 548–556, 2009.
- [Jol16] I.T. Jolliffe, J. Cadima, “Principal component analysis: A review and recent developments,” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2065, p. 20150202, 2016.

- [Kim11] S.K. Kim, H. Chang, E.T. Zellers, “Microfabricated Gas Chromatograph for the selective determination of trichloroethylene vapor at sub-parts-per-billion concentrations in complex mixtures,” *Analytical Chemistry*, vol. 83, no. 18, pp. 7198–7206, 2011.
- [Li10] M. Li, E.B. Myers, H.X. Tang, S.J. Aldridge, H. C. McCaig, J.J. Whiting, R.J. Simonson, N.S. Lewis, M. L. Roukes, “Nanoelectromechanical resonator arrays for ultrafast, gas-phase chromatographic chemical analysis,” *Nano Letters*, vol. 10, no. 10, pp. 3899–3903, 2010.
- [Lia23] W. Liao, D. Winship, I. Lara-Ibeas, X. Zhao, Q. Xu, H. Lu, T. Qian, R. Gordenker, Y. Qin, Y.B. Gianchandani, “Highly Integrated μ GC Based on a Multi-sensing Progressive Cellular Architecture with a Valveless Sample Inlet,” *Analytical Chemistry*, vol. 95, no.4, pp. 2157-2167, 2023.
- [Lov94] J.H. Love, “Gas chromatograph peak identification using statistically determined fuzzy logic membership functions,” *IEEE International Conference on Neural Networks (ICNN)*, Orlando, FL, USA, pp. 3283-3288, 1994.
- [Mat20] D.D. Matyushin, A.Y. Sholokhova, A.K. Buryak, “Deep Learning Driven GC- MS Library Search and its application for Metabolomics,” *Analytical Chemistry*, vol. 92, no. 17, pp. 11818–11825, 2020.
- [Men19] J.M. Mendel, *Uncertain Rule-Based Fuzzy Systems: Introduction and New Directions 2nd Edition*, Springer, 2019
- [Ott86] M. Otto H. Bandemer, “Pattern recognition based on fuzzy observations for spectroscopic quality control and chromatographic fingerprinting,” *Analytica Chimica Acta*, vol. 184, pp. 21–31, 1986.
- [Par00] J. Park, “Personal Monitoring Instrument for the selective simultaneous measurement of multiple organic vapors,” *American Industrial Hygiene Association journal*, 2000.
- [Par99] J. Park, W.A. Groves, E.T. Zellers, “Vapor recognition with small arrays of polymer-coated microsensors. A comprehensive analysis,” *Analytical Chemistry*, vol. 71, no. 17, pp. 3877–3886, 1999.
- [Qin16] Y. Qin Y.B. Gianchandani, “A fully electronic microfabricated gas chromatograph with complementary capacitive detectors for indoor pollutants,” *Microsystems & Nanoengineering*, vol. 2, no. 1, 2016.
- [Rea07] P. Rearden, P.B. Harrington, J.J. Karnes, C.E. Bunker, “Fuzzy rule-building expert system classification of fuel using solid-phase microextraction two-way gas chromatography differential mobility Spectrometric Data,” *Analytical Chemistry*, vol. 79, no. 4, pp. 1485–1491, 2007.

- [Reg18] B.P. Regmi M. Agah, “Micro Gas Chromatography: An overview of critical components and their integration,” *Analytical Chemistry*, vol. 90, no. 22, pp. 13133–13150, 2018.
- [Sen06] L.R. Senesac, P. Dutta, P.G. Datskos M.J. Sepaniak, “Analyte species and concentration identification using differentially functionalized microcantilever arrays and artificial neural networks” *Analytica Chimica Acta*, 558, pp. 94-101, 2006.
- [Sho22] A.Y. Sholokhova, O.I. Grinevich, D.D. Matyushin, A.K. Buryak, “Machine learning-assisted non-target analysis of a highly complex mixture of possible toxic unsymmetrical dimethylhydrazine transformation products with chromatography-mass spectrometry,” *Chemosphere*, vol. 307, p. 135764, 2022.
- [Ste99] S.E. Stein, “An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data,” *Journal of the American Society for Mass Spectrometry*, vol. 10, no. 8, pp. 770–781, 1999.
- [Ter79] S.C. Terry, J.H. Jerman, J.B. Angell, “A gas chromatographic air analyzer fabricated on a silicon wafer,” *IEEE Transactions on Electron Devices*, vol. 26, no. 12, pp. 1880–1886, 1979.
- [Wan19] J. Wang, N. Nuñoovero, R. Nidetz, S.J. Peterson, B.M. Brookover, W.H. Steinecker, E.T. Zellers, “Belt-mounted micro-gas-chromatograph prototype for determining personal exposures to volatile-organic-compound mixture components,” *Analytical Chemistry*, vol. 91, no. 7, pp. 4747–4754, 2019.
- [Wil20] 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,” *Analytical Chemistry*, vol. 92, no. 20, pp. 13953–13960, 2020.
- [Xu23] Q. Xu, Y. Qin, Y.B. Gianchandani, “A Rule-Based Automated Chemical Recognition Algorithm for a Multi-Cell Multi-Detector Micro Gas Chromatograph,” *Separations*, vol 10, no.11, p 555, 2023.
- [Zad65] L.A. Zadeh, “Fuzzy Sets”, *Information and Control*, vol 8, no 3, 1965.
- [Zha23] X. Zhao, R. Aridi, J. Hume, S. Subbiah, X. Wu, H. Chung, Y. Qin, Y.B. Gianchandani, “Automatic Peak Detection Algorithm Based on Continuous Wavelet Transform for Complex Chromatograms from Multi-detector, Micro-scale Gas Chromatographs,” *Journal of Chromatography A*, p. 464582, 2023.
- [Zel95] E.T. Zellers, S.A. Batterman, M. Han, S.J. Patrash, “Optimal coating selection for the analysis of organic vapor mixtures with polymer-coated surface acoustic wave sensor arrays,” *Analytical Chemistry*, vol. 67, no. 6, pp. 1092–1106, 1995.

Chapter 4 Conclusions and Future Work

This chapter provides reviews of the conclusions of the work performed for this dissertation and points the way to possible future improvements. In the conclusion section, the main contributions of this work are identified, the advancements towards μ GC automation in control and data analysis are recognized. The possible future improvements discuss the possible ways the control software can be improved to enhance autonomous and automated controls, and the possible ways the recognition algorithm can be improved to enhance chemical recognition.

4.1 Conclusions

This thesis describes the design, implementation and validation of an automated control software and automated chemical recognition algorithm for a μ GC. The contents of this work increase the level of automation of a highly complex μ GC system and enable fully automated *in situ* analysis of a highly complex μ GC. The μ GC test results generated from the automated control software are used directly by an automated peak detection algorithm. The automated chemical recognition algorithm in this work uses the results from the peak detection algorithm to generate a chemical recognition report. In particular, the automated control software uses a multithreaded architecture to support the concurrent controls of various μ GC components, and the chemical recognition algorithm uses a rule base expert system for automated recognition. The contents of this work are essential to the full automation of a multicell, multidetector μ GC system and show promise for border use in μ GC.

The main contributions of this work include:

1. Design and implementation of multithreaded control to be used in highly complex μ GC systems. This architecture solves the problems of:

- a. controlling components that utilize heterogeneous controls mechanisms.
 - b. controlling components with heterogeneous data rates.
 - c. closed loop control of both temperature and flow.
2. Showed a software design path for μ GC that can be applied to other μ GC systems as the independent threads used to control different components can be turned off or switched to other functions.
3. Design and implementation of rule based expert system for chemical recognition, which leverages expert knowledge of the system to create a recognition ruleset with a minimal training set.
4. Developed special rules for the recognition to support:
 - a. recognition using reference chemicals
 - b. recognition of chemicals with small signals
 - c. recognition of surface adsorptive chemicals

The challenges to automating control and data collection of highly complex μ GC systems were identified and investigated. Automated control software was developed for a representative highly complex μ GC. The software was developed to control and interrogate a variety of electronics that have heterogeneous control mechanisms and timing requirements. A UI was designed to interface with non-specialist users and provide real time control and updates. The software utilized multithreaded architecture to control all components of the μ GC. For the MPCA system, six threads were needed; the main thread which communicates with an external user interface and manages run parameters, the temperature control thread, capacitance detector readout thread, the valve control thread, the pump control thread, and the photoionization detector readout thread. The temperature control thread reported the temperature with an analog

to digital converter and controlled temperature either with pulse width modulation or by analog voltage control by controlling a digital potentiometer. The capacitive detector thread reads the capacitance with a capacitance to digital converter. The valve control thread controlled the latching valves by supplying pulses to the valve control circuit. The pump control thread controlled pumps with either an on off control or a hardware PWM pump. The closed loop control of the pump was realized by reading out differential pressure sensors and changing the pump signal frequency. The AiPD thread controlled a UVU lamp and read out the result from a transimpedance amplifier.

All the sensors and detectors had varying data rates. Uniform data rate is essential to control and post processing of data. To ensure data rate and data quality, a time buffer to the cycle time was implemented. A UI was implemented to provide remote control and remote monitoring. The software was validated with test runs. The software was able to achieve viable control for temperature and flow. The software also successfully generated viable chromatograms with high data rate uniformity. The relative standard deviations of all loops were all within <0.5% of the intended time interval between measurements. The control software also supported >1000 μ GC runs that analyzed various chemical mixtures, showcasing its reliability. The different runs were also conducted in various ambient temperatures (4.9°C – 30.2°C) and similar control results were observed. As the key components of MPCA undergo closed loop heating control to temperature greater than that of the ambient, the ambient temperature has minimal effect on the system.

The improved state of control automation moved μ GC closer to being able to perform *in situ* autonomous and automated analysis without human intervention. In particular, the success of the controls demonstrated the software architecture handles a variety of hardware control

mechanisms simultaneously; the successful control of timing demonstrated the software architecture supports the data acquisition requirements for automated analysis of a highly complex μ GC systems. The success of the control software not only showed advancements in the state of automation for a highly complex μ GC but also presented a pathway for the automation of future complex μ GC designs.

The challenges to automating a multi-sensing, multi-cell μ GC were investigated and an alternative approach to commonly used heuristic methods was explored. As a result, an automated algorithm for chemical recognition based on rule based expert systems was developed. The algorithm was designed specifically to address the challenges for automating chemical recognition for a representative μ GC for *in situ* tests. The algorithm was designed to take advantage of recent (and potentially future) hardware developments, incorporate expert knowledge, and reduce the needs for extensive hardware characterization. The algorithm used cell number, retention time, detector response patterns of the detector, and peak asymmetry as recognition parameters. A knowledge base library based on a minimal set of data acquired from calibration runs that any system would require was created. Rules for recognition for general case chemicals, surface adsorptive chemicals, small signals and recognition using reference chemicals were created based on expert knowledge.

By using coarse likelihood windows, the rules for DRPs addressed complex scenarios of DRP variations that might result from electronic noise, chromatogram baseline variation, or nonlinearity of detector responses to concentration. Therefore, the algorithm was applicable to a wide range of chemicals and concentrations. The expert rules for the system were obtained from both expert knowledge of the sensing principles (*e.g.*, CapDetB/CapDetA reflecting polarity) and the data collected in the calibration runs (*e.g.*, CapDetA/AiPD, CapDetB/AiPD).

The algorithm was tested on chromatograms with 21–31 peaks for each detector. Overall, the algorithm achieved 96.3% true positive rate, 94.1% true negative rate, 5.9% false positive rate, and 3.7% true negative rate. The overall reliability of the algorithm demonstrated it can be used to automate μ GC chemical recognition, especially for chemical screening and early warning applications.

Compared to heuristically used statistical and machine learning methods, one major advantage of the rule-based recognition algorithm is that it did not require massive amounts of training data. Only expert knowledge of the system and a minimal set of calibration data were needed. Removing the requirement of massive amounts of training data significantly speeds up the process of automation, as it removes the need for extensive experimental characterization, thereby reducing cost and time needed for automation.

Another advantage of a rule-based system is the ease of incorporating (and changing) special rules for recognition. Expert knowledge can be applied directly to special rules (*i.e.*, surface adsorptive chemicals and peaks with small signal-to-noise ratios.) In heuristically used statistical methods, the special rules may be difficult to directly incorporate into existing models and the special cases (sometimes considered outliers) can skew statistical analysis results. The ease of incorporating new rules can enable the rapid enhancement of automated recognition.

The rule-based recognition algorithm showed the viability of a different recognition paradigm. Instead of relying on statistics, the rule-based recognition algorithm relied more on mathematical expressions of expert knowledge. The success of this algorithm showed that this paradigm can be used in the automated recognition of high complex μ GCs.

Overall, the first half of this work provided automated control and data collection of μ GCs. The second half of this work provided a recognition algorithm that analyzes the peak

information of the chromatograms generated by the control algorithm. This work increased the automation level of μ GCs in hardware control and data analytics, improved the level of automation for μ GCs, and moved μ GCs closer to being used as a fully automated and autonomous gas phase analysis instrument suitable for *in-situ* analysis.

4.2 Future Work

The automated control software and automated chemical recognition algorithm demonstrate pathways for μ GC systems to becoming fully automated *in situ* test system. However, there are opportunities to improve the controls of automation and the success rate of recognition.

As reported in Chapter 2, the varying cycle wait time was not implemented for every thread. In the future, varying cycle wait time can be implemented for all control threads to improve data rate uniformity. A fully automated *in situ* system often operates on a limited power budget. Although the power budget is already considered in the control software, opportunities for further reducing power consumption exist. The current embedded computer is a multi-core platform and uses significantly more power than single core systems. As the current control software architecture already provisions for single core systems, the control software can be adapted to a single core system, and test runs can be done to analyze the performance of the algorithm. In addition, it is also possible to run the current EC in single core mode, which in theory could conserve power consumption. Testing for the single core mode could be tested and the performance characterized. Additionally, the current remote control of the μ GC is using wired ethernet connection. Remote control monitoring using wireless connections such as Wi-Fi or Bluetooth could improve remote sensing capabilities. As the control software shows great

promise for various μ GCs, additional control libraries can be expanded for additional electronics to suit a variety of μ GC setups, making the software more universal.

The recognition algorithm shows great promise for multicell multidetector μ GCs. However, the current likelihood windows are rough estimates based on expert knowledge. Although the coarse windows already show success, there are cases where recognition could be improved with finer likelihood windows. The likelihood windows are also simple step or rectangular functions. The windows can be fuzzified further to output more values or form continuous membership functions. The improved membership functions may improve recognition results. Both refining the windows and fuzzifying the windows can be done by recording more experimental data. Therefore, an algorithm to record the experimental parameters and to statically determine the likelihood windows could be developed. The weights for the scoring of the DRPs can also be investigated, and improving the weights for the scoring may also improve overall recognition. The algorithm to improve likelihood windows can be deployed while the current recognition algorithm is in use. The recognition algorithm can therefore be that of a combination of statistical tools and rules based on expert knowledge that leverages the advantages of each method to cover for the weakness of the other.

Currently, the chemical recognition algorithm has been running on a modern Windows laptop (as of 2024). The estimated run time for the most complex chromatograms for MPCA is 5-10 seconds. However, as the chemical library expands, and as the algorithm is incorporated into the EC for hardware control, the computational time for the algorithm may increase due to the need to look up a larger library and reduced computational power on an EC. The burden of searching in a larger library can be reduced by implementing improved search methods, better data indexing and better data structures (*i.e.*, *hashmaps*).

Machine learning techniques can sometimes be applied to extract features and rules for recognition as well. However, the extraction of these features and rules are usually reliant on massive amounts of training data. As more data is collected, machine learning techniques may be applied and a comparison between these two techniques can be made.

Appendix A: Other Publications and Manuscripts

The contents of this work have been used to support the development and evaluation of μ GCs. The publications that use this work are referred to in the following sections.

A.1 Highly Integrated μ GC Based on a Multisensing Progressive Cellular Architecture with a Valveless Sample Inlet

Using this work, our group was able to automate the MPCA system and demonstrate a pathway to a high performance and highly manufacturable μ GC [Lia23]. The author of this work also benchmarked the MPCA system using heuristically used techniques. The contents of this Section have been published in *Analytical Chemistry*. The title of the paper is “Highly Integrated μ GC Based on a Multisensing Progressive Cellular Architecture with a Valveless Sample Inlet.”

A2. An Enhanced-performance Multisensing Progressive Cellular μ GC: Design Advances and Blind Test Results

Using this work, our group was able to automate the MPCA system and analyze its reliability [Win24]. From a control software viewpoint, the reliability of the MPCA system was improved in part by using closed loop control of flow, and the reliability of the software provided a platform for repeatability tests. The title of the paper is “An Enhanced-performance Multisensing Progressive Cellular μ GC: Design Advances and Blind Test Results” and the paper is in review.

References

- [Lia23] W. Liao, D. Winship, L. Lara-Ibeas, X. Zhao, Q. Xu, H. Lu, T. Qian, R. Gordenker, Y. Qin, Y.B. Gianchandani, “Highly Integrated μ GC Based on a Multi-sensing Progressive Cellular Architecture with a Valveless Sample Inlet,” *Analytical Chemistry*, vol 95, no. 4, pp. 2157–2167, 2023.
- [Win24] D. Winship, W. Liao, H. Lu, I. Lara-Ibeas, X. Zhao, Q. X, T, Qian, R. Gordenker, Y. Qin, Y.B. Gianchandani, “An Enhanced-performance Multisensing Progressive Cellular μ GC: Design Advances and Blind Test Results,” in review.