Portable Micro-Gas Chromatography with Multidimensional Compound Identification Analysis

Arjun Sharma

Thesis submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Master of Science

in

Computer Engineering

Leyla Nazhandali, Chair

Masoud Agah

Wenjie Xiong

February 13, 2023

Blacksburg, Virginia

Keywords: Micro Gas Chromatography, Chromatogram Analysis, Multidimensional Compound Identification

Copyright 2023, Arjun Sharma

Portable Micro-Gas Chromatography with Multidimensional Compound Identification Analysis

Arjun Sharma

(ABSTRACT)

Gas Chromatography (GC) is an analytical technique in the chemistry field widely used to separate compounds present in a sample mixture. Conventional GC systems are an extremely versatile and powerful tool to perform complex separations. However, these systems come with the cost of being bulky and requiring a high amount of power for operation. With considerable research for over 40 years, the advent of Micro-Gas Chromatography (μGC) made it possible for miniaturized, compact, low-power, and field portable GC systems. This thesis presents a portable μGC system that enables real-time analysis of complex compound separations, made possible with the use of multiple separation columns and a novel multidimensional compound identification algorithm. The system architecture and the software design with multiple features enabling portability of the μGC system are discussed. A set of microfabricated separation columns (μSCs) and photoionization detectors (PIDs) are integrated to realize a fully functional μGC system that is tested with different types of complex compound mixtures. An in-depth analysis of processing the output chromatograms obtained from the setup for signal filtering and peak detection is described in this thesis. A multidimensional analysis for compound identification in complex mixtures is presented.

Portable Micro-Gas Chromatography with Multidimensional Compound Identification Analysis

Arjun Sharma

(GENERAL AUDIENCE ABSTRACT)

Volatile organic compounds (VOCs) are generally chemicals that have high vapor pressure and low boiling points used and produced in the processing of petroleum products, paint, refrigerants, pharmaceuticals, and adhesives. VOCs are emitted as gases from certain solids or liquids, some of which may have short- and long-term adverse health effects even with minute exposure. Gas Chromatography (GC) is a common analytical technique used to detect, identify, and quantify VOCs in the environment, and conventional GC Systems have been utilized for this purpose. The separation of compounds occurs inside an analytical column that has selective interaction between the column and the analytes passing through. However, these systems are expensive, bulky, consume high power, and require expertise to operate. Recently, advancements in the Microelectromechanical systems (MEMS) field has paved the way to create Micro-Gas Chromatography (μGC) systems with improved performance when compared to traditional systems. Active research is ongoing to improve the portability of μGC systems for reliable and quick on-field analysis. In this thesis, we present a μGC system that has a robust and scalable design that allows the development of a portable μGC system. The compound separation of complex mixtures is showcased using the portable μGC system setup. The output chromatograms obtained from the μGC system are pre-processed, which involves noise filtering and peak detection, followed by an analysis using a multidimensional compound identification algorithm.

Dedication

To my family.

Acknowledgments

I wish to acknowledge the help provided by several individuals throughout my graduate studies at Virginia Tech. Firstly, I would like to convey my profound gratitude to my research advisor and committee chair, Dr. Leyla Nazhandali. Thank you for providing me with the opportunity to work with you and gently guiding me throughout my graduate studies. I greatly value our discussions, and your pertinent input on skills such as presentation and effective oral communication would help me greatly as I transition to corporate life.

I wish to thank Dr. Masoud Agah and Dr. Wenjie Xiong for agreeing to be part of my Master's advisory committee. To Dr. Masoud Agah, I relished working in the VT MEMS Lab, and enjoyed our conversations on various topics. I wish to extend my special thanks to Mustahsin Chowdhury. I am immensely grateful for your support and guidance on technical matters. To the members of VT MEMS Lab, with a special mention to Nipun and Poonam, thank you for your cheerful companionship.

I thank my close friends Sumaiyyah, Pallavi, Radhika, and Nikhil for providing emotional support, creating fond memories, and always being there for me. Finally, I would like to thank my loving and supportive family - my parents, Gayathri and Shivaprakash, and my brother Aditya. Thank you for your patience in resolving all my problems, both academic and non-academic. My sincerest gratitude to all of you for always looking out for me and keeping me in your prayers.

Contents

List of Figures ix

List of Tables xii

1 Introduction 1

1.1 Background 1

1.1.1 Conventional GC Systems 2

1.1.2 Micro-Gas Chromatography System (μGC System) 4

1.2 Motivation 5

1.3 Contributions 6

1.4 Organization of Thesis 6

2 Hardware Specification 8

2.1 Hardware peripheral components 8

2.1.1 Valves 8

2.1.2 Pumps 10

2.1.3 Heaters and Temperature Sensors 10

2.1.4 Fans 10

2.1.5 Detectors 11

2.1.6 Modular μGC stack 11

3 Software Specification **14**

3.1 Embedded Software Design and Features 15

3.1.1 Wireless Communication Mode 15

3.1.2 Sensor Read Frequency 18

3.1.3 Parallel Heater Temperature Programming 20

3.1.4 Scheduling Algorithm 25

3.2 Graphical User Interface 27

3.2.1 Experiment Configuration Tab 28

3.2.2 Data Visualization Tab 29

3.2.3 Analysis Tab 29

4 System Integration and Chromatogram Analysis **31**

4.1 Experiment setup 31

4.2 Experimentation 33

4.3 Chromatogram Analysis 36

4.3.1 Output Chromatogram Processing 36

4.3.2 Signal filtering 37

4.3.3 Peak Detection 51

4.3.4 Results 54

5 Multidimensional Compound Identification Analysis **56**

5.1 Retention Time Library Generation 56

5.2 Multidimensional Compound Identification Algorithm 57

5.2.1 Identifying compounds from chromatogram 57

5.2.2 Compound List Merging 59

5.3 Testing and Results 60

5.3.1 Test scenario 1 61

5.3.2 Test scenario 2 64

6 Conclusion **67**

6.1 Future Work 67

Bibliography **69**

Appendices **73**

Appendix A Sample Mixtures **74**

Appendix B Retention Time Library **76**

List of Figures

1.1 Conventional GC system 2

1.2 GC detector output 4

2.1 Peripherals used in the μGC System 9

2.2 Modular μGC stack hardware architecture 13

3.1 Software components of portable μGC system 14

3.2 Network configurations for μGC systems enabled through Wi-Fi 17

3.3 Ring Buffer Implementation 19

3.4 Fixed packet structure for the modular μGC system 20

3.5 Heater Finite State Machine 23

3.6 Temperature programming profiles using multiple heaters in parallel 24

3.7 RTOS Scheduling algorithm flowchart 26

3.8 GUI Experiment Configuration Tab 28

3.9 GUI Data Visualization Tab 29

3.10 GUI Analysis Tab 30

4.1 Experiment setup 32

4.2 GC schematic diagram 34

4.3 GC system output chromatograms using complex compound mixtures 35

4.4 Output Chromatogram Processing Flow 36

4.5 Single-Sided FFT of the input signal chromatogram 37

4.6 Moving average filtering of sample chromatogram with window of 3 datapoints 39

4.7 Moving average filtering of sample chromatogram with window of 19 datapoints 39

4.8 Moving average filtering of sample chromatogram with window of 75 datapoints 40

4.9 Lowpass Butterworth filtering of sample chromatogram with cutoff frequency 2.5Hz and filter order 1 41

4.10 Lowpass Butterworth filtering of sample chromatogram with cutoff frequency 2.5Hz and filter order 2 42

4.11 Lowpass Butterworth filtering of sample chromatogram with cutoff frequency 2.5Hz and filter order 6 42

4.12 Gaussian filtering of sample chromatogram with standard deviation 1 44

4.13 Gaussian filtering of sample chromatogram with standard deviation 3 44

4.14 Gaussian filtering of sample chromatogram with standard deviation 15 . . . 45

4.15 Savgol filtering of sample chromatogram with window size 5 datapoints and filter order 1 46

4.16 Savgol filtering of sample chromatogram with window size 5 datapoints and filter order 2 47

4.17 Savgol filtering of sample chromatogram with window size 5 datapoints and filter order 4 47

4.18 Savgol filtering of sample chromatogram with filter order 1 and window size 5 datapoints 48

4.19 Savgol filtering of sample chromatogram with filter order 1 and window size 10 datapoints 48

4.20 Savgol filtering of sample chromatogram with filter order 1 and window size 50 datapoints 49

4.21 Varying peak prominence of peak detection in sample output chromatogram 52

4.22 Varying peak width of peak detection in sample output chromatogram . . . 53

5.1 Multidimensional compound identification algorithm 58

5.2 Compound list merging algorithm 60

5.3 Chromatograms from different stationary phases for test mixture 1 63

5.4 Chromatograms from different stationary phases for test mixture 2 66

List of Tables

2.1 Modular μGC Stack Specifications 12

4.1 Moving Average Filter 40

4.2 Lowpass Butterworth Filter (Cutoff Frequency = 2 Hz) 43

4.3 Lowpass Butterworth Filter (Cutoff Frequency = 2.5 Hz) 43

4.4 Gaussian Filter 45

4.5 Savitzky Golay Filter (Window = 5 datapoints) 49

4.6 Savitzky Golay Filter (Window = 10 datapoints) 50

4.7 Savitzky Golay Filter (Window = 20 datapoints) 50

4.8 Savitzky Golay Filter (Window = 50 datapoints) 50

4.9 Fixed chromatogram processing tuning parameters 54

4.10 Peak detection accuracy for six different types of chromatogram 55

5.1 Outcomes of compound identification in a given stationary phase 59

5.2 Compound identification examples 60

5.3 Test scenario 1 multidimensional compound identification result 62

5.4 Test scenario 2 multidimensional compound identification result 65

A.1 Sample Mixtures 74

B.1 Retention Time Library 76

Chapter 1

Introduction

The foundation for the development of gas chromatography (GC) began in the early 1950s after the introduction of liquid-gas chromatography. At the time, gas chromatography quickly gained acceptance as the demand for improved analytical controls boomed in the petrochemical industries. Since then, gas chromatography analysis has matured into a widely used analysis technique for separating and analyzing most organic compounds relevant to a wide range of industries and areas of the research study [11].

The GC process typically involves introducing vapors of a sample mixture into an analytical column which is coated with a specific type of stationary phase. As the mixture travels through the analytical column, the components get separated based on the interaction with the stationary phase. The compounds eluting from the column are detected, and a chromatogram is generated. The chromatogram can be further analyzed to obtain quantitative and qualitative information about the compounds detected in the sample mixture [6].

1.1 Background

In this section, we briefly describe the working principles of a conventional GC system. Further, we introduce the μGC system and describe its components in detail.

1.1.1 Conventional GC Systems

A conventional GC system typically has several components, such as the injector, separation column, detector, and data analysis software, as shown in Figure 1.1 [23].

The injector is used to introduce the sample into the system. There are several types of injectors, such as split/splitless, on-column, and programmed temperature vaporizers. The split/splitless injector is the most common type of injector used in GC systems which has two operational modes. In the split mode, it introduces a small portion of the sample into the column and discards the remaining portion, while in the splitless mode, it allows the entire sample to be injected into the column. The injector is an important component in the GC system that determines the quality and reproducibility of the analysis. It must be carefully maintained and operated to ensure minimal contamination to obtain accurate results.

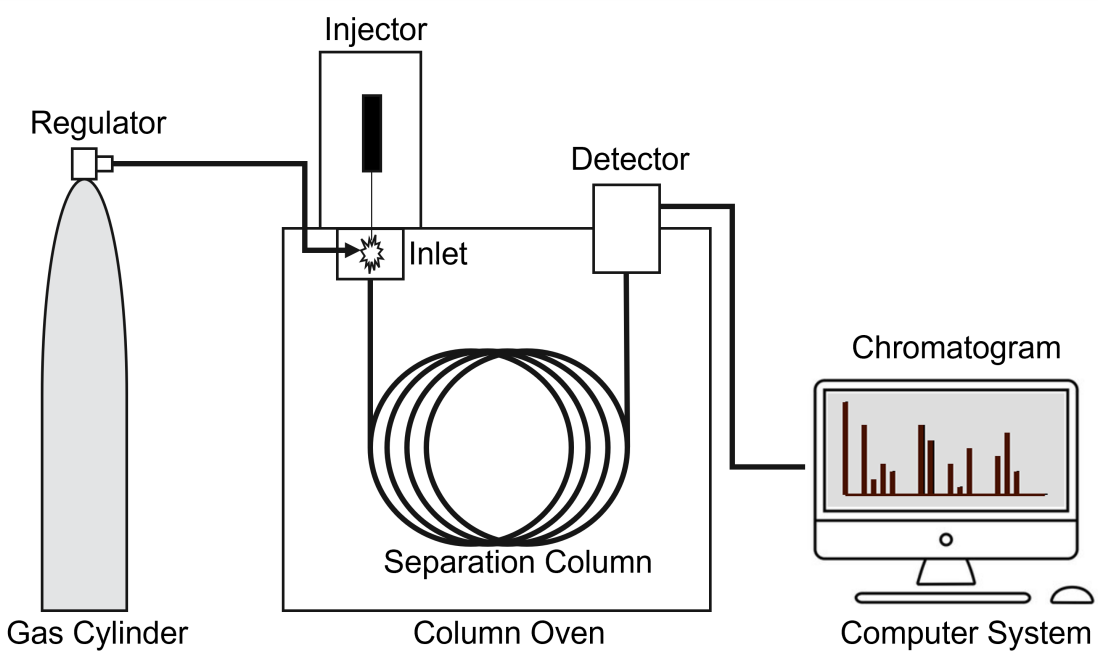


Figure 1.1: Conventional GC system

The carrier gas is used to carry the sample vapors through the separation column. Some

commonly used carrier gases are Hydrogen, Helium, and Nitrogen. The separation column is the heart of the GC system. It is a long, narrow tube packed with a stationary phase that separates the compounds in the sample. The stationary phase is a material that is packed inside the column that the compounds in the sample mixture interact with. As the compounds pass through the separation column, the different extent of interaction with the stationary phase results in varied elution times for the compounds [9], [7]. Different kinds of separation columns can be used in GC systems, namely, packed columns and capillary columns. The choice of selecting the separation column is dependent on the resolving power required for the target application. Additionally, the temperature of the separation column plays a crucial role in the accurate analysis of the compounds present in the sample mixture.

The detectors placed at the end of the separation columns are used to detect the elution of the compounds from the column. There are several types of detectors used in GC systems, such as Thermal Conductivity Detector (TCD), Flame Ionization Detector (FID), Electron Capture Detector (ECD), Mass Spectrometry Detector (MSD), and Photo Ionization Detector (PID) [11]. The nature of the compounds present in the sample mixture and the purpose of analysis determines the choice of detectors to be used in the system. Multiple detectors can be used to increase the amount of information obtained from the sample.

The data system is responsible for collecting and analyzing the data generated by the detector. It typically includes a computer and custom software that is used to perform qualitative and quantitative analysis. Additionally, the software can also be used to control the experimental conditions, such as the temperature of the separation column, the flow rate of the carrier gas, and settings for the injector and detector. The output generated by the detectors is called a chromatogram, which contains useful information pertaining to the detected compounds as they pass through the separation column. The software provides the ability to perform calculations such as peak area, peak height, and retention time. The retention

time is primarily used for compound detection, and the calculations of peak area and peak height provide information pertaining to the concentration of a compound in the mixture. Figure 1.2 displays a sample detector output and the different kinds of information obtained from a chromatogram.

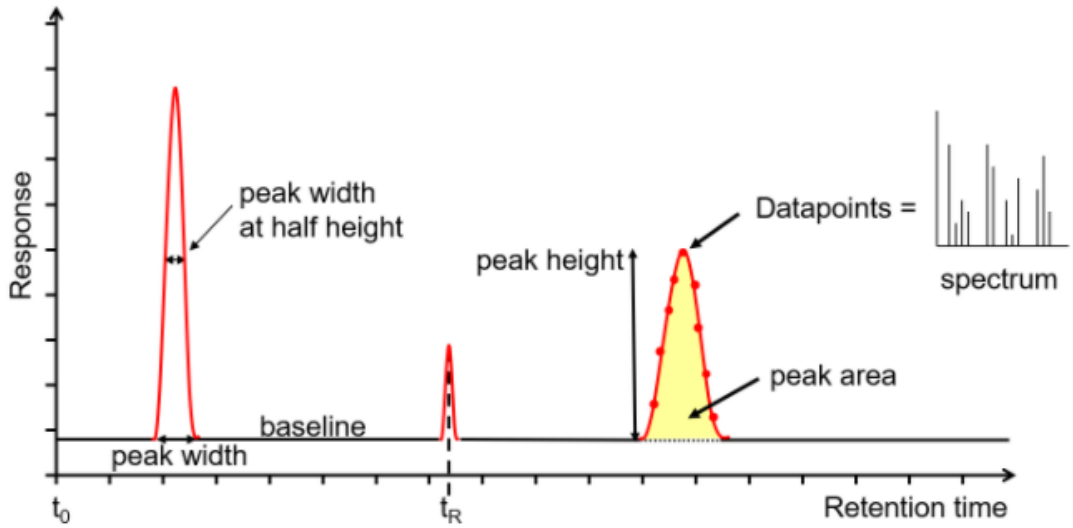


Figure 1.2: GC detector output

1.1.2 Micro-Gas Chromatography System (μGC System)

Conventional GC systems are accurate and provide high-precision results. However, such a system requires manual intervention, and the analysis process is often time-consuming. Moreover, these systems are bulky, power-hungry, and require supervision from trained technicians for their operation, confining them to laboratory environments.

μGC systems are miniaturized, field-portable, and energy-efficient versions of the conventional GC systems. They use miniaturized versions of components such as preconcentrator-injector [29], micro-scale separation columns [7], [14], micro-fabricated detectors [13], pumps, and valves, along with a data logging and analysis system. Their small form factor makes

them suitable for in-field applications requiring on-site and rapid on-demand analysis.

The μGC system can be used to periodically collect atmospheric samples and analyze them to generate relevant warnings and alerts against the presence of harmful and toxic gases in applications concerning environmental analysis, such as in transportation, mining, pharmaceutical, and petrochemical industries.

1.2 Motivation

The gas chromatography process involves passing the sample mixture through a single analytical column and analyzing the chromatogram. While μGC systems exhibit smaller sizes and increased portability, they suffer from deteriorated functional capacity for separation and analysis of compound mixtures due to factors like sample injection inefficiency and shorter column length arising due to miniaturization. It is difficult, tedious, and time-consuming to separate complex samples like diesel and gasoline, typically comprising more than a thousand compounds.

The selectivity and resolution of an analytical column enable the separation of compounds from a complex compound mixture. Using multiple analytical columns with different stationary phases effectively improves the compound separation capability of a μGC system [15], [14]. Additionally, in a complex compound mixture, two or more compounds can elute at similar times from the analytical column of a particular type of stationary phase. Such compounds are called coeluting compounds and a single analytical column is insufficient to separate these compounds [12]. With the use of a stationary phase that has different selectivity, these compounds could be separated distinctly. Thus, combining multiple separation columns and their outputs can aid in the identification of coeluting compounds. Additionally, the result of multiple separation columns that can separate a particular compound

present in the test mixture is a useful approach to assert the presence of hazardous compounds of interest. Detection in multiple stationary phases acts as a measure of confidence to assert the presence of such a compound within the test mixture.

The objective of this thesis is to develop a system that uses multiple separation columns to analyze a sample mixture, along with the development of an analysis mechanism for identifying compounds based on the results obtained from the multiple separation columns.

1.3 Contributions

The major contributions of this thesis are listed below:

- Implement a robust and scalable software design to enable the development of a portable μGC system.
- Preprocessing and conditioning the output chromatogram obtained from the μGC system by performing noise filtering, baseline correction, and subsequent peak identification as a part of the analysis system.
- Implemented a compound detection algorithm that can identify compounds from the output chromatograms of a multidimensional GC separation.

1.4 Organization of Thesis

This chapter reported the background of Gas Chromatography systems and the motivation for the developed system. The rest of this thesis is organized as follows:

Chapter 2 describes the detailed hardware architecture of the μGC system along with a brief description of all the hardware peripherals interfaced with the μGC system.

Chapter 3 reports the software design of the portable μGC system and the Graphical User Interface (GUI) used to interact with the μGC system. A detailed description of the software features and firmware of the μGC system is provided in this chapter.

Chapter 4 presents the integration of the hardware and software components to realize a functional portable μGC system with dual parallel separation columns. The chapter also discusses the results obtained from the setup, along with a thorough analysis of output chromatogram processing for peak detection and retention time calculation.

Chapter 5 presents a multidimensional analysis that utilizes information from the output chromatograms of multiple analytical columns to detect and identify compounds in a complex mixture.

Chapter 6 concludes the thesis work.

Chapter 2

Hardware Specification

This chapter discusses the hardware specifications of the components used to perform the gas chromatography experiments. A brief description of all the hardware peripherals used in the system is provided. This chapter also discusses the hardware specification of the modular μGC stack in detail at the end of the chapter.

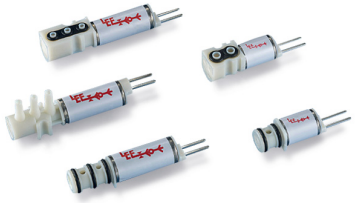
2.1 Hardware peripheral components

The equipment used to perform gas chromatography experiments comprises of the modular μGC system and a set of different peripherals actuated in various phases of an experiment. Figure 2.1 displays the entire set of hardware peripherals that can be interfaced with the system.

2.1.1 Valves

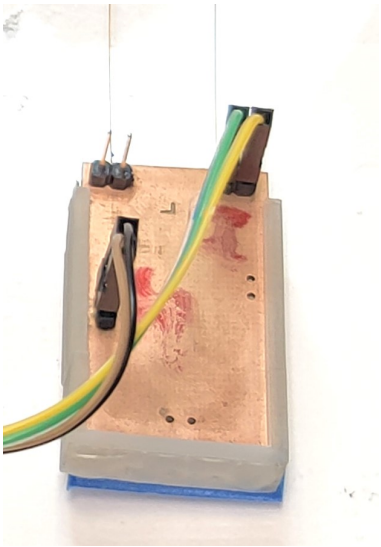
Gas flow control is needed to implement different GC experiment phases such as loading, injection, analysis, and cleaning, which is explained in detail in Section 4.2. This is facilitated through the use of multi-way valves.

Figure 2.1a shows a miniature 3-way latching solenoid valve (LHLA0521111H, The Lee Co.) that was selected due to its characteristic features of low fluidic resistance and low leakage

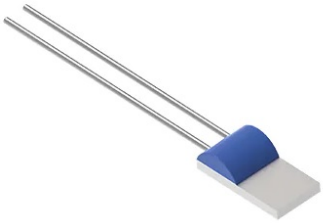


(a) Valve

(b) Pump



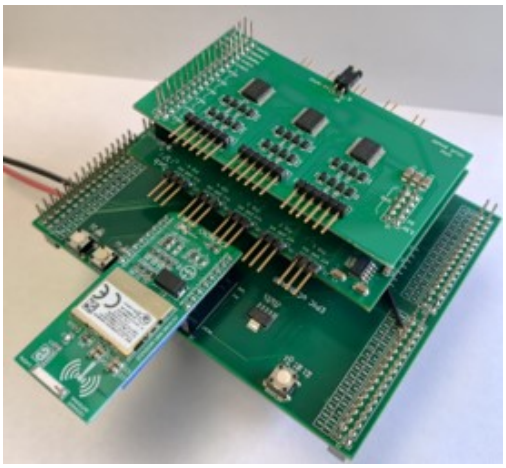
(c) Micro-heater



(d) RTD Pt1000

(e) DC Fan

(f) PID Detector



(g) Modular μGC stack

Figure 2.1: Peripherals used in the μGC System

rates. The valve can be operated between two states by applying a 5V DC pulse, wherein the polarity of the pulse controls the valve's position.

2.1.2 Pumps

Pumps are needed in μGC systems to control the flow of carrier gas and to maintain a constant pressure within the system. It enables the sample to move through the analytical column for the chromatographic process to occur. Additionally, pumps can provide flow rate control, which directly influences the resolution and efficiency of the separation process. Figure 2.1b shows the 5V low-power DC pump that was chosen for the μGC system.

2.1.3 Heaters and Temperature Sensors

The μGC system uses micro-fabricated heaters during the analysis phase of the GC experiment, where the separation column is heated with defined temperature programming profiles to enable fast and reliable compound separation.

Resistive temperature detectors (RTDs), such as Pt1000 [22], shown in Figure 2.1d, are used to sense the temperature of the heater device and provide feedback to the microcontroller to enable temperature control in a closed loop. Figure 2.1c shows an example of a micro-heater housed in a 3D-printed manifold along with a Pt1000 RTD embedded in it. This device is used to heat a micro-fabricated separation column during the experiment process.

2.1.4 Fans

Typical GC experiments involve different types of temperature control varying from isothermal profiles to having defined temperature programming profiles. The temperature of the

system is varied across a huge range for the desired separation of the compounds. Fans are utilized to expedite the cool-down process and bring the heaters back to room temperature. Figure 2.1e displays the 5V DC axial fans (F251R-05LLC, Digikey) that were selected due to their small form factor and low power consumption.

2.1.5 Detectors

The sensors used for the detection of effluents have a significant impact on the output chromatogram obtained from the system. Typically, a sensor with high sensitivity and voltage output is required to detect minuscule traces of effluent coming out of the separation column.

The μGC system can interface with Photoionization Detectors (PIDs) for the detection of compounds, as shown in Figure 2.1f. This detector requires a voltage supply of 5V and has an output voltage range of 0-2.5V. The small size and good Signal-to-Noise ratio (SNR) make it a good choice to be used in the system.

2.1.6 Modular μGC stack

The modular μGC stack is an electronic stack comprising 4 functional PCB modules. Figure 2.1g shows an image of the modular GC stack used for experimentation in this thesis, and Figure 2.2 shows a diagrammatic representation of the functional parts of this stack [20]. The function of each of the 4 modules is listed as follows:

- **Power Board:** This board accepts unregulated voltage in the range 18V-36V and converts it to regulated 3.3V, 5V, and 7.5V, which are supplied to different parts of the system. Additionally, this board also houses the heater control circuitry.

- **Microcontroller Board:** This board contains the microcontroller (Texas Instrument's TM4C129DNCPDT) and is responsible for issuing control signals for peripheral actuation and reading sensor data. This board also houses the Wi-Fi communication module and operates on 3.3V.

- **Sensor Board:** The sensor board contains Analog-to-digital converters (ADCs), which are used to interface the sensors and detectors with the microcontroller board. The ADCs (Analog Devices AD7793) communicate with the microcontroller over a Serial Peripheral Interface (SPI) channel.

- **Valve Control Board:** The valve control board uses motor driver circuitry (TB6612FNG Dual Motor Driver IC), which accepts digital input from the microcontroller board to actuate the valves, pumps, and fans interfaced in the system.

Table 2.1 summarises the specifications of the modular μGC stack.

Table 2.1: Modular μGC Stack Specifications

Component	**Description**
Input Voltage	18V - 36V
Output Voltages	3.3V, 5V, 7.5V, 18V
Resistance Temperature Detectors	5
Photoionization Detectors	5
Heaters	4
Valves / Pumps / Fans	6

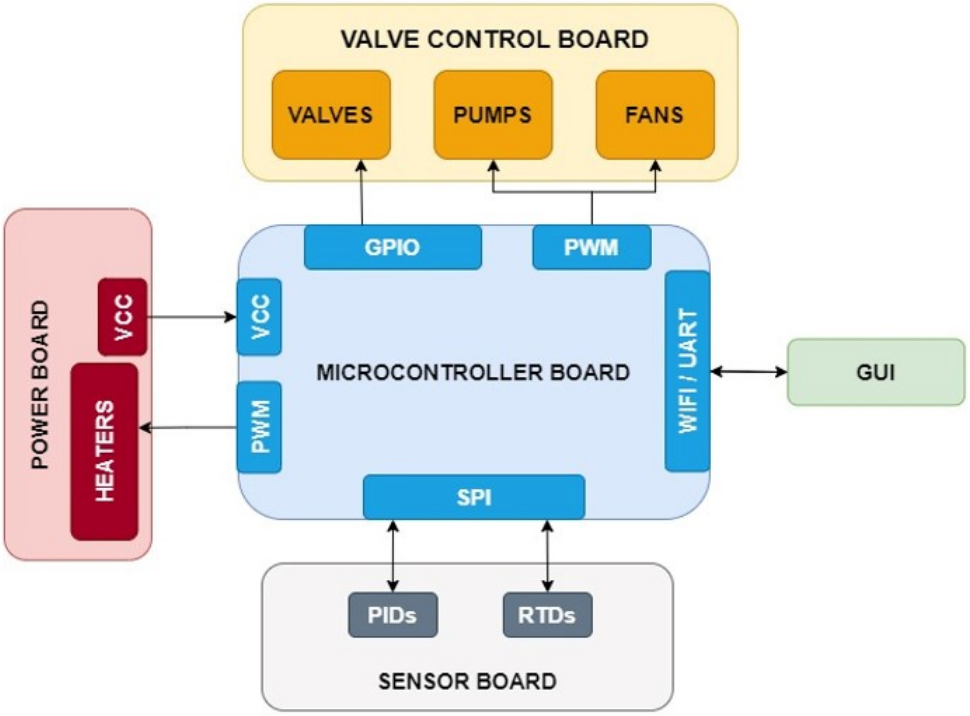


Figure 2.2: Modular μGC stack hardware architecture

Chapter 3

Software Specification

This chapter describes the software architecture implemented in the μGC system. The simple software design allows any user to develop various GC configurations as long as it is compatible with the GC hardware setup. Figure 3.1 shows an overview of the software components used in the portable μGC system. The first section of this chapter describes the firmware implemented in the embedded system and showcases some of the capabilities and features that enable a portable μGC system. The second section of this chapter describes the Graphical User Interface (GUI) tool used to interact with the modular μGC system.

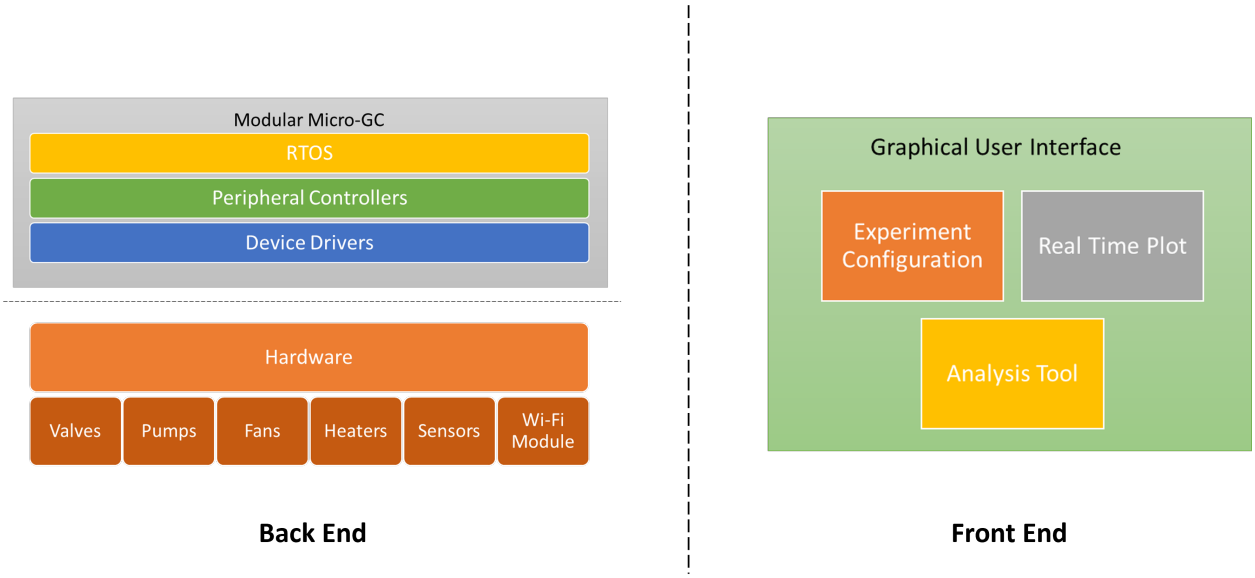


Figure 3.1: Software components of portable μGC system

3.1 Embedded Software Design and Features

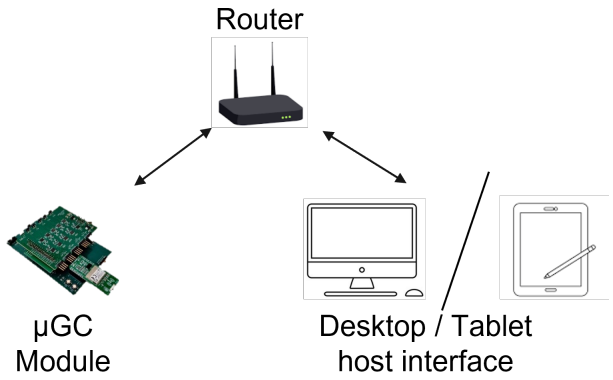
The firmware implementation of the embedded system is described in detail in this section. A μGC system can be assembled with various different configurations involving the usage of distinct hardware peripherals, some of which were described in Chapter 2. However, a fundamental sense of control of the peripherals is needed for the desirable execution of the μGC system. With most of the peripherals requiring electronic control, the μGC system forms an embedded system in which the micro-controller governs the operation of the system. Typically, for small-scale systems, the software running on the microcontroller is a bare-metal program that is dedicated to running a single application. However, μGC systems can comprise multiple peripherals performing different kinds of actuation simultaneously, thereby increasing the complexity of controls required for the correct execution of the system. Real-Time Operating System (RTOS) is an application that allows multiple tasks to run concurrently, which is more suited for μGC systems. As a part of this thesis, the embedded software for the μGC system was developed using Texas Instruments' TI-RTOS application [17], [18]. TI-RTOS provides different threading, timing, and synchronization modules that were utilized for controlling and actuating the hardware peripherals interfaced with the embedded system. One of the primary objectives of this thesis is to develop a portable μGC system that can be deployed for on-field analysis. In this section, we discuss the embedded software features that enable portability while maintaining the high performance of the μGC system.

3.1.1 Wireless Communication Mode

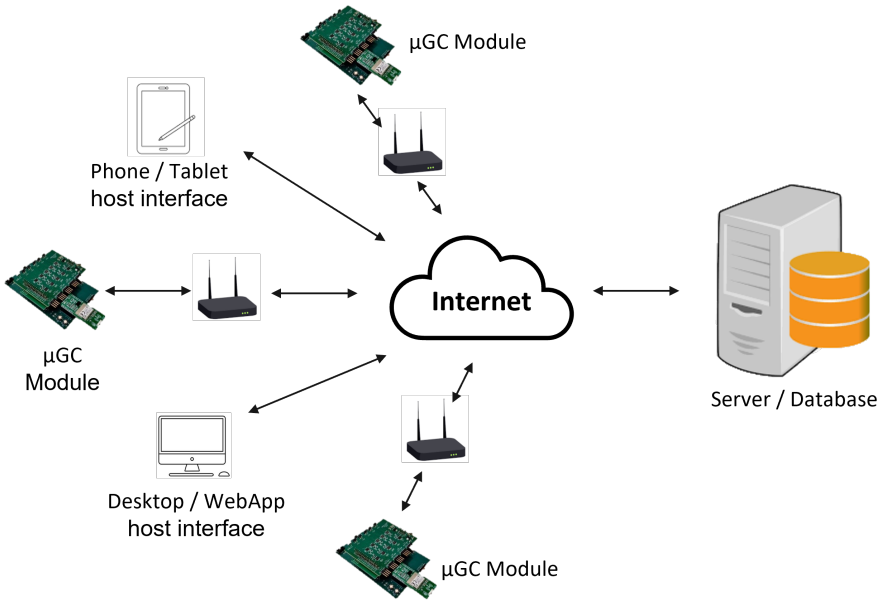
The implementation of Wireless Fidelity (Wi-Fi) communication mode for the transmission and receiving of data was one of the remarkable software features for the μGC system.

The CC3100 Click Wi-Fi Module was integrated with the microcontroller to provide Wi-Fi connectivity to the modular μGC system. This feature is highly scalable with the infrastructure available at the location where the μGC system is deployed. Figure 3.2 showcases two different network topologies enabled by this implementation:

- Figure 3.2a showcases the current network topology that can be established with wireless communication mode implemented on the modular μGC system. The host interface running on a desktop or tablet can interact with the μGC system by connecting to it within a Wireless Local Area Network (WLAN).
- Figure 3.2b showcases a possible network topology that can be useful in applications requiring high scalability. The modular μGC can connect to a server using HTTP over the internet. The end devices such as desktops, smartphones, and tablets can connect to the server over the internet and interact with the embedded systems. Such a network topology can be deployed for various environmental analysis applications.



(a) Wireless Local Area Network



(b) Hybrid Network of interconnected μGC systems

Figure 3.2: Network configurations for μGC systems enabled through Wi-Fi

3.1.2 Sensor Read Frequency

The capability to sample sensor output at high frequency enables the detection of minuscule compounds as they elute from the separation column. The sensor board used in the modular μGC system houses the AD7793 Analog-to-Digital Converters (ADCs) to interface the RTD and PID sensors with the microcontroller. The software configures ADCs in continuous conversion mode (CCM) at an update rate of 123Hz, wherein the ADC samples the sensor values and stores it in its data register. The microcontroller reads this value from the ADC's data register at a frequency of 50Hz (20ms), therefore yielding an effective sensor sampling frequency of 50Hz.

Ring Buffer

As described previously in Section 3.1.2, we were able to achieve a sensor read frequency of 50Hz. However, sending data from the embedded system to the GUI over Wi-Fi at this frequency can cause an overload on the system, resulting in unexpected behavior. In order to resolve this mismatch between the sensor read frequency and the Wi-Fi send data frequency, the sensor data is stored in a ring buffer before transmitting to the GUI.

As shown in Figure 3.3, each detector ADC is associated with a ring buffer, wherein the data is stored in a First-In-First-Out (FIFO) manner, and the least recently used data is overwritten in case the buffer is full. With such an implementation, we ensure all the sensor information is captured and transmitted reliably without overwhelming the system.

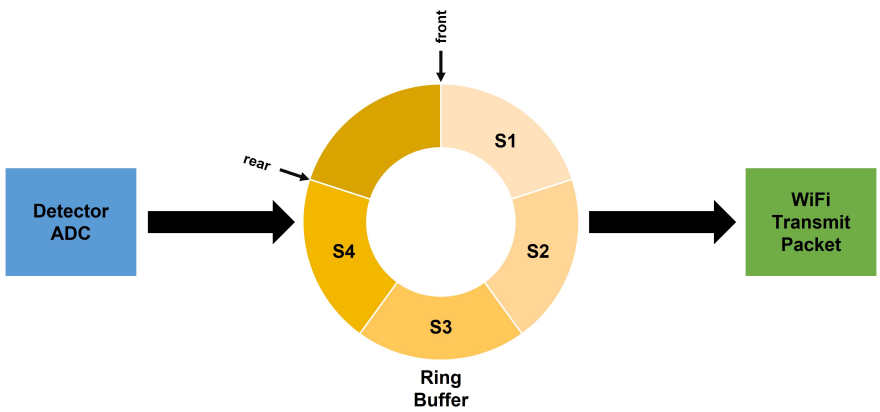


Figure 3.3: Ring Buffer Implementation

The ring buffer capacity can be calculated as shown in Equation 3.1. Therefore, for a desired sensor frequency of 50Hz and a Wi-Fi transmit frequency of 10Hz, the ring buffer capacity would be 5 sensor readings.

$$\text{Ring buffer capacity} = f_{sensor}/f_{transmit} \tag{3.1}$$

Fixed Packet Structure

The data communication between the modular μGC system and the host interface is an important design aspect since the GC experiment involves essential detector outputs that need to be recorded correctly for any form of analysis.

A fixed packet structure was developed to encompass the information about the current state of the system (experiment status and event number), RTD sensor information corresponds to four heaters, and detector information corresponds to five sensors. This information is obtained from the respective ring buffers. Figure 3.4 shows the format for the fixed packet structure.

The definition of a single fixed packet structure streamlined the process of encoding and decoding information that is exchanged between the μGC system and the host interface. The unification of all the sensor information into one packet allowed for simple data transmission using a single data transmit task scheduled at a period of 100ms. The transmitted packet is then decoded at the host interface for data recording/visualization.

Experiment Status (1 byte)	Event ID (2 byte)	Pad (1 byte)
Heater 1 Time Stamp		
Heater 1 Data		
...		
Heater 4 Time Stamp		
Heater 4 Data		
Detector 1 [0] Time Stamp		
Detector 1 [0] Data		
...		
Detector 1 [4] Time Stamp		
Detector 1 [4] Data		
...		
Detector 5 [0] Time Stamp		
Detector 5 [0] Data		
...		
Detector 5 [4] Time Stamp		
Detector 5 [4] Data		

Figure 3.4: Fixed packet structure for the modular μGC system

3.1.3 Parallel Heater Temperature Programming

The modular μGC system can be connected with a total of four heaters simultaneously at full capacity. Thus, the requirement for actuating multiple heaters in parallel with the capability of varied temperature programming profiles was necessary. In this section, the implementation of software PWM and Finite State Machine (FSM) along with various RTOS

resources utilized to realize the requirement for the actuation of multiple heaters is described in detail.

Software PWM

The hardware connections were designed such that two heater modules are actuated through PWM outputs, while the other two heater modules are actuated through GPIO outputs from the microcontroller. The heaters actuated using GPIOs effectively function as operated using ON-OFF control despite the presence of a PID control mechanism, thus yielding poor performance.

Software PWM was implemented in the system to provide PID-based temperature control for all four heaters in the system. The timer module from TI-SYS/BIOS was used to provide the timing requirement. The appropriate duty cycle is configured using the ON/OFF time by loading the corresponding timer period values. The state of the associated GPIO is subsequently switched each time the timer expires.

Heater Finite State Machine

A Finite State machine (FSM) was developed for the individual heaters to support parallel temperature programming capability in the μGC system, as shown in Figure 3.5. The different states of a heater are described as follows:

- **OFF:** This is the initial state for all heaters. In this state, there is no actuation of the heaters, and the PID control algorithm is effectively inactive. The only transition is to the state STEP, which occurs when a new user command is provided along with the set point and required ramp rate information.

- **STEP:** When a heater transitions to this state, the PID control algorithm is activated. The input set point and ramp rate information are used to perform the heater temperature ramping. When the temperature set point is reached, the state automatically transitions to MAINTAIN. The heater can transition to any of the other states on explicit user commands.

- **MAINTAIN:** In this state, the PID control algorithm works on maintaining the latest provided temperature set point. An explicit user command can trigger a transition to either OFF or STEP. In case of transitioning to STEP, the user provides the new setpoint and required ramp rate information.

- **BYPASS:** A transition to this state is typically required when multiple heaters are being actuated in parallel. This state is generally used when the current heater needs to continue its latest operation while another heater requires a change in its FSM state. In case the latest operation was a STEP, and the desired temperature set point has been reached, an automatic transition to MAINTAIN occurs. Otherwise, only explicit user commands can initiate transitions to STEP, MAINTAIN, or OFF.

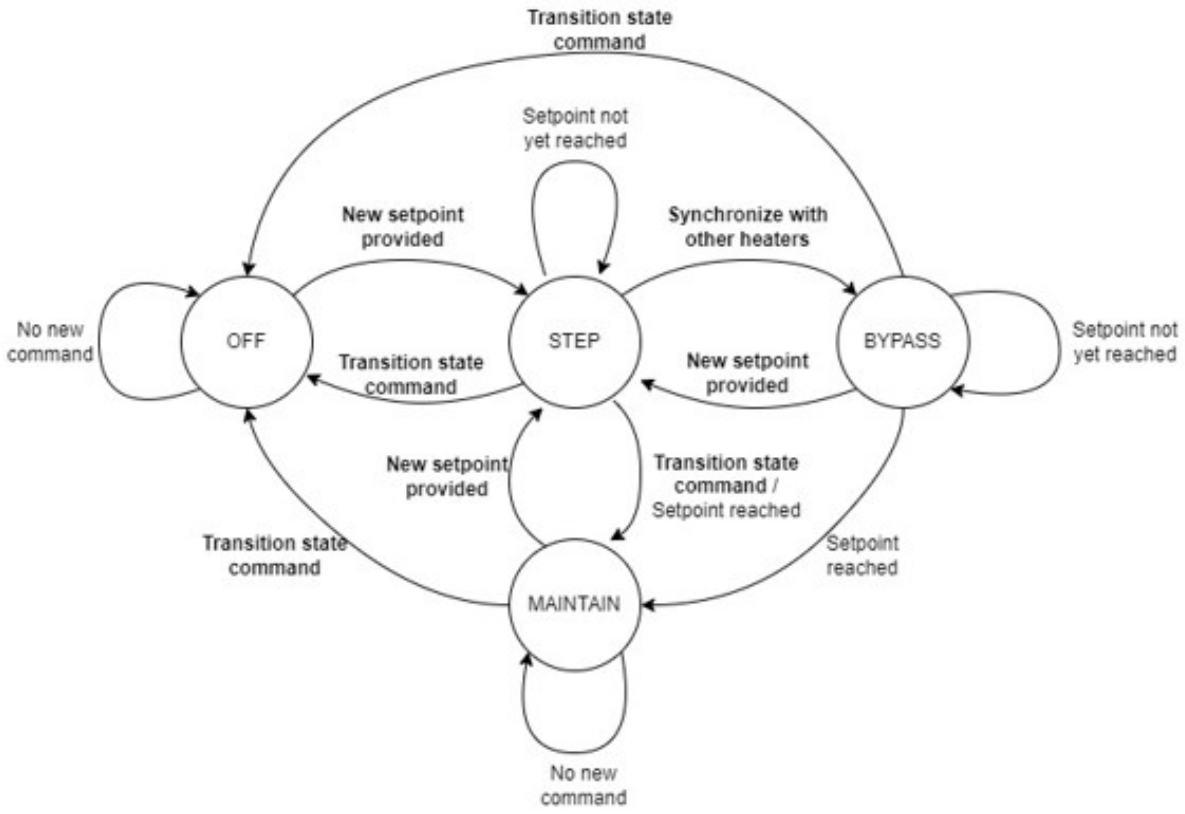
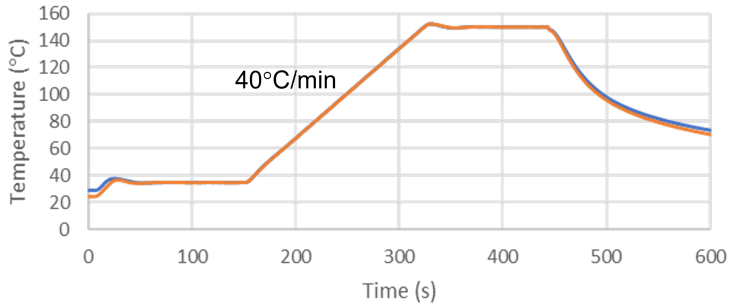


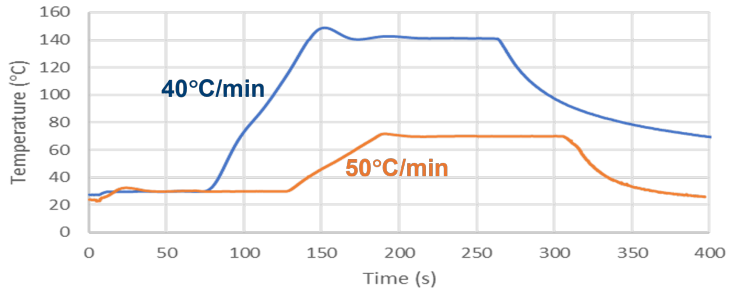
Figure 3.5: Heater Finite State Machine

Parallel Heater Profiles

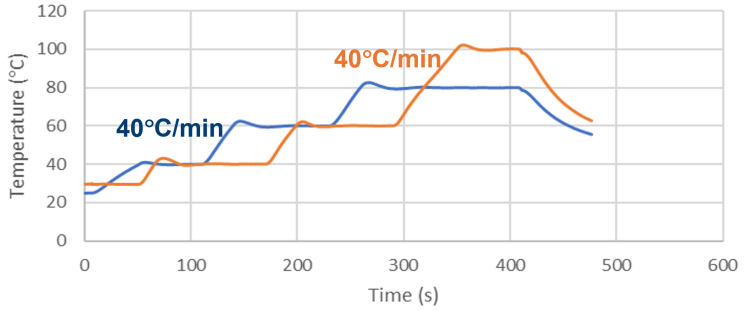
Different heater profiles were tested out by connecting multiple heaters to the μGC system. Figure 3.6 showcases the different heater profiles achieved for simultaneous actuation of two or more heaters. Through the implementation of this feature, varied temperature programming profiles for multiple heaters such as temperature ramping of two heaters simultaneously with similar ramp-rate, temperature ramping of two heaters with different ramp-rate and different actuation instances, and temperature ramping of two heaters with defined ramp-rate and intermediate setpoints were achieved.



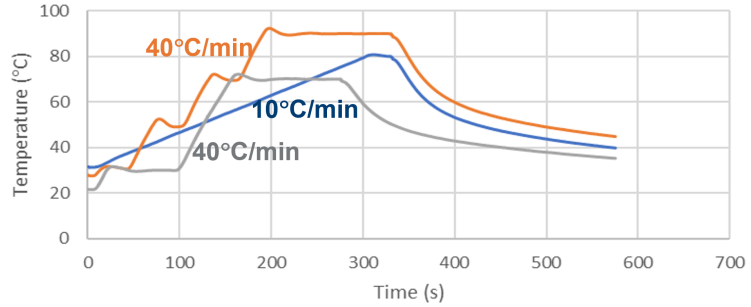
(a) Ramping two heaters simultaneously



(b) Ramping two heaters at different instances



(c) Ramping two heaters with intermediate setpoints



(d) Complex temperature programming profile

Figure 3.6: Temperature programming profiles using multiple heaters in parallel

3.1.4 Scheduling Algorithm

The design of the TI-RTOS application in the μGC system is based on the idea of having a centralized scheduler that is responsible for dynamically allocating relevant RTOS resources to peripheral objects based on the requirements of the experiment configuration. The scheduler is also responsible for delegating control action to different peripheral instances based on the completion of an event.

The μGC system uses a scheduling algorithm that provides precise timing and decision-making capabilities with the performance of the different aspects of the modular μGC system as its primary focus. Figure 3.7 shows the design of the scheduling algorithm, which incorporates a scheduler task responsible for scheduling the required tasks at a particular experiment stage. A single experiment is broken into multiple stages, and the start of each experiment stage acts as a scheduling point. The decision to execute the appropriate tasks responsible for peripheral actuation occurs at each scheduling point.

The different tasks and handlers that can be scheduled are listed below:

- **Scheduler Task:** The scheduler task is responsible for determining and delegating the scheduling of the appropriate tasks across different stages of an experiment. When the μGC system is powered on, it schedules the Experiment Fetch Task in anticipation of a connection to a host interface to fetch the experiment configuration information. When an experiment is in progress, it determines the appropriate tasks to be scheduled and dynamically allocates resources that are required for that particular experiment phase. Once an experiment completes, the scheduler dynamically frees all the RTOS resources and re-initializes the run-time variables to prepare the μGC system for a new experiment.

- **Experiment Fetch Task:** This task is executed when the μGC system is ready to

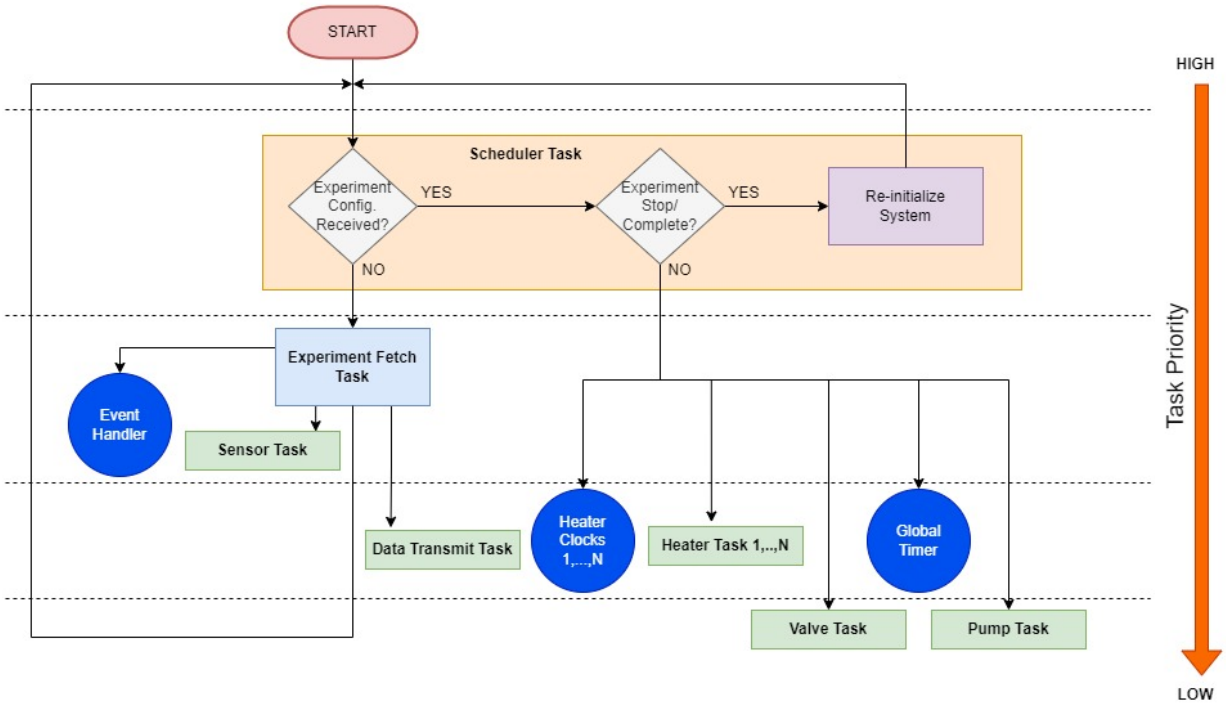


Figure 3.7: RTOS Scheduling algorithm flowchart

begin a new experiment. The μGC system fetches the experiment configuration as and when the host interface begins to transmit it over a Wi-Fi channel. After processing the experiment configuration, the task dynamically creates and initializes the resources required to perform the experiment.

- **Sensor Task:** Once the experiment is initiated, this task is scheduled for continuously sampling the heater and detector sensors connected to the μGC system. As described in Section 3.1.2, the detector sensor values are stored in a ring buffer in order to reduce the loss of any essential information.

- **Data Transmit Task:** This task is instantiated once the experiment begins and is responsible for transmitting the experiment information to the host interface. As described in Section 3.1.2, the experiment information is encoded within the fixed packet structure and transmitted over Wi-Fi communication mode.

- **Event Handler:** The event handler keeps track of the transitions along different stages of an experiment. The tracked changes are utilized by the scheduler task to determine the appropriate RTOS resources to be initialized for peripheral actuation.

- **Global Timer Handler:** The global timer handler is used to specify the duration of each experiment stage as defined in the experiment configuration.

- **Heater Tasks:** The heater task implements an FSM as described in Subsection 3.1.3 and a PID control logic to operate the heater connected to the μGC system. In the case of multiple heaters being connected to the μGC system, the scheduler creates multiple instances of the heater task to control the respective heaters. This task is typically scheduled during the separation phase of an experiment.

- **Valve Task:** This task is scheduled when the valves connected to the μGC system require actuation.

- **Pump Task:** This task is scheduled when the pumps or fans connected to the μGC system require actuation.

3.2 Graphical User Interface

The Graphical User Interface (GUI) acts as an interface for the user to interact with the μGC system. It has been developed using the PyQt5 library in Python3 [3]. The developed GUI consists of three tabs each serving a different purpose which are described in the subsequent sections.

3.2.1 Experiment Configuration Tab

This tab in the GUI allows the user to interact with the μGC system by specifying the experiment configuration and monitoring the status of the experiment.

The experiment configuration can be specified by providing the number of peripherals (valves, pumps, heaters, and detectors) used in the experiment, defining the experiment stage table, and also providing the temperature programming configuration for heaters. The experiment stages table is developed by splitting the experiment into distinct stages where each stage corresponds to the actuation of different peripherals.

The experiment configuration tab in the GUI also allows the user to save and load experiment configurations, change heater tuning parameters for heater control, and a sequence mode to run a series of back-to-back experiments with defined intervals.

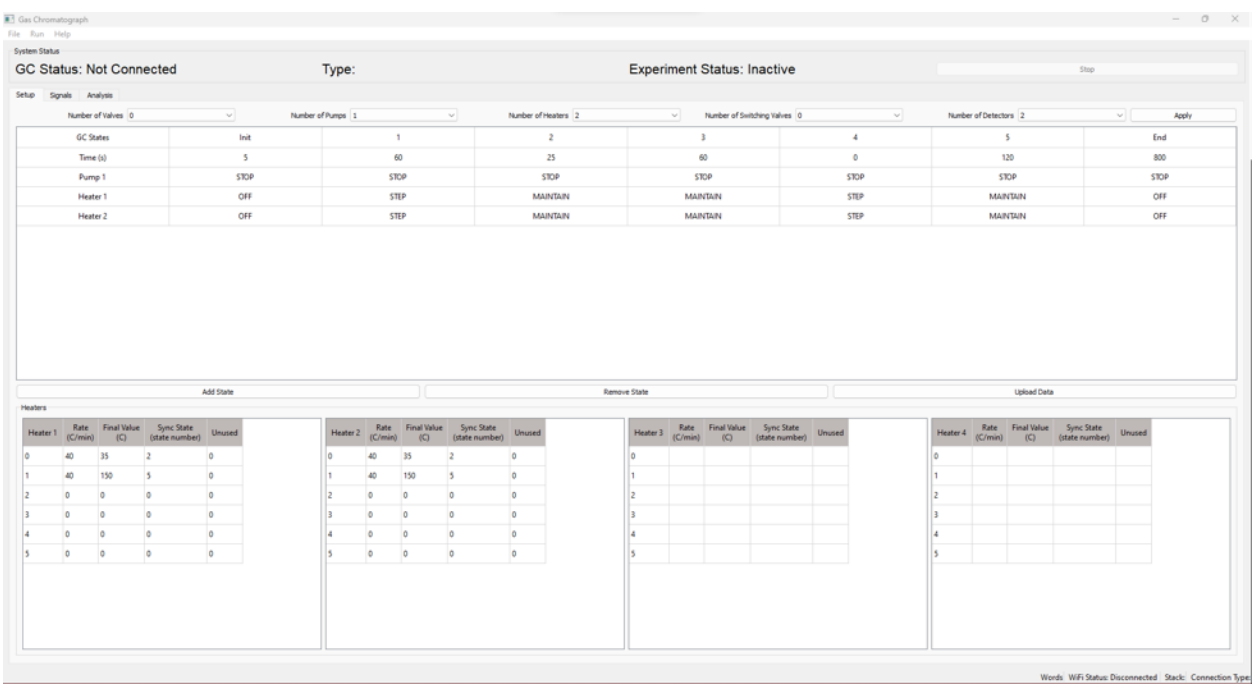


Figure 3.8: GUI Experiment Configuration Tab

3.2.2 Data Visualization Tab

The GUI also provides a tab to visualize the data received from the sensors in the μGC system, as shown in Figure 3.9.

The user can select the appropriate sensor or detector option by means of selecting one of the radio buttons provided on the left-hand pane of this tab. Based on the selection of the sensor, the real-time data is updated on the graph.

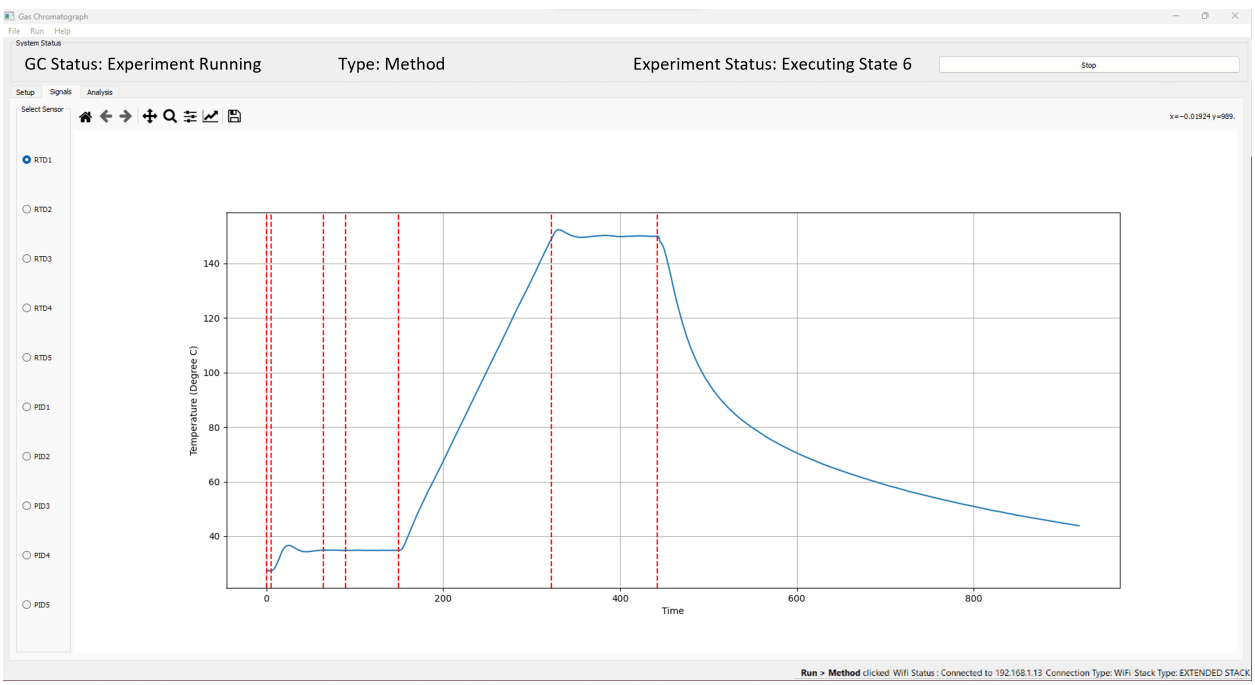


Figure 3.9: GUI Data Visualization Tab

3.2.3 Analysis Tab

The third tab in the GUI is the analysis tab, wherein the chromatograms obtained as a result of the experiments performed can be further analyzed. Figure 3.10 shows the analysis tab, which is used to perform the multidimensional compound identification process as described in Chapter 5. The output chromatograms of different stationary phases are obtained from

the input paths provided by the user, along with the retention time library information. When the ”START” button is pressed, the multidimensional compound identification algorithm is executed that processes the output chromatograms of the different stationary phases and displays the total list of identified and unidentified compounds. Additionally, the quantitative analysis parameters such as retention time of peaks, peak height, and peak width of the peaks obtained from the processed chromatogram are displayed on the right-hand side of the analysis tab.

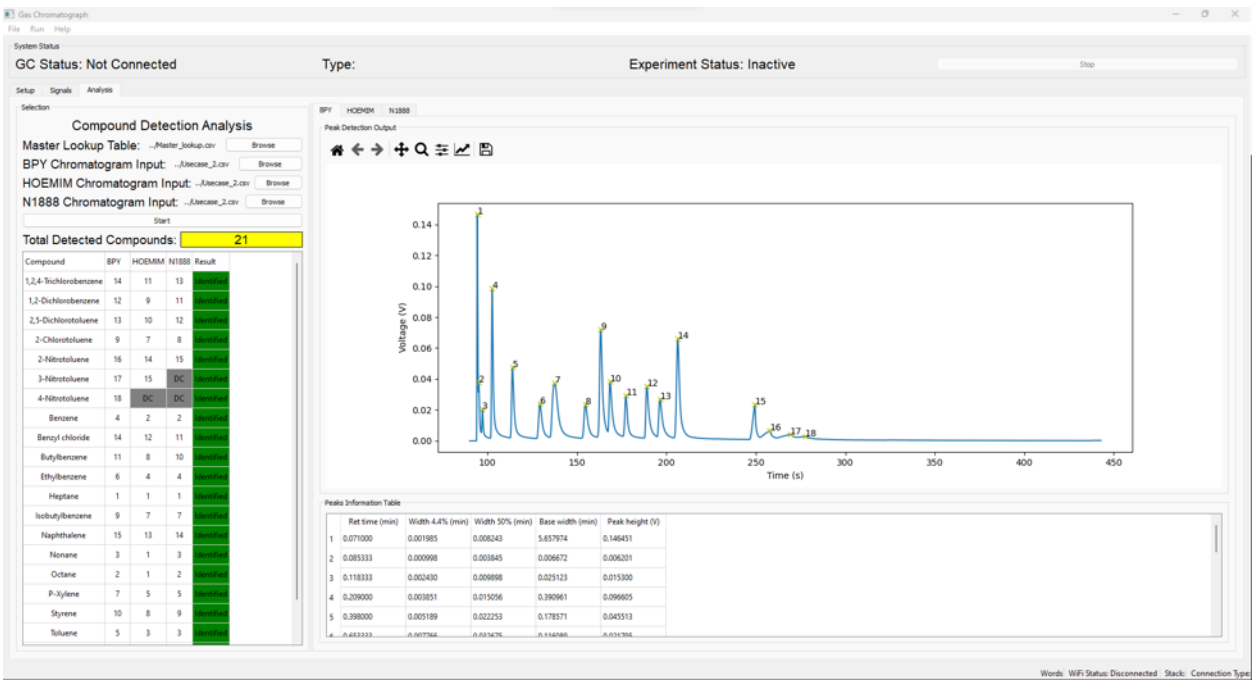


Figure 3.10: GUI Analysis Tab

Chapter 4

System Integration and Chromatogram Analysis

A prototype system was assembled using the modular μGC system and the hardware peripherals for running the GC experiments. This chapter reports the physical setup of the μGC system and describes the experimentation in detail. At the end of this chapter, we describe the steps of output chromatogram processing and analysis of different methods implemented in this process.

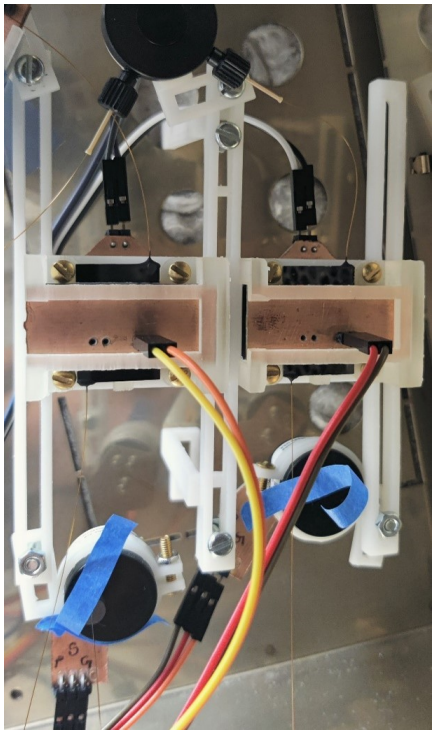
4.1 Experiment setup

The experiment setup for the μGC system uses a helium cylinder for carrier gas, the injector from a commercial GC system, a Y-connector, two micro-separation columns, two micro-heaters, two DC fans, and two PIDs. Figure 4.1 shows the image of the μGC system used for experimentation.

The sample is injected into the system by means of a commercial auto-injector. Helium is used as a carrier gas to transport the analyte through the system. A Y-connector is used to divide the flow of the analyte equally between the micro-separation columns.

The two micro-separation columns chosen are coated with distinct stationary phases but

(a) Setup showing conventional auto-injector, modular μGC stack, dual parallel micro-separation columns, and detectors



(b) Zoomed section showing Y-connector, parallel micro-separation columns, and photoionization detectors

Figure 4.1: Experiment setup

the same temperature programming profile is provided to them through the corresponding micro-heaters. The PIDs are placed at the end of each micro-separation column to detect the elution of compounds as they emerge from the columns. The μGC system uses Wi-Fi to interact with the host interface running on the desktop. The experiment configuration was developed on the desktop GUI, issued to the μGC system, and the experiment and sensor data were recorded in real-time. The μGC system also makes use of two DC fans, one for each micro-heater, to facilitate rapid cooling of the micro-heaters back to room temperature at the end of the experiment.

4.2 Experimentation

The experiment setup described in Section 4.1 was assembled together and the schematic diagram shown depicting the setup is shown in Figure 4.2. The experiment process performed using the setup can be divided into the following different phases:

1. **Initialization Phase:** The experiment process starts with the initialization of the different peripherals to an initial state. For example, the electronic μGC module is initialized, the micro-heaters are maintained at an initial temperature of 35°C, and the state of the DC fans is set to OFF state.

2. **Loading Phase:** During this phase, a portion of the sample mixture is collected into the auto-injector, while the μGC setup continues to maintain its initial configuration.

3. **Injection Phase:** The auto-injector injects the sample mixture into the μGC system. After entering the μGC system, the sample mixture is transported by the carrier gas, and the gets equally divided between the two micro-separation columns.

4. **Separation Phase:** During this phase, the micro-separation columns are heated to follow the temperature profile [8] as shown in Figure 3.6a. The columns are maintained at 35°C for one minute, followed by heating of the columns to a final temperature of 150°C at a defined ramp rate of 40°C/minute, and finally maintaining at 150°C for two minutes. The compounds gradually separate out as they travel through the micro-separation columns and get detected by the PIDs placed at the end of the columns.

5. **Cool-down Phase:** In this phase, the micro-heaters are turned off and the DC fans are set to ON state. This enables rapid cooling of the micro-separation columns back to room temperature to facilitate the start of a new experiment.

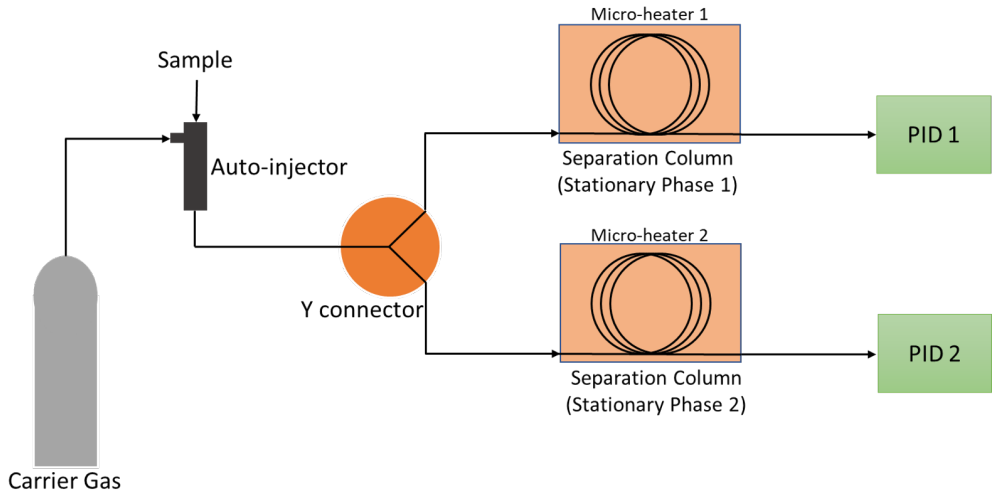
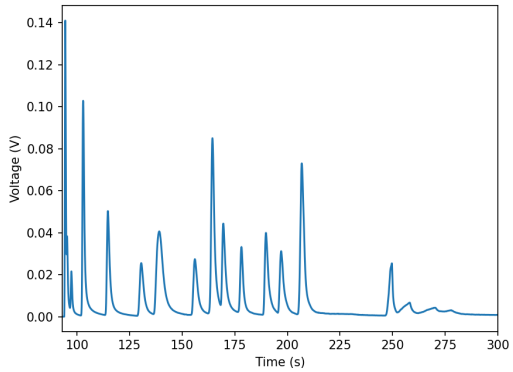


Figure 4.2: GC schematic diagram

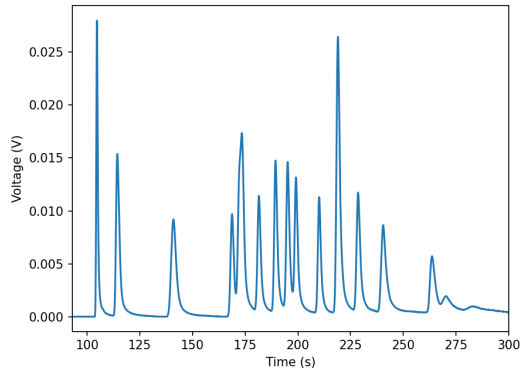
While the commercial auto-injector and the μGC system are two independent and disjoint systems, the experiment configuration was designed to ensure that the two systems worked in synchronization, therefore, avoiding any discrepancies during the injection phase.

Results

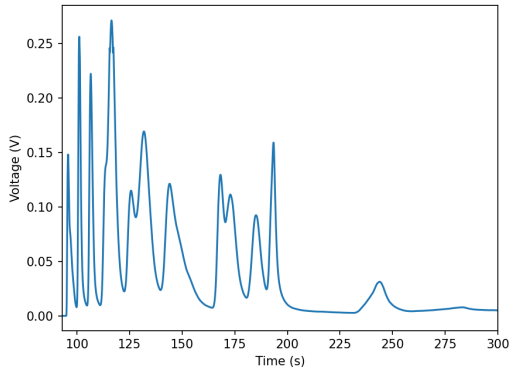
The experimentation process described in Section 4.2 was performed using three stationary phases, namely HOEMIM, N1888, and BPY, for two sets of sample mixtures whose composition has been described in Appendix A. The experiments were performed ten times with test mixture 1 and three times with test mixture 2 to gather data for all the stationary phases. Figure 4.3 shows the chromatograms obtained from the different stationary phases. Significant peaks being observed in the chromatograms show the outcome of the separation of compounds within the analytical column. The total time taken to separate the compounds was less than four minutes. These results exhibit that the μGC system can separate complex VOC mixtures with high efficiency and relatively low duration.



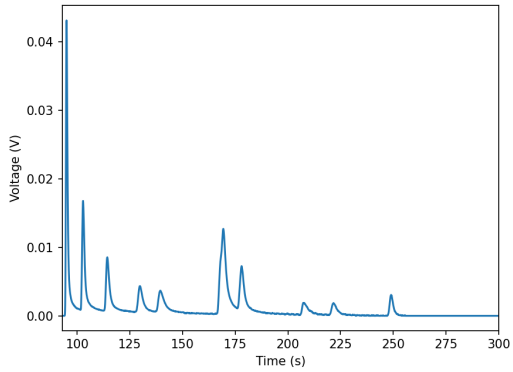
(a) Test mixture 1 chromatogram in BPY stationary phase



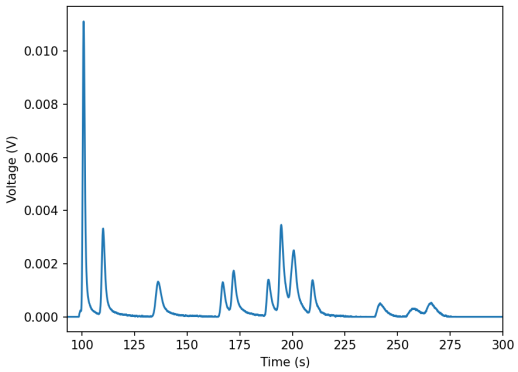
(b) Test mixture 1 chromatogram in N1888 stationary phase



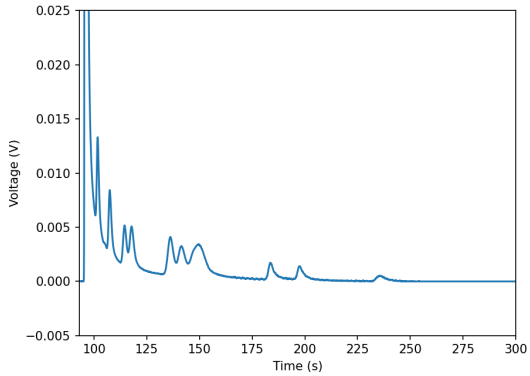
(c) Test mixture 1 chromatogram in HOEMIM stationary phase



(d) Test mixture 2 chromatogram in BPY stationary phase



(e) Test mixture 2 chromatogram in N1888 stationary phase



(f) Test mixture 2 chromatogram in HOEMIM stationary phase

Figure 4.3: GC system output chromatograms using complex compound mixtures

4.3 Chromatogram Analysis

This section describes the different steps involved in the processing of raw chromatograms to obtain useful information for qualitative and quantitative analysis. An analysis of different techniques in each step of chromatogram processing is reported.

4.3.1 Output Chromatogram Processing

Each output chromatogram obtained from a GC experiment run contains essential information about the compounds, such as retention time, concentration, volume, etc. However, determining the characteristics of the analytes in the mixture gets difficult due to the presence of environmental and system noise in the sensor data. Therefore, it is necessary to filter, smooth, and/or correct the baseline of the output chromatograms.

The retention time calculation is based on the time of injection of analytes. Thus, the output chromatogram is pre-processed to discard the initial sensor values before the point of injection of analytes by appropriately time-shifting the sensor data. The retention time can then be determined by extracting the peaks in the processed output chromatogram. The flow of output chromatogram processing is shown in Figure 4.4. The subsequent sections describe the different aspects of output chromatogram processing required to calculate the retention time of the compounds.

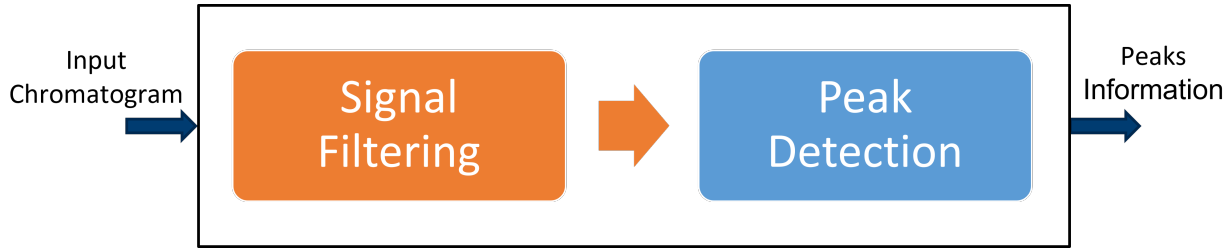


Figure 4.4: Output Chromatogram Processing Flow

4.3.2 Signal filtering

The sensor data obtained from the modular μGC system typically contains high-frequency noise. By observing the chromatogram signal in the frequency domain, it was observed that the useful signal information was present below the frequency 2.5Hz while frequencies above 2.5Hz predominantly contained noise. This knowledge was used to filter out the noise components and smoothen the signal. Figure 4.5 shows the input chromatogram signal in the frequency domain.

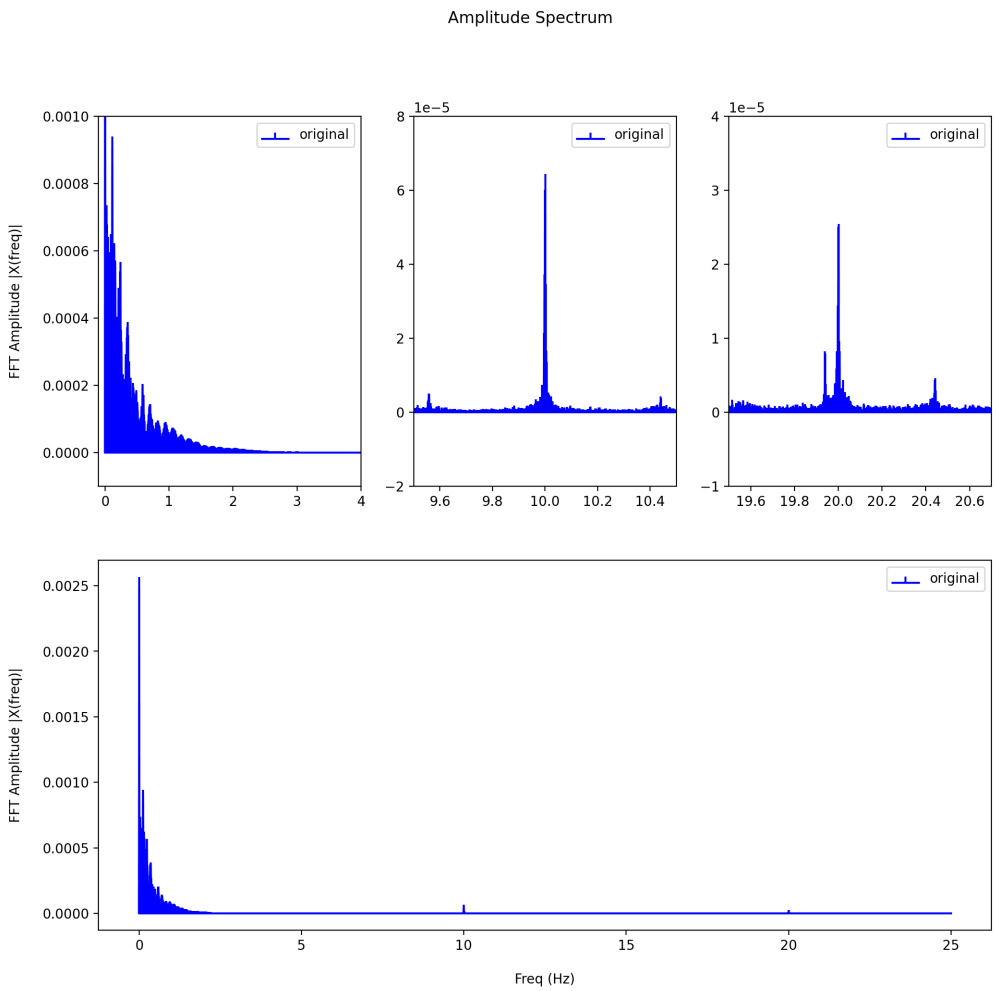


Figure 4.5: Single-Sided FFT of the input signal chromatogram

Different signal filtering techniques such as moving average filter, Lowpass Butterworth filter

[21], Gaussian filter [21], [27] and Savitzky-Golay filter [21] are suitable smoothing filters applied on the chromatographic outputs. A study of different filters and analysis of tuning parameters was performed to determine the optimal smoothing filter for the chromatogram data obtained from the experiments [28]. The suitability of the filters was evaluated using evaluation criteria such as the signal-to-noise ratio (SNR) of the input and output signals, the average shift in retention time, and the peak height caused by the filtering.

$$SNR = 20 \log_{10} \frac{|Signal|}{|Noise|} \tag{4.1}$$

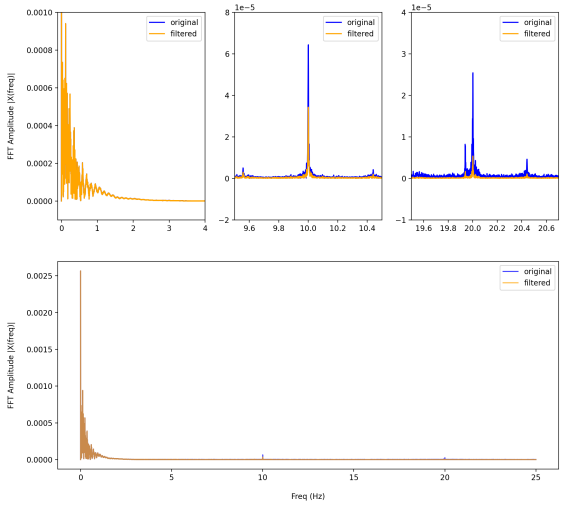
where, SNR = Signal to Noise Ratio

$|Signal|$ = Sum of magnitudes of FFT for frequency $\leq 2.5Hz$

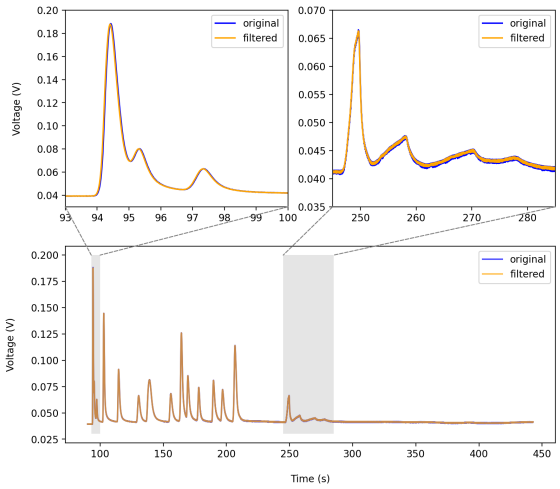
$|Noise|$ = Sum of magnitudes of FFT for frequency $> 2.5Hz$

Moving Average Filter

The moving average filter is a digital signal processing filter used to reduce random noise while retaining sharp step response [27]. The filtering is performed by taking the average of the values within a specified filter window size. The filter window size is a tunable parameter that impacts the intensity of the smoothing effect on the input signal. Figure 4.6, Figure 4.7, and Figure 4.8 shows the effect of varying the window size on the filtering of a sample chromatogram output of the μGC system.

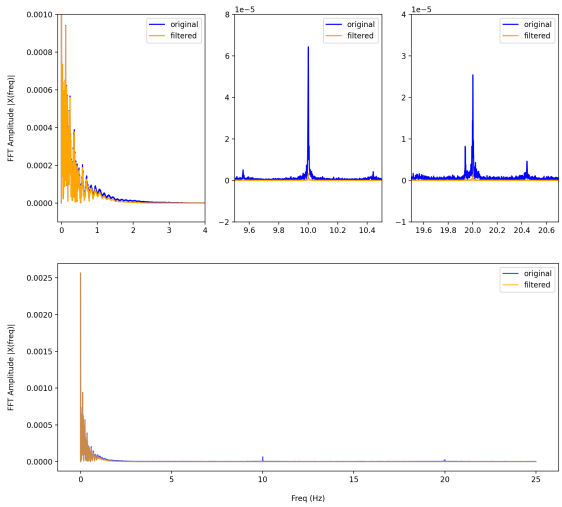


(a) Frequency domain

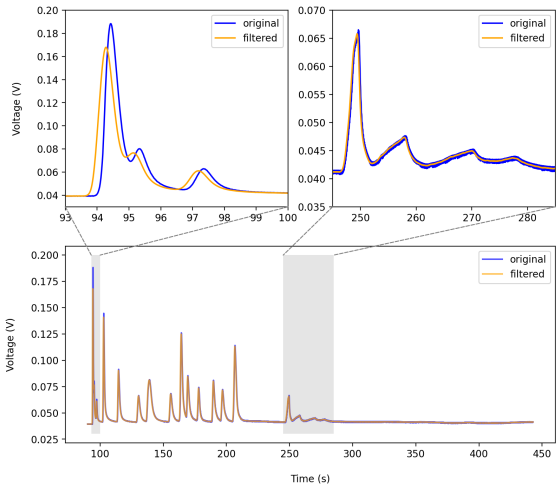


(b) Time domain

Figure 4.6: Moving average filtering of sample chromatogram with window of 3 datapoints

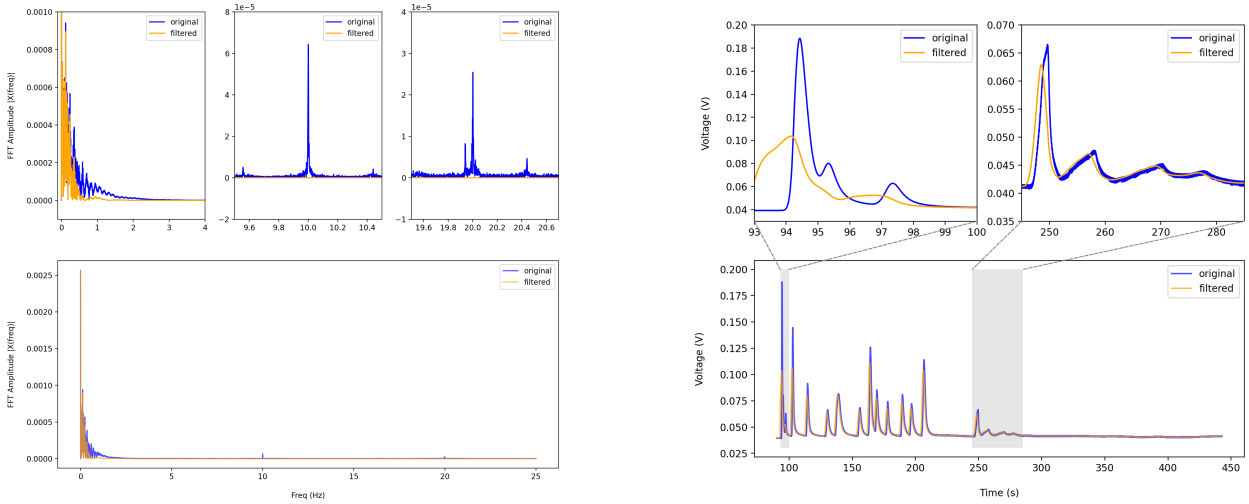


(a) Frequency domain



(b) Time domain

Figure 4.7: Moving average filtering of sample chromatogram with window of 19 datapoints



(a) Frequency domain

(b) Time domain

Figure 4.8: Moving average filtering of sample chromatogram with window of 75 datapoints

The filtering intensity increases as the filter window size is increased, thereby exhibiting an increase in the SNR of the filtered signal. The increase in filter window size also exhibits an increase in both average retention time shift and average peak height shift, which is an undesirable effect at higher filter window sizes. The smoothing is extremely aggressive at higher filter window sizes and there is a complete loss of peak shape and peak information. Table 4.1 summarizes the effect of varying the filter intensity on the different aspects like SNR, average retention time shift, and average peak height shift.

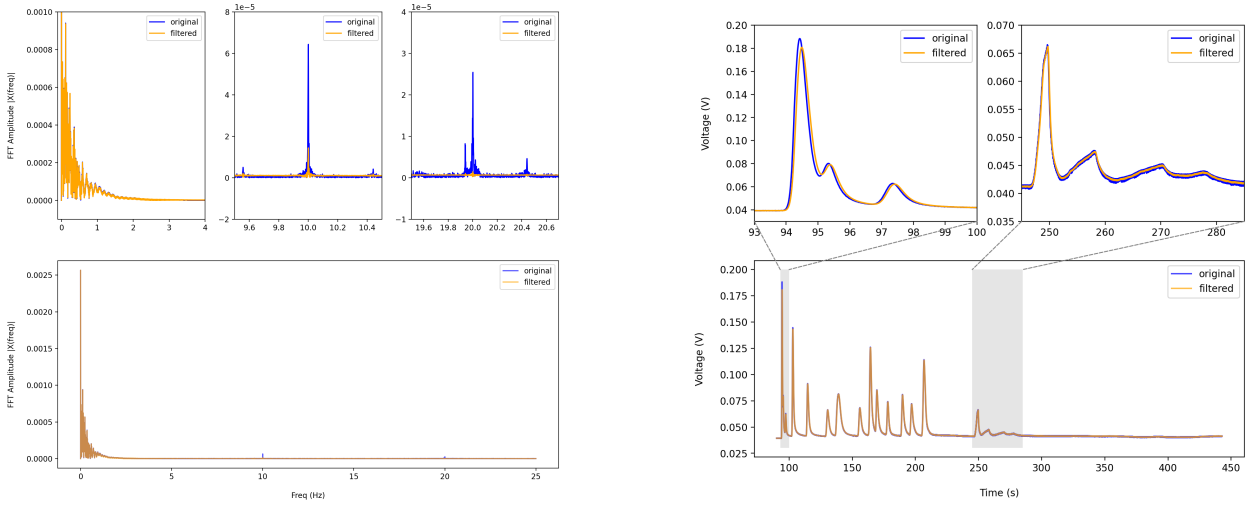
Table 4.1: Moving Average Filter

Window (data points)		3	5	9	19	49	75	99
Output SNR	(dB)	30.44	33.82	37.62	47.54	50.51	53.25	53.37
Average shift in retention time	(ms)	24.8	38	70	150	244.9	276	304
Average shift in peak height	(mV)	0.6	1.6	5.4	20.4	68.7	84.7	96.8

Lowpass Butterworth Filter

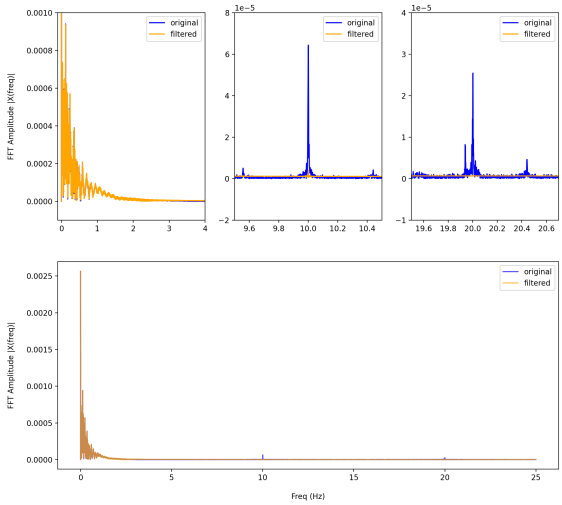
The Lowpass Butterworth filter is a classic infinite impulse response (IIR) filter that can be used to eliminate high-frequency noise from the input signal [10], [27], [2].

The parameters required to implement a Lowpass Butterworth filter are filter order and cutoff frequency. The order of the filter defines the efficiency of the filter output and the cutoff frequency defines the frequency of the noise to be eliminated from the signal. Figure 4.9, Figure 4.10, and Figure 4.11 shows the effect of varying the filter order on the filtering of a sample chromatogram output of the μGC system with cutoff frequency of 2.5Hz.

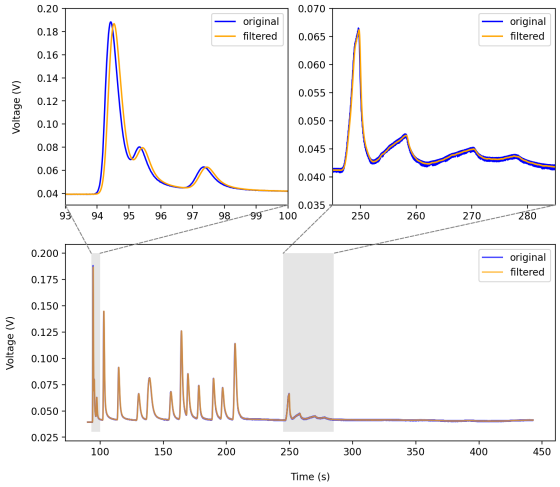


(a) Frequency domain (b) Time domain

Figure 4.9: Lowpass Butterworth filtering of sample chromatogram with cutoff frequency 2.5Hz and filter order 1

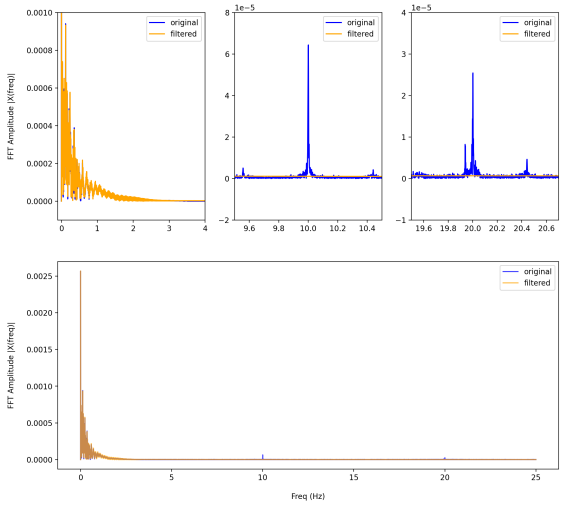


(a) Frequency domain

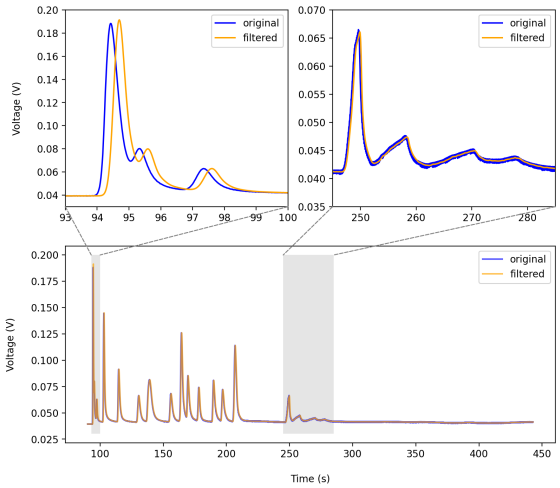


(b) Time domain

Figure 4.10: Lowpass Butterworth filtering of sample chromatogram with cutoff frequency 2.5Hz and filter order 2



(a) Frequency domain



(b) Time domain

Figure 4.11: Lowpass Butterworth filtering of sample chromatogram with cutoff frequency 2.5Hz and filter order 6

Increasing the filter order significantly increases the average shift in retention time, which is an undesirable effect on the filtered signal. Smoother filtering is achieved with a cutoff frequency chosen close to the signal frequency, wherein the high-frequency noise is eliminated

efficiently. Table 4.2 and Table 4.3 summarizes the effect of varying the filter order on the different aspects like SNR, average retention time shift, and average peak height shift for a cutoff frequency of 2Hz and 2.5Hz respectively.

Table 4.2: Lowpass Butterworth Filter (Cutoff Frequency = 2 Hz)

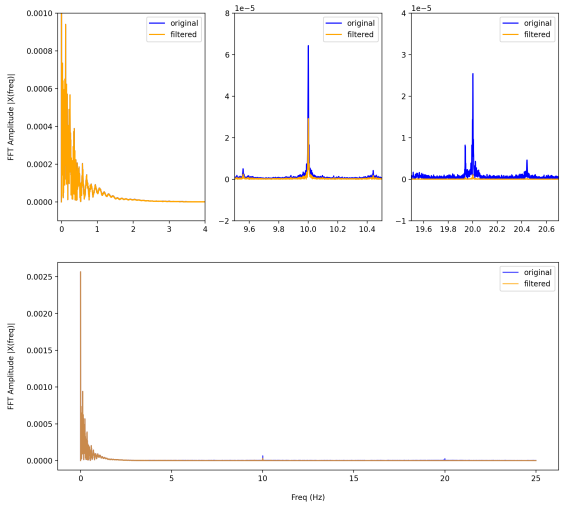
Filter Order		1	2	4	6	10
Output SNR	(dB)	44	37.9	41.3	45	42.9
Average shift in retention time	(ms)	79.7	130	246	350.9	573
Average shift in peak height	(mV)	11.1	4.15	0.4	0.9	0.9

Table 4.3: Lowpass Butterworth Filter (Cutoff Frequency = 2.5 Hz)

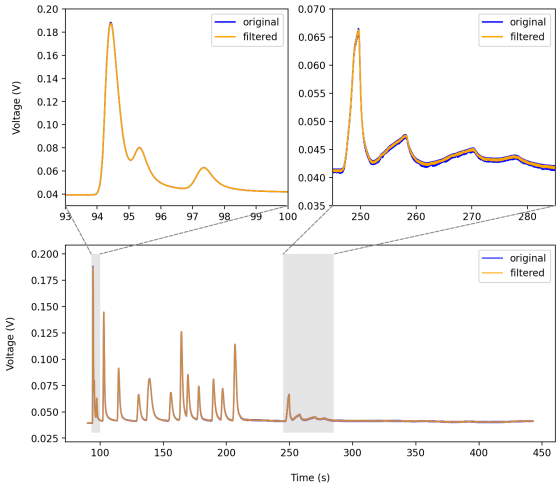
Filter Order		1	2	4	6	10
Output SNR	(dB)	44.3	37.4	37.2	45.6	40.2
Average shift in retention time	(ms)	71	110.5	190.9	271	432
Average shift in peak height	(mV)	7.6	1.6	1.7	3	4

Gaussian Filter

The Gaussian filter is another low-pass filter used for reducing high-frequency noise in the input signal [16], [1], [27]. The Gaussian filter performs a weighted average of sample values based on the Gaussian distribution. The standard deviation (σ) of the Gaussian filter kernel defines the smoothing intensity of the filter. Figure 4.12, Figure 4.13, and Figure 4.14 show the effect of varying the standard deviation on the filtering of a sample chromatogram output of the μGC system.

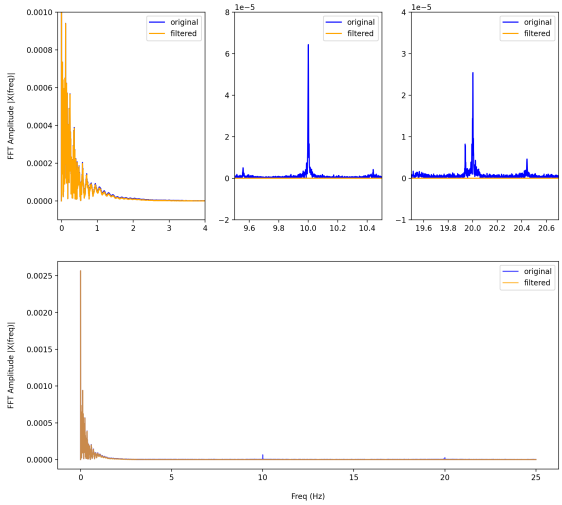


(a) Frequency domain

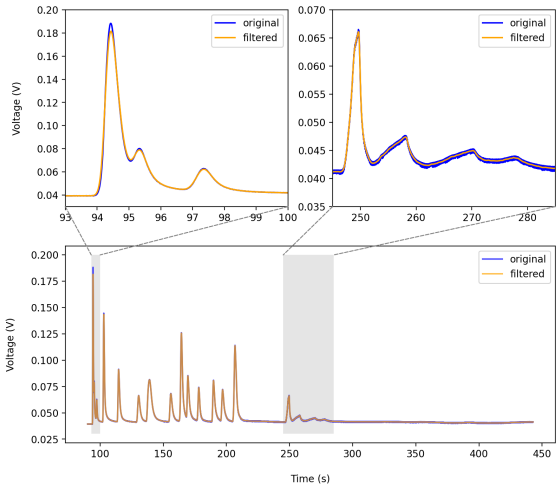


(b) Time domain

Figure 4.12: Gaussian filtering of sample chromatogram with standard deviation 1

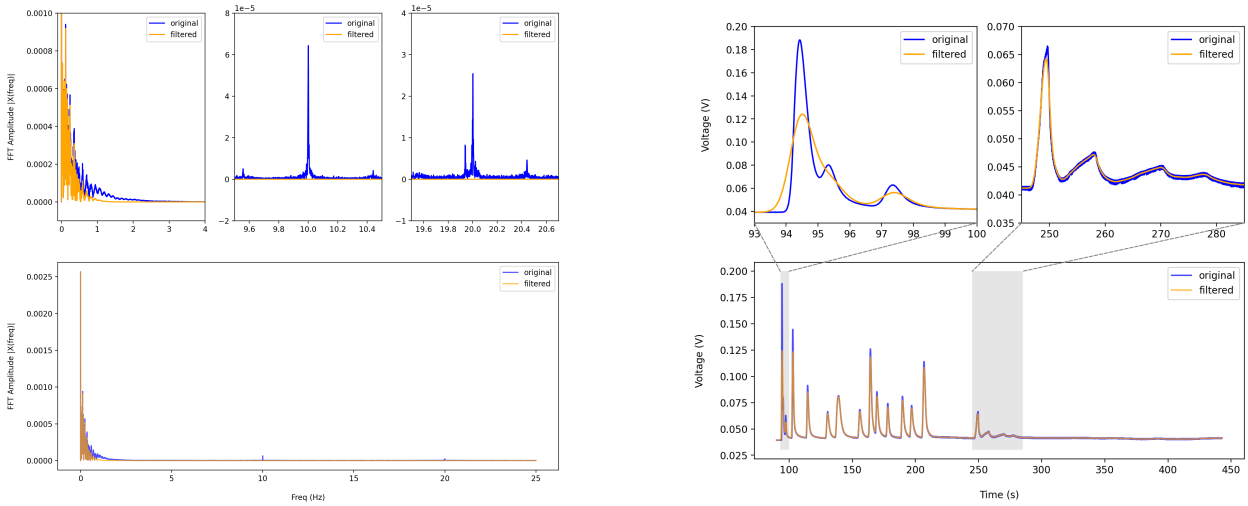


(a) Frequency domain



(b) Time domain

Figure 4.13: Gaussian filtering of sample chromatogram with standard deviation 3



(a) Frequency domain

(b) Time domain

Figure 4.14: Gaussian filtering of sample chromatogram with standard deviation 15

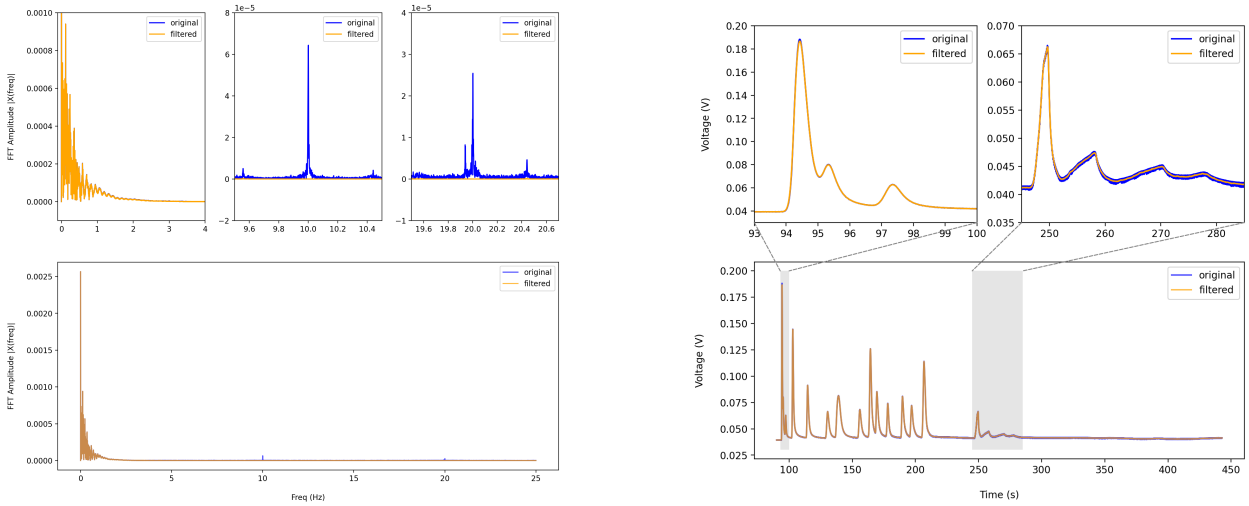
An increase in the standard deviation parameter exhibits a gradual increase in SNR, average retention time shift, and average peak height shift. Very low σ performs poor noise elimination. As the σ increases slightly, the noise elimination becomes more effective as can be seen in the increase in SNR. The average retention time shift and average peak height shift are also relatively low for this range of σ. Higher σ performs aggressive smoothing resulting in loss of peak information and an increase in average peak height shift. Table 4.4 summarizes the effect of varying the standard deviation on the different aspects like SNR, average retention time shift, and average peak height shift.

Table 4.4: Gaussian Filter

σ		0.5	1	1.5	2	2.5	3	4	5	10	15
Output SNR	(dB)	27.19	31.63	34.4	36.61	38.6	40.59	44.86	49.55	63.67	62.4
Average shift in retention time	(ms)	1.5	2.2	4.4	6.8	10.4	11.2	11.9	21	51.1	89.7
Average shift in peak height	(mV)	0.3	0.9	4.7	3.2	4.8	6.8	11.4	16.8	43.6	64.2

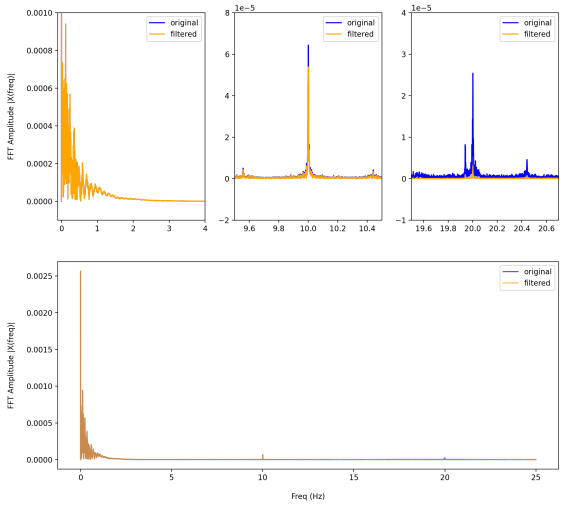
Savitzky-Golay Filter

The Savitzky-Golay (Savgol) filter, also called a least-squares smoothing filter, is a low-pass filter used to reduce high-frequency noise while improving the smoothness of a signal trend [25], [27], [4]. The tuning of the filter window and filter order enables the desired filtering of the input signal. Figure 4.15, Figure 4.15, and Figure 4.15 shows the effect of varying the filter order on the filtering of a sample chromatogram output of the μGC system with filter window of 5 datapoints.

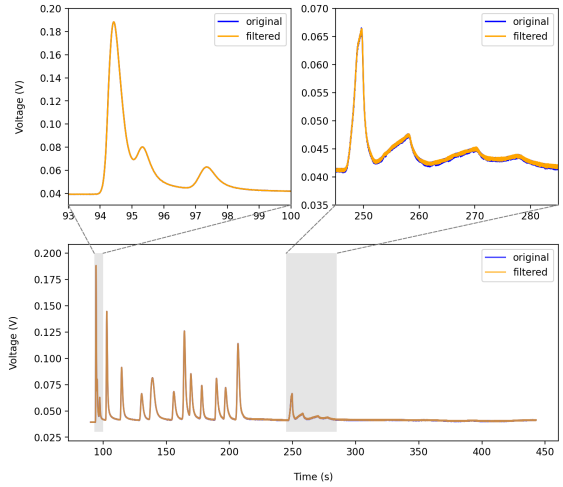


(a) Frequency domain (b) Time domain

Figure 4.15: Savgol filtering of sample chromatogram with window size 5 datapoints and filter order 1

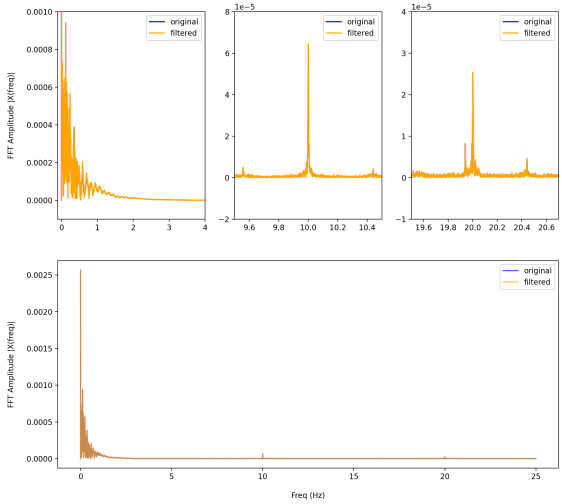


(a) Frequency domain

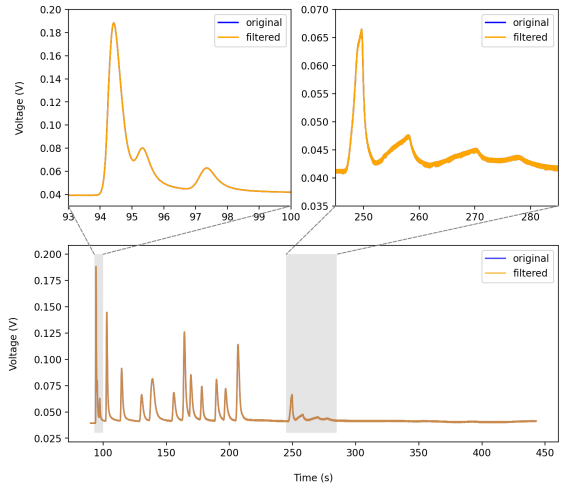


(b) Time domain

Figure 4.16: Savgol filtering of sample chromatogram with window size 5 datapoints and filter order 2



(a) Frequency domain

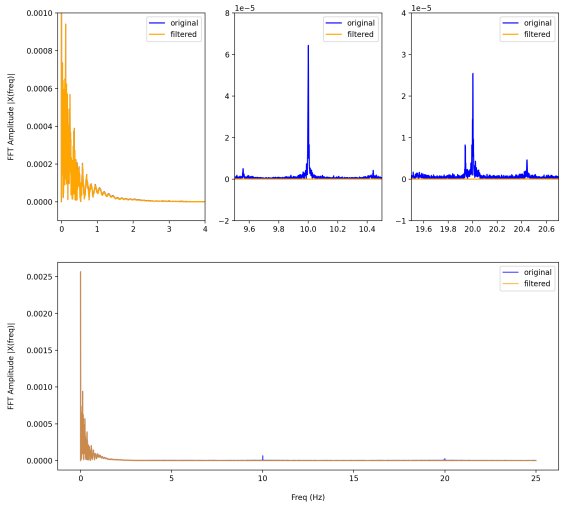


(b) Time domain

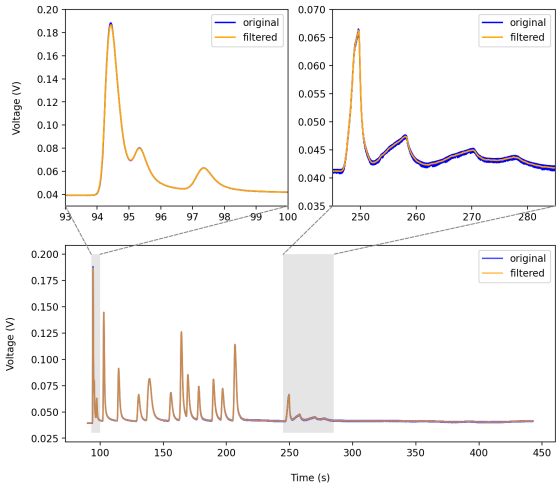
Figure 4.17: Savgol filtering of sample chromatogram with window size 5 datapoints and filter order 4

Low filter order displays desirable high-frequency noise elimination. As the filter order increases, the noise elimination deteriorates and the SNR was observed to decrease. The filter window was then varied by choosing a low filter order. Figure 4.18, Figure 4.19,

and Figure 4.20 shows the effect of varying the filter window on the filtering of a sample chromatogram output of the μGC system with filter order of 1.

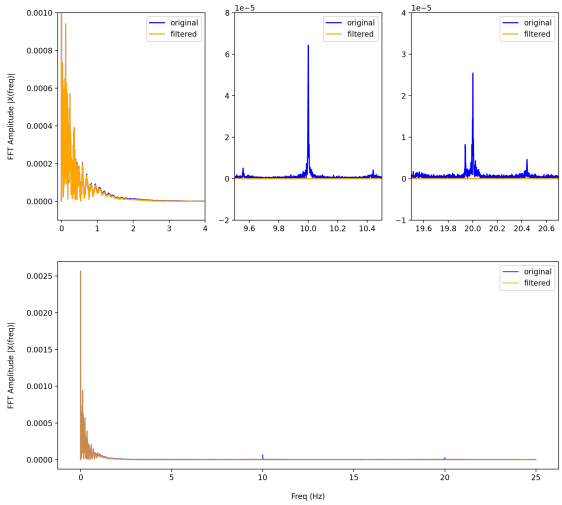


(a) Frequency domain

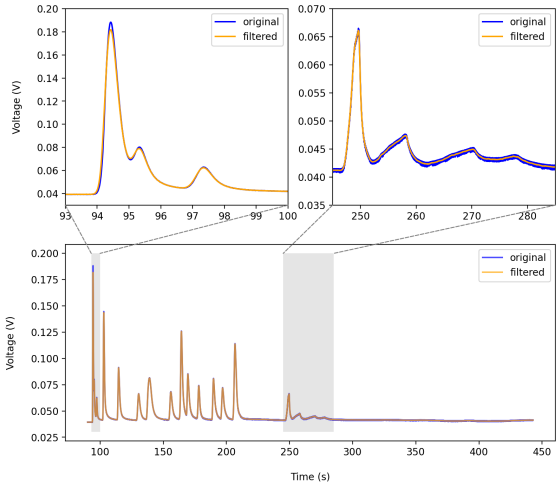


(b) Time domain

Figure 4.18: Savgol filtering of sample chromatogram with filter order 1 and window size 5 datapoints

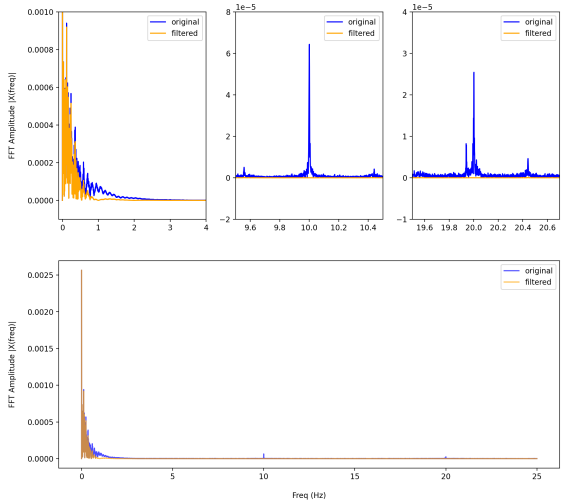


(a) Frequency domain

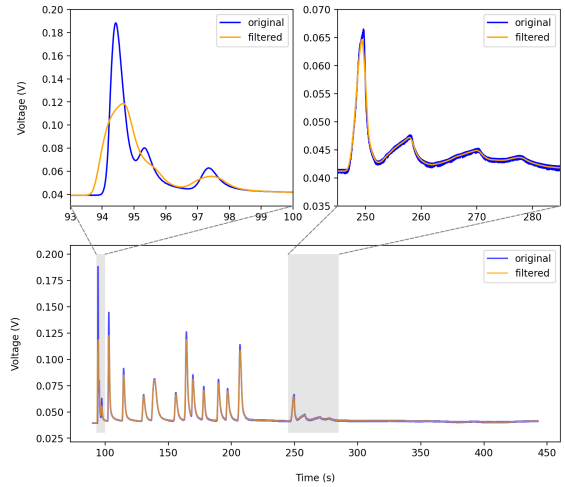


(b) Time domain

Figure 4.19: Savgol filtering of sample chromatogram with filter order 1 and window size 10 datapoints



(a) Frequency domain



(b) Time domain

Figure 4.20: Savgol filtering of sample chromatogram with filter order 1 and window size 50 datapoints

As the filter window is increased, there is a gradual increase in the average retention time shift and average peak height shift. Additionally, it smoothens out the signal aggressively, resulting in a loss of peak information. Table 4.5, Table 4.6, Table 4.7, and Table 4.8 summarizes the effect of varying the filter order on the different aspects like SNR, average retention time shift, and average peak height shift for filter window of 5, 10, 20, and 50 datapoints respectively.

Table 4.5: Savitzky Golay Filter (Window = 5 datapoints)

Filter Order		1	2	4
Output SNR	(dB)	33.83	28.85	25.53
Average shift in retention time	(ms)	3.1	1.0	0
Average shift in peak height	(mV)	1.7	0.1	0

Table 4.6: Savitzky Golay Filter (Window = 10 datapoints)

Filter Order		1	2	5
Output SNR	(dB)	39.33	32.83	30.31
Average shift in retention time	(ms)	0.7	9.1	9.0
Average shift in peak height	(mV)	6.5	0	0.1

Table 4.7: Savitzky Golay Filter (Window = 20 datapoints)

Filter Order		1	2	5
Output SNR	(dB)	47.60	38.92	34.30
Average shift in retention time	(ms)	28.6	8.7	8.8
Average shift in peak height	(mV)	22.2	1.4	0.1

Table 4.8: Savitzky Golay Filter (Window = 50 datapoints)

Filter Order		1	2	5
Output SNR	(dB)	51.18	50.59	47.48
Average shift in retention time	(ms)	235.7	51.7	20.0
Average shift in peak height	(mV)	69.7	25.2	6.2

4.3.3 Peak Detection

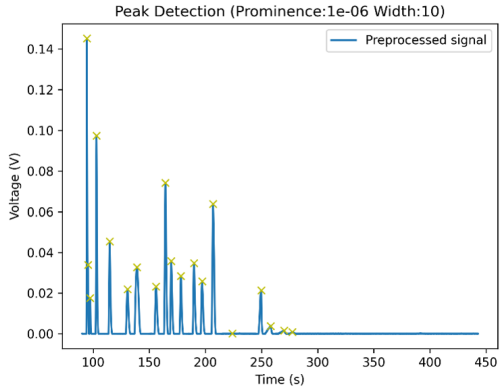
After filtering out the noise from the output chromatogram, the peak detection process is applied to calculate the retention time of the compounds present in the output chromatogram [28]. The Python library for signal processing was used to perform the peak detection [5]. The parameters such as prominence and width need to be tuned for the type of data the peak detection algorithm is applied to.

Prominence

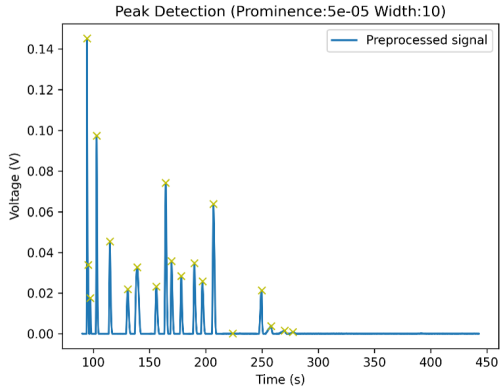
The peak prominence is a measure of the intrinsic height and location of the peak relative to other peaks in the input signal. Figure 4.21 shows the effect of varying the prominence on peak identification for a sample chromatogram output of the μGC system. An optimal value for the prominence should be chosen such that it can distinguish noisy peak amplitude from signal peak amplitudes.

Width

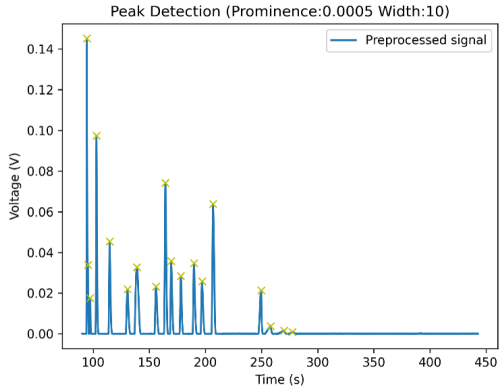
The peak width is a measure of the base width of the peaks identified in the input signal. Figure 4.22 shows the effect of varying the width on peak identification for a sample chromatogram output of the μGC system. Similar to prominence, choosing an optimal value of width should clearly distinguish noisy peaks from signal peaks for correct peak identification.



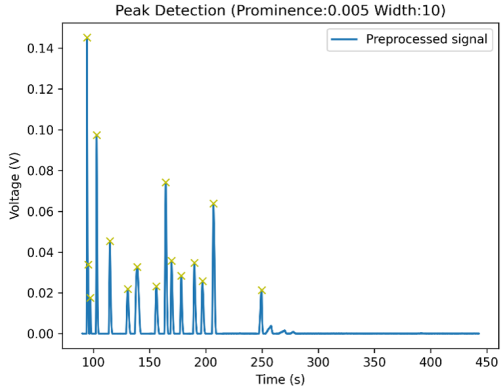
(a) Peak prominence 0.000005



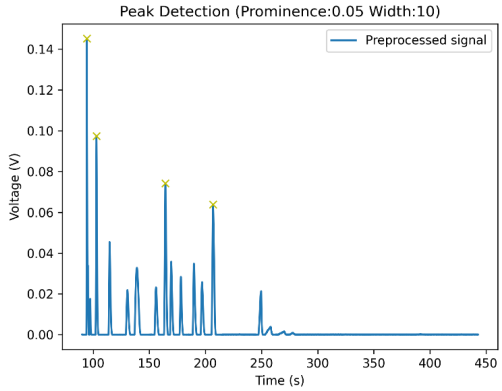
(b) Peak prominence 0.00005



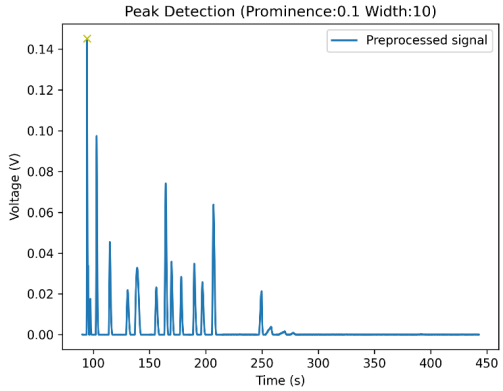
(c) Peak prominence 0.0005



(d) Peak prominence 0.005

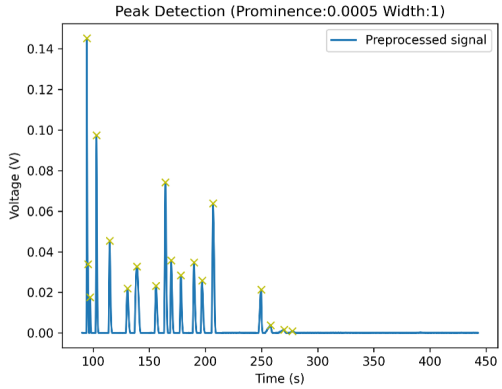


(e) Peak prominence 0.05

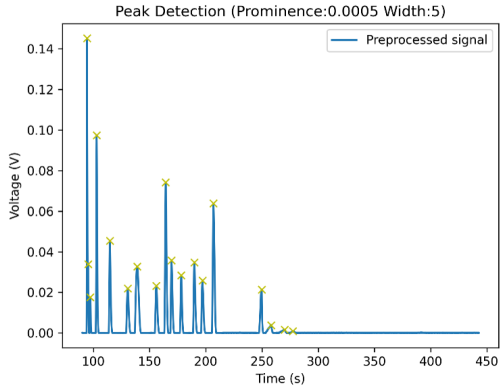


(f) Peak prominence 0.1

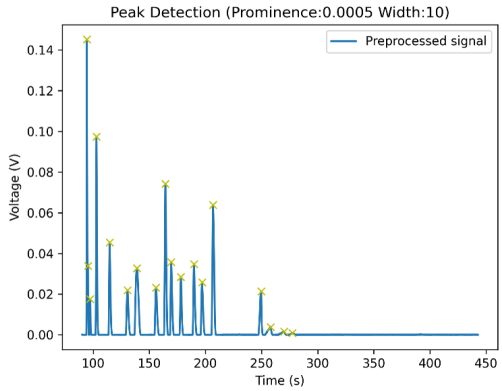
Figure 4.21: Varying peak prominence of peak detection in sample output chromatogram



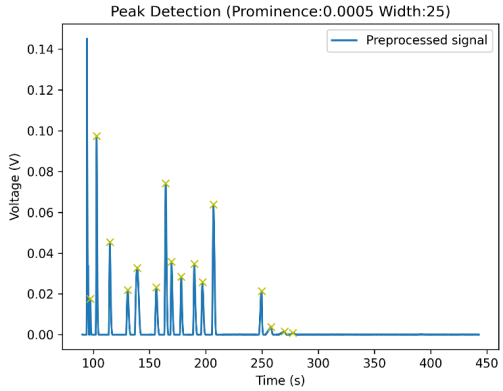
(a) Peak width 1



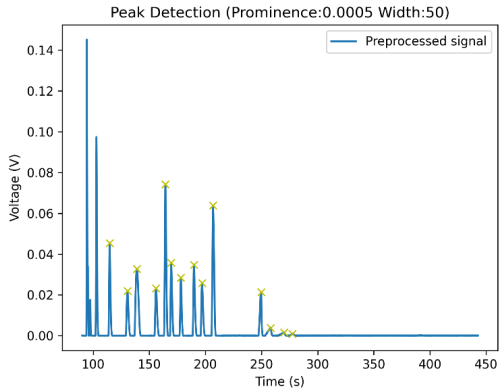
(b) Peak width 5



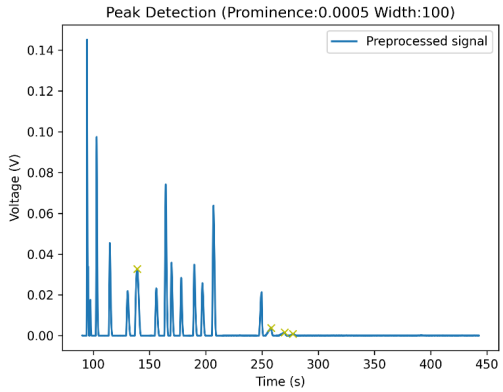
(c) Peak width 10



(d) Peak width 25



(e) Peak width 50



(f) Peak width 100

Figure 4.22: Varying peak width of peak detection in sample output chromatogram

4.3.4 Results

Based on the analysis performed for both signal filtering and peak detection for a sample chromatogram provides crucial information in selecting optimal parameters to get desired performance for the analysis of output chromatograms. The optimal signal filtering technique was chosen such that the filtered chromatogram exhibited adequate noise filtering, low average retention time shift, and low average peak height shift.

Table 4.9: Fixed chromatogram processing tuning parameters

Parameter	**Value**
Filter	
Type	Savitzky-Golay
Order	1
Window	10
Peak Detection	
Prominence	0.0005
Width	10

The optimal peak detection parameters were chosen based on observation from the analysis of prominence and width as described in Section 4.3.3. Table 4.9 displays the fixed parameters for the signal filtering and peak detection steps of the output chromatogram processing. Peak detection was then performed for six different chromatograms. Table 4.10 shows the peak detection accuracy for the six different chromatograms, while an average peak detection accuracy of 94.2% was achieved. The list of identified peaks directly corresponds to the retention time of compounds eluting from the analytical column. Thus, the peak detection output is used to extract the retention time information of the compounds detected in the output chromatogram.

Table 4.10: Peak detection accuracy for six different types of chromatogram

Chromatogram	**Peak Detection Accuracy (%)**
21 compound mixture (BPY)	98.36
21 compound mixture (HOEMIM)	96.42
21 compound mixture (N1888)	93.09
EPA VOC Mix 2 (BPY)	100
EPA VOC Mix 2 (HOEMIM)	97.5
EPA VOC Mix 2 (N1888)	79.86

Chapter 5

Multidimensional Compound Identification Analysis

Typical GC systems perform separation and analysis of sample mixture using a single analytical column. With advancements in μGC components, it is possible to develop a μGC system with two or more analytical columns in parallel. The analytical columns are generally chosen such that they have distinct polarities. Compounds present in a sample mixture can then be separated distinctly as they pass through the multiple parallel analytical columns. This effectively increases the range of compounds that can be separated based on the intrinsic properties of the chosen analytical columns.

A multidimensional analysis [19] is required to process the outputs from multiple analytical columns into a combined result. In this chapter, we first describe a search-based multidimensional compound identification algorithm developed to perform compound detection on the outputs obtained from the portable μGC system. Later, we discuss the results of the multidimensional compound identification for different test scenarios.

5.1 Retention Time Library Generation

A prerequisite for search-based algorithms is a master lookup library of compounds against different analytical columns and their corresponding retention times, containing all the useful

information used to perform an analysis. To develop the retention time library, multiple experiments as described in Section 4.2 were performed with the test mixture 1 (Appendix A) and output chromatograms for the stationary phases BPY, HOEMIM, and N1888 were gathered. For the set of output chromatograms obtained for each stationary phase, the following steps were performed:

- Process each chromatogram using the tuned parameters as described in Section 4.3.4 to obtain a list of retention times corresponding to the identified peaks, wherein each peak corresponds to a specific compound.
- For each compound, the mean and standard deviation of the retention times are calculated.

The combined result for all the stationary phases obtained from the above steps forms the retention time library for the given sample mixture, which is presented in Appendix B.

5.2 Multidimensional Compound Identification Algorithm

The multidimensional compound identification algorithm is described in Figure 5.1. The following subsections explain the algorithm in further detail.

5.2.1 Identifying compounds from chromatogram

The output chromatogram of a given stationary phase of a test run is processed using the fixed parameters as described in Section 4.3.4 to obtain a list of retention times corresponding

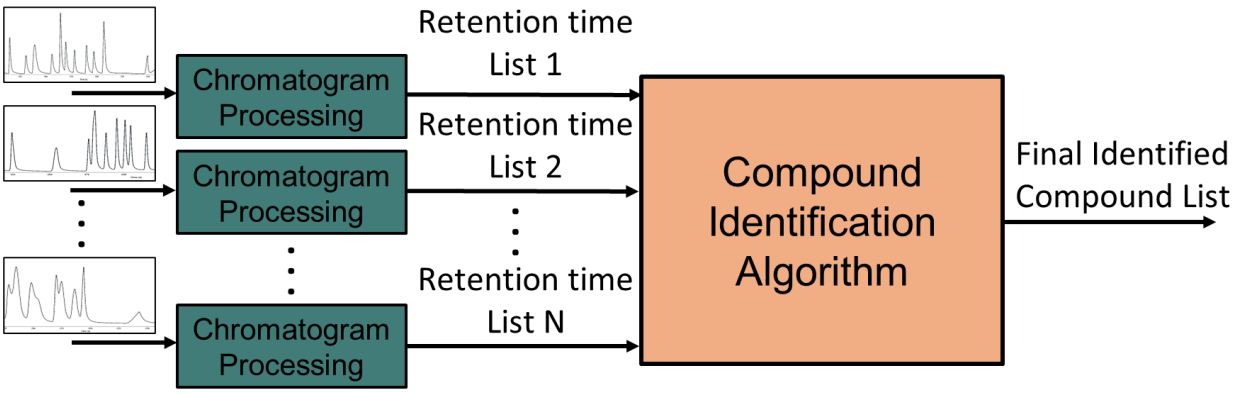


Figure 5.1: Multidimensional compound identification algorithm

to the identified peaks in the chromatogram. Each value in the retention time list is checked to be within the range shown in Equation 5.1 for all the compounds present in the master lookup library corresponding to the appropriate stationary phase.

$$[\mu - a \cdot \sigma, \mu + a \cdot \sigma] \tag{5.1}$$

It must be noted that in the range shown in Equation 5.1, the standard deviation multiplication factor (a) is a tunable parameter that is chosen as 3 based on the empirical rule [24] assuming that the dataset used to generate the retention time library is normally distributed.

In the case where multiple compounds are identified to be within the range shown in Equation 5.1, the coelution margin factor is used to determine the possibility of the peak being associated with coeluting compounds. The coelution margin factor is a tunable parameter that is chosen as 400 ms based on observation.

The possible outcomes for each identified peak after performing the above steps is a list of coeluting compounds, a unique compound, or no compound. The outcomes of all the identified peaks are combined to obtain a list of identified compounds for the given stationary phase.

5.2.2 Compound List Merging

The list of identified compounds is generated by following the procedure described in Section 5.2.1 for all the stationary phase chromatograms provided as an input to the multidimensional compound identification algorithm. The compound list generated for every stationary phase chromatogram is restructured to obtain a matrix wherein every compound present in the retention time library is associated with either of the three outcomes, Y, N, or DC, for each stationary phase. Table 5.1 elaborates on the significance of each of the outcomes.

Table 5.1: Outcomes of compound identification in a given stationary phase

Outcome	Description
Yes (Y)	Compound identified
No (N)	Compound not identified
Don't Care (DC)	Compound information not present in lookup library

The obtained matrix that correlates a given compound and its identification outcome in every stationary phase is evaluated based on the following set of rules to determine the final identification result for that compound:

- **All Yes (Y): Identified**

  If all the outcomes for a given compound are Yes (Y), the final result is evaluated as "Identified". An example of this is the case for compound A in Table 5.2.

- **At least one No (N): Not Identified**

  If at least one outcome for a given compound is No (N), the final result is evaluated as "Not Identified". An example of this is the case for compound B in Table 5.2

- Furthermore, if at least one of the outcomes is marked as Don't Care (DC), it is ignored during the decision-making process. Only the remaining stationary phase outcomes are

used to assert whether the compound is "Identified" or "Not Identified". An example of this is the case for compound C and compound D in Table 5.2.

Table 5.2: Compound identification examples

Cases	Stationary Phase 1	Stationary Phase 2	Stationary Phase 3	Result
Compound A	Y	Y	Y	Identified
Compound B	Y	Y	N	Not Identified
Compound C	Y	Y	DC	Identified
Compound D	DC	Y	N	Not Identified

Figure 5.2 shows the compound list merging algorithm, which is the last part of the multidimensional compound identification algorithm used to generate the final list of identified compounds in a test run.

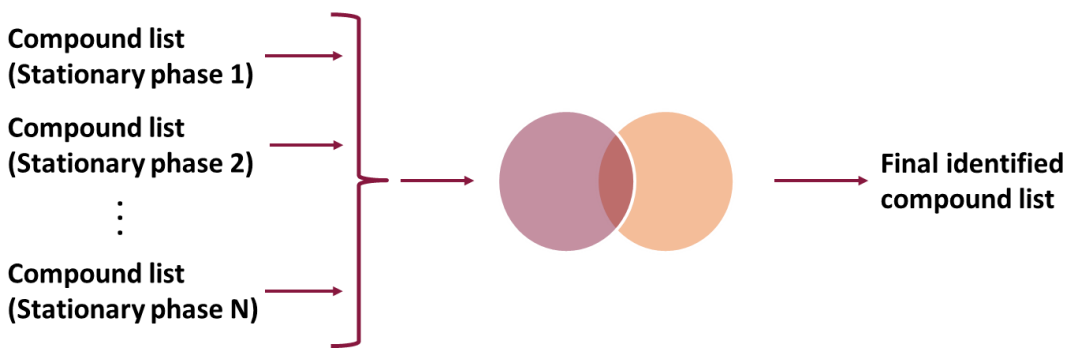


Figure 5.2: Compound list merging algorithm

5.3 Testing and Results

The multidimensional analysis was tested using two different sample mixtures (Appendix A). Section 4.2 describes the experiment that was performed using these sample mixtures.

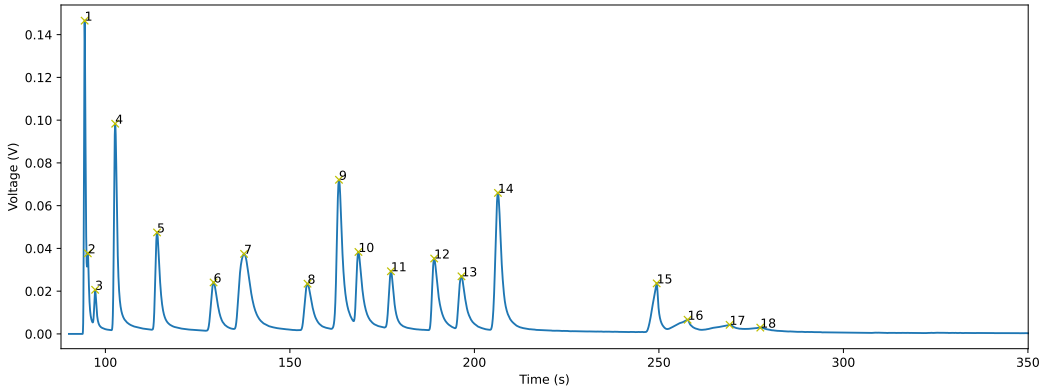
5.3.1 Test scenario 1

In this test scenario, we performed an experiment with the injection of test mixture 1, whose composition is listed in Table A.1, and gathered output chromatograms for the stationary phases BPY, HOEMIM, and N1888. The 21-compound mixture has complex properties with the presence of multiple coeluting compounds in each of the stationary phases. The multidimensional compound detection algorithm was used to determine the maximum number of compounds that could be identified from the complex test mixture.

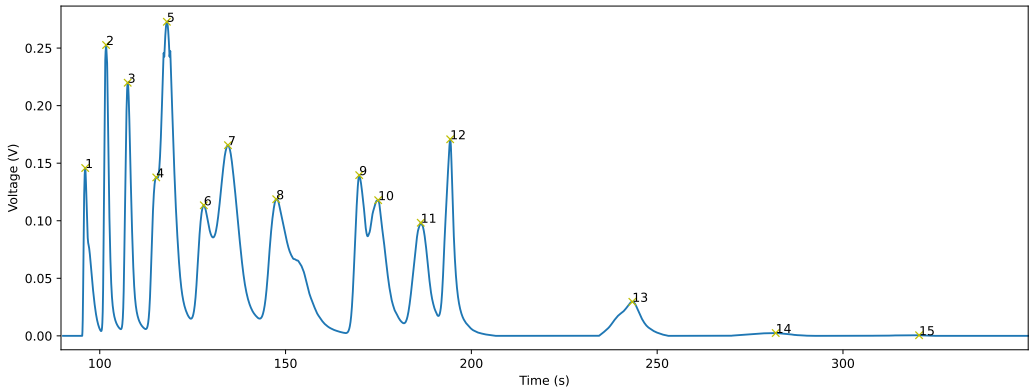
Figure 5.3 shows the output chromatograms obtained for all three stationary phases, and it can be observed that 18 peaks were identified in BPY, 16 peaks were identified in HOEMIM, and 15 peaks were identified in N1888, respectively. The multidimensional algorithm combines the individual chromatogram information to derive the final result of identifying all the 21 compounds present in the sample mixture, as shown in Table 5.3. To observe the identification of coeluting compounds, consider the compounds 2-Chlorotoluene and Isobutylbenzene. From Table 5.3, it can be observed that peak 9 in BPY and peak 7 in HOEMIM correspond to both these compounds. This is because the compounds 2-Chlorotoluene and Isobutylbenzene coelute in both the stationary phases BPY and HOEMIM. However, in the stationary phase N1888, these compounds are distinctly separated as can be observed from peak 8 corresponding to 2-Chlorotoluene and peak 7 corresponding to Isobutylbenzene. The combined information from all three stationary phases is utilized to correctly identify both compounds from the test mixture 1.

Table 5.3: Test scenario 1 multidimensional compound identification result

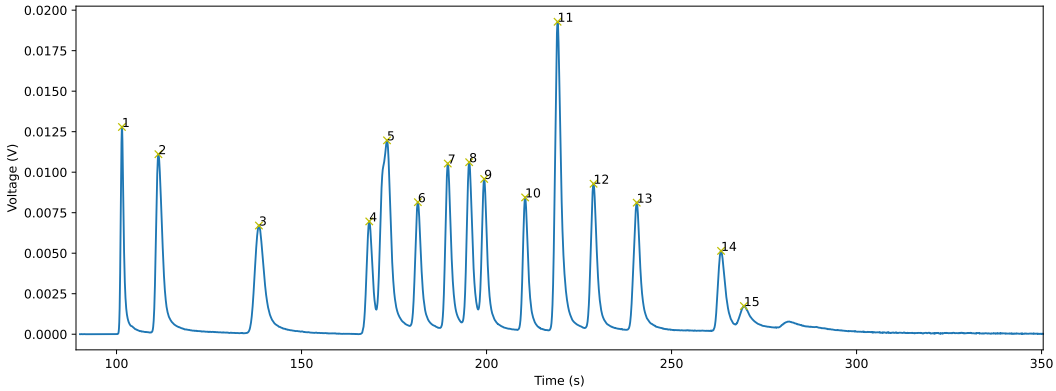
Compound	**BPY**	**HOEMIM**	**N1888**	**Result**
1,2,4-Trichlorobenzene	14	11	13	Identified
1,2-Dichlorobenzene	12	9	11	Identified
2,5-Dichlorotoluene	13	10	12	Identified
2-Chlorotoluene	9	7	8	Identified
2-Nitrotoluene	16	14	15	Identified
3-Nitrotoluene	17	15	DC	Identified
4-Nitrotoluene	18	DC	DC	Identified
Benzene	4	2	2	Identified
Benzyl chloride	14	12	11	Identified
Butylbenzene	11	8	10	Identified
Ethylbenzene	6	4	4	Identified
Heptane	1	1	1	Identified
Isobutylbenzene	9	7	7	Identified
Napthalene	15	13	14	Identified
Nonane	3	1	3	Identified
Octane	2	1	2	Identified
p-Xylene	7	5	5	Identified
Styrene	10	8	9	Identified
Toluene	5	3	3	Identified
m-Xylene	7	5	5	Identified
o-Xylene	8	6	6	Identified



(a) BPY chromatogram



(b) HOEMIM chromatogram



(c) N1888 chromatogram

Figure 5.3: Chromatograms from different stationary phases for test mixture 1

5.3.2 Test scenario 2

In this test scenario, we performed an experiment with the injection of test mixture 2, whose composition is listed in Table A.1, and gathered output chromatograms for the stationary phases BPY, HOEMIM, and N1888. The compounds Benzene, Butylbenzene, Ethylbenzene, Naphthalene, Styrene, Toluene, 1,2,4-Trichlorobenzene, and m-Xylene in the test mixture 2 are identifiable compounds whose information is present in the retention time library. The remaining compounds Bromobenzene, 1,2,3-Trichlorobenzene, and 1,2,4-Trimethylbenzene in test mixture 2 are not identifiable since their information is not present in the retention time library. This allows us to ascertain and verify the performance of the compound detection algorithm in determining the presence of the compounds of interest in a given sample which may also contain numerous other compounds whose identification might not be of interest.

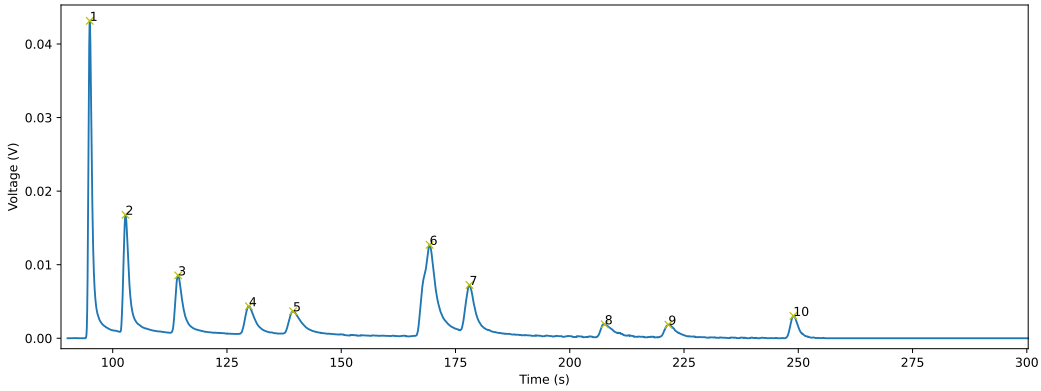
Figure 5.4 shows the output chromatograms obtained for all three stationary phases, and it can be observed that 10 peaks were identified in BPY, 11 peaks were identified in HOEMIM, and 11 peaks were identified in N1888, respectively. The multidimensional compound identification algorithm identifies 9 compounds as shown in Table 5.4. From the 8 identifiable compounds, the algorithm identified 7 compounds present in the test mixture. The algorithm did not identify the compound Naphthalene since there was no peak corresponding to this compound detected in the stationary phase HOEMIM. The solvent present in the test mixture was identified as Heptane due to the similar retention time of the compounds.

Another observation is that the multidimensional compound identification algorithm identified both m-Xylene and p-Xylene even though only m-Xylene is present in the test mixture. This is because m-Xylene and p-Xylene are isomers that have similar retention times in all the stationary phases used for experimentation. Information on such coeluting compounds

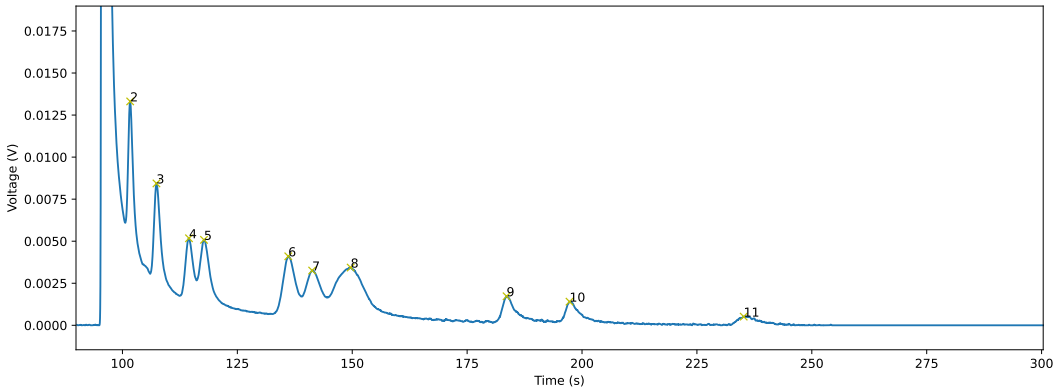
in a stationary phase where they are distinctly separated is necessary to correctly identify the compounds in a sample mixture.

Table 5.4: Test scenario 2 multidimensional compound identification result

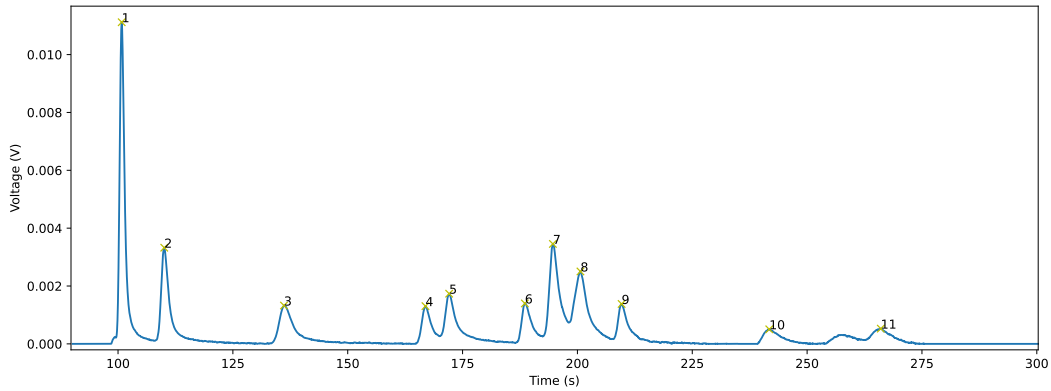
Compound	**BPY**	**HOEMIM**	**N1888**	**Result**
1,2,4-Trichlorobenzene	8	9	10	Identified
1,2-Dichlorobenzene	-	-	-	Not Identified
2,5-Dichlorotoluene	-	-	-	Not Identified
2-Chlorotoluene	-	6	7	Not Identified
2-Nitrotoluene	-	-	-	Not Identified
3-Nitrotoluene	-	-	DC	Not Identified
4-Nitrotoluene	-	DC	DC	Not Identified
Benzene	2	2	2	Identified
Benzyl chloride	8	-	-	Not Identified
Butylbenzene	7	7	9	Identified
Ethylbenzene	4	4	4	Identified
Heptane	1	1	1	Identified
Isobutylbenzene	-	6	6	Not Identified
Napthalene	10	-	11	Not Identified
Nonane	-	1	3	Not Identified
Octane	-	1	2	Not Identified
p-Xylene	5	5	5	Identified
Styrene	6	7	8	Identified
Toluene	3	3	3	Identified
m-Xylene	5	5	5	Identified
o-Xylene	-	-	-	Not Identified



(a) BPY chromatogram



(b) HOEMIM chromatogram



(c) N1888 chromatogram

Figure 5.4: Chromatograms from different stationary phases for test mixture 2

Chapter 6

Conclusion

In this thesis, the embedded software implementations enabling the portability aspects of a μGC system were described. The robust and scalable design of the modular μGC system is showcased. A μGC system with dual parallel separation columns was set up and the results of the μGC system demonstrate the ability to separate complex compound mixtures. An extensive analysis of the different steps in chromatogram processing like signal filtering and peak detection was performed as a part of this thesis. The optimal parameters chosen for the signal filtering and peak detection techniques were tested for six different types of chromatograms and an average peak detection accuracy of 94.2% was achieved.

A novel multidimensional compound identification algorithm was developed. A detailed process of generating a retention time library for a total of 21 compounds was described. Different test scenarios for the multidimensional compound identification were tested and the results highlight the advantages of a multidimensional analysis over conventional analysis techniques.

6.1 Future Work

The reported modular μGC system can be further improved by making the following design changes.

- Provisioning and implementing proportional valve control logic in the modular μGC system could facilitate flow control in GC experimentation.
- Providing internet connectivity to the modular μGC system could enable connectivity to servers and realize a network of highly-scalable field portable standalone GC systems.
- Introducing an infrastructure to run continuous GC experiments to collect a huge amount of data allows the development of an extensive pre-compiled library incorporating a vast set of compounds and their retention times. Additionally, it enables data collection for machine learning analysis.

Deliberate design considerations in the aspects of power supply through a battery management system, along with minor modifications in the modular hardware architecture, could pave the way to implement the modular μGC system in a wearable form factor. With a robust and scalable software architecture already in place, numerous applications could arise through the advent of a wearable μGC system.

Bibliography

[1] scipy.signal.gaussian_filter1d. https://docs.scipy.org/doc/scipy-0.14.0/reference/generated/scipy.ndimage.filters.gaussian_filter1d.html.

[2] scipy.signal.find_peaks. https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.iirfilter.html.

[3] Pyqt. https://riverbankcomputing.com/software/pyqt/intro. Accessed: 2021-01-90.

[4] scipy.signal.savgol_filter. https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.savgol_filter.html.

[5] scipy.signal.find_peaks. https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.find_peaks.html.

[6] *BASIC CONCEPTS AND TERMS*, chapter 2, pages 15–35. John Wiley Sons, Ltd, 2019. ISBN 9781119450795. doi: https://doi.org/10.1002/9781119450795.ch2. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119450795.ch2.

[7] M. Agah, J. A. Potkay, G. Lambertus, R. Sacks, and K. D. Wise. High-performance temperature-programmed microfabricated gas chromatography columns. *Journal of Microelectromechanical Systems*, 14(5):1039–1050, 2005. ISSN 1057-7157. doi: 10.1109/jmems.2005.856648.

[8] Agilent. *GC Method Development*.

[9] Syed Ali, Mehdi Ashraf-Khorassani, Larry T. Taylor, and Masoud Agah. Mems-based semi-packed gas chromatography columns. *Sensors and Actuators B: Chemical*, 141(1):

309–315, 2009. ISSN 0925-4005. doi: https://doi.org/10.1016/j.snb.2009.06.022. URL https://www.sciencedirect.com/science/article/pii/S092540050900495X.

[10] G. Bianchi and R. Sorrentino. *Electronic Filter Simulation & Design*. McGraw-Hill Education, 2007. ISBN 9780071494670. URL https://books.google.com/books?id=5S3LCIxnYCcC.

[11] Frank L. Dorman, Joshua J. Whiting, Jack W. Cochran, and Jorge Gardea-Torresdey. Gas chromatography. *Analytical Chemistry*, 82(12):4775–4785, 2010. doi: 10.1021/ac101156h. URL https://doi.org/10.1021/ac101156h. PMID: 20504041.

[12] Jason P. Dworkin. *Chromatographic Co-elution*, pages 302–302. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011. ISBN 978-3-642-11274-4. doi: 10.1007/978-3-642-11274-4_289. URL https://doi.org/10.1007/978-3-642-11274-4_289.

[13] Apoorva Garg, Muhammad Akbar, Eric Vejerano, Shree Narayanan, Leyla Nazhandali, Linsey C. Marr, and Masoud Agah. Zebra gc: A mini gas chromatography system for trace-level determination of hazardous air pollutants. *Sensors and Actuators B: Chemical*, 212:145–154, 2015. ISSN 09254005. doi: 10.1016/j.snb.2014.12.136.

[14] A. Gholizadeh, M. Chowdhury, and M. Agah. Parallel ionic liquid semi-packed microfabricated columns for complex gas analysis. *Anal Chem*, 92(15):10635–10642, 2020. ISSN 1520-6882 (Electronic) 0003-2700 (Linking). doi: 10.1021/acs.analchem.0c01721. URL https://www.ncbi.nlm.nih.gov/pubmed/32640785.

[15] J. Calvin Giddings. Two-dimensional separation: Concepts and promise. *Anal Chem*, 56, 1984. doi: 10.1021/ac00276a003,.

[16] Richard Haddad and Ali Akansu. A class of fast gaussian binomial filters for speech

and image processing. *Signal Processing, IEEE Transactions on*, 39:723 – 727, 04 1991. doi: 10.1109/78.80892.

[17] Texas Instruments. *TI-RTOS 2.20 User's Guide*. Texas Instruments.

[18] Texas Instruments. Ti rtos kernel. Report, 2020. URL https://www.ti.com/lit/ug/spruex3v/spruex3v.pdf?ts=1629080917657&ref_url=https%253A%252F%252Fwww.google.com%252F.

[19] Wei X Zhang X Kim S, Koo I. A method of finding optimal weight factors for compound identification in gas chromatography-mass spectrometry. *BMC Bioinformatics*, 212: 145–154, 2015. ISSN 09254005. doi: 10.1093/bioinformatics/bts083.

[20] Shaunak Manurkar. Modular gc: A fully integrated micro gas chromatography system. unpublished thesis, 2021.

[21] S. Niewiadomski. *Filter Handbook: A Practical Design Guide*. Elsevier Science, 2013. ISBN 9781483144627. URL https://books.google.com/books?id=o7MgBQAAQBAJ.

[22] Omega. Thin film din class a rtd elements. Report. URL https://assets.omega.com/pdf/test-and-measurement-equipment/temperature/sensors/rtds/F3105_3100_W2100_2200.pdf.

[23] B. P. Regmi and M. Agah. Micro gas chromatography: An overview of critical components and their integration. *Anal Chem*, 90(22):13133–13150, 2018. ISSN 1520-6882 (Electronic) 0003-2700 (Linking). doi: 10.1021/acs.analchem.8b01461. URL https://www.ncbi.nlm.nih.gov/pubmed/30359512.

[24] Sheldon M. Ross. Chapter 2 - descriptive statistics. In Sheldon M. Ross, editor, *Introduction to Probability and Statistics for Engineers and Scientists (Sixth Edition)*, pages 11–61. Academic Press, sixth edition edition, 2021. ISBN 978-0-12-

824346-6. doi: https://doi.org/10.1016/B978-0-12-824346-6.00011-9. URL https://www.sciencedirect.com/science/article/pii/B9780128243466000119.

[25] Abraham. Savitzky and M. J. E. Golay. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36(8):1627–1639, 1964. doi: 10.1021/ac60214a047. URL https://doi.org/10.1021/ac60214a047.

[26] sigmaaldrich. Epa voc mix 2. Report. URL https://www.sigmaaldrich.com/US/en/product/supelco/crm48777.

[27] Steven W. Smith. *The Scientist and Engineer's Guide to Digital Signal Processing*. California Technical Publishing, USA, 1997. ISBN 0966017633.

[28] Yu W Yang C, He Z. Comparison of public peak detection algorithms for maldi mass spectrometry data analysis. *BMC Bioinformatics*, 212:145–154, 2015. ISSN 09254005. doi: 10.1186/1471-2105-10-4.

[29] Junghoon Yeom. *Micro-preconcentrator Technology for Portable Gas Chromatography System*, pages 1–8. 09 2015. ISBN 978-94-007-6178-0. doi: 10.1007/978-94-007-6178-0_100964-1.

Appendices

Appendix A

Sample Mixtures

Two different sample mixtures used during the experimentation to collect data for this thesis are described in Table A.1.

Table A.1: Sample Mixtures

Test Mixture 1	**Test Mixture 2**
Heptane	Benzene
Octane	Bromobenzene
Nonane	Butylbenzene
Benzene	Ethylbenzene
Toluene	Naphthalene
Ethylbenzene	Styrene
P-Xylene	Toluene
m-Xylene	1,2,3-Trichlorobenzene
o-Xylene	1,2,4-Trichlorobenzene
Isobutylbenzene	1,2,4-Trimethylbenzene
Styrene	m-Xylene
Butylbenzene	
1,2-Dichlorobenzene	
2,5-Dichlorotoluene	
1,2,4-Trichlorobenzene	
Benzyl chloride	
Naphthalene	
2-Nitrotoluene	
3-Nitrotoluene	
4-Nitrotoluene	
2-Chlorotoluene	

Test mixture 1 is a 21-compound mixture. This mixture is used to generate the retention time library as described in Section 5.1. The significance of this mixture is that it contains compounds that coelute in different stationary phases like BPY, HOEMIM, and N1888 [14]. For example, the compounds 2-Chlorotoluene and Isobutylbenzene coelute in the stationary phases BPY and HOEMIM. Similarly, the compounds 1,2,4-Trichlorobenzene and Benzyl Chloride coelute in the stationary phase BPY. This allows comprehensive testing of the multidimensional compound identification analysis.

Test mixture 2 is a standard mixture called EPA VOC Mix 2 [26]. This mixture has a total of 11 compounds of which 8 are also a part of Test Mixture 1. This mixture was chosen to develop a test scenario for multidimensional compound identification analysis in the process of identifying compounds of interest from a complex sample mixture.

Appendix B

Retention Time Library

Table B.1: Retention Time Library

Compound	N1888		HOEMIM		BPY	
	μ (min)	σ (min)	μ (min)	σ (min)	μ (min)	σ (min)
Heptane	0.1817	0.0058	0.1009	0.0054	0.0758	0.0050
Octane	0.3458	0.0089	0.1009	0.0054	0.0904	0.0049
Nonane	0.8021	0.0242	0.1009	0.0054	0.1236	0.0047
Benzene	0.3458	0.0089	0.1932	0.0067	0.2167	0.0061
Toluene	0.8021	0.0242	0.2880	0.0090	0.4107	0.0117
Ethylbenzene	1.3019	0.0163	0.4173	0.0074	0.6723	0.0210
P-Xylene	1.3826	0.0159	0.4599	0.0134	0.8139	0.0256
m-Xylene	1.3826	0.0159	0.4599	0.0134	0.8139	0.0256
o-Xylene	1.5217	0.0148	0.6220	0.0180	1.0948	0.0210
2-Chlorotoluene	1.7537	0.0130	0.7281	0.0203	1.2338	0.0183
Isobutylbenzene	1.6563	0.0136	0.7281	0.0203	1.2338	0.0183
Styrene	1.8212	0.0122	0.9416	0.0268	1.3215	0.0163
Butylbenzene	2.0062	0.0114	0.9416	0.0268	1.4643	0.0149
1,2-Dichlorobenzene	2.1543	0.0114	1.3222	0.0129	1.6594	0.0128
2,5-Dichlorotoluene	2.3158	0.0107	1.4007	0.0134	1.7786	0.0115
1,2,4-Trichlorobenzene	2.5163	0.0116	1.5985	0.0103	1.9426	0.0114
Benzyl chloride	2.1543	0.0114	1.7313	0.0060	1.9426	0.0114
Naphthalene	2.9021	0.0139	2.5633	0.0143	2.6513	0.0117
2-Nitrotoluene	3.0137	0.0168	3.1879	0.0394	2.7830	0.0219
3-Nitrotoluene	0.0000	0.0000	3.8623	0.0325	2.9668	0.0305
4-Nitrotoluene	0.0000	0.0000	0.0000	0.0000	3.0880	0.0293

A retention time library was generated using test mixture 1 (Appendix A) and was utilized for the multidimensional compound identification analysis. Table B.1 presents the retention time library, which was generated by following the steps described in Section 5.1.