COMPUTER AIDED ANALYSIS OF

GAS CHROMATOGRAPHIC SPECTRA

COMPUTER AIDED ANALYSIS OF

GAS CHROMATOGRAPHIC SPECTRA

By

EDDIE HANS REBANE, B.Eng.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Master of Engineering

McMaster University

December 1983

MASTER OF ENGINEERING (1983)   McMASTER UNIVERSITY
(Engineering Physics)   Hamilton, Ontario

TITLE: Computer Aided Analysis of Gas Chromatographic
          Spectra

AUTHOR: Eddie Hans Rebane, B.Eng. (McMaster University)

SUPERVISOR: Dr. T. J. Kennett

NUMBER OF PAGES: xiii, 175

ABSTRACT

In this thesis, hardware and software was developed whereby the feasability of using pattern recognition techniques for gas chromatography could be quantitively explored. Processing methods were developed to overcome the distortion inherent in GC spectra. The reproducibility of peak retention times and areas was investigated from the processed spectra. Upon completion of this study, it was possible to amass a large quantity of experimental data for bacterial strains. With such data the exact form of the pattern recognition algorithm could be defined. A possible pattern recognition method is discussed.

ACKNOWLEDGEMENTS

I would like to express my appreciation to all the people that made my research possible and my stay at McMaster very happy and unforgettable. Special thanks to my research supervisor, Dr. T. J. Kennett whose continual encouragement, insight, enthusiasm, and jovial nature made my research and discussions with him pleasant and rewarding.

Thanks also goes to Dr. J. M. Rosenfeld, Dr. O. Hammerberg, and to all the people in their lab for helping me with the chemical and pathological aspects of this thesis.

Throughout my studies I received unbounded support from Ene-Ly, my parents, grandmother, and sister. This was very important to me.

Finally, I wish to thank my friends at NRB, to name a few, Tom, Gary, Colm, Ian, Kenrick, and Brian, all who were very helpful, interesting, and above all had a great sense of humor.

TABLE OF CONTENTS

vi

LIST OF ILLUSTRATIONS

LIST OF TABLES

CHAPTER I

INTRODUCTION

1.1   Introduction

The origin of the field of bacteriology is generally traced to the chemist Pasteur. Following his important findings, made during the middle of the last century, activity in this field which is of vital importance to mankind, has grown continuously. Unfortunately the laboritory techniques employed have not always reflected a corresponding rate of development. For example, until recently the methods of assaying for specific infections has closely resembled those used by early investigators, i.e. placement of a sample into a nutrient medium to increase the sample size through population growth. Once a sufficient sample size has been attained, physical and chemical procedures are used to assist in identification. Unfortunately this procedure is time consuming, generally requiring from one to three days. As a consequence, in a clinical environment patients with infections are often placed in jeopardy while awaiting identification of their infection.

Recent work by Brooks (1), Moss (2), Laforce (3), and others has the potential to alter the manner in which

identification of bacteria is made. If proven valid, the consequence will be that decision times are reduced from days to hours. These methods which utilize measurement of characteristic compounds found in the structure of bacteria themselves or in the fluids (eg. blood, cerebrospinal fluid) of the patients via gas chromatography (GC). This approach must be explored and tested to assess its specificity and reliability, requiring that GC spectra be amassed for a large number of bacterial types. A data base can thus be established provided that the spectra can be reliably analyzed into their constituents. From such a data base the reproduciblilty of sample preparation and also of the analysis methods used can be assessed. Following this, the degree of orthogonality between patterns can be explored in order to estimate what level in specificity can be achieved.

Particular algorithms are required must be developed in order to remove unwanted spectral components ("noise"), and to correct for any the temporal variation of the spectra. Reproducibility of various parameters such as peak location and intensity must also be explored if a pattern recognition algorithm is to be developed. Comparing the unknown spectra with a library of known strains could, if all factors were favourable, yield a reliable prediction for the identification of the unknown.

The present work was undertaken to provide the opportunity to conduct the evaluations thought necessary if the method was to be fully evaluated. To do this it was first necessary to design and construct an acquisition system and to interface it with an existing GC system in order that spectra could be collected. Once this task was accomplished, automated digital signal processing methods had to be devised to analyze the aquired data. From this point it would be possible to begin to establish the desired data base.

1.2   Review Of GC Operating Principles

The main component of a gas chromatography is a column, consisting of non volatile material that has an affinity for the class of compounds to be analysed. Over this column flows a stream of gas called the carrier gas (which is the mobile phase). When a sample is injected at the upstream end it will be vapourized and be swept along the column provided that the temperature of the environment is sufficiently high.

As the carrier gas, under pressure, moves the vapour sample through the column, the sample molecules are partitioned between the gas (i.e. mobile) phase and the stationary phase. The original mixtures of compounds is thus separated into its components. The components with no affinity for the stationary phase travel with the speed of the mobile phase while the components that have greater afinity are retarded by varying degrees depending upon the partitioning between the mobile and stationary phases. As the components are eluted from the column, they are swept into the detector. The retention time of the component is the transition time through the column to the detector.

Different components of a sample reach the detector at different times and are seen as peaks in a gas chromatograph spectrum. A sample GC output (for carboxyilic acids generated from E. Coli) is shown in figure 1.1.

The detector responds to changes in the composition of the gas flowing through it, and as each component reaches the detector, there will be a change in the detector output. The type and functional characteristics of the detector are described in greater detail in section 2.2. Figure 1.2 shows a schematic representation of a gas chromatographic system.
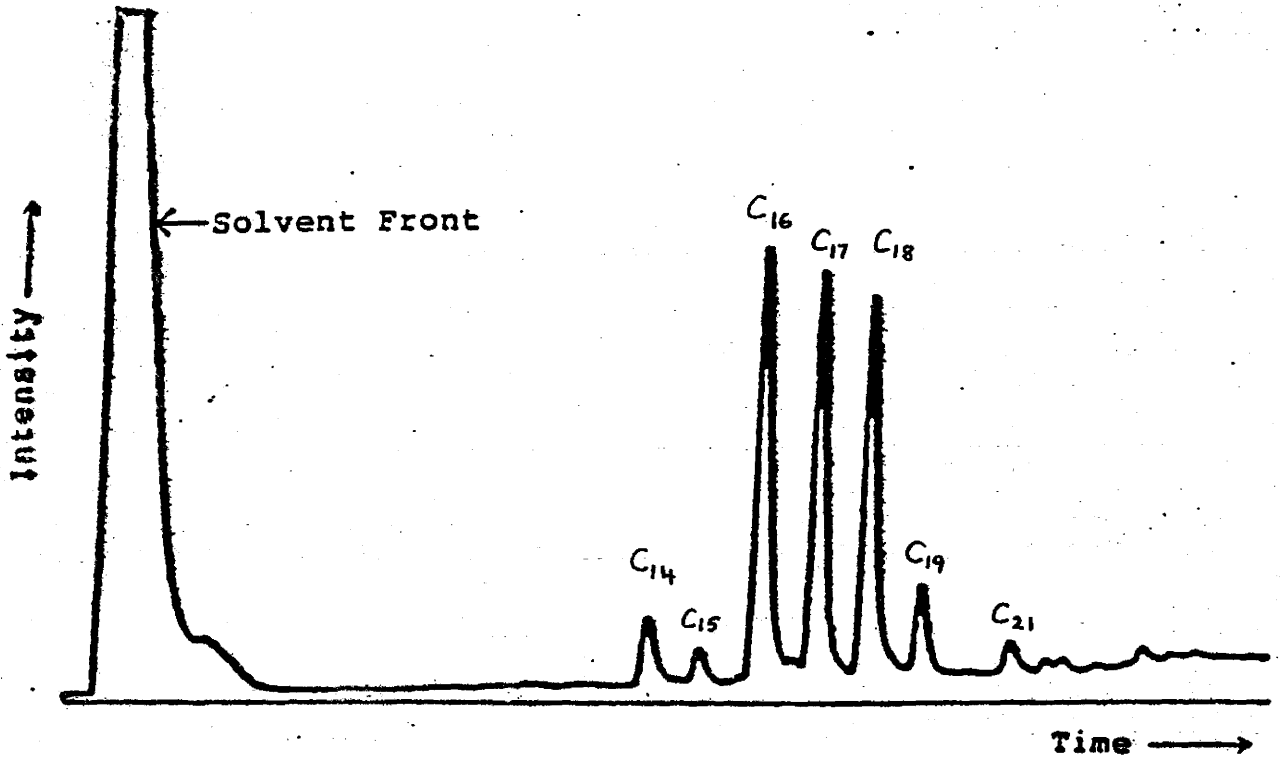
Figure 1.1

Sample of a GC spectra, (in this case of E. Coli)
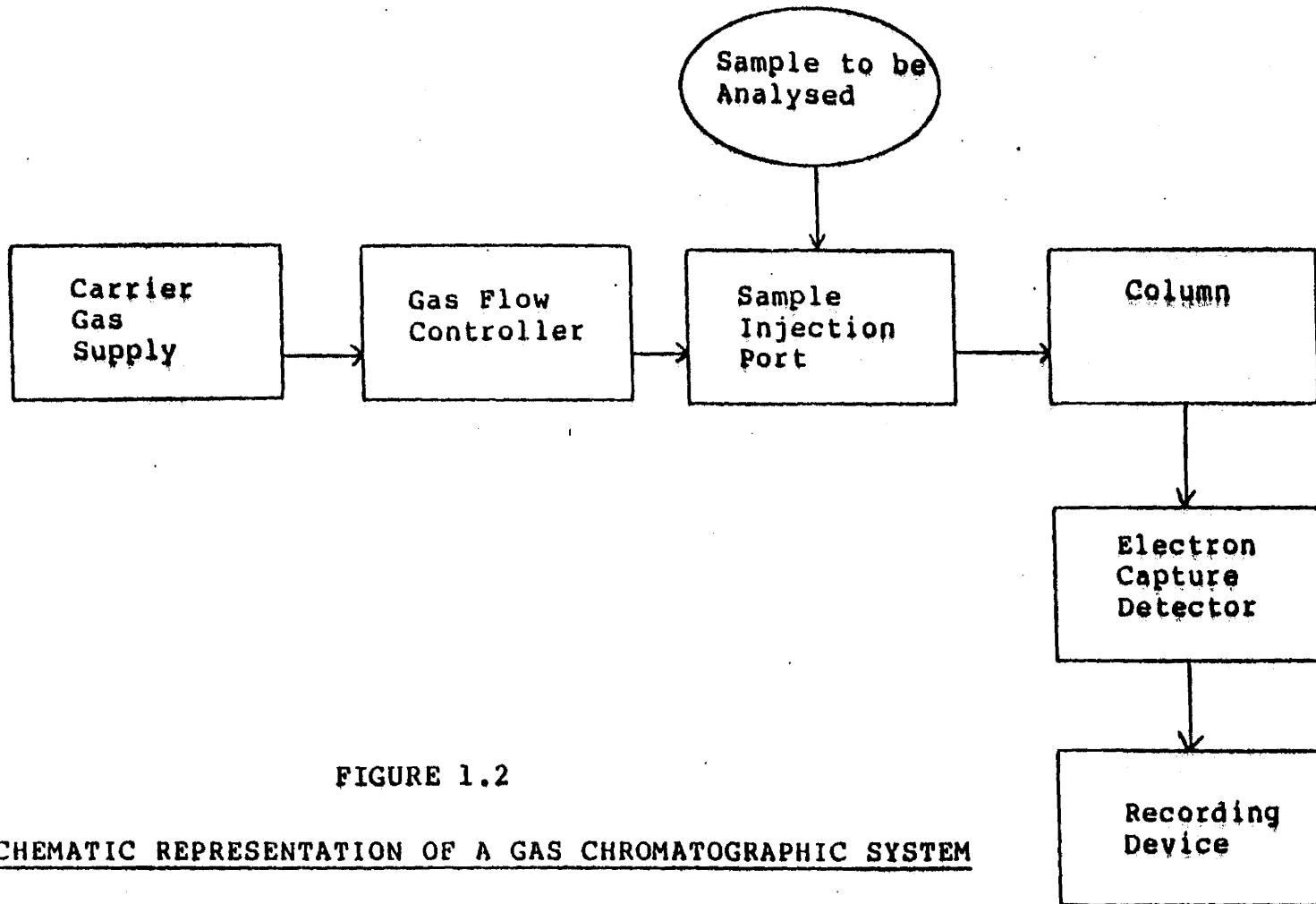
FIGURE 1.2

SCHEMATIC REPRESENTATION OF A GAS CHROMATOGRAPHIC SYSTEM

There are many different types of coatings used in columns, each with a different separating power. Many different detectors also exist, ranging from those that are highly selective, to types that respond to almost anything that passes through them. Therefore by using the appropriate column coating and detector, it is possible to separate and determine the concentration of any substance that can be vapourized.

There are two classes of columns that can be used, packed columns, and the newer capillary columns. Packed columns are usually glass tubes with approximately a 3 to 6 mm. inside diameter and filled with a specific packing material. (For the work discussed here this was an inert support material carrying the coating.) These columns have poorer resolving power than capillary columns. This is attributable to the relativly large inner diameter of the column which permits diffusion perpendicular to the flow direction, leading to peak broadening. Capillary columns usually have inside diameters of about 0.2 mm. with the active material coated on the inner wall of the column. These columns with their smaller diameters, effectivly suppress diffusion perpendicular to the flow direction, thereby producing narrower peaks. A disadvantage of capillary columns is that sensitivity problems occur when

observing minor peaks that are not well separated from major peaks (4). These sensitivity problems can result from the sampling rate used, cleanliness of the sample and the amount of sample that is eluted. To overcome these problems, capillary columns are overloaded but this produces fronting (i.e. a pre-peak skewing). Generally, observing trace elements with capillary columns requires fast scan speeds and a high sensitivity.

1.3   A Simple Physical Model For The GC System Response

Function.

Let us assume that a vapour is injected into a columnar section, with length much greater than the diameter, by the use of an inert carrier gas. If the transporting gas moves with velocity v then, in time Δt, any vapour in the volatile phase will move a distance $\Delta 1 = v\Delta t$ where for convenience the tube length L is divided into N such sections, ie. $L/\Delta 1 = N$. Given time to equilibrate, a fraction K of the vapour will condense as a longitudinally, immobile compound while the remaining fraction 1-K will remain in the gaseous phase. (The partition ratio involves $K' =$ concentration of solute in the stationary phase / concentration of solute in mobile phase i.e. $K' = K/K-1$. In the present case, the tube diameter and the wall material will determine the value for K.)

The selected model involves dividing the column into N equal intervals or "plates" in order to treat the problem in a discrete manner. Let us further assume that the equilibration time is small compared with the time $\Delta t = \Delta 1/v$ required to cross a plate with a velocity

equal to that of the carrier gas. Denoting $a = K$ and $b = 1-K$ the following sequence of events may be derived:

START **SHIFT** **EQUILIBRATE**

$$
\begin{array}{|c|c|c|c|}
\hline
b & 0 & 0 & 0 \\
\hline
a & 0 & 0 & 0 \\
\hline
\end{array}
\qquad
\begin{array}{|c|c|c|c|}
\hline
0 & b & 0 & 0 \\
\hline
a & 0 & 0 & 0 \\
\hline
\end{array}
\qquad
\begin{array}{|c|c|c|c|}
\hline
ab & b^2 & 0 & 0 \\
\hline
a^2 & ab & 0 & 0 \\
\hline
\end{array}
\qquad m = 0
$$

$$
\begin{array}{|c|c|c|c|}
\hline
ab & b^2 & 0 & 0 \\
\hline
a^2 & ab & 0 & 0 \\
\hline
\end{array}
\qquad
\begin{array}{|c|c|c|c|}
\hline
0 & ab & b^2 & 0 \\
\hline
a^2 & ab & 0 & 0 \\
\hline
\end{array}
\qquad
\begin{array}{|c|c|c|c|}
\hline
a^2b & 2ab^2 & b^3 & 0 \\
\hline
a^3 & 2a^2b & ab^2 & 0 \\
\hline
\end{array}
\qquad m = 1
$$

$$
\begin{array}{|c|c|c|c|}
\hline
a^2b & 2ab^2 & b^3 & 0 \\
\hline
a^3 & 2a^2b & ab^2 & 0 \\
\hline
\end{array}
\qquad
\begin{array}{|c|c|c|c|}
\hline
0 & a^2b & 2ab^2 & b^3 \\
\hline
a^3 & 2a^2b & ab^2 & 0 \\
\hline
\end{array}
\qquad
\begin{array}{|c|c|c|c|}
\hline
a^3b & 3a^2b^2 & 3ab^3 & b^4 \\
\hline
a^4 & 3a^3b & 3a^2b^2 & ab^3 \\
\hline
\end{array}
\qquad m = 2
$$

$$N = 1 \quad 2 \quad 3 \quad 4$$

For example at plate $N = 4$, at time interval $m = 2$ the equilibration is obtained by the following: the total at plate $N = 4$ is b^3 therefore since $a+b=1$ then $b^3(a+b) = b^3a + b^4$. The characteristic binomial pattern is clearly evidenced in the equilibrated distribution. Of interest for us is the form of the elution after passage through N such stages.

By inspection it can be seen that after N intervals $b^N = (1-K)^N$ of material is eluted. The next sample eluted has concentration $NK(1-K)^N$ or in general the output characterized by N and m is:

$$\frac{(N+m-1)!K^m(1-K)^N}{(N-1)!m!} \qquad \text{where } m = 0,1,2,\ldots\ldots\infty$$

1.0

This probability distribution function is normalized since

$$\sum_{m=0}^{\infty} \frac{(N+m-1)!K^m}{(N-1)!m!} = \frac{1}{(1-K)^N}$$

1.1

i.e. the binomial expansion of $1/(1-K)^N$.

Evaluation of equation 1.0 which describes the spectrum observed is most easily conducted using Sterling's approximation $K! \cong k^k e^{-k}\sqrt{2\pi k}$. In the present study the signal is given by $e^{S(m)}$ where

$$S(m) = m\,\ln[1+(N-1)/m] + (N-1)\,\ln[1+m/(N-1)] +$$
$$1/2\,\ln[(N+m-1)/(m(N-1))] - 1/2\,\ln[2\pi] + m\,\ln[K] +$$
$$N\,\ln[1-K].$$

1.2

Refer to figure 1.3 for elution profiles calculated from equation 1.0, for N=100, 1,000, and 10,000 with K = 0.1, 0.3, and 0.5. Comparison is most readily made if the spectra are plotted on a common scale therefore the data have been renormalized in the time domain in order to accomplish this.
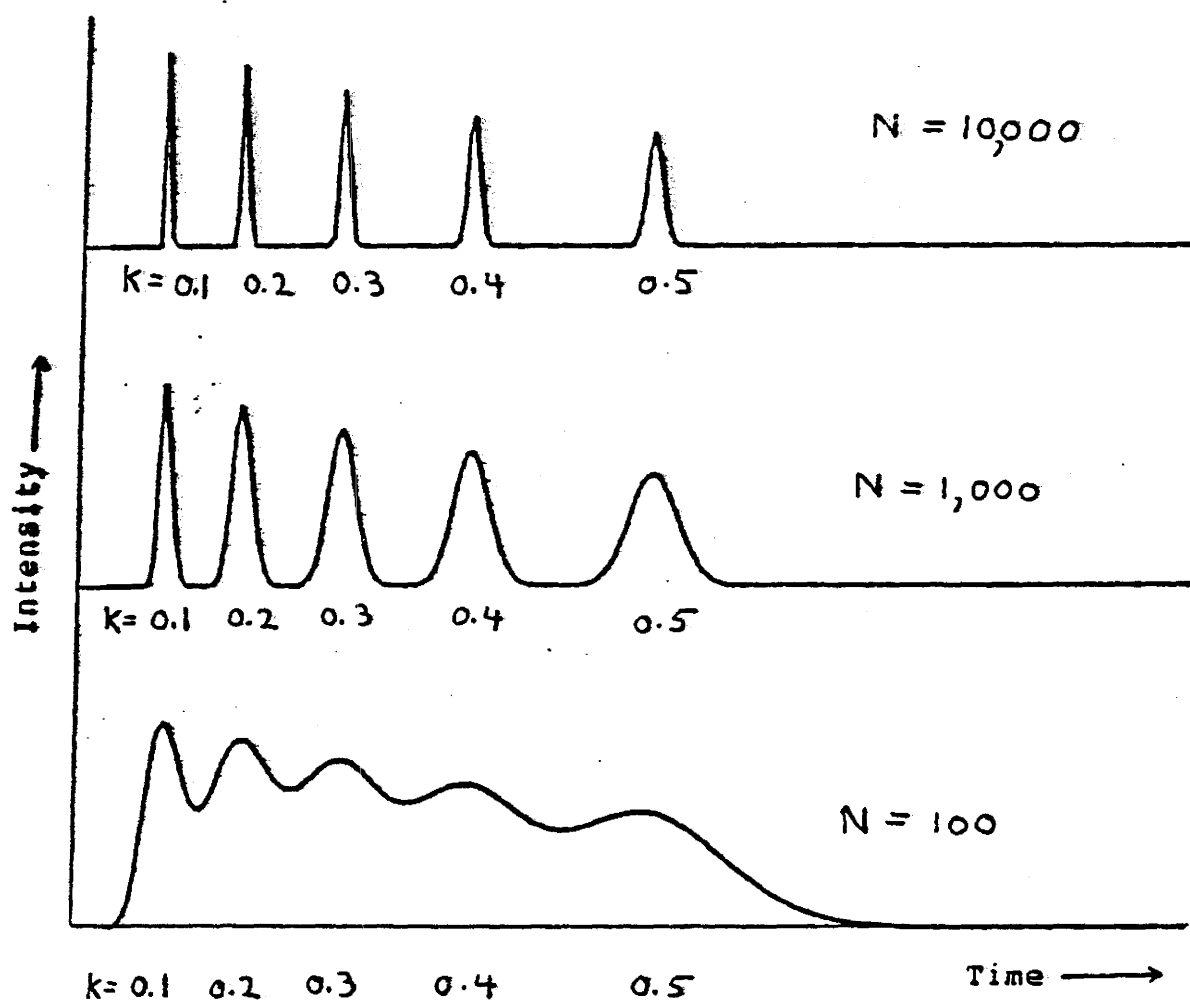
FIGURE 1.3

Elution profiles calculated from equation 1.0 using different values for N and K.

An examination of the output computed from the derived relationship shows a marked increase in line width with time and decreasing N and in addition very small skewing is displayed.

CHAPTER II

EXPERIMENTAL SYSTEM

2.1   System Overview

A schematic diagram of the complete experimental system is presented in figure 2.1. As can be seen, the system is primarily composed of three parts; the gas chromatograph, the electronics which interface the chromatograph to the microcomputer, and the computer system for the storage and subsequent processing of the data. A real-time data reduction system was not employed in order that greater flexibility in analysis techniques could be enjoyed. Ultimately a real-time system will be employed thereby reducing the size of the data fields and the analysis time. To preform a set of analyses, the operator is required to load samples to be tested onto the H.P. Automatic Injector and input into the microcomputer information that is relevent to each. This information is subsequently stored along with the corresponding spectrum for positive sample identification. Signals from the H.P. Intergrator/Plotter were used to start and stop the spectral data storage function by the microcomputer. A toggle switch provided the option of overriding these automatic control signals.

Gas Chromatograph System

Samples → Automatic Injector

Carrier Gas Supply → Gas Flow Control → Sample Injection Port → Column (in oven) → Electron Capture Detector

Oven Temperature Programmer → Column (in oven)

Linear ECD Control Unit → Electron Capture Detector

Frequency Signal

H.P. Integrator/Plotter

Computer System

Osborne Microcomputer (data storage on floppy diskettes)

GC/Computer interface electronics

Start/Stop Signals

Data transfer from diskettes to Nova Minicomputer at a different location. (Hard disc storage)

Magnetic Tape storage (after processing)

Interface electronics

Toggle Switch to disengage H.P Start/Stop signals

Manual Start/Stop signals

FIGURE 2.1

Schematic of the Experimental System used.

After sufficient data were acquired to fill the two floppy diskettes of the microcomputer transfer of the information from the diskettes to a Nova minicomputer (at another site) was performed. Then, with this minicomputer, all the spectra were processed and stored on magnetic tape for future reference.

2.2   The Detector

The three main types of detectors used for GC detection are the flame ionization detector (FID), the thermal conductivity detector (TCD), and the electron capture detector (ECD). The TCD and the ECD are concentration dependent, whereas the FID is mass flow dependent. In the experimental system used in this study, a frequency pulsed electron capture detector was used. These detectors are highly selective and respond only to those compounds capable of capturing electrons. The theory of operation of the ECD is as follows.

A 15 mCi Ni^{63} radioactive source is plated on the internal surface of one half of the EC cell. This source expells β particles (ie. fast electrons). As they move along their path, they collide with molecules of the argon/methane carrier gas and produce secondary electrons. If this liberated charge was collected it would

result in a current of $\sim$ 30 n amps. A reduced current and a sampling of the free electron population is obtained by applying a pulsed voltage between the anode and cathode. Normally a pulse frequency is selected that produces a "standing" current of $\sim$ 1 n amp. The value of this frequency is adjusted by a control circuit in such a way that the value selected for the standing current is maintained. Thus the presence of a compound with high affinity for electron attatchment in the detector chamber will reduce the electron population. The current, which is proportional to this population, can be retained at constancy by increasing the collection duty cycle or pulse frequency. Hence provided concentrations are selected which avoid saturation effects, the concentration of a compound is proportional to the pulse frequency.

If a substance A reacts with thermal electrons e^-, the yield of whatever the product P is will depend on the concentrations of A and e^-, the temperature, and the electron affinity of A.

$$A + e^- \longrightarrow P$$

Because the input and output of electrons into and from the detector chamber is constant (ie. a constant current), the remaining electrons must be removed by the compound (and

impurities in the gas, etc.). Therefore the concentration of the product is also constant.

$$K = [P]/[A][e^-]$$

and $$[A][e^-] = constant$$

The electron concentration in the cell increases with the interval between pulses, therefore it is inversely proportional to the pulse frequency f, and

$$[e^-] \quad \propto \quad 1/f$$

or $$[A] \quad \propto \quad f$$

Therefore a peak in the GC spectra can be measured as an increase in pulse frequency. This frequency signal is ideal for computer processing, and is the input signal to the GC/Computer interface circuitry. Refer to figure 2.2 for a block diagram of the ECD used.

The range of the detector was linear up to 10,000 pg of a highly electron capturing compound (Lindane), and had a minimum detectable level of 1 pg of Lindane. Hence a dynamic range of $> 10^4$ was attainable. With regard to the pulse frequency, detector saturation occured at approximately 0.6 MHZ.

Electron Capture
Cell

Frequency signal
to computer interface

Integ.
Op. Amp.

Exp.
Op. Amp.

Voltage to frequency
converter

Pulser

Divide by 4
(frequency divider)

+V

Balance

-V

Attenuation

Recorder output

+V

Zero

Current Source 1
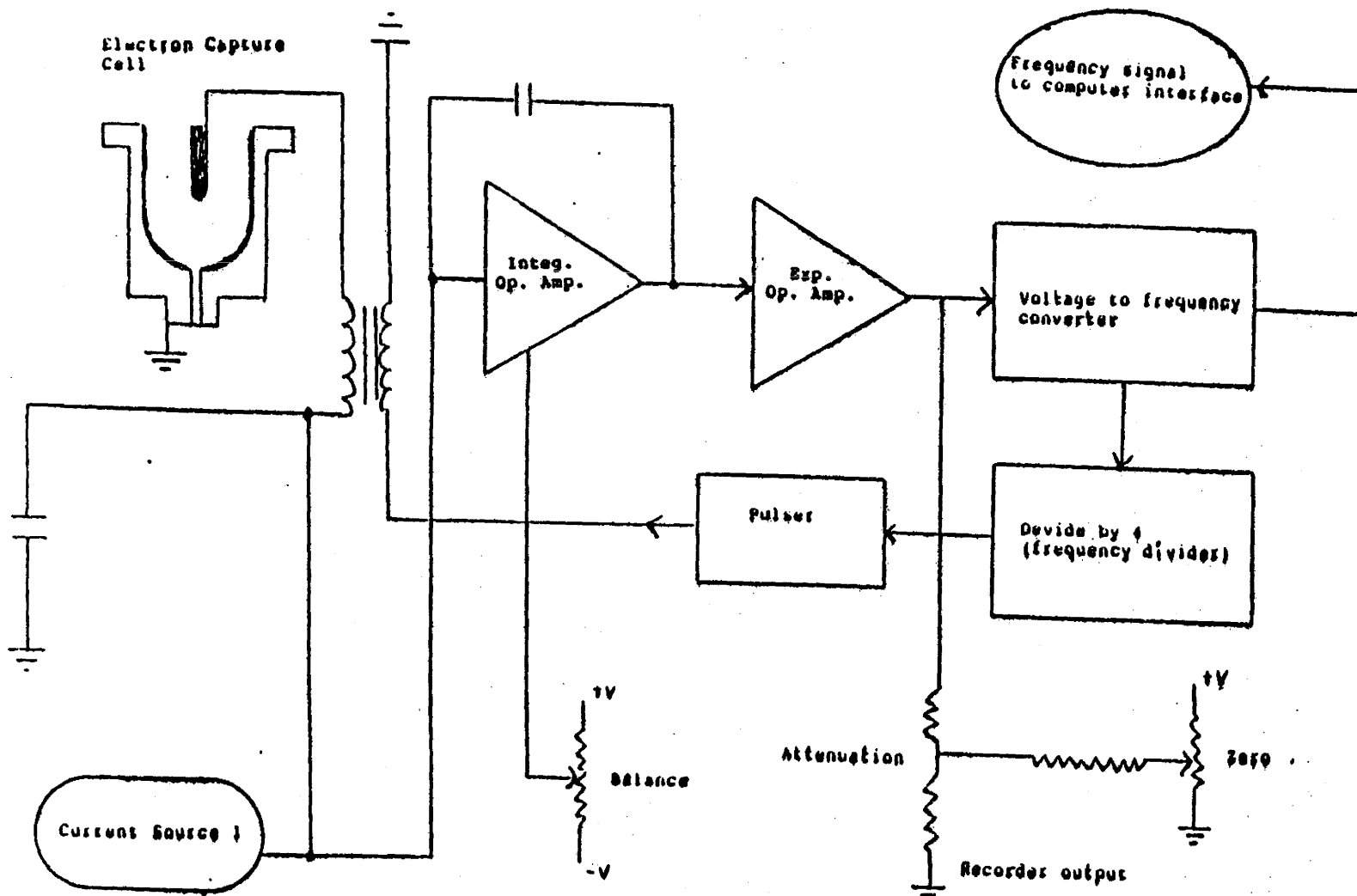
FIGURE 2.2   Block Diagram of the frequency pulsed
electron capture detector, and linear electron capture
control unit.

2.3 ·Column Temperature Programming

Under isothermal conditions, gas chromatographic analysis fails to provide adequate separation for samples that contain compounds with different values of K. As described earlier (in section 1.3), K is the fraction of the compound in the stationary phase. At low isothermal temperatures, highly retained components (usually but not always components with higher molecular weights) elute slowly thereby displaying long retention times with the associated peak broadening. At high isothermal temperatures the lighter components are not resolved from each other nor indeed from the solvent front.

An alternative approach to isothermal analysis is column temperature programming. Usually volatility depends upon molecular weight and temperature, therefore, if the temperature is varied monotonically progressively heavier substances will be eluted. Because each compound really only attains high mobility at a particular temperature, the value of K displayed at the time of elution is virtually constant. Thus, temperature programming effectivly produces a constant value of K for each compound in a sample. This in turn yields spectra which display translationally invariant response functions. The parametric relation between temperature and time and temperature and volatility

as a function of molecular weight results in an approximately linear mass scale when measured in time. Another important benefit of column temperature programming is that considerably lower starting temperatures can be used than can be tolerated under isothermal operation. This permits good separation to be obtained for quite volatile components. Figure 2.3 presents a comparison between isothermal and temperature programming upon GC spectra.

There are many different ways of temperature programming the column, with linear temperature programming and exponential programming being typical and common examples. For the experimental system in this study, the temperature was linearly programmed. Figure 2.4 illustrates the sequence used. For a temperature programmed column, the operating conditions of primary importance are the initial programmed temperature, heating rate, final temperature, and the carrier gas flow rate. In order to identify peaks in a GC spectra, these operating conditions must be accurately controlled and highly reproducible.

2.4   Gas Flow Rate Control

For programmed column temperature, the density and viscosity of the carrier gas must be kept at a constant value, as the temperature increases.
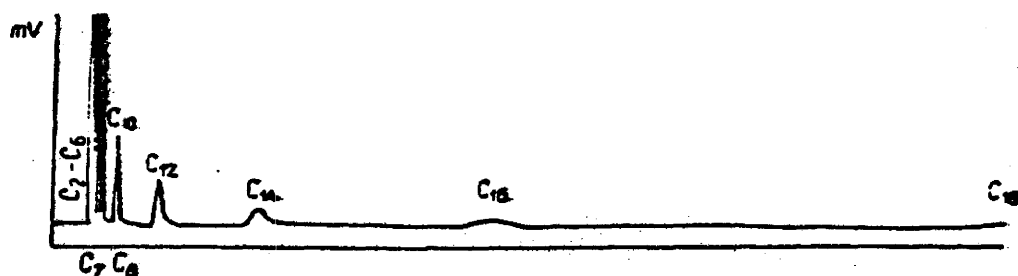
Isothermal temperature t= 175°C
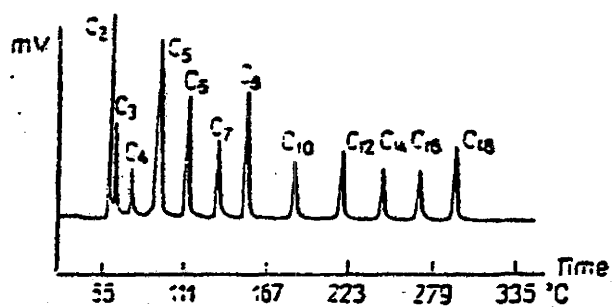
Programmed temperature t=65 ——→ 335°C

FIGURE 2.3

Comparison between isothermal and temperature programming of a GC spectra, (Fatty acid methyl ester).
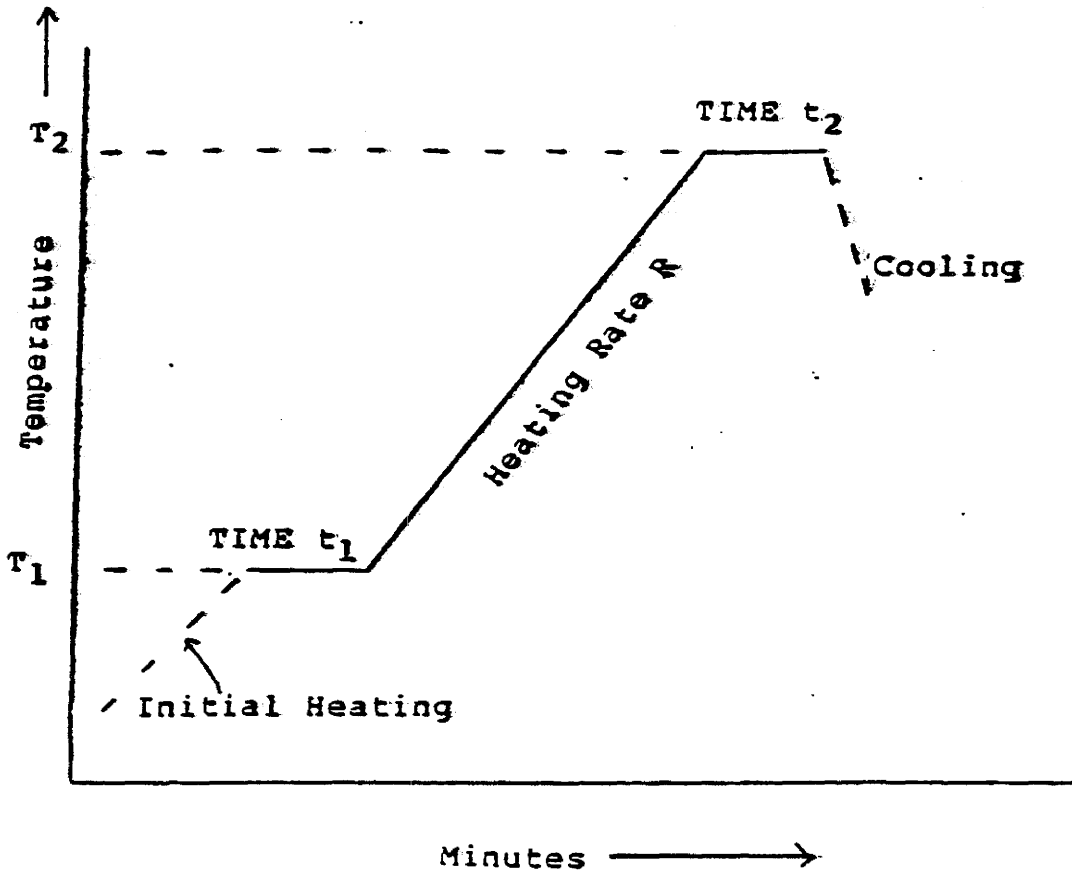
FIGURE 2.4 Linear temperature programming.

In this process, the oven is heated until it reaches the temperature T_1 (where it stays constant for a time t_1). Then, the oven temperature is increased at a constant rate R to a temperature T_2 (where it stays for a time t_2). After time t_2 the oven's door is opened and cooled down.

Changes in the flow rate can alter the peak shapes and their positions, and also affects the baseline. These effects can make the identification of the compounds difficult. Ambrose (5) states that when GC peaks are produced with different gas flow rates, the change in the observed peak areas is proportional to the reciprocal of the flow rate through the detector. The problem of maintaining a constant flow rate arises because of the fact that the flow resistance of the column changes with the physical state and analytical conditions of the column. Therefore to eliminate these effects due to column resistance, the flow rate of the carrier gas into the column must be adjusted in order to maintain a constant flow rate inside the column.

The experimental system in this study made use of a mechanical flow rate controller that provided a constant mass flow over a wide range of back pressure. Although this was adequate for the purpose at hand, greater accuracy and repeatability of GC information could be obtained by using an electronic flow controller (such as the H.P. model 19306 A/B/C).

CHAPTER III

DATA AQUISITION HARDWARE AND SOFTWARE

3.1   Consideration Of Possible Data Aquisition Methods

The stated problem of to devise a method of encoding and storing spectra obtained from a gas chromatograph which could be subsequently transfered to a minicomputer at another site. Development of the technique for spectral analysis was to be carried at the latter location.

Potential methods of data storage considered were cassette recorder, EPROMS, floppy discs, and a telecommunication link. Although quite a large amount of data can be stored on magnetic tape, noise is often a problem when large records are made, making the reliability of data integrity questionable.

Storing data in EPROMS or floppy discs is much more reliable than cassette recorders. The initial method employed involved storing data in EPROMS but subsequently an Osborne microcomputer was used and provided intelligence and the capability of floppy disc storage. The computer was used in order to provide guidance to the user as well as to supply continuous information about the present state of the recording process. It was also used because of the ease

26

with which data could be labelled, filed and stored on floppy discs.

3.2   Design Parameters

By examining the output from the gas chromatograph, it was observed that the modulation frequency spanned a range from 0.2 MHZ to 0.60 MHZ with the electron capture detector's (ECD) saturation occuring at 0.60 MHZ.   A HP integrater was used to integrate the modulated signal in order to to yield an output for the plotter. The spectral aquisition hardware used an analogous signal obtained by counting the modulated frequency.   By observing typical spectra, it was concluded that in order to obtain the desired resolution in the time domain, a sampling rate of about one point per second was required.

Since the chart speed of the integrator was 0.5 cm./min. , this sampling rate translates to 120 points/cm. along the x axis of the plot.   At the maximum frequency of 0.60 MHZ , the number of possible counts per sampling cycle was

$$0.60 MHZ \quad \times \quad 1.0 \ sec. \quad = \quad 600.0 \times 10^3$$

By prescaling with a $\div$ 10 , the number of counts per

cycle was 60,000. Therefore 16 bits per data point were required in order to handle the maximum frequency of the GC's signal, (since 2^{16} = 65,535).

In summary the essential design parameters required were 16 bits per data point and a sampling rate of once every second.

3.3   Initial Design

a.   Data Aquisition

The hardware consisted of two major parts, a timing circuit which controlled the pulse sequence necessary to store data on the EPROMS, and a circuit which contained the address lines and the data supplied to the EPROMS. Refer to figure 3.1 for a general schematic.   For more detail refer to Appendix 2.0 for the circuit diagrams.

The frequency signal from the GC was used to clock a binary counter, then after approximately a second the value at the output of the counter was transfered to the EPROMS. In this fashion the sampled spectrum was recorded.

The timing, (refer to circuit 2.0 in appendix 2.0) was generated through the use of a 3.8 MHZ crystal oscillator, which provided a stable clock.   By using a divider arrangement a nominal frequency of approximately 1 hertz was obtained.   This was then used to control the

sampling. In the control circuitry, monostables and D flip flops were used in order to provide correct pulse widths for such purposes as data set up times, data hold times, programming pulses and other timing requierments, (refer to figure 3.1). LEDs were used as indicators for the beginning and end of a GC run, and to show when the memory was exhausted.

To obtain greater detail along the x axis one would have to sample at a higher rate, (this is easily changed by using a smaller frequency divider). To gain greater resolution along the y axis, more bits per data point would have to be used (that's assuming that the sampling rate hasn't been changed from 1 Hz). According to the sampling theorem it can be possible to fully reconstruct a function from equally sampled intervals provided all signal frequencies are less or equal to half the sampling frequency.

b.   Data Transfer

This task was accomplished by enabling the appropriate EPROM to be read and then interrogating each location sequentially. The data were read into a Nova computer through the use of a short Basic program which continuously requested input. Each time a datum was read, the EPROM address register was incremented. The clock
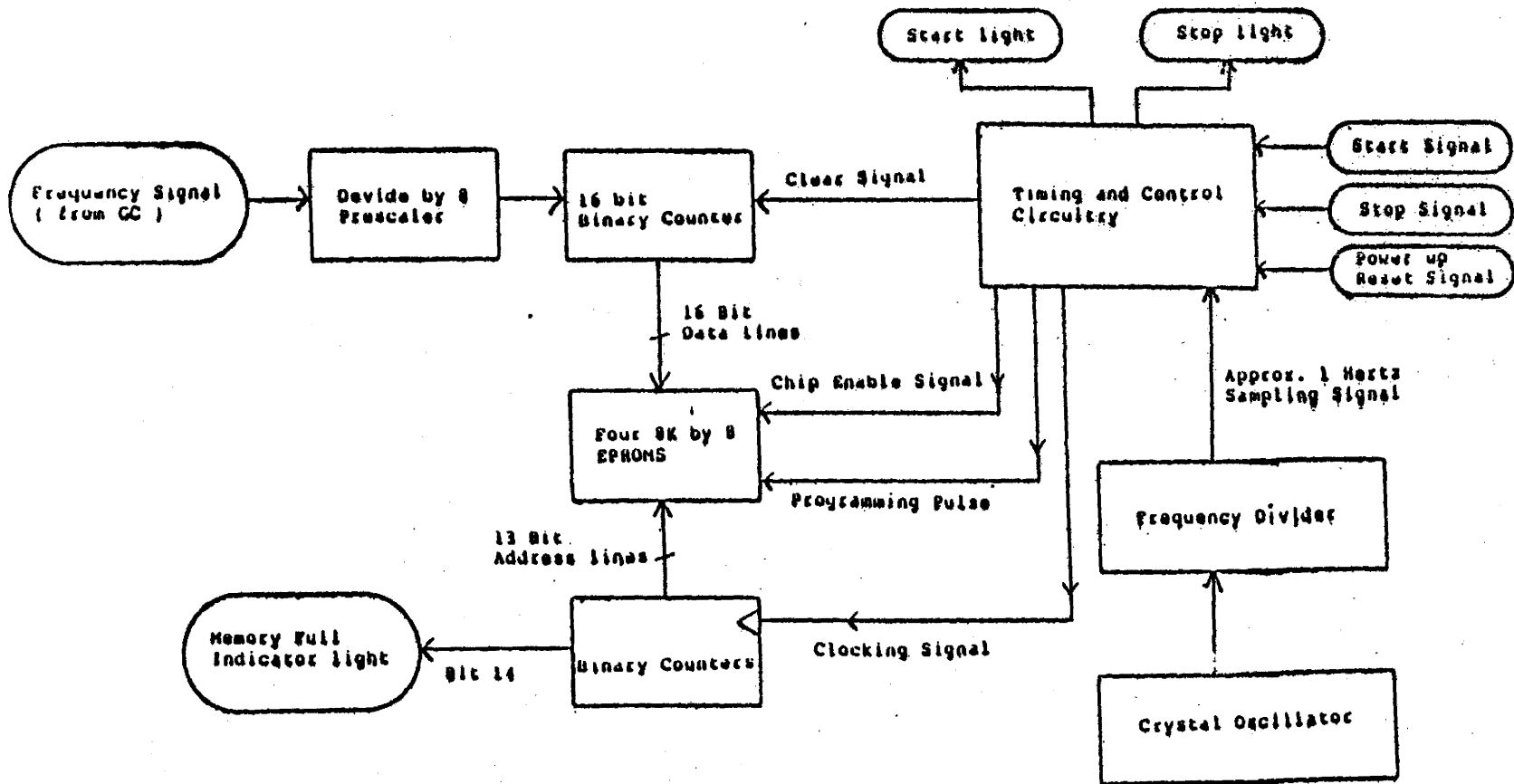
FIGURE 3.1

SPECTRAL STORAGE ON EPROMS

pulse to accomplish this was supplied by the computer. Figure 3.2 provides a general overview of the data transfer hardware. Circuit 3.0 in Appendix 2.0 is provided for greater detail.

3.4   Final Design

    a.   Hardware

The existing hardware for the EPROM spectral storage was expanded upon in order to allow an interface with the Osborne microcomputer. Once the data is input into the computer it can readily be stored on the floppy discs. Provisions to allow the computer to be disengaged from the GC was incorporated. Signals from the HP 3380A integrator were used to control the recording cycles duration.

The binary values from the 16-bit data counters were latched and then multiplexed 8 bits at a time into the microcomputer through it's parallel port to form a two-byte data word. Latching and multiplexing output signals were passed through the same input / output port both of which were computer controlled.

For a schematic of the hardware setup refer to figure 3.3 and for greater detail circuits 1,2,3 and 4 are provided in Appendix 2.0. A schematic of the Osborne's IEEE bus is provided in Appendix 2.0.
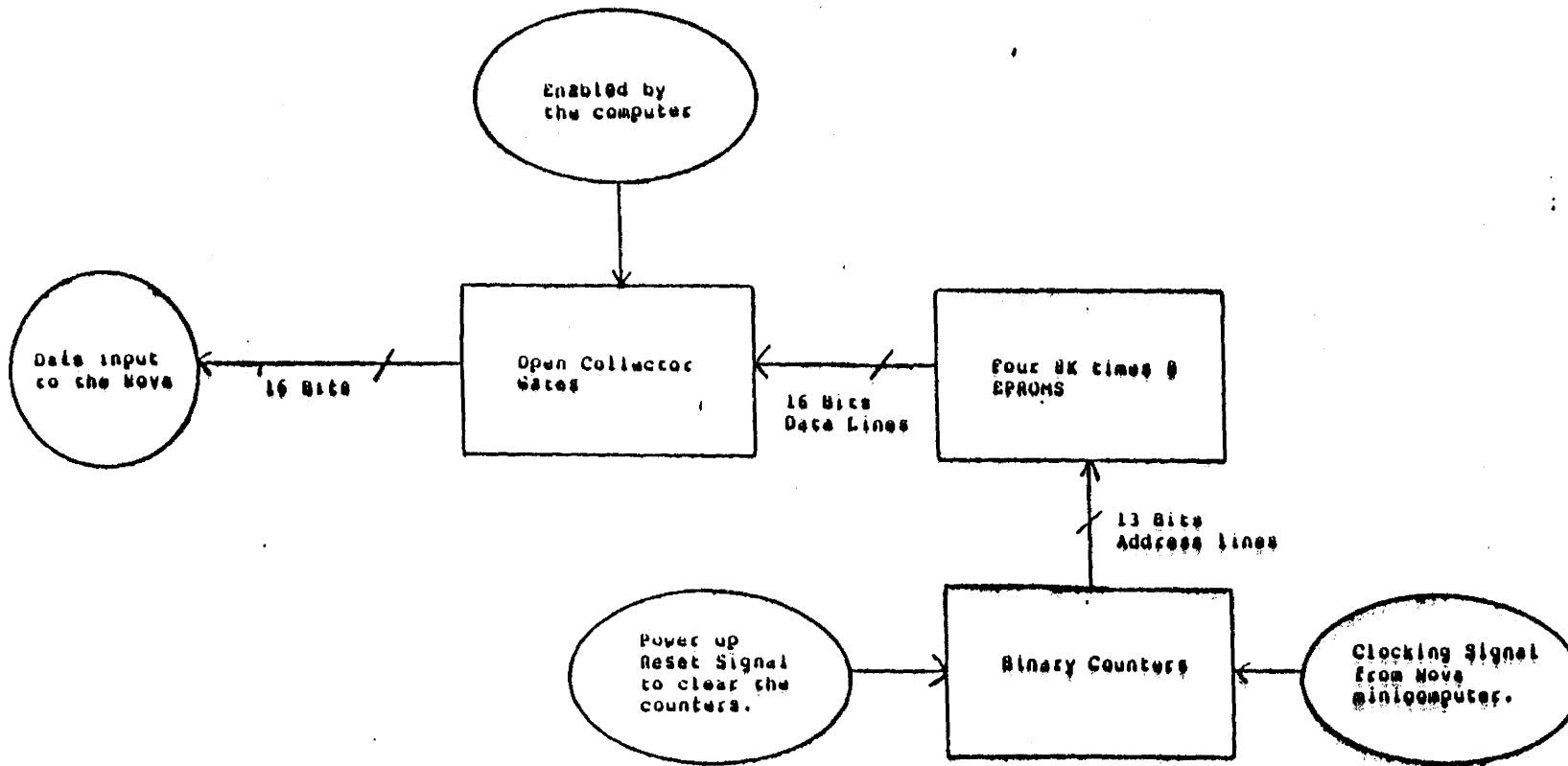
FIGURE 3.2

DATA TRANSFER FROM THE EPROMS TO THE NOVA MINICOMPUTER

b.   Software

The Osborne's IEEE port is controlled by an Intel 6821 programable interface adapter (PIA). This support chip is memory mapped into locations 2900H to 2903H in memory bank 2. Thus to control this parallel I/O port, bytes of information must be sent to and taken from these memory locations.

A flowchart of the machine code software that controls the actual transfer of data though the I/O port is provided in figure 3.4. A listing of this actual machine code routine is provided in the appendix.

During data acquisition the sampled points are stored into the computer's RAM at the memory locations A000H to BFFFH in bank 2. Bank 2 is a combination of the system ROM and memory mapped I/O ports, with the remainder of the 64K address space "mimicking" bank number 1. Once the run is completed, these memory locations are read onto floppy disc as a random access file.

This method was chosen as opposed to storing data point by point because of the excessive overhead in CPU time. Also since faster GC columns were anticipated in the future such an approach was deemed necessary. Because A000H to BFFFH encompases 8191_{10} locations, and since the sampling rate is approximately 1 second, the maximum length any one
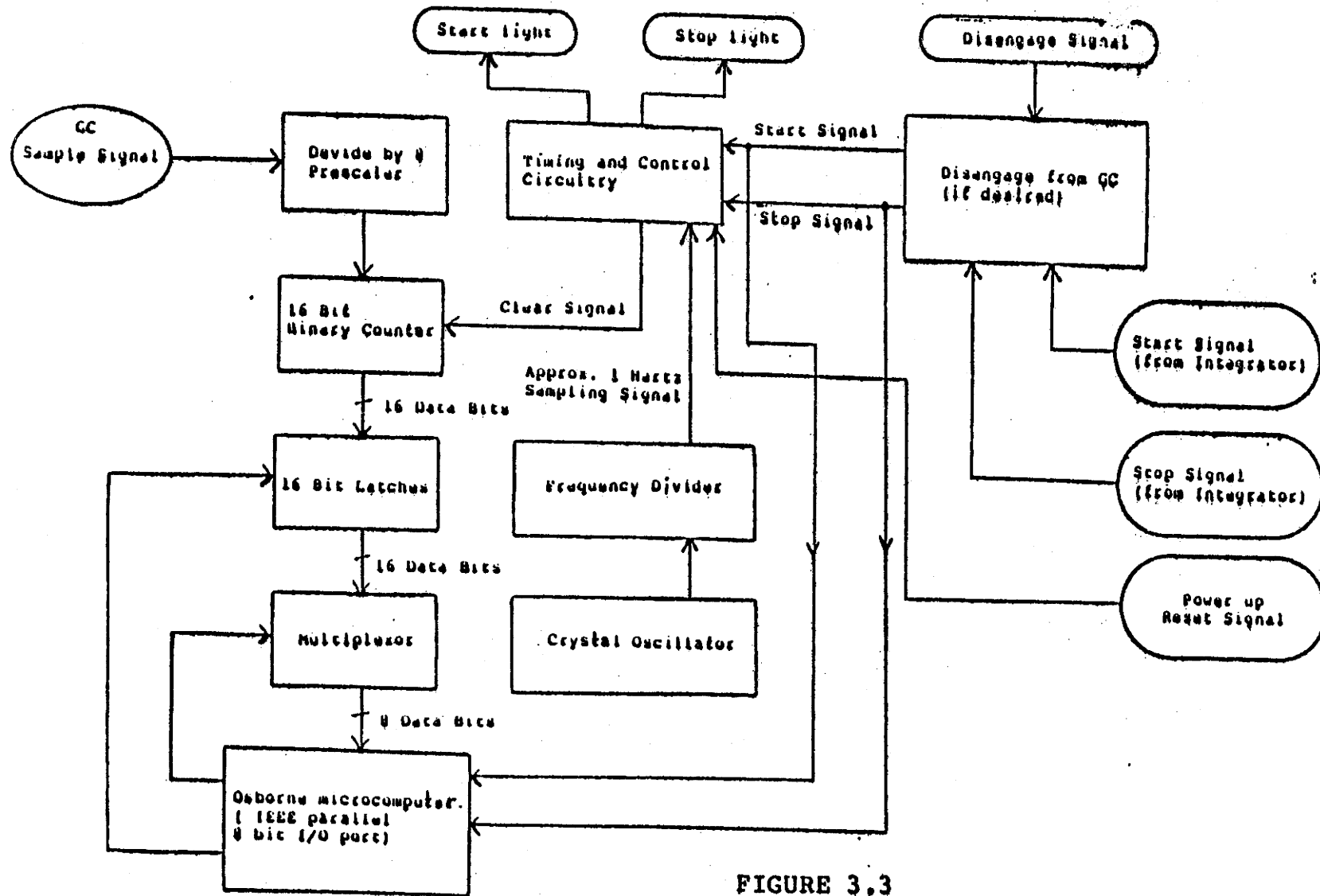
FIGURE 3.3

INTERFACE HARDWARE BETWEEN THE GAS CHROMATOGRAPH AND THE

OSBORNE MICROCOMPUTER

Call routine to remove disk write protection.

Initialization

Input number of runs desired and information about each run.

Store this on disc.

Is Disc full ? — yes

No

Call main I/O routine

Is the run over? — No

Yes

Store data on disc

Is disc full? — yes

No

Are all the runs finished? — No

Is there an error due to no more disc space? — No — Prompt user and stop.

yes

Are both discs full? — Yes — Prompt user and stop.

No

Switch to 2nd disc

Did an error occur during the storage of run information? — No

Yes

FIGURE 3.4

FLOWCHART OF THE PRINCIPLE CONTROLLING PROGRAM

35

run could occupy was limited to 68 minutes.

The floppy discs that were used were 5 1/4 inch single density, and have a storage capacity of 98 Kilobytes. Since each data point is stored as a single precision number which occupies four bytes, one disc can store approximately 6.8 hrs. of data (assuming a 1 second sampling rate). The system was set up to store data on both discs thereby allowing approximately 13.6 hours of data storage.

The entire data collection process was controlled through a program written in Basic. Refer to figure 3.5 which, through the use of a flowchart, illustrates the logic of this program. The CP/M operating system monitors the state of the disc drive's door. Opening and closing a door normally produces a flag which cannot be manually reset. To overcome this problem, the main Basic program calls a machine code routine which in turn calls a BDOS (Basic Disk Operating System) function 0Dh which clears the disk subsystem. This then eliminates the write protection on the discs.

A device is referred to as a "module" of the Osborne microcomputer. A "physical" device is one that exists physically on the computer, such as an I/O port. A "logical" device is a fictional device which the CP/M operating system uses for various functions. For example,

ASSEMBLY LANGUAGE SUBROUTINE   (STARTING AT THE RAM LOCATION 9000 hex)

```
┌─────────────────────┐        ┌──────────────────────────┐     ┌──────────────────────────────┐
│ Switch from         │        │ Set PB1 and PB0 HI,      │     │ Get LSB'S, then compliment   │
│ memory bank 1 to 2  │        │ therefore                │     │ it and store in memory at    │
│                     │        │ NDAC and NFRD are inputs │     │ the address given by         │
└─────────────────────┘        │ and register A are inputs│     │ the register pair BC.        │
                               └──────────────────────────┘     └──────────────────────────────┘

┌─────────────────────┐                  │                                    │
│ Assign the data     │                  ▼                      ┌──────────────────────────────┐
│ direction of the bus│              ╱───────╲                  │ Increment the register pair  │
│ by controlling 6821 │             ╱ Is there ╲                │ BC (ie. increment the        │
│ PIA IC through      │            ╱  a start    ╲   No         │ address)                     │
│ memory mapped       │            ╲  signal?    ╱──────►       └──────────────────────────────┘
│ locations 2900h to  │            ╲ (ie. Is    ╱                           │
│ 2903h               │             ╲ NDAC low?)╱                ┌──────────────────────────────┐
└─────────────────────┘              ╲───────╱                   │ Store the MSB'S into memory  │
                                        │ Yes                    └──────────────────────────────┘
┌─────────────────────┐                 ▼                                   │
│ Assign register A   │             ╱───────╲                   ┌──────────────────────────────┐
│ to the peripheral   │            ╱ Is the   ╲                 │ Store address register pair  │
│ register mode       │           ╱  stop      ╲   Yes          │ BC, B into location C000h,   │
└─────────────────────┘           ╲ signal     ╱──────►         │ and C into location C001h,   │
                                  ╲ pushed?    ╱                 │ for future reference.        │
┌─────────────────────┐            ╲(ie. Is   ╱                  └──────────────────────────────┘
│ Set register B      │             ╲NFRD low?╱                             │
│ data direction bits.│              ╲───────╱                   ┌──────────────────────────────┐
└─────────────────────┘                 │ No                    │ Switch memory bank 2 to 1    │
                                        ▼                        │ and go back into Basic       │
                          ┌──────────────────────────┐           │ program.                     │
┌─────────────────────┐   │ Get ready by sending word│           └──────────────────────────────┘
│ Assign register B   │   │ select and latch signals │
│ into the peripheral │   │ to multiplexor in order  │           ┌──────────────────────────────┐
│ register mode.      │   │ to read LSB'S.           │           │ Set the variable I% (in      │
└─────────────────────┘   └──────────────────────────┘           │ Basic) equal to 5,           │
                          ┌──────────────────────────┐           └──────────────────────────────┘
                          │ Load address register    │           ┌──────────────────────────────┐
                          │ pair with a value, and   │           │ Switch memory bank 2 to 1,   │
                          │ then increment it.       │           │ and then return to Basic.    │
                          └──────────────────────────┘           └──────────────────────────────┘
```

FIGURE 3.5

37

all printing goes to the logical LST device before it can go to a physical device. Data transfer from the Osborne microcomputer to the Nova minicomputer occured through the series RS232 I/O port. Since the physical CRT device was the RS232 serial port, it was necessary that the Osborne had it's logical LST device associated with the physical CRT device in it's IOBYTE before data transfer could occur. The IOBYTE location (03H) contains the association between the logical and physical devices presently in use. The baud rates of the receiver and transmitter also had to be the same, (a baud rate of 1200 was used).

CHAPTER IV

SAMPLE PREPARATION

4.1   Overview

Bacterial cell walls and mammalian cell membranes contain proteins, fatty acids, and sugars.   Mammalian cells contain fatty acids with only even numbers of carbon atoms, whereas bacterial fatty acids contain both even and odd numbers of carbon atoms.

In a sample generated from a bacterial preparation there are many different possible carboxyilic acid esters ("fatty acids").   Thus, the set of standards used for calibration consisted of a homologous series straight chain from C_7 (ie. with seven carbon atoms) to C_{24}. However there also exist cyclic structures and structures with different functional groups attached to them.   For analysis the acids are derivatized.   All these different structures exhibit their own unique chromatographic properties and hence their own unique elution times. It has been proposed that each bacteria has a specific spectrum when analyzed both in vitro and in vivo.   Furthermore, it appears that spectral differences between strains also exist.

The preparation of a sample for in vitro identification consists of several stages as illustrated in figure 4.1.
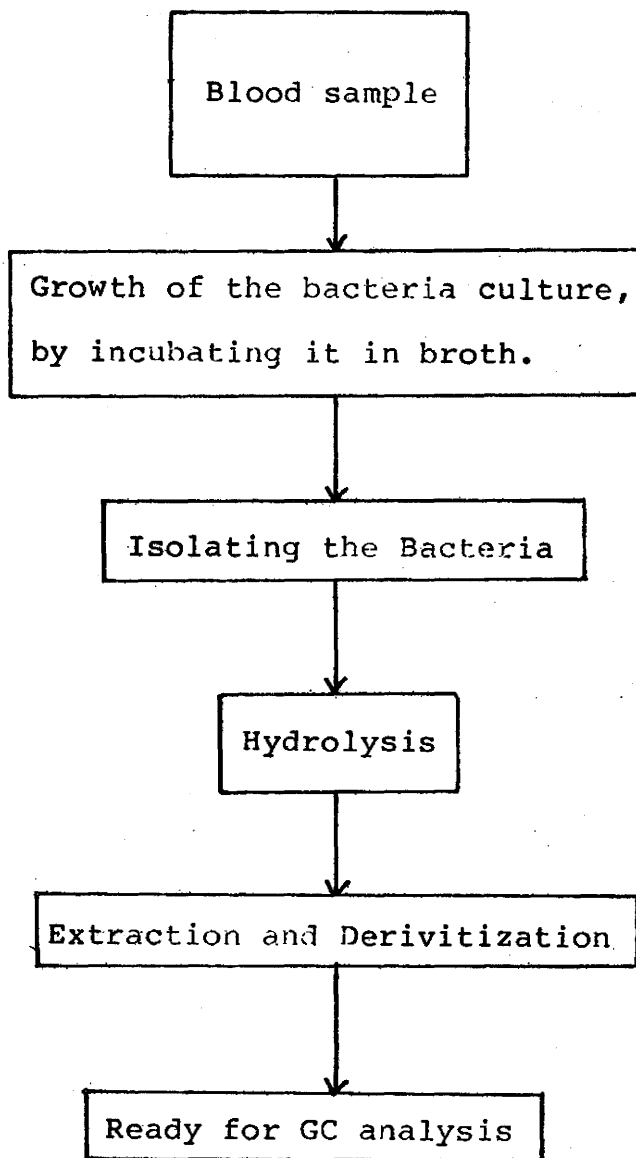
FIGURE 4.1

Preparation of a sample for in vitro identification.

This procedure may be represented as follows. The cell walls of bacteria contain alkyl acid esters bound to glycerol

$$R - C \overset{\displaystyle O}{\underset{\displaystyle O - R_G}{<}}$$
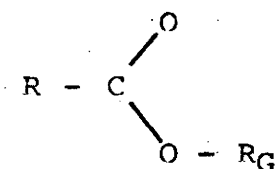
where R_G is glycerol.

These are then hydrolysed by the addition of KOH to give carboxylic acid and glycerol, i.e

$$R - C \overset{\displaystyle O}{\underset{\displaystyle O - K^+}{<}} \qquad\qquad + \qquad R_G$$
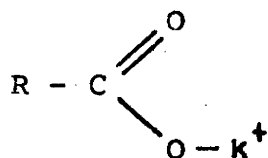
Carboxylic Acid   Glycerol.

The carboxylic acid is then extracted and derivitized to give a carboxylic acid ester.

$$R - C \begin{array}{c} \diagup\!\!\diagup O \\ \diagdown O - R_i \end{array}$$

where i = 1 to 3 and

R_1 = CH_3

R_2 = CH_2Cl_3

R_3 = PFB = $CH_2 -$ (pentafluorobenzyl ring with F substituents)
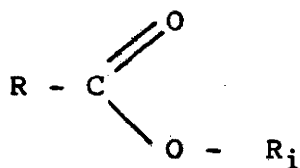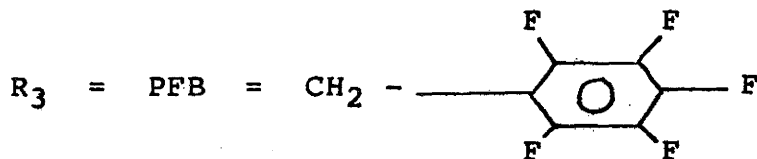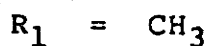
refer to ref.   (6)

These esters can be injected into the GC for analysis.

Flame ionization detection (FID) has been used for most in vitro studies (7), however ultimately it will be most beneficial to carry out such identifications directly from the body fluids (8,3). In this situation some components of primary interest might appear as trace elements hence requiring the sensitivity of an ECD to detect them. The minimum detectable quantity (MDQ) is defined as

$$MDQ = 2N/S$$

where N = noise level

S = detector sensitivity

Typical MDQ values for FID are 10^{-10} to 10^{-11} grams, for ECD they range from 10^{-9} to 10^{-12} grams. A prerequisite for electron capture detection is that the sample must be derivitized to form electrophilic derivitives.

4.2 Chemical Extraction and Derivitization Procedures

Carboxyilic acids in the aqueous phase represented as
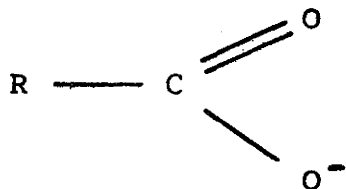
Ionized Carboxyilic Acid

(ie. Carboxyilic Acid in the aqueous phase)

can be extracted through many techniques. One possible method for extraction and derivitization is as follows. The aqueous phase is acidified with sulphuric acid (H_2SO_4) and then the unionized acids are partitioned between the aqueous

phase and chloroform. Two layers are present after this process: the chloroform layer containing the carboxyilic acids, and the aqueous layer. Following the separation of these, the organic layer is evaporated. The carboxyilic acid in the organic layer and represented by

$$R \longrightarrow C \overset{\displaystyle O}{\underset{\displaystyle OH}{\Big\langle}}$$
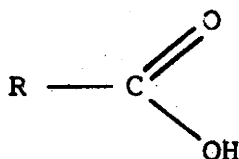
Carboxylic Acid

is then ready to be derivitized by a number of techniques. The derivitization process suggested by Brooks (1) uses trichloroethanol (CCl_3-CH_2OH) as the derivatizing agent and heptafluorabutynic anhydride (HFBA) as a catalyst.

$$R \longrightarrow C \overset{\displaystyle O}{\underset{\displaystyle OH}{\Big\langle}} + CCl_3CH_2OH + HFBA \longrightarrow R \longrightarrow C \overset{\displaystyle O}{\underset{\displaystyle O-CH_2-CCl_3}{\Big\langle}}$$
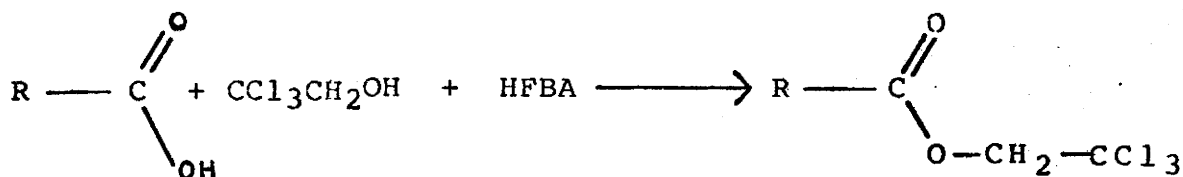
Subsequently the extraxts are cocentrated to a small volume (100 μL) and washed sequentially with aqueous, acidic, and basic solutions. Pentaflourobenzylbromide (PFBBr) is then added to produce a carboxyilic acid PFB ester.

Brooks (1) initially developed this method of extraction and derivitization of carboxyilic acids, preparotory to GC/ECD analysis and used this for microbiological identification. However the method is technically complex.

A newer method of extraction and derivitization as Rosenfeld (9) suggests seemed more appropriate and was used in this study. Rosenfeld's method is generally as follows. Firstly, PFBBr is impregnated onto the surface of a XAD_2 resin. This is a marcoreticular polystyrene divinylhenzene co-polymer available as 0.5 mm. beads. The beads are insoluble in acid, base and organic solvents. Then, an aqueous solution of ionized carboxylic acid is mixed with the impregnated beads. The acids become attached to the surface of the beads since they are lipophilic and a reaction between ionized carboxylic acid and alkylating agents occur.
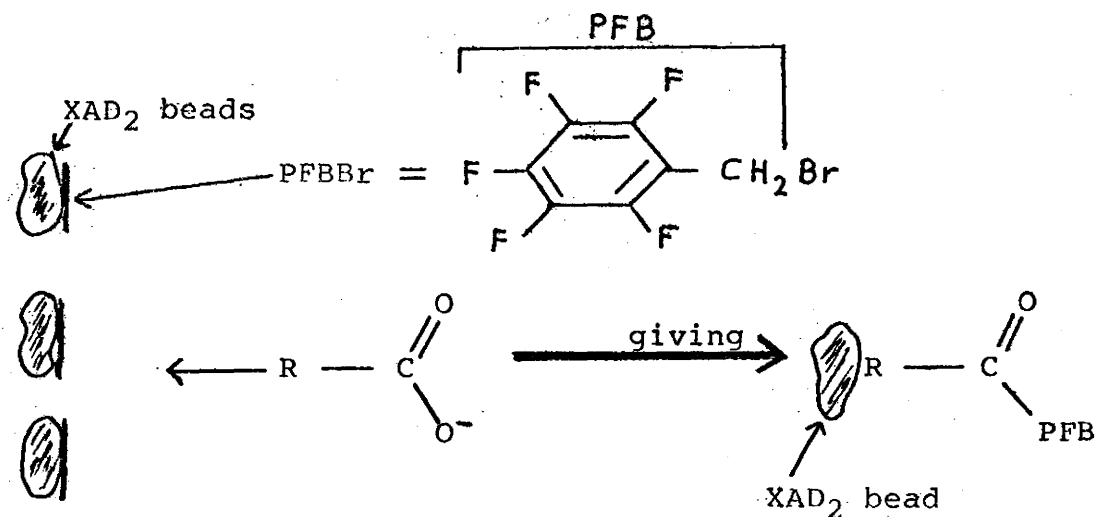
The beads, along with the carboxylic acid PFB ester are easily separated from the aqueous solution by filtration. Carboxylic acid PFB esters are then eluted from the bead with organic solvent. The ether is evaporated and the esters dissolved in toluene. This is then ready for gas chromatograph analysis. Because the entire process is based on filtration rather than solvent extraction it is more suited to automation.

CHAPTER V

DIGITAL SIGNAL PROCESSING

5.1 Peak Identification Parameters

The main parameters available to a gas chromatographer for peak identification are retention time, peak height and area. Retention data are usually used for structure identification. The peak's area or height is then used for quantitation. Under some circumstances it may be more appropriate to use peak heights rather than areas. In general, the choice of measurement technique is dictated by the affects against which one needs to preserve consistency in analyses.

If the carrier gas flow variability affects the relationship between the response area and the amount of component analysed, or if the baseline is fluctuating or the detector signal is noisy, then the peak height is generally a more reliable measurement. If peak shapes vary due to limitations in temperature programming or changing column characteristics, then peak area is a better measurement of concentration. In a GC system, if the gas carrier flow is not constant then the peak areas will change accordingly. Since the "raw" GC spectra had varying peak shapes, and since this system had a constant carrier flow, it became

evident that calculation of sample composition from peak areas would produce less variance than the use of peak heights.

Before spectral analysis could be undertaken (by obtaining retention and area data) it was necessary to remove the unwanted "noise" and distortion components which generally plague GC spectra. Noise may originate in the physical analysis apparatus and in the electronic measurement process. GC distortion is most commonly in the form of peak skewing and baseline drift (discussed in greater detail in the following sections).

Baseline is the condition when no sample fraction flows past the detector. A GC signal is at baseline when it is flat or slowly varying. Typically baseline drift is encountered during temperature programming and column bleed.

5.2 Asymmetrical (Skewed) Gas Chromatographic Peaks

a. Sources of Peak Skewing

When dealing with chromatographic peaks, it is almost always a mistake to assume a perfectly Gaussian shape for the peak shape. Experimentally observed GC peaks are produced from the combination of various components convolved together. The primary components are; the injection, the physical process of gas chromotography, and a skewing component refer to figure 5.1. If the sample injection had a relatively long duration this would widen all signals. If it were large enough, then other effects would be of little importance. However the injection time is very short in duration compared with the width of the near gaussian associated with transport through the column. Also since columns usually have quite a large number of theoretical plates, N, the near gaussian distribution calculated by equation 1.0 can be represented by a Gaussian. In general, a typical GC peak can be well approximated by a convolution of a Gaussian with an exponential (10). This type of GC peak is refered to as an exponentially modified Gaussian peak (EMGP). Skewing is not due to the chromatographic process itself but is a result of exponential modifiers acting on the gaussian.

In a chromatographic system an exponential modifier

Injection   "Near Gaussian"   Skewing Component

Observed GC Peak

where * represents a convolution
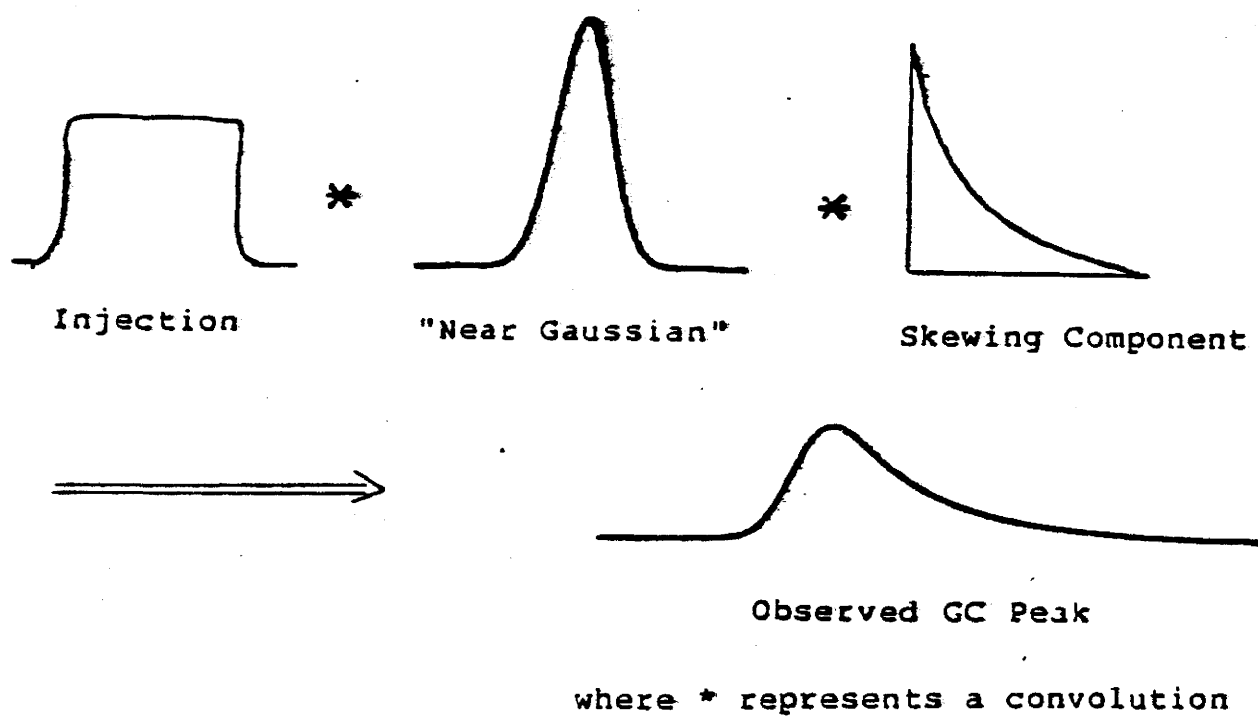
FIGURE 5.1

Principle components of an observed GC signal.

is any element that tends to delay a signal representing an elution profile and hence producing a skewed peak. There can be more than one modifier and they can occur anywhere in the GC system, (this is generally the case). Sternberg (11) has shown that if more than one exponential modifier is present in a system, then the order in which a signal is modified is not important in determining the peak shape.

Usually, an exponential modifier can be found in regions where detection occurs. For such a modifier the response function can be represented by $F/Ve^{-Ft/V}$ where $t>0$. Some exponential modifiers occur as undesirable effects in the column, for example, active sites of adsorbtion, channeling in the column packing, and slow lateral diffusion causing skewed dispersions. By using a capillary column, lateral dispersion and hence skewness can be significantly decreased (because of the small diameter of the columns) in certain instances.

Sternberg (11) has shown that various dead volume contributions produce peak skewing. In other words, dead volumes in the injector, detector or connecting tubing are also exponential modifiers. Other exponential modifiers may occur as a result of a slow response in either the detector, the amplifier, the recorder, or other electronic delays (12). Each one of these exponential modifiers has a certain

time constant $\Upsilon = 1/\alpha$ associated with it (12). The larger the time constant Υ ,the greater the skewness in the peak.

Peak asymmetry can also be produced by changes in operating conditions such as solvent viscosity, flowrate, or column temperature. By increasing the flowrate or solvent viscosity, or by decreasing the column temperature, the peak's skewness can be increased.

Imperfections in the GC apparatus may also cause skewing. For example, when a sample is injected onto the column packing, it partitions between stationary and mobile phases. The free volume in the column must be large enough to contain the vapourized part, because if any of it expands against the gas flow into a cooler region, it may condense and then re - evaporate slowly while most of the sample has already started moving down the column. The same kind of skewing effect can occur if there is adsorption on any of the components of the system (eg. injector port, column, or detector).

b. The Effects Of Peak Skewing On Information Recovery

One of the most important consequences of peak skewing ("fronting" or "tailing") is that the resolution is decreased. In addition peak maxima are shifted as may be seen in figure 5.2. This can result in an erroneous estimation of retention time and hence compound. Since the total area of the signal is invarient hence peak amplitude decreases as the degree of skewing increases.

Generally, there is more than one exponential modifier in a GC system, each characterized by a time constant Υ_i. The resultant effect of each modifier provides the input for the next modifier stage (13). By the central limit theorem, multiple convolutions of exponentials lead back to a symmetric gaussian. Therefore the peak becomes symmetric but displaced as may be seen in figure 5.3.

The mean of a peak is displaced by Υ and if several exponential modifiers are present the mean will be displaced by $\sum \Upsilon_i$. Clearly as the number of modifiers present increases, the greater will be the shift in the position of peak maxima. Consequently the retention time of the eluted components will be less precisely defined.

Figure 5.4 illustrates how distortion masks detail
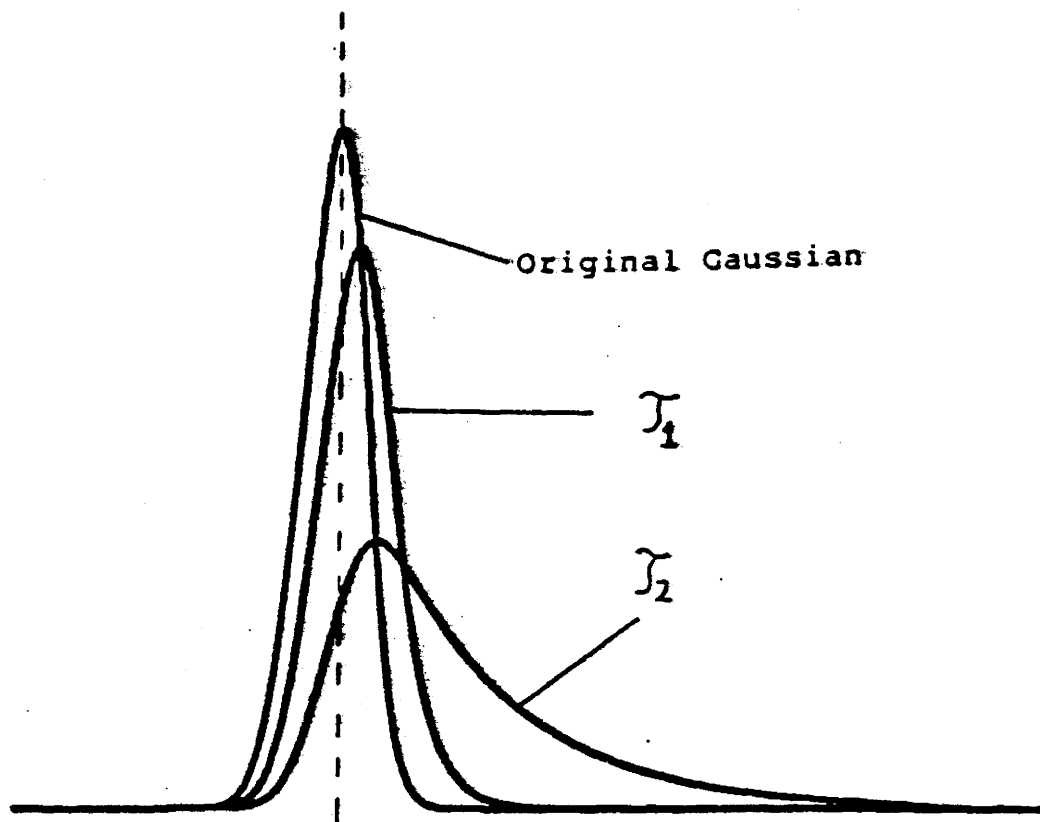
FIGURE 5.2

Peak skewness increases with increasing decay constants $\Im$ where $\Im = 1/\alpha$, (here $\Im_2 = 2 \times \Im_1 > \Im_1$). The greater the skew, the greater the centroid displacement from that of the original gaussian. Also, note that the area of the EMGP is always equal to that of the original gaussian.
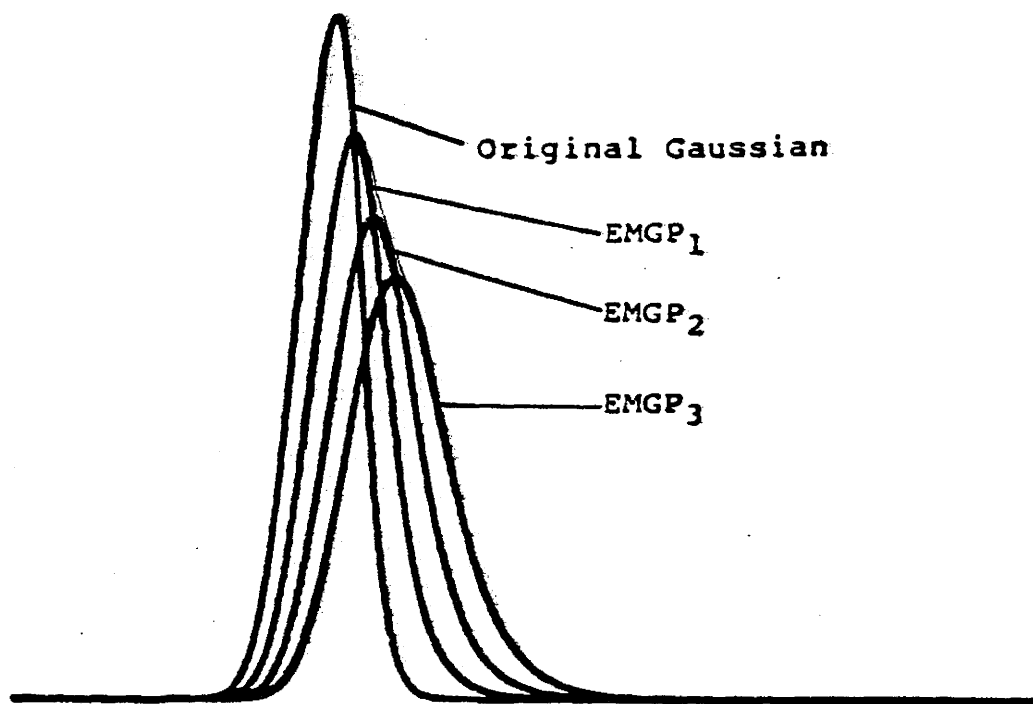
FIGURE 5.3

Cascading effect of the exponential modifiers. Each EMG peak serves as an input to the next exponential modifier stage. Shown here is the original gaussian, the first EMGP, second, and third. $EMGP_1$ was obtained after the gaussian was input into the first exponential modifier, $EMGP_2$ after $EMGP_1$ was input into the second exponential modifier, and so on.
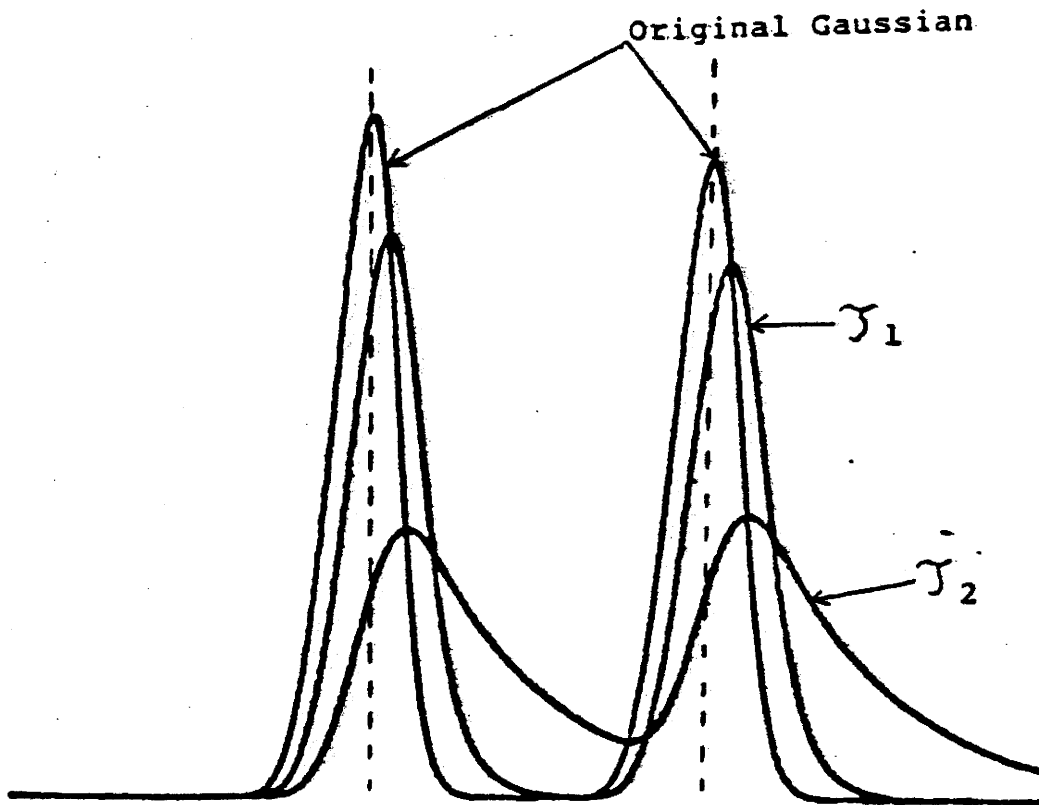
FIGURE 5.4

As the decay constant $\mathcal{T} = 1/\alpha$ of the exponential distorting function increases, the resolution decreases. (here $\mathcal{T}_2 > \mathcal{T}_1$). The two original Gaussians in this figure have different full width half maxima but the same areas.

when skewing is introduced for the case of two well resolved peaks. Because this form of distortion is asymmetric the loss of information is greater when a small peak arises after a large peak than for the converse case (12). This is illustrated in figure 5.5. In summary exponential modifiers lead to a lag in the arrival time of the signal and to resolution degradation.

By quantifying the degree of skew present in GC spectra, the preformance of the column can be deduced. Such information may be used to indicate the condition of the column and when replacement should be made.

Original Spectrum

Original Spectrum

Skewed Spectrum

Skewed Spectrum

Part A

Part B

(Minor peak precceds major one)

(Major peak preceeds minor one)
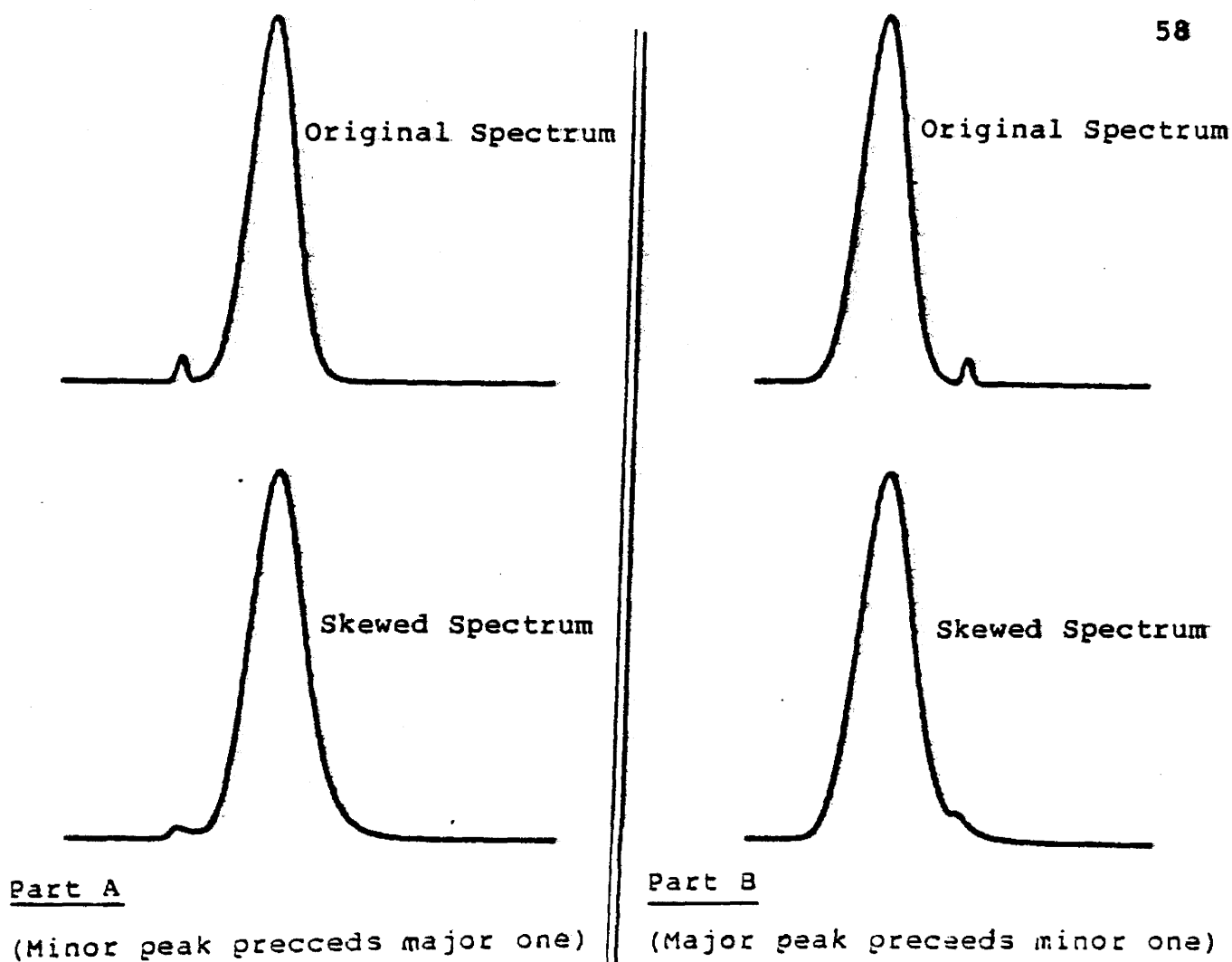
FIGURE 5.5

Both parts A and B have the same values for α (of the exponential distorting function $\alpha e^{-\alpha t}$). Also the area of the major peak is 75 times larger than that of the minor one. In part A, a minor peak preceeds a major peak, and in part B a major peak preceeds a minor one. It is evident that when a minor peak preceeds a major one then the loss of resolution with increasing α is very small compared to the case where a major peak preceeds a minor one.

c. Models for Representing Skewed GC Peaks

It is generally agreed that skewed GC peaks could best be represented as the convolution of a Gaussian resolution function with a one sided exponential decay function (13), (refer to figure 5.6). The skewed GC peak can therefore be represented by the following convolution integral I(t)

$$I(t) = \frac{\alpha A}{\sigma\sqrt{2\pi}} \int_0^t e^{-[(u-t_r)^2/2\sigma^2]} e^{-\alpha(t-u)} du$$

5.1

where A = the peak amplitude

σ^2 = variance of the Gaussian

α = $1/\Upsilon$ where Υ is the time constant of the exponential modifier

t_r = is the peak's centroid (retention time)

u = is a dummy variable of integration.

G(t)　　　　　*　　　　　R(t)

I(t)
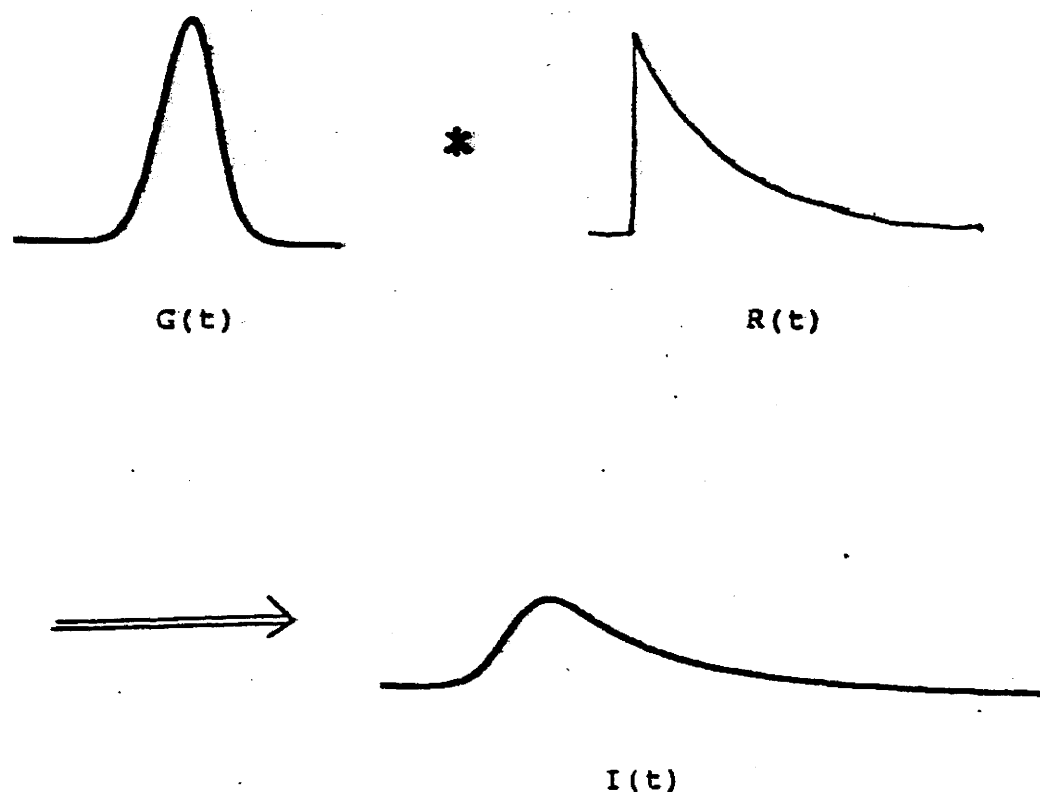
FIGURE 5.6

Model for representing a typical skewed GC peak.
Here $G(t) = A/\sigma\sqrt{2\pi}\ e^{-(t-t_r)^2/2\sigma^2}$ is the Gaussian
resolution function, $R(t) = \alpha e^{-\alpha t}$ is the exponential
distortion and I(t) is the observed result given by equation
5.1.

Other models for generating skewed Gaussians also exist. For example, Sumner (12), proposed a method which was based on a physical model of mixing dead volumes. For this method, the following algorithm could be used to generate EMGP.

$$g(n+1) = [g(n) \quad (V-1) + f(n)]/V$$

where f(n) is the original gaussian distribution function, V is the equivalent volume (where $V = \mathcal{T} X$ flow rate), $\mathcal{T}$ is the time delay constant, and n is the number of volume units that have passed through the mixing cell (ie. a dead volume).

d.   Properties Of Exponentially Modified peaks

An important property of the exponentially modified distortion is that the area under the EMGP is equal to the area of the undistorted signal.   A proof of this is as follows.   The observed responce is

$$I(t) = G(t)*R(t) = \int_{-\infty}^{\infty} G(u)R(t-u)du$$

where G(t) is the input function and R(t) is the system response. The total observed output signal at any time t' is

$$I(t') = \int_{-\infty}^{t'} G(u) R(t'-u) du. \qquad 5.2$$

Upon transforming equation 5.2 substittuting x = t'-u we get

$$I(t') = \int_{0}^{\infty} G(t'-x) R(x) dx. \qquad 5.3$$

Integrating equation 5.3 gives the area

$$\int_{-\infty}^{\infty} I(t') dt' = \int_{-\infty}^{\infty} G(t'-x) dt' \int_{0}^{\infty} R(x) dx. \qquad 5.4$$

The integral $\int_{-\infty}^{\infty} G(t'-x) dt' = \int_{-\infty}^{\infty} G(t') dt'$ since the limits are infinite and hence independent of x. We therefore get

$$\int_{-\infty}^{\infty} I(t') dt' = \int_{-\infty}^{\infty} G(t') dt' \int_{0}^{\infty} R(x) dx \qquad 5.5$$

R(x) is the solution of the differential equation

$$\frac{dR(x)}{dx} + \frac{1}{\mathcal{T}} R(x) = \frac{1}{\mathcal{T}} \delta(x) \qquad 5.6$$

which describes the exponentially distorting system. Here $\mathcal{T}$ is the time constant and $\delta(x)$ is a delta function.

Integrating equation 5.6 gives

$$\int_{-\infty}^{\infty} \frac{d}{dx} R(x)dx + \frac{1}{\mathcal{T}} \int_{-\infty}^{\infty} R(x)dx = \frac{1}{\mathcal{T}} \int_{-\infty}^{\infty} \delta(x)dx \qquad 5.7$$

The integral $\int_{-\infty}^{\infty} \delta(x)dx = 1$ from the property of the Dirac delta function. Also since $R(-\infty) = R(+\infty) = 0$, the integral on the left is equal to zero. The limits of integration on the second integral change to $\int_{0}^{\infty}$ since $R(x) = 0$ for $x < 0$, therefore equation 5.7 becomes

$$\int_{0}^{\infty} R(x)dx = 1 \qquad 5.8$$

It follows, that equation 5.5 becomes

$$\int_{-\infty}^{\infty} I(t')dt' = \int_{-\infty}^{\infty} G(t')dt' \qquad 5.9$$

The area of the undistorted curve is therefore equal to the area of the distorted curve, refer to figure 5.2. In a physical sence, this simply means that the total amount of sample that goes into the column will eventually come out of the column.

Through the use of Laplace transforms, additional information about the properties of EMGPs could be obtained.

The Laplace transform $F(p)$ of a function $f(x)$ is defined by

$$F(p) = \int_{-\infty}^{\infty} e^{-px} f(x)\, dx \qquad\qquad 5.10$$

The nth derivitive of the transform is

$$\frac{d^n}{dp^n} F(p) = \int_{-\infty}^{\infty} (-x)^n e^{-px} f(x)\, dx \qquad\qquad 5.11$$

and the nth moment, M_n, of $f(x)$ is defined by

$$M_n = \frac{\int_{-\infty}^{\infty} x^n f(x)\, dx}{\int_{-\infty}^{\infty} f(x)\, dx} \qquad\qquad 5.12$$

Comparing 5.11 and 5.12 shows that

$$M_n = (-1)^n \lim_{p \to 0} \left[\frac{1}{F(p)} \cdot \frac{d^n}{dp^n} F(p) \right] \qquad\qquad 5.13$$

The delay in retention time of a EMGP is given by M_1, the 1st moment about $x=0$. The varience of the EMGP, V_2, is defined as the 2nd moment about $x = M_1$

$$V_2 = \text{variance} = \frac{\int_{-\infty}^{\infty}(x-M_1)^2 f(x)\,dx}{\int_{-\infty}^{\infty} f(x)\,dx} \quad .$$

$$= M_2 - 2M_1 M_1 + M_1^2$$

$$= M_2 - M_1^2 \qquad\qquad 5.14$$

For an exponential distorting function, $R(x) = \dfrac{1}{\jmath} e^{-x/\jmath}$, the contribution to the observed peak's retention time is obtained from equation 5.13. The Laplace transform of

$$R(x) = \frac{1}{\jmath} e^{-x/\jmath} = \frac{1}{(1 + p\jmath)} = \text{transfer function}$$

hence $\quad M_1 = -\lim_{p \to 0} \left[1 + p\jmath\right] \dfrac{d}{dp} \dfrac{1}{(1 + p\jmath)}$

$$= \jmath^2$$

Therefore the introduction of an exponential modifier to the system delays each peak by $\jmath$ and increases the variance of each peak by $\jmath^2$.

When a signal is applied to an exponential modifier, the maximum of the modified signal will always fall on the contour of the input signal as illustrated in figure 5.2. This can be proved in the following way.

An exponentially modified system is defined by the differential equation

$$\frac{dI(t)}{dt} + \frac{I(t)}{\Upsilon} = \frac{G(t)}{\Upsilon} \qquad\qquad 5.15$$

where I(t) is the output of the system, G(t) is the input function, and Υ is the time constant. The output maximum occurs when

$$\frac{dI(t)}{dt} = 0 \ .$$

Therefore it follows from equation 5.15 that

$$I(t) = G(t)$$

An alternative proof involves Laplace transforms, and may be shown as follows. The Laplace transform of the output $\overline{I}(p)$ is related to the transform of the input $\overline{G}(p)$ by

$$\overline{I}(p) = \overline{G}(p)\,\overline{T}(p)$$

where $\overline{T}$ is the transfer function.

$$\overline{I}(p) = \overline{G}(p) \ . \ \frac{1}{(1 + p \ )}$$

and $\quad \overline{I}(p) + \overline{I}(p).p\mathcal{T} = \overline{G}(p).$ $\qquad\qquad$ 5.16

By the differentiation theorem for Laplace transforms

$$\frac{d}{dt} I(t) = p\overline{I}(p).$$

At the maximum of the output $\frac{d}{dt} I(t) = 0$ therefore $\overline{I}(p).p = 0$ and equation 5.16 becomes

$$\overline{I}(p) = \overline{G}(p)$$

Applying the reverse transformation yields

$$I(t) = G(t)$$

e. Technique for "De - Skewing" GC Peaks

Several methods for removing the distorting exponential response function from the observed data have been previously suggested (12). These methods usually involve lengthy calculations and an estimation of the time constants $\mathcal{T}$ of the exponential modifiers. These time constants $\mathcal{T}$ are usually estimated by statistical moment

calculations (10), although other methods have been suggested (14).

Since the GC provided virtually noise-free data, the following inverse filter technique proved to be a fast, straightforward procedure for obtaining the undistorted result. However, if the GC data were to contain a significant amount of noise, the inverse filter technique would not be applicable. In such a situation other methods such as Baysian deconvolution can be used (15).

The basis of the inverse filter technique may be seen through the following treatment. The observed spectrum $I(t)$ is given by

$$I(t) = G(t)*R(t) = \int G(u)R(t,u)du + N(t). \qquad 5.17$$

where $G(u)$ is the response of a system, $R(t,u)$ is the response of the distorting function, and $N(t)$ is the additive noise. Here the symbol $*$ indicates convolution.

The "de-skewing" process requires the deconvolution or removal of the effect of the distorting exponential response function $R(t,u)$ where

$$R(t,u)du = \alpha e^{-\alpha(u-t)}du. \qquad 5.18$$

Upon substitution of equation 5.18 into equation 5.17, taking the Laplace transform and employing the associated convolution theorem, the following result is obtained.

$$I(p) = \frac{\alpha}{p + \alpha} \, g(p) + n(p).$$

Then by solving for g(p) and taking the inverse transform the following result is produced.

$$G(t) = \left[\frac{1}{\alpha} \frac{d}{dt} + 1 \right] \cdot [I(t) + N(t)] \Big|_{t=u} \qquad 5.19$$

Since the GC data is relatively noise free, the term N(t) can be ignored and the original signal obtained by solving the following equation

$$G(u) = \left[\frac{1}{\alpha} \frac{dI(t)}{dt} + I(t) \right] \Big|_{t=u}. \qquad 5.20$$

Hence, as equation 5.20 suggests, "de-skewed" peaks are obtained if the derivative of the original data are divided by the appropriate α (characteristic of the specific exponential distorting function) and added to the original data I(t). Figures 5.7, 5.8, and 5.9 demonstrate the inverse filter technique and illustrate the contrast between the original GC spectra and the "de-skewed" spectra.
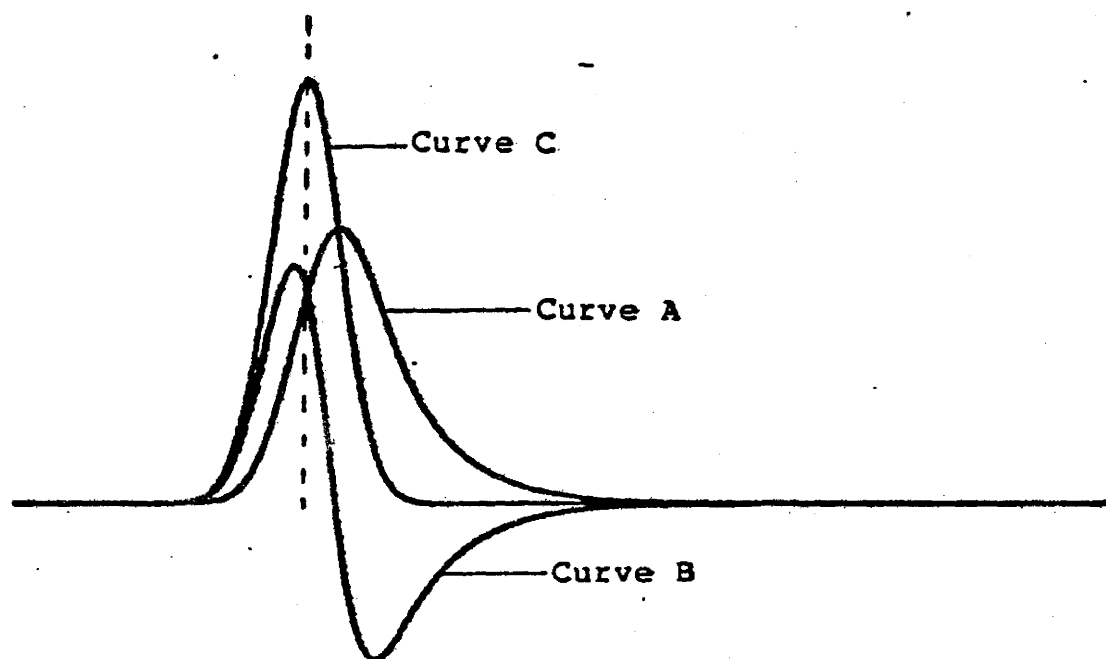
FIGURE 5.7

Curve A is the artificially produced skewed Gaussian peak by the convolution of a Gaussian with an exponential of $\alpha = .02$. Curve B is the derivative of A divided by $\alpha = .02$ and C is the "de-skewed" result (obtained by the addition of A to B). Note the displacement of the peak centroids for the skewed and unskewed peaks.

Part A

Part B

Part C
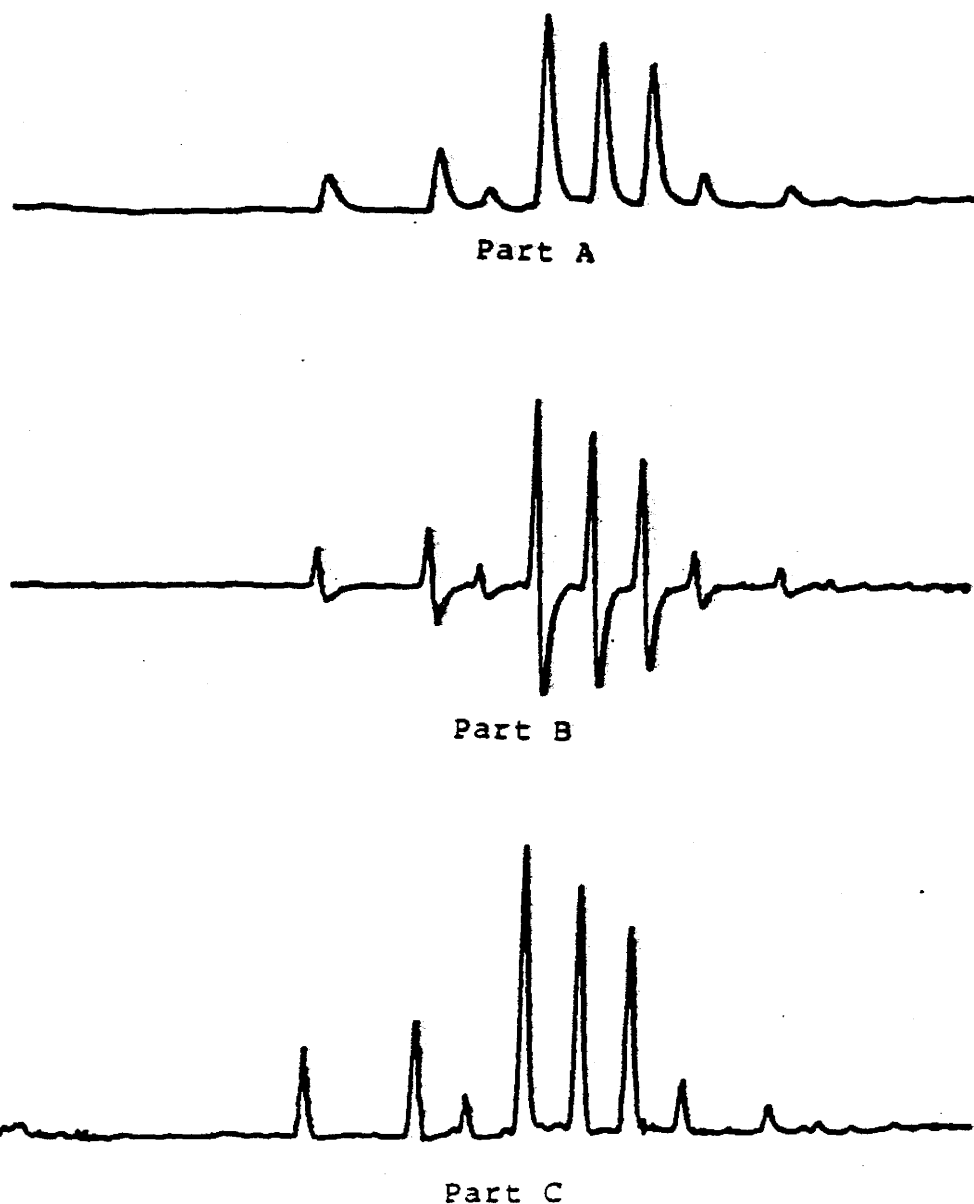
FIGURE 5.8

Process of "de-skewing" using the inverse filter deconvolution technique for a typical GC spectra (E. Coli). Part A demonstrates the partitioning of the original spectra, part B is the derivitive of part A devided by the appropriate value for α . and part C is the "de-skewed" result produced from the addition of B to A.
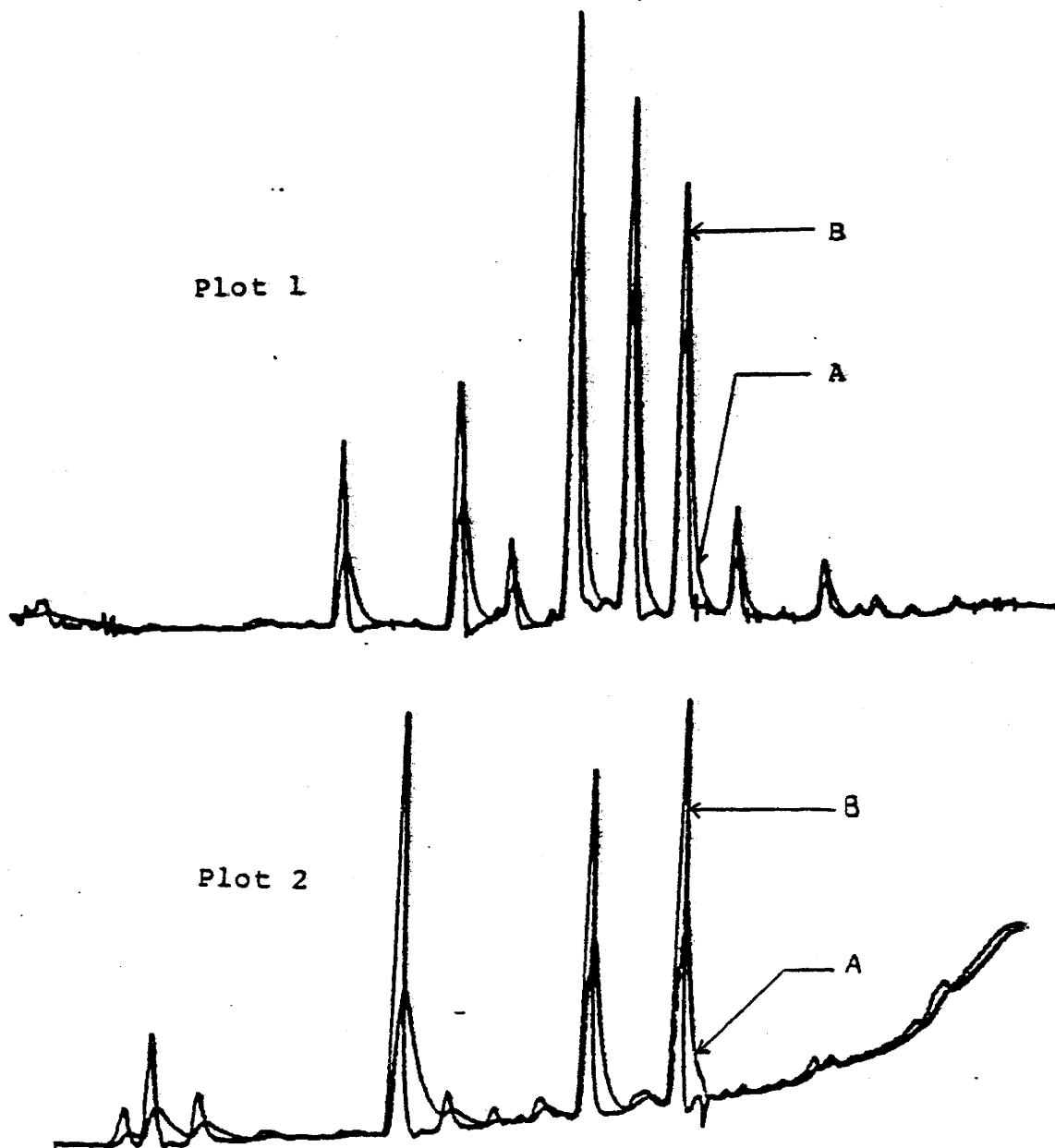
FIGURE 5.9

Two "de-skewed" spectra, (plot 1 and plot 2). Plot 1
is of E. Coli and plot 2 is of A. Antititratus. "A" is the
original GC spectra (for E. Coli), and "B" is the "de-
skewed" spectra. Note how the peak centroids of the "de-
skewed" spectra are displaced slightly to the left, and the
noticible improvement in resolution.

Sometimes in a GC system where a capillary column is used, skewing occurs on the leading edge of the peak instead of the trailing edge. In this case the peaks could be "de-skewed" using the same technique, except that α is replaced with $-\alpha$. In other words the derivative is subtracted.

In order to properly "de-skew" the EMG peaks, the appropriate value of α must be known. As can be seen from figure 5.10 if α is too large then the resultant peak (after employing the inverse filter deconvolution technique) still has a skewed tail. Conversly if α is too small, then too much of the skew is removed and a negative lobe will result.

If more than one exponential modifier had a major effect on the final skewed GC peak, then "de-skewing" once with a value α (where $\alpha e^{-\alpha t}$ is the skewing function) would not be sufficient. For example, if only two major exponential modifiers were present in a system then the final observed GC peak would be produced by first convolving the original "near gaussian" signal with an exponential of α_1 producing $EMGP_1$, $EMGP_1$ is then convolved with another exponential of α_2 producing $EMGP_2$.

To remove the skew properly from $EMGP_2$ it must be "de-skewed" twice, first with α_2 and then with α_1. This follows directly from the derivation of the inverse filter,
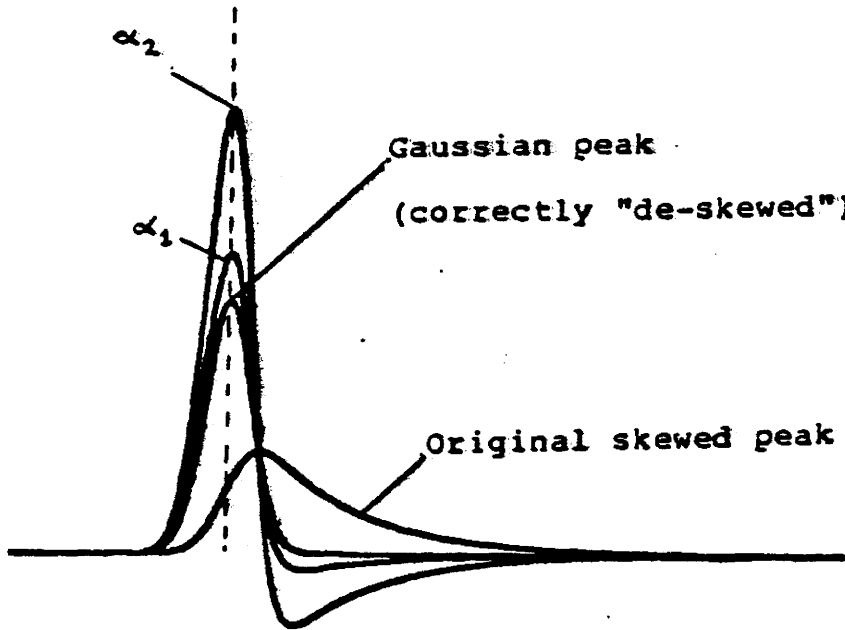
FIGURE 5.10

Varying α of the distorting exponential responce function $\alpha\, e^{-\,\alpha(u\,-\,t)}du$ effects the shape of the final "de-skewed" peak. If α is too small then negative lobes result, if α is too large then a skew is still present, until as the result becomes the original skewed peak $M(t)$ (since the inverse filter $\frac{1}{\alpha}\frac{dM}{dt} + M(t)$ in the "de-skewing" technique becomes $\longrightarrow M(t)$ as $\alpha \longrightarrow \infty$. Shown above is the original skewed peak, the correctly "de-skewed" peak (with the correct α) and two other skewed peaks with α too small, ($\alpha_1 < \alpha_1$).

equation 5.6, which G(u) and I(t) could be any function.
Only the distorting exponential function

$$R(t,u)\,du \;=\; \alpha e^{-\alpha(u-t)}\,du$$

was specified.   During the $1^{\underline{st}}$ stage of "de-skewing" G(u) is
$EMGP_1$ and the observed signal I(t) is $EMGP_2$.   For the $2^{\underline{nd}}$
stage G(u) is the original "near gaussian" and the observed
signal I(t) is $EMGP_1$.   Figure 5.11 demonstrates that $EMGP_2$
can not be totally "de-skewed" with a single α .   If α is
less than α_2 then a negative lobe will result and a skew is
still present, whereas when α gets increasingly larger than
α_2, then the result resembles $EMGP_2$ more.

Therefore if there are N exponential modifiers in a
GC system then the spectra must be "de-skewed" N times with
the appropriate α characteristic of each exponential
modifier.
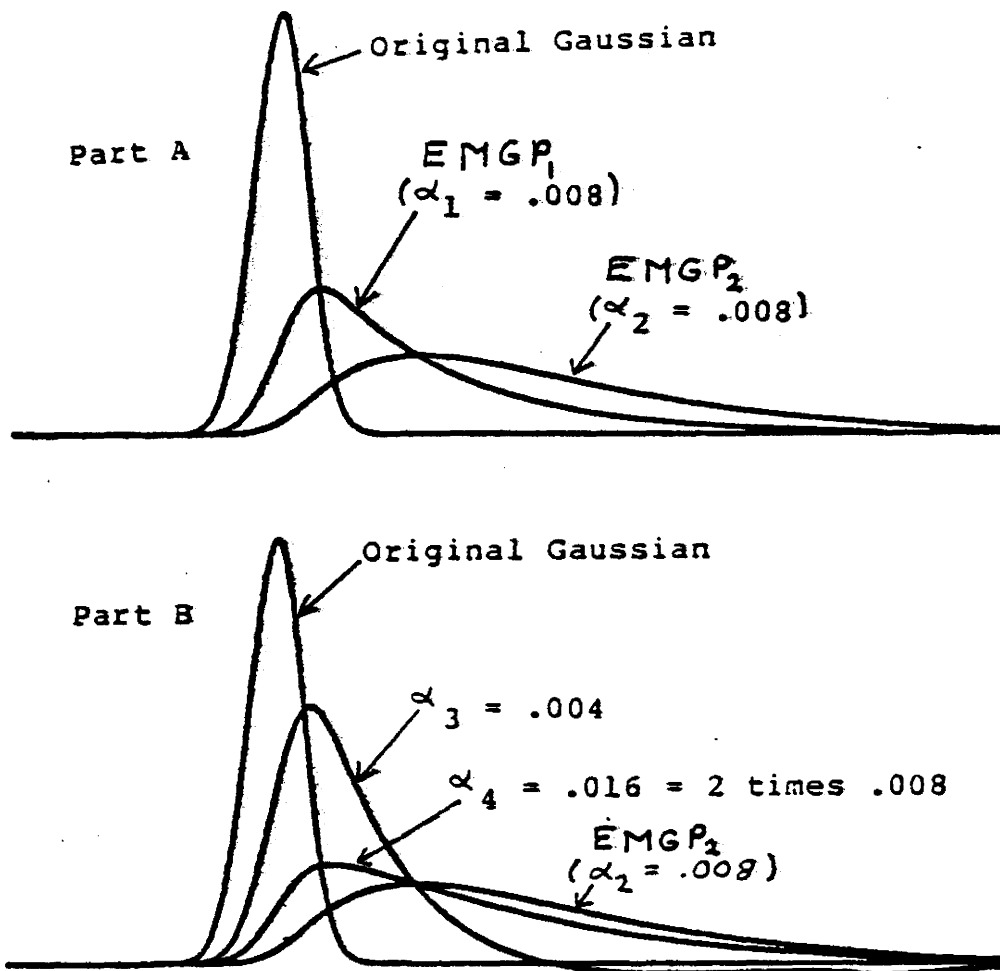
FIGURE 5.11

In part A, $EMGP_2$ was obtained by the cascading effect of the exponential modifiers with ($\alpha_1 = .008$ and $\alpha_2 = .008$). Part B illustrates the results obtained when $EMGP_2$ is "de-skewed" with $\alpha_3 = .004$ and $\alpha_4 = .006$. Since $EMGP_2$ was produced by more than one modifier, therefore it cannot be "de-skewed" to it's original form by employing the "de-skewing" procedure only once with a fixed value for α.

In the analysis undertaken, it was observed that "de-skewing" a spectrum once with a fixed value for α did not totally "de-skew" all of the peaks (refer to figure 5.12). However, "de-skewing" the spectrum once with varying α , which increased linearly along a ramp and then remained constant did remove the skews, (refer to figure below).
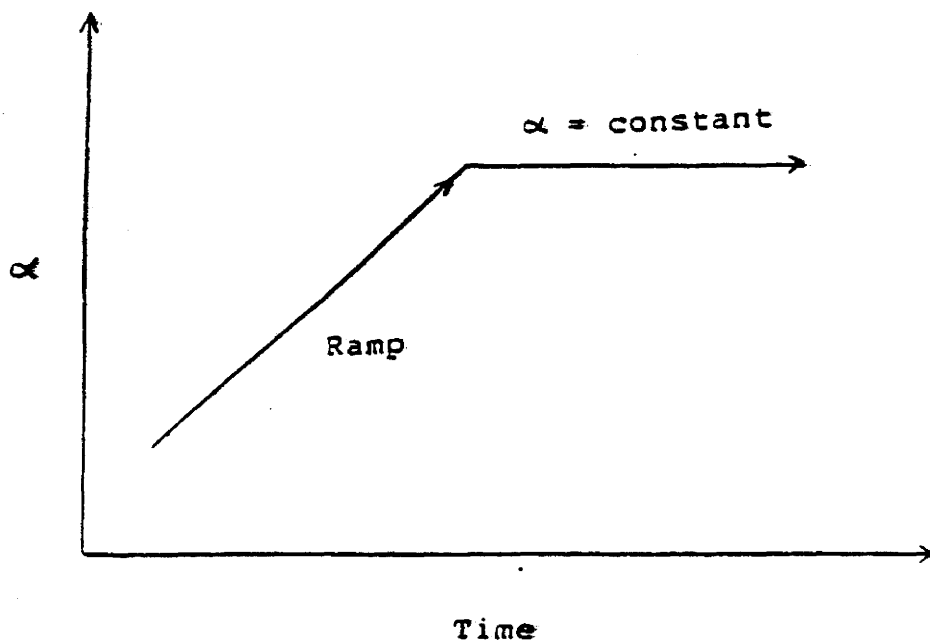
This suggests two things. Firstly that there is one dominant exponential modifier in the GC system and secondly

Original Spectrum

"De-skewed" spectrum with a fixed value for α , (α = .17)
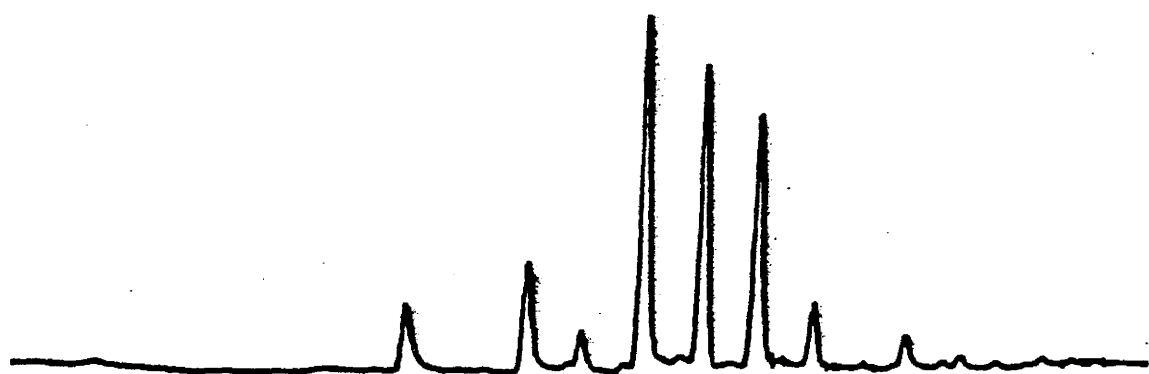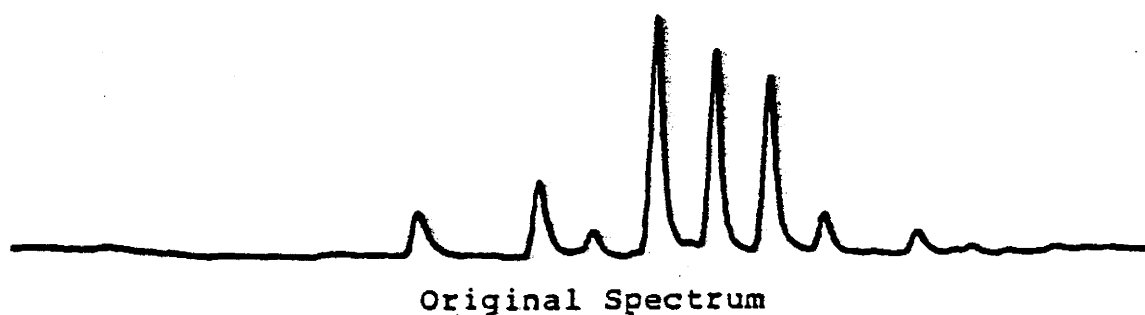
FIGURE 5.12

This figure illustrates the effect of "de-skewing" a typical GC spectra only once, with a fixed value for α. Skewed peaks are still present at the initial stages of the spectra.

that the linear temperature programming that was used affected the degree of skewing (since the temperature was also programmed linearly along a ramp and then held constant).

As the temperature increased, the time it takes for a component to travel through the column decreases, therefore there is less time for lateral diffusion and adsorption and hence the degree of skewing decreases.

In conclusion it appeared that only one exponential modifier was dominant in the GC system studied, and it was due to lateral diffusion and adsorption. Also for the "de-skewing" process, the way in which α is varied depends upon and corresponds to the temperature programming used in the system.

5.3   Baseline Drift

a.   Principle Cause of Baseline Drift (Column Bleed)

The temperature dependence of electron absorbtion by a compound is referred to as "column bleed".   This effect manifests itself through movement of the spectral baseline. For most of the GC spectra that had baseline drift, it appeared that the baseline increased in an exponential fashion as the temperature was linearly increased. (refer to figure 5.13).

Typically baseline drift has frequency components lower than the signal and thus may be removed by some form of filter.   The form of filter used here corresponded to a smoothed second derivative.

b.   Zero Area Filter Baseline Correction technique
(Theory of Operation)

The zero area digital filter is a correlation technique that offers many advantages over most other methods used to remove non constant baselines.   It is a simple algorithm, which requires little computation hence is fast and applicable to small computers.

This method essentially involves the filtering of a spectrum with a correlator F(x) where F(x) is a zero area rectangular foldin function, shown in figure 5.14.   Using

Drifting Baseline
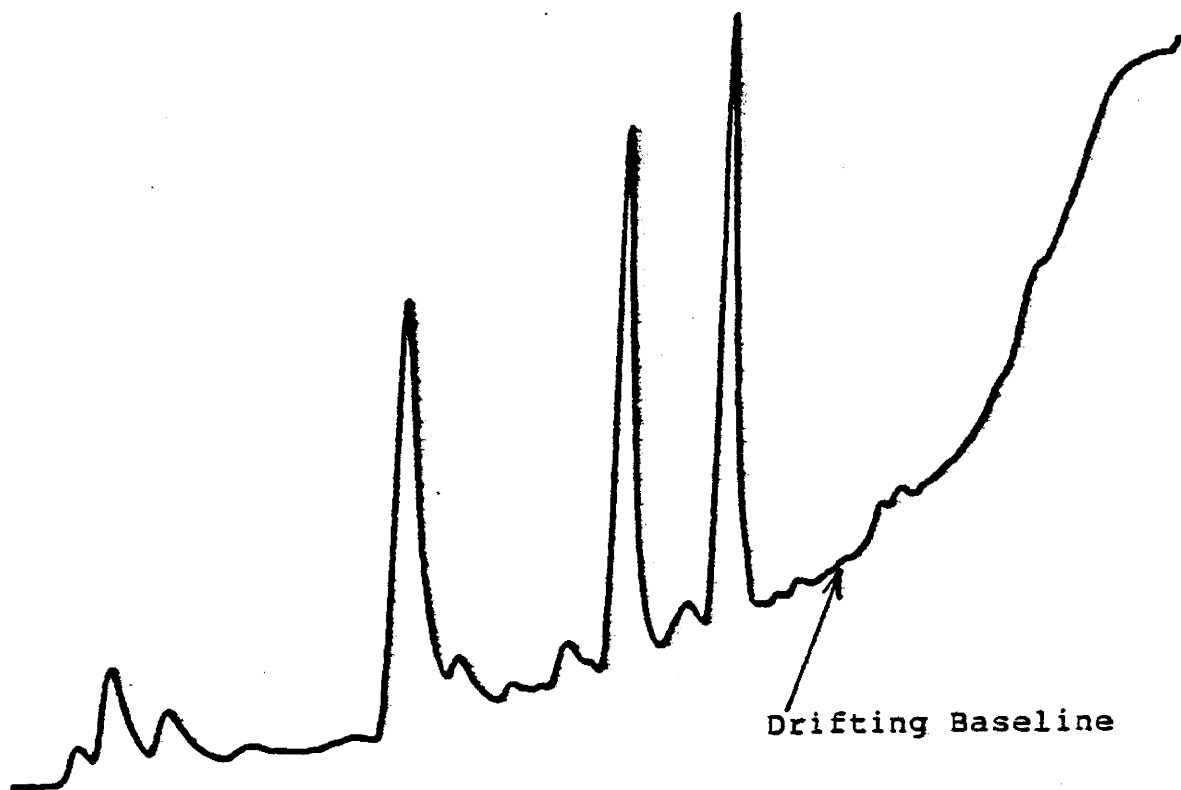
FIGURE 5.13

Illustration of baseline drift for a temperature programmed GC spectra.

G(x)   F(x)
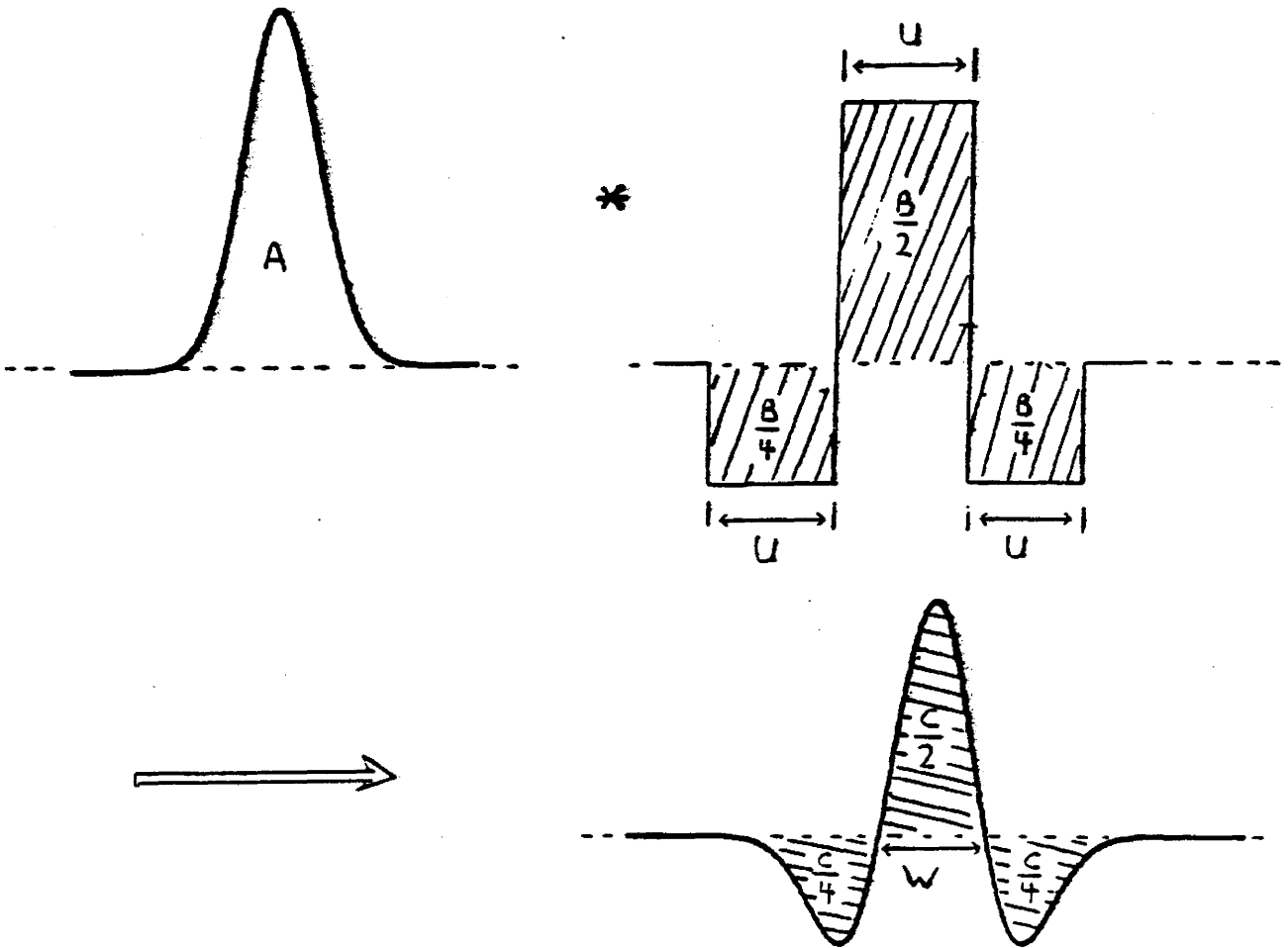
S(x) = second derivative of G(x)

FIGURE 5.14

The output S(x) is the result of the correlation of a gaussian peak G(x) with a zero area rectangular fold in function F(x). The area of the original gaussian is A, for the filter is B, and for the correlated signal is C. If the total area of the filter is 1 (i.e. B=1) then C=A.

this function F(x) is advantageous since the correlation only involves addition and subtraction of integers hence computational speed is fast.

The correlator F(x) has a total area of zero with negative wings which are each onehalf the area of the positive lobe. When a gaussian peak is correlated with this filter F(x) then a signal S(x) is produced which has two negative lobes and one positive lobe where the total area of the lobes equals that of the original gaussian peak, (refer to figure 5.14). Figure 5.15 provides a sample of a correlated spectrum produced by operating with an appropriate (i.e. matched) rectangular zero area filter.

To gain an understanding of how the filter width affects the appearance of the correlated signal, the following limiting cases may be considered:

Case (1):

If a very narrow filter F(x) is used (ie. if U $\longrightarrow$ 0, where U is the width of the filter lobes), then the result obtained after correlation is the same as taking the second derivative of the original gaussian. Thus correlating a gaussian signal $y = \dfrac{Ae^{-x^2/2\sigma^2}}{\sqrt{2\pi}\ \sigma}$ with such a filter produces

$$y'' = \frac{dy^2}{dx^2} = \frac{A\ [(x^2/\sigma^2) - 1]\ e^{-x^2/2\sigma^2}}{\sigma^3\ \sqrt{2\pi}} \qquad \text{5.21}$$

Curve A

(Original GC Spectrum)

Curve B

(The "de-skewed" spectrum)

Curve C

(Filtered Spectrum to remove the drifting baseline)
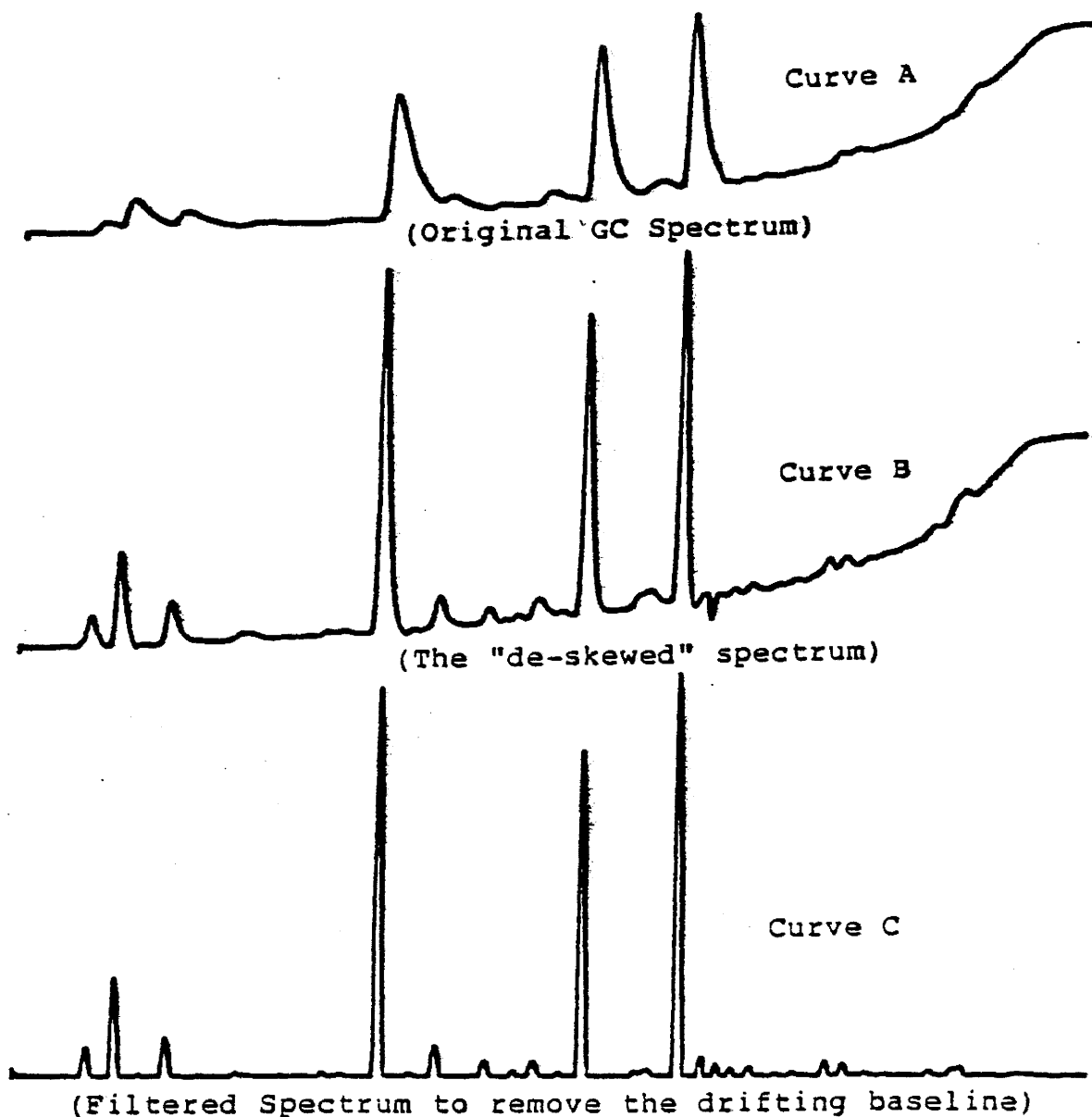
FIGURE 5.15

This is an illustration of the noise removal process for a typical GC spectrum. Curve A is the original spectrum, curve B is the "de-skewed" spectrum and curve C is the correlation spectrum, (produced by correlating with the zero area filter).

(refer to figure 5.16).

From equation 5.21 the width W between the zero crossings of the derivative, the height H, and the area A of the positive lobe can be determined. Setting $y'' = 0$ and solving for x gives $x = \pm\sigma$, hence $W = 2\sigma$, also setting x = 0 gives the height $H = (A/\sqrt{2\pi})(1/\sigma^3)$. Similarly the area of the positive lobe is equal to

$$A = \int_{-\sigma}^{+\sigma} y'' dx = 2A\sqrt{\frac{2}{\pi e}}\;\frac{1}{\sigma^2}$$

It is evident that if the baseline is slowly varying compared with the GC peaks, then by using a narrow filter, the resultant peak height will be reduced but the drifting baseline will be effectivly removed.

Case (2):

If the width U of the filter F(x) is equal to or greater than 10σ (99.9% of the gaussian lies between these bounds), then the height H of the positive lobe can be approximated by $H = A/U$. Figure 5.17 provides the correlated signal for when U is exactly equal to 10σ. As U becomes very large (ie. as $U \longrightarrow \infty$), the crossover width W approaches $\longrightarrow U$ since the distance AB relative to U decreases as illustrated by figure 5.18. In addition, the
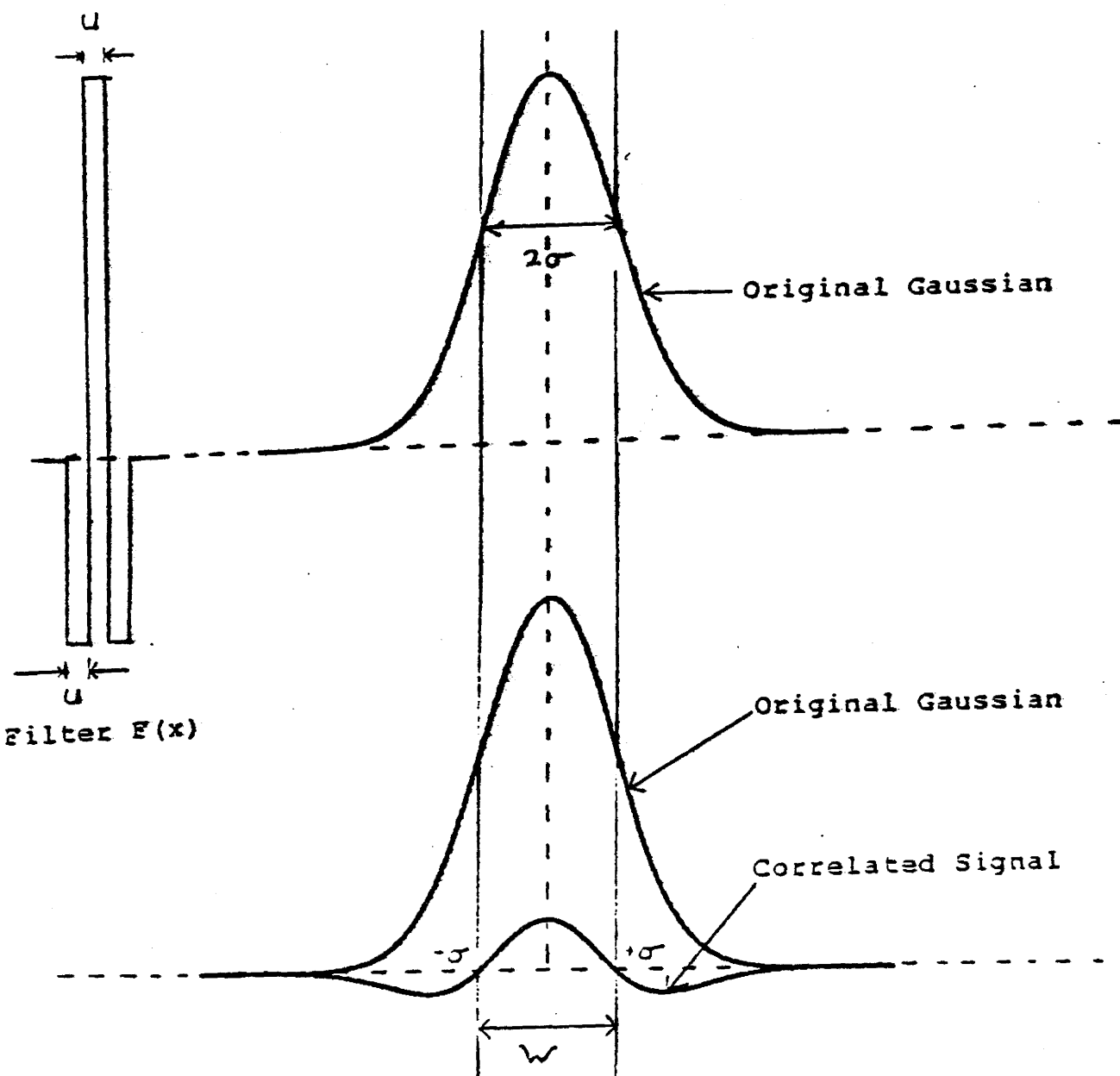
FIGURE 5.16

Case (I), correlating a gaussian peak with a very
narrow filter. The cross width $W = 2\sigma$ which equals the
width of the $e^{-1/2}$ points $(x = \pm \sigma)$ of the original gaussian.
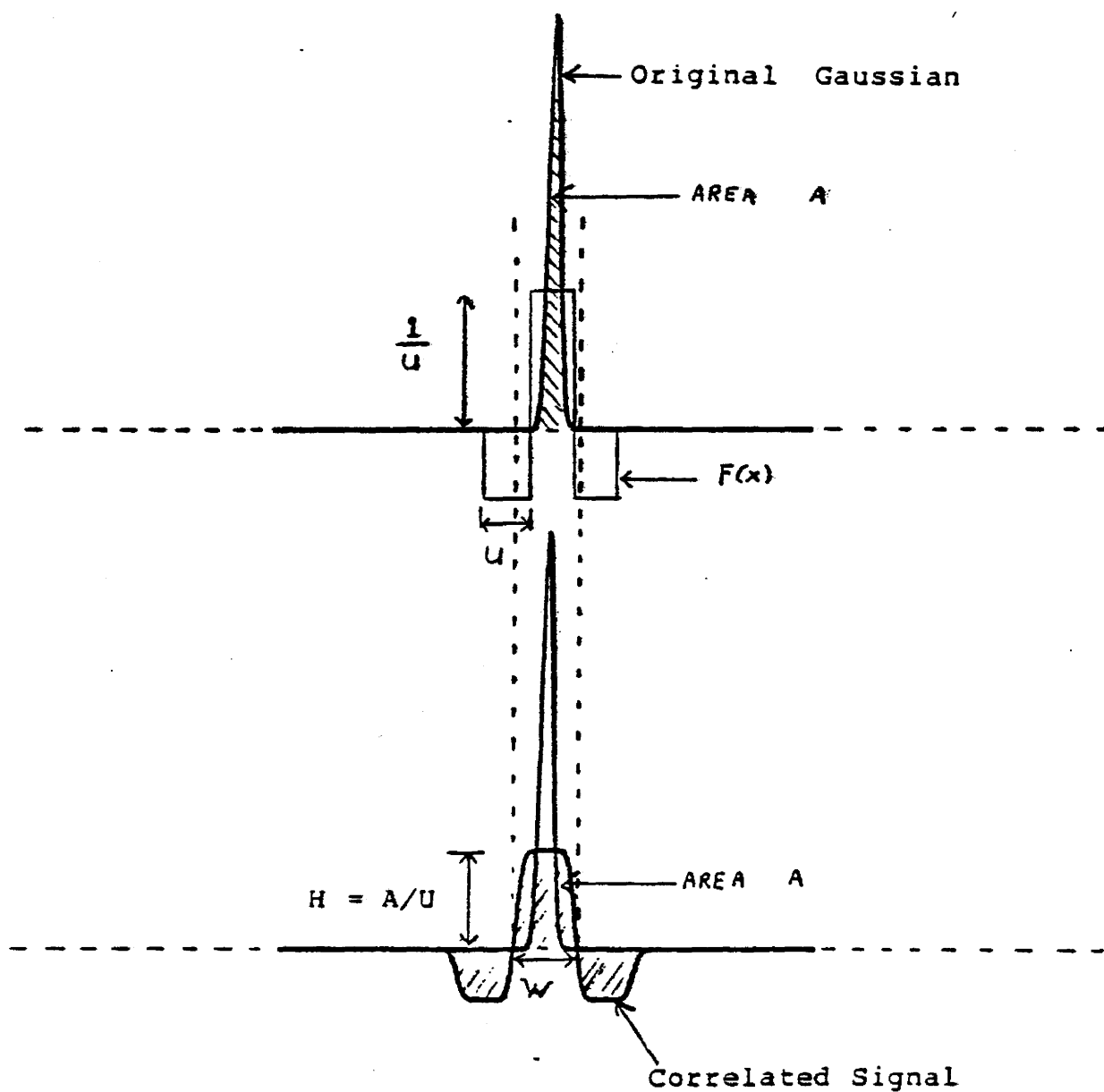
FIGURE 5.17

Correlation of a gaussian with a filter F(x) that
has a width U equal to the base width of the gaussian. Note
the cross width W is slightly larger than U.
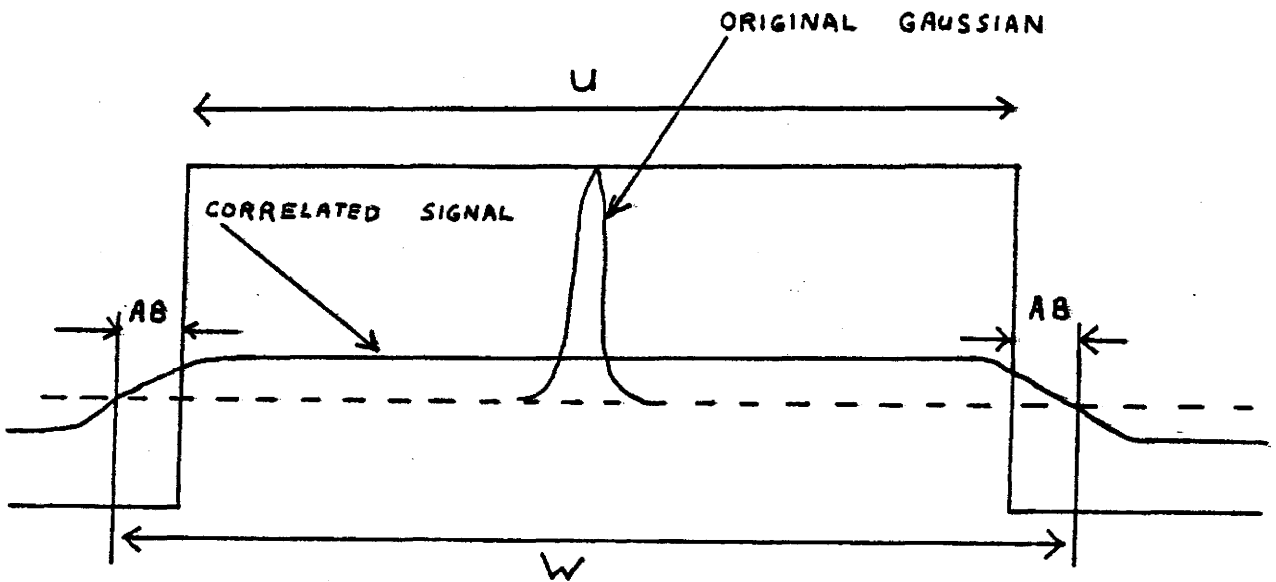
FIGURE 5.18

Case (2), correlating a gaussian peak with a very wide filter. The crosswidth W = U + 2AB, therefore as U⟶∞ then W ⟶ U.

positive lobe area approaches A and the height H of the positive lobe approaches zero since $H = A/U \longrightarrow 0$ as $U \rightarrow \infty$.

The following summarizes the two cases discussed. For the correlated signal, the range of possible positive lobe areas is

$$2A \sqrt{\frac{2}{\pi e}} \; \frac{1}{\sigma^2} \; \cdots \text{———} \longrightarrow \cdots \cdots \; A$$

for a very narrow   intermediate   very wide

filter width ($U \rightarrow 0$)   filter widths   filter width

($U \longrightarrow \infty$)

Similarly, the height of the positive lobe of the correlated signal ranges from

$$\frac{A}{\sqrt{2\pi}} \; \frac{1}{\sigma^3} \; \cdots\cdots\cdots \; \frac{A}{U} \; \cdots \text{———} \longrightarrow \cdots \; 0$$

for very narrow   intermediate   for infinitly

filter width   filter widths   wide filter

($U \longrightarrow 0$)   width ($U \rightarrow \infty$)

The cross width W of the correlated signal ranges from

$$2\sigma \quad \cdot\cdot \text{———} \longrightarrow \cdot\cdot \quad \tilde{=} \text{ filter width U } \cdot \text{———} \longrightarrow \cdot\cdot \; \infty$$

for very narrow   for wide filters   for infinitly

filters ($U \longrightarrow 0$)   wide filters

All data referred to in the remaining sections have
the following units. Retention data is given in units of
sampling intervals and peak height is given in units of
counts per sampling interval.

Figure 5.19 illustrates how the positive area of the
correlated signal varies with the width at the $e^{-1/2}$ points
$(x = \pm \sigma)$ of the gaussian. In this figure, the area of
the gaussian signal is normalized to A=1, and the filter
width is fixed at U=7. Alternatively, the variation of the
positive lobe area with the filter width U is illustrated by
figure 5.20. In this case the FWHM of the gaussian is fixed
and the filter width varied. It is evident from figures
5.19 and 5.20 that as the FWHM of the gaussian relative to
the filter width changes so does the area of the positive
lobe. Therefore, if the peak widths in a spectrum are not
equal, then upon application of a fixed zero area filter,
the relative areas of the positive lobes will not be
correct. Conversely if the peak widths in the spectra were
equal then the relative positive lobe areas (after
filtering) would be correct. Generally, the peak widths in
a GC spectra are not equal therefore it is imperative that
the positive lobe areas in the correlated spectrum be
corrected accordingly. This area correction is discussed in
section 5.3.d.

POSITIVE LOBE AREA VS. GAUSSIAN 2σ

GAUSSIAN 2σ
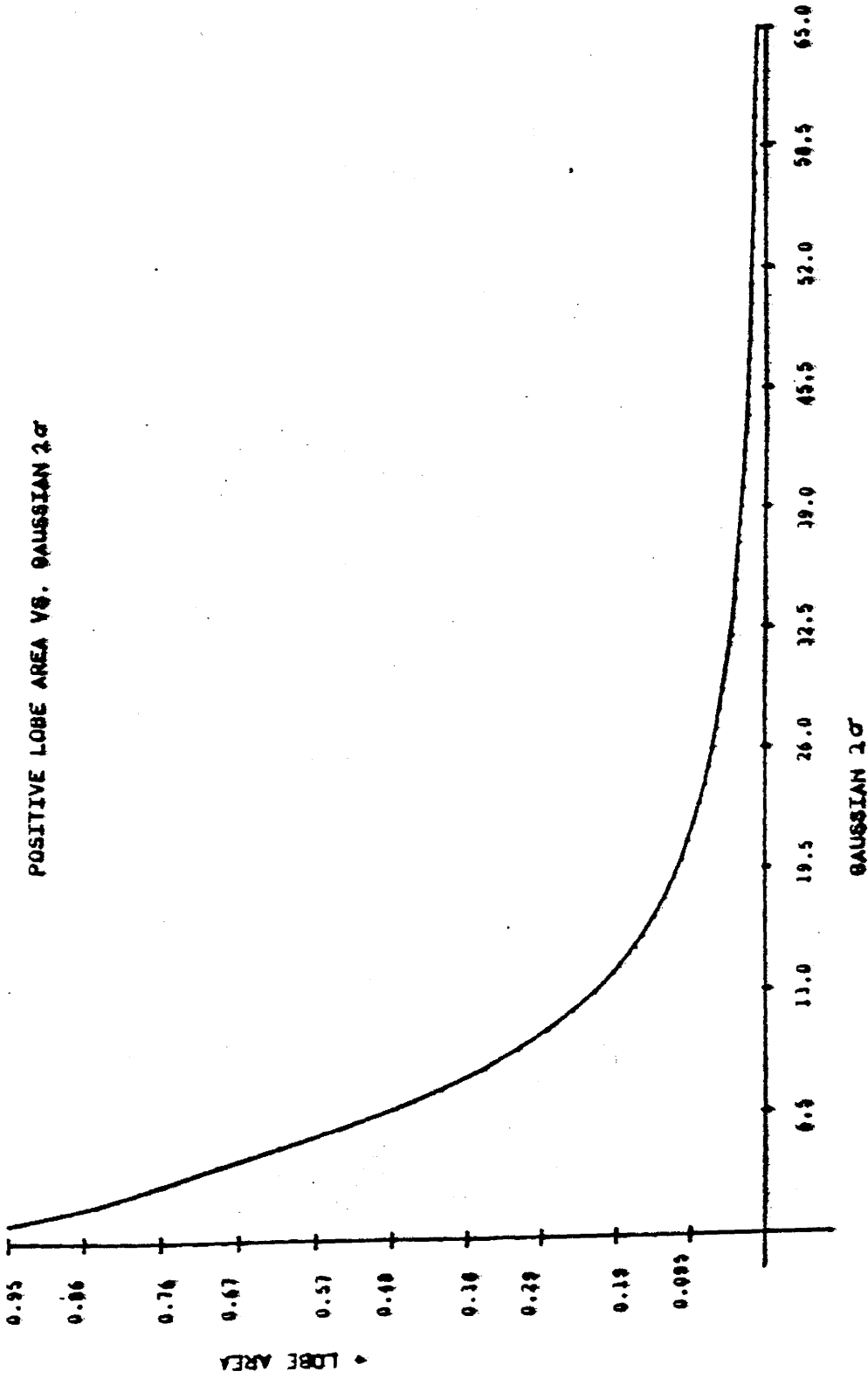
← LOBE AREA

Figure 5.19

[Illustration of how the positive lobe area (of the correlated signal) varies with the 2σ of the original gaussian. Note the greatest variations occur for gaussians with a 2σ less than 13. The values along the x axis are given in units of sampling intervals.

POSITIVE LOBE AREA VS. FILTER WIDTH

ZERO AREA FILTER WIDTH
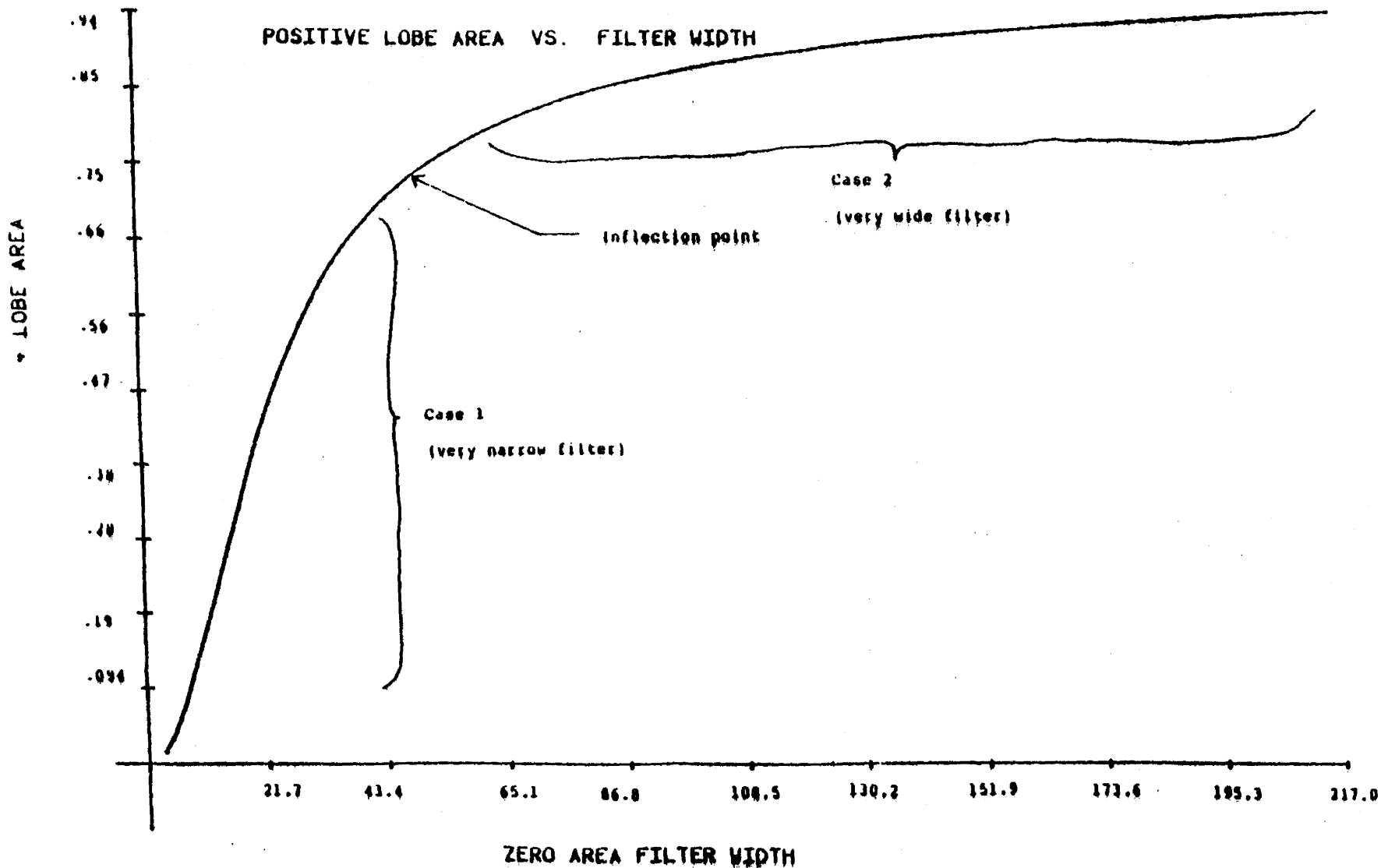
Figure 5.20

Plot of the positive lobe area of a filtered
gaussian peak versus the filtered width U. For the figure
the width of the gaussian was fixed at $2\sigma = 24$, while U was
varied.

To confirm how the filter width affects the outcome of the correlated signal, a gaussian peak was generated and then systematically correlated with filters of different width. The parameters of importance such as the positive lobe area, crossover width W, and the positive lobe height H, were tabulated for each filter width U, (refer to table 5.1 and figure 5.21 for a sample illustration).

The areas and the centroids of the positive peak lobes were calculated from the 0th and 1st moments respectivly, where

$$0^{th} \text{ moment} = \text{peak area} = \int_{-\infty}^{+\infty} I(t)dt = \sum I(t)$$

$$1^{st} \text{ moment} = \text{peak centroid} = \frac{\int_{-\infty}^{\infty} t.I(t)dt}{\int_{-\infty}^{\infty} I(t)dt}$$

$$= \frac{\sum t.I(t)}{\sum I(t)}$$

where I(t) is peak intensity with respect to time. A linear interpolation was used to estimate the zero crossover points and hence the zero crossover width.

Occasionally a situation can exist where a very small peak is close to a very large one. In such cases

FILTER WIDTH U	+ LOBE AREA	CROSSOVER WIDTH W	HEIGHT H
3.00	0.015	24.19	0.0010
5.00	0.040	24.53	0.0027
7.00	0.076	25.04	0.0050
9.00	0.12	25.69	0.0077
11.00	0.17	26.52	0.010
13.00	0.22	27.51	0.013
15.00	0.27	28.65	0.016
17.00	0.32	29.91	0.018
19.00	0.37	31.33	0.019
21.00	0.41	32.84	0.021
23.00	0.45	34.45	0.022
25.00	0.49	36.15	0.022
27.00	0.52	37.91	0.023
29.00	0.55	39.74	0.023
31.00	0.58	41.61	0.023
33.00	0.61	43.52	0.023
35.00	0.63	45.46	0.022
37.00	0.65	47.42	0.022
39.00	0.66	49.39	0.022
41.00	0.68	51.37	0.021
43.00	0.70	53.36	0.021
45.00	0.71	55.35	0.020
47.00	0.72	57.35	0.020
49.00	0.73	59.35	0.019
51.00	0.74	61.34	0.019
53.00	0.75	63.34	0.018
55.00	0.76	65.34	0.018
57.00	0.77	67.34	0.017
59.00	0.78	69.34	0.017
61.00	0.79	71.34	0.016
63.00	0.79	73.34	0.016
65.00	0.80	75.34	0.015
67.00	0.80	77.34	0.015
69.00	0.81	79.34	0.014
71.00	0.82	81.34	0.014

TABLE 5.1

Parameter variation due to correlation filter width U variations. The values were obtained by varying the filter width U and keeping the original gaussian peak width constant at $2\sigma = 24$.
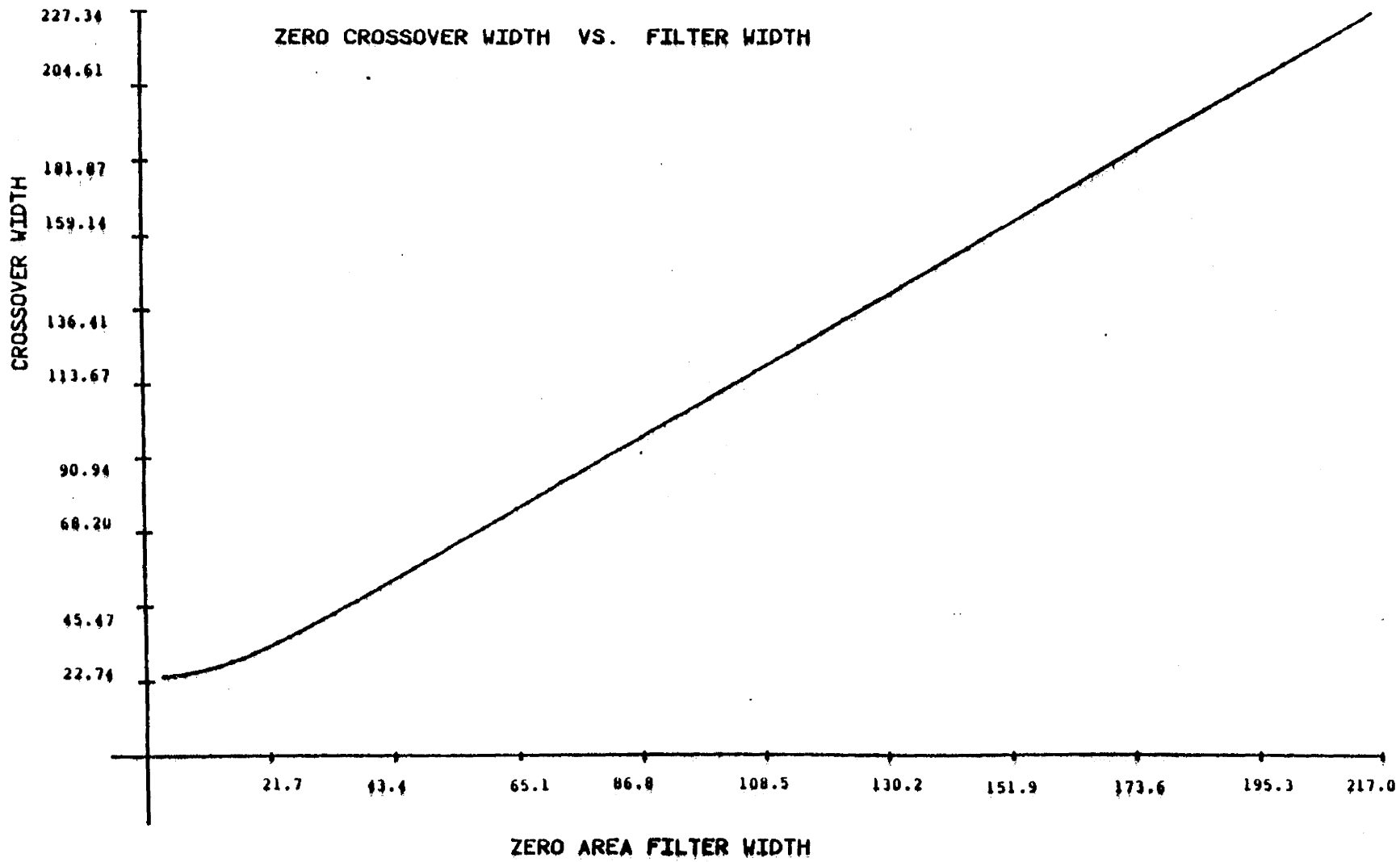
FIGURE 5.21

Plot of the zero crossover width W for a filtered gaussian peak versus the filter width U.

after correlating with a zero area filter the smaller peak
might not be visible in the resultant spectrum, (refer to
figure 5.22).

c.   Choosing The Optimal Correlation Filter

A paper by Robertson et.al (16) showed that for a
rectangular zero area filter with upper lobe width U and
lower lobe width L,
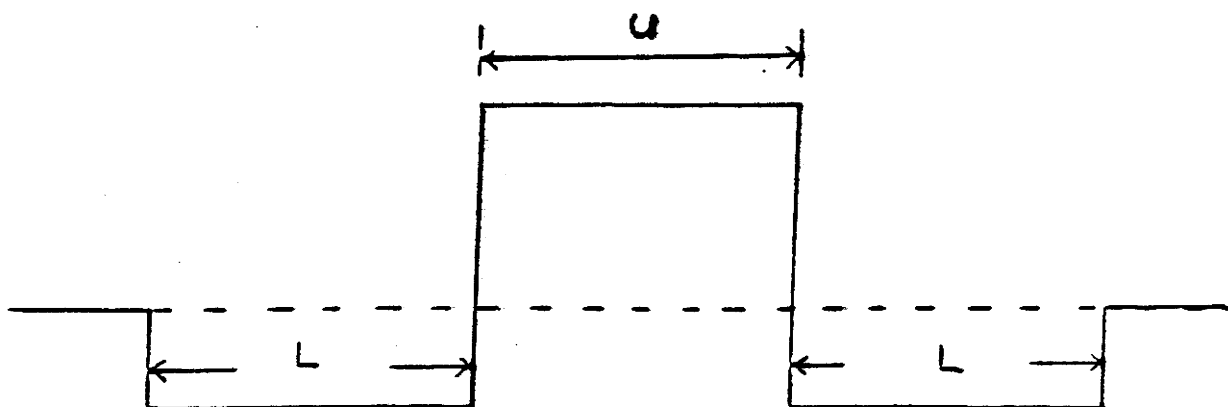
the optimal signal/noise ratio (S/N) in the correlated
spectrum, is realized if U ≃ L. Hence in the present study
U = L was chosen. In general, as L increases with respect
to U, the S/N ratio increases but the resolution decreases.

Figure 5.23 was obtained by correlating a gaussian
of fixed FWHM with a series of filters of varying widths
U=L. It suggests that there exists a filter for which the

Minor peak

Curve A

Curve B

Discrimination levels

Curve B

Magnified
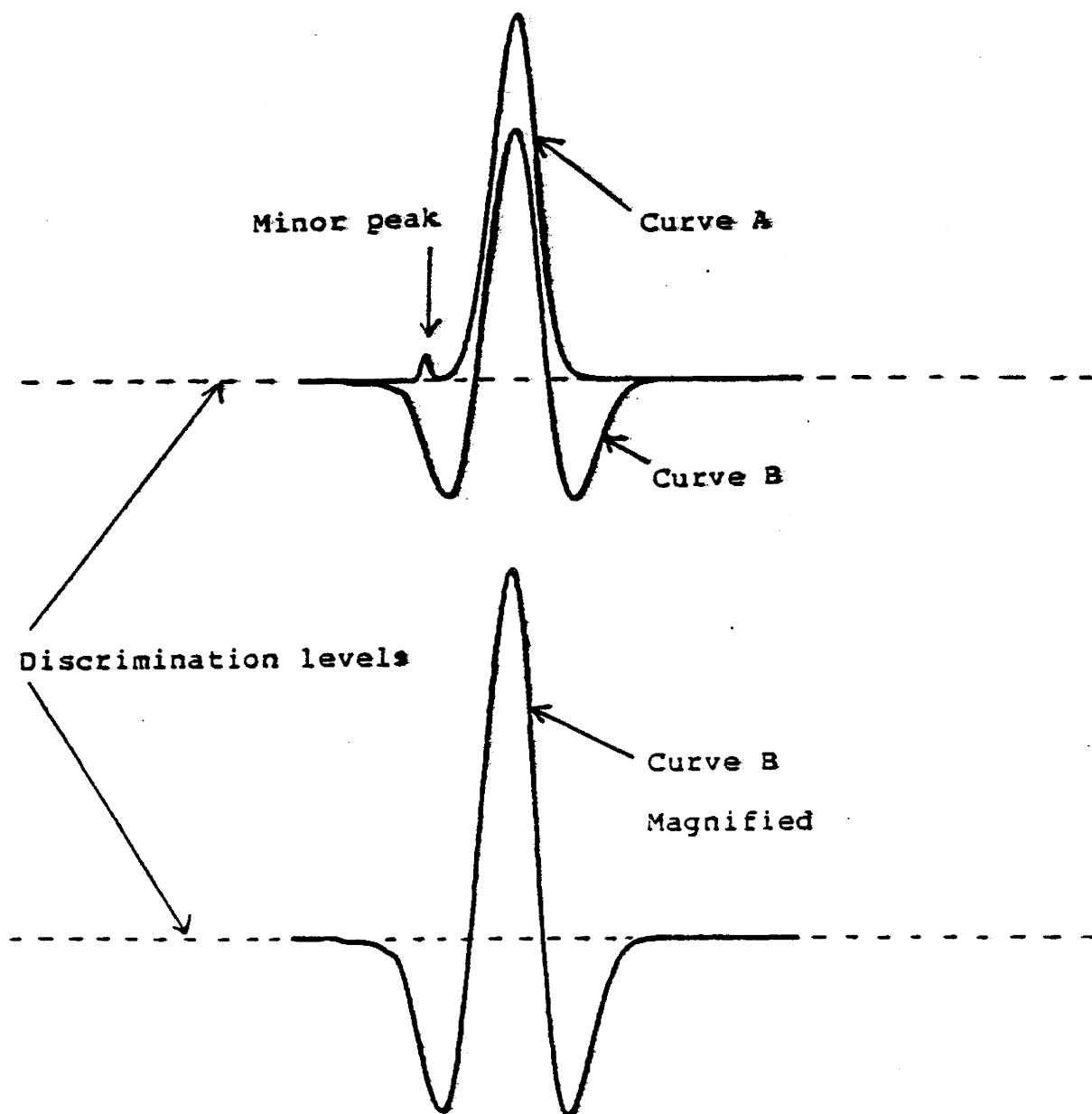
FIGURE 5.22

Effect of zero are filtering in the case where a very small peak resides close to a large one. Curve A is the original spectrum, and curve B is the correlation spectrum.

PEAK HEIGHT VS. FILTER WIDTH

Point of optimal
signal/noise ratio

ZERO AREA FILTER WIDTH IN UNITS OF 2σ (IE. $U/2\sigma$)
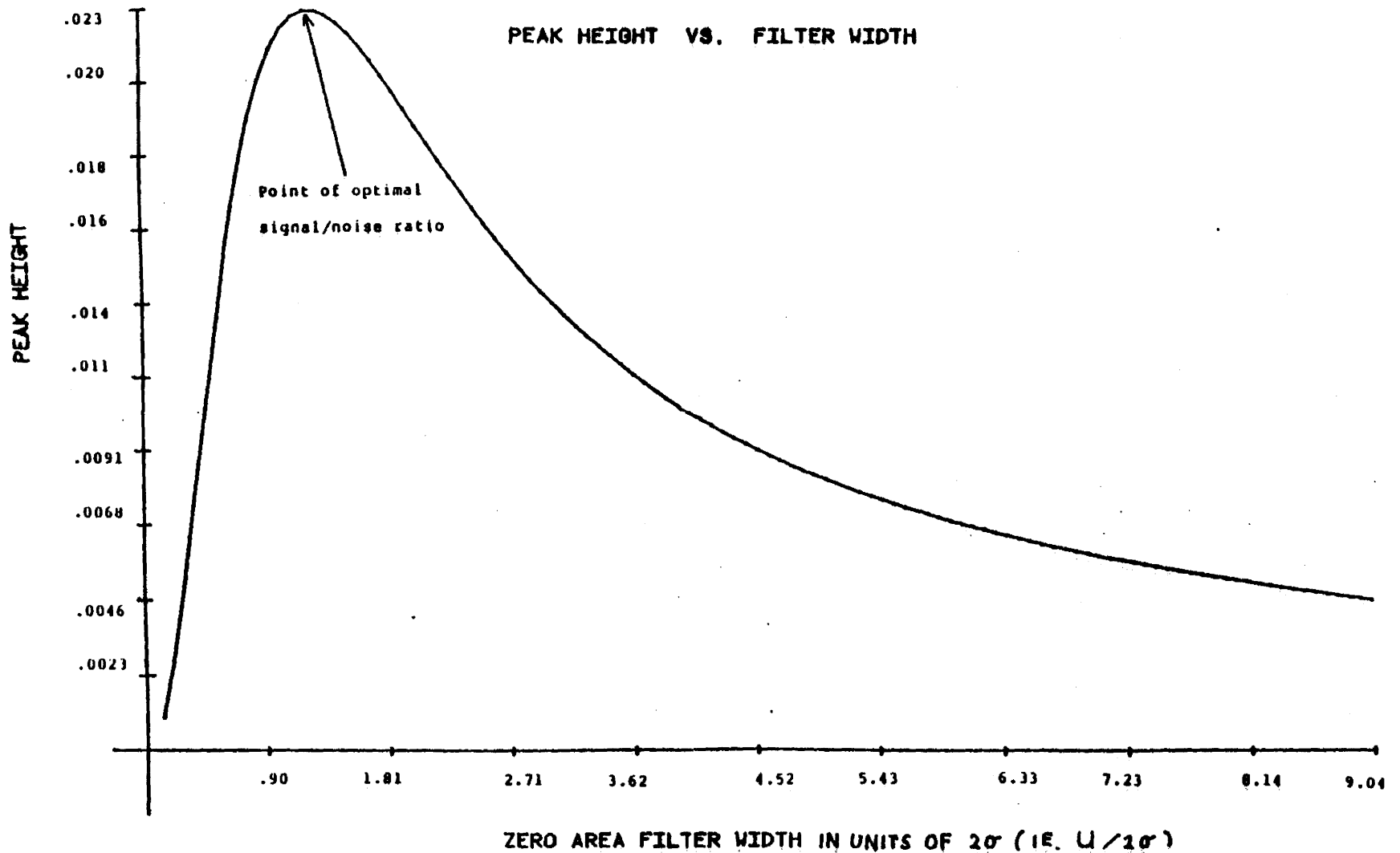
FIGURE 5.23

Plot of the peak height H of a filtered gaussian peak versus the filter width U. Note a maximum point on the curve exists where the signal/noise ratio is optimal.

height of the positive lobe of the correlated signal and hence the S/N is a maximum. Therefore for a given GC spectrum, an approximation of the peak widths should be obtained in order to determine the optimal filter.

d. Peak Area Correction Method

This method involves the measurement of the zero crossing width W of each of the peaks in the correlation spectrum. Once W is obtained, the appropriate correction ratio is assigned by which the positive lobe area of each peak must be multiplied in order to obtain the true area. By measuring the width W in terms of 2σ = width at the $e^{-1/2}$ points, only one calibration curve is required. Values for the calibration curve were obtained by generating a series of gaussians with varying full widths half maxima and of known area A. These were then correlated with a rectangular zero area filter of fixed width U. The zero cross widths and areas A' were measured for each of the correlated peaks from which the appropriate correction ratio

$$R = A/A' = \frac{(actual\ area)}{(observed\ area)}$$

was obtained. Figure 5.24.A provides the area correction calibration curve. As the ratio of the cross width W to the

CORRECTION RATIO VS. CROSSOVER WIDTH IN UNITS OF 2σ

CROSSOVER WIDTH IN UNITS OF 2σ (ιε. w/2σ)
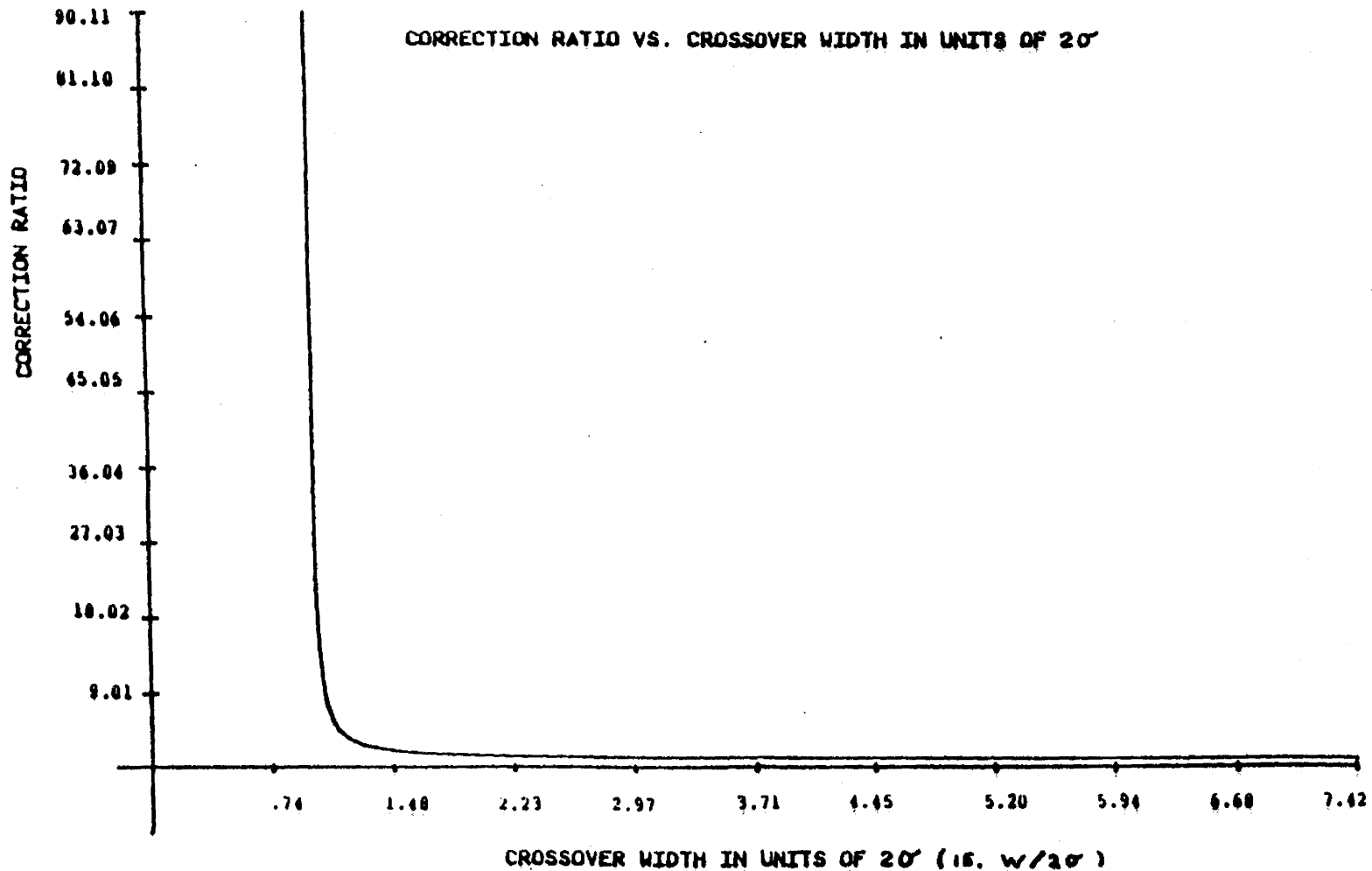
Figure 5.24.A

The area correction ratio versus the crossover width
W in units of 2σ .   Note the greatest variation of the
correction ratio occurs for crossover widths less that 1.48
times 2σ ,   (i.e. 1.7 times less than the FWHM of the
original gaussian).

2σ of the gaussian decreases, the positive lobe area decreases and hence the ratio A/A' increases. An alternative method for adjusting the positive lobe areas does not require the 2σ of the gaussian to be known but it does require one calibration curve for every filter used. The calibration curve consists of the correction ratio versus the crossover width W, refer to figure 5.24.B. Using a fixed filter width U and increasing the widths of the gaussians to be filtered, is the same as for Case (1) where a very narrow filter was used. Hence as gaussian widths increase, cross widths increase (where W = 2σ), and the positive lobe areas can be approximated by

$$A' = 2A \sqrt{\frac{2}{\pi e}} \quad \frac{1}{\sigma^2}$$

The correction ratio then becomes

$$R = A/A' = \frac{\sigma^2}{2\sqrt{\frac{2}{\pi e}}} = 1.0332\,\sigma^2$$

(ie. R is proportional to σ^2).

Linear interpolations were used when estimating the appropriate correction ratios from the calibration curves of R versus W
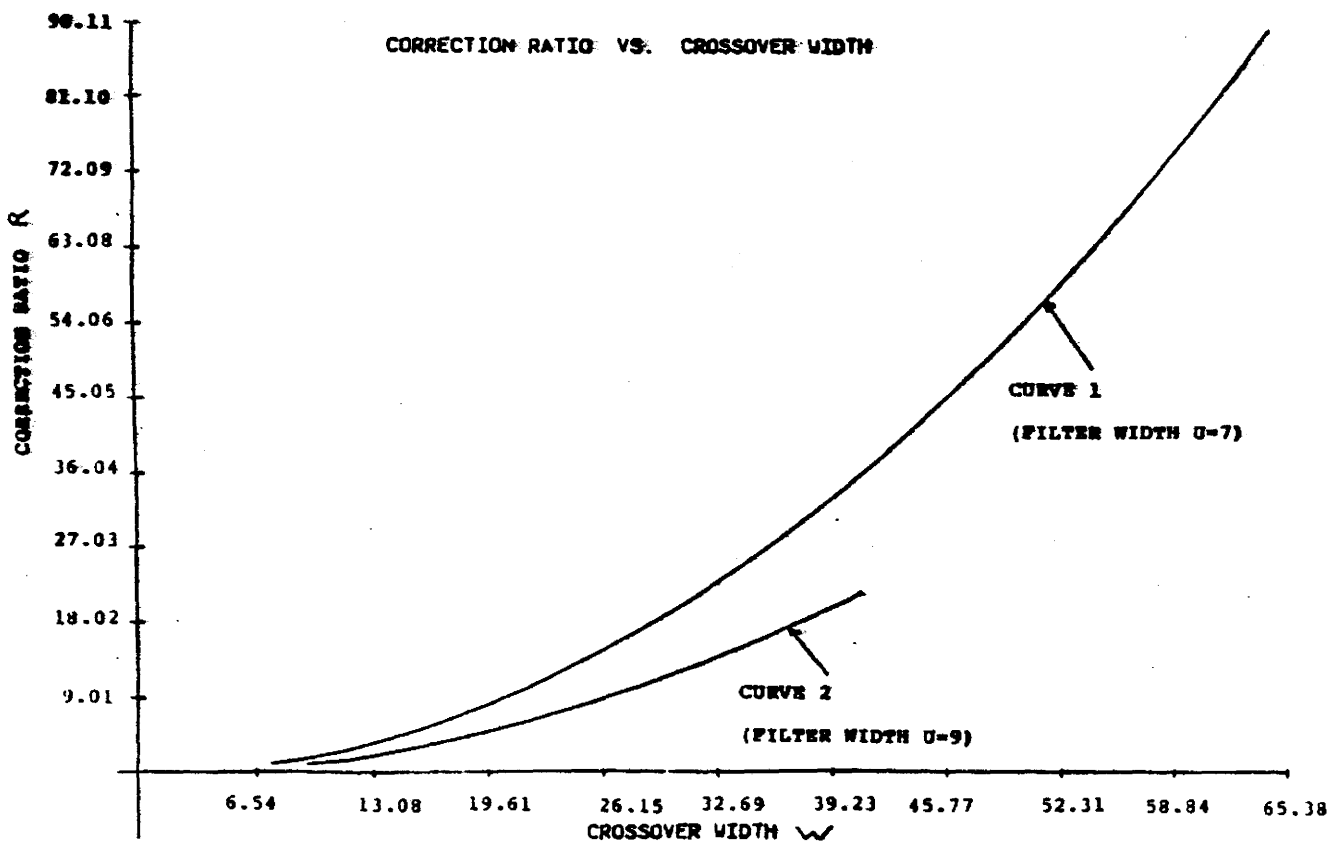
FIGURE 5.24.B

Calibration curves for the area corrections of filtered peaks. The ratio R=A'/A where A' is the positive lobe area of the filtered peak and A is the actual area of the original peak. Curve 1 is for a fixed filter width of U=7, and curve 2 is for a fixed filter width of U=9.

Figure 5.25 illustrates the effect of this area correction. Shown are two intensity vs. location graphs, one with no area correction and the other with the correction. The analysis was done for a typical GC spectrum, (A. Antitratus in this case). Table 5.2 provides the actual values for the intensities and retention times of these two graphs.

Peak Intensity vs. Retention Time

Plot 1
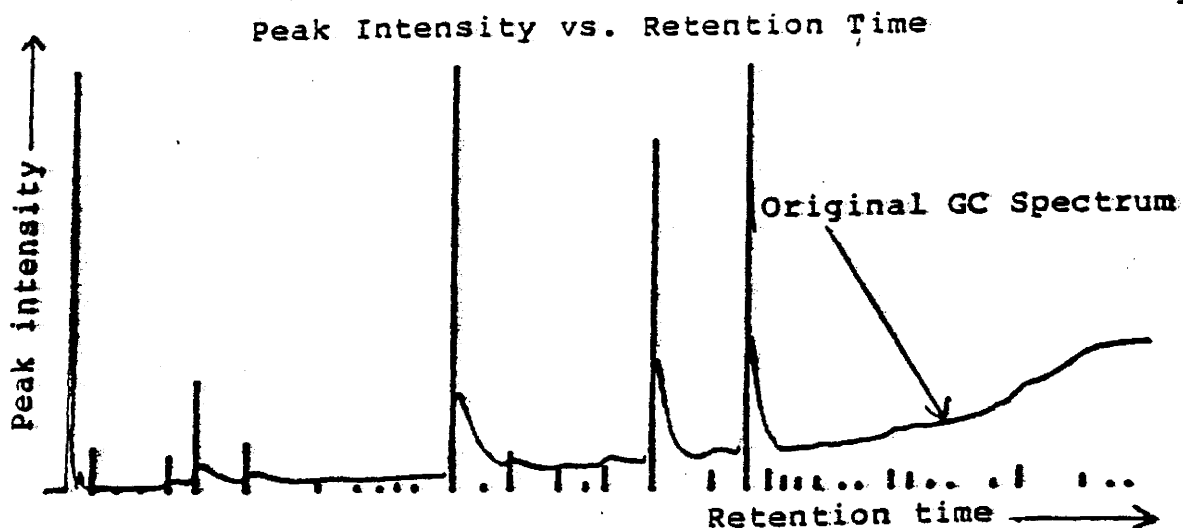
(From data obtained with no area correction)

Plot 2
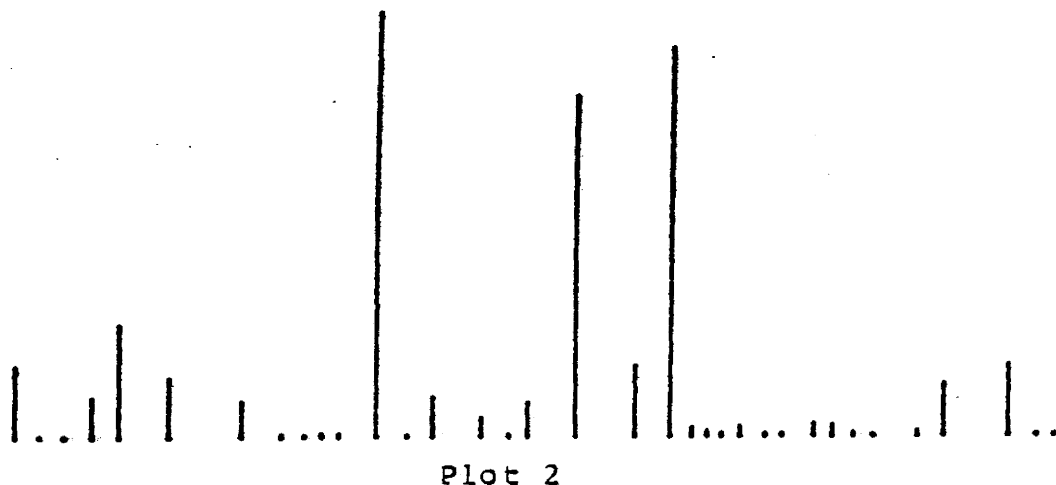
(From data obtained with the area correction)

FIGURE 5.25

The effect of the area correction technique on the correlation spectrum produced by a constant width correlator. The above spectrum is for A. Antitratus (a typical GC spectrum), with the actual values for the intensities and retention times listed by table 5.2.

RETENTION TIME	OBSERVED AREA	CROSSOVER WIDTH	CORRECTION RATIO	ACTUAL AREA
47.64	8516.94	13.72	3.90	33201.20
71.91	125.89	9.51	1.76	221.47
97.62	69.72	8.48	1.35	94.42
124.55	6762.09	11.55	2.72	18384.00
152.18	21923.40	10.92	2.39	52461.60
203.41	9270.12	12.01	2.96	27416.50
276.60	1352.67	24.08	12.22	16532.30
315.63	350.41	12.25	3.08	1080.12
338.59	142.64	8.45	1.34	191.63
356.91	777.45	9.50	1.75	1363.76
375.34	580.71	10.92	2.39	1388.73
412.88	85572.50	10.74	2.31	198040.00
444.78	547.50	7.66	1.13	617.10
470.16	6991.21	11.39	2.64	18441.00
519.98	3583.10	11.05	2.46	8808.38
548.27	990.66	7.77	1.16	1152.41
568.62	4116.65	13.66	3.86	15901.00
616.90	70643.90	10.58	2.24	158492.00
677.39	3030.39	22.72	10.85	32889.90
713.01	85466.50	10.30	2.12	180938.00
736.25	3363.51	7.76	1.16	3094.78
751.78	1884.87	7.75	1.16	2178.57
766.38	1261.51	8.24	1.28	1620.66
784.63	1826.45	10.48	2.20	4018.34
811.13	416.23	10.38	2.16	897.25
828.31	643.87	8.64	1.40	903.68
860.68	2878.24	9.77	1.88	5414.56
878.88	2324.68	10.26	2.10	4877.55
900.96	718.81	10.21	2.08	1494.75
920.05	237.37	7.51	1.08	255.36
965.41	952.39	10.58	2.24	2134.35
992.29	3519.85	18.08	6.85	24125.50
1056.96	1488.30	32.26	21.93	32638.60
1085.39	143.06	11.35	2.61	373.87
1102.93	96.47	8.57	1.38	133.01

TABLE 5.2

Intensity and retention time data for a typical GC spectrum (A. Antitratus). Note how the actual peak areas (ie. intensities) vary from the observed ones.

CHAPTER VI

DATA ANALYSIS

6.1   Retention Times

a.   Overview

After "de-skewing" and removing drifting baselines,
quantities such as retention time, the interval between
injection and detection and area are more conveniently
obtained.   There are three principal sources that contribute
to the dispersion observed in GC retention times.   These are
variations in the GC system, variation in chemical
preparation and the non-identity of the bacterial strain.
There are many factors that contribute to retention time
variation within the GC system.   As mentioned in section
3.2, peak skewing retards the retention time, hence all
mechanisms leading to peak skewing have an influence upon
this quantity. Large samples alter retention volumes and
therefore retention times (17), also the purity of the phase
influences the ability to duplicate results. However, the
greatest contributor to variation in retention time is
changing column conditions.

To obtain a measure of retention time variance
attributed to factors within the GC system, a mixture of
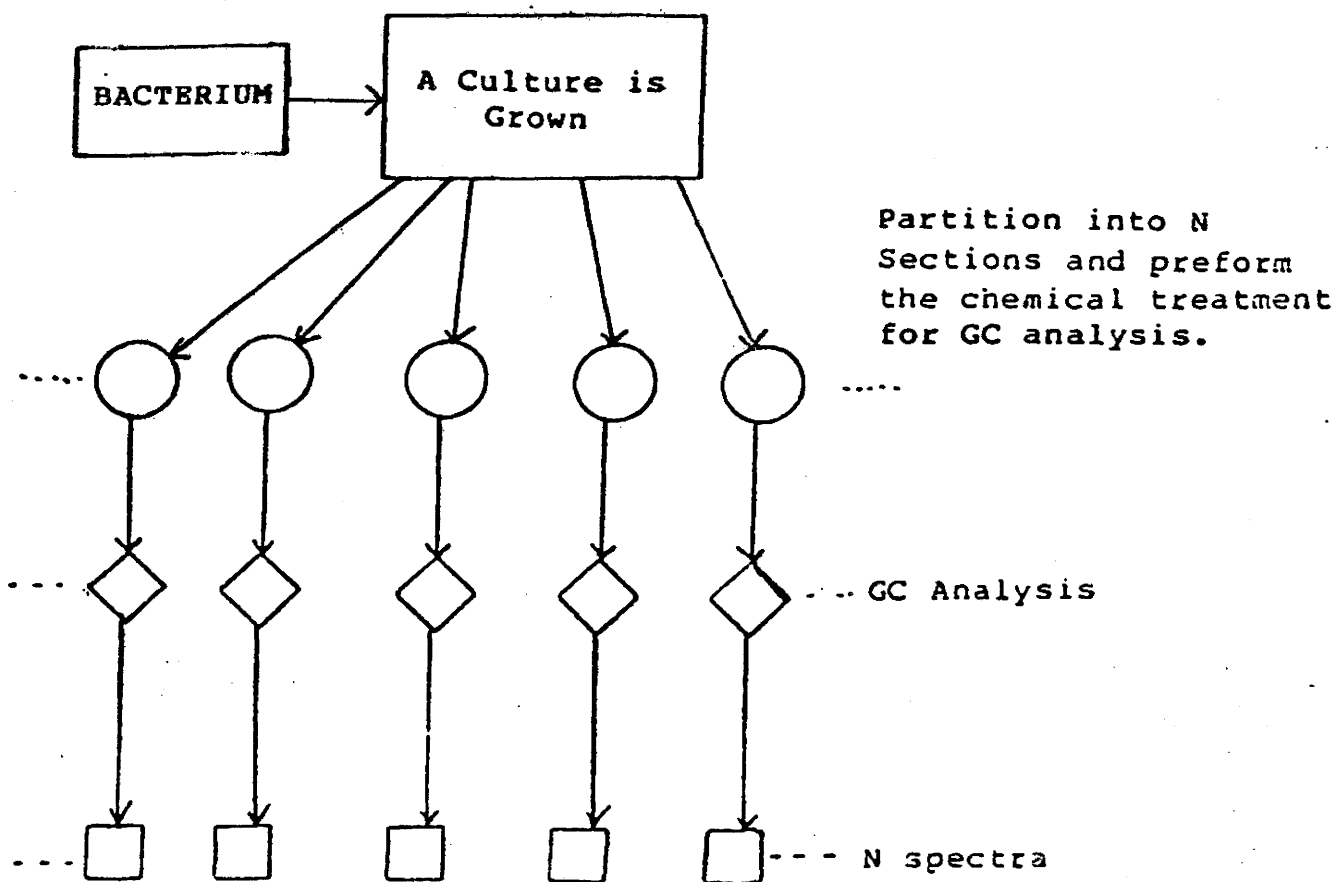
FIGURE 6.1

Process for obtaining N spectra which could then be

analysed to determine the lowest total variance limit due to

the GC system, chemical preparation and the "bug" itself.

prepared chemical standards consisting of pentafluorobenzyl esters (PFB esters) of carboxylic acids with 7 to 20 carbon atoms were injected and the spectrum obtained and analysed. Specific details are given in section 6.1 (c). This type of study produces only a lower limit of retention time variation since samples are identical, thereby displaying no dispersion for preparation or the non-identity of the bacterial strain. Similarly to obtain a measure of the total variance (involving the GC system, chemical variance and bacterial strain variance) a series of "bug" samples were prepared as illustrated in figure 6.1 adjusted and then tested to determine the lower limit for retention time variances.

b.   Offset And Slope Adjustment In Spectra

The retention times for several sets of standards were measured. If the system is preforming in a linear fashion then the emission of a substance with carbon number n will be

$$t(^nC) = t_o + \alpha\, t.$$

Here for the range of n of interest ($7 \longrightarrow 20$),   is linearly related to n. Both t_o and α are dependent upon sample preparation, size and column condition. For sets of

the same material comparison of the parameters t_0 and suggested both changed between runs. In order to minimize retention time dispersion, this type of variation can and should be removed. The method used to accomplish this proceeded as follows.

The retention times of spectrum 1 vs. the retention times of spectrum 2 (the compared spectrum) were initially plotted and a linear least squares fit was performed, (refer to figure 6.2). If the retention times between spectra were identical, then a line with a 45 degree slope and an intercept at the origin. If a change in α had occured, then the slope would not be 45 degrees. Similarly if a change in t_0 was present, then the defined line would not pass through the origin.

The change in α could be corrected by rotating the line along with the data points such that the slope equals 45°. The offset, t_0, was corrected by shifting the line to the origin. Numerically the offset and slope can be corrected as follows, (refer to figure 6.2).

$$P'_i = P_i + A$$

where P'_i = gain corrected point

P_i = original data point

A = correction term.

The slope of the original line is $S_2 = B/D \implies B = S_2 D$, and the slope of the corrected line is $S_1 = C/D \implies C = S_1 D$ therefore

$$A = C - B = S_1 D - S_2 D = D(S_1 - S_2)$$

and $P'_i = P_i + D(S_1 - S_2)$

To correct for the shift, the y intercept is subtracted therefore $P''_i = P'_i - I$ where P''_i are the shift and gain corrected points. This method is for an ideal case which requires that the peaks in both spectra are already identified, hence it provides the lowest limit for retention time variances.

c.   Determination Of The Minimal Retention Time
      Variance

After the spectra have been obtained and the offset and slope adjusted, the retention times of similar peaks were compared and the means and variances calculated. The observed mean and variance was calculated using the following equations:

$$\text{Mean} = \frac{\sum x_i}{N} = \overline{x} \qquad \text{where } i = 1,2,3,\ldots\ldots N$$

$$\text{Variance} = \sum \frac{(x_i - \overline{x})^2}{N-1}$$

where N-1 = the degrees of freedom.

It is usual to use the standard deviation (SD) to indicate by how much the retention times can vary (SD= $\sqrt{\text{variance}}$). After obtaining the means and standard deviations, an analysis was conducted to determine if this latter quantity was time dependent.

RETENTION TIMES

	PART A					PART B		
PEAK #	STD. #1	STD. #2	STD. #3	STD. #4	STD. #1	STD. #2	STD. #3	STD. #4
1	172.66	174.76	172.89	170.47	172.66	173.24	172.58	173.37
2	265.89	267.32	266.16	263.37	265.89	265.93	265.92	265.85
3	319.75	321.00	319.90	317.44	319.75	319.68	319.70	319.68
4	375.97	377.04	376.20	373.58	375.97	375.80	376.04	375.58
5	432.36	433.33	432.58	430.09	432.36	432.17	432.47	431.83
6	488.50	489.36	488.61	486.76	488.50	488.28	488.54	488.28
7	544.00	544.80	544.01	542.73	544.00	543.80	543.98	543.98
8	596.46	599.13	598.46	597.47	598.46	598.20	598.47	598.48
9	651.24	651.99	651.18	650.72	651.24	651.14	651.23	651.50
10	703.25	703.96	703.06	702.78	703.25	703.18	703.15	703.33
11	753.61	754.45	753.49	753.46	753.61	753.73	753.62	753.79
12	802.85	803.73	802.75	802.80	802.85	803.09	802.91	802.90
13	850.67	851.55	850.40	850.79	850.67	850.97	850.60	850.68

	MEAN RETENTION TIME	STANDARD DEVIATION	MEAN RETENTION TIME	STANDARD DEVIATION
1	172.70	1.75	172.96	0.40
2	265.68	1.67	265.90	0.037
3	319.52	1.50	319.70	0.031
4	375.70	1.48	375.85	0.21
5	432.09	1.40	432.21	0.28
6	488.31	1.09	488.40	0.14
7	543.89	0.86	543.94	0.097
8	598.38	0.68	598.40	0.13
9	651.28	0.53	651.28	0.15
10	703.28	0.50	703.23	0.077
11	753.75	0.46	753.69	0.086
12	803.03	0.47	802.94	0.10
13	850.85	0.49	850.73	0.16

TABLE 6.1

Part A provides the retention times, standard deviations and the means for similar peaks between 4 sets of chemical standards before any gain or shift corrections have taken place. Part B provides the same information as part A except that in this case the gain and shift adjustments have been preformed.

Four sets of standards were analysed by preforming a least squares fit of three sets to one reference set. It was found that in general the slope corrections were considerably less than the offset corrections. This is evident from the values of table 6.2, for which the gain corrections are proportional to $|1-B|$ and the shift corrections are proportional to A

The retention time of thirteen peaks for the four sets of standards were compared. Table 6.1 part A lists these retention times, the means and the standard deviations, before the gain and shift corrections. Part B lists these values after the gain and shift corrections have been made. It is evident that the standard deviations are considerably reduced for part B. Table 6.2 provides the parameters for the least squares fit used in the shift and gain corrections.

Figure 6.3 illustrates how the standard deviations versus the mean of the peak retention times varies. A line was fit to this data by performing a least squares fit. Table 6.3 provides the parameters of this fit. It was observed that a very small slope existed but the error of the slope was so large that it could be zero within error. The evidence, in general suggests it is safe to assume a constant standard deviation of peak retention time with

Y INTERCEPT A	SLOPE B	% ERROR IN THE Y INTERCEPT	% ERROR IN THE SLOPE
1.76	0.99	11.33	0.035
0.44	0.99	11.26	0.0087
-3.66	1.00	6.66	0.042

TABLE 6.2

Parameters of the linear least squares fit to the retention data of three different sets of chemical standards plotted against a "reference" set of standards. All values are in units of sampling intervals.

Y INTERCEPT A	SLOPE B	% ERROR IN THE Y INTERCEPT	% ERROR IN THE SLOPE
0.24	-1.68	32.45	79.68

TABLE 6.3

Parameters for the least squares fit of the standard deviations vs. the mean of the retention times, (for the set of chemical standards). All values are in units of sampling intervals.

STD. DEVS. VS MEAN

STANDARD DEVIATION (in units of sampling intervals)

MEAN RETENTION TIME

(in units of sampling intervals)
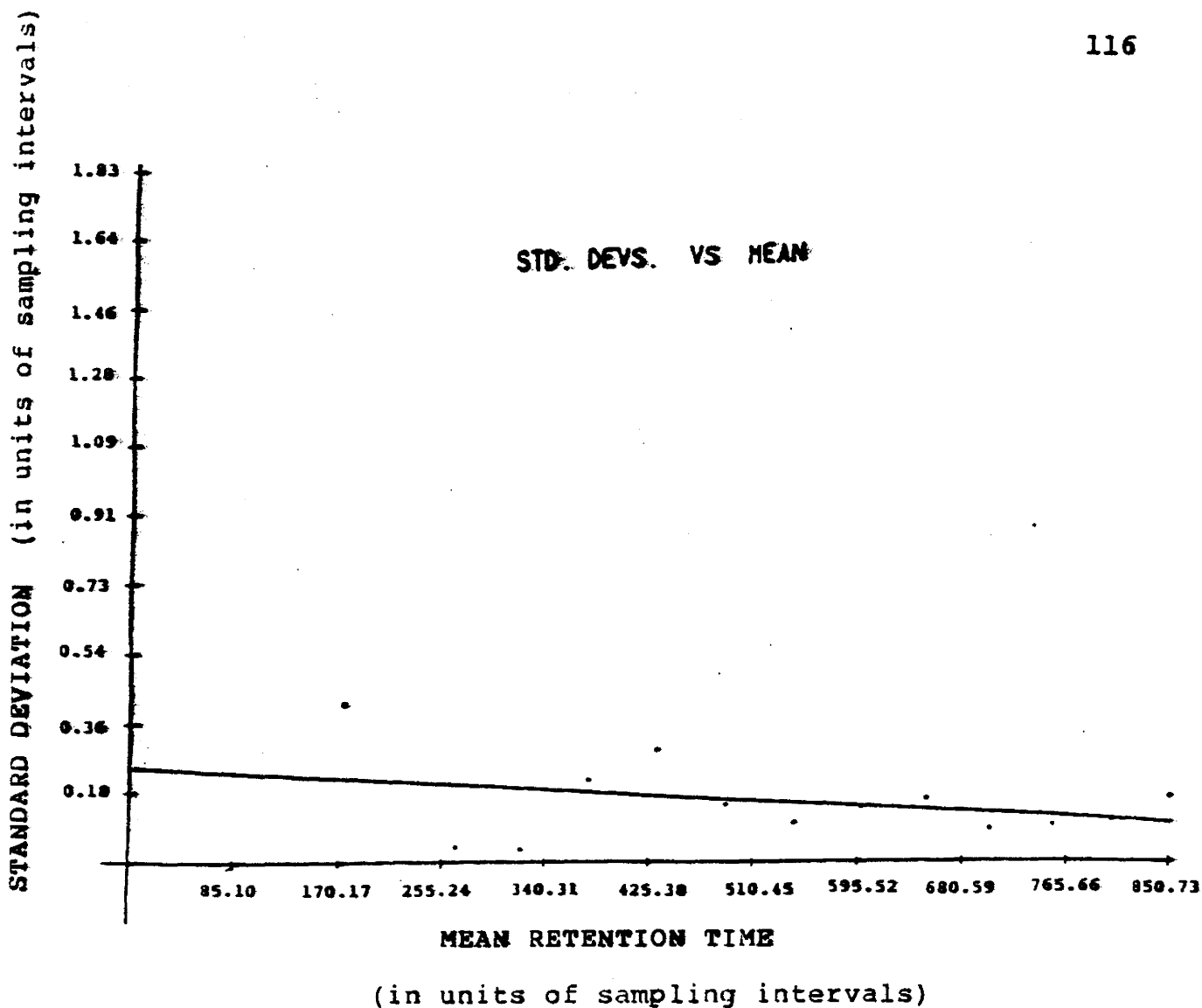
FIGURE 6.3

Plot of standard deviations vs. the means for retention time data of four sets of chemical standards. This demonstrates the minimum retention time varience due to the GC system. The error in the slope is 79%, hence it could be a constant.

respect to its location.

This constant standard deviation was obtained from table 6.1 part B. by taking the average of the standard deviations. The lowest limit for the standard deviation of the retention times for the set of standards (hence the GC system variance) was 0.15. Since the sampling rate was once every 1.015 seconds, this translates to 0.0025 minutes.

A similar analysis was performed for five spectra of a "bug" (E.Coli), demonstrating this minimum total variability in peak retention time due to all sources. Refer to table 6.4.A, 6.4.B, 6.5, 6.6, and figure 6.4. The lowest limit for the standard deviation of the retention times was found to be .86 which in turn translates to .015 minutes.

In summary the minimum peak retention time error (i.e. standard deviation) was;

.0025 minutes ——→ due to the GC system

.015 minutes——→ total variance due to all effects.

6.2   Peak Areas

a.   Overview

The following procedure was used to analyse the constancy of peak areas. The spectra was "de-skewed", filtered, and the positive lobe areas were compensated

RETENTION DATA

Peak #	E.Coli #1	E.Coli #2	E.Coli #3	E.Coli #4	E.Coli #5
1	105.78	107.19	109.88	106.53	106.16
2	132.75	131.60	135.61	133.03	132.57
3	181.63	180.28	183.54	181.66	180.73
4	242.74	243.30	243.89	242.69	240.82
5	264.99	265.67	266.12	263.65	263.34
6	294.34	292.16	295.90	294.06	292.34
7	391.54	387.07	393.56	390.50	390.14
8	454.71	447.87	452.30	454.17	452.05
9	495.11	491.94	497.51	494.12	494.85
10	546.68	544.48	548.70	545.85	547.25
11	597.41	595.42	600.00	596.86	599.02
12	629.40	627.73	630.64	628.54	631.71
13	651.12	649.47	652.82	650.81	653.92
14	700.60	699.09	701.91	700.16	703.89
15	752.34	751.27	752.99	752.01	757.22
16	799.33	798.78	799.52	798.89	804.96
17	864.84	863.47	864.68	864.27	870.73

Peak #	Mean Retention Time	Standard Deviation
1	107.1	1.6
2	133.1	1.4
3	181.5	1.2
4	242.6	1.1
5	264.7	1.2
6	293.7	1.5
7	390.5	2.3
8	452.2	2.6
9	494.7	2.0
10	546.5	1.5
11	597.7	1.8
12	629.6	1.5
13	651.6	1.7
14	701.1	1.8
15	753.1	2.3
16	800.2	2.6
17	865.5	2.9

TABLE 6.4.A

Retention data for five spectra of E. Coli before offset and slope corrections. All values are in units of sampling intervals.

RETENTION DATA

PEAK #	E.COLI #1	E.COLI #2	E.COLI #3	E.COLI #4	E.COLI #5
1	105.78	108.31	107.46	106.65	108.67
2	132.75	132.76	133.26	133.17	134.83
3	181.63	181.51	181.32	181.84	182.52
4	242.74	244.62	241.83	242.93	242.04
5	264.99	267.03	264.12	263.91	264.36
6	294.34	293.56	293.98	294.34	293.08
7	391.54	388.62	391.91	390.87	389.97
8	454.71	449.52	450.81	454.59	451.29
9	495.11	493.63	496.13	494.58	493.71
10	546.68	546.26	547.46	546.35	545.62
11	597.41	597.28	598.90	597.40	596.93
12	629.40	629.63	629.62	629.11	629.31
13	651.12	651.41	651.86	651.41	651.32
14	700.60	701.10	701.09	700.79	700.83
15	752.34	753.36	752.30	752.69	753.68
16	799.33	800.94	798.96	799.61	800.97
17	864.84	865.73	864.29	865.05	866.13

PEAK #	MEAN RETENTION TIME	STANDARD DEVIATION
1	107.37	1.18
2	133.35	0.85
3	181.76	0.46
4	242.83	1.10
5	264.87	1.26
6	293.85	0.54
7	390.57	1.32
8	452.18	2.34
9	494.63	1.05
10	546.47	0.66
11	597.58	0.76
12	629.41	0.21
13	651.42	0.27
14	700.88	0.21
15	752.87	0.61
16	799.96	0.93
17	865.20	0.72

TABLE 6.4.B

Retention data for five spectra of E. Coli after slope and offset corrections. All values are in units of sampling intervals.

Y INTERCEPT	SLOPE	% ERROR IN THE Y INTERCEPT	% ERROR IN THE SLOPE
-0.96	0.99	111.86	0.20
2.71	0.99	26.37	0.13
-0.03	0.99	857.15	0.05
-3.49	1.01	25.54	0.17

TABLE 6.5

Parameters for the linear least squares fit to the retention data of four sets of E. Coli spectra plotted against a reference E. Coli. All values are in units of sampling intervals.

Y INTERCEPT	SLOPE	% ERROR IN THE Y INTERCEPT	% ERROR IN THE SLOPE
1.17	-6.66	23.92	79.30

TABLE 6.6

Parameters for the least squares fit to the standard deviations vs. the mean of the retention times, (for the set of E. Coli spectra). All values are in units of sampling intervals.

STD. DEVS. VS MEAN

STANDARD DEVIATION

2.92
2.63
2.33
2.04
1.75
1.46
1.17
0.87
0.58
0.29

86.52   173.04   259.56   346.08   432.68   519.12   685.64   692.16   778.68   865.28

MEAN RETENTION TIME
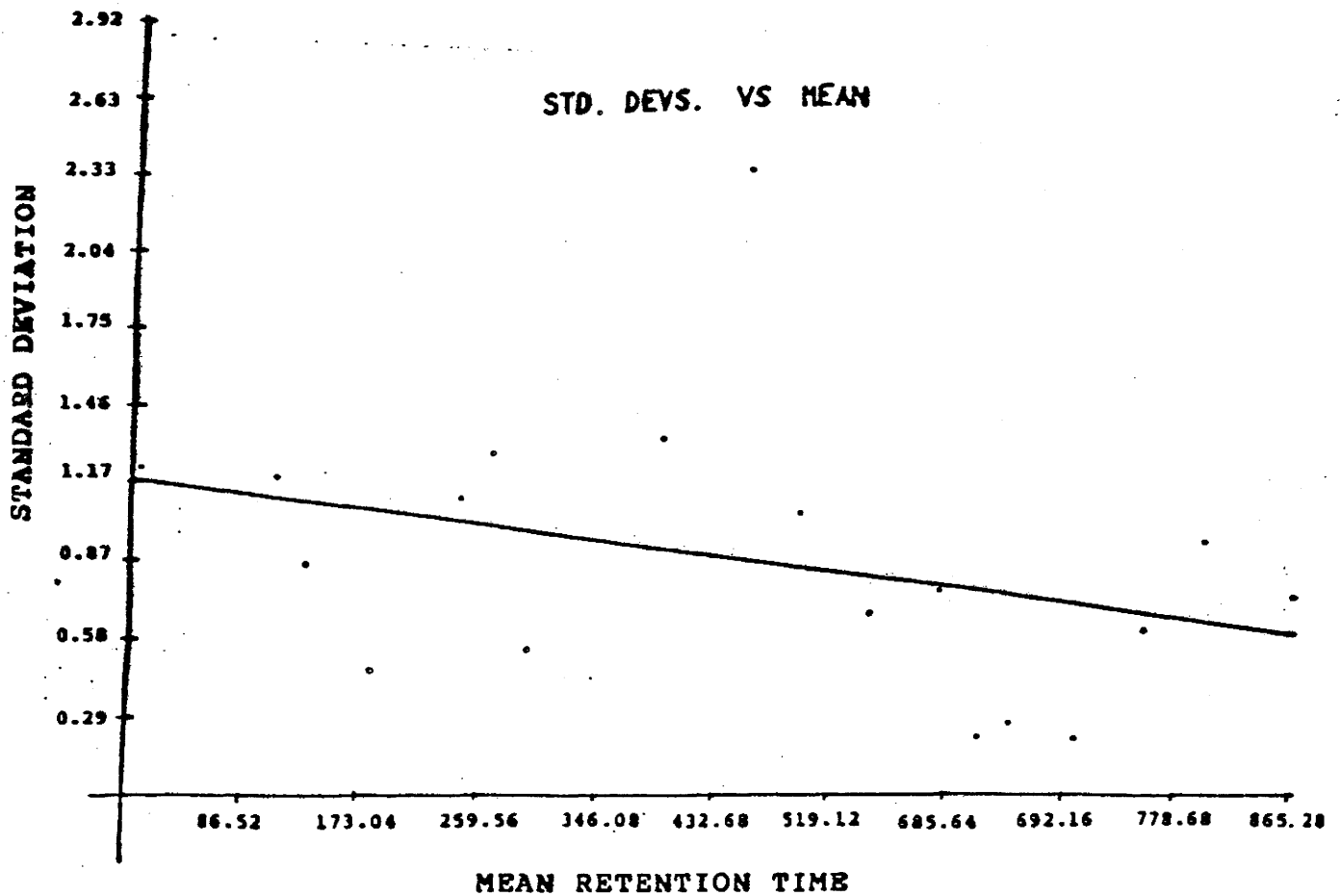
FIGURE 6.4

Plot of the standard deviations vs. the means for retention time data of five sets of E. Coli spectra. This illustrates the total minimum retention time variance. All values are in units of sampling intervals.

according to the crossover widths in the correlated spectrum. Each peak was then normalized by dividing by the total area of the spectrum, thereby determining each peak's area as a fraction of the total spectral area. These values were then compared for several spectra of the same material. Unlike the data used in the analyses of retention time, here only a shift correction had to be made. This was because when comparing normalized values between spectra, only deviation about the mean area was required.

As was done for retention time analysis, the offset was corrected for by first plotting a spectrum pair against one another, fitting the data by a linear least squares regression and then translating the points so that the Y intercept of the line passed through the origin (refer to figure 6.5). Initial peak identification is obtained from retention time analysis.

b. Determination Of The Minimal Peak Area Variance

Four spectra of E. Coli were compared to determine how the the peak areas varied. The area data for identical peaks between different spectra were compared, the standard
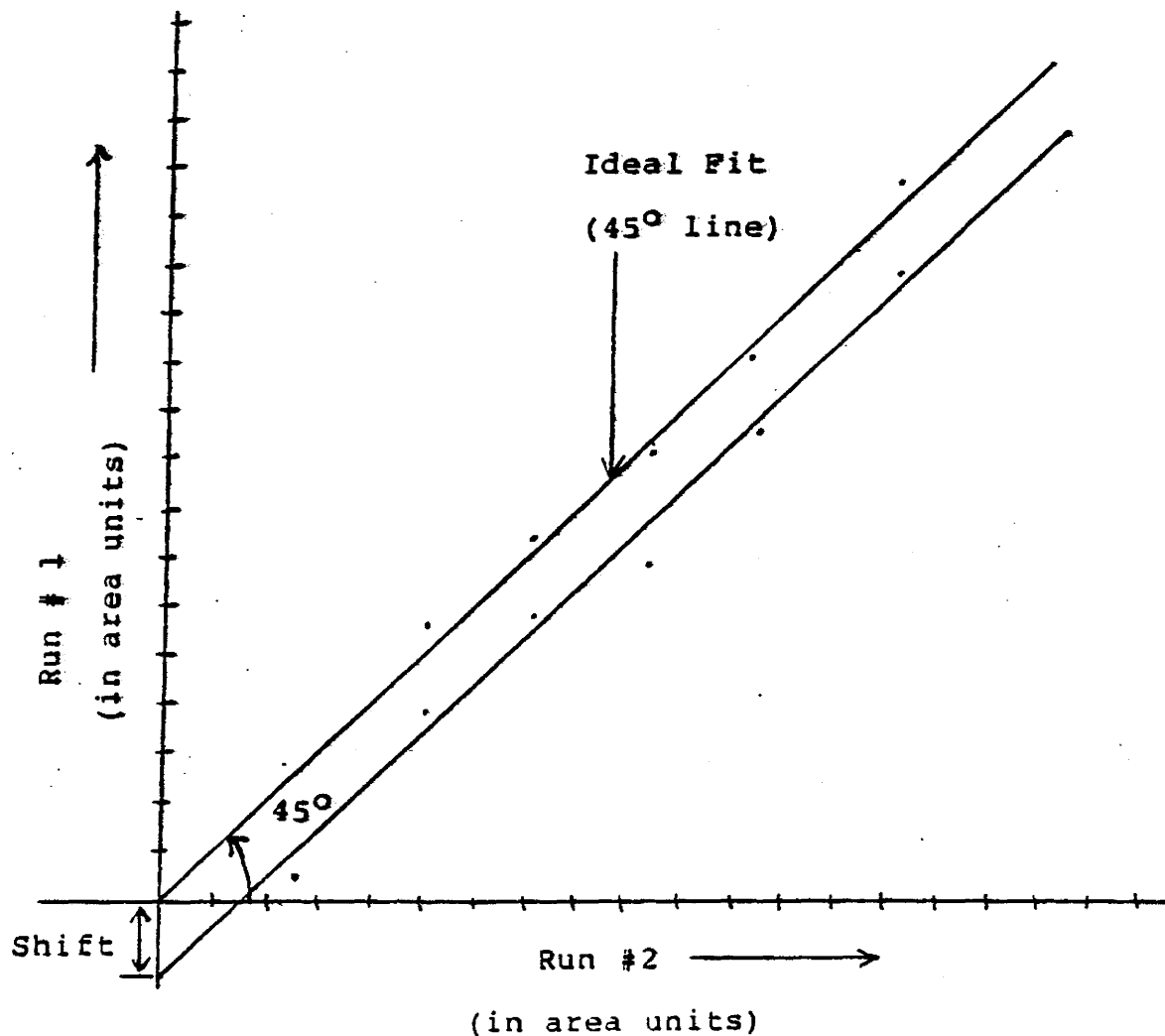
FIGURE 6.5

Shift correction for determining peak area variances. A linear least squares regression was preformed' followed by a shift such that the y intercept was equal to zero.

deviation and then the percentage error calculated as a function of peak location (i.e. time) . The standard deviation or the square-root of the variance was used in determining the percentage error for peak P_i via.,

$$\% \text{ Error} = E_i = \frac{(\text{Std. deviation of peak } P_i)}{(\text{Mean area of peak } P_i)} \times 100$$

Table 6.7.A and 6.7.B present the results obtained before and after a shift correction has been made. In general the standard deviation associated with each peak area seemed to be proportional to the magnitude of the peak area and not to the location. In other words

$$\frac{\sigma}{A} \propto A \qquad \text{therefore } \sigma = k = \text{constant.}$$

In order to determine this relationship, the variation of E_i with P_i was analysed. The percentage error E_i versus the mean area of peak P_i (for i = 1,2,.....N) was plotted and a linear least squares fit preformed (refer to figure 6.6). Since the errors of the line parameters in this fit were large, the percentage error can be considered to be a constant. Therefore with the assumption that the percentage error was constant with respect to location, the mean of the percentage error values was used as the lower limit for

% OF THE TOTAL SPECTRAL AREA

PEAK #	E.COLI #1	E.COLI #2	E.COLI #3	E.COLI #4
1	8.02	6.72	7.87	8.76
2	8.48	7.40	8.49	8.73
3	9.76	9.20	9.76	10.26
4	8.70	8.37	8.75	8.87
5	8.15	7.99	8.46	8.46
6	8.50	8.77	8.62	8.61
7	8.54	8.59	8.53	8.32
8	6.09	6.40	6.14	5.98
9	10.93	11.84	10.94	10.61
10	8.04	8.63	8.01	7.63
11	9.29	10.08	8.99	8.65
12	3.39	3.72	3.36	3.19
13	2.10	2.28	2.07	1.94

PEAK #	MEAN	STD. DEVIATION	% ERROR
1	7.84	0.84	10.76
2	8.28	0.59	7.18
3	9.74	0.44	4.48
4	8.67	0.21	2.46
5	8.26	0.23	2.81
6	8.63	0.11	1.26
7	8.50	0.11	1.40
8	6.15	0.18	3.01
9	11.08	0.53	4.77
10	8.08	0.41	5.12
11	9.25	0.61	6.59
12	3.42	0.22	6.52
13	2.10	0.14	6.71

THE MEAN OF THE STANDARD DEVIATIONS IS 0.36

TABLE 6.7.A

Peak area data used to determine how the normalized peak areas vary with respect to location (ie. time). The above results were from the unadjusted spectra. All data is in area units.

% OF THE TOTAL SPECTRAL AREA

PEAK #	E.COLI #1	E.COLI #2	E.COLI #3	E.COLI #4
1	8.02	6.61	7.87	8.93
2	8.48	7.29	8.49	8.89
3	9.76	9.08	9.76	10.43
4	8.70	8.26	8.75	9.03
5	8.15	7.88	8.46	8.62
6	8.51	8.66	8.62	8.77
7	8.54	8.48	8.53	8.49
8	6.09	6.29	6.14	6.13
9	10.93	11.72	10.93	10.77
10	8.03	8.52	8.01	7.79
11	9.29	9.96	8.98	8.81
12	3.39	3.61	3.36	3.35
13	2.10	2.16	2.07	2.10

PEAK #	MEAN	STD. DEVIATION	% ERROR
1	7.86	0.95	12.13
2	8.29	0.69	8.37
3	9.76	0.55	5.62
4	8.68	0.32	3.68
5	8.27	0.33	3.97
6	8.64	0.11	1.27
7	8.51	0.03	0.37
8	6.16	0.09	1.43
9	11.09	0.43	3.87
10	8.09	0.31	3.79
11	9.26	0.51	5.48
12	3.43	0.12	3.05
13	2.11	0.04	1.87

THE MEAN OF THE STANDARD DEVIATIONS IS   0.34

TABLE 6.7.B

Peak area data used to determine how the normalized peak areas vary with respect to location (i.e. time). The above data have undergone a correction for offset error. All the above data is in area units.

X ERROR   VS.   MEAN

Least Squares Fit Parameters:

Slope = .2847
Y Intercept = 2.067
% Error in the slope = 132.7
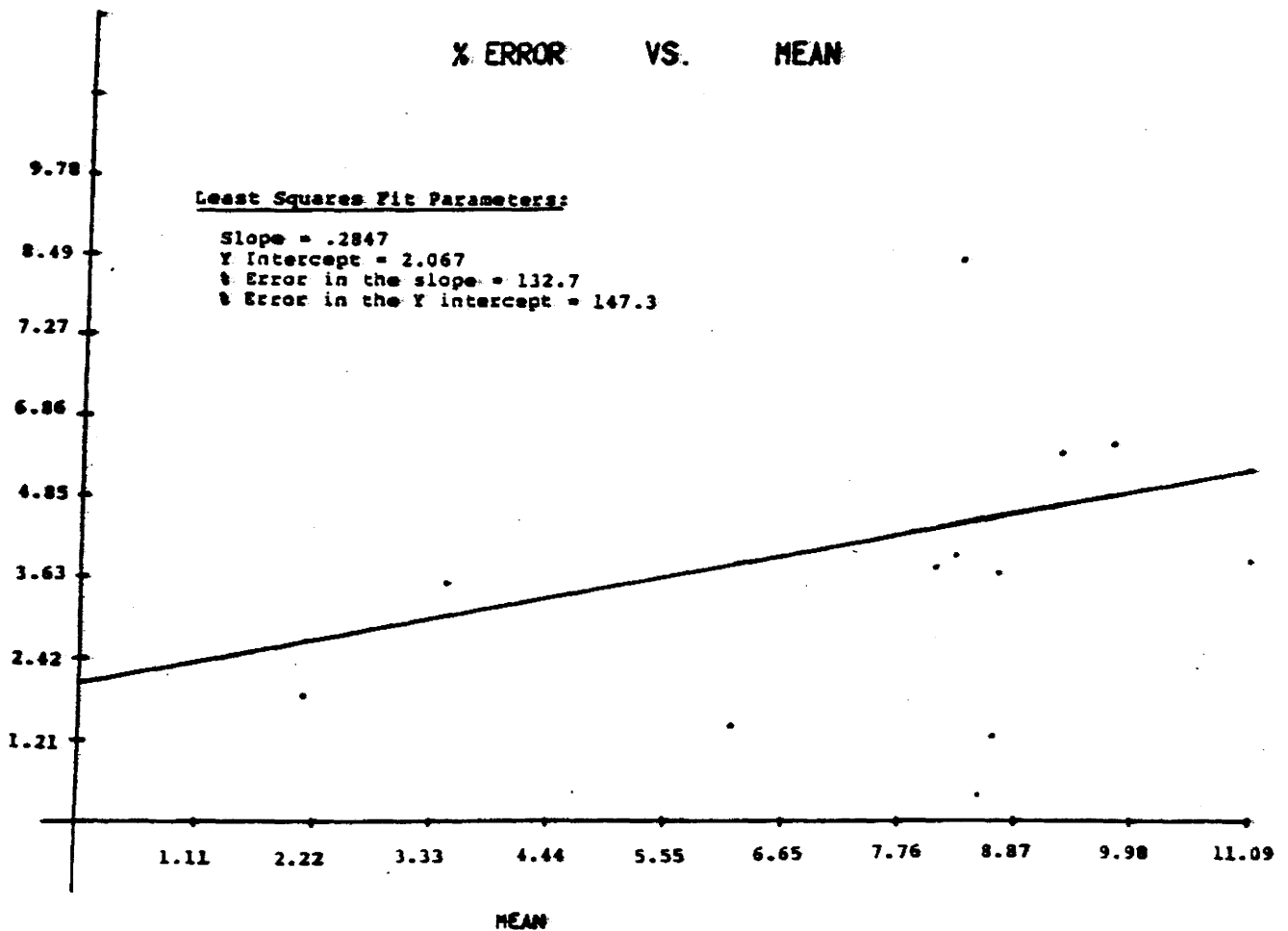% Error in the Y intercept = 147.3

MEAN

FIGURE 6.6

By comparing the normalized peak area data, the following plot was obtained. Since the error in the slope and y intercept of the plotted line is large, it is possible that the % error could be a constant with respect to peak location.

indicating how the peak areas varied. The value obtained

was, mean percentage error = 4.25 %.

6.3   Carbon Atom Scale

a.   Determination Of Linearity

Components with fewer carbon atoms elute sooner
than those with more carbon atoms. By heating, the value of
K (which is the fraction of vapour that condenses as a
longitudinally immobile compound) can be varied and hence
any particular species can be vapourized and thereby rapidly
proceed through the column. If the function

$$f(T) = \frac{dT}{dt} = \text{the temperature programming function}$$

is properly chosen,   where t is the time at which
temperature T occurs, the release of molecules will proceed
linearly with respect to their carbon number.

For the GC spectra used in the present study, the
linearity of such a carbon atom scale was experimentally
established. Several spectra of standards were examined,
for which the number of carbon atoms associated with each
component was known. For each of these,   retention times
were determined,   after all spectral interference and
destortion had been removed.   These values were inter
compared between spectra and the mean retention time for N-
carbon atom components (where N = 7,8,...28), determined.   A

plot of mean retention time versus the number of carbon atoms was obtained to which a linear least squares fit was performed. Figure 6.7 illustrates the results obtained.

It is evident that a linear relation between retention time and N exists. Changing the temperature programming in the GC system will alter the slope of the line but should not affect the linearity. In general it is good practice to obtain such a curve on a daily basis in order to assure consistent results.

In order to translate the retention times of the unknown spectra to the number of carbon atoms, an internal standard is required. Because the retention time and the number of carbon atoms of such a standard is known, a direct translation to a carbon atom scale is possible by simply using the slope of the calibration curve obtained.

b. Translation To The Carbon Atom Scale From
   Retention Ratios

Peak locations can also be represented in terms of retention ratios R' where $R' = R_1/R_2$, (refer to figure 6.8). An internal standard should be present in each spectrum in order for this method to be used. Similarly the retention ratios for each carbon atom peak can be obtained (refer to figure 6.8). Using the retention ratio versus the

MEAN CHANNEL # VS. # OF CARBON ATOMS
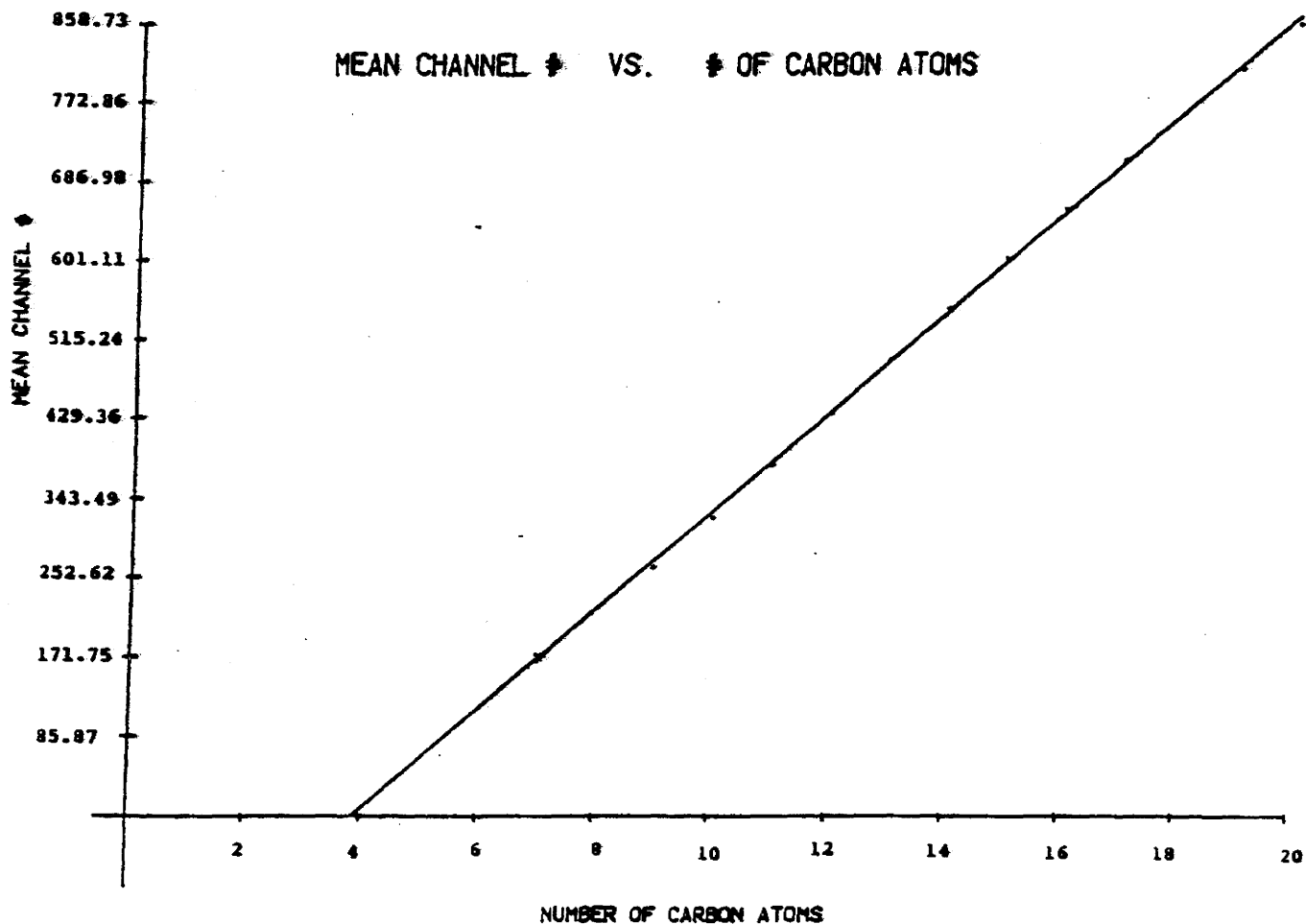
FIGURE 6.7

Illustrating the linearily of a retention time vs. the number of carbon atoms plot. For the above plot the slope of the line is 53.25 with an error of .73%, and the y intercept is -206.14 with an error of 2.7%.
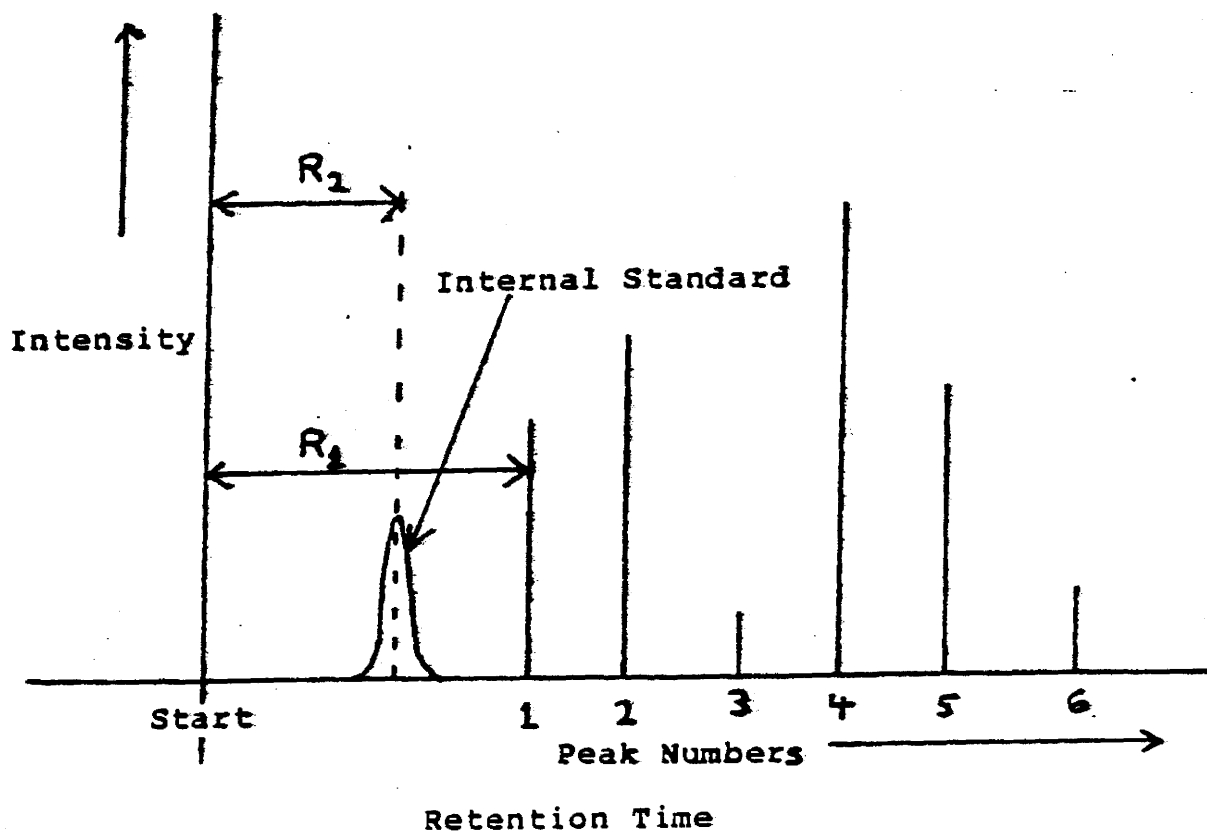
FIGURE 6.8

The peak retention ratio is defined as $R' = R_1/R_2$.
Therefore each peak is identified by it's retention ratio.
Similarly if the number of carbon atoms for each peak is
known (as it is for a spectrum of standards) then a curve of
retention ratios vs. the number of carbon atoms can be
obtained.

carbon number scale, a translation from the retention ratios to the number of carbon atoms can be obtained for any peak.

The use of retention ratios compensate for any offset in retention time. However the slope is not compensated for. Since the slope in retention time is very small (refer to section 6.1.c) the use of retention ratios proves to be a simple and effective tool for peak identification.

6.4   Basics Of GC Pattern Recognition

In order to determine the contents of a sample, the peaks of the GC spectrum must be identified. After the spectrum has been "de-skewed" and filtered, many small peaks are left in the spectrum in addition to the major ones. These very small peaks are usually due to noise present in the system or residuals from the filtering process. Therefore it is necessary to establish a noise discrimination level for peak intensities, below which peaks will not be considered in any subsequent analysis (refer to figure 6.9). The retention ratios and the carbon number of the remaining peaks can then be determined.

Retention time repeatability is crucial to the matching process, therefore retention ratios are used in order to correct for minor shifts. It was determined in

Typical GC spectrum without a noise descrimination level.

Intensity

Retention Time ———→

Intensity

The above GC spectrum with a 5% noise descrimination level.
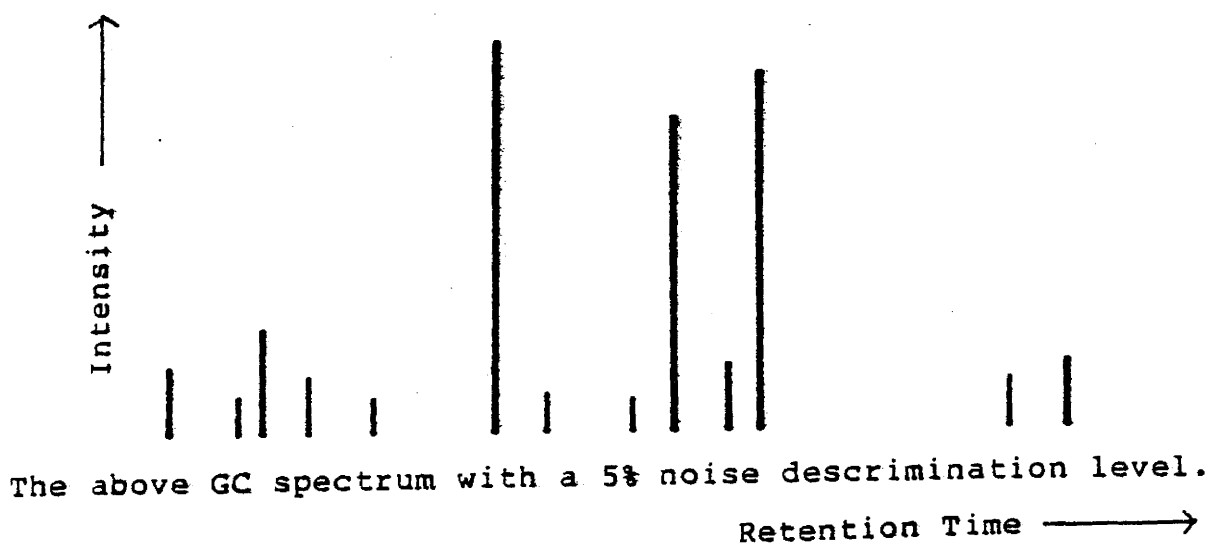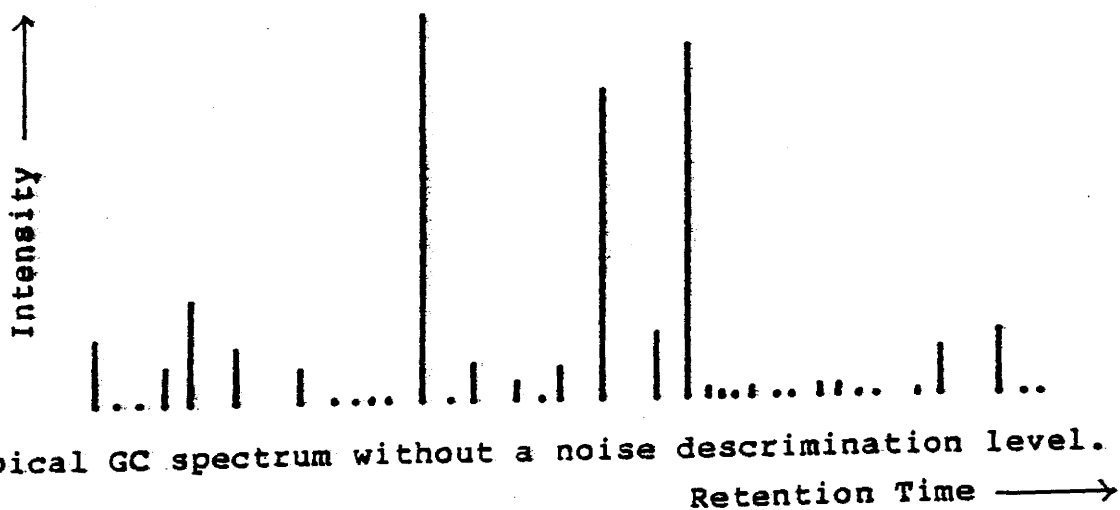
Retention Time ———→

FIGURE 6.9

Illustration of the application of a noise descrimination level.

section 6.1.c that, under typical conditions, the uncertainty in peak retention times was $\sigma = 0.86$ of a sampling interval. The minimum uncertainty of the

$$\text{retention ratio} = R' = \frac{R_1}{R_2}$$

could then be determined by

$$\frac{\sigma_{R'}^2}{(R')^2} = \frac{\sigma_{R_1}^2}{(R_1)^2} + \frac{\sigma_{R_2}^2}{(R_2)^2}$$

where $\sigma_{R'}$ is the standard deviation of the retention ratio R', and $\sigma_{R_1} = \sigma_{R_2} = 0.86$ of a sampling interval are the standard deviations of the retention times R_1 and R_2 respectively. This standard deviation $\sigma_{R'}$ can be used for assigning a confidence level, to the retention ratio. By using such ratios we can test to ensure that peaks will be correctly matched in the pattern recognition process.

Peaks with similar retention ratios are compared between an unknown spectra and a library of standard spectra. The peak areas in the unknown spectra are divided by the corresponding peak areas in the standard spectra to

provide an area ratio

$$A' = \frac{A_i}{A_{si}}$$

where A_i is the area of the unknown peak and A_{si} is the area of the standard. The consistency of a match is determined by calculating χ^2 of the area ratios A' for all the peaks encluded in the set. χ^2 is a statistic that estimates the deviation between the data and the parent distribution function which in this case is a gaussion. For a gaussion parent distribution function

$$\chi^2 = \sum \frac{(A' - \bar{A})^2}{\sigma_{A'}^2}$$

where $\bar{A} = \dfrac{\sum (A'/\sigma_{A'}^2)}{\sum (1/\sigma_{A'}^2)}$ = is the weighted mean.

The spectrum that has the smallest value for χ^2 will be the closest match to the unknown spectrum.

CHAPTER VII

DISCUSSION OF RESULTS

The goal in undertaking this work was to develop hardware and software whereby the feasibility of using pattern recognition techniques for gas chromatograms could be quantitatively explored. To attain this objective, it has been necessary to devise and assemble a data acquisition system that not only is able to monitor an existing GC unit but which possesses sufficient capacity to permit immediate use with newly evolving chromatographs. (Just as chromatographs continue to change, the availability of specialized and complex integrated circuits grows. As a result, the interfacing subsystem could now be greatly simplified in parts, though not in function.)

The software used in the GC system can be divided into three sections; data aquisition, communications, and data analysis, (refer to figure 7.1). A brief description of programs associated with these areas are provided in Appendix 1.0 along with the operating procedures. The environment in which the GC system is located and the tasks required of the analyst dictated that the logging system be

Program Name	CPU used	
PORTC	Osborne Microcomputer	
DISKE	" "	Data Aquistion
MASTER	" "	
RNDO	Osborne Microcomputer	Communications
RNDI	Nova Minicomputer	
CRTD	Nova Minicomputer	
DESKW	" "	
FLT	" "	Data Analysis
CALC	" "	
PLOT	" "	

FIGURE 7.1

Organization of the GC system software. Three basic groups are envolved; data aquisition, communications, and data analysis.

made virtually transparent to the technician. Completion of this phase of the work permitted the collection of digitized spectra to be undertaken.

The path which leads to the successful and effective application of pattern recognition methods is rough indeed, particularly if one does not possess, or cannot produce through digital signal processing methods, good quality data. At the outset, the spectral data were highly distorted, displaying peak skewing, poor resolution and baseline drift. Processing methods were devised to overcome these factors and much improved spectra were produced. While the methods developed here will continue to have application, improved chromatographs make signal recovery a much less onerous task. The extraction of most of the information contained within spectra permitted investigation of the reproducability typically attained in both sample proparation and sample analysis. This study was conducted with a view to assessing the quantities that would be of most use in any pattern recognition procedure. A high degree of reproducibility enhances ones ability to identify a sample with a reference within a standard library of possibilities. Small errors and good reproducibilty are the two factors which decreases ambiguity.

A careful study of several spectra and an assessment

of associated error indicated that the random variables corresponding to retention time and peak area are potentially good candidates for keys in any pattern recognition method. The errors displayed by these quantities are small so that use of retention time and area will be sufficient to permit a significant reduction in the ambiguity inherent in bacteria specification. Now that this phase has been completed, it is now necessary to proceed to amass a large quantity of experimental data for the bacterial strains most commonly encountered. With such data the exact form that the pattern recognition algorithm must take can be defined and, with luck, implemented.

The nature of development work usually permits one to break problems into portions and to deal with these in the most expedient and effective way possible. We have done that here but, throughout the entire project, the idea of producing a real time system has not been lost. Once the last phase of the program has been completed, i.e. the identification procedure, there appears to be no reason why a complete and automated "infection diagnostic machine" cannot be realized.

APPENDIX 1.0

SOFTWARE

a.   Data Aquisition Programs

Three programs control the data aquisition process, they are PORTC, DISKE and MASTER.   PORTC should be executed first, followed by DISKE and MASTER respectively.   PORTC and DISKE poke machine language programs into memory.   The program poked into memory by PORTC controls the input/output through the Osborne microcomputer's parallel port.   The program DISKE allows the user to use both disk drives for data storage.

The main controlling program is MASTER.   This program uses the subroutines provided by PORTC and DISKE in order to control I/O through the parallel port.   It also controls the storage of data onto the diskettes and inicates the status of the storage process to the user.

b.   Communication Programs

In order to transfer data from the Osborne microcomputer to the Nova minicomputer, two programs RNDO and RNDI have to be executed.   RNDI is executed first by

the Nova minicomputer followed by RNDO (which is executed by the Osborne microcomputer). The program RNDI inputs the serial data through the Nova serial port and writes it into random access files named RUN# (where # = an integer 1,2,3,4,5...). RNDO reads data from a random access file on the Osborne's diskettes and then outputs it through its serial port.

c.   Data Analysis Programs

The program CRTD displays the raw spectra on a graphics terminal. The data processing programs are DESKW, FLT, and CALC, and should be executed in the order tabulated. DESKW, as the name implies, de-skews the raw spectra. FLT filters the deskewed spectra in order to remove any baseline drift. The program CALC calculates the centroids and areas of all the peaks in the spectra. This information is then used for spectral pattern recognition. The program PLOT plots any selected spectra, if a permanent record is desired.

```
1 REM
2 REM    PROGRAM PORTC
3 REM
4 REM    THIS PROGRAM CONTROLS THE FLOW
5 REM    OF DATA THROUGH THE SERIAL I/O
6 REM    PORT
7 REM
8 REM
10  POKE &H9000,&HF3
20  POKE &H9001,&HD3
30  POKE &H9002,&H0
40  POKE &H9003,&H3E
50  POKE &H9004,&H0
60  POKE &H9005,&H32
70  POKE &H9006,&H8
80  POKE &H9007,&HEF
90  POKE &H9008,&H3E
100 POKE &H9009,&H30
110 POKE &H900A,&H32
120 POKE &H900B,&H1
130 POKE &H900C,&H29
140 POKE &H900D,&H3E
150 POKE &H900E,&H0
160 POKE &H900F,&H32
170 POKE &H9010,&H0
180 POKE &H9011,&H29
190 POKE &H9012,&H3E
200 POKE &H9013,&H34
210 POKE &H9014,&H32
220 POKE &H9015,&H1
230 POKE &H9016,&H29
240 POKE &H9017,&H3E
250 POKE &H9018,&H30
260 POKE &H9019,&H32
270 POKE &H901A,&H3
280 POKE &H901B,&H29
290 POKE &H901C,&H3E
300 POKE &H901D,&H3F
310 POKE &H901E,&H32
320 POKE &H901F,&H2
330 POKE &H9020,&H29
340 POKE &H9021,&H3E
350 POKE &H9022,&H34
360 POKE &H9023,&H32
370 POKE &H9024,&H3
380 POKE &H9025,&H29
390 POKE &H9026,&H3E
400 POKE &H9027,&H3
410 POKE &H9028,&H32
420 POKE &H9029,&H2
```

```
430 POKE &H902A,&H29
440 POKE &H902B,&H3A
450 POKE &H902C,&H2
460 POKE &H902D,&H29
470 POKE &H902E,&HE6
480 POKE &H902F,&H40
490 POKE &H9030,&HFE
500 POKE &H9031,&H40
510 POKE &H9032,&HC2
520 POKE &H9033,&H2B
530 POKE &H9034,&H90
540 POKE &H9035,&H0
550 POKE &H9036,&H0
560 POKE  &H9037,&H0
570 POKE &H9038,&H0
580 POKE &H9039,&H0
590 POKE &H903A,&H0
600 POKE &H903B,&H0
610 POKE &H903C,&H0
620 POKE &H903D,&H0
630 POKE &H903E,&H0
640 POKE &H903F,&H3A
650 POKE &H9040,&H2
660 POKE &H9041,&H29
670 POKE &H9042,&HE6
680 POKE &H9043,&H80
690 POKE &H9044,&HFE
700 POKE &H9045,&H80
710 POKE &H9046,&HCA
720 POKE &H9047,&H77
730 POKE &H9048,&H90
740 POKE &H9049,&H3E
750 POKE &H904A,&H3
760 POKE &H904B,&H32
770 POKE &H904C,&H2
780 POKE &H904D,&H29
790 POKE &H904E,&H3E
800 POKE &H904F,&H23
810 POKE &H9050,&H32
820 POKE &H9051,&H2
830 POKE &H9052,&H29
840 POKE &H9053,&H1
850 POKE &H9054,&H0
860 POKE &H9055,&H0
870 POKE &H9056,&H3
880 POKE &H9057,&H3A
890 POKE &H9058,&H0
900 POKE &H9059,&H29
910 POKE &H905A,&H2F
920 POKE &H905B,&H2
930 POKE &H905C,&H3E
940 POKE &H905D,&HB
950 POKE &H905E,&H32
```

```
960 POKE &H905F,&H2
970 POKE &H9060,&H29
980 POKE &H9061,&H3E
990 POKE &H9062,&H2B
1000 POKE &H9063,&H32
1010 POKE &H9064,&H2
1020 POKE &H9065,&H29
1030 POKE &H9066,&H3
1040 POKE &H9067,&H3A
1050 POKE &H9068,&H0
1060 POKE &H9069,&H29
1070 POKE &H906A,&H2F
1080 POKE &H906B,&H2
1090 POKE &H906C,&H7B
1100 POKE &H906D,&H32
1110 POKE &H906E,&H0
1120 POKE &H906F,&HC0
1130 POKE &H9070,&H79
1140 POKE &H9071,&H32
1150 POKE &H9072,&H1
1160 POKE &H9073,&HC0
1170 POKE &H9074,&HC3
1180 POKE &H9075,&H84
1190 POKE &H9076,&H90
1200 POKE &H9077,&HEB
1210 POKE &H9078,&H1A
1220 POKE &H9079,&H4F
1230 POKE &H907A,&H13
1240 POKE &H907B,&H1A
1250 POKE &H907C,&H47
1260 POKE &H907D,&H3E
1270 POKE &H907E,&H5
1280 POKE &H907F,&H2
1290 POKE &H9080,&H3
1300 POKE &H9081,&H3E
1310 POKE &H9082,&H0
1320 POKE &H9083,&H2
1330 POKE &H9084,&HD3
1340 POKE &H9085,&H1
1350 POKE &H9086,&H3E
1360 POKE &H9087,&H1
1370 POKE &H9088,&H32
1380 POKE &H9089,&H8
1390 POKE &H908A,&HEF
1400 POKE &H908B,&HFB
1410 POKE &H908C,&HC9
```

```
1 REM
2 REM    SUBROUTINE DISKE
3 REM
4 REM    THIS PROGRAM REMOVES THE WRITE PROTECTION
5 REM    FROM THE DISKS AFTER A DISK DRIVE DOOR HAS
6 REM    BEEN OPENED AND THEN CLOSED.
7 REM
8 REM
10 POKE &H9200,&HE
20 POKE &H9201,&HD
30 POKE &H9202,&HCD
40 POKE &H9203,&H5
50 POKE &H9204,&H0
60 POKE &H9205,&HC9
```

```
10 REM
20 REM    PROGRAM MASTER
30 REM
40 REM    THIS PROGRAM CONTROLS THE ENTIRE DATA
50 REM    AQUISITION PROCESS BY THE OSBORNE MICROCOMPUTER
60 REM
70 REM    IT SHOULD BE EXECUTED AFTER PORTC
80 REM
90 REM
100 REM...CALLING THE SUBROUTINE DISKE
110 DS=&H9200
120 CALL DS
130 REM.............................
140 FL8=0
150 FL1=0
160 MC1=0
170 R=0
180 V=0
190 MC=0
200 E1$="B:DATA"
210 CLOSE #3
220 OPEN "O",#3,E1$
230 FOR Z1=1 TO 500
240 PRINT CHR$(7)
250 NEXT Z1
260 PRINT "EVERYTHING IS READY, JUST ANSWER THE"
270 PRINT "FOLLOWING QUESTIONS............."
280 PRINT
290 PRINT "P.S.  DON'T FORGET, TO ENGAGE THE "
300 PRINT "INTEGRATOR WITH THE COMPUTER THE TOGGLE"
310 PRINT "SWITCH MUST BE FLICKED UP AS MARKED"
320 PRINT
330 PRINT
340 PRINT
350 INPUT "HOW MANY RUNS ARE YOU GOING TO RUN ";Q!
360 FOR W=1 TO 25
370 PRINT
380 NEXT W
390 E9=V
400 FOR J=1 TO Q!
410 E9=E9+1
420 E$="RUN      "
430 Q$=STR$(E9)
440 MID$(E$,4)=Q$
450 PRINT
460 PRINT TAB(7);"THIS INFORMATION IS FOR ";E$
470 PRINT TAB(31);"-----"
480 PRINT TAB(33);"^"
490 PRINT TAB(33);"!"
500 PRINT TAB(33);"!"
510 PRINT TAB(31);"LOOK !"
520 FOR W=1 TO 8
530 PRINT
```

```
540 NEXT W
550 PRINT "INFORMATION ABOUT BOTTLE NUMBER ";J
560 PRINT ".............................. "
570 INPUT "NAME OF THE SAMPLE ";A1$
580 INPUT "CREATION DATE ";B1$
590 INPUT "DATE OF THE RUN ";C1$
600 PRINT
610 L2$=STR$(J)
620 L1$="BOTTLE NUMBER        "
630 MID$(L1$,15)=L2$
640 PRINT
650 INPUT "IS THE ABOVE DATA CORRECT ";P$
660 FOR W=1 TO 7
670 PRINT
680 NEXT W
690 IF LEFT$(P$,1)="N" THEN GOTO 550
700 PRINT #3,L1$;",";A1$;",";B1$;",";C1$
710 NEXT J
720 CLOSE #3
730 OPEN "I",#3,E1$
740 FOR J1=1 TO Q!
750 INPUT #3,H1$,A1$,B1$,C1$
760 POKE &H9054,&HFF
770 POKE &H9055,&H9F
780 F=0
790 PT=0
800 IZ=9
810 BZ=VARPTR(IZ)
820 ST=&H9000
830 CALL ST(BZ)
840 IF F=1 GOTO 1000
850 GOTO 1400
860 FOR AZ=1 TO 25
870 PRINT
880 NEXT AZ
890 R=R+1
900 R$="RUN        "
910 S$=STR$(R)
920 MID$(R$,4)=S$
930 PRINT TAB(15);R$;"HAS STARTED"
940 PRINT TAB(15);"-----"
950 PRINT TAB(15);H1$;" INJECTED"
960 F=1
970 FOR K=1 TO 12
980 PRINT
990 NEXT K
1000 IF IZ=5 GOTO 1080
1010 CZ=PEEK(&HC001)
1020 AZ=PEEK(&HC000)
1030 POKE &H9054,CZ
1040 POKE &H9055,AZ
1050 PT=PT+1
1060 HC=HC+1
```

```
1070 GOTO 800
1080 FOR I=1 TO 12
1090 PRINT
1100 NEXT
1110 N=256*(PEEK(&HC000))+PEEK(&HC001)
1120 I=2
1130 V=V+1
1140 IF FL1=1 THEN GOTO 1160
1150 N$="B:RUN            "
1160 G$=STR$(V)
1170 G$=MID$(G$,2)
1180 MID$(N$,6)=G$
1190 CLOSE #1
1200 OPEN "R",#1,N$,100
1210 FIELD #1,24 AS HA$,26 AS AB$,25 AS AC$,25 AS AD$
1220 LSET HA$=H1$
1230 LSET AB$=A1$
1240 LSET AC$=B1$
1250 LSET AD$=C1$
1260 PUT #1,1
1270 ON ERROR GOTO 1670
1280 CLOSE #1
1290 OPEN "R",#1,N$,4
1300 FIELD #1,4 AS D$
1310 I=26
1320 FOR J=40960! TO N STEP 2
1330 A!=256*(PEEK(J+1))+PEEK(J)
1340 LSET D$=MKS$(A!)
1350 PUT #1,I
1360 ON ERROR GOTO 1670
1370 I=I+1
1380 NEXT J
1390 CLOSE #1
1400 FOR K=1 TO 25
1410 PRINT
1420 NEXT K
1430 IF F=1 GOTO 1460
1440 F=1
1450 GOTO 860
1460 FOR K=1 TO 25
1470 PRINT
1480 NEXT K
1490 PRINT N$
1500 PRINT
1510 PRINT "THE RUN IS OVER AND THE DATA SAVED"
1520 PRINT "ON THE DISK"
1530 PRINT "THE NUMBER OF POINTS SAVED WERE: ";PT
1540 PRINT
1550 PRINT
```

```
1560 PRINT
1570 PRINT "INFORMATION ABOUT ";N$
1580 PRINT "..............................."
1590 PRINT H1$
1600 PRINT "NAME OF THE SAMPLE: ";TAB(25);A1$
1610 PRINT "CREATION DATE: ";TAB(25);B1$
1620 PRINT "DATE OF THE RUN";TAB(25);C1$
1630 NEXT J1
1640 FOR ZA=1 TO 1500
1650 NEXT ZA
1660 GOTO 210
1670 IF ERR=61 THEN GOTO 1700
1680 PRINT "AN ERROR OCCURED, LET EDDIE KNOW"
1690 STOP
1700 FL1=1
1710 IF FL8=1 THEN GOTO 1870
1720 FL8=1
1730 N$="A:RUN        "
1740 FOR J=1 TO 25
1750 PRINT
1760 NEXT J
1770 PRINT "NO MORE ROOM ON DISK B THEREFORE"
1780 PRINT "SWITCHING TO DISK A"
1790 GOTO 1140
1800 IF ERR<>61 THEN GOTO 1680
1810 PRINT "NO MORE ROOM ON DISK B THEREFORE"
1820 PRINT "SWITCHING TO DISK A"
1830 PRINT "INPUT INFORMATION ABOUT ALL THE RUNS "
1840 PRINT "AGAIN"
1850 LET E1$="A:DATA"
1860 GOTO 230
1870 PRINT "NO MORE DISK ROOM ANYWHERE "
1880 PRINT "PHONE EDDIE REBANE EXT.2010"
1890 END
```

```
10 REM
20 REM
30 REM     PROGRAM RNDO
40 REM
50 REM     THIS PROGRAM OUTPUTS THE SPECTRAL
60 REM     DATA THROUGH THE SERIAL I/O PORT
70 REM     TO THE NOVA MINICOMPUTER
80 REM
90 REM     RUN THIS AFTER RNDI HAS BEEN EXECUTED
100 REM
110 DS=&H9200
120 CALL DS
130 INPUT "HOW MANY RUNS? ",V
140 FL5=0
150 N$="B:RUN        "
160 FOR J=1 TO V
170 C=0
180 G$=STR$(J)
190 G$=MID$(G$,2)
200 MID$(N$,6)=G$
210 PRINT
220 PRINT
230 PRINT "INFORMATION ABOUT ";N$
240 PRINT "..............................."
250 OPEN "R",#1,N$,100
260 FIELD #1,24 AS X$,26 AS E$,25 AS F$,25 AS G$
270 GET #1,1
280 LPRINT X$
290 LPRINT E$
300 LPRINT F$
310 LPRINT G$
320 PRINT X$
330 PRINT E$
340 PRINT F$
350 PRINT G$
360 CLOSE #1
370 OPEN "R",#1,N$,4
380 FL3=0
390 FIELD #1,4 AS H$
400 C1=2
410 I=26
420 K=1
430 GET #1,I
440 L!=CVS(H$)
450 IF FL3=1 THEN GOTO 470
460 IF L!=0 THEN C1=3
```

```
470 IF L!=0 THEN C=C+1
480 IF C=C1 THEN GOTO 580
490 LPRINT L!
500 PRINT L!,K
510 FOR K1=1 TO 3
520 NEXT K1
530 ON ERROR GOTO 620
540 I=I+1
550 FL3=1
560 K=K+1
570 GOTO 430
580 CLOSE #1
590 FOR H=1 TO 9000
600 NEXT H
610 NEXT J
620 IF ERR<>61 THEN GOTO 670
630 N$="A:RUN           "
640 IF FL5=1 THEN GOTO 690
650 FL5=1
660 GOTO 170
670 PRINT "AN ERROR OCCURED "
680 STOP
690 PRINT "BOTH DISKS ARE EMPTIED"
700 PRINT "NUMBER OF FILES EMPTIED WERE ";V
710 STOP
```

```
0010 REM
0020 REM
0030 REM     PROGRAM RNDI
0040 REM
0050 REM     THIS PROGRAM IS USED TO INPUT DATA
0060 REM     FROM THE OSBORNE MICROCOMPUTER TO THE NOVA
0070 REM     MINICOMPUTER
0080 REM
0090 REM     EXECUTE THIS BEFORE RNDO
0100 REM
0110 DIM D[2300]
0120 DIM D$[23]
0130 DIM N$[20],M$[5]
0140 DIM A$[23],B$[23],C$[23]
0150 INPUT "FROM WHAT RUN TO WHAT RUN? ",A,B
0160 FOR J=A TO B
0170    CLOSE
0180    LET N$="RUN          "
0190    LET M$=STR$(J)
0200    LET N$[4]=M$
0210    INPUT A$,B$,C$,D$
0220    LET C=0
0230    LET F=0
0240    LET K=1
0250    INPUT D[K]
0260    IF F=0 THEN GOTO 0280
0270    IF D[K]=0 THEN LET C=C+1
0280    IF C=1 THEN GOTO 0320
0290    LET F=1
0300    LET K=K+1
0310    GOTO 0250
0320    OPEN FILE[1,0],N$,24
0330    WRITE FILE[1,0],A$
0340    WRITE FILE[1,1],B$
0350    WRITE FILE[1,2],C$
0360    WRITE FILE[1,3],D$
0370    CLOSE
0380    OPEN FILE[1,0],N$,4
0390    FOR L=25 TO K+25
0400       WRITE FILE[1,L],D[L-24]
0410    NEXT L
0420    CLOSE
0430 NEXT J
```

```
0010 REM
0020 REM
0030 REM     PROGRAM CRTD
0040 REM
0050 REM     THIS PROGRAM DISPLAYS THE COLLECTED SPECTRA
0060 REM     ON THE MONITOR
0070 REM
0080 DIM X[2300]
0090 DIM P$[10]
0100 DIM N$[20],M$[5]
0110 DIM A$[23],B$[23],C$[23],D$[23]
0120 LET F9=0
0130 INPUT "FILE PREFIX ? ",P$
0140 INPUT "FROM WHAT RUN TO WHAT? ",A,B
0150 INPUT " Y SCALE EXPANSION ? 1X, 2X ETC  ",X9
0160 FOR J=A TO B
0170    CLOSE
0180    LET N$="             "
0190    LET N$[1,4]=P$
0200    LET M$=STR$(J)
0210    LET N$[4]=M$
0220    OPEN FILE[1,0],N$,24
0230    READ FILE[1,0],A$
0240    READ FILE[1,1],B$
0250    READ FILE[1,2],C$
0260    READ FILE[1,3],D$
0270    CLOSE
0280    PRINT
0290    PRINT A$
0300    PRINT "NAME OF SAMPLE "; TAB(20);B$
0310    PRINT "CREATION DATE "; TAB(20);C$
0320    PRINT "DATE OF THE RUN "; TAB(20);D$
0330    PRINT
0340    FOR P=1 TO 1000
0350    NEXT P
0360    LET C1=1
0370    OPEN FILE[1,0],N$,4
0380    LET C=0
0390    LET F=0
0400    LET K=25
0410    LET L=0
0420    READ FILE[1,K],E
0430    IF F=0 THEN GOTO 0450
```

```
0440    IF E=0 THEN LET C=C+1
0450    LET X[C1]=E
0460    IF E>L THEN LET L=E
0470    IF C=1 THEN GOTO 0520
0480    LET F=1
0490    LET K=K+1
0500    LET C1=C1+1
0510    GOTO 0420
0520    CLOSE
0530    LET Y1=X9*470
0540    CALL 17
0550    CALL 14,0,0,0
0560    FOR M=1 TO C1
0570       LET X=(500/C1)*M
0580       LET Y=(Y1/L)*X[M]
0590       IF Y>470 THEN GOTO 0640
0600       IF Y=0 THEN GOTO 0650
0610       IF Y<0 THEN LET F9=1
0620       IF Y<0 THEN LET Y=0
0630       CALL 14,X,Y,1
0640       CALL 14,X,Y,0
0650    NEXT M
0660    IF F9=1 THEN PRINT "NEGATIVE NUMBERS PRESENT"
0670    FOR P=1 TO 3000
0680       NEXT P
0690 NEXT J
```

```
0010 REM
0020 REM
0030 REM    PROGRAM DESKW
0040 REM
0050 REM    THIS PROGRAM DE-SKEWS A SPECTRUM
0060 REM    USING A VARIABLE TIME CONSTANT
0070 REM
0080 DIM X[1300]
0090 DIM P$[10]
0100 DIM N$[20],M$[5]
0110 DIM A$[23],B$[23],C$[23],D$[23]
0120 LET P$="RUN"
0130 INPUT "FROM WHAT RUN TO WHAT? ",A,B
0140 FOR J=A TO B
0150    LET F5=0
0160    CLOSE
0170    LET W$="DRW"
0180    LET N$="                "
0190    LET N$[1,4]=P$
0200    LET M$=STR$(J)
0210    LET N$[4]=M$
0220    LET W$[4]=M$
0230    OPEN FILE[1,0],N$,24
0240    READ FILE[1,0],A$
0250    READ FILE[1,1],B$
0260    READ FILE[1,2],C$
0270    READ FILE[1,3],D$
0280    CLOSE
0290    PRINT
0300    PRINT A$
0310    PRINT "NAME OF SAMPLE "; TAB(20);B$
0320    PRINT "CREATION DATE "; TAB(20);C$
0330    PRINT "DATE OF THE RUN "; TAB(20);D$
0340    PRINT
0350    LET C1=1
0360    OPEN FILE[1,0],N$,4
0370    LET C=0
0380    LET F=0
0390    LET K=25
0400    LET L=0
0410    READ FILE[1,K],E
0420    READ FILE[1,K+1],G
0430    IF EOF(1) THEN GOTO 0510
0440    LET X[C1]=G-E
0450    IF E>L THEN LET L=E
```

```
0460    LET F=1
0470    LET K=K+1
0480    GOTO 0730
0490    LET C1=C1+1
0500    GOTO 0410
0510    CLOSE
0520    LET C1=C1-1
0530    OPEN FILE[1,0],U$,24
0540    WRITE FILE[1,0],A$
0550    WRITE FILE[1,1],B$
0560    WRITE FILE[1,2],C$
0570    WRITE FILE[1,3],D$
0580    CLOSE
0590    LET S5=.13/630
0600    FOR Z=1 TO C1
0610      LET K0=S5*Z
0620      IF Z>630 THEN LET K0=.17
0630      LET X[Z]=(1/K0)*X[Z]
0640    NEXT Z
0650    CLOSE
0660    LET K=25
0670    OPEN FILE[1,0],N$,4
0680    FOR M=1 TO C1
0690      READ FILE[1,K],C
0700      LET X[M]=X[M]+C
0710      LET K=K+1
0720    NEXT M
0730    CLOSE
0740    OPEN FILE[1,0],U$,4
0750    FOR G=25 TO C1+25
0760      WRITE FILE[1,G],X[G-24]
0770    NEXT G
0780    WRITE FILE[1,G],0
0790    CLOSE
0800 NEXT J
0810 LET S5=4E-02/6330
```

```
0010 REM
0020 REM
0030 REM    PROGRAM FLT
0040 REM
0050 REM    THIS PROGRAM FILTERS A SPECTRA
0060 REM    IN ORDER TO REMOVE A VARYING BASELINE
0070 REM    AND THEN CORRECTS THE PEAK AREAS ACCORDINGLY
0080 REM
0090 DIM R[250]
0100 DIM C[50]
0110 DIM X[1500]
0120 DIM P$[10]
0130 DIM N$[20],M$[5]
0140 DIM A$[23],B$[23],C$[23],D$[23]
0150 LET P$="TRU"
0160 INPUT "FROM WHAT RUN TO WHAT? ",A,B
0170 INPUT "UPPER, MUST BE ODD",U5
0180 INPUT "LOWER",L5
0190 FOR J=A TO B
0200    LET S=0
0210    CLOSE
0220    LET W$="FRU        "
0230    LET N$="          "
0240    LET N$[1,4]=P$
0250    LET M$=STR$(J)
0260    LET N$[4]=M$
0270    LET W$[4]=M$
0280    OPEN FILE[1,0],N$,24
0290    READ FILE[1,0],A$
0300    READ FILE[1,1],B$
0310    READ FILE[1,2],C$
0320    READ FILE[1,3],D$
0330    CLOSE
0340    PRINT
0350    PRINT A$
0360    PRINT "NAME OF SAMPLE "; TAB(20);B$
0370    PRINT "CREATION DATE "; TAB(20);C$
0380    PRINT "DATE OF THE RUN "; TAB(20);D$
0390    PRINT
0400    LET C1=1
0410    OPEN FILE[1,0],N$,4
0420    LET C=0
0430    LET F=0
0440    LET K=25
0450    LET L=0
0460    READ FILE[1,K],E
```

```
0470    IF EOF(1) THEN GOTO 0530
0480    LET X[C1]=E
0490    IF E>L THEN LET L=E
0500    LET K=K+1
0510    LET C1=C1+1
0520    GOTO 0460
0530    CLOSE
0540    LET C1=C1-2
0550    OPEN FILE[1,0],W$,24
0560    WRITE FILE[1,0],A$
0570    WRITE FILE[1,1],B$
0580    WRITE FILE[1,2],C$
0590    WRITE FILE[1,3],D$
0600    CLOSE
0610    LET U=U5
0620    LET L=L5
0630    FOR K=1 TO U
0640       LET R[K]=-1/(2*L)
0650       LET R[K+(2*U)]=-1/(2*U)
0660       LET R[U+K]=1/U
0670    NEXT K
0680    LET N=U+(2*L)
0690    LET M=C1-N+1
0700    CALL 32,R[1],X[1],N,M
0710    FOR K=C1 TO 1 STEP -1
0720       LET V5=L+INT((U/2)-.5)
0730       LET X[K+V5]=X[K]
0740    NEXT K
0750    FOR K=1 TO V5
0760       LET X[K]=0
0770    NEXT K
0780    FOR K=(C1-1)-V5 TO C1
0790       LET X[K]=0
0800    NEXT K
0810    CLOSE
0820    OPEN FILE[1,0],W$,4
0830    FOR K=1 TO C1
0840       WRITE FILE[1,K+24],X[K]
0850    NEXT K
0860    WRITE FILE[1,K+25],0
0870    CLOSE
0880 NEXT J
```

```
0010 REM
0020 REM
0030 REM    PROGRAM CALC
0040 REM
0050 REM    THIS PROGRAM CALCULATES THE PEAK CENTROIDS AND AREAS
0060 REM    AND STORES THEM ON A FILE FOR FUTURE REFERENCE
0070 REM
0080 DIM C[50]
0090 DIM X[1500]
0100 DIM P$[10]
0110 DIM N$[20],M$[5]
0120 DIM A$[23],B$[23],C$[23],D$[23]
0130 LET P$="FRU"
0140 INPUT "FROM WHAT RUN TO WHAT? ",A5,B5
0150 FOR J=A5 TO B5
0160    LET E5=1
0170    LET S=0
0180    CLOSE
0190    LET W$="CRU"
0200    LET N$="            "
0210    LET N$[1,4]=P$
0220    LET M$=STR$(J)
0230    LET N$[4]=M$
0240    LET W$[4]=M$
0250    OPEN FILE[1,0],N$,24
0260    READ FILE[1,0],A$
0270    READ FILE[1,1],B$
0280    READ FILE[1,2],C$
0290    READ FILE[1,3],D$
0300    CLOSE
0310    PRINT
0320    PRINT A$
0330    PRINT "NAME OF SAMPLE "; TAB(20);B$
0340    PRINT "CREATION DATE "; TAB(20);C$
0350    PRINT "DATE OF THE RUN "; TAB(20);D$
0360    PRINT
0370    PRINT "  LOC"," + AREA","CR WIDTH"," RATIO","   AREA"
0380    LET C1=1
0390    OPEN FILE[1,0],N$,4
0400    LET C=0
0410    LET F=0
0420    LET K=25
0430    LET L=0
0440    READ FILE[1,K],E
0450    IF EOF(1) THEN GOTO 0500
0460    LET X[C1]=E
0470    LET K=K+1
0480    LET C1=C1+1
0490    GOTO 0440
0500    CLOSE
```

```
0510    LET C1=C1-2
0520    OPEN FILE[1,0],W$,24
0530    WRITE FILE[1,0],A$
0540    WRITE FILE[1,1],B$
0550    WRITE FILE[1,2],C$
0560    WRITE FILE[1,3],D$
0570    CLOSE
0580    OPEN FILE[1,0],W$,4
0590    LET I=1
0600    LET F1=0
0610    LET M1=0
0620    LET M2=0
0630    LET M3=0
0640    LET F2=0
0650    IF X[I]>0 THEN GOTO 0950
0660    IF X[I]=0 THEN GOTO 0950
0670    LET F1=1
0680    IF X[I]<0 THEN GOTO 0950
0690    IF X[I]=0 THEN GOTO 0950
0700    IF F2=1 THEN GOTO 0750
0710    LET F2=1
0720    LET M=(X[I]-X[I-1])
0730    LET B=X[I]-(M*I)
0740    LET S=-B/M
0750    IF X[I]<0 THEN GOTO 0810
0760    LET M1=(I*X[I])+M1
0770    LET M2=X[I]+M2
0780    LET I=I+1
0790    IF I=C1 THEN GOTO 0990
0800    GOTO 0700
0810    LET M=(X[I]-X[I-1])
0820    LET B=(X[I]-(M*I))
0830    LET F=-B/M
0840    LET M3=M1/M2
0850    LET A1=F-S
0860    IF A1<7.42377 THEN GOTO 0940
0870    GOSUB 1050
0880    PRINT M3,M2,A1,R3,R3*M2
0890    CLOSE
0900    OPEN FILE[1,0],W$,4
0910    WRITE FILE[1,24+E5],M3
0920    WRITE FILE[1,25+E5],(R3*M2)
0930    LET E5=E5+2
0940    GOTO 0600
0950    LET I=I+1
0960    IF I=C1+1 THEN GOTO 0990
```

```
0970    IF F1=1 THEN GOTO 0680
0980    GOTO 0650
0990    CLOSE
1000    OPEN FILE[1,0],W$,4
1010    WRITE FILE[1,24+E5],Q
1020    CLOSE
1030 NEXT J
1040 STOP
1050 LET L$="DAT3"
1060 CLOSE
1070 OPEN FILE[1,0],L$,4
1080 LET S1=1000
1090 LET K=1
1100 READ FILE[1,K],V
1110 IF EOF(1) THEN GOTO 1150
1120 IF ABS(V-A1)<S1 THEN LET S1=ABS(V-A1)
1130 LET K=K+8
1140 GOTO 1100
1150 LET K=1
1160 READ FILE[1,K],V
1170 IF ABS(V-A1)=S1 THEN GOTO 1200
1180 LET K=K+8
1190 GOTO 1160
1200 LET N=K
1210 READ FILE[1,K+8+4],R1
1220 READ FILE[1,K+4],R2
1230 LET R=R1-R2
1240 READ FILE[1,K+8],C5
1250 READ FILE[1,K],C2
1260 LET C=C5-C2
1270 LET R3=R2+((R/C)*(A1-C2))
1280 IF A1>C2 THEN GOTO 1340
1290 READ FILE[1,K-8+4],R1
1300 LET R=R2-R1
1310 READ FILE[1,K-8],C5
1320 LET C=C2-C5
1330 LET R3=R1+((R/C)*(A1-C5))
1340 RETURN
```

```
0010 REM
0020 REM
0030 REM     PROGRAM PLOT
0040 REM
0050 REM     THIS PROGRAM PLOTS THE SPECIFIED SPECTRA
0060 REM     ALONG WITH ITS BAR GRAPH
0070 REM
0080 DIM Z[500]
0090 DIM E$[50]
0100 DIM X[1500]
0110 DIM P$[10]
0120 DIM N$[20],M$[5]
0130 DIM A$[50],B$[50],C$[50],D$[50]
0140 LET F9=0
0150 INPUT "FROM WHAT RUN TO WHAT? ",A,B
0160 INPUT "HIT 1 FOR UP ",U
0170 INPUT "INPUT 1 IF PLOT TOTAL SPECTRA ",A8
0180 IF A8=1 THEN INPUT "INPUT WHICH RUN",F$
0190 IF A8=1 THEN GOTO 1020
0200 INPUT " HIT 1 IF YOU DONT WANT TO INPUT TITLE NAME ",A9
0210 FOR J=A TO B
0220    CLOSE
0230    LET N$="CRU "
0240    LET M$=STR$(J)
0250    LET N$[4]=M$
0260    PRINT
0270    PRINT
0280    LET C1=1
0290    OPEN FILE[1,0],N$,4
0300    LET C=0
0310    LET F=0
0320    LET K=25
0330    LET L=0
0340    READ FILE[1,K],E
0350    IF F=0 THEN GOTO 0380
0360    IF E=0 THEN LET C=C+1
0370    IF EOF(1) THEN GOTO 0440
0380    LET X[C1]=E
0390    IF C=1 THEN GOTO 0440
0400    LET F=1
0410    LET K=K+1
0420    LET C1=C1+1
0430    GOTO 0340
0440    CLOSE
0450    IF EOF(1) THEN LET C1=C1-2
0460    FOR K=2 TO C1-1 STEP 2
0470       IF X[K]>L THEN LET L=X[K]
0480    NEXT K
```
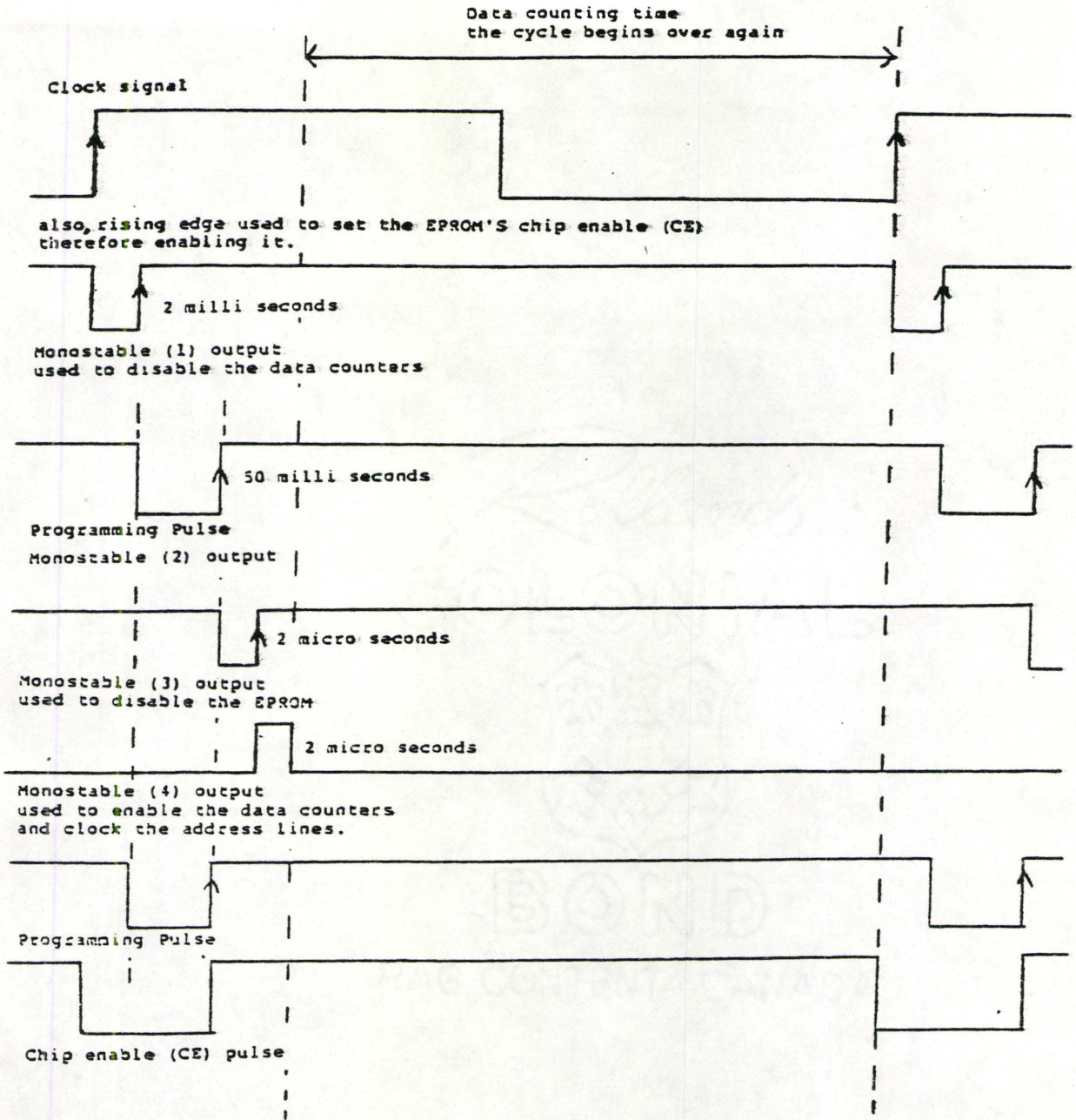
```
0490    FOR K9=1 TO 500
0500       LET Z[K9]=X[K9]
0510    NEXT K9
0520    MAT X=(0)*X
0530    INPUT "PERCENTAGE ",P
0540    LET C=2
0550    FOR K=2 TO C1-1 STEP 2
0560       IF Z[K]<((P/100)*L) THEN GOTO 0600
0570       LET X[C-1]=Z[K-1]
0580       LET X[C]=Z[K]
0590       LET C=C+2
0600    NEXT K
0610    LET C=C-2
0620    LET H=C1-1
0630    IF A9<>1 THEN INPUT "INPUT TITLE OF GRAPH ",E$
0640    IF U=1 THEN INPUT "INPUT FIGURE 1(A) TITLE ",A$
0650    IF U<>1 THEN INPUT "INPUT FIGURE 1(B) TITLE ",C$
0660    LET C9=0
0670    CALL 9,2
0680    CALL 12,35,50,90
0690    CALL 10,593,680,0
0700    CALL 11,E$
0710    FOR M=1 TO C STEP 2
0720       LET Y=(1928/1200)*X[M]
0730       LET Y=Y+482
0740       IF U<>1 THEN GOTO 0800
0750       LET X=1911
0760       CALL 10,X,Y,0
0770       LET X=1911-((1303/L)*X[M+1])
0780       CALL 10,X,Y,1
0790       IF U=1 THEN GOTO 0840
0800       LET X=3536
0810       CALL 10,X,Y,0
0820       LET X=3536-((1303/L)*X[M+1])
0830       CALL 10,X,Y,1
0840    NEXT M
0850    IF U<>1 THEN GOTO 0930
0860    LET B$="FIGURE    1(A)"
0870    CALL 12,35,50,90
0880    CALL 10,2044,869,0
0890    CALL 11,A$
0900    CALL 10,2097,1000,0
0910    CALL 11,B$
0920    IF U=1 THEN GOTO 0990
0930    LET D$="FIG     1(B)"
0940    CALL 12,35,50,90
0950    CALL 10,3669,869,0
0960    CALL 11,C$
```

```
0970    CALL 10,3722,1000,0
0980    CALL 11,D$
0990    CALL 9,0
1000 NEXT J
1010 STOP
1020 CLOSE
1030 OPEN FILE[1,0],F$,4
1040 LET L5=0
1050 LET K=25
1060 LET C5=1
1070 READ FILE[1,K],E
1080 IF EOF(1) THEN GOTO 1140
1090 LET X[C5]=E
1100 IF X[C5]>L5 THEN LET L5=X[C5]
1110 LET C5=C5+1
1120 LET K=K+1
1130 GOTO 1070
1140 LET C5=C5-2
1150 CALL 9,0
1160 CALL 9,2
1170 FOR M=1 TO C5-1
1180    LET Y=(1928/1200)*M
1190    LET Y=Y+482
1200    IF U<>1 THEN GOTO 1260
1210    LET X=1911
1220    IF M=1 THEN CALL 10,X,Y,0
1230    LET X=1911-((1303/L5)*X[M])
1240    CALL 10,X,Y,1
1250    IF U=1 THEN GOTO 1300
1260    LET X=3536
1270    IF M=1 THEN CALL 10,X,Y,0
1280    LET X=3536-((1303/L5)*X[M])
1290    CALL 10,X,Y,1
1300 NEXT M
1310 CALL 9,0
1320 STOP
```

*

Data counting time
the cycle begins over again

Clock signal

also, rising edge used to set the EPROM'S chip enable (CE)
therefore enabling it.

2 milli seconds

Monostable (1) output
used to disable the data counters

50 milli seconds

Programming Pulse
Monostable (2) output

2 micro seconds

Monostable (3) output
used to disable the EPROM

2 micro seconds

Monostable (4) output
used to enable the data counters
and clock the address lines.

Programming Pulse

Chip enable (CE) pulse

TIMING DIAGRAM FOR THE EPROM PROGRAMMER CIRCUIT

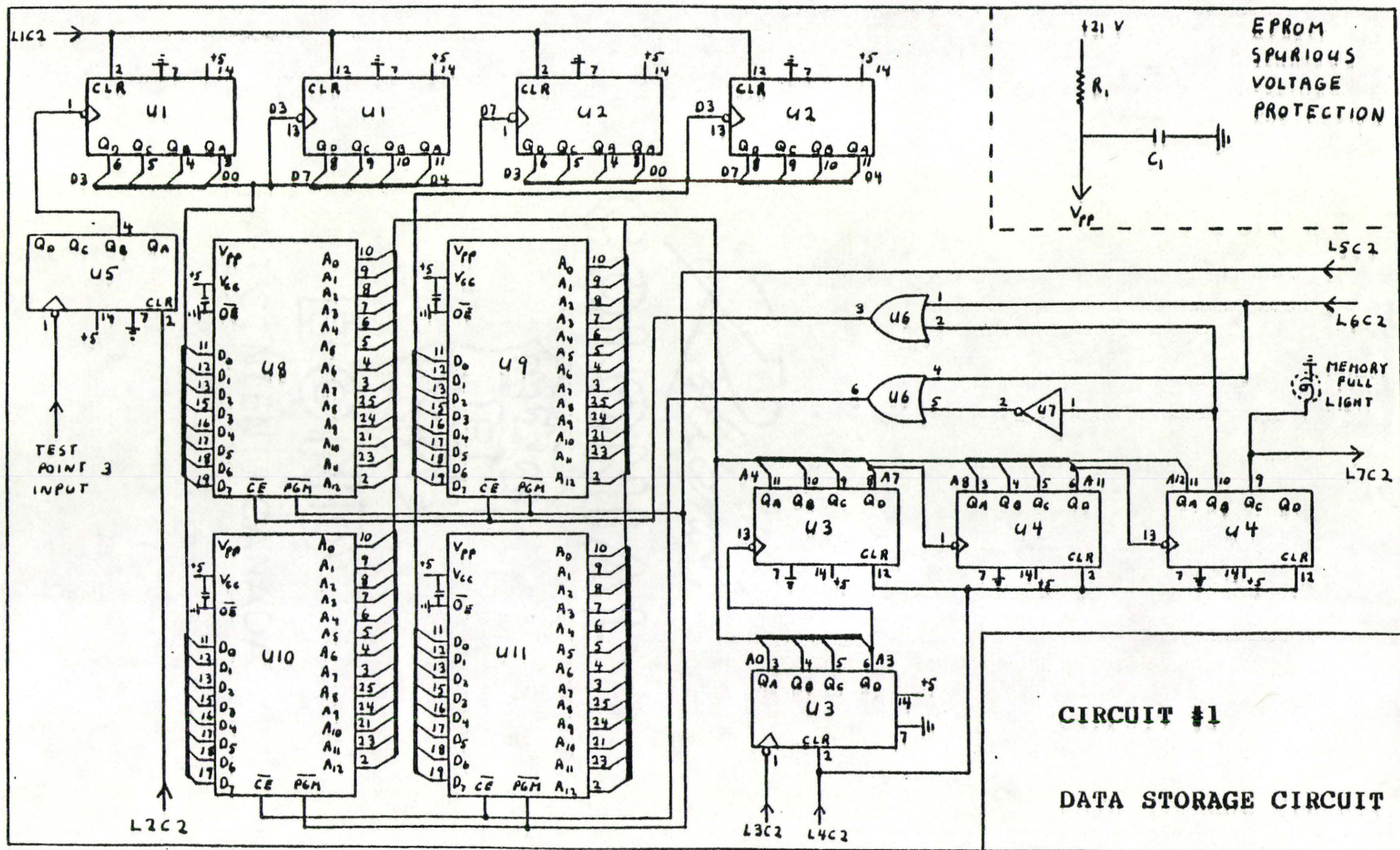(Gives the timing of the signals which are of major
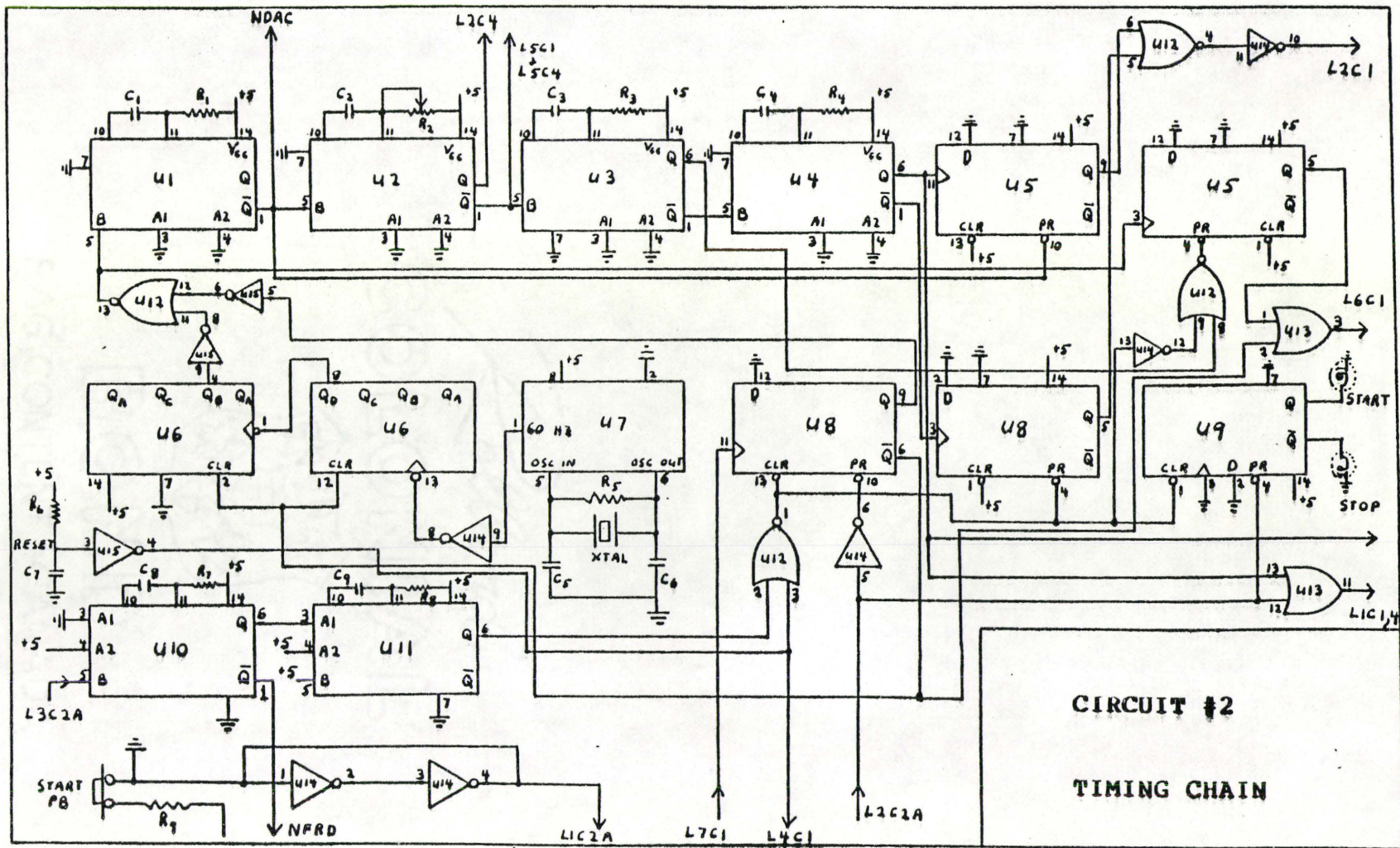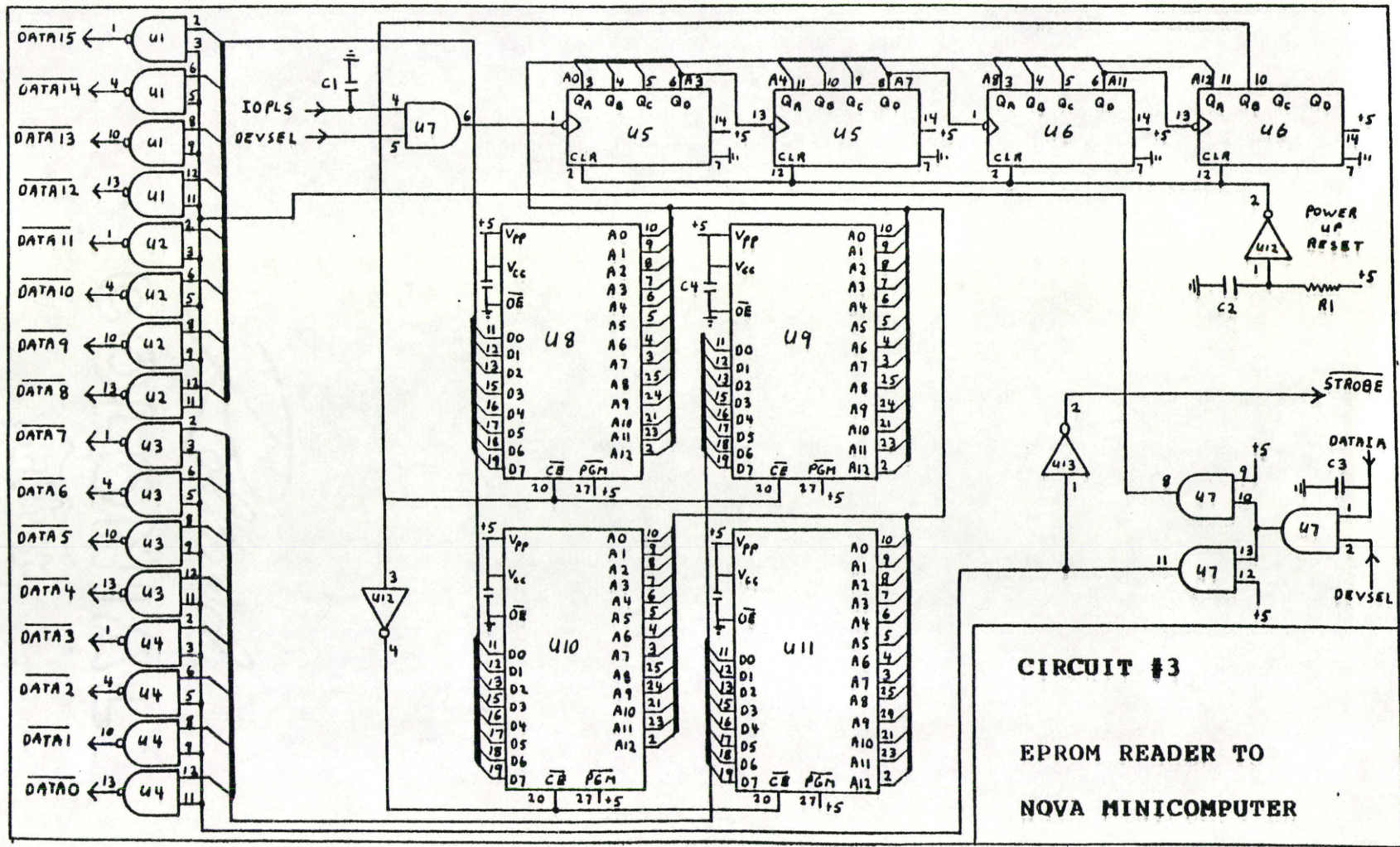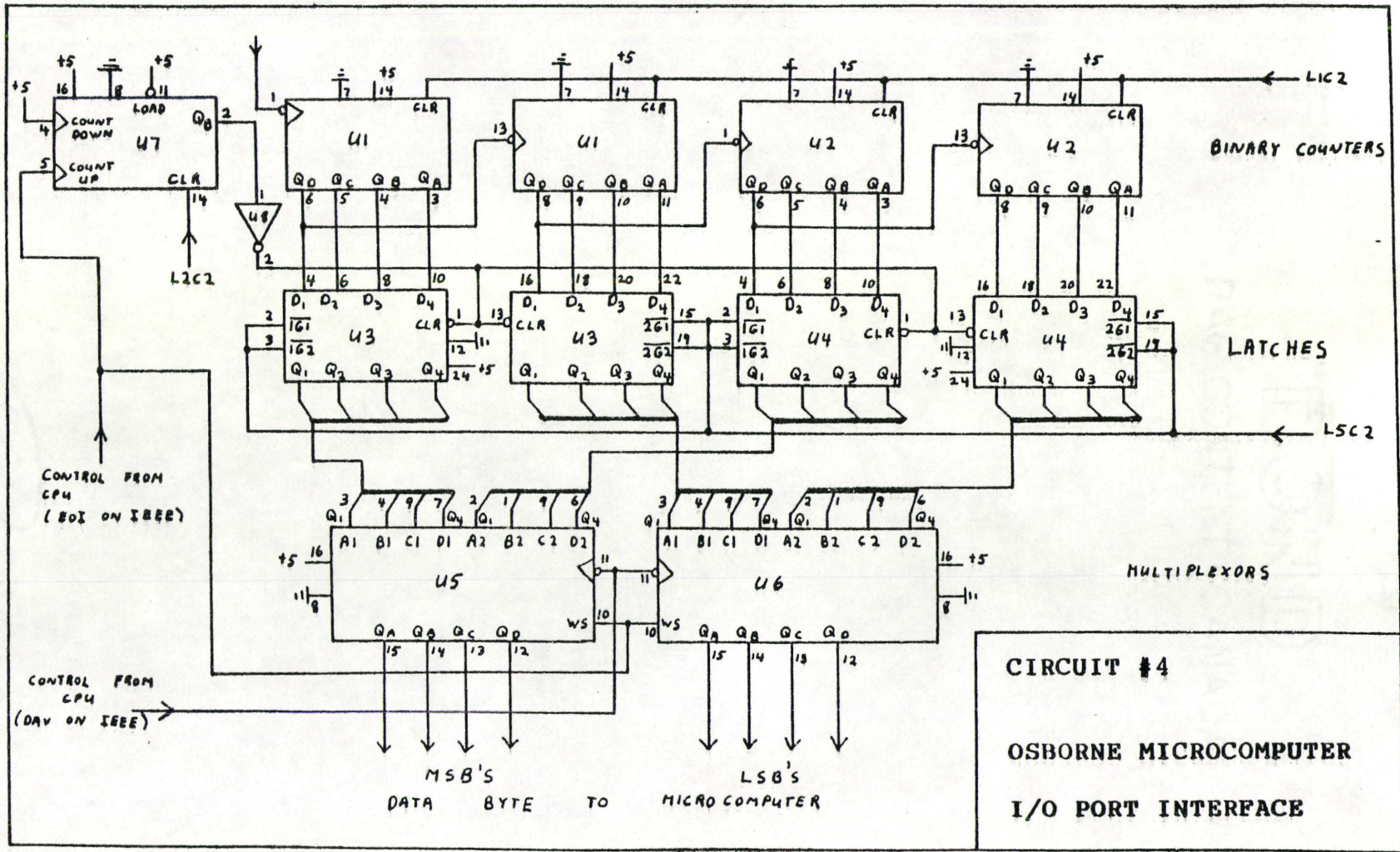
importance)

FIGURE A.1

FIGURE A.2

FIGURE A.3

FIGURE A.4
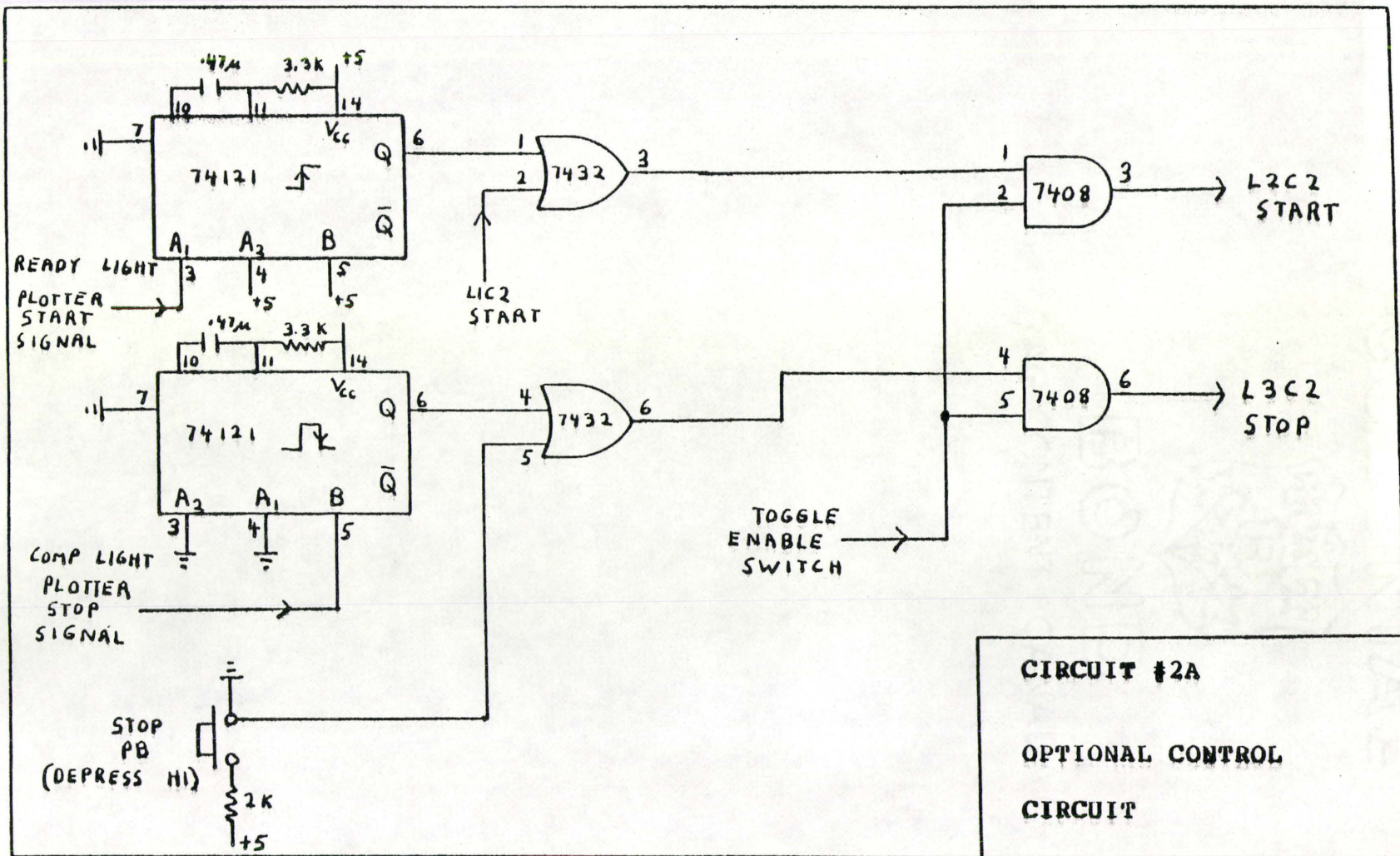
FIGURE A.5

FIGURE A.6

PARTS LIST

INTEGRATED CIRCUITS

CIRCUIT #3

U1-U5	74393	DUAL 4-BIT BINARY COUNTERS
U6	7432	QUADRUPLE 2-INPUT POSITIVE-OR GATES
U7	7404	HEX INVERTERS
U8-U11	2764	(8K X 8) UV ERASABLE PROMS

CIRCUIT #2

U1-U4,U10-U11	74121	MONOSTABLE MULTIVIBRATORS
U5,U8,U9	7474	DUAL D-TYPE FLIP-FLOPS
U6	74393	DUAL 4-BIT BINARY COUNTERS
U7	MM5369	CLOCK/DIVIDER
U12	7428	QUADRUPLE 2-INPUT POSITIVE -NOR BUFFERS
U13	7432	QUADRUPLE 2-INPUT POSITIVE -OR GATES
U14-U15	7404	HEX INVERTERS

CIRCUIT #3
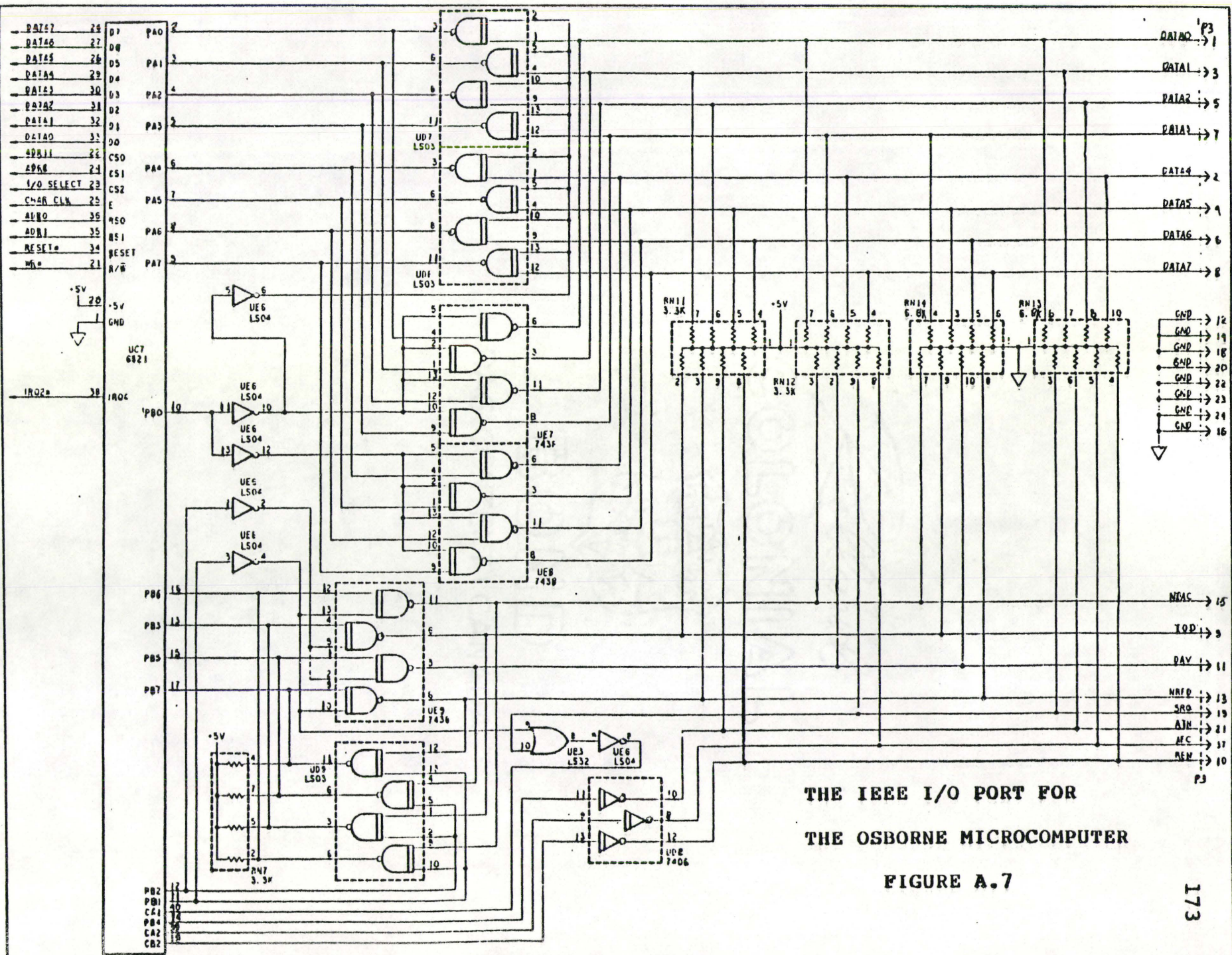
U1-U4	7401	QUADRUPLE 2-INPUT POSITIVE -NAND GATES
U5-U6	74393	DUAL 4-BIT BINARY COUNTERS
U7	7408	QUADRUPLE 2-INPUT POSITIVE -AND GATES
U8-U11	2764	(8K X 8) UV ERASABLE PROMS
U12	7404	HEX INVERTERS

CIRCUIT #4

U1-U2	75393	DUAU 4-BIT BINARY COUNTERS
U3-U4	74116	DUAL 4-BIT LATCHES
U5-U6	74298	QUAD 2-INPUT MUX'S WITH STORAGE
U7	74193	SYNCHRONOUS UP/DOWN DUAL CLOCK COUNTERS

TABLE A.1

THE IEEE I/O PORT FOR

THE OSBORNE MICROCOMPUTER

FIGURE A.7

173

REFERENCES

(1)   J.B. Brooks, "Frequency-Pulsed Electron Capture Gas-Liquid Chromatography and the Tryptophan Color Test for Rapid Diagnosis of Tuberculous and Other Forms of Lymphocytic Meningitis". Journal of Clinical Microbiology, Vol.12, No.2, p 208-215, (1980)

(2)   C.W. Moss, "Gas-Liquid Chromatography as an Analytical Tool in Microbiology". Journal of Chromatography, Vol.203, p 337-347, (1981)

(3)   F.M. LaForce, "Diagnosis of bacterial Meneingitis by Gas-Liquid Chromatography. II. Analysis of Spinal Fluid". The Journal of Infectious Diseases, Vol.140, No.4, p 453-463, (1979)

(4)   J.J. Vrbanac and C.C. Sweeley, "Automated Metabolic Profiling Analysis of Urinary Steroids by a Gas Chromatography Mass Spectrometry Data System". Biomedical Mass Spectrometry, Vol.10, No.3, p 155-161, (1983)

(5)   Ambrose."Gas Chromatography".Second Edition, London Butterworths.

(6)   D.R. Knapp, "Handbook of Analytical Derivitization Reactions". p 146-224, A Wiley-Interscience Publication, John Wiley & Sons.

(7)   C.W. Moss, "Analysis of Short-Chain Acids from Bacteria by Gas-Liquid Chromatography with a Fused-Silica Capillary Column".Journalof Clinical Microbiology, Vol.15, No.2, p 308-310, (1982)

(8)   J.L. Brice, "Major Articles". The Journal of Infectious Diseases, Vol.140, No.4, p 443-451, (1979)

(9)   J. Rosenfeld, M. Russell, A. Phatak, "Simultaneous Extraction and Derivatization of Organic Acids". Journal of Chromatography, in press.

(10)   E. Grushka, "Characterization of Exponentially
         Modified Gaussion Peaks in Chromatography".
         Analytical Chemistry, Vol.44, No.11, p 1733-1738,
         (1972)

(11)   J.C. Sternberg, "Advances in Chromatography".
         Volume 2, J.C. Giddings and R.A. Keller, Ed.,
         Marcel Dekker, New York, N.Y., (1966)

(12)   I.G. McWilliam, and H.C. Bolton, "Instrumental
         Peak Distortion 1. Relaxation Time Effects".
         Analytical Chemistry, Vol 41, p 1755, (1969)

(13)   C.H. Lochmuller and M. Sumner, "Estimation
         of Extra-Column Dead Volume Effects Using a
         Mixing Cell Model". Journal of Chromatography
         Science, Vol.18, p 159-165, (1980)

(14)   W.W. Yau, "Characterizing Skewed Chromatographic
         Band Broadening". Analytical Chemistry, Vol.49,
         No.3, p 395, (1977)

(15)   T.J. Kennett and W.V. Prestwich, "On The
         Deconvolution of Exponential Responce Functions".
         Phys. Med. Biol., Vol.24, No.6, p 1107-1122,
         (1979)

(16)   A. Robertson, "An Automatic Peak-Extraction
         Technique". Nucl. Instr. and Meth., No.100,
         p 317, (1972)

(17)   W.A. Dietz and B.F., Jr., Dudenbastel, "Advances
         in Gas Chromatography". A.C.S., p 175,
         (Sept. 1957)