AD NUMBER

ADB174940

NEW LIMITATION CHANGE

TO

Approved for public release, distribution unlimited

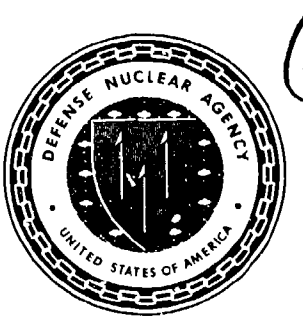
FROM

Distribution authorized to U.S. Gov't. agencies and their contractors; Administrative/Operational Use; 24 Feb 93. Other requests shall be referred to Defense Nuclear Agency, 6801 Telegraph Road, Alexandria, VA 22310- 3398.

AUTHORITY

DTRA ltr, 1 Aug 2001

Defense Nuclear Agency
Alexandria, VA 22310-3398



DNA-TR-92-196

Task 2 Report
Algorithm Development and Performance Analysis

Robert V. Mustacich
William Foreman
Paul M. Holland
General Research Corporation
P.O. Box 6770
Santa Barbara, CA 93160-6770

July 1993

Technical Report

CONTRACT No. DNA 001-91-C-0146

DESTRUCTION NOTICE:

FOR CLASSIFIED documents, follow the procedures in DoD 5200.22-M, Industrial Security Manual, Section II-19.

FOR UNCLASSIFIED, limited documents, destroy by any method that will prevent disclosure of contents or reconstruction of the document.

Retention of this document by DoD contractors is authorized in accordance with DoD 5220.22-M, Industrial Security Manual.

PLEASE NOTIFY THE DEFENSE NUCLEAR AGENCY, ATTN: CSTI, 6801 TELEGRAPH ROAD, ALEXANDRIA, VA 22310-3398, IF YOUR ADDRESS IS INCORRECT, IF YOU WISH IT DELETED FROM THE DISTRIBUTION LIST, OR IF THE ADDRESSEE IS NO LONGER EMPLOYED BY YOUR ORGANIZATION.

DISTRIBUTION LIST UPDATE

This mailer is provided to enable DNA to maintain current distribution lists for reports. (We would appreciate your providing the requested information.)

NOTE:
Please return the mailing label from the document so that any additions, changes, corrections or deletions can be made easily.

☐ Add the individual listed to your distribution list.

☐ Delete the cited organization/individual.

☐ Change of address.

NAME: _____

ORGANIZATION: _____

OLD ADDRESS	CURRENT ADDRESS
_____	_____
_____	_____
_____	_____

TELEPHONE NUMBER: (___) _____

DNA PUBLICATION NUMBER/TITLE	CHANGES/DELETIONS/ADDITIONS, etc. *(Attach Sheet if more Space is Required)*
_____	_____
_____	_____
_____	_____

DNA OR OTHER GOVERNMENT CONTRACT NUMBER: _____

CERTIFICATION OF NEED-TO-KNOW BY GOVERNMENT SPONSOR (if other than DNA):

SPONSORING ORGANIZATION: _____

CONTRACTING OFFICER OR REPRESENTATIVE: _____

SIGNATURE: _____

CUT HERE AND RETURN

DEFENSE NUCLEAR AGENCY
ATTN: TITL
6801 TELEGRAPH ROAD
ALEXANDRIA, VA 22310-3398

DEFENSE NUCLEAR AGENCY
ATTN: TITL
6801 TELEGRAPH ROAD
ALEXANDRIA, VA 22310-3398

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	930701	Technical 920701-921220

4. TITLE AND SUBTITLE
Task 2 Report
Algorithm Development and Performance Analysis

5. FUNDING NUMBERS
C - DNA 001-91-C-0146
PE - 62715H
PR - TA
TA - TA
WU - DH316290

6. AUTHOR(S)
Robert V. Mustacich, William Foreman, and
Paul M. Holland

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
General Research Corporation
P.O. Box 6770
Santa Barbara, CA 93160-6770

8. PERFORMING ORGANIZATION REPORT NUMBER
CR-92-1331

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Defense Nuclear Agency
6801 Telegraph Road
Alexandria, VA 22310-3398
OPAC/Fox

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
DNA-TR-92-196

11. SUPPLEMENTARY NOTES
This work was sponsored by the Defense Nuclear Agency under RDT&E RMC Code B4613D TA TA 00066 2600A 25904D.

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Distribution authorized to U.S. Government agencies and their contractors; Administrative or Operational Use, 24 February 1993. Other requests for this document shall be referred to the Defense Nuclear Agency, 6801 Telegraph Road, Alexandria, VA 22310-3398.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This report covers "Task 2: Algorithm Development and Performance Analysis," for the program "Development of a Handheld Chemical Detector Using Microchip GC." The study described in this report addresses the algorithmic performance requirements for the concept handheld instrument for vapor detection of CW-related materials using microchip GC. Included in this report are example signal processing and algorithmic analyses using actual microchip GC data. An overall algorithmic approach for the concept handheld CW detection instrument is developed and presented. Approaches investigated in this study include: pre-processing of chromatograms by different smoothing methods; the application of various peak detection methods; chromatographic correlations using CW data reference tables; and algorithms for merged peak analysis. To best illustrate the robustness of the recommended approach, our signal processing approach is successfully applied to several microchip GC data sets which confound existing methods.

14. SUBJECT TERMS
CW Agent Detection
Microchip Gas Chromatography
Correlated Gas Chromatography
CW Treaty Verification

15. NUMBER OF PAGES
56

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAR

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18
298-102

CLASSIFIED BY:

N/A since Unclassified.

DECLASSIFY ON.

N/A since Unclassified.

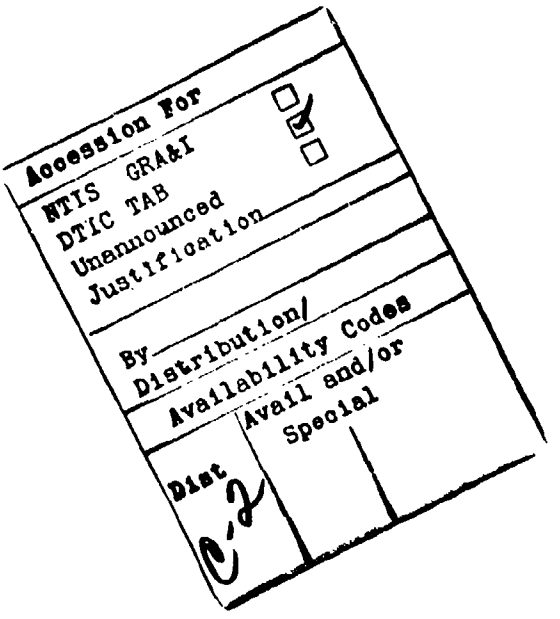
CONVERSION TABLE

Conversion factors for U.S. customary to metric (SI) units of measurement.

To Convert From	To	Multiply By
angstrom	meters (m)	1.000 000 X E -10
atmosphere (normal)	kilo pascal (kPa)	1.013 25 X E +2
bar	Kilo pascal (kPa)	1.000 000 X E +2
barn	meter^2 (m^2)	1.000 000 X E -28
British thermal unit (thermochemical)	joule (J)	1.054 350 X E +3
cal (thermochemical)/cm^2$	mega joule/m^2 (MJ/m^2)	4.184 000 X E -2
calorie (thermochemical)$	joule (J)	4.184 000
calorie (thermochemical)/g$	joule per kilogram (J/kg)*	4.184 000 X E +3
curie$	giga becquerel (GBq)†	3.700 000 X E +1
degree Celsius‡	degree kelvin (K)	$t_K = t°_C + 273.15$
degree (angle)	radian (rad)	1.745 329 X E -2
degree Fahrenheit	degree kelvin (K)	$t_K = t°_F + 459.67)/1.8$
electron volt$	joule (J)	1.602 19 X E -19
erg$	joule (J)	1.000 000 X E -7
erg/second	watt (W)	1.000 000 X E -7
foot	meter (m)	3.048 000 X E -1
foot-pound-force	joule (J)	1.355 818
gallon (U.S. liquid)	meter^3 (m^3)	3.785 412 X E -3
inch	meter (m)	2.540 000 X E -2
jerk	joule (J)	1.000 000 X E +9
joule/kilogram (J/kg) (radiation dose absorbed)$	gray (Gy)*	1.000 000 X E +91
kilotons$	terajoules	4.183
kip (1000 lbf)	newton (N)	4.448 222 X E +3
kip/inch^2 (ksi)	kilo pascal (kPa)	6.894 757 X E +3
ktap	newton-second/m^2 (N-s/m^2)	1.000 000 X E +2
micron	meter (m)	1.000 000 X E -6
mil	meter (m)	2.540 000 X E -5
mile (international)	meter (m)	1.609 344 X E +3
ounce	kilogram (kg)	2.834 952 X E -2
pound-force (lbf avoirdupois)	newton (N)	4.448 222
pound-force inch	newton-meter (N•m)	1.129 848 X E -1
pound-force/inch	newton/meter (N/m)	1.751 268 X E +2
pound-force/foot^2	kilo pascal (kPa)	4.788 026 X E -2
pound-force/inch^2 (psi)	kilo pascal (kPa)	6.894 757
pound-mass (lbm avoirdupois)	kilogram (kg)	4.535 924 X E -1
pound-mass-foot^2 (moment of inertia)	kilogram-meter^2 (kg•m^2)	4.214 011 X E -2
pound-mass/foot^3	kilogram-meter^3 (kg/m^3)	1.601 846 X E +1
rad (radiation dose absorbed)$	gray (Gy)*	1.000 000 X E -2
roentgen$	coulomb/kilogram (C/kg)	2.579 760 X E -4
shake	second (s)	1.000 000 X E -8
slug	kilogram (kg)	1.459 390 X E +1
torr (mm Hg, 0°C)	kilo pascal (kPa)	1.333 22 X E -1

*The gray (Gy) is the accepted SI unit equivalent to the energy imparted by ionizing radiation to mass of energy corresponding to joule/kilogram.
†The becquerel (Bq) is the SI unit of radioactivity; 1 Bq = 1 event/s.
‡Temperature may be reported in degree Celsius as well as degree kelvin.
$These units should not be converted in DNA technical reports; however, a parenthetical conversion is permitted at the author's discretion.

TABLE OF CONTENTS



FIGURES

FIGURES (Continued)

SECTION 1
INTRODUCTION

The draft Chemical Weapons Convention (CWC) completed by the Conference on Disarmament on 3 September 1992, and scheduled for signature on 13 January 1993, addresses the significant threat to peace posed by the proliferation of chemical weapons technology. A key part of this treaty is the provision for on-site challenge inspections for the presence of prohibited Chemical Weapons (CW) and pre-cursor chemicals.

The proposed language of the treaty specifies that, "where possible, the analysis of samples shall be performed on-site." Since the allowable time for challenge inspections under the treaty is limited, this places a premium on rapid, accurate analysis for treaty prohibited materials. It is expected that challenge inspections will be requested by foreign governments at U.S. sites where proprietary industrial and sensitive military research and/or production are being carried out, and thus the analysis of samples with general purpose techniques such as Gas Chromatography/Mass Spectrometry (GC/MS) might be deliberately used to probe for important information about commercial and military projects unrelated to chemical weapons. This situation poses significant legal and security questions concerning U.S. industrial competitiveness and national security, and places a high premium on the development of inspection technologies which will provide accurate analyses for treaty prohibited materials *without* revealing the identity of other chemical compounds.

The present program to develop a handheld detector utilizing microchip GC (μGC) addresses the needs for portability, speed of operation, sensitivity, and chemical selectivity through the use of miniature GC components, preconcentration, and the use of correlated gas chromatography for chemical selectivity. General Research Corporation's (GRC's) concept combines these functional components to provide a handheld detector for treaty verification applications. Special signal processing is then used to analyze the results of the correlated chromatograms to determine the presence of any treaty prohibited materials.

An illustration of two-dimensional correlated gas chromatography and the resulting chemical detection selectivity is shown in Figure 1-1. In this example, a vapor sample is split and simultaneously analyzed on two GC columns having differing separation properties. Each chromatogram displays the signal from the column's detector as a function of time. For example, compound 1 passes through the column most rapidly while compound 9 is the slowest. The time for a compound to travel through a GC column is referred to as its elution or retention time. Importantly, both the elution times and relative order of the compounds in chromatogram A can vary from those in chromatogram B. These differences arise from differences in the chemical compositions of the polymeric coatings inside the two GC columns. Deliberate choice of coatings which provide contrasting separations can provide a high degree of chemical selectivity in correlated gas chromatography. This is illustrated by the two-dimensional representation at the right in Figure 1-1 which shows each compound in the correlated chromatogram plotted using its pair of elution times. This approach provides a large "detection space" in which each chemical target compound is represented by its specific coordinates in this space.

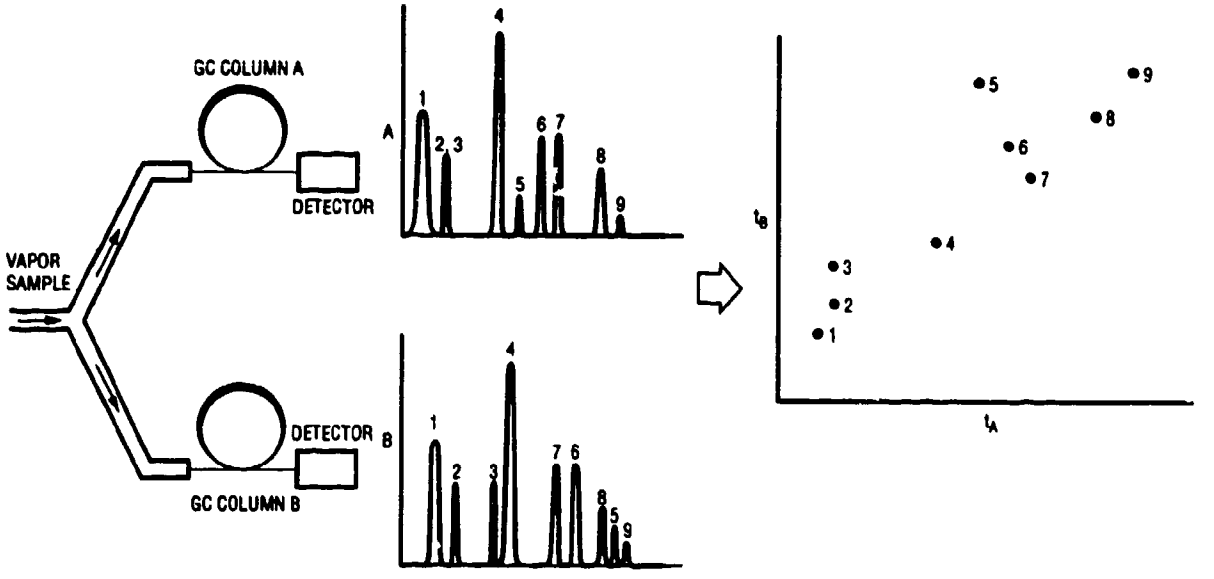


Figure 1-1. Correlation of two different chromatograms for a sample and the corresponding two-dimensional detection space.

The advantage of this enlargement of the detection space by correlated chromatography can be seen as follows. While ordinary, one-dimensional GC is an excellent chemical separation technique, in situations where the chemical backgrounds are complex as environmental sampling, only 30-40% of the peaks in the chromatogram may be single components. It is then up to the detector to distinguish between chemical components in the remaining overlapping (multicomponent) peaks. This leads to much of the ambiguity that can arise in GC/MS with trace environmental samples, for example. Correlated chromatography, on the other hand, separates the same complex mixture of compounds in two or more different ways. If, for example, a number of compounds happen to co-elute on one column, they are unlikely to co-elute on a second column which has different separation properties. As an example, peaks 2 and 3 in Figure 1-1 are not separated by column A, but are well separated by column B. Thus, the use of a second correlated column can greatly reduce the overlap ambiguity between different compounds. Extending this approach to three or more dimensions can further enhance selectivity. This can result in chemical selectivity similar to that of MS/MS using correlated gas chromatography.

The search for target CW compounds in an unknown sample using correlated chromatography is made by checking the known retention (or elution) time coordinates in detection space. This report presents and discusses the algorithms required to accomplish this task along with performance testing of the proposed algorithms using μGC correlated chromatography data for representative CW precursors in Schedules II and III. Following an overview discussion of the detection algorithm, specific algorithmic needs are presented in the following sections. These include the pre-processing and the detection of chromatographic components, the calculation and correlation of retention indices, the confirmation of potential detections, and the assessment of algorithm performance using μGC data.

SECTION 2
DETECTION ALGORITHM OVERVIEW

The detection algorithm for correlated μGC is expected to be less complex than the illustration in Figure 1-1 might suggest. It is useful to conceptualize the two- or higher-dimensional detection space to explain the high degree of chemical selectivity attainable with this technique, but the actual detection algorithm need not extensively correlate the chromatograms. Because the CW target signatures themselves consist of correlated data, it is possible to independently scan each chromatogram for peaks contained in the CW target signatures and then check any peaks found for proper correlation according to the CW signature table. This process might be thought of as analogous to placing the "grading key" of a multiple choice test over an answer sheet to check for possible "correct" answers. Thus, rather than a detailed construction of detection space, it may be most expedient to "screen" each chromatogram for possible CW signature components and then check a short list of any such matches by correlation. This approach can be thought of as a "table driven" algorithm that focuses only on the detection targets.

The overall process is depicted in Figure 2-1 and summarized by the steps listed in Table 2-1. In the first step, the chromatograms are first smoothed using routines to reduce the noise. Following these routines, the peaks corresponding to different compounds in the chromatograms are located. Apex identification for these peaks is a relatively simple process. It does not appear necessary at this stage of the automated processing of the chromatograms to determine peak starting and ending points. Since peaks may overlap, the location of the ending point for a current peak can be complicated due to new peaks starting before the current peak ends. This can require a "recursive" search for a new peak apex and end point before locating the current peak end point. This becomes an important issue for later post processing such as peak integration, for example, but in the initial processing step may not be required. An example chromatogram is shown in Figure 2-1 where the apexes of the peaks in the chromatograms are indicated by the triangles. These peak apexes are numbered for reference. Unlike the case in Figure 1-1, use of the same numbers in labeling the peaks of both chromatograms does not imply these are the same compounds.

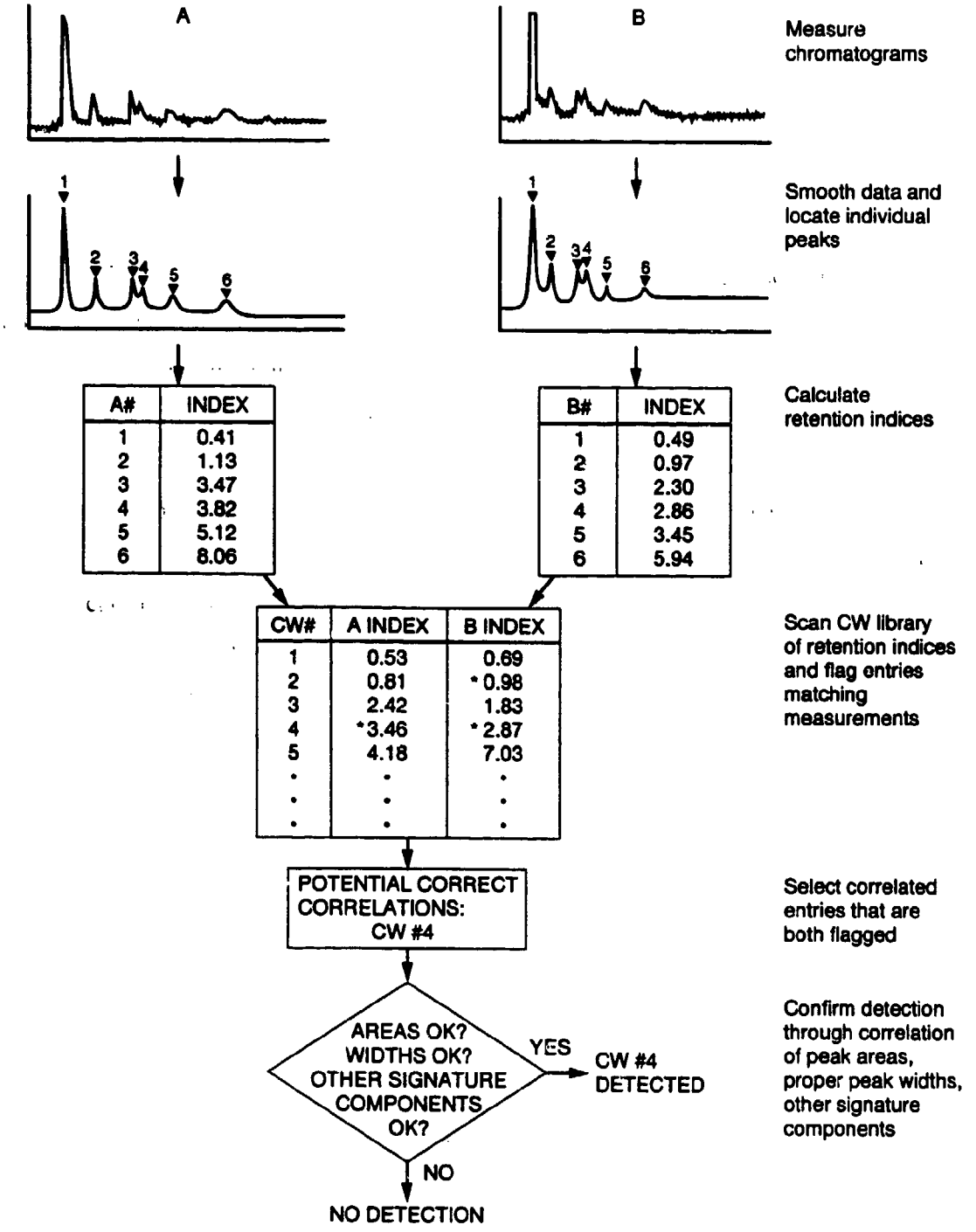


Figure 2-1. Overview of the signal processing approach proposed for the handheld CW detector.

The second signal processing step involves the calculation of retention indices for the peaks. The concept μGC instrument will need periodic calibration with reference compounds. This calibration could be made using either an occasionally administered external standard or an internal standard. An external standard might consist of a sample having a known vapor composition that the handheld instrument is trained to recognize and accordingly use for calibrating its response to this vapor. An internal standard could depend on the use of compounds present in the background in many or all of the chromatograms for retention time calibration. By either approach, the measurement of the retention times for some calibrant compounds provides corrections needed to adjust for chromatographic performance variations due to small changes in gas pressures, temperatures, or the GC columns themselves. The result for each column is a list of retention indices for the peaks observed.

In the next step of the proposed detection process, the library of retention indices for the target CW materials is scanned to determine if any of the observed peak retention indices coincide with these values. The variability of the instrument's response after calibration is used determines the closeness of values required for a match. For example, allowing retention index differences of 0.02, results in one match for the A column chromatogram and two matches for the B column chromatogram as illustrated in Figure 2-1. These matches are highlighted by asterisks in the CW table. Note that no computational processing has yet been required for correlation purposes.

By restricting the consideration of correlation only to peaks that first match retention indices for target materials in each chromatographic analysis, the number of components which need to be further analyzed is minimized. Thus, the correlation between columns consists of analyzing only the flagged entries in the CW library of retention indices for correct correlation. If retention indices of a given CW compound are flagged in both columns, then there is a potential observation of this particular compound. Two cases are illustrated in Figure 2-1. In the first, correlated observations for CW compound #4 indicate a potential detection of this compound. In the second case, an uncorrelated observation for CW

compound #2 in column B only fails to indicate the presence of compound #2. Thus, the correlation is simplified by utilizing the CW library of correlated retention indices.

The final step in the process is the confirmation of the detection of a given compound through the correlation of integrated areas, the examination of expected peak widths or shapes, and the consideration of other signature components such as the possible appearance of associated chemical compounds. This is the most computationally intensive step of the overall detection algorithm. Importantly, the use of this approach limits the amount of data which must be post-processed in the final step for confirming correlated detections.

The fact that the samples are split between the pair of GC columns requires that correctly correlated peaks have areas with a ratio which corresponds to the splitting factor. For example, if the splitting factor is calibrated to be 60:40, then the corresponding ratio of the areas should be 1.5 within experimental error. A substantially different ratio would invalidate the detection. Similarly, the expected peak widths for different compounds can be included in the CW library for various operating conditions and compared with the observed widths for confirmation of the detection. Again, poor correspondence of an observed width with that expected can invalidate the detection. An example of an errant peak type which may result from peak locating algorithms is a spurious "peak" resulting from changes in the chromatographic baseline. Such "peaks" are easily removed from consideration by examining their large widths. Other signature elements which also may be considered are the presence of additional signature components which are associated with certain CW materials. Examples are the simultaneous appearance in chromatograms of known decomposition products or impurities related to the chemical synthesis of the CW compounds.

A more detailed discussion of the steps followed in the proposed algorithms summarized in Table 2-1 will be presented in the sections that follow. In each section, μGC experimental data will be included for demonstrating and testing the algorithmic steps and discussing potential alternative approaches. The features used in the proposed algorithm will be discussed in more detail and assessments of the algorithmic performance using actual μGC data as the different steps are presented. The occurrence of signatures with multiple

components may provide the need for artificial intelligence algorithms to analyze multiple features. This will also be discussed in a later section in this report.

Table 2-1. Proposed algorithmic steps.

1. Smooth the chromatographic data, locate individual peaks, and determine the corresponding retention times
2. Calculate the corresponding retention indices according to the retention times of reference compounds
3. Scan the list of the retention indices for any times which coincide with the known retention indices of the target CW materials for either column
4. Compare any observations coinciding with CW targets for each column for correct correlation with observations on the other column
5. Confirm correlated observations for CW targets by integration, peak width, and the presence of possible coincident compounds in the signatures

SECTION 3
DETECTOR OUTPUT AND DATA SMOOTHING

It is usually advantageous to smooth chromatographic data as a prelude to automated processing. Importantly, the routines used for the detection of peak features such as starts and apexes operate more robustly and with greater sensitivity using smoothed data to provide a greatly increased number of peak detections. The degree of smoothing and the specific approaches applied are typically determined by the instrumentation used and the expected application. For applications seeking sensitive detection, the noise levels from the components of the instrument and the chemical background noise are combined to determine the practical noise levels for the threshold detection of target compounds. The chief source of noise for chromatographic detection by the μGC hardware in its present form is detector noise which is probably dominated by noise levels in the motherboard layout of the MTI instrument. It is also important to note that the specific choice of the chromatographic detector can determine the relative importance of many common forms of chemical background noise. For example, the noise floor typically encountered with ion trap mass spectrometers as detectors for GC is typically set by chemical background due to relatively constant "bleed" of chemicals from the columns and nonmetallic structural components of the ion trap.

The thermal conductivity detectors (TCD) in the μGC instrument appear to be more dominated by electrical noise within the instrument according to experimentation conducted at Louisiana State University with μGC instruments (Overton, 1992). The baseline sensitivity of the miniature TCDs is about 1 ppm (volume/volume) (Lee et al., 1989). Similar to the μGC injector system, the TCDs is a product of microfabrication using silicon. They consist of two electrical filaments suspended in each detection channel. A closeup photograph of the detector housing a matched pair of TCDs is shown in Figure 3-1. Advantages of silicon microfabrication of TCDs include: the ability to optimally scale the detector size for concentration sensitivity and removing the need to dilute samples with make-up gas; the ability to match filaments since they are fabricated from the same thin film deposition; and good thermal conductivity of silicon which affects warm up, thermal stability, and the linear

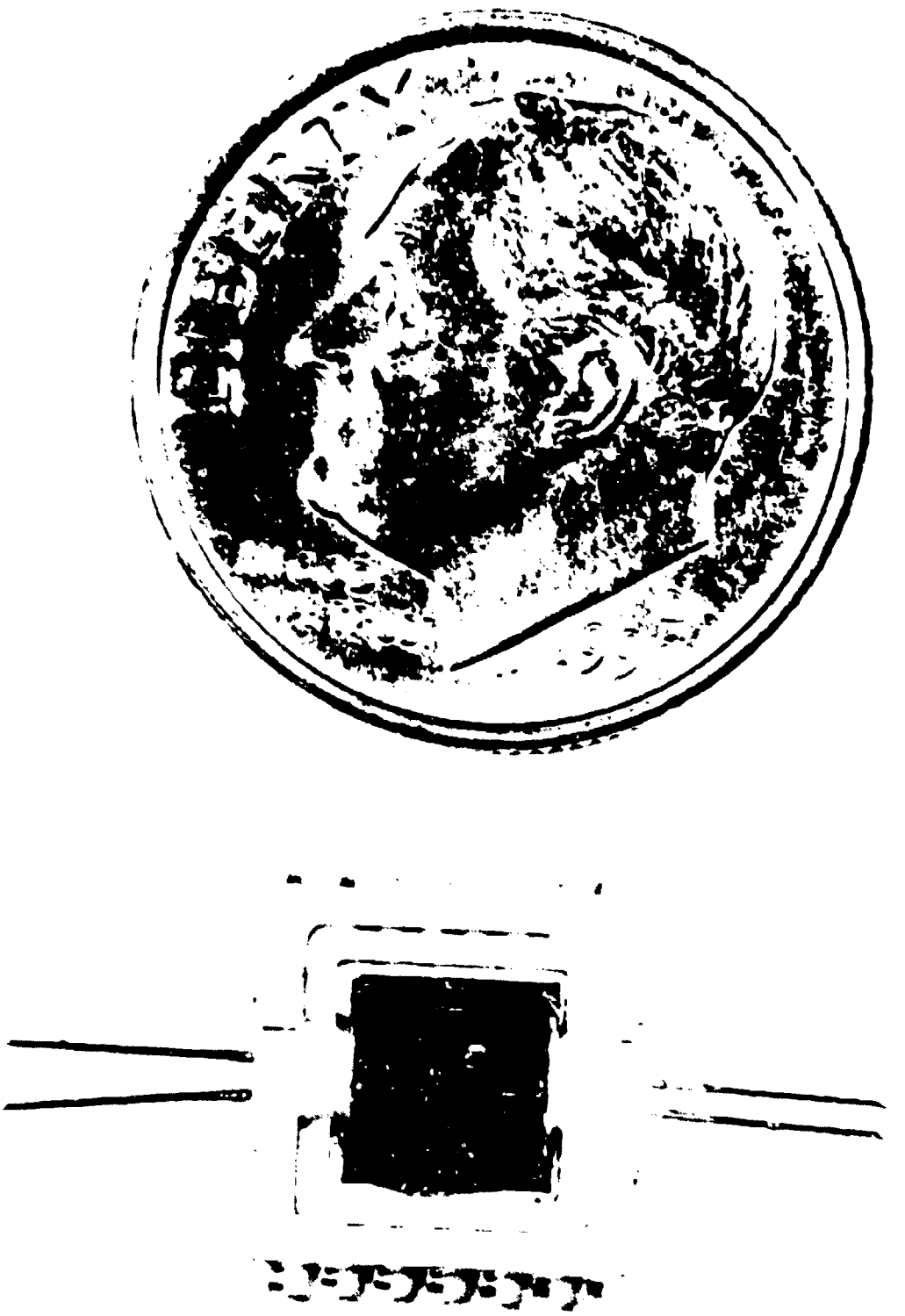


Figure 3-1. Closeup photograph of a pair of microfabricated thermal conductivity detectors.

dynamic range. The TCDs illustrated in Figure 3-1 have a dead volume of 2 nL and a time constant of less than 10 msec. The response is highly linear over many orders of magnitude (Lee et al., 1989). With respect to injections of 100 nL, typical analytes are measured in the femtogram range. With preconcentration in the form of thermal desorption of vapors from adsorbent traps, the effective sensitivity of the TCD is driven typically into the 10's of ppb range (Overton, 1992).

The sensitivity of the μGC modules in terms of electrical noise according to Microsensor Technology Inc. (MTI) is typically about 1200 μV (Microsensor Technology Inc, 1992). By screening the production lots, MTI has been able to select modules with noise levels in the 400-600 μV range (Microsensor Technology Inc., 1992). These noise levels apply to the "high gain" operation of the MTI units, one of three amplifier gain settings. The high gain setting corresponds to amplification of the detector signal by a factor of 500. Thus, this corresponds to a noise level of about 1 μV at the detector. According to LSU, it may be possible to reduce the noise level significantly through isolation of the detection circuitry from the remainder of the motherboard (Overton, 1992).

A number of different approaches have previously been tested for smoothing chromatographic data before automated peak detection. The short time constant of the TCDs used with the μGC allows data rates of 100 data points/sec per chromatogram. This results in a substantial number of points relative to the number necessary for accurately describing the peak shapes. This excess of data points can also be exploited for noise reduction as well. Three primary techniques which have been used singly or in combination for digital filtering of chromatograms are as follows:

a. Data Bundling
b. Moving Average Filters
c. Linear Least Squares Fits to Line Segments

Any of these smoothing techniques may be adequate for μGC data analysis, and they are often used in combination (Woerlee & Mol, 1980). These techniques are described in the following paragraphs.

Data bundling refers to grouping the data points and summing the groups to result in a new representation of the data using fewer points. Group sizes in powers of two are typically used for this bundling. This summation reduces the noise by a factor of $\sqrt{2}$ for every doubling up of data points. Since a large data rate is possible with the TCD because of its short time constant (less than 10 msec), an "excessive" data collection rate can be advantageously used to reduce the noise by this data bundling technique. It appears that the software developed for μGC instrumentation by LSU performs data bundling with groups of four data points based upon our measurements of the MTI data rate and the data density of the files produced by LSU's software (25 points/sec). This degree of data bundling with GC chromatograms was previously discussed to be advantageous for automated analysis (Woerlee & Mol, 1980).

An example of 4-point data bundling applied to μGC data is shown in Figure 3-2. The resulting smoothness of the curve which results from the bundling is readily apparent. More importantly, this small peak in the chromatogram goes undetected without the bundling because the noise at the start of the peak invalidates the peak start detection routines. Because these routines are very noise sensitive, they examine the trend of a number of points for determining the departure from the baseline that establishes the start of the peak. Smoothing, however, greatly reduces this noise and allows the routines to correctly see the required successive lift of the points from the chromatographic baseline. Similarly, the routines misplace the apex of this peak due to fluctuations near the apex if smoothing is not used.

Moving average filters also sum the data points over a specified window but differ from data bundling in that the window and its resulting sum moves along the chromatogram point by point. A four point moving average, for example, would begin by summing the first four points. Next, the moving average would move over one point by subtracting the first

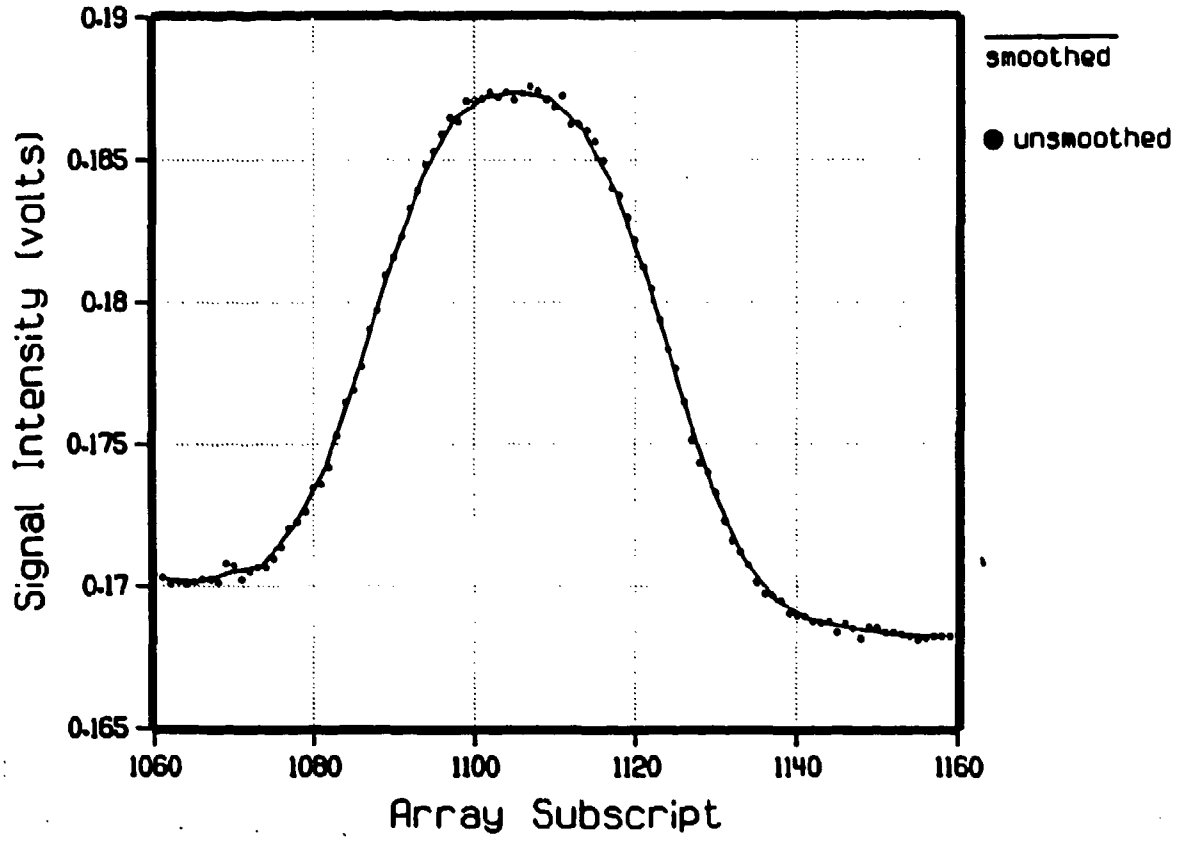


Figure 3-2. Example μGC data showing 4-point data bundling (smooth curve).

point and adding the fifth point from the moving average to result in the sum of points 2-5. Similarly, the average moves next to sum the points 3-6. The features in this type of moving average lag the actual data due to the effect of averaging in earlier data values. For this reason, the moving average is "centered" by positioning the window such that equal numbers of data points behind and ahead of the window center are averaged. This positions the moving average to coincide with the data features.

Another effect of simple moving averages is that the edge values are as important as the other values in computing the average. The fluctuations in the leading edge values and similar fluctuations in the trailing edge or "drop off" values combine additively. This can sometimes be unfavorable, especially with short moving averages. To damp such edge effects on the computation of the moving average, the averages are often "center weighted" by applying weights to the points in the computation of the moving average which favor the center points in the window. Examples of the favorable application of center-weighted

averages to chromatograms are provided in the literature (Woerlee and Mol, 1980; Fozard et al., 1972).

An example of a 7-point weighted moving average to μGC data is shown in Figure 3-3. This example uses the same data of the small peak discussed in Figure 3-3. The weights are described by the seventh row of Pascal's triangle and are identical to the binomial expansion coefficients for $(a+b)^6$: 1,6,15,20,15,6,1. The underweighting of edges can lead to small negative valleys near peak starts. This can be partially compensated by appropriate adjustments in the peak start position (Woerlee and Mol, 1980). Again, smoothing results in a large reduction in the noise and a large improvement in the operation of the peak detection algorithms. The overall curve is well described by the moving average, but is not quite as smooth as the result using 4-point bundling. In practice, it may be advantageous to apply this type of moving average after bundling the data.

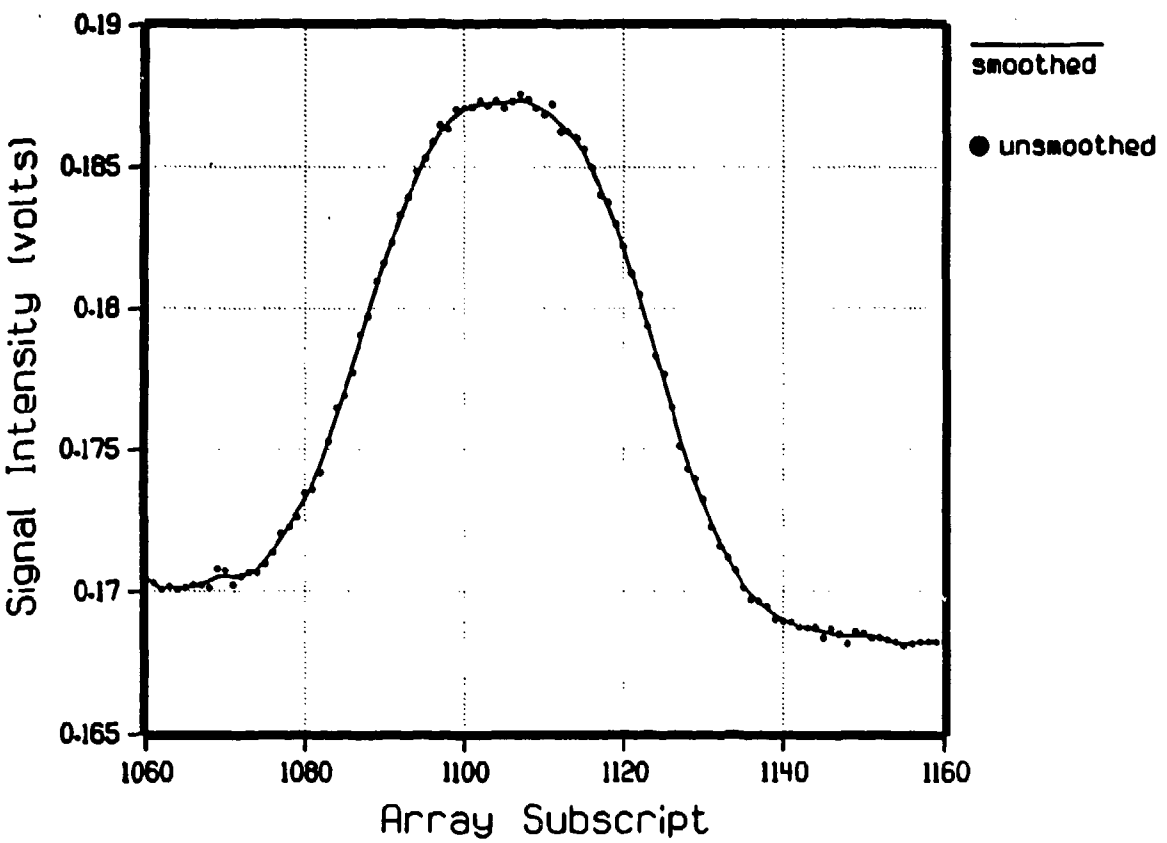


Figure 3-3. Smoothed μGC data using 7-point Pascal-weighted moving average.

A third type of smoothing routine is the least squares fitting of data points to lines. Such an approach using a five point window has been tested previously (Li et al., 1987). The smoothed value for the central point of the five data points was taken as the value of the linear fit at that point. These investigators did not compare their approach with previously published methods described above. Their overall approach appears to be less sophisticated than the other approaches referenced above, especially with regard to resolving merged peaks and handling baseline drifting. This approach appears to be computationally more cumbersome, and in view of the well documented successes with the other methods, we have not subjected this approach to comparative testing.

Our analysis of these smoothing approaches suggests that data bundling and/or center-weighted moving averaging provide a large degree of noise reduction. This noise reduction greatly facilitates the detection of peak starts and apexes with regards to both accuracy and sensitivity. Increased sensitivity results in more detections by the algorithms for enhanced detection results. Figure 3-2 provides an example of a small peak with proper shape which was not detected without smoothing.

SECTION 4
PEAK LOCATION METHODS

Compared to determining the location of peak starts and peak endings, the determination of the peak apex is a simpler task. In automated GC analyses directed at the entire chromatograms, the chromatograms are typically analyzed from left to right, i.e., as a function of their time evolution. In this approach, a baseline start is normally presumed and the initial search target is for baseline departures which indicate peak starting points. Once a peak start is located, the search continues to the right to locate the peak apex. Once the peak apex is found, then a return to baseline is analyzed. The return to baseline is often complicated by the presence of "rider" peaks on the tail of the peak being analyzed. This complication can be avoided on the first signal processing pass if the detection algorithm focuses only on peak apexes which correspond to potential detection targets of interest. These peak apexes can be identified relatively easily in the chromatograms, and if there are no potential target components present, the unnecessary detailed analysis of the chromatograms can be avoided.

A criterion often used for the determination of a peak apex is a trend reversal consisting of the first observation of four successively decreasing points (Woerlee and Mol, 1980). Low frequency noise and different peak forms are possible complications in this process, especially with flat peaks. The rejection of maxima and valleys on flat peaks is especially important to avoid faulty splitting of the peak. An example of peak detection using this criterion with μGC data is shown in Figure 4-1. Following the large "air spike" seen initially by the detector, four peaks are detected. Note that this data is not smoothed, and the small peak near 1100 is not detected. This peak is detected after smoothing the data and is the peak which was shown in Figures 3-2 and 3-3 to demonstrate the effects of smoothing.

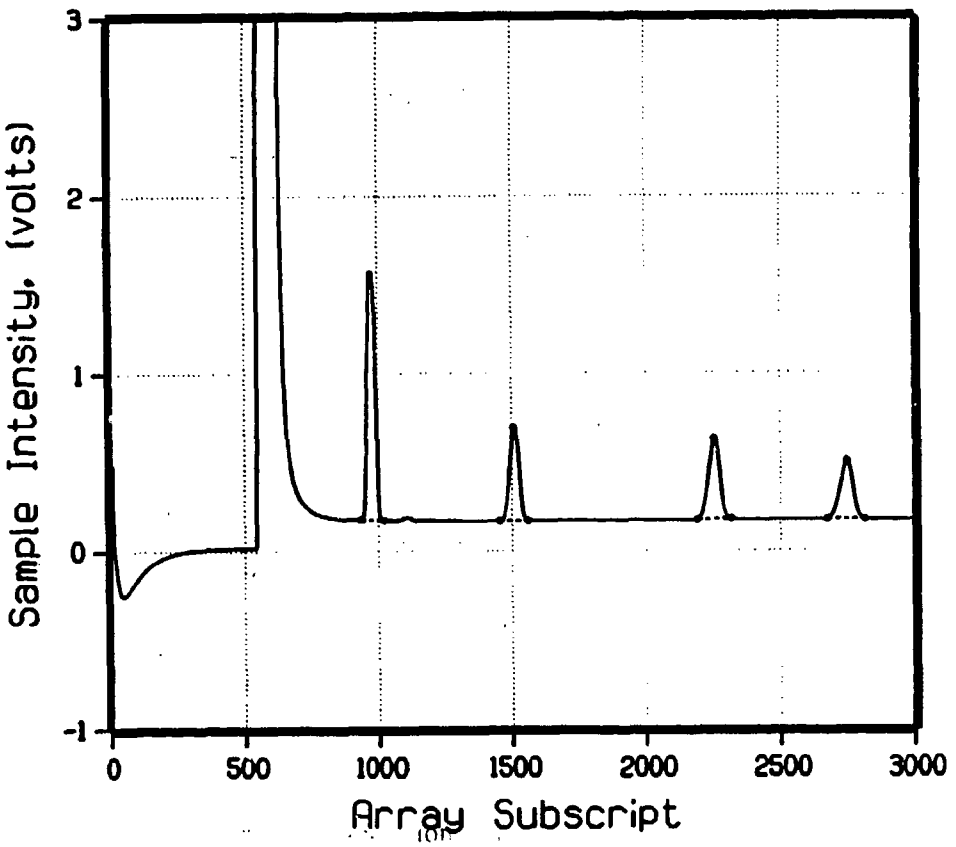


Figure 4-1. Example demonstrating peak detection with μGC data.

SECTION 5
THE CALCULATION OF RETENTION INDICES

The standard method for the analysis of chromatographic components utilizes retention indices. While a variety of different retention indices can be defined, the most commonly used retention index method is the Kovat's Retention Index system (for review, see Budahegyi et al., 1983). In this method the retention index of a compound describes its chromatographic elution relative to a series of normal alkanes (saturated hydrocarbons). The number of carbons in the alkane times 100 defines the retention index of the hydrocarbon. For example, pentane and hexane have retention indices of 500 and 600, respectively.

Isothermal retention times of the n-alkanes tend to follow an exponential function of the carbon number for carbon numbers of 4 and greater. Here, the elution of the n-alkanes can be described by the equation:

$$\log (t_R(z)-t_M) = bz + a \qquad (5.1)$$

where $t_R(z)$ is the retention time of the alkane having carbon number z, t_M is the gas holdup time, b is the slope, and a is the intercept. Sometimes the difference $t_R(z)-t_M$ is referred to as the adjusted retention time $t'_R(z)$. The slope, b, can also be expressed as

$$b = \log t'_R(z+1) - \log t'_R(z). \qquad (5.2)$$

While minor variations occurring in chromatographic conditions day to day will cause retention times to vary, these variations can be accounted and corrected for by periodic chromatography of calibrant materials. If, for example, a set of n-alkanes was chromatographed, a fit of equation 5.1 to the measured retention times would give current values of a and b for calculating retention times for compounds eluting within the range of the calibration curve. The need for instrument recalibration would be established according to the inherent stability of the instrument and its regimen of use.

Retention indices for compounds which are not n-alkanes can be determined by exponential interpolation between values of the n-alkanes as shown:

$$I_s^{st,ph,T} = 100 \left[\frac{\log t_R(s) - \log t_R(z)}{\log t_R(z+1) - \log t_R(z)} + z \right] \tag{5.3}$$

Compounds other than n-alkanes can also be used for calibrants using their retention indices in place of the integer values of the n-alkanes in equation 5.1. This provides flexibility in designing calibrations for field instrumentation using the most practical calibrants.

While the gas holdup time can be approximated by the air peak in the chromatograms as detected by the TCD (see Figure 4-1), the gas holdup time is more accurately determined by the retention times of three calibrants (Peterson and Hirsch, 1959). For example, using three consecutive n-alkanes, gas holdup time is calculated as

$$t_M = \frac{t_R^2(z+1) - t_R(z)t_R(z+2)}{2t_R(z+1) - t_R(z) - t_R(z+2)} \tag{5.4}$$

Errors in retention indices can arise from a number of sources. For correct GC identification of target materials using our algorithms, it is necessary to minimize possible sources of error and understand where these can enter into the analysis. Especially important is variation in the reproducibility of retention time measurements for the target compounds. This will determine the width of the data acceptance windows for potential matches between the retention indices calculated for the chromatograms and the reference retention indices for the CW target materials. Major sources of error in retention indices include (Budahegyi et al., 1983):

a. Errors in the determination of retention times (measurement error, influence on the amount of sample, operator errors);

b. Errors in the measurement or calculation of the gas holdup time (e.g., calculation is carried out in the nonlinear range of n-alkanes);

c. Situations where the stationary phase is not uniformly coated, or is active; there are wall effects in the capillary column;

d. Situations where the stationary phase is inhomogeneous;

e. Fluctuations of instrumental and/or gas chromatographic parameters.

Equations to derive the maximum expected error in the determination of retention times have been developed. Such calculations can be used to determine the necessary acceptance windows for matching retention indices calculated from measurements with the reference retention indices in the CW target material tables. A standard expression for maximum expected error in retention index calculation is given by (Takacs and Kralic, 1970):

$$\frac{E_1}{100} \leq \left| \frac{1}{[\overline{t_R}(S) - \overline{t_M}] \log \left[\frac{\overline{t_R}(z+1) - \overline{t_M}}{\overline{t_R}(z) - \overline{t_M}} \right] 2.30258} \right| \cdot \Delta t_R(S) +$$

$$+ \left| \frac{\log \left[\frac{\overline{t_R}(z+1) - \overline{t_M}}{\overline{t_R}(S) - \overline{t_M}} \right]}{[\overline{t_M} - \overline{t_R}(z)] \log^2 \left[\frac{\overline{t_R}(z+1) - \overline{t_M}}{\overline{t_R}(z) - \overline{t_M}} \right] 2.30258} \right| \cdot \Delta t_R(z) +$$

$$+ \left| \frac{\log \left[\frac{\overline{t_R}(S) - \overline{t_M}}{\overline{t_R}(z) - \overline{t_M}} \right]}{[\overline{t_M} - \overline{t_R}(z+1)] \log^2 \left[\frac{\overline{t_R}(z+1) - \overline{t_M}}{\overline{t_R}(z) - \overline{t_M}} \right] 2.30258} \right| \cdot \Delta t_R(z+1) +$$

$$+ \left| \frac{\frac{\overline{t_R}(S) - \overline{t_R}(z)}{[\overline{t_R}(S) - \overline{t_M}][\overline{t_R}(z) - \overline{t_M}]} \cdot \log \left[\frac{\overline{t_R}(z+1) - \overline{t_M}}{\overline{t_R}(z) - \overline{t_M}} \right] - \frac{\overline{t_R}(z+1) - \overline{t_R}(z) \cdot \log \left[\frac{\overline{t_R}(z+1) - \overline{t_M}}{\overline{t_R}(z) - \overline{t_M}} \right]}{[\overline{t_R}(z) - \overline{t_M}][\overline{t_R}(z+1) - \overline{t_M}]}}{\log^2 \left[\frac{\overline{t_R}(z+1) - \overline{t_M}}{\overline{t_R}(z) - \overline{t_M}} \right] 2.30258} \right| \cdot \Delta t_M \qquad (5.5)$$

where E_1 equals total error of the retention index determination (index units) under the conditions of classical gas chromatography (i.e. constant temperature, flow rate, and inlet pressure):

$$\Delta t_R(X) = \frac{\left| t_{R,1}(X) - \overline{t_R}(X) \right| + \left| t_{R,2}(X) - \overline{t_R}(X) \right| + \ldots + \left| t_{R,g}(X) - \overline{t_R}(X) \right|}{m}$$

$$\overline{t_R}(X) = \frac{t_{R,1}(X) + t_{R,2}(X) + \ldots + t_{R,g}(X)}{m}$$

$\overline{t_R}(X)$ = average retention time;

X = x, z and $z + 1$;

g = serial number;

m = number of parallel measurements;

t_R	=	retention time;
t_M	=	time of passage of the inert substance through the column (min);
s	=	substance;
z and $z + 1$	=	n-alkanes with carbon numbers z and $z + 1$, respectively;
z	=	carbon number.

By combining the above parameters with the experimental measurements, the range of potential error in the retention indices can be projected. For a wide range of compounds measured under different experimental conditions, equation 5.5 results in errors that are typically on the order of 0.01-1.0 retention index unit depending upon the class of compound, temperatures, and specific instrumentation (Budahegyi et al., 1983).

An example calculation for phosphorus pentachloride on DB-5 at 50°C is shown below. The data used in the calculation are from some preliminary measurements using our μGC evaluation system. In these preliminary results the retention index for PCl_5 is determined to be 732.3 with $\overline{t_M}$ = 7.15 sec; Δt_M = 0.020 sec; $\overline{t_R}$(n-C_7) = 31.81 sec; $\overline{t_R}$(PCl_5) = 39.93 sec; $\overline{t_R}$(n-C_8) = 64.26 sec; Δt_R(n-C_7) = 0.0356 sec; Δt_R(PCl_5) = 0.0889 sec; Δt_R(n-C_8) = 0.151 sec. The example calculation results in an estimate of error for the retention index of 0.76 index units.

$$\frac{E_I}{100} \leq \left| \frac{1}{[39.93 - 7.15] \log \left[\frac{64.26 - 7.15}{31.81 - 7.15}\right] 2.30258} \right| \cdot 0.0889 +$$

$$+ \left| \frac{\log \left[\frac{64.26 - 7.15}{39.93 - 7.15}\right]}{[7.15 - 31.81] \log^2 \left[\frac{64.26 - 7.15}{31.81 - 7.15}\right] 2.30258} \right| \cdot 0.0356 +$$

$$+\left|\frac{\log\left[\frac{39.93-7.15}{31.81-7.15}\right]}{[7.15-64.26]\log^2\left[\frac{64.26-7.15}{31.81-7.15}\right]2.30258}\right|\cdot 0.151+$$

$$+\left|\frac{\frac{39.93-31.81}{[39.93-7.15][31.81-7.15]}\cdot\log\left[\frac{64.26-7.15}{31.81-7.15}\right]-\frac{64.26-31.81\cdot\log\left[\frac{64.26-7.15}{31.81-7.15}\right]}{[31.81-7.15]\,[64.26-7.15]}}{\log^2\left[\frac{64.26-7.15}{31.81-7.15}\right]2.30258}\right|\cdot 0.020$$

$$E_1 \leq (0.00323 + 0.001135 + 0.00107 + 0.0022)\cdot 100 = 0.76 \qquad (5.6)$$

SECTION 6
CORRELATION OF DETECTED PEAKS WITH CW TARGET SIGNATURES

Our selected approach for correlating detected peaks with CW target signatures is based on a multi-step algorithm. This begins with the "table driven" concept for CW compound detection using microchip GC which was illustrated in Figure 2-1 (overview). Here the list of retention indices resulting from correction of the measured retention data is compared with the target CW retention indices for each column. This is envisioned to proceed first on the (uncorrelated) individual chromatograms from each column. In this process, the calculated retention indices from the components of a chromatogram are each checked against the CW retention indices with reference to the acceptance window for each CW retention index. These acceptance windows are determined by the experimental measurements and the calculations described in Section 5, and are tabulated along with target CW retention indices in a table of predetermined constants.

When apparent matches are found between the measured retention indices and the CW target retention indices, these matches are "flagged" as possible detections of target compounds. In software, this could correspond to the setting of a flag bit associated with the particular CW retention time from its binary 0 (false) to a 1 (true). Proceeding through the measured array of retention indices will result in a series of flag bits being set for CW-related compounds in the table for each column. This process is done independently with respect to each of the measured chromatograms -- i.e., no correlation requirement between the GC columns has yet been invoked.

Once all of the retention indices have been flagged that were found to be present in the chromatograms, the correlation process can take place. At this stage of the process, the correlation focuses on potential target observations where only those target CW materials having both retention indices flagged are of interest. Computationally, this is a simple task in which the CW target table is scanned for entries having coincident true flags for both chromatographic columns. Importantly, this approach avoids unnecessary analysis of the chromatograms and focuses the processing on potential targets.

The observation of correlated flags, then, passes the particular target CW compound's information to further stages of processing to confirm the detection through further analysis of the appropriate regions of the chromatograms. Example information might include a code for the compound identity, the retention indices for each chromatographic column used in the instrument, the acceptance window for each retention index, detection flags, input splitting ratio for the chromatographs with reference to the compound, peak widths, peak asymmetry values, instrument response factors, and additional information regarding the possible presence of associated compounds which may further confirm the identification.

SECTION 7
DETECTION CONFIRMATION METHODS

The descriptions of the detection algorithm in the preceding sections provide a computationally streamlined approach to the detection of correlated GC component signatures. This streamlining is possible through the relatively simple detection of peak apex values and their conversion to retention indices. Once potential correlations of retention indices indicative of CW target compounds are identified as described in Section 6, the proposed algorithmic approach should focus on the confirmation of these detections. For example, the algorithm should not accept the correlation of a large peak in one chromatogram with a noise peak in the other. Similarly, a "peak" due to a low frequency hump in the baseline should not constitute a detection. The suggested confirmation methods include checks of: the correspondence between the integrated intensities of the correlated peaks; the peak widths and shapes; and the correspondence between any additional features such as the expected presence of associated decomposition products, for example. These considerations are discussed in the subsections that follow.

Correspondence Between Integrated Intensities of the Correlated Peaks. We view integration as one of the key post processing activities of the algorithm. Besides being used to confirm a potential detection, it is also the source of quantitation information for the compound if the detection is confirmed. However, automated integration of chromatograms is a complex operation subject to inherent difficulties. This has been recognized by a number of authors and the development of a robust approach for all possible situations is not an easy task. Many software approaches consist of algorithms which rely on various "patches" to respond to a wide variety of different data and instrument pathologies as they are encountered. Our approach is to focus the more challenging software aspects of the analysis only on the chromatographic regions of potential interest, by first identifying these regions using correlated retention indices.

Conventional integration approaches nearly all utilize what might be considered a "turtle graphics" approach in that the analysis proceeds from beginning (time 0) to the end of

the chromatogram in a point by point fashion. The "turtle" doesn't know what lies ahead and must make decisions as it travels the length of the chromatogram. A conventional detection then consists of a peak start detection from the baseline, a search for the peak apex, and then a search for a return to the baseline. The situation becomes complicated after a peak apex if the turtle encounters a valley at a higher signal level (a new peak start) rather than simply returning to the baseline. Since peaks frequently occur near each other with some degree of overlap, this situation is often encountered. One approach to resolve this is the use of a recursive search for a new peak apex and ending when a new peak start is located before finding the ending point for the current peak. More complex situations can occur when an additional peak start is found before the expected apex of the first peak is reached. Software decisions must also be made regarding whether two peaks should be split or whether one peak should be considered as "riding" on the shoulder of the other.

Our alternative approach is to look at the problem from the standpoint of first gaining information on a selected peak or peaks, and then determining where more detailed analysis must be applied to properly integrate relevant peaks. This detailed analysis for the integration of chromatograms consists of three activities, each representing a subgoal of the overall integration task:

1. Partition the segment of the chromatogram into its component peaks by identifying all peak starts, apexes, and ends
2. Analyze the resulting peak structure to determine the relationships between the peaks, i.e., parent or child ("rider") peaks and the appropriate splitting of merged peaks
3. Integrate the peaks according the peak structure and relationships

These activities are described in the following paragraphs.

The partitioning of a chromatogram trace into its component peaks is accomplished by locating the three parts of the trace which define each peak:

(1) a start-of-peak (SOP),

(2) an apex or top-of-peak (TOP), and

(3) an end-of-peak (EOP).

Thus, a "peak" can be regarded as a data structure containing the above information. The peak data structure also needs to include other attributes such as the relationship between the peak and closely spaced peaks. In integration, the sequence of the peak data structures determines how the area under the chromatogram is allocated to each peak. From an algorithmic point of view, integration is the process that generates this sequence of data structures which are used to allocate the area under the chromatogram trace to the different peaks.

For location of the start-of-peak (SOP), the so-called tangent search method developed by Woerlee and Mol (Woerlee and Mol, 1980) is recommended. This search method for SOPs is illustrated in Figure 7-1. The rationale behind this can be understood as follows. Methods which rely on first derivatives have decreased sensitivity to slowly accelerating peaks on positive sloping baselines. While methods using the second derivative can correct this problem, they are very sensitive to noise fluctuations. The increasing tangent search method is basically a second derivative test which has a decreased sensitivity to noise. A peak start is detected when the "tangents" from a specified point to the next four points successively increase as shown in Figure 7-1. While the use of the term tangent as illustrated in Figure 7-1 may seem confusing, the successive series of chords connecting B_1 with the next points are used to approximate the series of actual tangents to the curve. For example, the chord B_1B_3 is approximately parallel to the tangent to point B_2 and is correspondingly labelled α_2. Similarly, the line through B_1B_4 (labelled α_3) is approximately the tangent to B_3. The series of successively increasing tangents illustrated in Figure 7-1 are indicative of a positive second derivative. The occurrence of a positive result in an increasing tangent search starting from baseline is interpreted as the start of a peak. These conditions can be codified by the following equations:

$$\tan(\alpha_2) > \tan(\alpha_1) \text{ or } ((B_3-B_1)/2) > ((B_2-B_1/1)$$
$$\text{or } (B_1-2\cdot B_2 + B_3) > 0 \qquad (7.1)$$

$$\tan(\alpha_3) > \tan(\alpha_2) \text{ or } ((B_4-B_1)/3) > ((B_3-B_1)/2)$$
$$\text{or } (B_1-3\cdot B_3 + 2\cdot B_4) > 0 \tag{7.2}$$

$$\tan(\alpha_4) > \tan(\alpha_3) \text{ or } ((B_5-B_1)/4) > ((B_4-B_1)/3)$$
$$\text{or } (B_1-4\cdot B_4 + 3\cdot B_5) > 0 \tag{7.3}$$

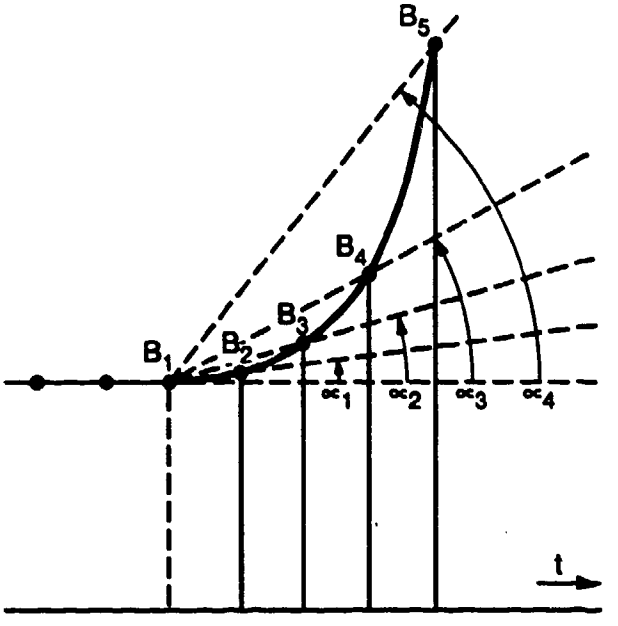


Figure 7-1. Peak start detection with the increasing tangent search.

In a number of situations, the results of the increasing tangent search are significantly improved by also examining the second derivative. This is suggested in the literature as a possible conditional check on the behavior of the increasing tangent search (Woerlee & Mol, 1980). The second derivative is most easily approximated by first approximating the first derivative of the chromatogram by differencing point-by-point, and then taking the second differences. We have found that some slight smoothing is desirable to reduce the noise levels of the second differences. By setting a threshold on the smoothed second differences, significant thrusts of the chromatographic trace away from the baseline can be detected. To determine the threshold levels, the distribution of the second differences for the chromatographic baseline can be considered. An example of the smoothed second differences for a baseline region of a chromatogram obtained with the μGC together with the distribution

of the second differences is shown in Figures 7-2 and 7-3. The distribution is shown in its integrated form in Figure 7-2. Smoothed difference levels which set thresholds for the strongest 2% and 5% of the values are drawn on the chart. These threshold levels are also included in Figure 7-3 showing the occasional thrusts of the smoothed second differences above these levels.

An example of the second differences used to confirm the increased tangent search finding of an SOP for a small peak is shown in Figures 7-4 through 7-6. Figure 7-4 shows the smoothed peak located at approximately 340 points on the time scale. The vertical lines indicate the approximate SOP, TOP, and EOP values for this particular peak. Figure 7-5 shows a plot of the first differences for this same data. The SOP, TOP, and EOP values are close to the zero crossings of this approximation to the first derivative, but these zero crossings are not very useful because of frequent shifts resulting from sloping baselines. In Figure 7-6, the second differences and the smoothed second differences are shown for the same data. Because of the small size of the peak, the noise in the second differences is quite substantial and the effects of smoothing are readily apparent. A significant lift in the smoothed second differences is clearly observed within the window of the increased tangent search (5 points) located at the true SOP. Thus, this "breakout" of the smoothed second differences within the window of the tangent search correctly confirms the result of the increased tangent search for this weak chromatographic peak.

The use of second differences to suppress unwanted detections due to baseline noise is shown in Figures 7-7 and 7-8. Figure 7-7 shows the faulty triggering of peak detections using the increasing tangent search method with a very noisy baseline as determined by the LSU analysis software. While the LSU software does very well in general, this is an example situation which created problems for their algorithm. Figure 7-8 shows the effect of coupling the increasing tangent search with the breakout of the second differences as discussed in the examples above. Substantial suppression of unwanted detections is accomplished, while the peaks in the earlier portion of the chromatogram are properly detected.

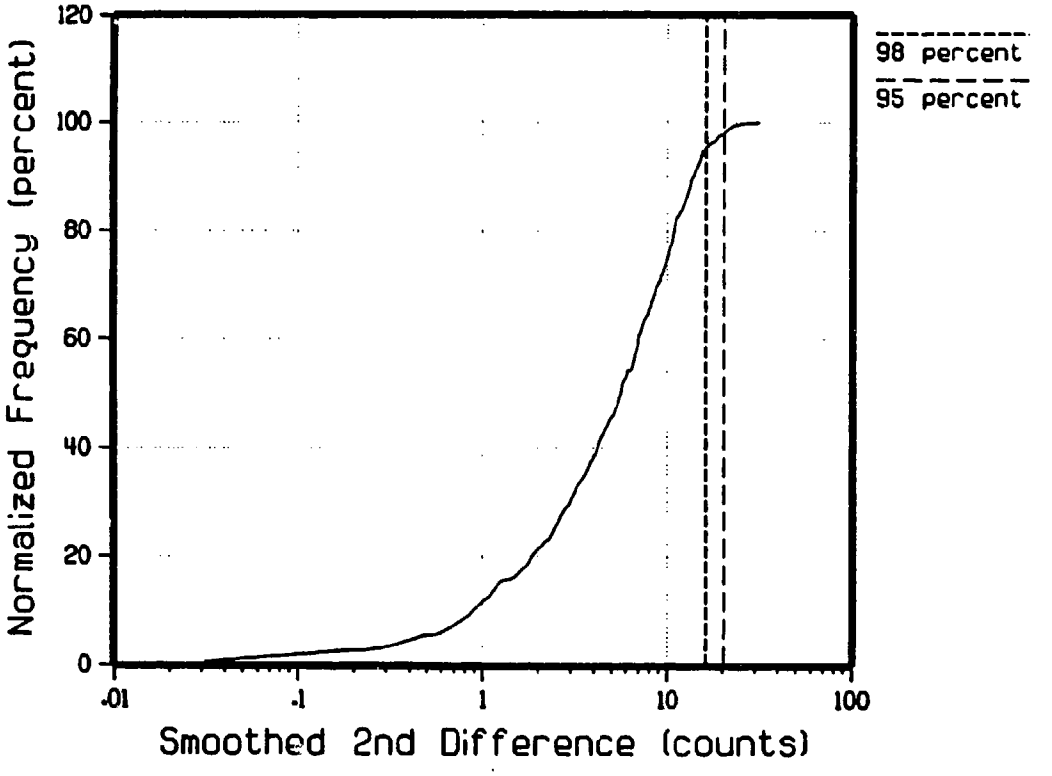


Figure 7-2. **Integrated distribution of absolute values of the smoothed second differences showing threshold levels for values in the top 2% and 5%.**

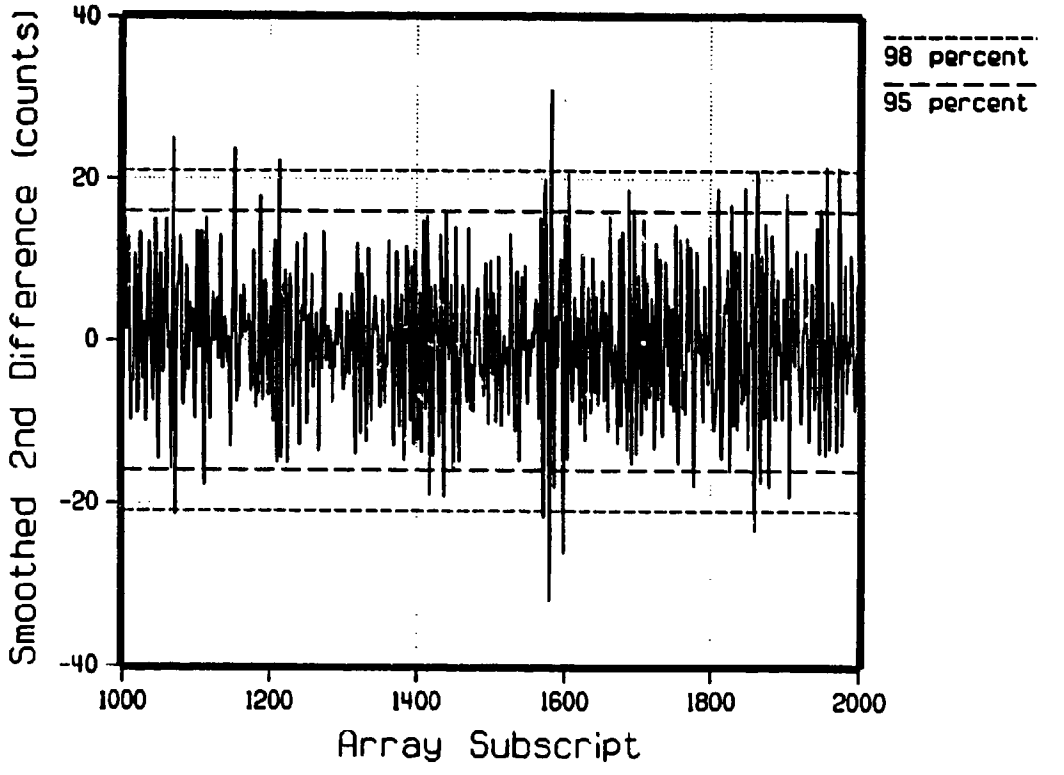


Figure 7-3. **Smoothed second difference data from a baseline region of a μGC chromatogram. Threshold values are shown for the top 2% and 5% levels.**

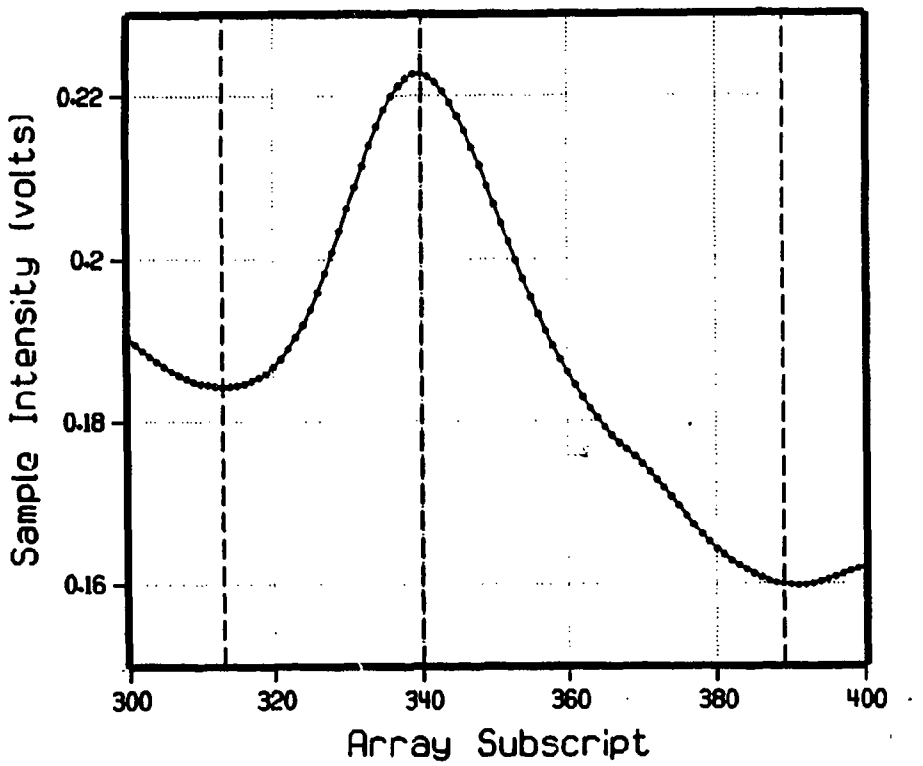


Figure 7-4. Example small μGC peak (smoothed) Vertical lines indicate approximate SOP, TOP and EOP values.

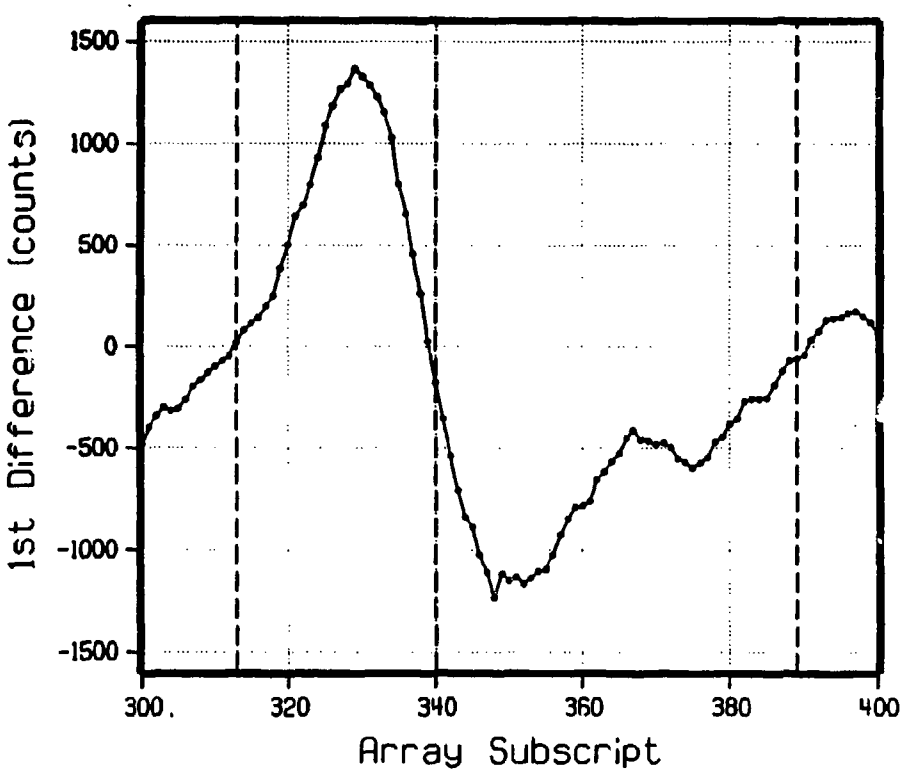


Figure 7-5. First differences calculated for peak in Figure 7-4 showing approximate zero-crossings at the SOP, TOP and EOP values.

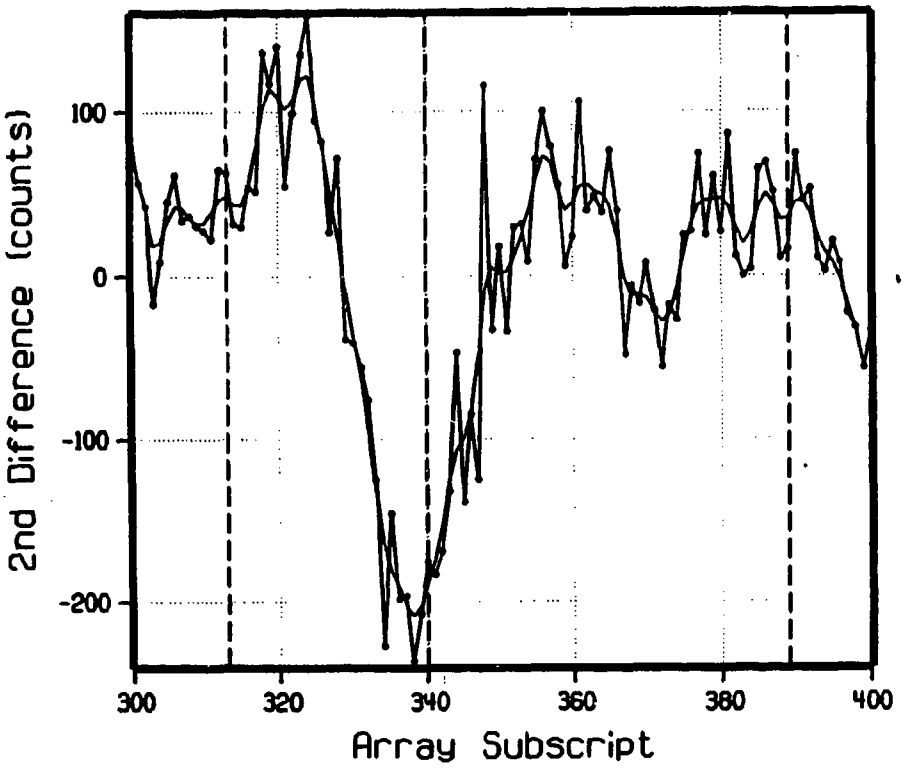


Figure 7-6. Second difference (connected points) and smoothed second difference (line only) for the peak in Figure 7-4 showing large positive value within five points of the SOP (the width of the increasing tangent search).

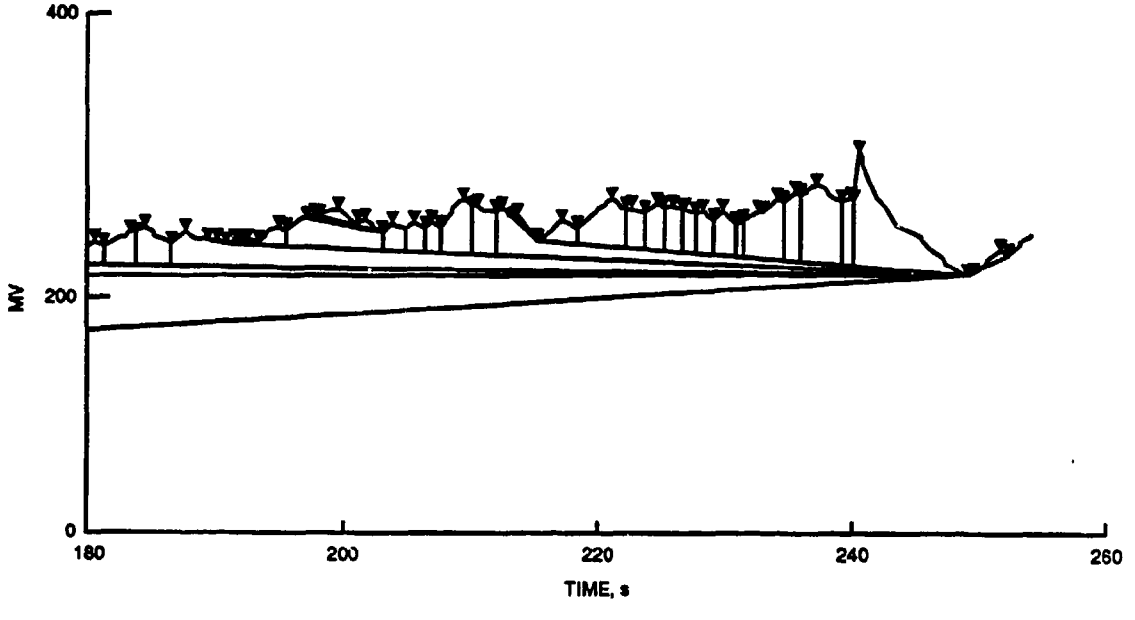


Figure 7-7. Example showing faulty peak detection and splitting of a noisy μGC baseline using other methods.

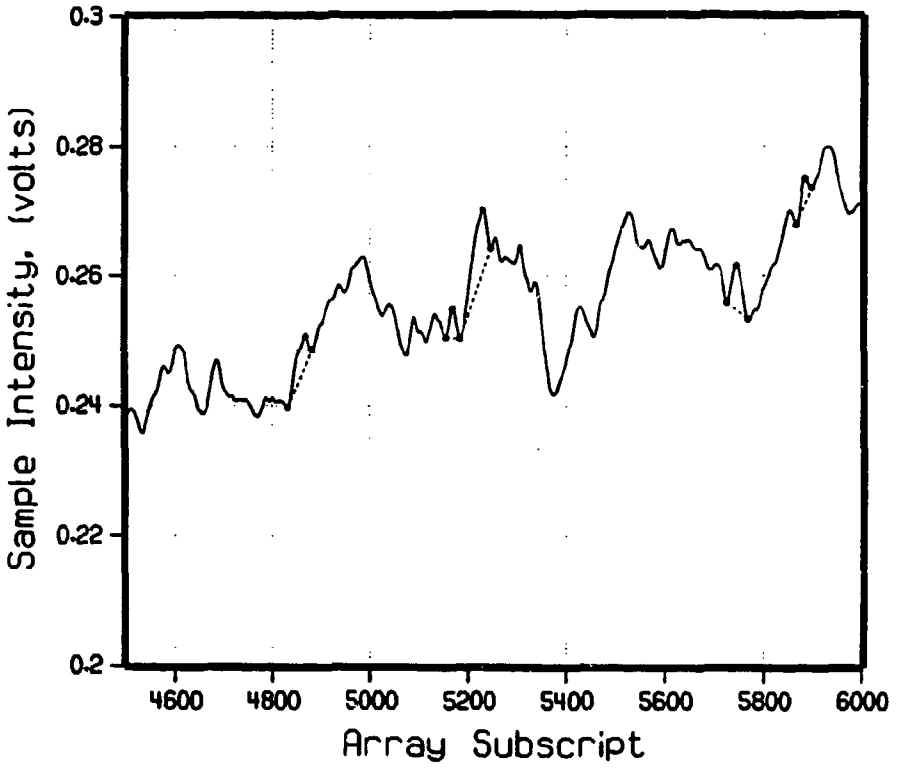


Figure 7-8. **Same data as Figure 7-7 showing an increasing tangent search combined with thresholded second differences to reduce unwanted noise detections in the baseline.**

An additional example of suppression of unwanted peak detections is provided by the chromatogram in Figures 7-9 and 9-10. This chromatogram shows the separation using μGC of a mixture of phosphites including the trimethyl, triethyl, and diethyl phosphites in the CW schedules. Again, the unwanted detections in Figure 7-9 are suppressed in Figure 7-10 by the inclusion of the requirement for a breakout of the second differences in conjunction with the increasing tangent search. Only the peaks with visible areas were detected by inclusion of the second differences requirement.

The peak ending, or EOP, is either determined by the presence of a valley, a return to the original baseline, or the location of a tangent which approximates the end of a peak "riding" on the shoulder of another peak. Peaks are frequently observed to overlap, or "merge." In the general situation, a number of peaks can overlap to result in "nesting" of the SOP, TOP, and EOP values of the different peaks. In the simple situation of a single,

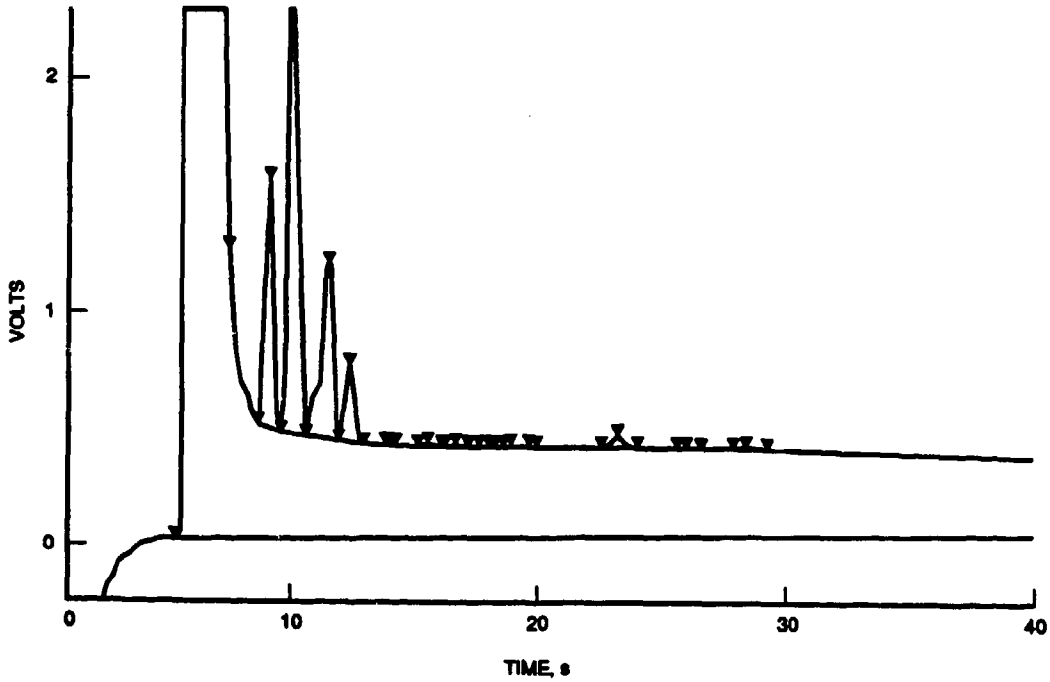


Figure 7-9. Example μGC chromatogram showing excessive detections using other methods.

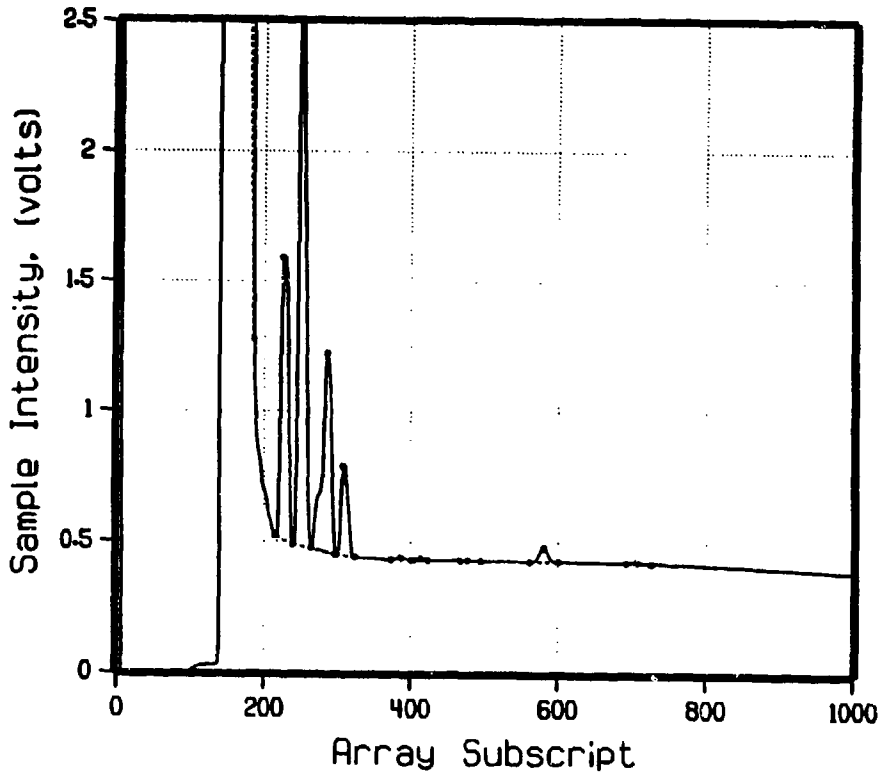
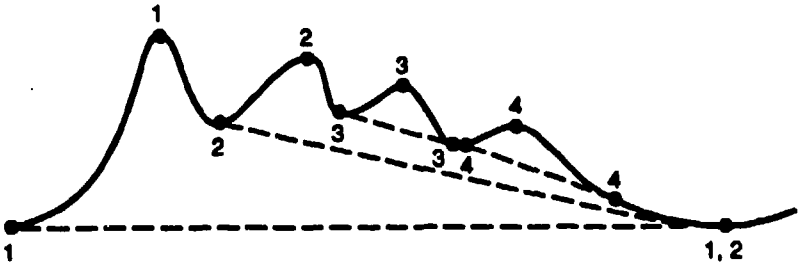


Figure 7-10. Same data as Figure 7-9 showing reduced baseline detections by combining thresholded second differences with the increasing tangent search method.

isolated peak, the peak can be represented by its (SOP,TOP,EOP) set of values. Figure 7-11 illustrates an example merging four peaks with the corresponding nesting of the SOP,TOP and EOP values. The nesting is represented by parentheses. Peaks 2,3 and 4 are nested between the TOP and EOP of peak 1, the "primary" peak. Peak 1 would be considered "un-nested." In the terminology of nested algorithms, peaks that are nested are considered "children" peaks of one unique parent peak. For example, peak 2 is a child of peak 1 in Figure 7-11. Children can be parents, and parents can also have multiple children. Peaks 3 and 4 are examples of children of peak 2.



$(SOP_1\ TOP_1\ (SOP_2\ TOP_2\ (SOP_3\ TOP_3\ EOP_3)\ (SOP_4\ TOP_4\ EOP_4)\ EOP_2)\ EOP_1)$

Figure 7-11. Illustration of four merged peaks with the nesting of the SOP, TOP and EOP values for the four peaks.

Decisions to change the splitting of the peaks in the second step of the algorithm can alter these relationships. This is accomplished in an algorithmic pass which checks the consistency of the peak structures and nesting. For example, the chromatogram should not cross lines connecting the SOP and EOP values of each peak. Further, the parent-child assignments of the peaks must also be consistent with the peak structure data.

Two possible splittings of a pair of merged peaks is illustrated in Figures 7-12 through 7-14. As shown in Figure 7-12, the valley point, V, is determined by a new peak start, and the point E_1 is determined by a return to the baseline. Both tangential ("tangent skimmed") and perpendicular separations ("valley drops") are possible. Since children peaks are typically small "rider" peaks on a parent peak, it is reasonable to initially assume that all children peaks will be tangent skimmed.

For tangent skimming, a line directly connecting V and E_1 would partly lie above the chromatogram trace. For this reason, a closer-lying point E_2 is sought which provides a tangential connection to V. This is illustrated in Figure 7-13. The areas of the peaks are then calculated by computing the area of the tangent skimmed peak above the line VE_2 and subtracting this area from the total area under the two peaks.

The separation of the two peaks by the valley drop method is shown in Figure 7-14. In this case, the perpendicular to the baseline from point V determines point E_3. The decision on whether to valley drop instead of tangent skim can be based upon the relative size of the child peak. For relatively even peak sizes, the integration errors would be smaller by splitting the peaks by valley dropping.

An example using microchip GC data showing a separation by valley dropping is shown in Figure 7-15. The valley between the two peaks to the right in this chromatogram is sufficiently small relative to the two peaks that this separation provides the least error in the integration of the two peaks. Other peaks in the chromatogram are tangent skimmed because they are small relative to their parent peaks.

Once the peak data structures are established, the integration is a straightforward operation. The area of each peak is initially obtained by ignoring the parent-child relationship. The final areas are obtained by subtracting the area of each peak from its parent.

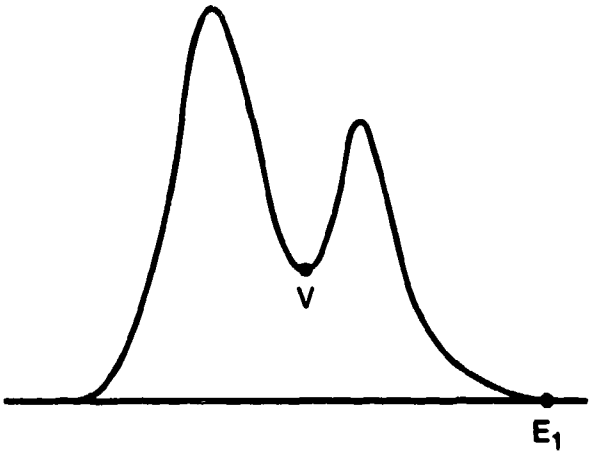


Figure 7-12. Pair of peaks showing valley (V) and return to basline point E_1.

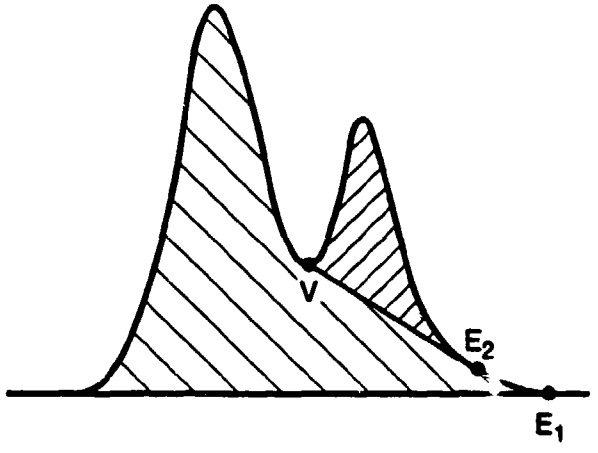


Figure 7-13. "Tangent skimmed" separation of a merged pair of peaks by locating tangential point E_2.

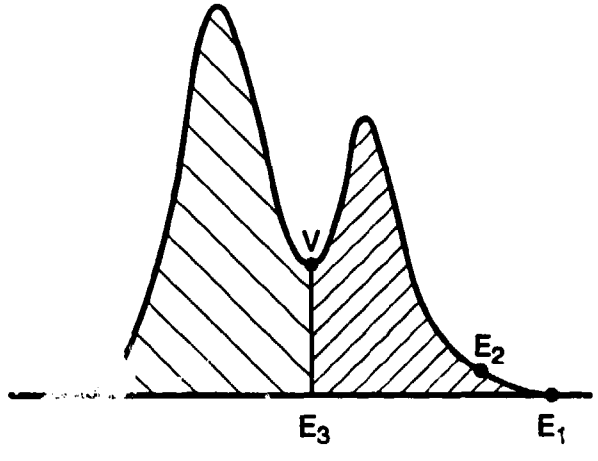


Figure 7-14. "Valley dropped" separation of a merged pair of peaks by perpendicular from point V to baseline (point E_3).

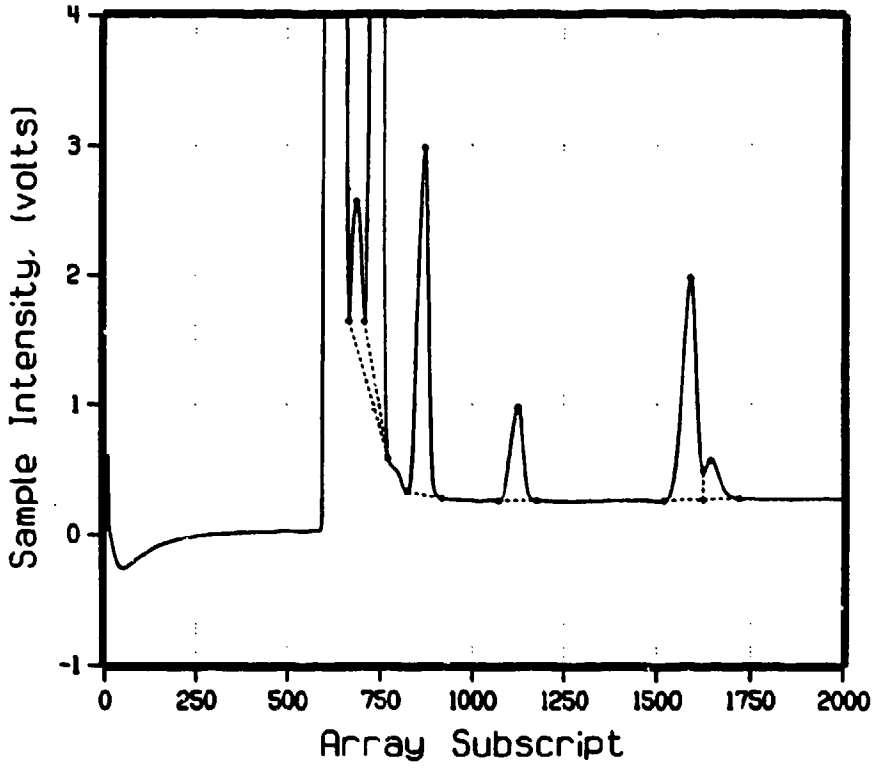


Figure 7-15. Example μGC data showing both tangent-separated and valley drop-separated peaks.

Some final examples of this algorithmic approach were applied to μGC data which was pathological to the LSU software due to irregular baselines. These examples begin with Figures 7-16 and 7-17 which compare the performance of the LSU software with the proposed approach for a mixture of phosphites including CW schedule 3 phosphites. The earliest peaks are correctly tangent skimmed on the large "air spike" which elutes first. The software adjusts to a dip in the baseline rather than tangent skimming beneath a large extent of the baseline in the middle of the chromatogram. To allow for the shifting baseline, the tangent condition on the EOP of the final peak was not required, but this can easily be modified.

Figures 7-18 and 7-19 show similar data for weak peaks on a more irregular baseline. Again the proposed approach appears to properly detect the important features of the

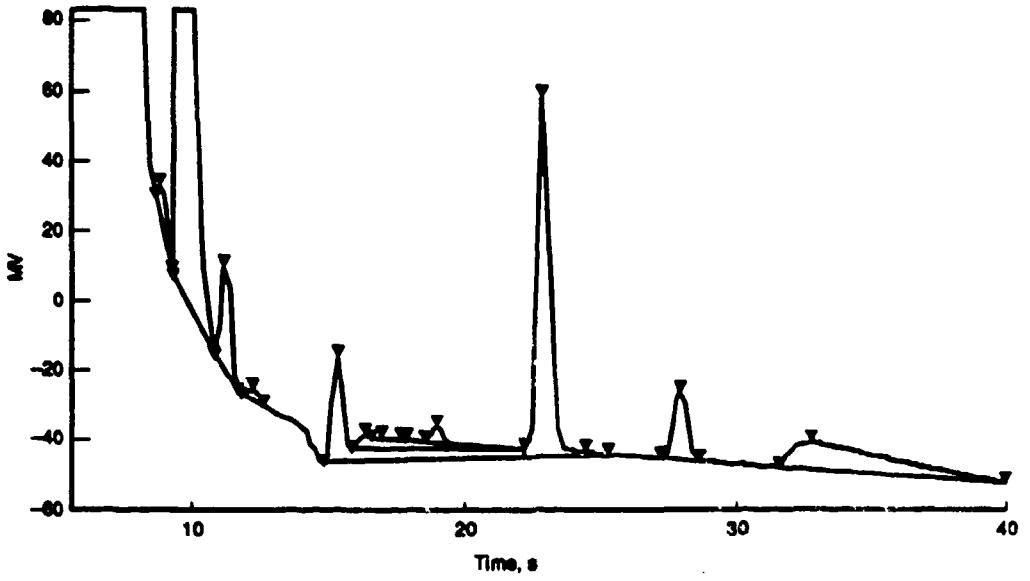


Figure 7-16. Example μGC data for schedule 3 phosphites showing an analysis method which integrates below the apparent baseline.

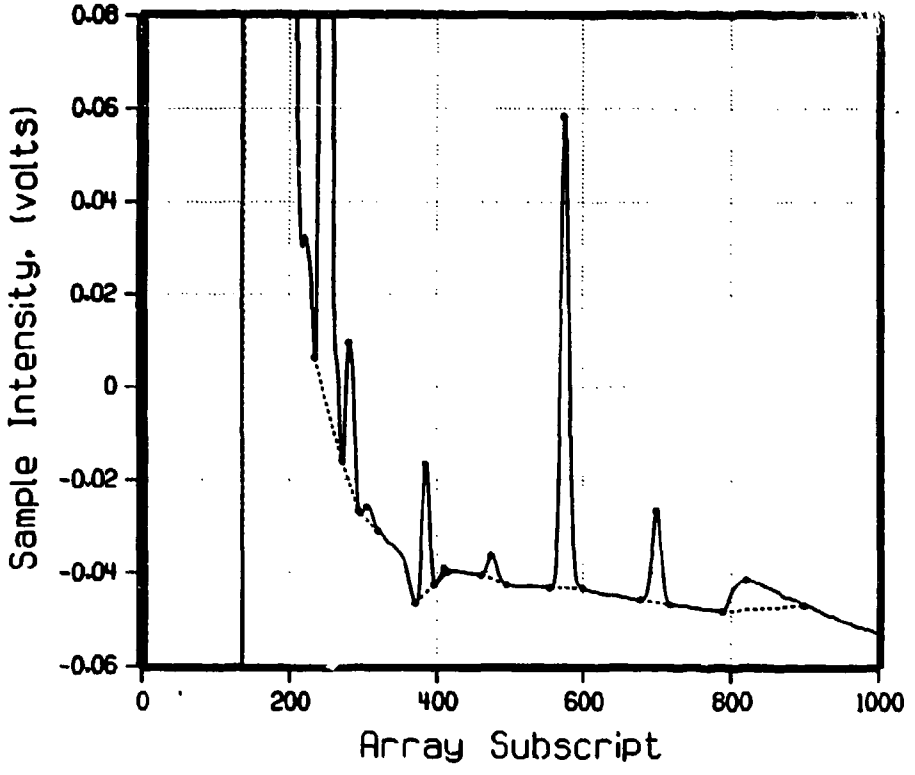


Figure 7-17. Same data as Figure 7-16 showing the modified increasing tangent search method which more closely follows the baseline.

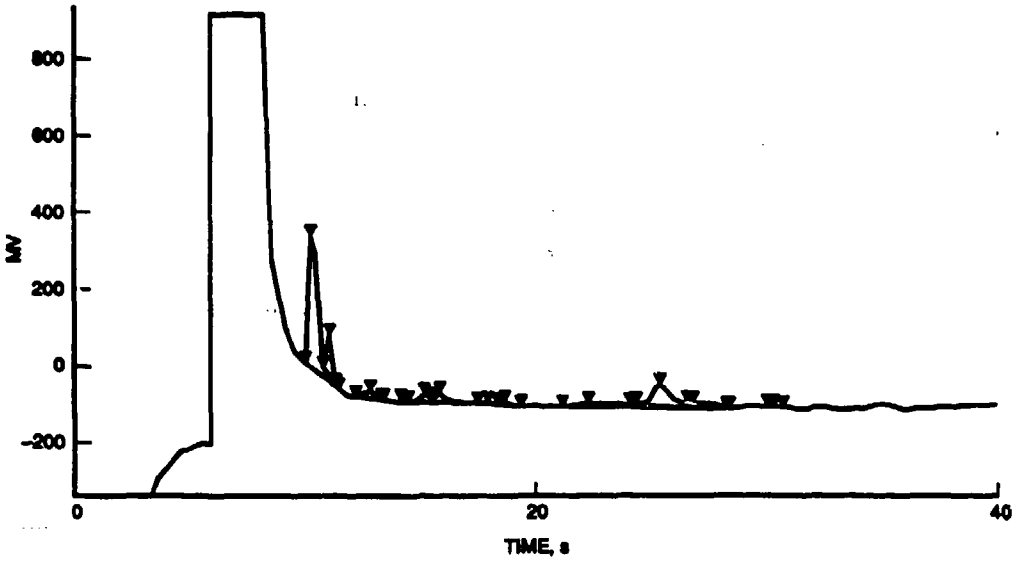


Figure 7-18. Example μGC data for schedule 3 phosphites showing an analysis method resulting in unwanted baseline detections.

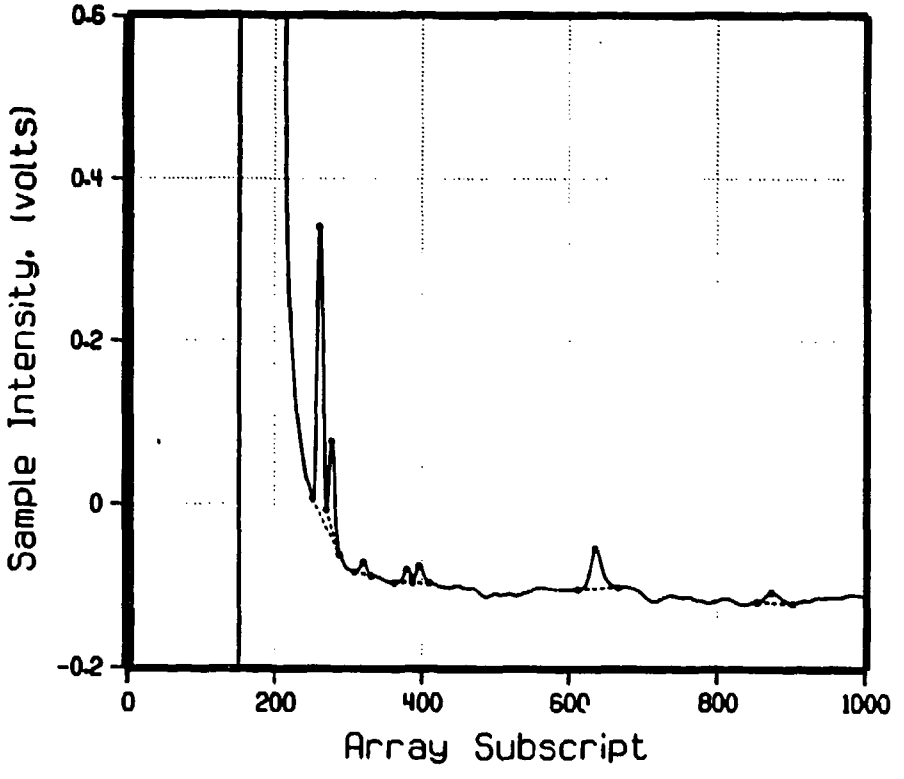


Figure 7-19. Same data as Figure 7-18 which shows the modified increasing tangent search method which reduces unwanted detections due to baseline irregularities.

chromatogram including a possible late peak which was not detected by the LSU software. Again, the integrations resulting from the proposed approach do not invoke large extents of tangent skimming beneath the baseline. Further, the detected peaks appear reasonable.

As a final example, Figures 7-20 and 7-21 show the successful detection of a peak on an unusual baseline having a large rise. This is a difficult situation with the tangent search method because of the steep slope. The peak near 800 on the horizontal axis is properly detected by the proposed approach under these baseline conditions. Further, noise in the baseline does not trigger detections in the two large rises in the baseline. A small feature is overlooked in the middle of the baseline which appears to be the result of local noise structure on this small feature.

Peak Widths and Shapes. As suggested in Section 6, expected values for the peak widths and shapes can be included in the library of CW target information. Widths and shapes can be concentration dependent, so it may be necessary to store several values so that an appropriate reference width can be selected according to the approximate integrated area of the peak. By selecting the appropriate reference width, a spurious chromatogram "peak" arising from a wide "hump" in the baseline of the chromatogram can be eliminated from consideration. As illustrated in Figure 7-20 and 7-21, the use of the smoothed second differences in conjunction with the increasing tangent search in the proposed approach also helps alleviate such occurrences.

Similarly, peak shape information can be stored for comparison, if a particular compound is known to have a characteristic leading or trailing edge. Leading or trailing edges typically arise from concentration-dependent distribution coefficients between the mobile and stationary chromatographic phases (Oi et al., 1983). Since these effects are concentration dependent, a series of reference widths spanning a range of concentrations (integrated areas) could be stored in the CW target reference tables. Then, using the integrated area for a suspect CW target peak, the reference width, if concentration dependent, could be determined by interpolation from the table values.

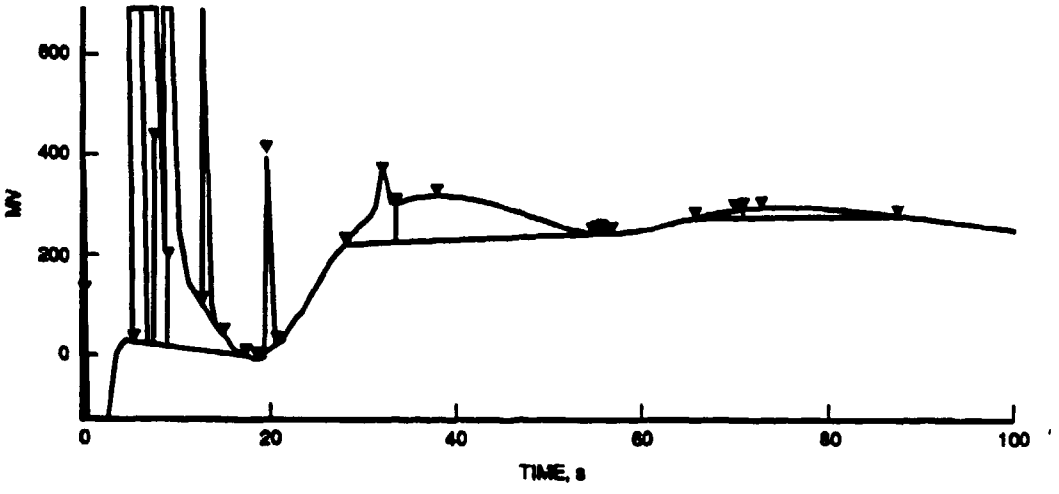


Figure 7-20. Unusual μGC baseline showing an analysis method which fails on a rising baseline slope.

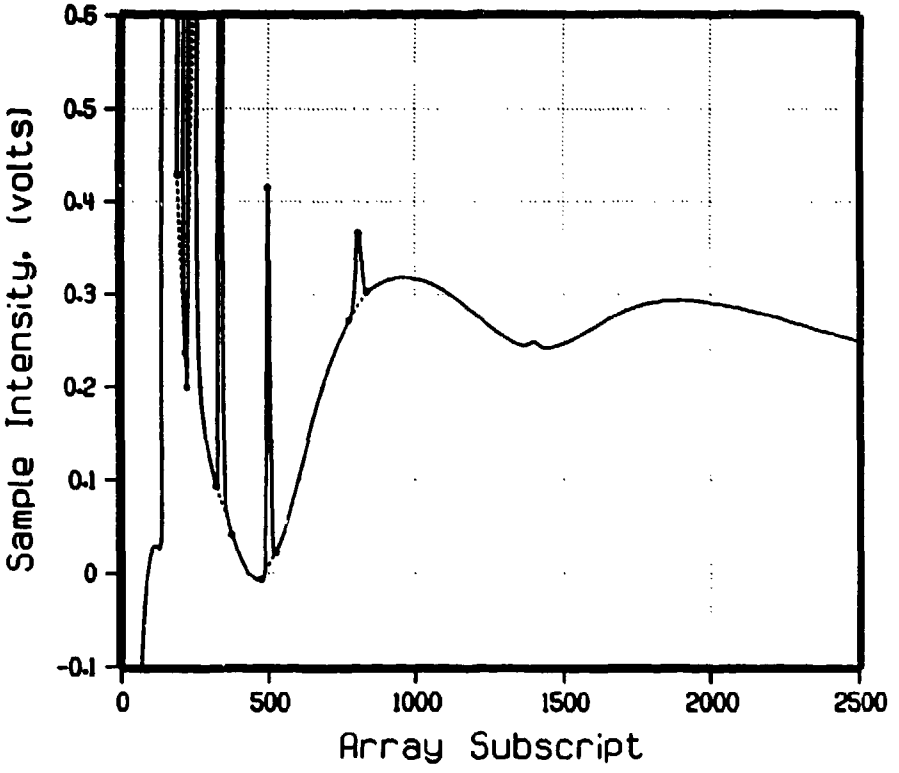


Figure 7-21. Same data as Figure 7-21 with which the modified increasing tangent search method shows the correct integration of a peak on a rising baseline slope.

Correspondence Between Additional Features such as Associated Peaks Due to Decomposition Products or Expected Contaminants. It is possible that certain CW target materials may have signatures consisting of multiple peaks. Although we have not yet encountered this situation in our testing of representative CW schedule materials, our detection approach should make use of such occurrences as the database becomes fully developed for the CW target materials.

Two likely sources of multiple peaks in CW target signatures are decomposition products and contaminants. Decomposition products included oxidation products, hydrolysis products (possibly from environmental exposure to water), and thermal degradation products. Thermal desorption of the adsorbent vapor traps used for preconcentration may be result in the significant thermal degradation of certain compounds, especially if they are chemically reactive to the trap materials (Witkiewicz, 1990). In cases involving thermal decomposition by the thermal desorption of the trap, the decomposition products become potential signature components for the detection of the CW target material. Contaminants, if these were found to be important in a CW signature, would probably be by-products in the synthesis of the CW target material.

The joint appearance of multiple signature components can be used to confirm the detection of a CW target material. If, for example, peak 2 always appears along with peak 1, and peak 3 sometimes occurs with peak 1, then the detection of peak 1 would be confirmed by the detection of peak 2 or invalidated by the failure to detect peak 2. The detection of peak 3 would strengthen the detection of peak 1, but failure to detect peak 3 would not invalidate the detection of peak 1.

While some direct logic can be applied to detection algorithms to handle simple cases such as the case in the previous paragraph, more complex signatures with more components can be easily accommodated by signal processing techniques such as neural nets. Neural nets can be trained to associate certain data conditions with a desired outcome and then be used to recognize these data conditions. Thus, a neural net can potentially be trained to recognize a complex CW target signature and then be successfully applied to recognize this signature.

GRC has neural net application and training tools which can be applied to CW detection signatures for the handheld μGC instrument, but the signature processing analysis has yet to dictate a need for such complex processing. While the chemical background encountered by a field instrument can be very complex and highly variable, the CW target signatures studied so far are not complex. Indeed, these appear in most cases to consist of a single chromatographic peak. Since the signal processing is driven by recognition of the CW target components rather than background, the need for recognition approaches for complex signatures is yet to be demonstrated.

There are a number of situations in which neural nets which can recognize complex signatures may apply to program needs e.g., complex CW target signatures may yet arise; refinement of the signal processing algorithms may benefit from the application of neural nets to certain algorithmic tasks such as the peak separation decisions; and neural nets might be used to control instrument settings for different field applications. The last application borders on the use of expert systems, in which software tools other than neural nets might also apply. For example, GRC has commercial expert system tools which could be used for automatically controlling instrument operation. This approach could respond to different detection scenarios that an inspector might select, or be used to automatically adjust instrument response to enhance CW target detection. Opportunities for such applications to facilitate inspector-friendly operation of the handheld instrument will become apparent when the program reaches the stage of fielding a prototype instrument. These tools are available and can be readily applied to the advanced signal processing needs of the instrument.

SECTION 8
CONCLUSIONS AND RECOMMENDATIONS

The results of these signal processing studies using actual μGC data delineate the basic signal processing requirements expected for the handheld CW detector. A range of μGC data was tested with a specific focus on developing automated approaches which can robustly respond to the full variety of possible μGC data variations. The key conclusions and recommendations based upon the results in this report are as follows:

- **An algorithm which is table-driven for the detection of the CW target compounds is recommended.** By using a one-dimensional approach independently with each column for the detection of CW target peaks, the correlation between the results for the two columns becomes a simple table-oriented computation. If the correlation of the tabulated findings results in a potential detection of a CW target compound, then more detailed signal processing analysis can quantitate the separate findings and validate the detection.

- **The use of a retention index representation of the CW targets is recommended.** By converting reference data and measurements to retention indices, the instrument can be readily calibrated and measurements can be compared to reference data regardless of normal instrument variations. Instrument variation which can be readily compensated for by calibration includes: day to day variation of the instrument performance to include variations in pressure, temperature, and flow rates; longer-term variation in column materials; longer-term variation in the sample injection and splitting of samples between multiple chromatograph modules; and substitutions or repair of μGC instrumentation components.

- **A specific method is recommended for determining the exact retention index matching criteria for each CW target compound.** An error

propagation analysis was presented which bounds the expected variation in measured retention indices for a given compound. This is based upon a number of experimental μGC parameters and corresponding experimental measurements. This provides an estimate of the expected variation in retention index measurement for specified conditions and can be used to set the target matching criteria.

- **Integration methods using increasing tangent searches combined with smoothed second differences are recommended.** Our testing of integration methods for quantitation of chromatographic peaks indicates that use of the second derivative (smoothed second differences) adds substantial improvement to the performance of the increasing tangent search methods for locating peak starts. This was especially clear using data sets selected for their pathological results with other approaches. This should be easy to implement based upon the results we have obtained using a wide assortment of actual μGC data.

Based upon these results and recommendations, we expect to carry forward our proposed algorithmic approach into the functional elements evaluation and the concept instrument design. We find the above results using simplified algorithms very encouraging since they represent the use of compact and fast algorithms which can potentially provide improved performance over existing automated GC quantitation methods.

SECTION 9

REFERENCES

Budahegyi, M. V., E. R. Lombosi, T. S. Lombosi, S. Y. Meszaros, S. Nyiredy, G. Tarjan, I. Timar, and J. M. Takacs, "Twenty-fifth Anniversary of the Retention Index System in Gas-Liquid Chromatograph," (U) J. Chromatography, Vol. 271, 1983, pp. 213-307. (UNCLASSIFIED)

Fozard, A., J. J. Franses, and A. Wyatt, "An Analysis of the Errors in Computerized Gas Chromatograph," (U) Chromatographia, Vol. 5, 1972, pp. 377-81. (UNCLASSIFIED)

Lee, G., C. Ray, R. Siemers, and R. Moore, "Recent Developments in High Speed Gas Chromatograph," (U) American Laboratory, February, 1989, pp. 14-19. (UNCLASSIFIED)

Conference on Disarmament, "Draft Convention on the Prohibition of the Development, Production, Stockpiling and Use of Chemical Weapons and Their Destruction," (U) 3 September 1993. (UNCLASSIFIED)

Li, B. Q., S. Siu, and J. W. Evans, "Microcomputer Processing of Chromatographic Data," (U) J. Chromatographic Sci., Vol. 25, 1987, pp. 281-285. (UNCLASSIFIED)

Microsensor Technology, Inc., personal communication with Steve Santy, 1992. (UNCLASSIFIED)

Oi, T., H. Kakihana, M. Okamoto, and M. Maeda, "Numerical Analysis of Chromatographic Band Development of One-Component Systems," (U) J. Chromatography, Vol. 270, 1983, pp. 7-16. (UNCLASSIFIED)

Overton, E.B., personal communication, 1992. (UNCLASSIFIED)

Takacs, J., and D. Kralik, J. Chromatography, Vol. 50, 1970, p. 379. (UNCLASSIFIED)

Witkiewicz, Z., M. Mazurek, and J. Szule, "Chromatographic Analysis of Chemical Warfare Agents," (U) J. Chromatographic Sci, Vol. 145, 1990, pp. 293-357. (UNCLASSIFIED)

Woerlee, E. F. G., and J. C. Mol, "A Real time Gas Chromatographic Data System for Laboratory Applications," (U) J. Chromatographic Sci., Vol. 18, 1980, pp. 258-266. (UNCLASSIFIED)

DISTRIBUTION LIST

DNA-TR-92-196

DEPARTMENT OF DEFENSE

DEFENSE NUCLEAR AGENCY
4 CY ATTN: OPAC J FOX
2 CY ATTN: OPAC LTCOL B PATE
2 CY ATTN: TITL

DEFENSE TECHNICAL INFORMATION CENTER
2 CY ATTN: DTIC/OC

OASD ATOMIC ENERGY
2 CY ATTN: LTC PATE
10 CY ATTN: S LEIBBRANDT

OFFICE OF THE JOINT CHIEF OF STAFF
ATTN: LTC DUPLANTIER

OFFICE OF THE UNDER SECRETARY OF DEFENSE
ATTN: S BUCKLEY

ON-SITE INSPECTION AGENCY
ATTN: COL GILBERT

OSD (DDR&E)
ATTN: BG ARTHUR JOHNSON

DEPARTMENT OF THE ARMY

DEPARTMENT OF THE ARMY
ATTN: DAMO/SWC MS KOTRAS

DEPARTMENT OF THE ARMY
ATTN: DAMO/SWC DR BOYLE

U S ARMY CHEMICAL SCHOOL
ATTN: ATZN-CM-FTO MR SHEHEANE

U S ARMY MATERIAL COMMAND
ATTN: AMCAN-CN COL NIKAI

USA CML & BIOLOGICAL DEFENSE AGENCY
ATTN: AMSCB-TC MR COLBURN
2 CY ATTN: MR FLAMM
5 CY ATTN: SMCCR-MUM DR HUTCHINSON

DEPARTMENT OF THE NAVY

ASSISTANT SECRETARY OF THE NAVY
ATTN: MR GRATTON

OFFICE OF CHIEF OF NAVAL OPERATIONS
ATTN: OP-65 CAPT JOHNSON

DEPARTMENT OF THE AIR FORCE

USAF/XOXXI
ATTN: COL ENGSTROM

DEPARTMENT OF ENERGY

DEPARTMENT OF ENERGY
ATTN: DP 5.1 MS CASEY
ATTN: DP-5.1 G DUDDER

OTHER GOVERNMENT

ARMS CONTROL & DISARMAMENT AGENCY
ATTN: DR SEIDERS
ATTN: MR MIKULAK
ATTN: MR STAPLES

CENTRAL INTELLIGENCE AGENCY
ATTN: DCI/ACIS/TMC MR SPALDING

DEPARTMENT OF DEFENSE CONTRACTORS

GENERAL RESEARCH CORP
ATTN: DR JOHN REDICK
2 CY ATTN: P HOLLAND
2 CY ATTN: R MUSTACICH
2 CY ATTN: W FOREMAN

JAYCOR
ATTN: CYRUS P KNOWLES

KAMAN SCIENCES CORP
ATTN: DASIAC

KAMAN SCIENCES CORPORATION
ATTN: DASIAC

Defense Threat Reduction Agency
8725 John J Kingman Road MS 6201
Ft Belvoir, VA 22060-6201

TDANP-TRC August 1, 2001

MEMORANDUM TO DEFENSE TECHNICAL INFORMATION CENTER
ATTN: OCQ/MR LARRY DOWNING

SUBJECT: DOCUMENT CHANGES

The Defense Threat Reduction Agency Security Office reviewed the following documents in accordance with the Deputy Secretary of Defense Memorandum entitled, "Department of Defense Initiatives on Persian Gulf War Veterans' Illnesses" dated 22 March 1995, and determined that the documents were unclassified and cleared for public release:

DNA-TR-93-84, AD-B244408, Acoustic Resonance Spectroscopy in CW Verification Tooele Field Trial (August 1992).
DNA-TR-93-129-V1, AD-B192045, Global Proliferation – Dynamics, Acquisition Strategies and Responses, Volume 1 – Overview.
DNA-TR-93-129-V2, AD-B192046, Global Proliferation – Dynamics, Acquisition Strategies and Responses, Volume 2 – Nuclear Proliferation.
DNA-TR-91-216, AD-B163637, Harmonizing the Chemical Weapons Convention with the United States Constitution.
DNA-TR-92-180, AD-B175230, Evaluation of the Concept of a List for the BWC.
DNA-TR-92-61, AD-B167663, Basic State Party Functions and Skills Under CWC.
DNA-TR-92-66, AD-B167357, Domestic Reporting Requirements for Chemical Industry.
DNA-TR-91-213, AD-B163260, Analysis of the Interactions Between Treaties.
DNA-TR-93-70, AD-B177262, Chemical Weapons Convention Inspections of Private Facilities Application of United States Environmental and Safety Laws.
DNA-TR-92-182, AD-B173450, Commercial Products from Demilitarization Operations.
DNA-TR-91-217-V3, AD-B169350, Chemical Weapons Process Parameters, Volume 3 – Users' Guide.
DNA-TR-92-116-SUP, AD-B175292, Technical Ramifications of Inclusion of Toxins in the Chemical Weapons Convention (CWC), Supplement.
DNA-TR-92-128, AD-B175452, Task 1 Report Target Vapor Identification and Database Development.
DNA-TR-92-196, AD-B174940, Task 2 Report Algorithm Development and Performance Analysis.
DNA-TR-93-68, AD-B178109, CW Detection Instrument R&D Design Evaluation.

Enclosed is a copy of the referenced memorandum. If you have any questions, please call me at 703-325-1034.

Ardith Jarrett

ARDITH JARRETT
Chief, Technical Resource Center