

Lawrence Berkeley National Laboratory

Recent Work

Title

REAL-TIME ANALYSIS OF GAS CHROMATOGRAMS WITH A MICROCOMPUTER SYSTEM

Permalink

<https://escholarship.org/uc/item/7c81f14s>

Author

Bobba, G.M.

Publication Date

1976-07-01

Submitted to Journal of Chromatographic
Science

LBL-5411
Preprint C. |

REAL-TIME ANALYSIS OF GAS CHROMATOGRAMS WITH A
MICROCOMPUTER SYSTEM

G. M. Bobba and L. F. Donaghey

July 1976

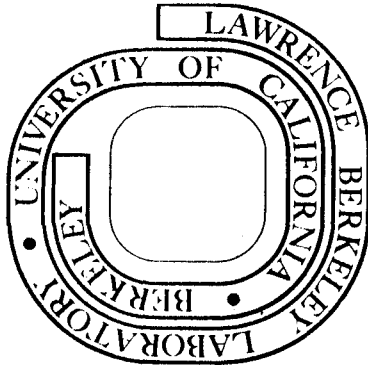
RECEIVED
LAWRENCE
BERKELEY LABORATORY

OCT 19 1976

LIBRARY AND
DOCUMENTS SECTION

Prepared for the U. S. Energy Research and
Development Administration under Contract W-7405-ENG-48

For Reference
Not to be taken from this room



LBL-5411
c1

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.

Real-Time Analysis of Gas Chromatograms
with a Microcomputer System

G. M. Bobba and L. F. Donaghey

Materials and Molecular Research Division, Lawrence Berkeley Laboratory
and Department of Chemical Engineering, University of California,
Berkeley, California 94720

July, 1976

Abstract

The real-time performance of a microcomputer system designed for dedicated control of gas chromatography and real-time analysis was studied to determine operating characteristics and accuracy of the peak-area integration algorithm. The software program developed to obtain peak areas used statistical criteria for the detection of the start, maximum value, and end of a chromatographic peak. The accuracy of the system was tested with propane-helium mixtures and by comparison to results obtained with off-line data reduction. The study showed that the real-time microcomputer system gave peak areas which were lower than the off-line results by less than 0.6 percent. The difference was shown to arise from truncation error in the double-precision octal mathematical subroutines, and to error in the assignment of a baseline at the peak-end.

Introduction

During the past twenty years the gas chromatograph has evolved extensively as an analytical device for the chemical industry. Since the development of the computer, many types of computer-assisted chromatographs have been developed to analyze and reduce gas chromatographic data, and to control the operation of the instrument.

Quantitative gas chromatography depends ultimately upon the accuracy of chromatographic peak-area measurements, from which the concentrations of components present in the sample can be determined. The problem is compounded by signal noise and baseline drift. Inaccuracies in the determination of the start and end of each chromatographic peak are significantly affected by these problems.

Computer systems for chromatography which are on-line offer advantages of rapid chromatogram reporting and low operator assistance. Electronic, analog, and hybrid integrators have been developed for automatic peak-area integration [1]. While these methods are low in cost, they are often limited in performance and are inherently inflexible [2]. Computing integrators rely on noise filtering of the detector signal to decrease the error in assigning the peak baseline. In that approach the filter characteristics must be fixed, whereas an optimum noise filtering involves a trade-off between the peak resolution and the signal slope resolution, so that baselines are frequently assigned improperly.

The minicomputer has also been utilized for the on-line analysis of chromatograms, with the variety of conceptual approaches and

systems in current use [3-6]. The minicomputer systems are characterized by high performance and multichannel capability [7], but often at very high cost. These systems utilize background programming methods to analyze the chromatogram which is stored in the computer memory, and thus do not operate in real-time. Although the cost of minicomputer systems and of semiconductor memories has decreased at nearly an annual rate of 30 percent since 1967, these systems remain beyond the reach of most small laboratory users of gas chromatographs.

A significant recent advancement in the automatic monitoring and control of gas chromatographs has been made feasible by the development of the microprocessor, a large-scale integrated circuit on a single silicon chip which is the central processing unit for a microcomputer. Microprocessor technology has made it possible to develop gas chromatograph systems with the performance of existing minicomputer systems, but at the cost of the limited, electronic integrators. A microprocessor-based system for the control of gas chromatographs and for the analysis of chromatograms has been developed by the authors [8]. Microcomputers have the advantages of high reliability and a high noise immunity, as well as small size and power requirements in continuous, dedicated service. This approach to on-line chromatographic analysis promises to efficiently solve the problem of gas-chromatograph automation for the small laboratory user.

The development of software programs for the implementation of microcomputers in real-time analytical applications of gas chromatography is the objective of the present study. A particular objective is the

development of a real-time analytical program for peak-area integration which utilizes statistical methods for reducing uncertainties arising from noise and baseline variation, similar to the methods utilized in off-line data reduction [9-10]. In this paper, an approach to the real-time integration of peak areas is presented and tested for sources of error.

The Microcomputer System and Software

The microcomputer system studied utilizes an Intel 8008 microprocessor in a bin-oriented structure with 7.6 by 12 cm modular printed-circuit boards containing the microcomputer control and interface circuits [8,11]. Together with a 4 1/2-digit digital voltmeter, (DVM), and control-relay modules for valve control and power supplies, the system occupies a volume less than one cubic foot. A low-noise amplifier is used to amplify the detector signal of the chromatograph to the ± 2 volt range of the digital voltmeter. This system was constructed in the Electronics Shop of the College of Chemistry from modules obtained from the Lawrence Livermore Laboratory [12].

A special-purpose operating system was developed for the system to facilitate software programming and data acquisition [13]. The operating system contains an interactive teletype monitor and incorporates useful features such as an automatic breakpoint facility which is able to store the contents of memory registers in the microprocessor -the program status - in a push-pop (last-in, first-out) memory stack.

A floating-point mathematical package (FPP) makes possible 24-bit, binary floating-point arithmetic (16-bit mantissa, 8-bit sign and exponent), including square root. The FPP allows the performance of real-time data reductions with an operational accuracy of about $2 \times 10^{-3}\%$. Relatively sophisticated data processing can be implemented at some reduction in computational speed. The relatively slow Intel 8008 CPU with its 20 μ s cycle time can sample and process data at a rate up to two samples per second even with statistical subroutines in real-time operation. On the other hand, the newer Intel 8080 microprocessor has a 2 μ s cycle time and is "upward compatible" with the 8008 at the source-code level. In a similar control system the Intel 8080 CPU can run the same software 10 times faster.

In addition to the operating system and floating-point mathematics subroutines, a set of flexible, easy-to-sequence subroutines was is stored in programmable read-only memory (PROM). These subroutines constitute the gas chromatography package (GCP). The GCP includes subroutines to perform on-line data acquisition from the DVM at a predetermined sampling rate controlled by the real-time clock, and to perform real-time data reduction and statistical analysis. Other routines provided column temperature control and the plotting of the chromatogram on a teletype printer.

Program features of the system are shown in Table I, along with memory requirements and memory storage mode. All programs were written in the Intel 8008 octal language.

Table I

Program Features of the Microcomputer System
for Gas Chromatography

<u>Sub-Program</u>	<u>Memory Storage</u>	<u>Total Memory Occupancy (bytes)</u>
Restart and Breakpoints	PROM	256
Operating System	PROM	1024
Floating-Point Mathematics (24 bit words, 16 bit mantissa)	PROM	1536
Gas Chromatography Analytical Package	PROM	1024
Data Storage	RAM	256
Heater Control	RAM	512
Plotter Routines	RAM	256
Main Control Program	RAM	1024

The main program for control, analysis and data reporting is shown in Figure 1. The computer memory registers for control status are first cleared and initial control parameters are input from a teletype. Initial controls such as valve positions and the column heater temperature are then established, and a time delay is introduced for noise abatement.

In each computational cycle, typically 0.5 s, a new sample is acquired via the digital voltmeter in BCD format. This datum is converted into ASCII format for optional print-out on the teletype, and then to a floating-point representation. When the program is first started, a sequence of detector signal samples are recorded during N computational cycles and the mean and standard deviations are computed along with the baseline slope. If either the standard deviation of the noise or the baseline slope exceeds an input limit, an optional teletype message is generated, and the program returns to the delay cycle for noise abatement, as shown in Figure 1.

The flow chart for peak-start detection is shown in Figure 2. The detector signal is read and the mean (\bar{x}) and standard deviation (σ) of the preceding N samples are computed. If the condition,

$$x_n > \bar{x} + 2\sigma \quad \text{Eq. 1}$$

is met, a correction is made for the lost area of the peak preceding the satisfaction of the above condition, and an integration subroutine is started, based on the trapezoidal method. Otherwise, new \bar{x} and σ values are computed to establish a moving statistical window with a

sample number denoted by WINDOW. The size of this sample number is kept low to avoid excessive errors due to a declining or rising baseline.

One interesting feature of this start-of-peak test is its implicit noise-compensation property. If the input signal noise increases, then the admissible dynamic band ($\bar{x} \pm 2\sigma$) for the next sample x_n is increased. A smaller start-of-peak admissible dynamic band is set by a low-noise level.

When the integration is started, the last value \bar{x}_s is saved in memory. The program then begins to check for the peak maximum, as shown in Figure 3. Following the peak maximum, the program tests for the end-of-peak condition which requires a nonnegative instantaneous slope, γ_n , and an instantaneous slope exceeding the baseline-projected slope, γ_b , or

$$\gamma_n > 0 \quad \text{and} \quad \gamma_b > \gamma_n . \quad \text{Eq. 2}$$

In order to minimize the effect of noise, two successive data samples satisfying the end-of-peak condition are needed to terminate the integration. An alternate end-of-peak condition requires a return to the baseline and a minimum instantaneous slope, or

$$(x_n < \bar{x}_s + 2\sigma) \quad \text{and} \quad \gamma_n > \text{MSLOP} . \quad \text{Eq. 3}$$

The flow chart for the peak-end detection block is shown in Figure 4. Performing the tests with noncontiguous samples improves the noise-immunity of the algorithms.

Experimental

The accuracy of chromatographic peak integration was studied with propane-helium mixtures of known concentrations. Chromatograms were produced with a Beckman GC-2A gas chromatograph with a thermal conductivity detector, whose 0-1 mV output was amplified with an extremely stable, high-impedance differential amplifier and input to the 4 1/2-digit digital voltmeter. Peak integrals were obtained with the gas chromatographic subprogram, with a normalization factor of unity. The number of data points used in the moving statistical window was 16.

The program operation with a typical gas chromatogram is shown in Figure 5. The first COUNT samples was acquired without testing for the start-of-peak condition to compensate for any start-up noise or abnormality in the first samples, and to provide some statistical inertia in high-noise environment. To avoid unwanted program instability or lack of small-peak detection, it was found that the sampling rate should be preset to a value related to the expected noise characteristics of the signal. A 1-byte program variable introduced for this purpose was set by the initialization routine to act as the time constant of a digital filter.

Comparison to off-line data processing was achieved by recording the detector signal values on paper tape for input to a CDC 6400 computer. The off-line analytical program utilized the linear regression analysis of 30 successive detector signal values to establish the baseline prior to start of peak integration [5]. A third-degree

convolution of a group of nine successive values was then used to define the data-smoothed instantaneous value of the mid-point within each group, and to allow precise calculation of the first time derivative of the detector signal. The baseline slope at the start of each chromatographic peak was obtained by averaging the first derivative at each of the previous 22 data points of the smoothed baseline. The threshold value for the start of a chromatographic peak was set equal to the moving-baseline average plus twice the standard deviation of the previous 22 baseline values. The peak area was then obtained from Simpson's method, with the baseline defined by back-extrapolation to the smoothed tangent points on each side of the peak.

Results

The results of the integration of gas chromatographic data are summarized in Table II for propane-helium mixtures. The results showed that the present microcomputer integration algorithm gave integrals which are 0.2 to 0.6 per cent lower than those calculated by off-line computer analysis with data smoothing, for all samples where the peak area is much larger than the root-mean square integrated baseline noise (denoted by Integral Noise in Table II). The correlation of both analytical methods with the amount of propane injected was within 0.3% for major peaks. For minor peaks with areas near the integral noise values, the real-time analytical method showed significant error. Both the microcomputer algorithm and the off-line program were susceptible to the identification of false peaks.

Table II

Comparison of Analytical Results

Propane Injected (μmol)	Real-Time Microcomputer Analysis				Off-Line Analysis		
	Standard Deviation of Baseline Noise (mV)	Integral Noise (mV-s)	Peak Area (mV-s)	Correlation with Amount Injected	Standard Dev. of Smoothed Baseline Noise (mV)	Peak Area (mV-s)	Difference (mV-s)
866.0	5.99×10^{-4}	0.047	109.740	0.9988	3.69×10^{-4}	110.075	-.335
868.8	7.03×10^{-4}	0.056	109.980	0.9969	3.34×10^{-4}	110.658	-.678
321.9	1.51×10^{-3}	0.071	40.852	1.0004	3.69×10^{-4}	40.958	-.106
0.20	8.80×10^{-4}	0.004	0.003	0.1183	3.69×10^{-4}	0.196	-.192
0.10	5.78×10^{-4}	0.004	0.009	0.7096	3.69×10^{-4}	0.010	-.010
Noise	8.17×10^{-4}	0.003	Not Detected	-	3.69×10^{-4}	0.014	-.014
Noise	5.54×10^{-4}	0.002	0.007	-	3.34×10^{-4}	Not Detected	0.007

In many runs the entire digital data set was recorded by one of two methods, either by storing the data report in memory for print-out following the run, or by decreasing the sample rate below 2 Hz. An exemplary print-out in graphical form on a teletype is shown in Figure 6. The character "I" was printed after each data point during the integration routine to define the range of the detected peak. The relatively slow printing rate of the teletype, 11 characters per second, limited this method of chromatogram recording to exemplary use, however. Following the plot, a summary of output data was printed. Finally, an optional statistical summary was printed.

The performance of the peak-area integration algorithm on peaks of short time duration was tested by using the output from a signal generator. Simulated peaks with time durations as short as two sampling intervals could be readily integrated, but with some error. The start of the integration was initiated as soon as the input signal exceeded the sum of the running mean value of the signal and twice the running standard deviation, and as a consequence, peaks with a duration as short as two sampling periods could trigger the integration subroutine. While peaks of only a few sampling periods in length were easily detected, the limited number of intensity samples taken during them did not allow high accuracy in measured peak area.

Discussion

The difference between the off-line and the real-time results can be attributed partly to round-off error of the mathematical subroutines, and partly to the method used to detect the integral end-point. In the off-line program with data smoothing, the base-line is computed from the tangent fitted to the smoothed data, whereas in the present microcomputer algorithm the integration is terminated at the first data point where the peak-end condition is met. This end-point condition introduces an error of approximately half the integral noise, shown in Table II, or approximately $\sim 0.08\%$. A significantly larger error is introduced as a truncation error by the present floating-point mathematics subroutines, approximately 0.002% per computation. A comparison of input and stored data showed that the truncation error for the conversion from BCD to floating-point representation is approximately $\leq 0.01\%$. Consequently, the accumulated error for peaks containing many data points can be large. This source of error can be significantly reduced by extending the mathematical routines to handle triple-precision floating-point data, at a nominal sacrifice in computation speed.

The tests performed showed that the peak area measured by real-time microcomputer analysis is sensitive to the statistical criteria used for peak-start and peak-end whenever the baseline drift is much larger than the standard deviation of the baseline noise.

Part of this variation is introduced by the process of correcting in real time for lost area at the start of the peak before the baseline

slope was known. A second contributing factor is a variation in the assignment of the peak-end point without recomputing the statistics of the baseline.

A critical limitation of the present microcomputer algorithm arises from the simplified end-of-peak conditions currently in use. The GCP routines are not able to handle efficiently several possible chromatogram peak shapes. In particular the resolution and analysis of merged and trailing-peaks is not presently available in real-time. This is not necessarily an inherent limitation of the real-time approach and can be overcome by some additional software development. For example a nonnegative slope encountered after a first maximum is reached causes one type of end-of-peak condition explored in the present study. This condition could be used in a special subroutine which applies a drop-line separation between two merged peaks. Also, the storage of a "moving window" of N sequential detector samples, including the instantaneous sample, could be maintained in memory, from which a statistically smoothed baseline tangent point could be determined. More complex algorithms (such as preprogrammed tangent slope tests) are necessary to resolve trailing peaks by "tangent skimming" [2].

The use of a threshold of 2σ above the running baseline as the criterion for starting peak integration causes some sampling noise to be interpreted as false peaks. From a statistical viewpoint, one sample in every 44 will exceed the running baseline average of 2σ and trigger the integration subroutine. In the present analytical scheme, the false peak is not terminated at the next sample lying below the

threshold $\bar{x}_s + 2\sigma$, because the baseline is projected from the value \bar{x}_s . Therefore false peaks usually have a duration of three to five sampling periods. As an example, a noise peak of five sampling periods duration and an integral of 0.0056 mV was triggered by a sample of 0.006 mV when the running baseline average and standard deviation were 0.0034 and 0.0012 mV, respectively.

The detection of false peaks is a minor problem with the present microcomputer algorithm. The interpretation of noise peaks as real data can be avoided by several means. For example, the microcomputer program could be easily modified by ignoring peaks with area less than a minimum fixed value, by increasing the threshold for peak-start detection, or by applying a digital filtering technique. The technique of digital filtering with an 8-bit microcomputer has been tested by Seim and shown to be important for noise reduction [14].

The continuing advances in the microelectronics technology and microprocessor architecture toward lowering the cost while increasing the size and the sophistication of system components holds promise for expanded future feasibility of the microcomputers in a wide variety of chemical processes analysis and control applications.

Conclusion

An on-line real-time microcomputer approach to the problem of automatic analysis of gas chromatography has been developed. Experimental studies with the system have shown that measured peak areas are in good agreement with off-line analytical methods using data smoothing.

The flexibility of the bus-oriented microcomputer system used was found to facilitate the special-purpose hardware implementations needed for automated control, and the expanded instruction set of the Intel 8008 and 8080 microprocessors was found sufficient for expedient, software developments such as the operating system, the floating-point mathematics package and the peak integration package.

Due to the large time constants involved in the gas chromatographic control and analytical processes, even a relatively slow 8-bit microprocessor such as the Intel 8008 was able to execute control, and analytical subroutines to carry out a wide variety of preprogrammed, discrete functions. The system was shown to be able to perform real-time statistical analysis, tolerate a noisy input signal, and correct for baseline drift. The program developed was able to determine the integral of the peak area in real-time with accuracy limited mainly by truncation error of mathematical subroutines, and by baseline fitting at the peak end.

Acknowledgment

The financial assistance of the U.S. Energy Research and Development Administration and of the duPont de Nemours Company are gratefully acknowledged.

References

1. Bartle, E. R., and Meckstroth, E. A., Anal. Chem. 39, 273 (1967).
2. Hettlinger, J., "Comparison of Computing Integrators and Conventional Integrators for Chromatography," Tech. Bull. 111-74, Spectra Physics, Santa Clara, CA (1974).
3. Gill, J. M., J. Chromatog. Sci. 7, 731 (1969).
4. Leung, A. T., and Gill, J. M. Research/Development 25, 10, 36 (1974).
5. Dobbs, C. R., McGinnis, C. M., and Peddie, J. G., Amer. Lab. 6 (2), 69 (1974).
6. Karasek, F. W., Research/Development 25, 3, 28 (1974).
7. Hendrickson, J. L., Wallace, D., and Baumann, F., "Multi-User Gas Chromatography Data System with Foreground-Background Processing," Tech. Bull. 139-71, Varian Aerograph, Walnut Creek, CA (1971).
8. Donaghey, L. F., Bobba, G. M., and Jacobs, D., "A Microcomputer System for Real-Time Monitoring and Control of Gas Chromatographs," J. Chromatog. Sci. 14, 6, 274 (1976).
9. Hegedus, L. L., and Petersen, E. E., J. Chromatog. Sci. 9, 551 (1971).
10. Hendrickson, J. L., Wallace, D., and Baumann, F., "Multi-User Gas Chromatography Data System with Foreground-Background Processing," Tech. Bull. 139-71, Varian Aerograph, Walnut Creek, CA (1971).
11. A block diagram and a simplified logic drawing are available from the authors on request.

12. Fisher, E. R., "A System Approach for Processor Implementation,"
Tech. Rept. UCRL-75671, Lawrence Livermore Laboratory, Livermore,
CA, Oct. 1974.
13. Rubin, X. K., "RUTS: a Tutorial," Undergraduate Project Report,
College of Chemistry, University of California, Berkeley, CA (1975).
14. Seim, T. A., "A Microprocessor Sampled Data Process Controller,"
Tech. Rept. BNWL-1795, Battelle Pacific Northwest Laboratories.

List of Figures

Figure 1 Flow Chart of the Analytical Program for Real-Time Analysis of Gas Chromatograms.

Figure 2 Flow Chart of the Peak-Start Detection Block.

Figure 3 Flow Chart of the Peak-Maximum Detection Block.

Figure 4 Flow Chart of the Peak-End Detection Block.

Figure 5 Integration of a Typical Gas Chromatographic Peak.

Figure 6 Sample of Plotted Output and Data Report.

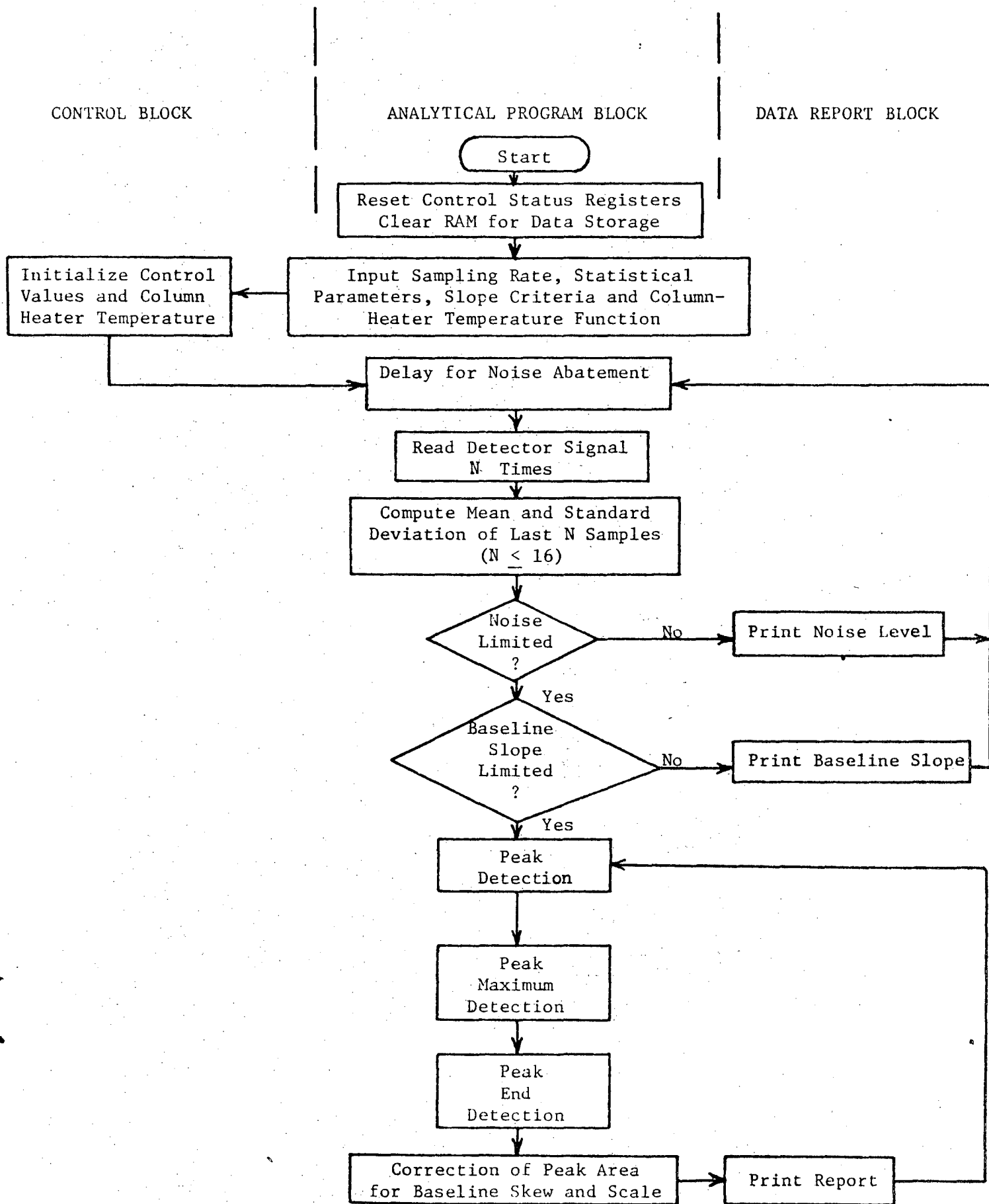
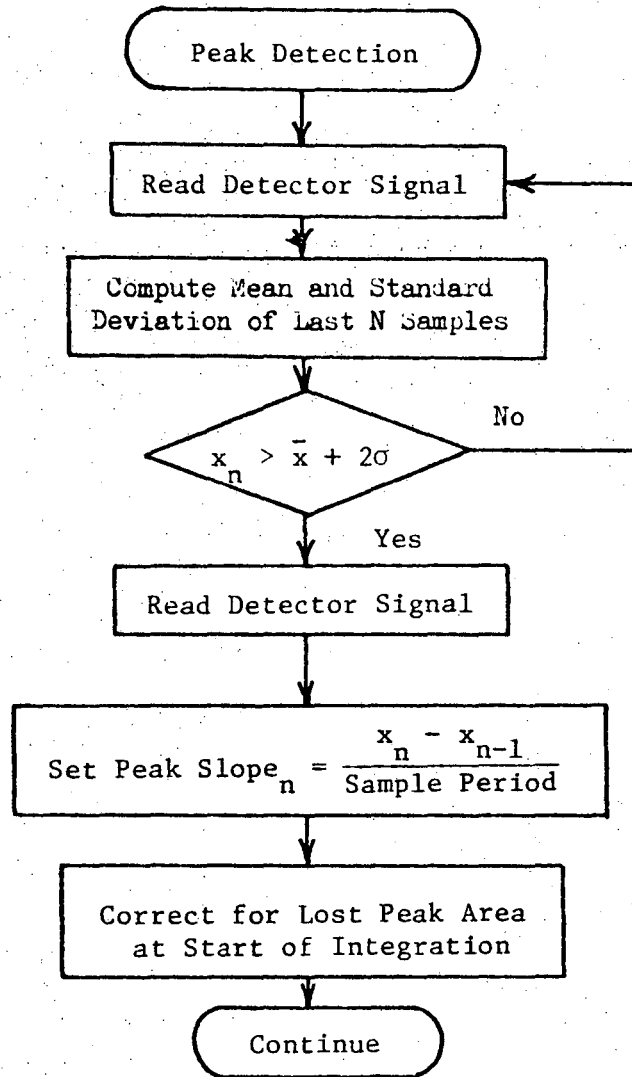
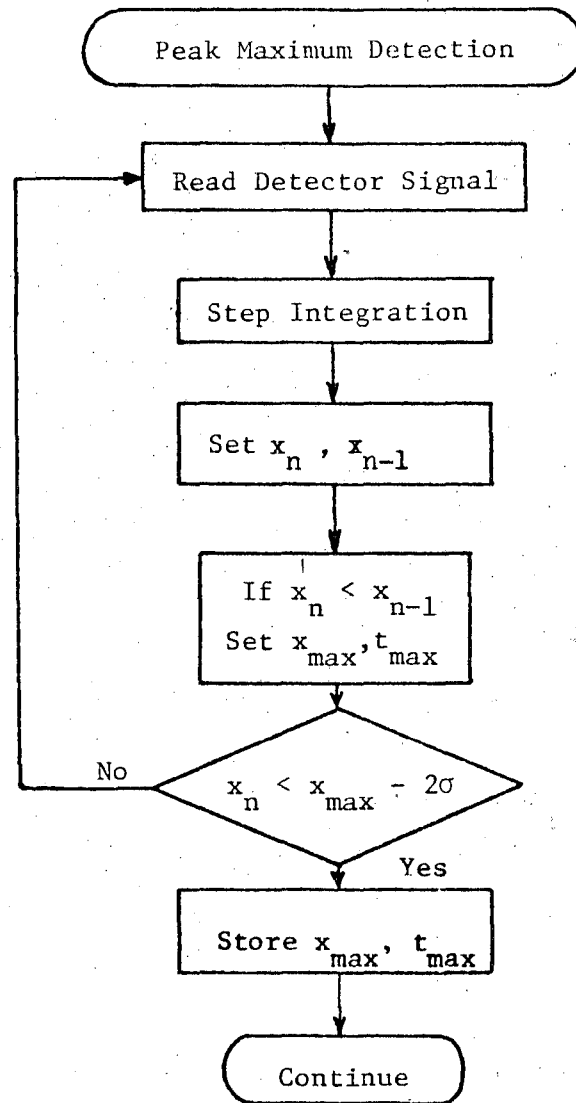


Fig. 1 Flow Chart of the Analytical Program for Real-Time Analysis of Gas Chromatograms



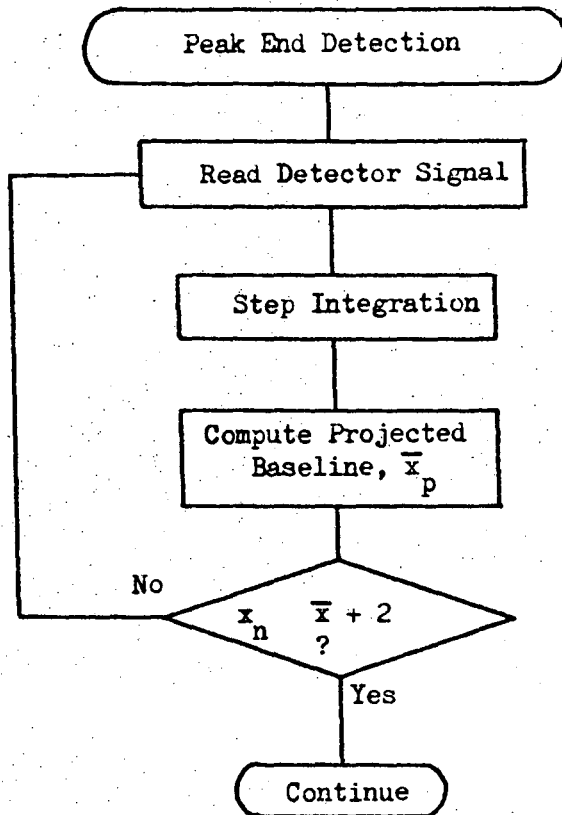
XBL 767-8733

Fig. 2 Flow Chart on the Peak-Detection Block



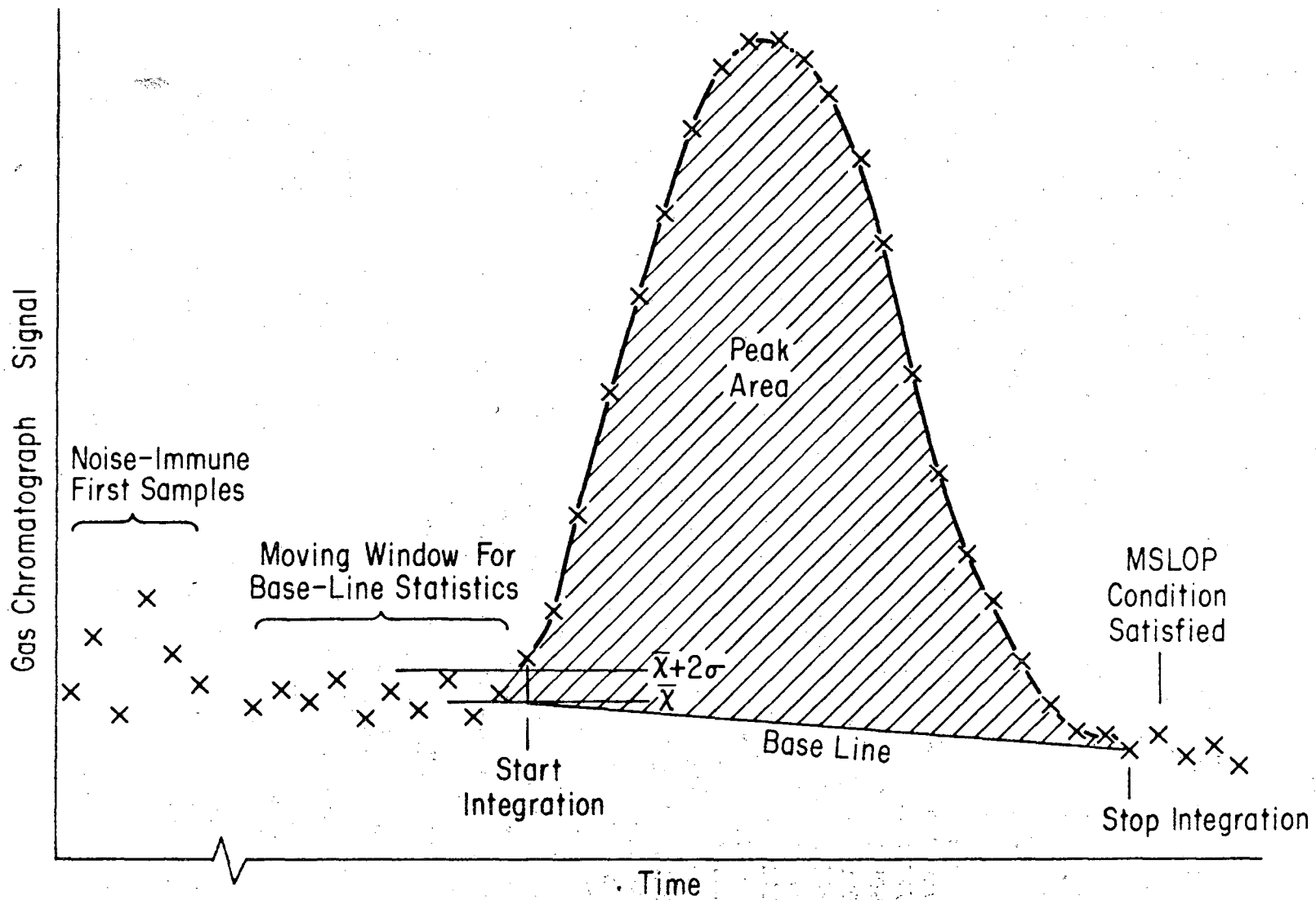
XBL 767-8734

Fig. 3 Flow Chart of the Peak-Maximum Detection Block



XBL 767-8735

Fig. 4 Flow Chart of the Peak-End Detection Block



00004602608

Fig. 5 Integration of a Typical Gas Chromatographic Peak

:+0.1837 H:*****
:+0.1610 H:*****
:+0.1598 H:*****
:+0.1804 H:*****
:+0.1623 H:*****
:+0.1543 H:*****
:+0.1390 H:*****
:+0.1293 H:*****
:+0.1176 H:*****
:+0.1129 H:*****
:+0.1043 H:*****
:+0.1759 H:*****
:+0.2727 H:*****
:+0.3908 I:*****
:+0.5006 I:*****
:+0.5945 I:*****
:+0.6678 I:*****
:+0.7648 I:*****
:+0.8104 I:*****
:+0.8787 I:*****
:+0.8958 I:*****
:+0.9211 I:*****
:+0.9051 I:*****
:+0.8374 I:*****
:+0.8332 I:*****
:+0.7277 I:*****
:+0.6850 I:*****
:+0.6136 I:*****
:+0.5725 I:*****
:+0.5297 I:*****
:+0.4673 I:*****
:+0.4182 I:*****
:+0.3491 I:*****
:+0.3080 I:*****
:+0.2725 I:*****
:+0.2287 I:*****
:+0.1804 I:*****
:+0.1431 I:*****
:+0.1137 I:*****
:+0.1041 I:*****
:+0.0798 I:****
:+0.0688 I:***
:+0.0586 I:***
:+0.0585 I:***

INTEGRAL	START	MAX	STOP	SCALE
9.6374E 01	1.2000E 01	2.3000E 01	4.4000E 01	1.2500E-01
:				
10.				

Fig. 6 Sample of Plotted Output and Data Report

This report was done with support from the United States Energy Research and Development Administration. Any conclusions or opinions expressed in this report represent solely those of the author(s) and not necessarily those of The Regents of the University of California, the Lawrence Berkeley Laboratory or the United States Energy Research and Development Administration.

TECHNICAL INFORMATION DIVISION
LAWRENCE BERKELEY LABORATORY
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720