United States
Environmental Protection
Agency

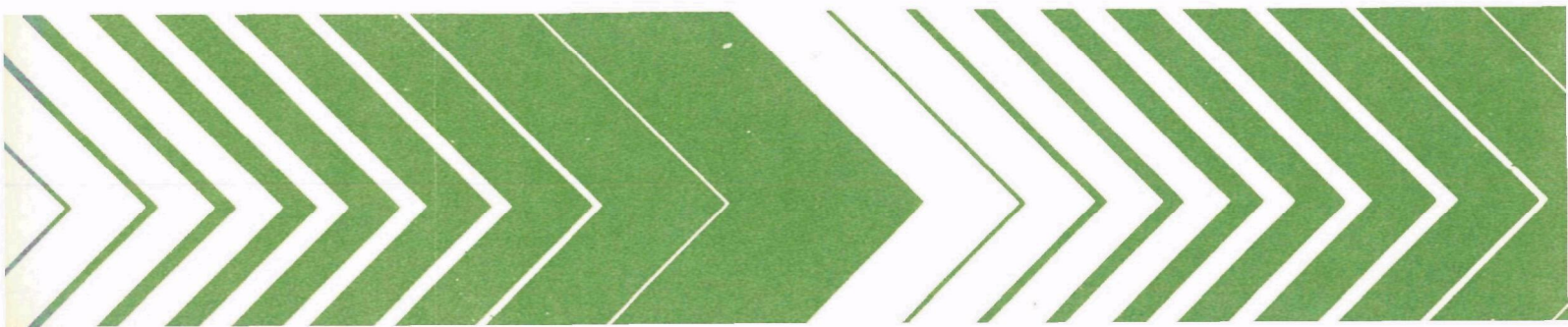
Environmental Monitoring and Support
Laboratory
Cincinnati OH 45268

EPA-600/4-79-038
June 1979

Research and Development

EPA

Functional Specifications for an Advanced Chromatography Automation System



RESEARCH REPORTING SERIES

Research reports of the Office of Research and Development, U.S. Environmental Protection Agency, have been grouped into nine series. These nine broad categories were established to facilitate further development and application of environmental technology. Elimination of traditional grouping was consciously planned to foster technology transfer and a maximum interface in related fields. The nine series are:

1. Environmental Health Effects Research
2. Environmental Protection Technology
3. Ecological Research
4. Environmental Monitoring
5. Socioeconomic Environmental Studies
6. Scientific and Technical Assessment Reports (STAR)
7. Interagency Energy-Environment Research and Development
8. "Special" Reports
9. Miscellaneous Reports

This report has been assigned to the ENVIRONMENTAL MONITORING series. This series describes research conducted to develop new or improved methods and instrumentation for the identification and quantification of environmental pollutants at the lowest conceivably significant concentrations. It also includes studies to determine the ambient concentrations of pollutants in the environment and/or the variance of pollutants as a function of time or meteorological factors.

This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.

EPA-600/4-79-038
June 1979

FUNCTIONAL SPECIFICATIONS FOR AN ADVANCED CHROMATOGRAPHY AUTOMATION SYSTEM

by

Jonathan E. Kopke
SouthWestern Ohio Regional Computer Center
University of Cincinnati
Cincinnati, Ohio 45220

Contract No. GS-05S-10458

Project Officer

John M. Teuschler
Physical and Chemical Methods Branch
Environmental Monitoring and Support Laboratory
Cincinnati, Ohio 45268

ENVIRONMENTAL MONITORING AND SUPPORT LABORATORY
OFFICE OF RESEARCH AND DEVELOPMENT
U.S. ENVIRONMENTAL PROTECTION AGENCY
CINCINNATI, OHIO 45268

DISCLAIMER

This report has been reviewed by the Environmental Monitoring and Support Laboratory-Cincinnati, U.S. Environmental Protection Agency, and approved for publication. Approval does not signify that the contents necessarily reflect the views and policies of the U.S. Environmental Protection Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

FOREWORD

Environmental measurements are required to determine the quality of ambient waters and the character of waste effluents. The Environmental Monitoring and Support Laboratory - Cincinnati, conducts research to:

+ Develop and evaluate techniques to measure the presence and concentration of physical, chemical, and radiological pollutants in water, wastewater, bottom sediments, and solid waste.

+ Investigate methods for the concentration, recovery, and identification of viruses, bacteria and other microbiological organisms in water; and to determine the responses of aquatic organisms to water quality.

+ Develop and operate an Agency-wide quality assurance program to assure standardization and quality control of systems for monitoring water and wastewater.

+ Develop and operate a computerized system for instrument automation leading to improved data collection, analysis, and quality control.

This report was developed in the Advanced Instrumentation Section of the Environmental Monitoring and Support Laboratoary in the interest of advancing laboratory techniques and quality control through computerization.

Dwight G. Ballinger
Director
Environmental Monitoring and Support
Laboratory - Cincinnati

ABSTRACT

This document contains a project definition, a set of functional requirements, and a functional design for a system which will link a commercial chromatography data system to the EPA Laboratory Automation System.

A Varian 220L Chromatography Data System was selected as the prototype system to be extended in this project, although these specifications can be adapted to other commercial systems. The current methods of using the Varian system are briefly described in this report. The bulk of the report is a detailed list of the additional functions to be performed by the EPA Laboratory Automation System. These functions include multi-point calibration, calculation of concentrations, identification of compounds, calculation of relative retention times, and calculation of quality control statistics. A general plan for the proposed system is provided.

This report was submitted in partial fulfillment of Contract No. GS-05S-10458 by SouthWestern Ohio Regional Computer Center. This report covers work done between November 13, 1978 and January 9, 1979.

CONTENTS

ACKNOWLEDGMENTS

The following people contributed greatly to the compilation of the functional requirements in this document: Thomas Bellar, Denis Foerst, William Potter, Leown Moore, and William Budde. Advice on systems development methodology was received from John Teuschler and Dennis Ryan. The assistance of these colleagues is gratefully acknowledged.

SECTION 1

INTRODUCTION

These functional specifications describe a prototype system for linking a commercially available chromatography data system to the EPA Laboratory Automation System.

A chromatography data system (such as the Varian 220L, the Perkin-Elmer Sigma 10, or the Hewlett-Packard 3380 or 3385) is an efficient system for real time data acquisition from one or more chromatographs, and for identifying compounds and determining their peak areas. However, programming any of the above commercial systems to perform additional or more sophisticated data reduction and to communicate with the EPA Sample File Control System would be a formidable, if not impossible, task.

Therefore, the system described in this document is designed to accept data from a chromatography data system, perform a variety of processing with the data, and ultimately report the results to Sample File Control, as shown in Figure 1. Although these functional specifications are written specifically for the Varian 220L at EMSL, Cincinnati, they can be applied in general to other commercial chromatography data systems.

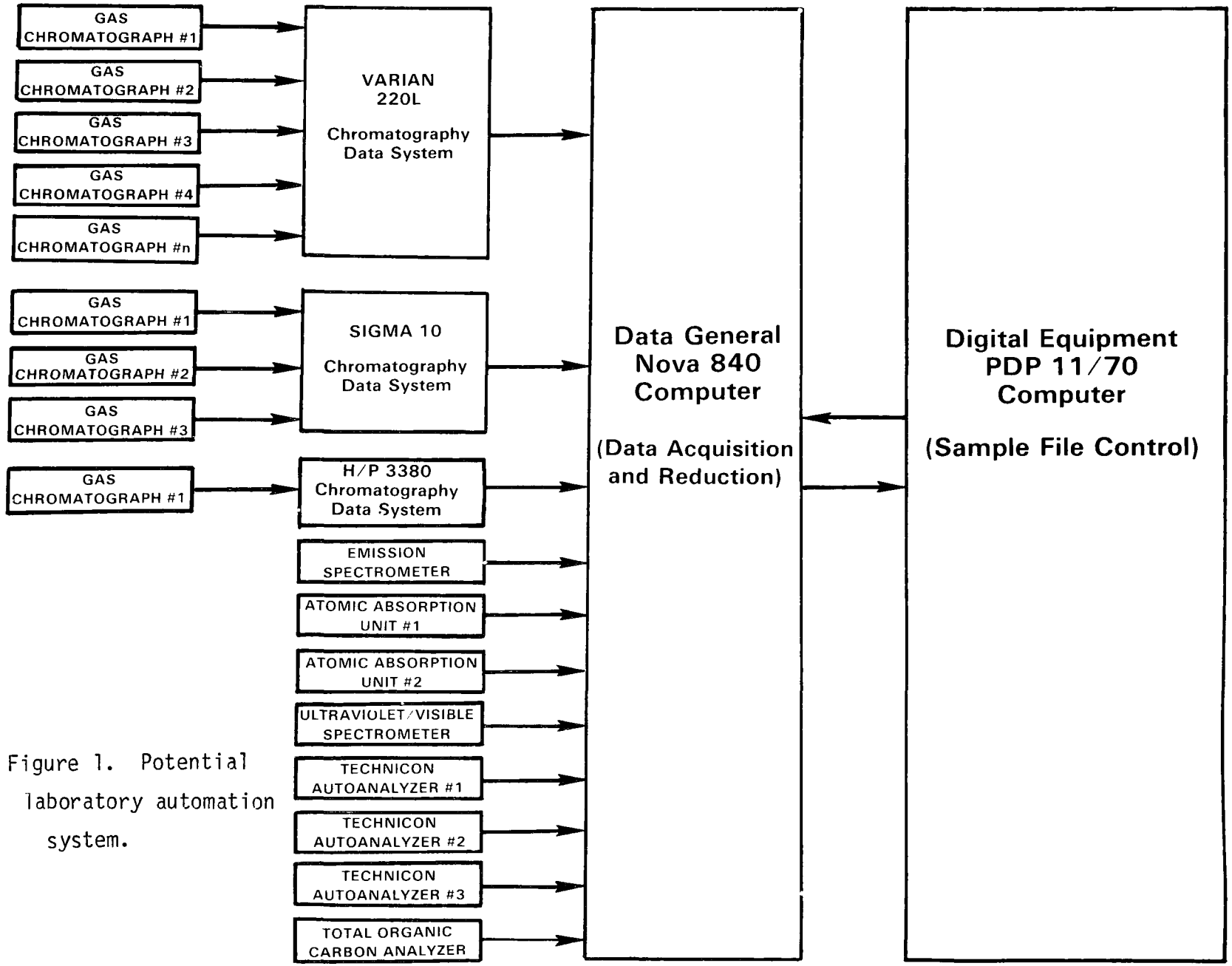


Figure 1. Potential laboratory automation system.

SECTION 2

PROJECT DEFINITION

The goal of this project is to develop a computer automation system which will accept data from a commercial chromatography data system and do additional processing of the data. The system will allow for:

+ unattended automatic transfer of data from the commercial system to the EPA Laboratory Automation System,

+ calibration using regression on multiple internal or external standards,

+ calculations of concentrations and relative retention times,

+ identification of compounds,

+ timely EPA standard quality control checks,

+ dissimilar analysis confirmation,

+ replicate statistics on samples,

+ detailed printed reports and calibration plots, and

+ transfer of results to the Sample File Control System.

SECTION 3

FUNCTIONAL REQUIREMENTS
OVERVIEW OF THE PROJECT

SCIENTIFIC PRINCIPLES OF CHROMATOGRAPHY

Chromatography is a technique of separating the various components of a sample so that they can be identified and measured. Basically, a chromatograph consists of a tube filled or coated inside with a material (the stationary phase) through which a steady stream of gas or liquid (the moving phase) is forced. When a sample is introduced into the tube, the moving phase forces the sample through the stationary phase and out the end of the tube. However, the stationary phase impedes the passage of some compounds more than others, and therefore the different compounds in the sample arrive at the end of the tube at different times. A detector at the end of the tube measures the amount of matter leaving the tube, and that variable is plotted against time on a recording device.

The retention time of a compound is the length of time that the compound is retained in the stationary phase. A compound may be identified in chromatography by comparing its retention time to the retention times of known compounds. Furthermore, the concentration of a compound in a sample is calculated by comparing the area of its recorded peak to the area of the peak produced by a known concentration of the same compound.

INSTRUMENTS TO BE AUTOMATED

Tracor 560 Gas Chromatograph, EMSL-CI, Room 588
Infotronics 2400 Gas Chromatograph, EMSL-CI, Room 588
Tracor MT220 Gas Chromatograph, EMSL-CI, Room 586
Tracor MT222 Gas Chromatograph with auto-injector, EMSL-CI, Room 586
Hewlett-Packard 5700 Gas Chromatograph, EMSL-CI, Room 586
Waters Associates 440 Liquid Chromatograph, EMSL-CI, Room 574
Varian Aerograph 1400 Gas Chromatograph, MERL, Room B22
Varian Aerograph 2100 Gas Chromatograph, MERL, Room B22
Two Tracor MT222 Gas Chromatographs, MERL, Room B22
Perkin-Elmer 900 Gas Chromatograph, HERL, Room 670.

All of these instruments are to be automated using the Varian Chromatography Data System 220L in Room 574 for initial processing of data. All room numbers refer to the Environmental Research Center, Cincinnati.

DESCRIPTION OF THE EXISTING SYSTEM

The chromatographs in EMSL, MERL, and HERL have already been automated to some extent. Currently, several gas chromatographs and a high pressure liquid chromatograph are interfaced to a Varian computer. To use this automation system, an analyst must first generate a method in the Varian. This involves providing the Varian with the retention times and eight-character identifiers for the compounds which can be expected in the standards and samples. The method can be retained by the Varian and used repeatedly.

Standards and samples are prepared in three different ways for different applications. The type of preparation influences the calculations which must be done to find true concentrations. The three types of preparation are:

Liquid/Liquid Extraction

The compounds of interest are extracted from the aqueous sample by an organic solvent. The extract is then concentrated, and a small portion of it is injected into the chromatograph. (A standard may be extracted from an aqueous solution, or prepared at a known concentration in the organic solvent.)

Purge and Trap Method

The compounds of interest are extracted from the aqueous sample by an inert gas. The compounds are trapped in a short column and then backflushed into the chromatograph. (In this case, a standard is prepared at a known concentration in an aqueous solution and is treated in exactly the same way as a sample.)

Direct Aqueous Injection

A small portion of the actual aqueous sample is injected directly into the chromatograph. (A standard is treated in the same way, although it may be dissolved in an organic solvent rather than in water.)

After any method of preparation and injection, a sample or standard travels through the chromatograph tube. The various compounds are separated and ultimately pass through a detector. The signal from the detector is transmitted to the Varian and digitized. The Varian determines the retention time of each peak, the total area of all of the peaks, and the percent of the total area for each peak. From the retention times provided by the analyst during method generation, the Varian determines what compounds are present in the sample or standard, and it prints the eight-character identifier for each compound found. The Varian also provides a variety of other information including a simple calculation of the concentration of each compound based on a one-point calibration.

To calculate concentrations more accurately, an analyst must first develop a calibration curve for each of the compounds covered by the method. A calibration curve for chromatography relates peak area to the

concentration (or mass) of the compound. To establish a calibration curve, an analyst normally uses several calibration standards, and may use several replicates of each of the standards. In a single standard, the concentrations of the various compounds are not necessarily the same. A calibration curve is most often a straight line; however, other curves are sometimes employed.

The concentrations of compounds in samples are calculated using the calibration curves, although other factors must be taken into account such as the volume of the injection and (in liquid/liquid extraction) the volumes of the extract and the original aqueous solution.

Calibration standards are also used as control (check) standards. They are compared to the analyst's previous experience to determine if the instrument is in control.

There are several other complications unique to chromatography. Sometimes one peak represents the total of several compounds. In such a case, the analyst must match the eight-character identifier on the Varian report to the names of several compounds for the final report. Also, in some cases, one pesticide may be a mixture of several compounds. This implies that several peaks represent one pesticide. In such a case, the analyst must total several peaks in the calibration standards, calibrate on that basis, and then total the corresponding peaks from the samples to calculate the concentration of one pesticide.

ADDITIONAL EXISTING HARDWARE AND SOFTWARE TO BE UTILIZED

This chromatography project is one branch of a much larger laboratory automation project. Therefore, some of the basic decisions about system hardware and software have already been made within the context of the larger project. Thus, these functional requirements are not strictly hardware and software independent.

This project will use an existing Data General Nova 840 Minicomputer with the operating system RDOS (Revision 6.2) and the language Extended BASIC (Revision 4.3). Varian output reports will be transferred to the Nova for further data processing.

TIME AND FINANCIAL CONSIDERATIONS

The first version of the gas chromatography automation system should be operational for testing by March 1, 1979. The system described in the functional requirements should be completed and documented by June 30, 1979. No more than $30,000 should be spent on developing the system.

SUPPORT OF THE COMPLETED SYSTEM

Support of an automated system is divided into software and hardware support. Since this project will involve two computer systems, the hardware and software support of each system will be discussed.

The Advanced Instrumentation Section will provide assistance in the selection and diagnostic analysis of computer terminals used with both systems; however, when it is cost effective, repairs to the terminals will be made by the manufacturer's representatives. The purchase order for such repairs will be completed by and billed to the user's section or branch.

The Varian computer system hardware is supported by the manufacturer under a maintenance contract which covers all hardware except the terminals. If additional hardware is necessary to implement this project, the custom hardware will be documented by the Advanced Instrumentation Section in such a way as to allow repair on a module-swapping basis, and spare cards will be provided with any custom hardware installation.

The system software on the Varian has had no maintenance for several years. The bugs which exist in the system are known and have been avoided procedurally.

The Nova computer system also has a maintenance contract for all vendor-supplied hardware. Custom hardware on the Nova system is maintained by Advanced Instrumentation personnel.

The BASIC software and system software for the Nova computer are supported by the Data General Corporation. Any custom modifications to the BASIC software (or assembly language code for a microprocessor) will be supported by the Advanced Instrumentation Section, as will the applications software.

The maintenance of the chromatograph instruments is the responsibility of the users of those instruments.

FUTURE CONSIDERATIONS

The goals of this automation project are limited to those stated in the functional requirements. However, the system should be designed to allow for future enhancements to be easily implemented.

A major future consideration is that dedicated computers other than the Varian will eventually be interfaced with the system. These other computers include the Perkin-Elmer Sigma 10, the Hewlett-Packard 3380, the Hewlett-Packard 3385, and the Autolab chromatography system. The software of this system will be modularized to facilitate the addition of other dedicated computers, and the mode of transferring information from the dedicated computers to the Nova will be generalized.

Another future consideration is that of the internal standards calibration method. This automation system will be designed to include modules for the internal standards method. However, the design and implementation of these modules will be deferred until the rest of the project is operational and until the internal standards method is more rigorously defined for conventional gas chromatography.

SECTION 4

FUNCTIONAL REQUIREMENTS
REQUIREMENTS OF THE SYSTEM

DATA TRANSMISSION AND STORAGE REQUIREMENTS

Eleven chromatographs will be connected to the Varian. Servicing all of these instruments, the Varian will produce at most 115 reports per day during normal working hours, plus 40 more reports during the night. Each report will be no more than 100 lines long, and 30 lines will be the average length.

There are currently about 28 Varian methods in use, although in the future this number will increase. Each method deals with no more than 60 named compounds, although the average number will be about 15. Presently, fewer than 1000 different compounds are being identified by chromatography, although this number will also increase substantially in the future.

Data transmission from the Varian will be accomplished as quickly as possible, since the Varian will lock a user out of its system until that user's previous report has been transmitted. This data transmission will occur without human intervention to accommodate an auto-injector running during the night.

The Nova will save the 100 most recent Varian reports for each chromatograph, and will automatically delete old reports. The Nova will keep a list of the ID numbers for the existing reports, and this list will be available for the analyst to examine.

All Varian reports will be saved in one programming area in the Nova, since they can be distinguished by their ID numbers.

DESCRIPTION OF THE VARIAN REPORT

Normal Report

Figure 2 shows a typical Varian report, as described below.

LINE	COLUMN	CONTENTS
1	1-2	blank
	3-4	GC
	5-26	blank

	27-34	date, left justified
	35-47	blank
	48-52	time, left justified
2	1-70	blank
3	1-2	blank
	3-6	INST
	7-10	blank
	11-12	instrument number, right justified
	13-24	blank
	25-30	METHOD
	31-32	blank
	33-36	method number, right justified
	37-39	method number modifier, usually -ES
	40-47	blank
	48-55	operator identification, left justified
4	1-70	blank
5	1-16	blank
	17-46	title, left justified
6	1-70	blank
7	1-70	blank
8	1-2	blank
	3-6	PEAK
	7-10	blank
	11-14	NAME
	15-17	blank
	18-21	TYPE
	22-28	blank
	29-32	COMP
	33-37	blank
	38-41	RETN
	42-45	blank
	46-49	CORR
	50-56	blank
	57-60	AREA
	61-64	blank
	65-70	FACTOR
9	1-2	blank
	3-4	NO
	5-28	blank
	29-30	UG (or other two-character unit abbreviation)
	31-37	blank
	38-41	TIME
	42-45	blank
	46-49	TIME
	50-56	blank

	57	%
10	1-2	blank
	3-6	peak number, right justified
	7-8	blank
	9-16	compound identification (or UNEXP)
	17-18	blank
	19-20	peak type, left justified
	21-22	blank
	23-33	calculated concentration, right justified
	34-35	blank
	36-41	retention time, right justified
	42-43	blank
	44-49	corrected retention time, right justified
	50	& (or blank)
	51-52	blank
	53-62	percent area, right justified
	63-64	blank
	65-70	response factor, right justified

Lines 11 to 10+N are identical to line 10. N is the number of peaks.

11+N	1-70	blank
12+N	1-22	blank
	23-33	total concentration, right justified
13+N	1-70	blank
14+N	1-2	blank
	3-12	TOTAL AREA
	13-14	blank
	15-23	total area, right justified

The Varian is capable of producing a variety of other formats for output reports, but only the format described above will be used for this project.

Unavoidable Exceptions

Figure 2 shows a typical Varian report, but Figures 3 and 4 show unavoidable exceptions to the rules. In Figure 3, line 9 is a message from the Varian system. Such a message may occur at any point in a report. A message always begins with the word MESSAGE in columns 1 through 7, and includes one or more message numbers which may be quite important to an analyst.

The format of the report in Figure 3 shows a second unavoidable exception to the rules. When the Varian is unable to apply a method to a sample, it defaults to this format to report the results. Such a report is not useful in this project.

A third exception to the rules is the paging which occurs when a Varian report exceeds 11 inches in length. In Figure 4, after the thirty-ninth

GC 11/1/78 14:10

INST 1 METHOD 100-ES AP&T

1 DAY SPIKE #2 CFT

PEAK NO	NAME	TYPE	COMP UG	RETN TIME	CORR TIME	AREA %	FACTOR
1	UNEXP	P	0.5065	764	42	1.0049	1000
2	CNCL	PB	0.0000	790	68	3.8631	0
3	UNEXP	P	1.8531	1046	325	3.6770	1000
4	UNEXP	P	2.2701	1057	336	4.5044	1000
5	11CL2C=C	P	0.0000	1095	374&	9.0605	0
6	CCL4	PB	42.9654	1369	640&	14.0213	6080
7	12CLC3+	P	8.6511	1405	676&	17.1650	1000
8	UNEXP	P	7.2651	1415	686	14.4149	1000
9	CL3CH=C	P	50.6580	1487	766&	10.7719	9331
10	CHBR2CL	P	61.7029	1530	808&	11.9360	10257
11	UNEXP	P	2.7222	1568	835	5.4012	1000
12	2BRCLC3@	P	0.0000	1605	872&	3.8208	0
13	CHBR3	PB	2.8358	1690	958&	0.3587	15685
			181.4301				

TOTAL AREA 503990

Figure 2. Normal Varian report

GC 11/21/78 09:17

INST 2 METHOD 205-ES P&T

40/1000 STD #2

MESSAGE 00, 23,

PEAK NO	NAME	TYPE	RETN TIME	CORR TIME	AREA %
1			18		16.4304
2		P	64		0.7135
3		P	93		4.3922
4		P	135		0.9373
5		P	165		1.7541
6		P	211		4.2019
7		P	318		3.3525
8		P	474		6.5665
9		P	542		5.2780
10		P	594		5.1760
11		P	629		0.7541
12		P	675		0.2889
13		P	688		0.3839
14		P	747		6.1907
15		P	807		2.0400
16		P	889		5.6351
17		P	907		3.8524
18		P	941		0.7136
19		PB	985		2.7393
20		P	1038		1.7316
21		P	1082		0.1564
22		P	1157		7.1984
23		P	1275		3.4223
24		P	1293		4.8589

Figure 3. Error message and default format.

27	123CLC3	P	0.2947	1782	1073&	4.1181	20
28	CL4C=C	P	259.4929	1808	1099&	3.2643	22207
29	12CL2C4	P	92.0854	1857	1145&	2.3729	10841
30	UNEXP	P	20.2071	1960	1249	5.6450	1000
31	CLBZ	P	154.2618	2006	1295&	4.9745	8663
32	UNEXP	P	12.1952	2015	1304	3.4068	1000
33	UNEXP	P	13.5698	2046	1336	3.7908	1000
34	UNEXP	P	14.2975	2104	1394	3.9941	1000
35	UNEXP	P	10.5835	2152	1443	3.9566	1000
36	UNEXP	P	9.5953	2181	1472	2.6805	1000
37	UNEXP	P	4.4091	2571	1865	1.2317	1000
38	UNEXP	P	5.0859	2616	1911	1.4208	1000
39	UNEXP	PB	3.5579	2676	1971	0.9939	1000

PEAK NO	NAME	TYPE	COMP UG	RETN TIME	CORR TIME	AREA %	FACTOR
40	UNEXP		0.5064	2735	2035	0.1414	1000
			1195.8286				

TOTAL AREA 3579640

Figure 4. Paging feature of the Varian report.

peak, the Varian skipped 12 lines, typed a new set of column headings, and then continued the report. Possibly this paging feature can be disabled during system generation on the Varian. (See the Varian manual, page 2-5, item 5.)

SIGNIFICANT DIGITS

On the Varian reports, only the three high order digits of each number are significant.

FLAGS FOR USE IN DATA TRANSMISSION

The following flags will be used to identify the beginning of a Varian report, the end of the report, and the format of the report.

The first line of a Varian report always begins with the letters GC in columns 3 and 4.

The last line of a report always begins with the words TOTAL AREA in columns 3 through 12. If for some unexpected reason, a report does not contain a line beginning with the words TOTAL AREA, then the next appearance of the letters GC in columns 3 and 4 would indicate that the end of the report has been missed, and another report is beginning.

Only one report format will be acceptable in this project. The flag to be used to identify valid reports is the line of column headings which reads,

PEAK NAME TYPE COMP RETN CORR AREA FACTOR

as described previously. If any report reaches the Nova in a format other than this, a notation will be made of its arrival, but the contents of the report will be ignored.

INJECTION ID NUMBERS

Injection ID numbers will be eight-digit numbers of the form IIMMDDSS, where II = instrument number, MM = month, DD = day, and SS = sequence number. The Nova will automatically assign an ID number to an injection when its Varian report reaches the Nova. The instrument number will be obtained from the Varian report. The month and date will be acquired from the Nova system clock, except in a case noted below. The sequence number will be generated by a counter maintained for each instrument by the Nova. The counter will be automatically reset to 1 when the date of an injection is different from the date of the previous injection. Thus, for example, the fifth injection into instrument number seven on January 23 would be assigned ID number 07012305.

Using this system, ID numbers will be assigned by the Nova without human intervention and without the analyst having to type them into the Varian. Nevertheless, the analyst will be able to manually determine the ID number of each injection immediately when he injects it (or when he loads it into the sample wheel of the auto-injector).

This system requires one further refinement to keep injection ID numbers organized when a series of injections continues for more than one calender day (an auto-injector running past midnight, for example). As mentioned above, the sequence number counter for an instrument will be automatically reset to 1 when the date changes. However, if an analyst specifies a date (in the form DATEMMDD) in the "operator identification" field of the Varian report, the Nova will use that date in the ID numbers it generates, instead of the date in the system clock. Thus, the sequence number counter will not be reset until the analyst changes the date specified. Analysts who do not need this option will simply not begin the "operator identification" field with the keyword "DATE."

EDITING OF THE VARIAN REPORT

The entire Varian report will be printed on paper and will be transferred to the Nova. However, the report can be pruned considerably before it is stored in the Nova. The programs in the Nova must store the following information for each sample:

1. instrument number
2. method number
3. operator identification
4. title
5. total area

The Nova programs must also store the following information for each peak within a sample:

1. name of compound
2. type of peak
3. retention time
4. percent area

Furthermore, any MESSAGE from the Varian system must be saved in the Nova and displayed when the report is used so that the analyst does not have to check the original Varian report for messages.

None of the other information on the Varian report is needed for the calculations to be done by programs in the Nova.

EXTERNAL STANDARDS CALIBRATION METHOD

The existing calibration procedure in the Varian will not be used in this project. The Varian will deal with all calibration standards as if they were ordinary samples, and the Varian reports for calibration standards

will be stored in the Nova until the analyst is ready to execute the calibration procedure.

The Nova programs will allow for up to 15 calibration standards, including replicates. Curve fitting will be accomplished using one of the following methods, at the option of the analyst: 1) least squares linear regression using three or more standards, 2) least squares quadratic regression using four or more standards, or 3) linear interpolation using the origin and one or more standards. The analyst will be able to recalibrate at any time by deleting standards and/or introducing new standards.

Two variations of external calibration will be available to allow for the different methods of sample preparation.

Purge and Trap Method

When an analyst is ready to run the calibration procedure in the Nova for standards prepared by the Purge and Trap method, it will be necessary to enter into the Nova from the keyboard the injection ID number for each of the calibration standards and the concentration of every named compound within each standard. The analyst will be permitted to enter the concentration of standards in micrograms per liter. The program in the Nova will acquire the necessary peak areas from the Varian reports, and will calculate a set of calibration curve equations relating the peak area for each compound to its concentration. The calibration curve equations will be developed by linear regression, quadratic regression, linear interpolation, or linear regression forced through the origin, at the option of the analyst. A fitting error (percent relative standard error of estimate) will be calculated so that the analyst can objectively judge the goodness of fit.

Liquid/Liquid Extraction and Direct Aqueous Injection Methods

The calibration procedure for standards prepared by Liquid/Liquid Extraction or Direct Aqueous Injection differs from the first procedure in that it is based on the mass of the standards rather than concentration. To initiate a calibration process, an analyst will enter from the keyboard into the Nova program the injection ID number of each standard, the injection volume in microliters for each standard, and the concentration in micrograms per liter. The program will then multiply the injection volume times the concentration times 1000 to get the mass in nanograms for each compound. Next, the program will develop a set of calibration curve equations relating the peak area for each compound to its mass, and a fitting error will be calculated as described above.

After either calibration procedure, the set of calibration curves for a method will be stored in the Nova for repeated use, and the calibration points will also be stored for plotting purposes.

CALIBRATION CURVE PLOTTING

The analyst will have the option of plotting a calibration curve for each compound of a method. The plot will contain heading information

including the name of the analyst, the name of the compound, the date of calibration, the type of fit, the fitting error and the regression equation. The plot itself will be a graph with concentration (or mass) on the horizontal axis and peak area on the vertical axis. Axes will be labeled and scaled. A table of residuals will also be provided.

All plotting will be done on existing Tektronix 4000 Series Computer Display Terminals.

CONCENTRATION CALCULATIONS USING EXTERNAL CALIBRATION

Varian reports for samples will be stored in the Nova until the analyst is ready to calculate the concentrations for a set of them. Three different types of concentration calculations will be available for the three methods of sample preparation.

Purge and Trap Method

When an analyst is ready to calculate concentrations for samples prepared by the Purge and Trap method, the injection ID numbers will be entered into the Nova from a keyboard. The program running in the Nova will calculate the concentrations in micrograms per liter for every named compound within each sample by simply substituting the peak area from the Varian report into the appropriate calibration equation.

Liquid/Liquid Extraction

Concentration calculations for the Liquid/Liquid Extraction method will require the analyst to enter from a keyboard not only the ID number for each injection, but also the volume of the injection in microliters, the volume of the extract in milliliters, and the volume of the water extracted in liters. The program running in the Nova will use the peak areas from the Varian reports and the calibration equations to calculate the mass in nanograms of each compound. It will multiply this mass by the volume of the extract, and then divide by the product of the volume of the injection and the volume of the water extracted to determine the concentration of each compound in micrograms per liter.

Direct Aqueous Injection

To calculate the concentrations of samples run by Direct Aqueous Injection, the analyst will enter into the Nova from a keyboard the ID number and injection volume in microliters for each injection. Using the peak area from the Varian report and the set of calibration curves, the program in the Nova will calculate the mass in nanograms of every compound in each sample, and will then divide by the injection volume to get the concentration in nanograms per microliter. Dividing by 1000 will produce concentration in micrograms per liter.

IDENTIFICATION OF COMPOUNDS

The identification of compounds measured by gas chromatography is cur-

rently hampered by the fact that the NAME field on the Varian output report is only eight characters wide, while the chemical name of a compound may be 30 or more characters long. The following procedure will be used to enable the Nova program to print the full chemical names on final output reports:

When an analyst generates a method in the Varian, each of the compounds involved will be identified by a name consisting of eight or fewer alpha-numeric characters. For example, an analyst may refer to bromodichloro-methane as CHBRCL2. No conflict will result if two analysts happen to use the same identifier for different compounds; the only restrictions will be that 1) within a Varian method, no two compounds will be allowed to have the same identifier, and 2) within two Varian methods which will be used for dissimilar analysis confirmation, identifiers will have to be consistent.

The analyst will then create a corresponding method in the Nova. As part of the Nova method-generation procedure, the analyst will be asked to type the identifiers used in the Varian, and the matching chemical names and Chemical Abstracts Services (CAS) Registry numbers. For example, the analyst would have to tell the Nova program that CHBRCL2 represents bromo-dichloromethane, and that the CAS Registry number for that compound is 75-27-4. The analyst will not be required to use any particular chemical name for a compound, but rather will be able to use any synonym he prefers. For example, dichlorobromomethane could be used instead of bromodichloro-methane. In the event that one peak in a Varian method represents more than one compound, the analyst will use one identifier for that group of com-pounds, and will supply the Nova program with the names and CAS Registry numbers for all of the individual compounds associated with that identifier.

Each analyst will manually look up the CAS Registry numbers needed in the Chemical Abstracts Ninth Collective Index (1972-1976). Volumes 76 through 85 of that work are the Formula Index, which seems to be the most convenient source of CAS numbers currently available. The user's manual for this system will include a list of EPA priority pollutants and corresponding CAS numbers.

When an analyst types a CAS Registry number into the Nova program, the program will recalculate the value of the check digit as described in the CAS Registry Handbook. This process will be used to check the number for typing errors, but it will not be used to verify that the analyst has used the correct CAS number for the compound named.

The final reports from the Nova program will include the full names and CAS Registry numbers for the compounds measured, but will not include the identifiers used by the Varian.

Because CAS Registry numbers may often be superfluous, the analyst will have the option of not entering them into the Nova method at all. However, CAS numbers may be required for further processing of the data in other EPA computer systems, specifically the Sample File Control System.

RELATIVE RETENTION TIMES AND CAPACITY RATIOS

As part of the method generation procedure on the Nova, the analyst will elect either to have relative retention times or capacity ratios calculated for samples. If the analyst chooses relative retention times, it will be necessary to enter the Varian identifier for the compound to whose retention time all other compounds are to be compared. Then when the analyst runs the concentration calculation procedure, the Nova program will automatically calculate the relative retention time for each compound using the retention times from the Varian report and the formula:

$$\text{Rel. Retention Time} = \frac{\text{Retention Time}}{\text{Retention time of Reference}} \times 100.$$

If during Nova method generation, the analyst chooses to use capacity ratios instead of relative retention times, it will be necessary to enter the dead volume time for the method. Then when concentrations are calculated, the Nova will calculate the capacity ratio for each compound using the retention times from the Varian report and the formula:

$$\text{Capacity Ratio} = \frac{\text{Retention Time}}{\text{Dead Volume Time}} - 1.$$

PROCESSED DATA REPORT

Processing an injection will involve determining the full chemical name of each compound, calculating the relative retention time or capacity ratio for each compound, and calculating the concentration of each compound. After an injection has been processed, a report will be printed. This report will contain the following heading information: 1) name of the analyst, 2) the instrument number, 3) the method number, 4) the date of the report, 5) the title of the injection from the Varian report, if any, and 6) the injection number and date. The report will also include the following data for each peak: 1) the full chemical name(s) and CAS Registry number(s) for the compound(s) associated with the peak, 2) the peak type from the Varian report, 3) the peak area, 4) the retention time, 5) the relative retention time or capacity ratio, and 6) the concentration of the compound in micrograms per liter. Unexplained peaks will be listed in the report, but will not be matched to a chemical name nor assigned a concentration.

QUALITY CONTROL STATISTICS

Quality control statistics will be calculated for control (check) standards, spiked samples, duplicate samples, and surrogate spikes.

Control or Check Standards

The analyst will enter from the keyboard into the Nova program the ID number of a control (check) standard and the prepared concentration in micrograms per liter of each compound within the standard. The Nova program will print a report showing the prepared concentration, measured concentration, and percent recovery for each compound. Percent recovery for stand-

ards will be calculated by the formula:

$$\text{Percent Recovery} = \frac{\text{Measured Concentration}}{\text{Prepared Concentration}} \times 100.$$

Spiked Samples

The analyst will enter from the keyboard into the Nova program the ID number of the unspiked sample, the ID number of the spiked sample, and the amount of spike in micrograms per liter for each compound within the sample. The Nova will print a report showing the measured concentration of the unspiked sample, the measured concentration of the spiked sample, the amount of spike, and the percent recovery of the spike for each compound. Percent recovery for spikes will be calculated by the formula

$$\text{Percent Recovery} = \frac{\text{Spiked Conc. - Unspiked Conc.}}{\text{Spike}} \times 100.$$

Duplicate Samples

The analyst will enter from the keyboard into the Nova program the ID numbers of both members of the duplicate pair. The Nova will print a report showing both of the measured concentrations and the absolute value of the difference between them for each compound.

Surrogate Spikes

A surrogate spike is a known amount of a pure compound that is added to a sample, but which was not previously present in the sample. The percent recovery of a surrogate spike can be used as an estimate of the recoveries of the compounds in the sample when actual spikes are not used. The procedure for processing a surrogate spike will be the same as that for a check standard, except that only one compound will be involved.

DISSIMILAR ANALYSIS CONFIRMATION

Dissimilar analysis confirmation (also called multi-analysis merging) involves verifying the presence of a compound in a sample by showing that the compound has been detected by two or three different methods. This process must also take into account the fact that a sample may be separated into as many as three fractions by Florisil Column Adsorption Chromatography before injection. In such a case, each of the fractions is injected into the chromatograph separately, but all of the fractions are analyzed by the same method.

To perform a dissimilar analysis confirmation, the analyst will be required to enter from the keyboard into the Nova program the ID numbers of the appropriate injections, and the Nova program will simply print a chart similar to that in Figure 5. The Nova program will allow for either two or three methods, and for either no fractioning, two fractions, or three fractions within each method.

The list of compounds will include every compound appearing in any of the methods. Into each position of the chart, the Nova program will print one of the following: 1) the measured concentration, 2) a dash to indicate that the compound was not detected in the injection, or 3) an asterisk to indicate that the method does not apply to the compound.

REPLICATE STATISTICS

To calculate replicate statistics on a sample, the analyst will enter from the keyboard into the Nova program the ID numbers for every replicate injection of the sample. The Nova will print a report showing the average (arithmetic mean) concentration of each compound within the sample.

SAMPLE FILE CONTROL INTERACTION

A user of the Gas Chromatography Automation System will be able to get from the Sample File Control (SFC) system a list of samples which are to be run and will be able to return to the SFC system the measured concentrations of compounds in those samples.

The SFC backlog list will be obtained by the user directly from the SFC PDP-11/70 computer in the form of a lineprinter or terminal printout. After all data processing is complete for the required samples, a module of the Gas Chromatography system will be used to construct an SFC run results file, as defined in the document, "EPA Sample File Control Functional Design," (specifically, Section I - Foreground BASIC SFC/Nova Design). The method/parameter code for the header record of the run results file will be entered manually by the user, who can copy it from the backlog list. The method/parameter codes for the results records will be the CAS numbers of the compounds measured. There will be one result record for each compound measured. The user will have to manually match the SFC identification number for each sample with the corresponding injection ID number. The program will then collect the necessary data from the processed data files and method file in the Nova, and fill the run results file.

The run results file will then be transmitted to the SFC system by other programs which are not part of the Gas Chromatography Automation System.

	Method One			Method Two		
	Fraction A	Fraction B	Fraction C	Fraction A	Fraction B	Fraction C
Compound 1	1.5	-	-	1.6	-	-
Compound 2	*	*	*	2.7	-	-
Compound 3	-	3.2	-	-	3.1	-
Compound 4	-	7.5	-	-	7.1	-
Compound 5	-	-	0.2	-	-	0.2
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
Compound n	-	3.1	0.1	*	*	*

Figure 5. Dissimilar analysis confirmation chart.

SECTION 5

FUNCTIONAL DESIGN

DESCRIPTION OF AN AUTOMATED RUN

In operating this system, there will be no need for the analyst to interact with both the Varian and the Nova at the same time. Rather, it will be possible for the analyst to work with the Varian alone, and then sign on to the Nova to do data processing after all of the necessary injections have been completed.

The Varian will be operated according to its normal procedures with the only exception being that the Varian calibration procedure will not be used. As each injection is processed by the Varian, a report will be printed on the analyst's terminal and also automatically transmitted to the Nova. The analyst's only responsibility to the Nova at this time will be to note the ID number of each injection.

When the analyst signs on to the Nova, there will be a choice of 10 options: 1) sign off from the Nova, 2) generate a method in the Nova, 3) calibrate, 4) plot a calibration curve, 5) process injections, 6) perform quality control calculations, 7) perform dissimilar analysis confirmation, 8) calculate replicate statistics, 9) list the names of all extant data files, or 10) interact with Sample File Control. The analyst will be able to select any one of these options immediately if the prerequisite options have been executed during a previous session. For example, if a Nova method has been generated and a calibration curve has been formed, the analyst will be able to immediately process an injection and evaluate it as a control (check) standard.

Generating a method in the Nova (option 2) will involve entering heading information for reports and the Varian identifiers, full chemical names, and CAS Registry numbers for the compounds covered by the method. Once a method has been generated in the Nova, it will be available for use or for alteration indefinitely.

Processing injections (option 5) will include calculating concentrations, identifying compounds with their full chemical names and CAS Registry numbers, calculating relative retention times or capacity ratios, and printing reports.

All of the other options have been described previously. After any option has been completed, the analyst will again be confronted with the choice of all ten options.

PROGRAM MODULES

There will be 21 programs in this system as shown in Figure 6. Each of the programs will be independent of the others to the extent that one program can be altered or replaced without affecting the others. All data transfer from one module to another will be done via disk files.

FILE DESCRIPTIONS

There will be four types of disk files used by this system. They are described as follows:

Arrival File

TYPE: Binary sequential

NUMBER: Exactly one

CONTENTS: This file will hold the Varian report exactly as it is transmitted.

USES: This file will hold the Varian report when it is received from the Varian. A program will then look through this file for the instrument number and data which are needed to generate the name of data file. The program will then compact the information in this file as it copies it into a data file.

EXISTENCE This file will be created when a report arrives from the Varian, and will exist only until it is overwritten by the next arrival file.

Data File

TYPE: Binary random access

NUMBER: One for every sample or standard

CONTENTS:

1. Before processing:
    a. Varian method number
    b. 30-character title
    c. identifier of each compound
    d. type of peak for each compound
    e. retention time of each compound
    f. area of the peak for each compound
    g. any error messages from the Varian
2. After processing:
    a. all of the above
    b. relative retention time of each compound

c. concentration of each compound

USES: This file will store unprocessed data for any sample or standard. In the case of a calibration standard, this file will provide the information needed by the calibration routine. In the case of an unknown sample or QC sample, this file will provide the data needed to identify the compounds, and to calculate concentrations and relative retention times for the compounds. After being processed, this file will contain the information necessary for doing quality control calculations, dissimilar analysis confirmation, and replicate statistics, and for reporting to SFC.

EXISTENCE: This file will be created from an arrival file. It will exist in the Nova until 100 more samples have been run on the same instrument.

ID Number List File

TYPE: Binary random access

NUMBER: One for each instrument

CONTENTS: This file will contain the ID numbers, in reverse chronological order for the 100 most recent samples run on the corresponding instrument.

USES:

1. When the 101st sample is run on the corresponding instrument, the Nova will delete the oldest file recorded in this list.

2. When the Nova assigns an ID number to a sample, it will check the newest entry in this file to determine what sequence number should be assigned.

3. If an analyst forgets the ID numbers of his samples or standards, he will be able to see a copy of this list.

EXISTENCE: This file will be automatically created the first time the Nova encounters a sample from an instrument. The file will be initialized with 100 dummy entries, and will then gradually fill with real entries. It will exist indefinitely.

Nova Method File

TYPE: Binary random access

NUMBER: One for each method in use in the Varian

CONTENTS:

1. Analyst's name and other heading information.
2. Calibration points and equations.
3. Table of identifiers, true names, and CAS numbers.
4. Identifier of the compound to whose retention time all other retention times are relative.

USES:

1. Headings for all reports.
2. Calibration equations are used in calculating concentrations of unknowns.
3. Calibration points and equations are used in plotting calibration curves.
4. True names of compounds and their CAS numbers are determined using this file.
5. Relative retention times are calculated from the retention time of the compound named in this file.

EXISTENCE: This file is created by an analyst during the Nova method generation procedure. It is altered whenever the analyst recalibrates, and possibly at other times. This file exists indefinitely.

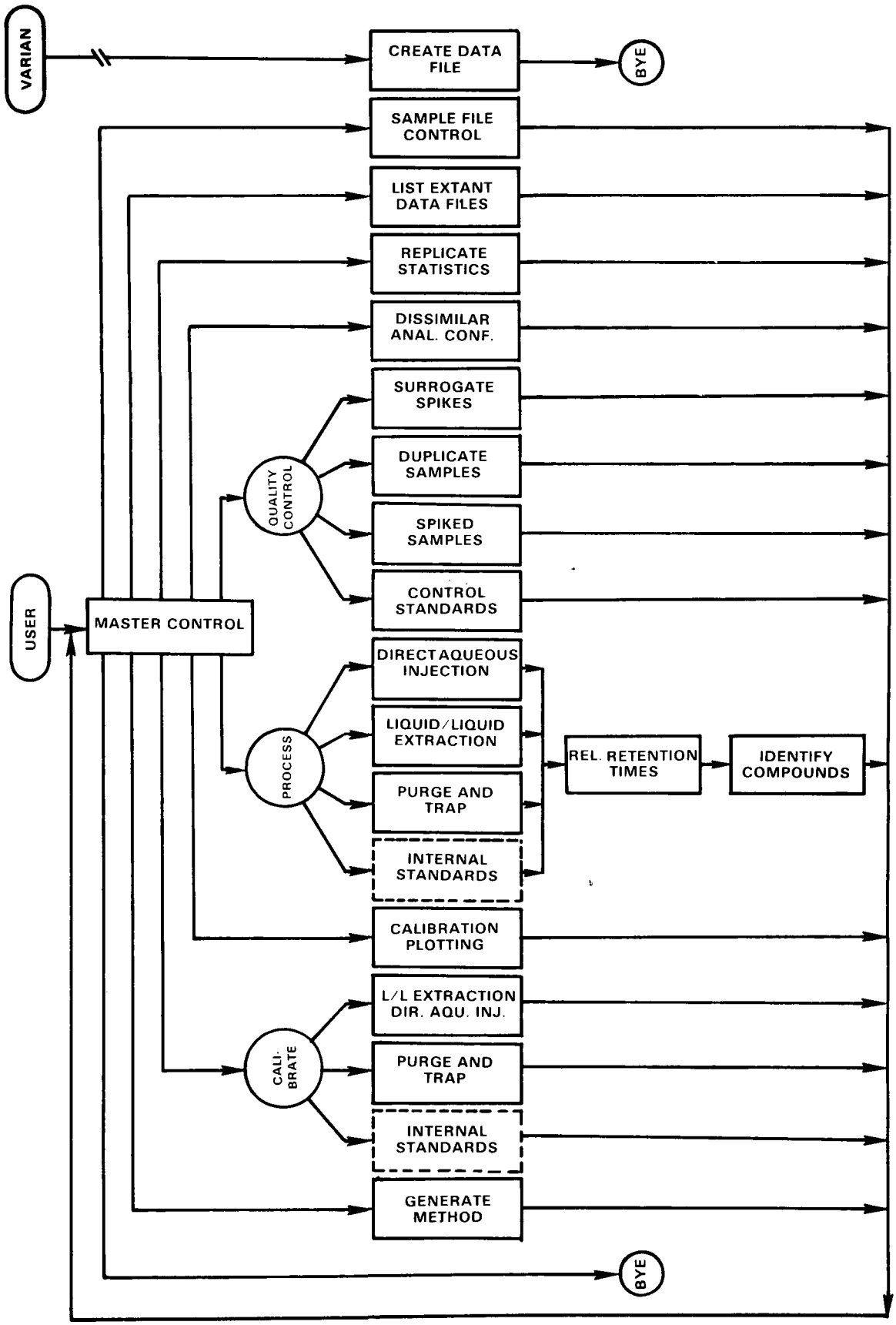


Figure 6. Program modules and their chaining relationships.

SECTION 6

SIGNOFF SHEET

These documents, which describe a proposed automated chromatography data system but do not constitute an implementation design, are approved by the undersigned interested parties with the understanding that changes in detail are likely and that implementation of all features depends upon the availability of funding and personpower.

Requestors: _____ Thomas Bellar

_____ Denis Foerst

_____ Leown Moore

_____ William Potter

Implementor: _____ Jonathan Kopke

Gas Chromatography System Project Officer: _____ John Teuschler

Laboratory Automation Project Officer: _____ William Budde

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before completing)

1. REPORT NO.	2.	3. RECIPIENT'S ACCESSION NO.
EPA-600/4-79-038		

4. TITLE AND SUBTITLE	5. REPORT DATE
Functional Specifications for an Advanced Chromatography Automation System	June 1979 issuing date
	6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S)	8. PERFORMING ORGANIZATION REPORT NO.
Jonathan E. Kopke	
9. PERFORMING ORGANIZATION NAME AND ADDRESS SouthWestern Ohio Regional Computer Center University of Cincinnati Cincinnati, Ohio 45220	10. PROGRAM ELEMENT NO.
	11. CONTRACT/~~GRANT~~ NO. GS-05S-10458
12. SPONSORING AGENCY NAME AND ADDRESS Environmental Monitoring and Support Lab-Cincinnati Office of Research and Development U.S. Environmental Protection Agency Cincinnati, Ohio 45268	13. TYPE OF REPORT AND PERIOD COVERED Extramural
	14. SPONSORING AGENCY CODE EPA/600/06

15. SUPPLEMENTARY NOTES

16. ABSTRACT

This document contains a project definition, a set of functional requirements, and a functional design for a system which will link a commercial chromatography data system to the EPA Laboratory Automation System.

A Varian 220L Chromatography Data System was selected as the prototype system to be extended in this project, although these specifications can be adapted to other commercial systems. The current methods of using the Varian system are briefly described in this report. The bulk of the report is a detailed list of the additional functions to be performed by the EPA Laboratory Automation System. These functions include multi-point calibration, calculation of concentrations, identification of compounds, calculation of relative retention times, and calculation of quality control statistics. A general plan for the proposed system is provided.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
Gas chromatography Calibrating Quality assurance Computers		09/B 14/B 07/C

18. DISTRIBUTION STATEMENT	19. SECURITY CLASS *(This Report)*	21. NO. OF PAGES
Release to Public	Unclassified	35
	20. SECURITY CLASS *(This page)*	22. PRICE
	Unclassified	

☆ U.S. GOVERNMENT PRINTING OFFICE: 1979 — 657-060/5315