

THE APPLICATIONS OF COMPUTER TECHNIQUES IN CHEMICAL RESEARCH

*Proceedings of a conference organized by the Institute of Petroleum
Hydrocarbon Research Group and held at the Midland Hotel, Manchester,
15-17 November 1971.*

Edited by
PETER HEPPLÉ

*Printed in Great Britain by
Adlard & Son Ltd
Bartholomew Press, Dorking*

Contents

	Page
Opening Address H. POWELL	i
Data Acquisition in Mass Spectrometry K. BIEMANN	5
A Low-Cost Data Acquisition System for Use in Gas Chromatographic-Mass Spectrometers J. A. HARDY AND R. I. REED	23
A Simple Data Handling System for Gas Chromatography R. L. ERSKINE AND N. G. McTAGGART	29
Real-Time Peak Detection and Area Apportionment of Unknown Chromatograms A. FOZARD, J. J. FRANCES, AND A. J. WYATT	41
A GLC/Quadrupole MS/Computer System J. BARGERY, J. FARREN, R. HALLETT, J. HAWKE, D. A. SMITH, AND J. W. YOUREN	50
Fourier Transform Spectroscopy RICHARD R. ERNST	61
The Application of Fourier Transformation NMR to the Measurement of Nuclear Relaxation Times D. SHAW	76
A Computer-Controlled Gas Chromatograph-Quadrupole Mass Spectrometer System E. J. BONELLI, S. H. MULLER, AND R. D. SMITH	87
A Fully Automated NMR Spectroscopy System G. MICHEL, H. P. SAUTER, AND A. STUBLI	101
Organization and Programming Techniques in a Petrochemical Laboratory under Real-Time Computer Control C. B. LANCASTER, P. MITCHELL, AND A. R. L. MOSS	112
Information Systems for Chemists D. C. VEAL	129
Computer Techniques at the Mass Spectrometry Data Centre W. M. SCOTT AND R. G. RIDLEY	148
Manipulation of Chemical Structural Formulae on Computers: the Present Position M. A. T. ROGERS	155

The Computer and Chemical Structure Classification: Applications of the Wiswesser Line-Formula Notation	174
E. HYDE AND G. PALMER	
On Machine Intelligence Applied to Chemical Problems	189
THOMAS L. ISENHOUR AND PETER C. JURIS	
Infra-red Band Shapes and their Interpretation	206
R. P. YOUNG	
Numerical Simulation of Physico-chemical Systems	215
L. V. WOODCOCK	
"Monte Carlo" Computer Simulation of Chain Molecules. IV. Equilibrium Shapes of <i>n</i> -Alkane Molecules	234
M. LAL AND D. SPENCER	
The Use of Computers in the Analytical Laboratory	253
J. M. SKINNER	
An Operational Multi-Instrument Computer System	259
A. C. POPPE	
A Time-Sharing Computer System for Laboratory Automation	270
J. G. HALLETT, P. A. LAWSON, D. RICHARDS, AND H. M. STANIER	

Opening Address

By H. POWELL

(Chairman, Institute of Petroleum Hydrocarbon Research Group)

FIRST of all, I would like to welcome you to this conference, which we hope will be both interesting and useful. I am especially pleased to welcome the delegates from overseas.

The conference has been arranged by the Hydrocarbon Research Group of the Institute of Petroleum, a body which originated during the war years to give financial support to academic research on topics of interest to its member companies. It has for most of its time operated through three panels, the Spectroscopic, Mass Spectrometric, and Hydrocarbon Chemistry Panels. The panels are made up of representatives of each of the member companies and members of the universities receiving support from the Group. The technical interests of the Group have expanded considerably during its life, but the strong links between university and industry have been maintained and will, it is hoped, continue into the future.

During recent years there has been growing interest among the members of the Group in computers and computing techniques and their applications to chemical research. The Group has arranged conferences on a wide range of topics in the past, and the subject of "Applications of Computer Techniques in Chemical Research" seemed to be one of wide interest which would provide a useful and timely exchange of ideas between chemists with differing applications for computers.

Some aspects of the subject are quite controversial, and we are hoping that there will be a lively exchange of views on these. I am referring particularly to the conflict of interest within a laboratory when computing requirements are being assessed. Manufacturers have put considerable efforts into the production of laboratory systems, but these have to be limited by cost considerations and must inevitably involve compromises. Some users advocate the use of small dedicated computers rather than large central computers.

The last ten years have seen the growth of computer applications in chemical research. It stemmed from the chemists' realization that computers would be necessary to handle the huge volume of experimental data and information that modern techniques were beginning to generate. High-resolution mass spectrometry was one of the first techniques to experience this need. Without computers it would not have been possible to consider the combination of high-resolution mass spectrometry and gas chromatography. We are very fortunate in having Professor Biemann of MIT, one of the first to make use of computers in the field of mass spectrometry, as our opening invited lecturer.

Computers were also needed for Fourier transform spectroscopy and we welcome Professor Ernst of Zurich as our main speaker on this subject.

X-ray diffraction used computers initially for computation and then developed

towards computer-controlled data acquisition. Other techniques involving a large amount of data, such as gas chromatography, have not been slow to take advantage of the development of computers for the acquisition and processing of data.

Chemical information has been increasing exponentially in recent years, and its storage and retrieval is a growing problem. Computer techniques are again a necessity if we are to overcome this situation.

Interpretation of data is another area in which computers are playing an increasing role. A new concept here is that of "training" the computer to make the correlations on which predictions are based. We very much look forward to Professor Isenhour's presentation on this topic.

Another topic we will be discussing is the use of computers for *ab initio* calculation of physical and chemical properties from basic molecular data.

In a recent survey of computer applications in the chemistry laboratory, Perone, in a paper in *Analytical Chemistry* this year, classifies the applications as passive or active according to whether or not the computer exercises some significant measure of control over the experiment. Data acquisition and processing, file searching and pattern recognition, and display of data or results are passive applications. Active applications include automation using computer control, real-time computer interaction with experiments, iterative optimization of experiments, and computer-user interactive experimentation. The trend is towards an increasing proportion and degree of active applications. This subject will be discussed at tomorrow morning's session.

Several of these potential applications may occur together as the requirements of a particular laboratory and be linked to the general requirements for scheduling work, allocating costs, and reporting results. Laboratory computing systems have been devised to meet such complex requirements; usually, so far, these have been highly individual systems. Today, more and more computer manufacturers and software houses are developing packages which can be combined to meet the specific requirements of individual laboratories; this is indeed a welcome move which should help to keep down the cost of the final system. We will be hearing presentations describing the way in which different laboratories have approached their requirements for a computer system and this will be covered by the final session.

Today's chemist needs to be able to use computers, as with their aid he can do more, and do a better job than without. We must remember, however, that the use of computers is a means to an end, not an end in itself.

We will be reviewing over the next few days the state of the art of computing techniques applied to a wide range of aspects of modern chemistry. We will also be assessing trends in both computer technology and its application to our fields of interest. We hope also to have many contributions to the discussion from the floor. These contributions will be published, together with the papers.

We also have a number of displays and demonstrations in the equipment exhibition, and the manufacturers' representatives are available to help us as well. We hope you will make full use of all these facilities.

DATA ACQUISITION

Chairman: J. BEYNON

Data Acquisition in Mass Spectrometry

By K. BIEMANN

*(Department of Chemistry, Massachusetts Institute of Technology,
Cambridge, Massachusetts 02139, U.S.A.)*

Of the two modes of producing complete high-resolution mass spectra of organic compounds—scanning across a narrow slit in front of an electron multiplier or exposing an ion-sensitive photographic plate placed in the focal plane—the latter is less direct. However, the signal produced by the electro-optical scanning device is more suitable to on-line data processing, if accuracy, reliability, sensitivity, and dynamic range are to be optimized. The second area of computer application is the on-line data acquisition from a gas chromatograph-low resolution mass spectrometer system. It permits the continuous monitoring of the effluent of the gas chromatograph and generates a three-dimensional set of data which allows the application of novel interpretative techniques for the identification of the composition of even very complex mixtures.

The vast amount of data which a modern mass spectrometer can produce within a short time has, over the past few years, forced the chemist to rely more and more on computers to ensure the full and efficient exploitation of these data. The techniques fall into three general categories: the high speed by which mass spectra are recorded in certain situations requires on-line data acquisition; the large volume of raw data (mainly time-intensity pairs) requires fast and efficient processing (to mass-intensity pairs and, in the case of high-resolution data, further to elemental composition-abundance pairs); and, finally, interpretative methods to identify the compound giving rise to the spectrum or to aid the chemist in this interpretation.

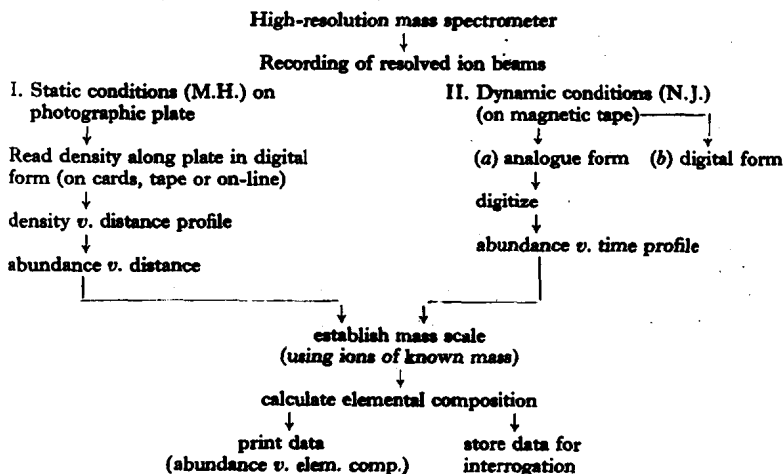
RECORDING OF HIGH-RESOLUTION MASS SPECTRA

Historically, the first instance of at least semi-automatic data acquisition and full-scale processing of the resulting data on a large computer was triggered by the desire to fully exploit the output of a high-resolution mass spectrometer used in the determination of structures of complex organic molecules.¹ The accurate measurement of the position of a few hundred lines representing a spectrum recorded on a photographic plate using a Mattauch-Herzog type double-focusing mass spectrometer by far exceeded the patience of a densitometer operator and led to the design of a semi-automated densitometer with punch-card output. Further evolution went through the stages of digital tape recording² and, finally, on-line operation of the densitometer.^{3, 4} Similar techniques were developed for fast scanning spectrometers of the Nier-Johnson geometry, which do not possess a focal plane. These instruments thus cannot make use of a photographic plate and must record on-line into the computer.^{5, 6} The primary data obtained in either case is a parameter related to the mass of a particular ion giving rise to the signal, as well as its abundance. This parameter distance along the photoplate in one case and time during the scan of the

magnetic field in the other. Scheme 1 outlines the various steps involved in the generation of a high-resolution mass spectrum in terms of the elemental composition of the ions.

The on-line approach is, of course, the most direct and certainly most fashionable technique, but it has a number of practical disadvantages. First, the data acquisition system, which has to be a reasonably sophisticated one to be able to handle a complete high-resolution spectrum of a compound of the size and complexity normally encountered in natural products, must be available and operated while the sample is introduced into the mass spectrometer under the appropriate conditions. Secondly, it must have the wide dynamic range which is required in a situation where one simultaneously records the mass spectra of two substances, namely, the one under consideration and the mass standard, both of which produce ions of widely varying abundance but equal significance. Thirdly, ions of low abundance when recorded at the relatively high scan rates and resolution practically necessary will give a signal that does not correspond to a smooth Gaussian envelope but consists of a few spikes due to the occasional arrival of an ion while the very weak beam is scanned across the collector slit. Nevertheless, both relatively simple as well as more elaborate on-line computer systems have been developed⁶ that deal with this situation adequately as long as one is satisfied with a compromise in scan speed, accuracy, sensitivity, and dynamic range optimized for the problem at hand.

SCHEME 1



For reasons that are mainly related to our need for obtaining reliable high-resolution data over a wide dynamic range and a wide mass range on compounds of which often very little is known and whose thermal behaviour and volatility is not established prior to the experiment, we prefer photographic recordings of the high-resolution mass spectrum and its later measurement using an automatic densitometer, a technique which has been refined considerably in recent years.

Its main practical advantage is, first, that the spectrum can be recorded without paying any attention to the data system while the sample is vaporized into the ion source. Secondly, since the spectrum is integrated during the entire exposure time over the entire mass range, fluctuations in sample pressure are of no consequence. Thirdly, the technique is independent of the statistical variation of the number of ions arriving at a given time in the focal plane because of the integrating properties of the photographic system. Fourthly, additional smoothing of the data results in the photometric system which employs a relatively long but very narrow slit and a relatively slow scan rate compared with the scan rate required for real-time recording of a mass spectrum.

Most important, by taking consecutive exposures during the vaporization of the sample into the mass spectrometer, one obtains a series of line spectra on the photographic plate which can be easily inspected using a low-power microscope. This visual inspection permits one at a glance to select that spectrum during which the compound in question vaporized at the most appropriate conditions, to judge the quality of the spectrum, to deduce whether any fractionation from impurities took place or whether thermal decomposition occurred. Thus the selection of the particular spectrum to be actually measured does not involve any computer processing of the data. Most important is, however, the smoothness of the signal produced when scanning the optical density profile along the spectrum with the automatic densitometer, which thus makes the processing of the data in terms of exact line centres of even weak ion signals, as well as the further resolution of unresolved multiplets, a much simpler task.

A schematic block diagram of the densitometer operated on-line with the computer is shown in Fig 1.³ In brief, the table, on which the photographic plate containing up to 45 mass spectra rests, is driven by a motor and a precision screw on the axis of which is mounted a pulse generator which produces 2000

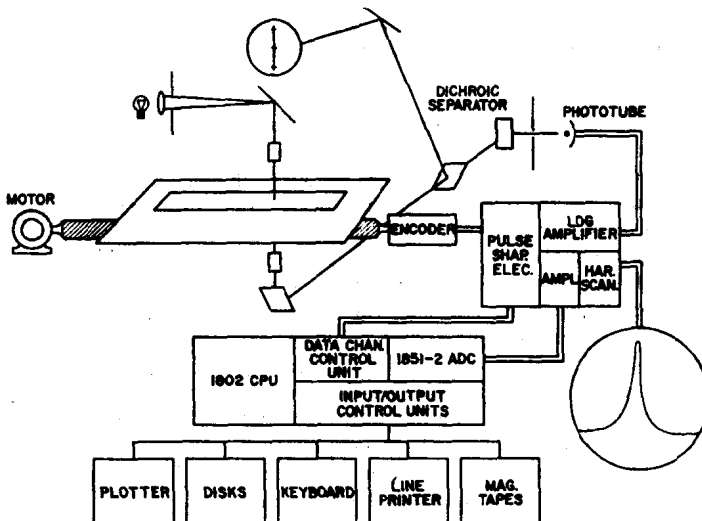


Fig 1. Schematic representation of an automatic densitometer

pulses per revolution, each pulse corresponding to a motion of the table by 0.5 microns. Each pulse triggers the computer to read and digitize the output of the phototube, which is mounted behind the slit of the optical system. Fig 2 shows a plot of this signal in the molecular ion region of methyl myristate. The entire spectrum was recorded in the mass spectrometer from mass 20 to 600 by exposing that section of the plate for 3 sec and read by the densitometer in 5 minutes. Both the monoisotopic molecular ion at mass 242, as well as the isotope peaks at 243 and 244, are clearly visible.

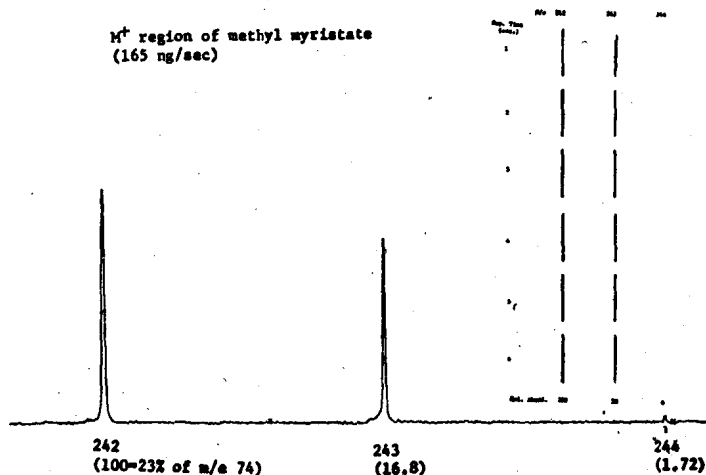


FIG 2. Plot of density of a 3-sec exposure of methyl myristate (region of m/e 242-244 only shown). Upper right: consecutive exposures (1, 2, 3, 4, 5, and 6 sec).

The profile of the signal at mass 243 and 244 is shown in Fig 3 in expanded form, each x-mark corresponding to a data point read at 0.5 micron intervals. Note the very smooth signal for the ^{12}C isotope peak while the ion beam at mass 244 is partly resolved due to the occurrence of molecular ions containing two ^{13}C atoms or one ^{18}O , respectively. Because of the much lower density of the lines at mass 244, they are plotted at an expanded scale. Of the two horizontal lines, the lower one corresponds to the background level which is computed anew for each one-millimetre segment along the plate, while the upper line corresponds to the threshold setting, which is calculated by the computer on the basis of the standard deviation of the data points in regions free of peaks. From the data shown in Fig 3 it is clear that unresolved multiplets of components differing by only a fraction of a line width can easily be detected when applying various simple parameters, such as monitoring of the line width at various fractions of total peak height or the disagreement in line position applying differing methods for its calculation such as centroid v. peak top.

Most important, however, is the wide dynamic range over which lines can be detected and their position accurately determined. The data in Fig 2 show a

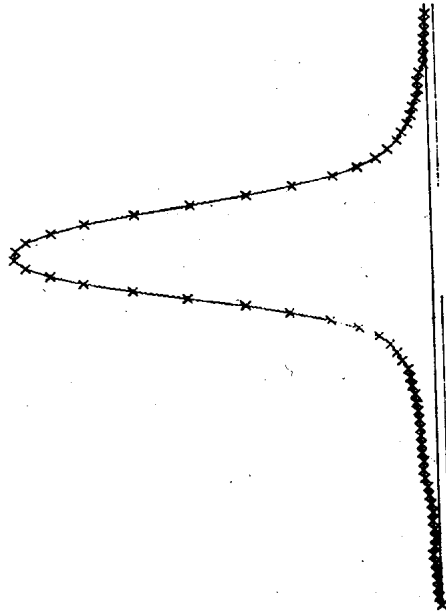
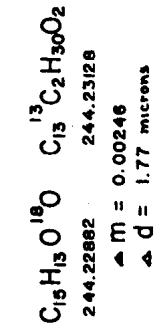
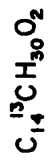


FIG 3. Expanded plot of m/e 243 and 244 of Fig 2.

dynamic range of 1 in 1000 based on the peak due to the ^{18}O isotope line of the molecular ion v , the most abundant peak in the spectrum, mass 74. This dynamic range becomes particularly important for mass measurement at high mass (mass 1000 and beyond) where the abundance of ions due to the mass standard are sometimes extremely low.

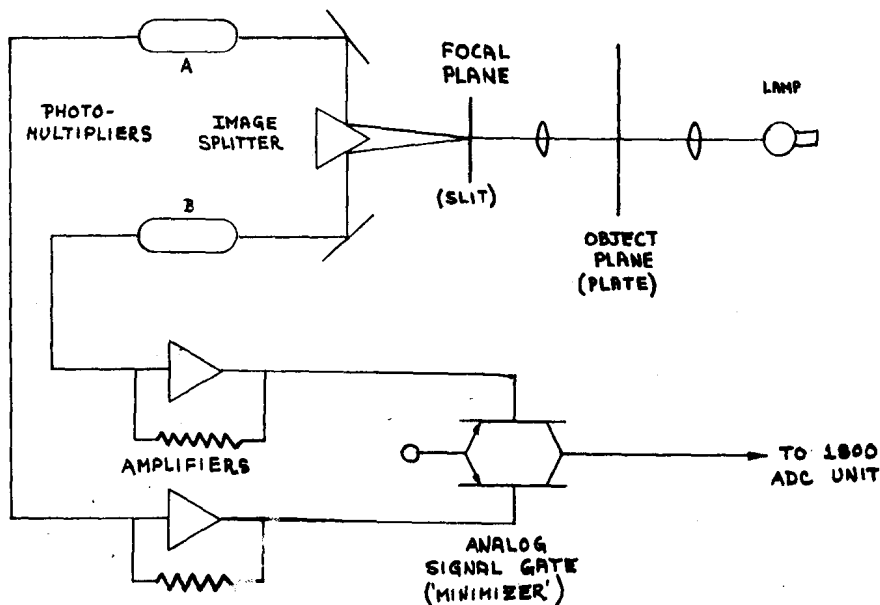


FIG. 4. Schematic of dual beam system for the minimization of optical imperfections on photographic plates read on the automated comparator shown in Fig 1.

In order to indeed utilize the wide dynamic range and capability of measuring very weak lines, as well as treating unresolved multiplets on the basis of their unsymmetrical line shape, it became important to eliminate the influence of any minor imperfection in the optical quality of the photographic plate which otherwise would lead to spurious line centres and thus to incorrect elemental compositions or unassignable lines. For this purpose, the light beam passing through the projection slit is split into two components, one representing the upper half of the line, the other the lower half. Both light beams are deflected on one of a pair of matched photo-multiplier tubes, the signals amplified separately, and then compared by an appropriate circuit ("minimizer") which transmits only that one which corresponds to the larger amount of light falling on the phototube (Fig 4). Thus any imperfection, such as a dust corn or scratch that occurs only in one half of the spectrum, will not be noticed (Fig 5). For a long scratch the system will always read that side of the light beam which is not obscured, assuming that the scratch is not exactly parallel to the lines of the spectrum, a situation which is very rarely encountered. This technique permits one to

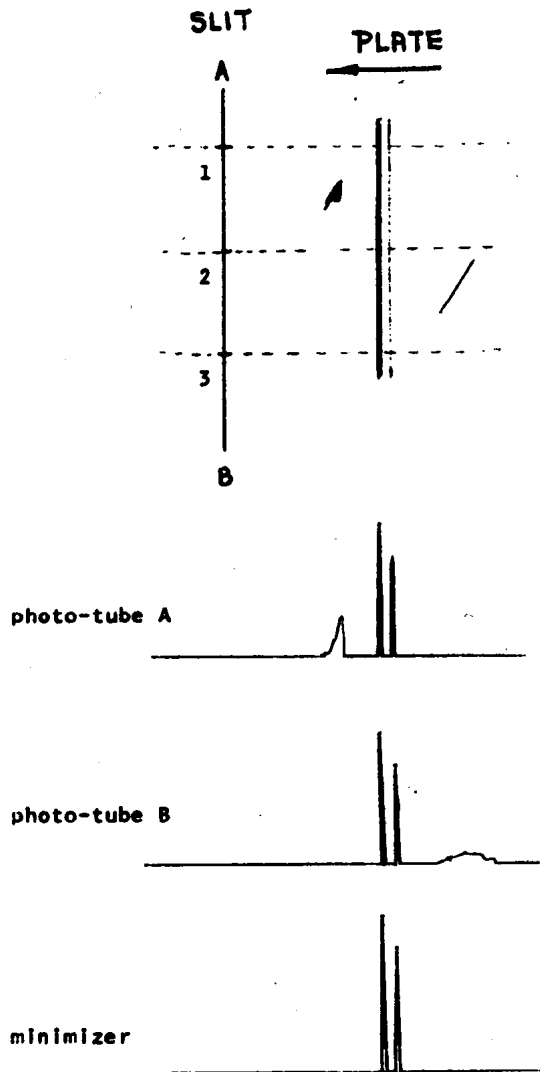


FIG 5. Principle of the elimination of optical imperfection from the output data (for details, see text).

select a threshold for processing of the data that lies very close to the background level, and thus utilizes sufficient data points for accurate line centre calculation even in the case of very weak lines.⁷

It should be noted at this point that all the results described above were obtained with evaporated silver bromide plates,⁸ rather than with the silver bromide emulsion plates (Ilford Q2) used in earlier work. The absence of a gelatin emulsion increases the dimensional stability of the light-sensitive layer, which is only a fraction of a micron thick, and vastly increases its sensitivity towards ions of high mass, a problem, which previously presented a serious handicap in the utilization of photographic plates for the mass spectrometry of large molecules. Indeed, accurate mass measurements of ions covering a wide dynamic range at mass 1640 to 1653 have been reported elsewhere, utilizing this system.⁸

CONTINUOUSLY SCANNED LOW-RESOLUTION MASS SPECTRA OF GAS CHROMATOGRAPHIC EFFLUENTS

A second area in which on-line data acquisition has become an important technique is the recording of the output of a mass spectrometer that is directly coupled to a gas chromatograph in order to identify the components of complex mixtures. In this case, the vast majority of the problems do not require a high resolution mass spectrum, but rather a conventional one that is scanned sufficiently fast to produce a complete spectrum during a fraction of the time over which the gas chromatographic fraction emerges. Unless capillary columns of high resolving power are used, a scan time of 2 to 4 sec suffices. The data problem comes into play when one considers that a mixture of compounds isolated from natural sources may easily contain 20, 50, or 200 components and that those will not be equally well resolved and thus give rise to gas chromatographic peaks which encompass not only one but a number of individual components. For this reason, one generally wishes to scan the mass spectrometer very often during the gas chromatogram. The general practice is to initiate the scan at the top of each gas chromatographic peak in order to obtain a mass spectrum of maximum intensity and to scan again at discernible shoulders in an effort to record also unresolved components. This method requires considerable information concerning the compounds present and prior knowledge of the retention times of the fractions are of interest.

However, once one has succeeded in coupling the gas chromatograph to a mass spectrometer, and the latter in turn to a computer, there is indeed no need for human intervention for the purpose of scan selection. Obviously, one could conceive of an algorithm or a hard-wired interface which would trigger the mass spectrometer to scan whenever a gas chromatographic peak top or shoulder is encountered, but it is much simpler to eliminate this human or electro-mechanical decision-making process and simply record all the spectra produced by a continuously scanning mass spectrometer. This is even more so because peak tops and peak shoulders may not at all be the most appropriate portions of the gas chromatogram to be scanned if not all components are completely resolved from each other. The peak top of the signal representing two fractions of equal abundance but resolved by less than one peak width at half height, results in a mass spectrum that corresponds to an equal mixture of the two components. On the other hand,

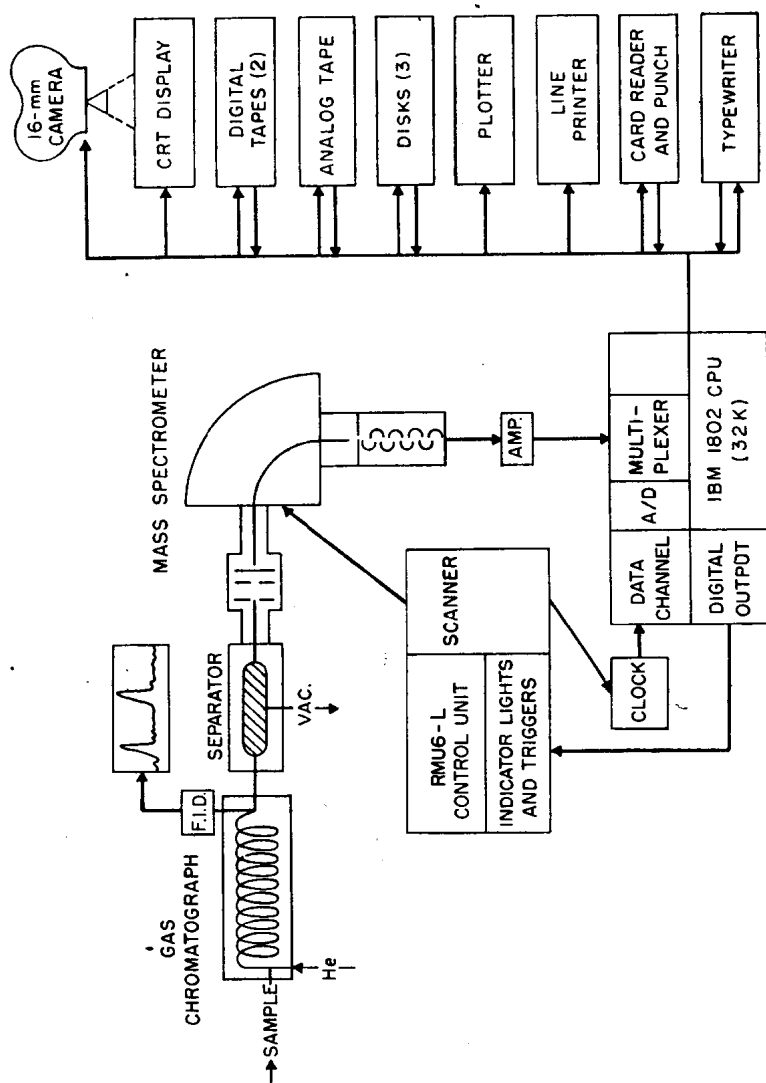


FIG 6. Block diagram of a gas chromatograph-mass spectrometer-computer system with ancillary input-output devices (the computer is physically the same as that shown in Fig 1).