Computer Processing of Gas Chromatographic Data: Quantitative Analysis of Fatty Acid Methyl Esters Using an Off Line Magnetic Tape System[1]

GARY J. NELSON, Bio-Medical Division, Lawrence Radiation Laboratory, University of California, Livermore, California 94550

ABSTRACT

This paper describes the off line processing of chromatographic data from fatty acid methyl ester mixtures by a large digital computer. The output of the gas chromatograph is recorded and stored on magnetic tape; the tape can then be processed at any convenient time. The system takes the analog signal from the electrometer and converts it to digital form, adds a time base and records the data using an incremental magnetic tape recorder. The data sampling rate can be varied from 0.05 to 5 times per second. The total number of samples per tape is limited only by the length of the tape. A 1200 ft reel can store up to 30 hr of gas chromatographic runs. The tape is processed by a CDC-6600 computer. The computer locates all peaks, determines the retention time and calculates the mass percentage of each peak. Along with the numerical output, a cathode ray tube plot of the data is also generated by the computer. The program has provisions for baseline corrections and peak overlap. The program also incorporates corrections for detector nonlinearity and variation in detector response to different molecular species. The accuracy of the system is within 5% based on analysis of fatty acid methyl ester samples of known composition. Computation time for a 1200 ft reel of tape is approximately 20 min.

[1]One of seven papers to be published from the Symposium 'Computer Systems and Applications in the Oil and Fat Industry,' presented at the ISF-AOCS Congress, Chicago, September 1970.

FIG. 1. Photograph of digital data station. The voltmeter is on the bottom of the rack, the clock scaler is on the right of the second row, and the master control module is next to it. The tape coupler is mounted below the incremental tape recorder. The indicative data are entered on the tape by the thumb wheel switches in the center of the tape coupler unit.

INTRODUCTION

The use of computers to process gas chromatographic data is becoming common practice (1-5). Most current systems (2,3,5) preprocess the data output of the gas chromatograph in some way, e.g., with an electronic digital integrator, before feeding it to the large computer for final processing. In many cases this is not a disadvantage and the programming for the computer is greatly simplified. Another approach in which the data from the gas chromatograph is processed directly by the computer utilizes time-sharing (2,4), in which several instruments are connected to a single computer, and the data from any one instrument is sampled at a predetermined rate. In these systems, usually highly expensive, the data is processed in much the same way as by the electronic integrator, because of limitations in the memory capacity of the computer.

An alternative method, discussed in this presentation, is to record the output signal of one or more gas chromatographs on magnetic tape off line from the computer. The tape is then processed batchwise, taking advantage of the full capability of the computer. This approach allows the computer to perform more extensive data-processing, but alternatively involves detailed computer programming tailored specifically to the particular analysis and gas chromatograph in the system.

The system described here is a dual channel system that records the output of one or two gas chromatographs digitally on magnetic tape. No preprocessing of the data is done. The computer, a CDC-6600, locates all peaks and calculates the mass percentage of all components present in

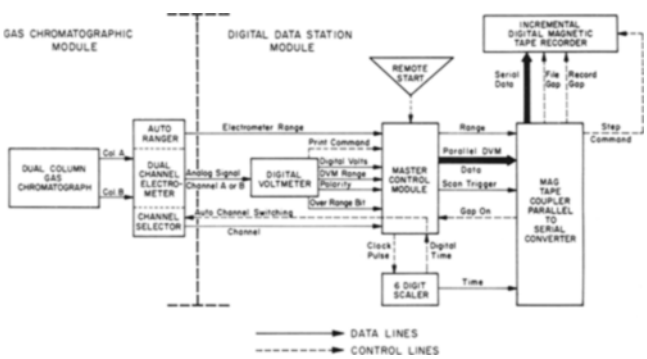

FIG. 2. Block diagram of digital data station. The vertical dotted line indicates which parts are physically located in either the gas chromatograph or the digital data station rack.

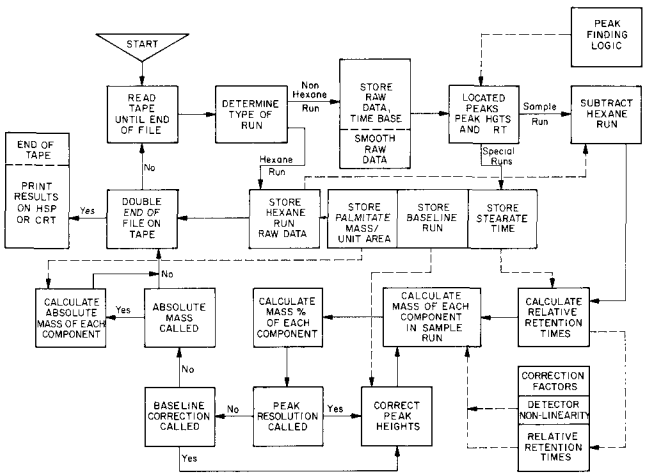

FIG. 3. The computer program flow diagram. Solid lines: main logic pattern and direction of data-processing. Dotted lines: temporary storage and subroutines.

the sample, and yields a numerical printout and cathode ray tube (CRT) plot of the data.

DESCRIPTION OF HARDWARE

The digital data station shown in Figure 1 is a free-standing unit mounted in a standard 19 in. electronics rack, connected to the gas chromatograph by cables only. The unit consists of five components:

1. A digital voltmeter (DVM) that digitizes the analog signal from the electrometer of the gas chromatograph.

2. A digital clock scaler that provides a time base for the chromatographic run and the timing pulse which controls the rate at which the data station samples the data from the DVM.

3. A master control module that integrates the operation of the DVM and clock scaler. The length of the run is selected by a counter on this module, as is also the sampling rate at which the data are placed on the tape.

TABLE I

Reproducibility of Multiple Injections of 10 μg Samples of Methyl Palmitate[a]

Injection no.	Elution time, min	Peak height, in.
1	7.29	6.34
2	7.24	6.37
3	7.24	6.30
4	7.23	6.29
5	7.27	6.40
6	7.34	6.16
7	7.40	6.20
8	7.32	6.19
9	7.32	6.21
10	7.38	6.25
Average	7.305	6.271
± S.D.	± 0.057	± 0.078

[a] 0.005 ml of hexane, containing 10.03 μg methyl palmitate.

4. A parallel-to-serial tape coupler that controls the operation of the tape recorder and also allows indicative data such as run number and other information to be inserted on the tape manually at the start of the run.

5. The incremental magnetic tape recorder, which holds 1200 ft reels of IBM compatible ½ in. magnetic recording tape; reels of this size are sufficient to record up to 32 hr of gas chromatograph runs using a sampling rate of 2.5 times per second.

Figure 2 is a block diagram of the digital data station as it is connected to the gas chromatograph. The electrometer in the gas chromatograph was designed in-house to provide autoranging and dual channel operation, and hence is an integral part of this system. However, this is an optional feature and is not absolutely necessary in this type of system. The data station as it stands could be connected to any gas chromatograph and operated essentially as described here for single channel recording. For this reason the electrometer will not be discussed further. (Details of the construction and operation of the electrometer are available from the author on request.)

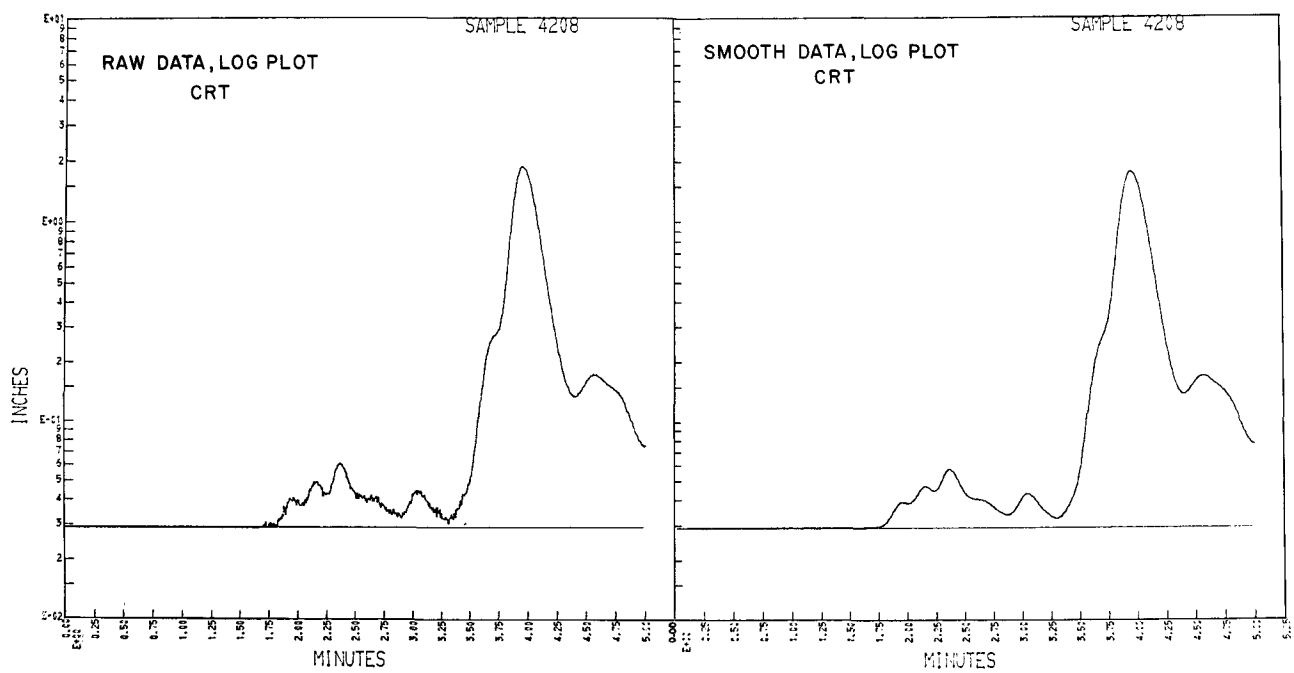

FIG. 4. Examples of raw and smoothed data from the CRT plot of a typical fatty acid methyl ester sample. The ordinate is a log plot, which tends to suppress noise at high levels and emphasize noise at low levels. Actual noise level is constant regardless of signal magnitude. The straight line is a computer-calculated baseline. The first 2 min of the CRT are blank because the computer has suppressed the hexane peak. Note that smoothing routines do not noticeably alter the shape of the curve or the peak heights.

TABLE II

Comparison of Retention Times in Minutes of Fatty Acid Methyl Esters Determined Manually (M) and by Computer Processing (C)

Fatty acid methyl ester	Run 4130		Run 4131		Run 4137		Average	
	C	M	C	M	C	M	C	M
14:0	4.35	4.33	4.33	4.30	4.23	4.20	4.30	4.29
16:0	7.63	7.61	7.59	7.57	7.39	7.34	7.54	7.51
16:1	8.95	8.94	8.90	8.90	8.70	8.66	8.85	8.83
18:0	13.73	13.72	13.61	13.60	13.26	13.21	13.53	13.51
18:1	15.71	15.70	15.56	15.54	15.18	15.13	15.48	15.46

For simplicity, the operation of the data station will be described in the single channel mode.

The analog signal is digitized by the DVM at a rate of 10 times per second to five significant figures. The maximum sensitivity of the DVM is 10 μvolts. The digital signal along with its polarity is sampled by the control module normally every 0.4 sec (the maximum rate is every 0.2 sec). The data from the DVM and clock scaler, which are in parallel form, are converted to a serial form by the tape coupler, which allows the data to be transferred to the tape recorder in a suitable binary converted decimal (BCD) format.

In this system a maximum of 12 characters per word is recorded on the tape. A word can be as short as six characters if the time base is ignored. In practice we use a 10-character word, retaining the four least significant figures and discarding the two most significant. The computer can count words to provide a time base, provided the recording rate is constant, which it is in this system. The first five characters are the data from the DVM, while the sixth character provides the information as to the polarity of the signal, which electrometer channel is recording, and what range the electrometer is set on.

The digital voltmeter, clock scaler, tape coupler and tape recorder were commercial units incorporated into the system essentially without modification. The master control unit was designed as built in this laboratory, as was the electrometer.

At the start of a run, a single indicative data word is inserted on the tape. The run is given a number and is identified as to type. Data processing options are also selected and a baseline tie point given. A record gap is placed on the tape manually after the indicative data word. The injection of the sample into the gas chromatograph is made, and simultaneously the operator presses a switch to start recording the signal from the electrometer. The data station then records continuously for a predetermined period. The length of the recording is determined by the operator by a record counter on the control module. For convenience, we use a record containing 1000 words. When recording at a sampling rate of 2.5 times per second, a record is 6.67 min long; hence, the total time of a run must be a multiple of 6.67 min.

At the end of each sample run, the operator manually inserts a file gap on the tape. Before a tape is demounted from the data station for processing by the computer, two file gaps are inserted after the last sample run. The double file gap tells the computer that the tape is finished and to rewind the tape.

Details of the construction and assembly of this system, including complete schematic diagrams, are available from the author on request.

DESCRIPTION OF SOFTWARE

Figure 3 is a flow diagram of the computer program. The actual program is written in Fortran IV and a listing is available from the author on request. The tape is read one sample at a time, and each sample is processed completely before the next run is read by the computer. The first four runs on each tape are special runs; a hexane blank, a pure methyl stearate sample, a pure methyl palmitate sample and, if available, a baseline blank. The first two special runs

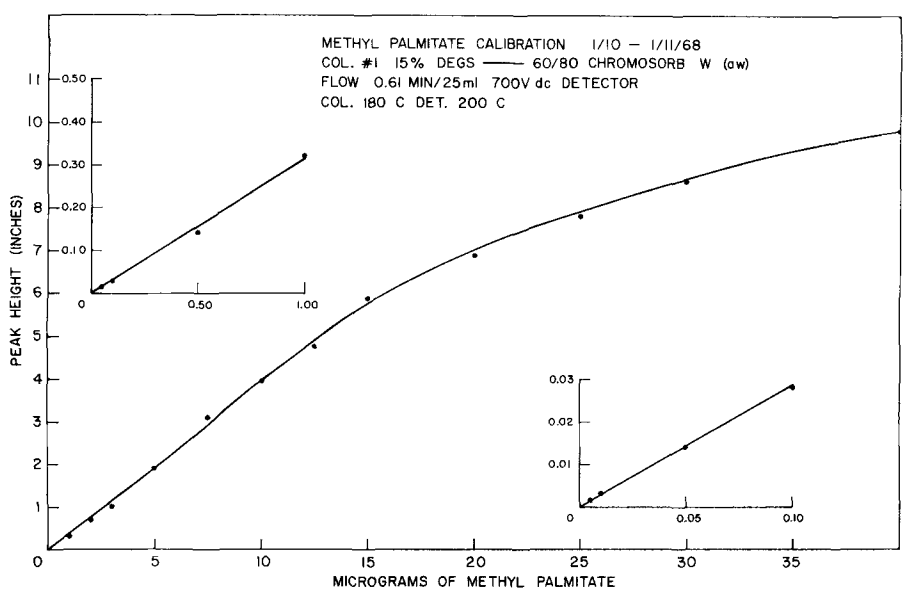

FIG. 5. Typical detector calibration curve for the Sr^{90}-argon ionization detector. The experimental points shown are average values of three injections. The solid curve is the best fit through the points as calculated by the computer. The inserts show the extension of the calibration curve to submicrogram amounts of methyl palmitate on a magnified scale.

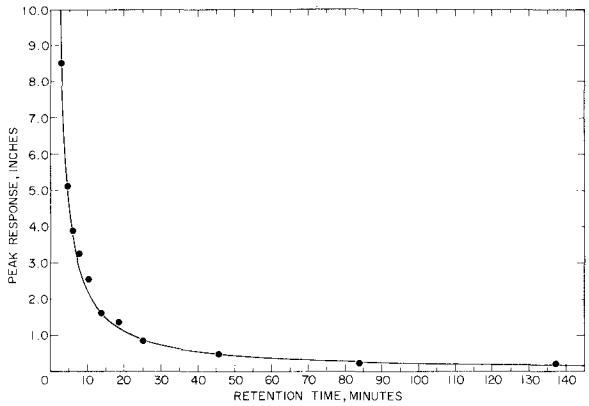

FIG. 6. Maximum detector response versus absolute retention time for pure fatty-acid methyl-esters of different chain length, C-12 through C-26, for a 10 μg sample. Dots: experimental points. Solid curve: best fit hyperbola through these points. This figure shows that in isothermal operation the maximum detector response to any sample is inversely proportional to its retention time on the column. See the text for the application of this fact to the calculation of the individual mass of any component in a sample.

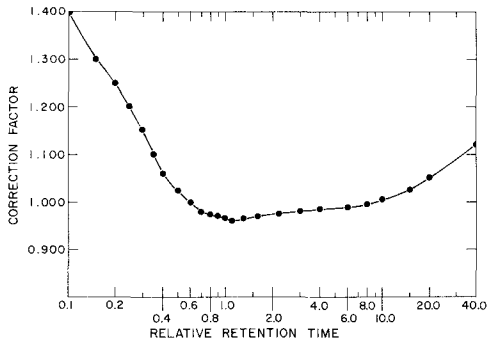

FIG. 7. Second correction factor applied to peak heights on the basis of relative retention time. Experimental points determined by trial and error on computer runs using standard fatty acid mixtures of known composition. Solid curve: a computer-calculated best fit through the experimental points.

are mandatory, while the methyl palmitate and baseline blanks are optional. The purposes of these special runs are listed below.

1. The hexane blank: All samples are injected into the gas chromatograph in 5 μl of pure hexane. This causes an initial large hexane peak at the beginning of each run. Thus the computer stores this information and subtracts the tail of the hexane peak from all subsequent sample runs.

2. The stearate run: This run allows the computer to locate the absolute retention time of methyl stearate. It uses this information to locate the stearate peaks in subsequent runs and to calculate the relative retention times (RRT) of all sample peaks compared to methyl stearate. If no peak corresponding to methyl stearate is found in a sample run, the retention time from the methyl stearate run is artibrarily assigned to calculate the relative retention time for the peaks in that sample. The RRT is essential because the second correction factor (discussed below) is applied on the basis of the relative retention time of each peak.

3. The palmitate run: This run is optional. It is used to find the absolute mass of each peak, in micrograms, on the basis of the area under the curve for that peak. A known mass of methyl palmitate is injected and the computer calculates a mass-unit area. This information is stored in memory for use in subsequent sample runs.

4. The baseline blank: This is also an optional run. A sample blank is carried through the entire sample preparation and injected in a minimum volume (5 μl) of hexane into the gas chromatograph. Any peaks present in the blank are located by the computer. Their peak heights and retention times are stored in memory and subsequently subtracted from any peaks found in samples that have corresponding relative retention times.

Special runs 3 and 4 are not used in the program calculation unless they are called in the indicative data word.

The main body of the computer program consists of the curve-smoothing and peak-finding logic. In this program the raw data from the samples are fed into the core memory of the computer. Three files are created; the first contains the DVM readings, the second the time of each reading, and the third, the smoothed DVM data. The third file is obviously created from the first. A typical sample run may have 30,000 data points, hence for this program a computer with over 90,000 words of core memory is required. Certain modifications can be made to reduce the amount of core needed to 35 or 40,000 words for most runs, however.

Two algorithms are used to smooth the raw data. The first algorithm is a wild point rejection. The computers find the standard deviation, σ, of a best fit parabola for 25 data points. The midpoint (13th point of raw data) is rejected if it falls outside of 0.9 σ. The data block is then incremented by one data point, the parabola and σ recalculated and the midpoint tested. Rejected points only are replaced by the best fit point from the computed parabola. The process is continued for all data points in the run. Then the process is repeated in a second pass through the data.

After the second pass of the wild point rejection all the data points are smoothed by a similar algorithm which finds the best fit, 25 point parabola and replaces all 25 raw data points by the corresponding points from the best fit parabola. Three passes are made with this algorithm. These routines were selected to minimize the disturbance to the shape of the curve. Figure 4 shows an example of the smoothing of a typical sample run. Note that the CRT presentation is logarithmic on the ordinate; this exaggerates the noise at low levels and depresses noise at higher levels. This figure also illustrates the effect of subtraction of the hexane peak from the early part of the curve.

Peak-finding is done only on the smoothed data. The inflection points are located by simply determining whether a point on the curve is greater than either of its neighbors. This point is assumed to be a peak unless it is eliminated by the following three criteria:

1. The point must have an arbitrary value, Δ, above the baseline; Δ is usually set to reject a peak which would contribute less than 0.01% of the mass in a sample.

2. The computer looks at the nearest neighbor peaks and keeps only the highest peaks if peaks are too close. The criterion here is based on the relationship between retention time and the gaussian shape of the peaks. An equation is set up which informs the computer if two peaks with retention times, t_1 and $t_1 + \Delta t$, can be resolved by the particular column and gas chromatograph in use.

TABLE III

Comparison of Peak Heights (In.) of Fatty Acid Methyl Esters Determined Manually (M) and by Computer Processing (C)

Fatty acid methyl ester	Run 4130		Run 4131		Run 4137	
	C	M	C	M	C	M
14:0	4.14	4.19	3.92	4.00	0.032	0.032
16:0	5.08	5.08	4.85	4.88	0.040	0.040
16:1	0.93	0.96	0.91	0.93	0.009	0.009
18:0	3.05	3.06	2.93	2.95	0.028	0.028
18:1	3.36	3.37	3.24	3.24	0.026	0.026

TABLE IV

Comparison of Retention Times (Min) and Peak Heights (In.) of Fatty Acid Methyl Esters Determined Manually (M) and by Computer (C) Processing

Fatty acid methyl ester	Retention times				Peak heights			
	Run 4136		Run 4141		Run 4136		Run 4141	
	C	M	C	M	C	M	C	M
14:0	4.37	4.32	4.30	4.14	1.79	1.82	0.172	0.172
16:0	7.66	7.60	7.52	7.43	1.95	1.95	0.194	0.193
18:0	13.79	13.73	13.46	13.37	1.89	1.91	0.199	0.198
20:0	25.21	25.19	24.51	24.45	2.31	2.30	0.240	0.238
22:0	46.83	46.79	45.11	45.01	2.09	2.10	0.215	0.210
24:0	89.00	88.93	84.46	84.35	1.60	1.62	0.181	0.182

TABLE V

Accuracy of Gas Liquid Chromatography of Computer Data Reduction System on Standard Fatty Acid Mixtures (NHI Mixture D)

Fatty acid methyl esters	Relative retention times	Mass per cent of each component							
		4244[a]	4245	4246	4247	4248	4249	4250	4251
14:0	0.32	11.97	12.06	11.44	11.45	11.34	11.41	11.32	11.56
16:0	0.56	23.25	23.63	23.26	23.95	23.41	23.41	23.26	23.04
16:1	0.65	7.11	7.12	6.96	6.94	6.72	6.80	6.54	7.48
18:0	1.00	13.56	13.74	13.38	13.42	13.29	13.27	13.08	13.95
18:1	1.14	44.10	43.45	44.95	44.90	45.24	45.10	45.81	43.98
Total,[b] μg		79.82	82.67	42.67	43.63	8.93	9.73	0.82	0.88

[a]Run number. [b]Total mass injected.

```
1                               NO BASELINE CORRECTION USED

 10-26- 8-70
 SET NUMBER=.   4208              BHT CORRECTION....0=NO BHT CORR.
 BHT=    0                                          1=BHT CORR.

              NSET     ALMP     U        V  TC PR BC  DIFF
 CONTROL      4208    10.00   5.0     100.  1  1  0   0.98
                                   REL                  PCT        PCT        PCT
   TIME      PH       BASE        TIME       MASS      MASS      MASS2      MASS3      ABM1       ABM2       ABM3
   2.38  0.03066  0.02901        0.191      0.041      0.04       0.04       0.         0.03       0.03       0.
   3.05  0.31501  0.02897        0.245      0.025      0.02       0.02       0.         0.02       0.02       0.
   3.95  1.79712  0.02891        0.319      2.839      2.66       2.66       0.         1.75       1.75       0.
   4.57  0.14058  0.02887        0.368      0.301      0.28       0.28       0.         0.19       0.19       0.
   5.19  0.07511  0.02883        0.418      0.181      0.17       0.17       0.         0.11       0.11       0.
   6.08  0.05950  0.02877        0.489      0.178      0.17       0.17       0.         0.11       0.11       0.
   6.90  0.36454  0.02872        0.555      1.125      1.06       1.06       0.         0.70       0.70       0.
   8.08  3.93240  0.02865        0.650     11.019     10.33      10.33       0.         6.81       6.81       0.
   9.46  0.19601  0.02856        0.761      0.835      0.78       0.78       0.         0.52       0.52       0.
  10.50  0.22617  0.02849        0.844      1.063      1.00       1.00       0.         0.66       0.66       0.
 -12.43  0.07065  0.02837        1.000      0.408      0.38       0.38       0.         0.25       0.25       0.
  14.15  4.13735  0.02825        1.138     19.938     18.70      18.70       0.        12.32      12.32       0.
  15.94  0.07975  0.02815        1.282      0.575      0.54       0.54       0.         0.36       0.36       0.
  17.43  0.25884  0.02805        1.402      2.013      1.89       1.89       0.         1.24       1.24       0.
  20.15  0.02415  0.02788        1.620      0.220      0.21       0.21       0.         0.14       0.14       0.
  22.95  0.09440  0.02770        1.846      0.988      0.93       0.93       0.         0.61       0.61       0.
  25.27  1.19482  0.02755        2.032     12.187     11.43      11.43       0.         7.53       7.53       0.
  27.10  0.18819  0.02744        2.180      2.317      2.17       2.17       0.         1.43       1.43       0.
  31.29  0.02580  0.02717        2.517      0.376      0.35       0.35       0.         0.23       0.23       0.
  33.31  0.01775  0.02704        2.680      0.266      0.25       0.25       0.         0.16       0.16       0.
  36.44  0.00559  0.02684        2.931      0.093      0.09       0.09       0.         0.06       0.06       0.
  41.14  0.03422  0.02654        3.309      0.637      0.60       0.60       0.         0.39       0.39       0.
  45.38  0.29834  0.02627        3.650      6.225      5.84       5.84       0.         3.85       3.85       0.
  54.62  0.46039  0.02568        4.393     11.559     10.84      10.84       0.         7.14       7.14       0.
  74.83  0.01439  0.02440        6.019      0.496      0.47       0.47       0.         0.31       0.31       0.
  85.63  0.29727  0.02371        6.888     12.965     12.16      12.16       0.         8.01       8.01       0.
  98.34  0.03839  0.02290        7.910      1.745      1.64       1.64       0.         1.08       1.08       0.
 113.23  0.25168  0.02195        9.108     16.026     15.03      15.03       0.         9.90       9.90       0.

                                          106.640                          0.        65.90      65.90       0.
THE END
   93:05:00        146978816
```

FIG. 8. A typical computer printout of the results from a sample of fatty acid methyl esters. Sample number is listed at the top as a set number along with the date. The indicative data are listed on the first line. The first column lists the absolute time of each peak found, the second column its peak height in inches (arbitrary). Column three lists the value for the baseline at the retention time of the peak maximum. Column four is the relative retention time compared to methyl stearate (1.000). Column five is the mass (arbitrary units) calculated for each peak along with its sum. The next three columns give the mass percentage for each peak; mass percentage 2 is corrected for peak overlap while mass percentage 3 is corrected for blank contaminates if available. In this particular run no blank run was available; hence, zeros are printed in this column. The absolute mass of each peak, in micrograms, is given in the last three columns corresponding to mass percentages 1, 2 and 3.

TABLE VI

Accuracy of Gas Liquid Chromatography Computer Data Reduction System on Standard Fatty Acid Mixtures; Statistical Treatment of Results

Fatty acid methyl ester	Average, n=8	Standard deviation	Theoretical value	Difference, Avg.–Theo.	Coefficient of variation
14:0	11.57	0.27	11.8	−0.1	0.023
16:0	23.37	0.20	23.6	−0.2	0.009
16:1	6.96	0.27	6.9	+0.1	0.039
18:0	13.46	0.26	13.1	+0.4	0.019
18:1	44.69	0.73	44.6	+0.1	0.016

3. The last criterion eliminates peaks on the basis of the expected basewidth of a peak at a given retention time. If the observed basewidth is less than 75% of the expected value the peak is rejected. In this system basewidth is linearly related to retention time (1,6,7).

Shoulders, not located by this procedure, are handled manually at present. This will be discussed later.

After the peaks are located by retention times, the true peak heights are determined by referring to the raw data DVM reading, and the relative retention time for each peak is calculated. At this time the raw data, the smoothed data, and time base file are cleared from the core. The computer is now ready to calculate the mass and mass percentages for each peak located in the sample run.

CALIBRATION AND CALCULATION OF RESULTS

The gas chromatograph used in this work has an argon ionization detector. This particular detector is notoriously nonlinear in response; hence, it is necessary to carefully prepare calibration data. Figure 5 is a curve of detector response (maximum peak height) to the injection of a known mass of methyl palmitate. Since peak basewidth is a linear function of retention time (6,7) for isothermal operation, it was decided to use only peak height to calculate the mass of each component in a sample. The calibration curve (Fig. 5) was reoriented to make mass the ordinate and peak height the abscissa. An eight-power polynomial fit was made to this curve between zero and the saturation level of the detector (11 in. on the recorder). Thus for any response of the detector it was possible to calculate an equivalent mass of methyl palmitate. Thus the mass, M_{t_i}, of any component with any retention time, RT_i, is given by the equation

$$M_{t_i} = \frac{RT_i}{RT_a}(M_{e_i})$$

where RT_a is the retention time at which the methyl palmitate calibration curve was obtained, and M_{e_i} is the equivalent palmitate mass for a component with a retention time RT_i.

Figure 6 is an experimental verification of this relationship showing the experimental response to 10 μg samples of various fatty acid methyl esters of widely differing retention times. The solid line is the best fit hyperbola drawn through these points. The equation above describes a hyperbola.

After the mass of each component has been calculated, a second correction factor is applied to the mass on the basis of the relative retention time of the peak. Figure 7 is a plot of this correction factor versus relative retention time. The second correction factor compensates for the differential responses of compounds of differing molecular weight and chemical identity as well as loss of longer chain compounds on the column.

After the mass of each component is determined, they are summed and the mass per cents calculated. A peak resolution correction can be applied at this point in the program if it has been called on the indicative data word. This correction assumes a gaussian shape for the peak and corrects the peak height by subtracting a contribution from nearest neighbor peaks, selecting each pair from the beginning of the run and incrementing by one peak through all peaks in the sample.

Finally the baseline correction is applied, if called. Each mass percentage is printed out. If the absolute mass is called, a separate column is printed corresponding to mass percentages, resolution-corrected mass percentages, and baseline-corrected mass percentages, depending on whether these options were also called. Figure 8 is a typical computer printout of such results on a sample of methyl esters from a total lipid extract from seal red blood cells.

SHOULDER LOCATION AND CALCULATIONS

The computer program currently does not locate shoulder peaks that do not have a definite valley separating them from their nearest neighbors. Referring to Figure 4,

TABLE VII

Accuracy of Gas Liquid Chromatography Computer Data Reduction System on Standard Fatty Acid Mixture (NHI Mixture F)

Fatty acid methyl ester	Relative retention time	Mass per cent of each component						
		4136[a]	4139	4140	4141	Average, ± S.D.	Theoretical value	Per cent error
14:0	0.32	2.55	2.54	2.37	2.37	2.48 ± 0.08	2.58	−3.9
16:0	0.56	4.36	4.27	4.19	4.18	4.25 ± 0.07	4.41	−3.6
18:0	1.00	7.36	7.13	7.41	7.35	7.31 ± 0.10	7.50	−2.5
20:0	1.82	16.06	15.53	16.20	16.12	15.98 ± 0.26	15.92	+0.4
22:0	3.35	27.63	27.20	26.80	26.98	27.15 ± 0.30	26.50	+2.4
24:0	6.28	42.03	43.34	43.03	43.00	42.85 ± 0.49	43.05	−0.5
Total,[b] μg		131.6	126.8	14.7	16.2			

[a]Run number.
[b]Total mass injected.

the peaks at 2.6, 3.6 and 4.7 min would not be located. Several subroutines using rate of change of slope have been explored, but so far none has proved satisfactory. Research is continuing on this point.

In the meantime the retention times of shoulders are inserted into the computer manually. A sample tape is fed to the computer with a program that asks only for a CRT printout of the raw data. A technician then scans the printout for shoulder peaks, noting the retention time. The retention times are punched onto IBM cards along with the run number. Alternatively if a strip chart recording of the run is made, the shoulders can be read off the strip chart, alleviating the need of the preliminary computer run.

The computer reads the retention times for the shoulders off the cards and finds the peak associated with that retention time. It also checks to see whether it found a peak at that time to avoid duplication. The shoulders are then handled as discrete peaks in the remainder of the program.

RESULTS

Table I gives some data on the reproducibility of repeated injections of 10 μg of methyl palmitate. The deviations are representative of the performance of the gas chromatograph rather than the variation in the computer. This Table is primarily a measure of the limits of accuracy that can be expected using the data station coupled to this particular gas chromatograph.

Table II compares results of manually measured and computer-calculated retention times for a standard mixture of fatty acid methyl esters. The average values are essentially identical. Table III compares manual and computer values for the peaks in the same runs. Again the agreement is excellent. In Table IV similar comparisons are made for a different fatty acid mixture containing compounds of much longer retention items. The agreements are still good. These Tables show that this system can, without manual intervention, locate retention time and peak height with as good or better accuracy than can be obtained by manual methods.

These results are an absolute necessity for such a system before any attempt to calculate mass percentages can be made. Any degradation of the data at this stage would invariably reduce the accuracy of the final results.

Tables V through VII give results obtained on NHI standard mixtures. Table V presents results on duplicate analysis over a 100-fold concentration range for a five-component mixture. Inspection of these data indicates that there is no loss of accuracy over this range, although the low mass injections, runs 4250 and 4251, show somewhat poorer agreement than the other pairs. Table VI gives a statistical treatment of these runs and the theoretical values. The standard deviations, differences and coefficients of variation are all well within criteria set for this system when it was originally designed, i.e., components present below 5% of the total must be analyzed within 10% and those over 5% must be within 5% of the true value.

Table VII gives data on the accuracy of analysis using NHI mixture F. The accuracy is again within the desired limits.

DISCUSSION

The system described here is now being used routinely in this laboratory for the analysis of fatty acid mixtures from natural samples. Since immediate turn-around of the data is not usually necessary, 1000 to 1200 ft of tape are recorded (representing between 10 and 15 sample runs) before the tape is transferred to the computer for processing. The 1200 ft of tape is equivalent to 30 hr of gas chromatograph operation, which is roughly three to four laboratory days.

The tape is processed currently on a Control Data Corporation 6600 digital computer. A 1200 ft tape can be completely processed by this computer in 20 min, operating in batch mode at a 1:1 priority. The cost of operating a 6600 is approximately $100/hr. Thus for a typical tape the cost per sample is about $2. On a more efficient machine such as the CDC-7600, this cost per sample can be reduced by a factor of 3. These figures are for quite long runs where the gas chromatograph is operated isothermally. A typical run may last 2 to 3 hr. The processing time is directly proportional to the length of the run. Short runs can be processed much more quickly at considerably lower cost.

The advantage of this system is primarily in the saving on technical man hours compared to either manual or semiautomatic gas chromatographic data reduction. A 3 hr run can be processed by the computer in less than 1 min; manual data reduction on the same run would take up to 1 hr or more, semiautomated techniques using a digital integrator would take at least 10 min. In addition, the mathematical treatment of the data can be considerably more sophisticated; noise problems such as spikes, slope changes and baseline drift can all be handled much more easily by a computer than a digital integrator.

Another feature that we hope to incorporate into the system is an identification subroutine (8) that will, by the use of such data as RRT (9), equivalent carbon chain length (10), separation factors (11), column polarity, age and temperature (12) and standard runs, identify the various fatty acids in an unknown sample. While such identifications would not be absolute, they should prove of great help to the investigator in routine fatty acid analysis.

The accuracy of the system can be improved further. Using a more modern gas chromatograph with a flame ionization detector, accuracies better than 1% for all components should be easily attainable. The major difficulty remaining in this system is the handling of unresolved peaks or shoulders. It is certainly not insolvable, but rather a matter of deciding which of several options to use. The best solution would be to use column phases that give good resolution to all components in a given sample. As this is probably an urealistic goal, some compromises are necessary. The programming of shoulder subroutine involves considerable testing and debugging. Hence, we have delayed incorporating such a subroutine into our program until we are completely satisfied with its performance.

ACKNOWLEDGMENTS

The electronics designs for the data station were made by G. Holladay. The computer programming was done by M. Uyemura. This work was performed under the auspices of the U.S. Atomic Energy Commission.

REFERENCES

1. Tandy, R.K., F.T. Lindgren, W.H. Martin and R.D. Wills, Anal. Chem. 33:665-669 (1961).
2. Bauman, F., E. Herlicska, A.C. Brown and J. Blesch, J. Chromagogr. Sci. 7:680-684 (1969).
3. Hancock, H., and I. Lichtenstein, Ibid. 7:290-292 (1969).
4. Tochner, M., J.A. Magnuson and L.Z. Soderman, Ibid. 7:740-744 (1969).
5. Castor, W.O., in "Methods of Biochemical Analyses," Vol. 17, Edited by D. Glick, Interscience Publishers, New York, 1969, p. 135-188.
6. Carroll, K.K., Nature 191:377-378 (1961).
7. Brandt, A.E., and W.E.M. Lands, Lipids 3:178-181 (1968).
8. West, C.E., and T.R. Rowbotham, J. Chromatogr. 30:62-76 (1967).
9. Farquhar, J.W., W. Insull, P. Rosen, W. Stoffel and E.H. Ahrens, Nutr. Rev. (Supplement) 17:1-30 (1959).
10. Miwa, T.K., K.L. Mikolajczak, F.R. Earle and I.A. Wolff, Anal. Chem. 32:1739-1742 (1960).
11. Ackman, R.G., JAOCS 40:564-567 (1963).
12. Jamieson, G.R., and E.H. Reid, J. Chromatogr. 42:304-310 (1969).

[Received October 30, 1970]