Compound Identification in Mass Spectrometry by Optical Coincidence*

C. H. O'DONOHUE, M. A. MANZELLI, and W. F. KUHN
Philip Morris U.S.A. Research Center, Richmond, Va. 23206

W. A. POWELL
University of Richmond, Richmond, Va.

Received January 28, 1969

The process of identifying unknown compounds by their mass spectra has been facilitated by the development of an optical coincidence system that permits the storage and retrieval of relevant data. New features of this system are the correlation of the relative intensity of the peak to its mass number and the rapid association of other instrumental analytical data to the mass spectra. Compounds are identified by associating the mass number and relative intensity with chemical classification and/or molecular weight. Initial results have shown this system to be an efficient, time-saving means of identifying compounds by their mass spectra, especially in cases where the spectra contain no detectable molecular ions.

The mass spectra of compounds have been appearing in ever-increasing numbers in the literature. Several organizations have compiled some of these spectra, but at present, there is no central depository. In the Philip Morris Research Center mass spectral laboratory, identification of a compound involved searching through three major sources—American Petroleum Institute certified spectra, Dow Chemical uncertified spectra, and its own unpublished spectra—as well as many minor sources. Manually searching these files was a cumbersome and time-consuming chore which was made even more difficult when no detectable molecular ion was present in the unknown spectrum. Various systems have been devised to facilitate the searching process, and the more commonly used ones will be briefly discussed.

Edge-punched cards[1] have been used, but the number of cards became so great that quick-sorting was no longer possible. IBM punched cards and a sorter are not practical.[2,3]

Computer-based magnetic tape systems[4] have been devised to process the data and to give quick, reliable answers. However, if computer time is not freely available and work is done in batches only, then a computer-based system is not practical. A system must be found which is under the control of the user and can be located within the laboratory.

One type that meets these requirements is an optical coincidence system, based on an inverted data processing mode; that is, each record unit handles a single characteristic, and the identities of the information items are coded in the appropriate units.

To retrieve an item in this system, the user chooses the characteristics that best describe it. The cards representing the chosen characteristics are superimposed over a light source which has numerical coordinates. Wherever the holes coincide—i.e., the light beam passes through all of the cards—the information item represented by that particular serial number has all of the chosen characteristics. Its number is obtained by reading the coordinates of the X- and Y- axes.

Basically, the optical coincidence system consists of a light-source box and plastic cards, 9⅝-inch × 11½-inch, each of which has a capacity of 10,000 compound accession number positions. Input to these plastic cards[5] is by tabulating cards prepared from logsheets (Table I).

The logsheet heading contains the compound name, the accession number, the original source (i.e., the API files, the Dow uncertified file, the Philip Morris files, journal reference, etc.), the structure, and the molecular weight.

The 10 strongest peaks and their pattern coefficients (relative intensities) are selected from the original source. The base peak is listed first, followed in descending order by the mass numbers, m/e, of the other peaks, with their corresponding relative intensities, which are rounded off to whole numbers. If a half mass occurs, it is coded as the next lower number, with its corresponding relative intensity. For example, peak 54.5 with an intensity of 29% is coded as 54 with an intensity of 29%. The personnel of the mass spectral laboratory are responsible for selecting the peaks and their relative intensities, as well as choosing the correct chemical classification coding.

The data on the logsheets are keypunched directly into tabulating cards (Figure 1). The first 10 fields contain the peaks and their intensities. Columns 1 to 3 hold the base peak, which does not have its relative intensity entered, since it is always 100%. Each of the other nine fields is allotted three columns for mass number and two columns for relative intensity. There is a one-column space after each field.

Columns 59 to 66 comprise the chemical classification field. Digital punches are used to indicate the individual

* Presented before the Division of Chemical Literature, ACS, Atlantic City, N. J., September 13, 1968.

Table I. Logsheet

Compound _____ Acc. No. _____
Source _____
Structure _____ Mol. Wt. _____

Peak	Intensity	Hydrocarbon	Nonmetallic
		Nitrogen	Natural Products
		Sulfur	Oxygen
		Special	

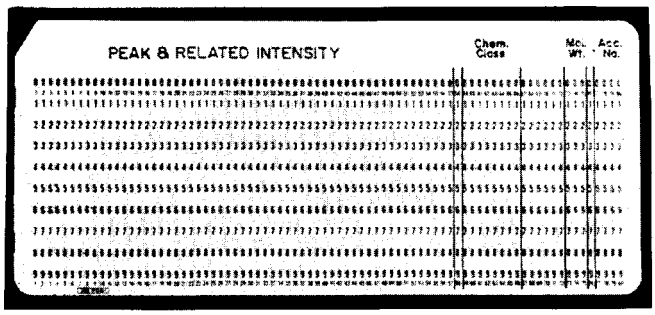

Figure 1. Tabulating card for Termatrex input

Table II. Chemical Classification Cards

Blue 0	Hydrocarbon	Blue 80	Nonmetallic
1	alip satd	81	boron
2	alip unsatd	82	bromine
3	cyclic	83	chlorine
4	aromatic	84	fluorine
5	arom fused	85	iodine
6	C & H only	86	phosphous
Blue 20	Nitrogen	87	others
21	amides & amines	Yellow 0	Natural Products
22	imides & imines	1	alkaloids
23	nitrites	2	amino acids
24	nitros	3	carbohydrates
25	heteros	4	lipids
26	others	5	terpenes
Blue 40	Sulfur	6	others
41	sulphones	Yellow 20	Oxygen
42	thiocyanates	21	acids
43	thiols	22	alcohols
44	heteros	23	aldehydes
45	others	24	ethers
Blue 60	Special	25	ketones
61	metallic org	26	phenols
62	pesticides	27	heteros
63	silylethers	28	anhydrides
64	inorg	29	esters
65	polymers	30	lactones
66	deuterium	31	others

chemical classes. Thus, some columns will have multiple punches. The molecular weight is entered in columns 73 to 75 as a whole number. The assigned accession number is in the last field, columns 77 to 80.

Concurrently, one additional tabulating card is prepared. This tabulating card contains the compound's name, the original source, the molecular weight, and the accession number. Two listings are prepared from this card: one by accession number and another by molecular weight. After viewing the appropriate optical coincidence cards to delineate the unknown, the accession number listing is used to give the user readily the compound's name, its molecular weight, and the original source of the complete spectrum. Listing of all compounds by molecular weight allows research personnel outside the mass spectral laboratory to determine quickly whether or not a compound of interest has been identified by mass spectrometry previously and entered in this system.

The system as designed involves a total of 423 Termatrex cards, distributed as follows:

Mass number	300 cards
Chemical classification	54 cards
Molecular weight	29 cards
Peak number and related intensity	40 cards

Peaks of 0 to 99 m/e are coded in a green edge-coated deck with the number edge tab matching the mass number—e.g., peak 39 is assigned green edge-coated card No. 39. Peaks of 100 to 199 m/e are coded in a white edge-coated deck with the number edge tab matching the last two digits of the peak—e.g., peak 136 is assigned white edge-coated card No. 36. Peaks of 200 to 399 m/e are coded in a black edge-coated deck, where each card represents two peaks—e.g., card No. 0 is 200 and 201 m/e. Peaks of 400 to 600 m/e can be coded in another deck, where each card would represent two peaks.

Blue and yellow edge-coated cards (Table II) have been assigned to the chemical classification coding. Each of the major groupings—i.e., hydrocarbon, nitrogen, sulfur, special, nonmetallic, natural products, and oxygen—is allotted 20 cards, which permits 19 specific subgroupings. Expansion is possible, for in no group have all the positions been assigned. A spectrum's accession number is entered into the major grouping card, and into the appropriate subgroup cards as well.

Digital coding is used for the molecular weight input, which utilizes 29 orange edge-coated cards. There are sets of nine cards for the hundreds unit, 10 for the tens unit, and 10 for the ones unit.

Colored transparencies are used for relating a mass number to its intensity. The system employs five sets of transparencies, one for each of the following ranges of intensities: 0 to 16% (red), 15 to 27% (green), 25 to 53% (yellow), 50 to 79% (blue), and 75 to 100% (red). The overlapping of the intensity ranges allows for a 5% variation in the intensity value. Based on daily laboratory runs of n-butane, the relative deviation is 4.5% or less for the individual intensity. It is clearly understood by the authors that this figure of 5%, when using various sources, is open to question. However, the utility of the system affords the user a rapid screening of the spectra on file, even though the relative intensities do not match

Table III. Coding of Specimen Compound, 1,3-Dimethyl-5-Isopropylbenzene

Description	Deck	Card Number
Chemical Class	Blue	0, 1, 4, 6
Mol. wt. 148	Orange	1, 14, 28
10 strongest peaks	Green	15, 27, 39
	(0–99 *m/e*)	41, 77, 91
	White	5, 33
	(100–199 *m/e*)	34, 48

Peak and Intensity		Transparency Color	% Intensity Range	Mass Range
M/e	%			
133	100	Red	75–100	120–139
148	26	Green	15–27	140–159
		Yellow	25–53	140–159
91	11	Red	0–16	80–99
15, 27, 39	10	Red	0–16	1–39
41	10	Red	0–16	40–49
134	10	Red	0–16	120–139
105	9	Red	0–16	100–119
77	7	Red	0–16	60–79

within the arbitrary value. Through the use of the transparency cards, the user can arrive at a suggested chemical structure. There are eight transparencies in each of the preceding intensity sets—i.e., one each for the following mass number ranges: 1 to 39, 40 to 49, 50 to 59, 60 to 79, 80 to 99, 100 to 119, 120 to 139, and 140 to 159 *m/e*. Transparencies for higher mass numbers are not included because studies[6] of 15,746 peaks from the spectra of 4036 compounds showed that the first, second, and third most abundant ions were found below 160 *m/e*.

Table III illustrates the techniques of coding a compound, in this case, 1,3-dimethyl-5-isopropylbenzene. Its accession number, 0675, will be entered into the indicated Termatrex cards.

To identify a compound in this system, the cards representing the known characteristics of the compound are chosen (Figure 2). If, when these cards are superimposed over a light source, there remain too many points of optical coincidence, the transparencies that correlate the mass number with its per cent relative intensity are used to limit the number of spectra that must be examined. Since there is an overlap in the per cent intensity range, some peaks with their corresponding intensities will require two transparencies, while in some cases one transparency will suffice for two or more peaks. For peaks 15, 27, and 39, which have an intensity of 10, the red transparency card (0 to 16%) for the mass range 1 to 39 *m/e* will be used. The green transparency card (15 to 27%) and the yellow transparency card (25 to 53%) for the mass range 140 to 159 *m/e* will be used for peak 148, which has an intensity of 26%. The remaining points of optical coincidence will identify compounds with similar chemical characteristics and mass spectra as the unknown. If, after placing the transparencies on the viewer, there are no points of optical coincidence remaining, the user looks for points where just one primary color is showing. This indicates a compound whose spectrum is similar to that of the unknown, with a variance in intensity within the mass range represented by the transparency which prohibits the light passage.

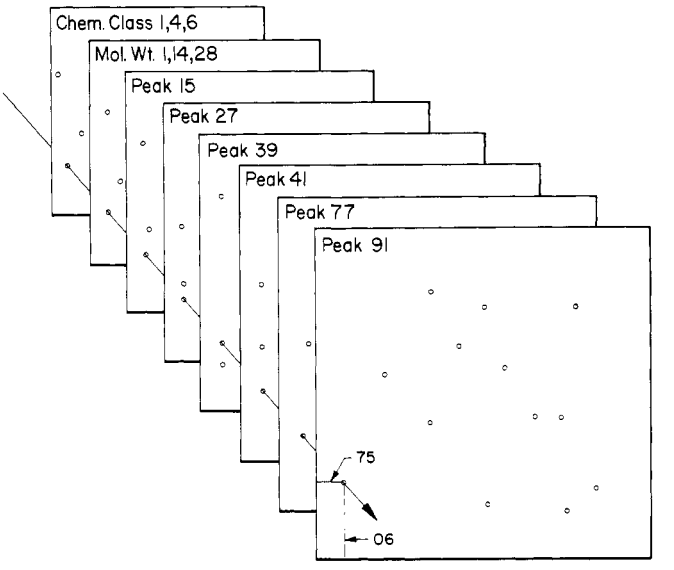

Figure 2. Termatrex search technique

DISCUSSION

The system which has been described provides the Philip Morris mass spectral personnel with a practical and economical laboratory method of storage and retrieval of mass spectral data. This system incorporates all of the characteristics that the Philip Morris Research Center personnel felt were necessary for their specific needs. Through the employment of the Termatrex method, the stored data is on a minimum number of input cards and requires very simple retrieval equipment. The file of data is kept current by expanding the data input on the existing cards, rather than by adding new cards to the system. Thus, each updating does not necessitate additional laboratory space for storage, nor does it increase the time required for searching the data. The data on the Termatrex cards can be transferred automatically to paper tape, tabulating cards, or magnetic tape later for use in a computer system.

One drawback to this type of storage and retrieval system is the relatively high data input time. The coding of five to 10 spectra requires approximately one hour. However, if all of the desirable Termatrex features were present in another system, the input time for coding would be of the same magnitude.

Compared with previously employed methods, the described system offers a significant reduction in the time needed to locate spectra for the identification of unknown compounds. Studies to date have shown that it takes approximately 5 minutes or less to group the cards containing the coded characteristics of an unknown compound and to view the superimposed Termatrex cards. The former practice of searching several individual sources, such as the American Petroleum Institute files and the Dow Chemical uncertified file, required a minimum of 15 minutes per spectrum.

The major advantage of this system is that a mass spectrum may be correlated with the chemical data obtained on the unknown from other instrumental analyses. Up to now, when there was no detectable molecular ion present in the spectrum, there was no way to

enter the files. The files previously used in the laboratory had the known spectra arranged by molecular weight, and great difficulty was encountered in identifying a sample if the molecular weight was unknown. With the present system, the chemical class can be combined with the peaks present in the spectrum to locate similar spectra of knowns. In addition, unknowns can be identified readily by nonprofessional laboratory assistants.

Chemists from non-mass spectral laboratories can use this system to check confirming spectra of their compounds. The use of the system does not require knowledge of mass spectrometry, and thus it can be used by anyone.

LITERATURE CITED

(1) Zemany, P. D., *Anal. Chem.* **22**, 920 (1950).
(2) Haefele, C. R., and J. F. Tinker, J. CHEM. DOC. **4**(2), 112 (1964).
(3) McLafferty, F. W., and R. S. Gohlke, ASTM E-14 Committee Meeting on Mass Spectrometry, New Orleans, La., 1954.
(4) Crawford, L. R., and J. D. Morrison, *Anal. Chem.* **40**, 1464 (1968).
(5) Termatrex, Jonker Corp., Gaithersburg, Md.
(6) McLafferty, F. W., "Mass Spectral Correlations," Advances in Chemistry Series, No. 40, American Chemical Society, Washington, D. C., 1963.

An Evaluation of TEXT360 for Producing Reports*

HERMAN SKOLNIK and RUTH E. CURTISS
Hercules Research Center,** Hercules Incorporated, Wilmington, Del. 19899

Received February 7, 1969

TEXT360, designed by IBM for use with Operating System/360, is a text-processing system with capabilities for entering, updating, correcting, and rearranging information for the production of printouts in a variety of page formats. TEXT360 is oriented primarily to the printing industry, and the program provided by IBM has imbedded in it the standards of the printing industry. Consequently, we found it necessary to convert these standards to those we need for the production of manuals, bibliographies, and other documents in a research environment which requires the 8½ x 11-inch page size rather than the different sizes of the printing industry. The several problems we encountered in the use of TEXT360 are discussed, with suggestions for solutions to the problems.

Although computers have been designed for and applied in numerical operations primarily, there has been considerable activity in the use of computers with text material.[1,2] This activity, for the most part, has been directed to typesetting environments, such as for the production of newspapers, books, and journals. An excellent example of a technical journal which is photocomposed on the Photon Photographic Type Composing machine using computer-generated tape is the American Chemical Society's *Journal of Chemical Documentation*.[3] All applications with text material reported so far have had typographic needs as the objective. Consequently, most of the features for which the computers have been programmed are related to type font and style. These features are of minor importance in a research environment.

There are many operations in a research environment, such as at the Hercules Research Center, that require the production of manuals, reports, bibliographies, and other text material on a continuing and updating basis. Because changes must be made, such as new information added, old information deleted or corrected, or because the information needs to be communicated in different formats, these operations are relatively time-consuming and expensive. We have been interested for some time, therefore, in applying our computers to text processing of in-house literature. The sophisticated programming required for such text-processing became available with the advent of IBM's TEXT360.[4]

MACHINE REQUIREMENTS

TEXT360 requires for input an IBM 29 keypunch and for output a minimum computer configuration of a System/360 Mod 40 with 128K of core storage and the Universal Instruction Set, and an IBM 1403 printer with Universal Character Set. The type of print train or chain used on the printer is determined by the output requirements, e.g., all caps, upper and lower case, and 64- or 120-character sets.

FEATURES OF TEXT360

The several standard features of TEXT360 are:

1. Automatic hyphenation of words at end of line.
2. Right margin justification by spacing between words and by hyphenation.
3. Specification of line width and page length.
4. Automatic numbering of pages.
5. Automatic printing of running heads.
6. Provision for prescribed indentions.
7. Choice of one- and two-column formats.

Other functions provided by the system, although more complex to apply and to execute, include the generation of horizontal and vertical lines for tables and charts, the

* Presented at Middle Atlantic Regional Meeting, ACS, Washington, D. C., February 14, 1968.

** Hercules Research Center Contribution Number 1491.