The Documentation of Molecular Spectra.

By H. W. THOMPSON.

DURING the past fifteen years there has been a marked growth of spectroscopic work in relation to chemical problems. This was stimulated, primarily, by a better theoretical interpretation of the spectra of molecules in terms of the molecular geometry or nuclear structure, and of the molecular energy levels. It led naturally to technical advances in methods of quick, direct, and automatic recording of spectra, so that data on a vast number of molecules have accumulated rapidly. This output of spectral knowledge, already large, is still increasing. There is a particular emphasis on infrared absorption and Raman data, which have become indispensable for much research in organic chemistry involving the study of complex natural or synthetic products. Indeed, the vibrational spectrum of a molecule is now generally recognised as a highly characteristic property, more useful for identification, for assessment of purity, and for other purposes than most other physical properties previously employed. In addition, the spectral data are required in connexion with many physical problems concerned with molecular dynamics.

Anyone working in these fields knows that, for most purposes, it is necessary to have at hand, not only the spectrum of some individual compound, but also a collection of such spectra covering many structural types, for comparison and correlation. The task of making such a collection is formidable, and if carried out by every individual laboratory it would involve an undesirable duplication of effort, if indeed the work could be done effectively at all. Another difficulty arises since rising costs are now making it impossible to publish in the scientific periodicals all the spectra of pure compounds and standard technical products, and many results which might normally be available for general use are accumulating in different laboratories where they lie inaccessible and unknown to the wider circle of potential users.

The compilation of spectral data in a form in which they can be used by different workers to the maximum advantage itself involves some problems. For example, the data are not well suited to tabular presentation, and even if set out in this way they would be very cumbersome. Also, in the case of infrared and Raman spectra, quantitative measures of the band intensities which have absolute significance are, as yet, only rarely available. Whatever method of compiling infrared spectra is adopted, it must be designed so that answers can be quickly obtained to a variety of questions of which the following are examples :

(i) What is the spectrum of a particular compound ?
(ii) With what substance is a measured spectrum associated ?
(iii) What substances give rise to a particular absorption band, or what set of spectral characteristics are connected with some particular atomic grouping or structural feature ?
(iv) What impurities are present in a given substance ?
(v) Which recently published papers deal with some special technical problem ?
(vi) What recent theoretical work has dealt with a particular topic ?
(vii) What were the main points made by a particular author in an important recent paper ?

The most promising kind of catalogue and index for these purposes can be devised by using punched cards, which record the relevant spectral and structural data and can be sorted by means of an appropriate code. The means for setting up such a system have therefore been explored by the Infrared Absorption Data Joint Committee, which was set up several years ago by the Chemical Society to advise on the more general problem of the presentation of spectral data. The following were members of this Committee, representing the bodies indicated :

H. W. Thompson (Chemical Society and Royal Society, chairman), R. S. Cahn (Editor, Chemical Society), E. D. Hughes and A. E. Martin (Chemical Society), A. C. Menzies (Physical Society), N. Sheppard (Faraday Society), H. Powell (Institute of Petroleum), W. C. Price (Photoelectric Spectrometry Group), J. E. Page (Society of Public Analysts), E. F. G. Herington (Department of Scientific and Industrial Research).

The Committee had the benefit of consultations with a number of those who were potential users, or who had already gained experience with indexes of this kind, including Prof. Sir Alexander R. Todd, Prof. E. R. H. Jones, Dr. H. B. Henbest, Mr. R. Wood [1] (Atomic Weapons

---

[1] R. Wood, United Kingdom Atomic Energy Authority, AWE Report H26/52.

Research Establishment, Aldermaston), Dr. L. J. Bellamy (Explosives Research and Development Establishment, Waltham Abbey), Prof. E. A. Braude, Prof. J. W. Cook, and Dr. F. L. Rose; a census of views on some points was also taken at meetings of the Infrared Discussion Group and of the Hydrocarbon Research Group of the Institute of Petroleum.

Two kinds of punched card have recently been introduced for spectral data. One is an IBM card, on which the positions of the main absorption bands are shown by slots on a rectangular matrix.[2] These cards, which are being issued by the American Society for Testing Materials, Philadelphia, are sorted by electrically operated Hollerith machines. The other type is a keysort card, prepared under the editorship of Dr. C. Creitz for a Committee of the National Research Council, Washington;[3] this has holes around the edges, coded and slotted according to the spectral and structural characteristics of the compound concerned. These cards are sorted by manual mechanical methods. The main advantage of the IBM cards is that a very large number of cards can be sorted rapidly by the machines. On the other hand, they carry only a limited amount of information, and the inclusion of more would require much larger cards and lead to other difficulties. Keysort cards are able to record much more information, and their main disadvantage is the need for sorting by manual methods. The latter may not, in fact, be a serious shortcoming in the present case, for it is probable that workers in special fields will usually retain groups of cards in defined lots, thus simplifying the sorting operations.

After detailed examination the Committee decided that a scheme involving keysort cards would at present best meet British and European requirements, especially if it could provide rather more information than that furnished by the existing schemes. The extent and kind of material which should be included on the cards, together with detailed methods of coding, were next considered. The Committee was then informed by Dr. H. Kaiser of the Institut für Spektrochemie, Dortmund, about a similar project under his direction sponsored by the Gesellschaft für Spektrochemie und angewandte Spektroskopie. This group, working in conjunction with the Chemisches Zentralblatt in Berlin, planned to provide a literature survey of spectral data on keysort cards, with a code which was more extensive than others used previously, and which was designed specifically to deal with the sort of questions listed above. It became clear that a union of the German scheme, with its literature survey, and the British scheme, which had been directed primarily at the spectra of pure compounds from published and unpublished sources, would be mutually beneficial and might have operational and technical advantages leading to a collection of universal value. The joint plan which has now been agreed upon is outlined below.

General Plan and Coding System. Two identical editions of cards will be published, in English and German respectively, the publishers being Butterworths Scientific Publications, London, and Verlag Chemie, Weinheim. For the present, they will deal only with infrared and Raman spectra, with an emphasis on the former, although it is hoped to extend the scheme later to ultraviolet spectra. Two separate cards will be issued in each edition, namely, S (spectral) and L (literature) cards. S cards will be rose-coloured (infrared spectra) and white (Raman spectra), and the L cards will be yellow. In order to advance from the start along a broad front of the widest possible interest, compounds will be chosen from many classes, such as hydrocarbons, general organic chemicals, fluorocarbons, steroids, carcinogens, heterocyclics, organic derivatives of silicon, phosphorus, and sulphur, pyrimidines, carbohydrates, amides and peptides, polymers, etc. Spectral data will also be included on inorganic materials such as salts and minerals, and on a limited number of standard, technologically important substances of indefinite structure.

(a) *Spectral cards.* The two sides of an S card are illustrated in Figs. 1 and 2. On the front (Fig. 1) the name of the compound is printed, together with its main physical properties, the molecular and structural formulæ, the author or source of the spectrum, and the conditions under which the spectrum was measured. On the reverse side of the card (Fig. 2) the spectrum is shown, with a Table of frequencies of the chief bands and their rough apparent intensities. The infrared spectra are drawn on a standard format with abscissa linear in wave numbers, covering the range 200—4000 cm.$^{-1}$. A break of scale is made at 2000 cm.$^{-1}$, and the positive ordinate is percentage transmittance.

Around the edges of the card there are two rows of holes, with 203 holes in all. These can be cut so as to make deep, shallow, or " middle " slots. The upper and right-hand sides (Figs. 1 and 2, 1—31 and a—v) are used for coding the structural features of the compound, and the bottom and left-hand sides (32—61 and A—V) are used for marking the positions of the main

[2] L. E. Kuentzel, *Analyt. Chem.*, 1951, **23**, 1413.
[3] See W. Brugel, " Einführung in die Ultrarotspektroskopie," Steinkopff, Darmstadt, 1954, pp. 224—226.
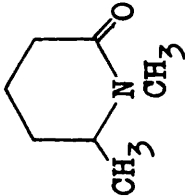
	HYDRO-CARBON	DEEP	HETEROCYCLIC NUCLEUS	MIDDLE	SHALLOW OVERLAP

$C_7H_{13}ON$

	AUTH.CODE/YEAR		MOL. WT.
	EDW-O-SIN/54	I	127.2

LIT. CARD NO.

BEILSTEIN SYSTEM NO.

3179

- X

	HYDRO-CARBON	HETEROCYCLIC NUCLEUS	OVERLAP
	·Cl	·I, ·IO$_n$	-Br
	-CN	·NCO, -NC, ·SCN, -N$_3$	-F
	-NO$_2$	NITRAMINE, NITROSAMINE	-NO
	-O-	ACETAL, OZONIDE ORTHOESTER	-S-
	-(SO$_2$)-	-X<(O, S, Se)	-(SO)-
	-N=N-, -N=(NO)-	-N$_2$	-P<
	-N<	-X<(N,P,As)	-(XO)<(N,P,As)
	-H(c-c)	-OR', -SR'	-H(c=c)
	-R(c-c)	-NR$_2$, -Hal.	-R(c=c)
	-V[O]	-V[N-T]	-V[S]
	-R[O]	-R[N-T]	-R[S]
	-R	T: -R, -H; CYCL. KETONE	-H
	-OH, -OOR', -SH, -OM	T: -OH; CYCL. ANHYDRIDE	-OCOR'
	-OR'	T: -OR'; LACTONE	-SR'
	-NR$_2$	T: -NR$_2$; LACTAM	-Hal., -N$_3$

SUBSTITUENT CONJUGATED TO DOUBLE BOND

	IN RING	AT RING	IN CHAIN
	BETAINE, MESOION.COMPD	MOLECULAR COMPOUND	
	MINERAL ACID DERIV., ALCOHOLATE	ISOTOPE	METALLORGANIC COMPOUND
		RADICAL COMPLEX	

-X-

-X·Z

m.p

b.p. 1 80-85°

II

-U [Y]

STATE	SPECTRAL RANGE
liquid	600-3600cm^{-1}

CONC.PRESS.TEMP.	PRISM/EXCITING LINE
	NaCl

U · C · Y

V · V

n_D^{24} 1.4802

d

THICKNESS	APPARATUS
film	P. E. 21

PURITY	DIFFERENTIAL
elem. anal.	

1000
11/55

1:6-Dimethylpiperid-2-one

AUTHOR

O. E. Edwards and Tara Singh

JOURNAL

Canad. J. Chem. 32. (1954.) 683-91

Fig. 1.

poly I □

poly I II

<3700

3700
3600
3500
3400
3300
3200
3100
3000
2900
2800
2600
2400
2200
2000
1800
1750
1700
1650
1600
1550
1500
1450
1400
1350
1300
1250
1200
1150
1100

<250
250
350
450
550
650
700
725
750
775
800
825
850
875
900
925
950
975
1000
1050
1100

Absorption %

← Wave number (cm⁻¹)

Wavelength (μ) →

Transmittance → %

BAND TABLE		
cm⁻¹	% A	
3460	52	
2950	70	
1637	87	
1490	65	
1475	64	
1451	63	
1422	55	
1402	71	
1382	57	
1337	73	
1311	46	
1250	57	
1186	52	
1140	40	
1101	30	
1056	60	
1026	35	
1000	21	
909	22	
850	20	
688	32	
657	38	
645	39	

	DEEP	MIDDLE	SHALLOW	OVERLAP
	SILANE	OTHER	OTHER Si-COMPD.	

CHAINS

MULTIPLE BONDS

RING DISPOSITION AND SIZE

	3,4	≥8	7
	5		5·5· ...
	6	6·6·6· ...	6·6

ISOCYCL. RINGS

| | SPIRO | |
| | CYCL. POLYENE | |

HETEROCYCL. RINGS

	O·N IN	O·O· ...
	S·N SAME	S·S· ...
N	N·N· RING	N·N· ...

CHAIN SUBSTITUTION
mono | mono + poly | poly

RING SUBSTITUTION
mono | mono + poly | poly

| 2 | p- | o- | m- |
| 3,4 | asymm. | symm. | vic. |

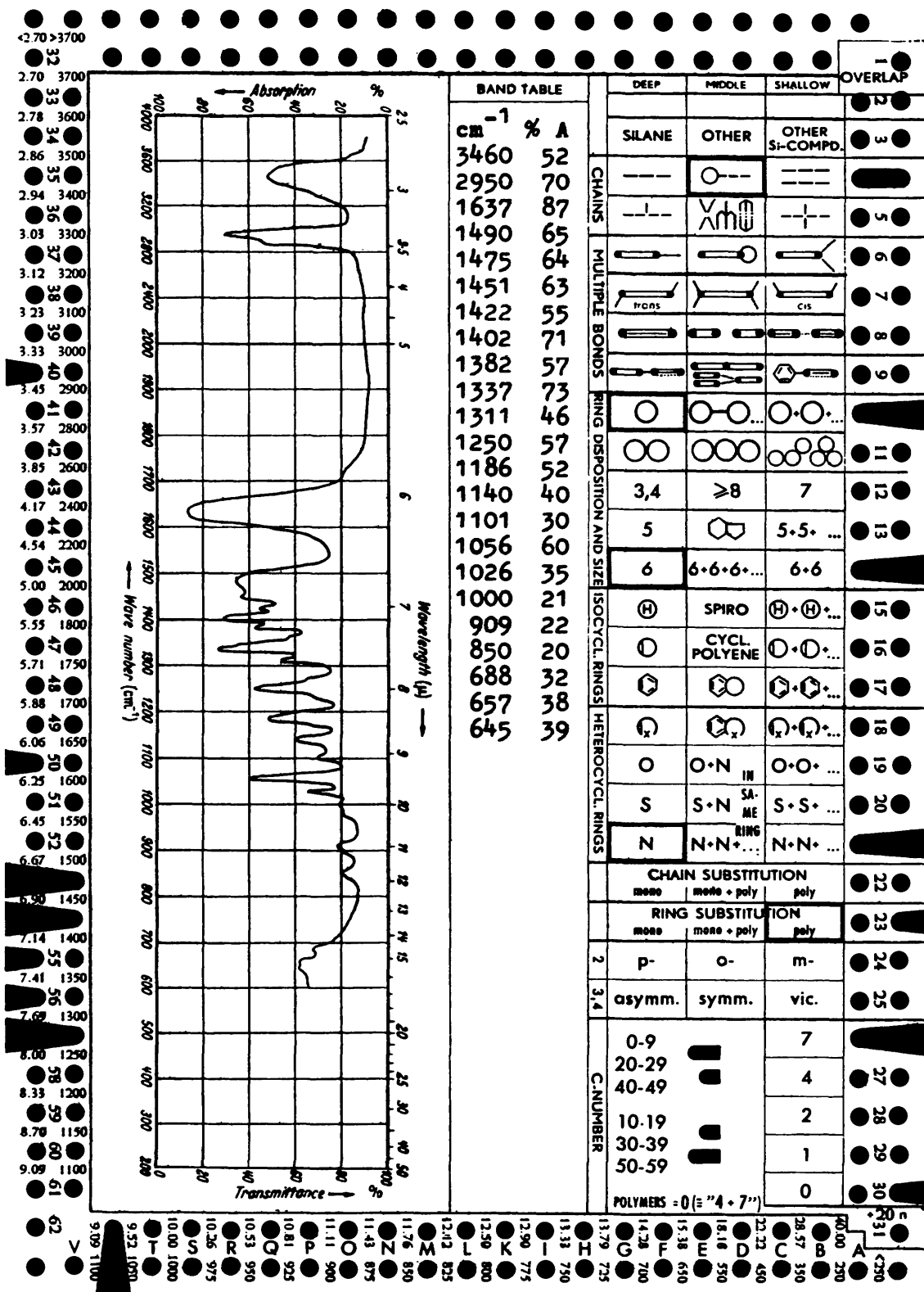| C·NUMBER		
0-9		7
20-29		4
40-49		
10-19		2
30-39		1
50-59		0

POLYMERS = 0 (≡ "4 + 7")

Fig. 2.

bands, up to eight or ten strong bands being marked. The decision as to which bands should be chosen as being the most significant from a spectroscopic standpoint is difficult, since for some purposes the weaker bands may be more informative than the stronger. For this reason, therefore, it is proposed at first to indicate by black markings the holes which should be slotted according to the present system, but not to cut the slots. This will enable any user who prefers to use another system of slotting, which may be more appropriate to his special needs, to do so. The spectral cards are also linked to the literature cards, where relevant, by a defined code.

The code used to record the structural characteristics has been designed to cater for all likely organic compounds, and is therefore necessarily somewhat elaborate. It will be described in detail in a manual supplied to subscribers, and also in a handy abridged form for day-to-day use. It is based upon four structural features of the compound, namely, the number of carbon atoms, the basic skeleton, and the types and positions of substituent groups. The molecule is

Some examples of coding.

d = deep, s = shallow, m = middle slotting.

I, Chloromycetin. II, Strychnine. III, Sulphathiazole.
IV, *isopseudo*Irone. V, *N*-Benzyloxycarbonyl-*O*-D-glycyl-DL-serine hydrochloride.
VI, 3-Cyano-4-ethoxymethyl-2-hydroxy-6-methyl-5-nitropyridine.
VII, Acetic acid. VIII, Tigogenin acetate.

Number of carbon atoms.

Hole no.	26	27	28	29	30	31
I				s	d	
II				d	s	s
III	d		s			
IV		s			d	
V			s	d		
VI	s	d				
VII				d	s	
VIII	d		s			s

Basic skeleton.

Hole no.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
I										d				d			d					s		d	
II	s			m							s	s	s	m	d		d	s	d				d	d	
III									d	s			d	d			d	d		m		d	m	d	
IV	s				d	s	d	m	d	d				d			d					d	d		
V				m						d				d								m	s		
VI				m						d				d				d			d	d	s		
VII				d																		d			
VIII	s			m	m					m	s		s	m	m				s			d	s		

Substituents.

Hole no.	a	b	c	d	e	f	g	h	i	k	l	m	n	o	p	q	r	s	t	u	v
I	s		d		d	d			d				d				d	d			
II													d				m				
III				d			d		d	s	m							d			
IV									m	d	d		d	d		d	d	s			
V									m	d	d	d				d	d	d			
VI			d	d	d					s	d					d		d			
VII													d			d					
VIII						m							d					d			

regarded as being built up from structural units, which have been selected, as far as possible, to possess chemical and spectroscopic significance. For example, the basic skeleton may be characterised (holes 4—21) by the degree of branching or of unsaturation of a carbon chain, by the number and size of rings—fused or unfused—by the presence of hetero-atoms in a ring, or by the
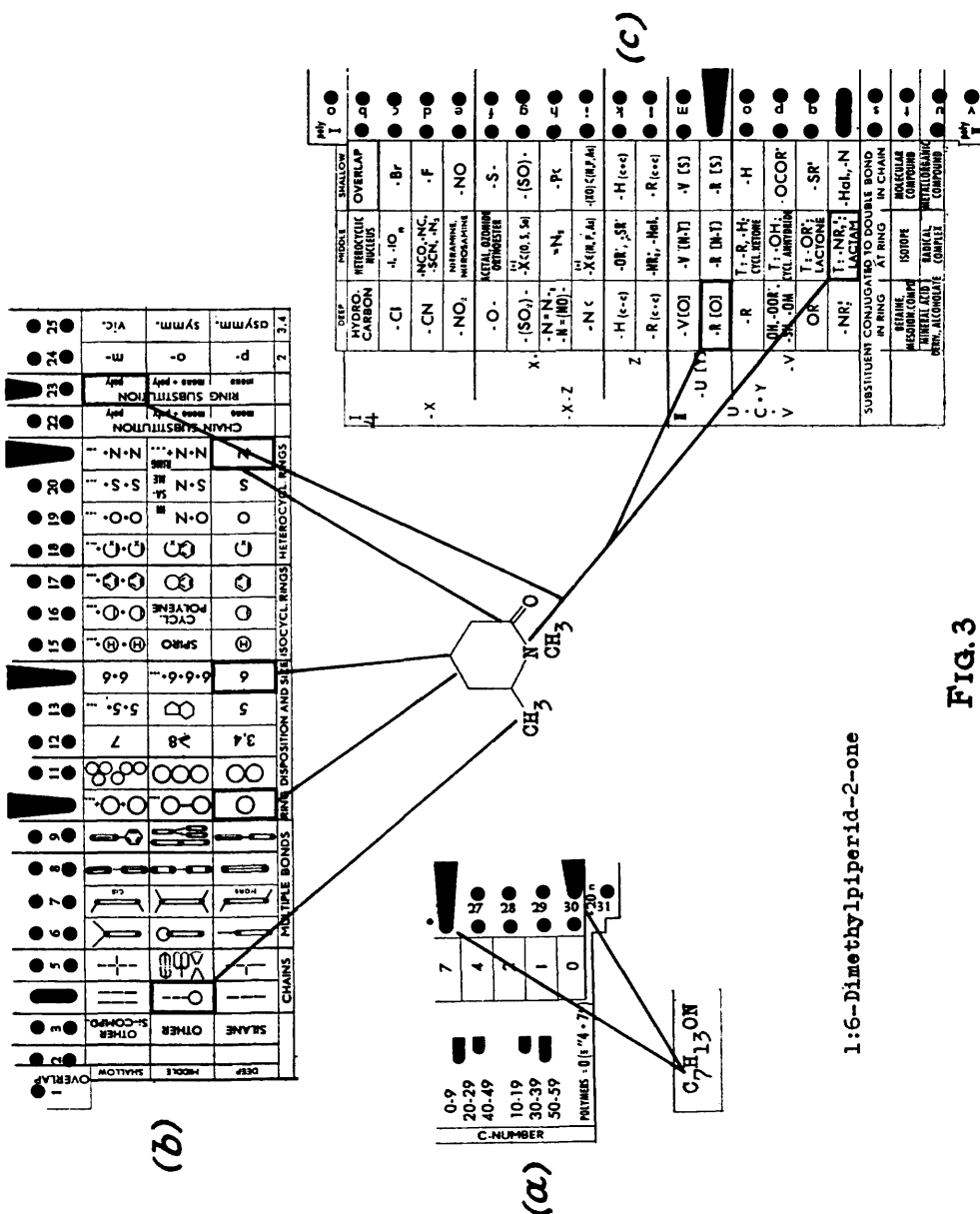
Fig. 3

1:6-Dimethylpiperid-2-one

AUTHOR
I. Gaunt

JOURNAL
J. sci. Instrum. 31. (1954.) 315–18

A simple grating spectrometer for use in analysis.

A simple inexpensive grating spectrograph (2400 lines per inch, blaze at 2.95 μ) is described. The receiver is either an uncooled lead sulphide cell with 400 c.p.s. amplifier or a thermocouple with 10 c.p.s. amplifier. Over-lapping orders are eliminated by means of filters of coloured alkali halide crystals (F centre filters). The characteristics of such filters are given. The instrument is roughly comparable in performance at 6 μ with a rock salt prism spectrometer, but much better at 3 μ. As an example of its use the analysis of D_2O is described. (Diagrams are given for the regions 2.6 - 2.9 μ, and 6.0 - 7.0 μ; the spectrum of HCl near 3.4 μ, and the mercury arc emission spectrum from 1 - 2 μ).

METHOD		SPECTRAL RANGE		SUBSTANCE	
UV	(UV)	2000-3500 Å	< 2000 Å	SOLID	
VISIBLE	(VISIBLE)	3500-9000 Å	0.9-3μ	LIQUID	CRYST.
IR	(IR)	3-3.8μ	0.9-3μ	GASEOUS	EMULSION, DISPERSION
MICRO-WAVE	(MIKRO-WAVE)	3.8-16μ	PARTIAL	SOLVENT	PELLET
RAMAN	(RAMAN)	16-50μ	> 50μ		FILM
LUMI-NESCENCE	EMISSION			PURIFICATION OF SUBSTANCE	SOLVENT
M.S.	(M.S.)				

LIT. NO. 0000

DATE — DECADE: 0 1 2 4 7 — YEAR: 0 1 2 4 7

CHEM. ZENTRAL-BLATT

CHEM. ABSTRACTS 1955 2790 d

LITERATURE: HIST. | GENERAL
REVIEW JOURNAL | BOOK
REVIEW SPECTRA | CONSTIT.
SPECTR. PUBL. | SPECTR. CARD

APPLICATION: ANALYSIS QUAL. | ANALYSIS GEN. | ANALYSIS QUANT. | TEST OF PURITY | MULTICOMP. ANALYSIS | COMPUTER | PLANT CONTROL

NAME

SURNAME . 1ST LETTER | SURNAME . 2ND LETTER | SURNAME . 3RD LETTER | 1ST AUTHOR: 1ST INITIAL

FIG. 4.

SURNAME: 1ST LETTER

2ND LETTER

1ST AUTHOR ▌▌ (DEEP)
2ND AUTHOR ▲▲ (SHALLOW)

3RD LETTER

1ST AUTHOR ▲ (SHALLOW)

1ST INITIAL OF
1ST AUTHOR ▲ (SHALLOW)

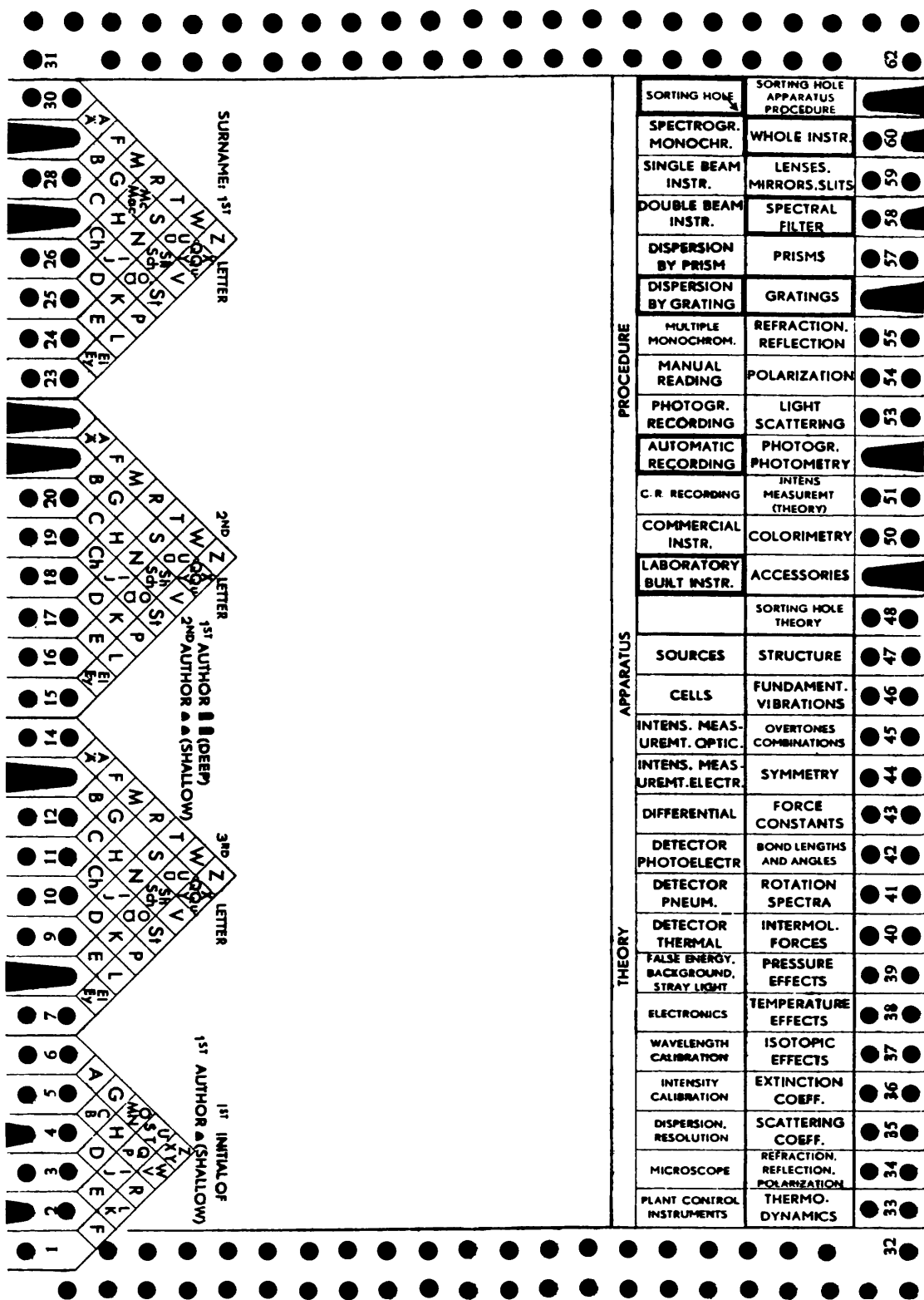|	
SORTING HOLE	SORTING HOLE APPARATUS PROCEDURE
SPECTROGR. MONOCHR.	WHOLE INSTR.
SINGLE BEAM INSTR.	LENSES. MIRRORS. SLITS
DOUBLE BEAM INSTR.	SPECTRAL FILTER
DISPERSION BY PRISM	PRISMS
DISPERSION BY GRATING	GRATINGS
MULTIPLE MONOCHROM.	REFRACTION. REFLECTION
MANUAL READING	POLARIZATION
PHOTOGR. RECORDING	LIGHT SCATTERING
AUTOMATIC RECORDING	PHOTOGR. PHOTOMETRY
C. R. RECORDING	INTENS MEASUREMT (THEORY)
COMMERCIAL INSTR.	COLORIMETRY
LABORATORY BUILT INSTR.	ACCESSORIES
	SORTING HOLE THEORY
SOURCES	STRUCTURE
CELLS	FUNDAMENT. VIBRATIONS
INTENS. MEAS-UREMT. OPTIC.	OVERTONES COMBINATIONS
INTENS. MEAS-UREMT. ELECTR.	SYMMETRY
DIFFERENTIAL	FORCE CONSTANTS
DETECTOR PHOTOELECTR	BOND LENGTHS AND ANGLES
DETECTOR PNEUM.	ROTATION SPECTRA
DETECTOR THERMAL	INTERMOL. FORCES
FALSE ENERGY. BACKGROUND. STRAY LIGHT	PRESSURE EFFECTS
ELECTRONICS	TEMPERATURE EFFECTS
WAVELENGTH CALIBRATION	ISOTOPIC EFFECTS
INTENSITY CALIBRATION	EXTINCTION COEFF.
DISPERSION. RESOLUTION	SCATTERING COEFF.
MICROSCOPE	REFRACTION. REFLECTION. POLARIZATION
PLANT CONTROL INSTRUMENTS	THERMO. DYNAMICS

PROCEDURE

APPARATUS

THEORY

FIG. 5.

presence of exocyclic double bonds. The positions of substituent groups (holes 22—24) are indicated, for example, for the various possible arrangements around a ring system, and for *cis–trans*-isomers. Designation of the nature of substituent groups (holes a—s) requires a rather more intricate coding system, especially for dealing with aliphatic structures. It has been found convenient to distinguish between three structural types, namely $-X$, $-X-Z$, and $U-C-V$
$$\overset{\|}{Y}$$
and by specifying the groups X, Y, Z, U, and V to characterise all classes of compound, or parts of a compound. The cards also make provision for recording other classes such as molecular compounds, radicals, isotopic species, and organometallic compounds. The use of three kinds of slot—deep, shallow, and middle—makes it necessary to follow certain rules and priorities in sorting. Thus, deep slots have the highest priority, and middle slots the lowest. The most important and commonly occurring groupings have been allocated to the deep slots.

Fig. 3 shows the coding of a card for 1 : 6-dimethylpiperid-2-one, (*a*) shows the number of carbon atoms, (*b*) the basic skeleton, and (*c*) the substituents. The annexed Table shows how the coding is applied to a number of compounds belonging to very different classes. The code has been tested as a whole on a very large number of substances, and will certainly prove satisfactory. However, it is possible that some users will prefer for special reasons to employ their own code, or make some modification of the present one. For this reason too, the holes which should be slotted will be marked on all cards, but not cut at present.

Certain substances cannot be classified completely by the present code, sometimes because their structure is indefinite. Among these, for example, are a number of natural substances with trivial names, certain mixed polymers and condensates, and technical solvents. In such cases a card will be used on which the code to the basic skeleton (1—25) is replaced by an alphabetical key, the remainder of the card being unaltered.

(*b*) *Literature cards.* The L cards, the front and back of which are shown in Figs. 4 and 5, may relate either to spectra recorded on S cards, or to important developments of technique or theory. Where relevant, the S and the L card are linked together by a cross-reference. On the front face of the L cards there is a short abstract of the publication concerned, with full references. The holes along the edges are used to record names of authors, year of publication, an indication of the nature of the work (*e.g.*, analysis), the spectral range studied, the state of aggregation of the sample, and details of the equipment, method, or theoretical aspects discussed in the paper. By appropriate slotting, it is possible to select a card or cards dealing with specific matters of these kinds.

(*c*) *Blank cards.* The above code is of general value for the classification of organic compounds on punched cards and would naturally be applicable to the recording of other molecular properties. Also, some spectroscopists may wish to introduce into the present collection of spectral cards certain other data of their own. For this purpose, arrangements are being made to furnish subscribers with blank cards printed with the structural code but otherwise carrying no specific data. For instance, there may be cases where very small shifts of a particular band, say of a carbonyl group in a series of compounds, are highly characteristic and important in some laboratory, and such small shifts may be too slight to be differentiated on the standard cards. In such cases a refined set could easily be prepared.