known as *STAR*. The periodical published literature is covered in AIAA's *International Aerospace Abstracts*. Both *STAR* and *IAA* are issued semimonthly on alternate weeks. Each issue is arranged in two major sections: the first presents informative abstracts; the second includes the subject, corporate source, author, and report number indexes. In addition to the indexes in each semimonthly issue, cumulative indexes are distributed which integrate and cumulate the indexes quarterly, semiannually, and annually. Both *STAR* and *IAA* follow the same format, the same abstracting and indexing techniques, and the same indexing terms to provide the individual user with compatible bench tools.

In addition to these two compatible services, NASA also assists in the support of the Aerospace Medicine Bibliography Project at the Library of Congress, the publication of abstracts in the *Journal of Aerospace Medicine*, and the relevant abstracts in *Meteorological and Geoastrophysical Abstracts*. The interest here is in complementing or supplementing the aerospace coverage with abstracting and indexing in depth by experts in the subject matter field itself—the "identifiable area" approach.

Prior to each issuance of *STAR*, the NASA system provides for automatic distribution of the reports covered in the journal to local centers where they can be available to the individual user as he identifies them in the announcement journal. The distribution is either in macroform (that is, full-sized printed copies) for reports published by NASA, and in microform for NASA and non-NASA reports covered, except those subject to reproduction limitations.

The next step is the forthcoming implementation and evaluation within the NASA complex of a system for selectively disseminating report literature. This involves the matching of profiles of user needs and interests with profiles of report content. Such a system does not presume to "prepackage" what information it will provide, it does not attempt to impose a classification or indexing system beyond what is needed by the information service to retrieve the documents in its collections, and it is primarily determined by the characteristics of the user.

Certainly much of what has been described in this paper is not unique. What is regarded as significant, however, is that the whole system is deliberately geared to current awareness and local access by the individual scientist and engineer. The emphasis is on helping the individual user to identify what is relevant to him so that he may project and extrapolate as befits the subject matter expert.

And this is appropriate to servicing such an area as Bioastronautics. An information service must gear itself to the characteristics of the area to be served lest it find itself in the same relation to its users as science fiction has to the products of sober research.

Searching X-Ray Diffraction Powder Data with an Inverted Coordinate Index*

By F. W. MATTHEWS
Central Research Laboratory, Canadian Industries, Ltd.,
McMasterville, Quebec, Canada
Received April 22, 1963

The X-ray diffraction powder method of identification of solid substances has been described as a fingerprint method in that each substance in the solid state gives a unique pattern. As with the fingerprint method of identification, it depends on matching the pattern of the unknown with established data. This use of the powder diffraction pattern was clearly envisaged by Hull, who in the *Journal of the American Chemical Society* for the year 1919[10] described this use of the pattern with examples of how it could be employed. However, the method was not applied for lack of data which was in part due to the lack of hardware to record the patterns conveniently. It took almost 20 years to overcome these hurdles. The method took on practical importance with the publication by Hanawalt and Rinn[9] of the Dow Chemical Company in 1936. These authors realized that application of the method was dependent on setting up a convenient method for searching data. They described a ledger type of index in which data was entered as it accumulated. Their publication in 1938 of diffraction data on 1000 substances initiated the use of the method on a practical basis.

These data have grown under the sponsorship of a Joint Committee of The American Society for Testing Materials, the American Crystallographic Society, and the British Institute of Physics so that now over 10,000 substances have been recorded, and the file is growing at the rate of about a thousand patterns a year. Cooperative schemes, which are sponsored by the Joint Committee, for collecting and producing data are in operation in several countries. The method found most ready acceptance by mineralogists and metallurgists to whom crystallography was a basic subject, and it has been extended to wide coverage of inorganic chemicals and the more common organic substances, which now number about 5000.

* Presented before the Division of Chemical Literature, 142nd ACS National Meeting, Atlantic City, N. J., September 12, 1962.

As the data have grown, the problems of searching have become more difficult, and a number of schemes have evolved to meet the problem at various stages. The first published index was on 3 × 5-in. cards with three cards for each entry. The three entries were based on the three strongest lines of the pattern. This is illustrated in Fig. 1 which shows a typical powder diffraction pattern and the numerical measurements derived from it. These are the interplanar or d spacing of the crystal planes of the substances measured in angstroms and the relative intensity of the diffracted beams from these planes (I/I_1.). In actual practice, the d spacings, as measured from an unknown, may differ from that recorded due to error in measurement or solid solution effects, and the relative intensity can be considered only as a guide and not as a requirement. Some information on qualitative chemical composition of the unknown may be available and, if so, is very useful in guiding the search.

Since some X-ray diffraction workers were not well schooled in retrieval methods, they objected to examining 3 × 5 cards, drawer by drawer, and soon came to the conclusions that there must be a better way. This thinking led to a Keysort edition of the cards[13] followed by IBM card editions.[12] The use of typed moveable line index strips (Linadex) to prepare an annual edition of the up-dated book index has proved effective for 15 years, but more automated methods will now be employed. As the volume of data grew there was increasing need for an improved index which could enable the user to combine the diffraction data with knowledge of chemical composition without using expansive equipment and which could be sold and up-dated at reasonable cost.

The work on Peek-a-Boo Indexing [3, 4, 6, 7, 12, 16, 17] suggested that this could be an answer. Two experimental versions of the X-ray powder diffraction data index were set up using IBM cards as an optical coincidence index.[15] These showed considerable promise, but progress was stopped for four years until a large capacity card became available which could be reproduced in multiple copies. The Termatrex System of the Jonker Business Machine Co.[11] provided an answer to these difficulties, and an index using Termatrex cards is now available for some 6000 inorganic compounds listed in the X-Ray Data File.

A typical card is shown in Fig. 2. This marks with a hole those compounds having in their diffraction patterns a strong line in the range indicated. It is an inverted index in which the title of the card is a "descriptor" or a characteristic, and all items, *i.e.*, diffraction patterns having that characteristic, are recorded on the card. Holes are punched in spaces reserved for or dedicated to the compounds. A numerical key gives the name of the compound and the serial number of the diffraction pattern for each hole number. When cards are superimposed on a lighted background, patterns having characteristics in common are identified by the coincidence of holes.

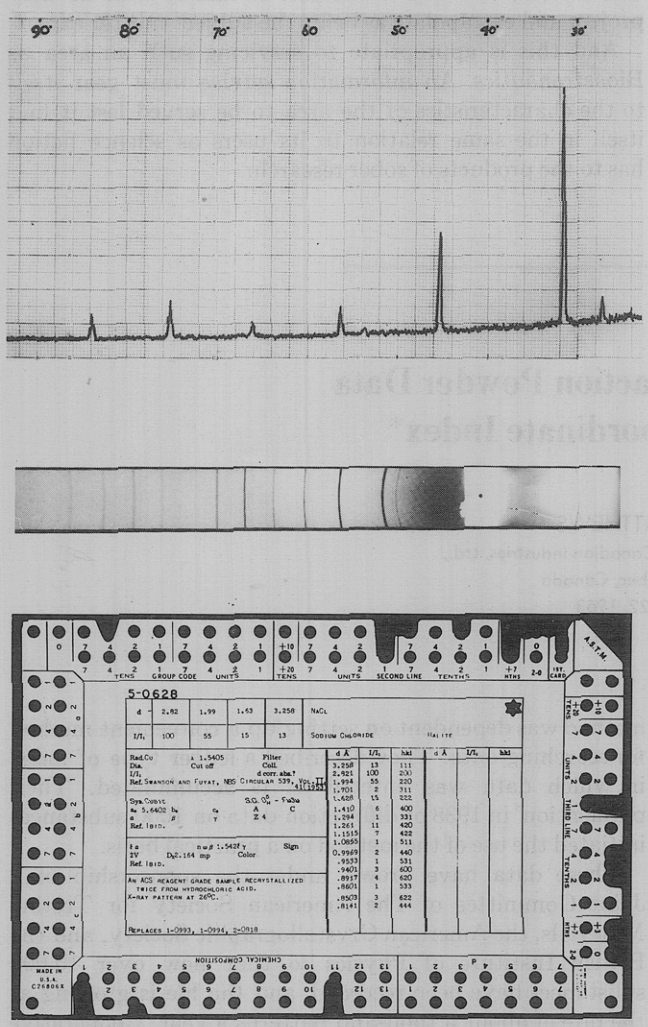

Fig. 1.—X-Ray defractometer recording of pattern from sodium chloride (top); X-ray diffraction powder camera record of the same pattern (middle); X-ray data file card for sodium chloride (bottom) giving d spacing (d Å.) and relative intensity $1/1_1$. of the above pattern. The three most intense and the innermost lines are given in the upper left corner of the card.

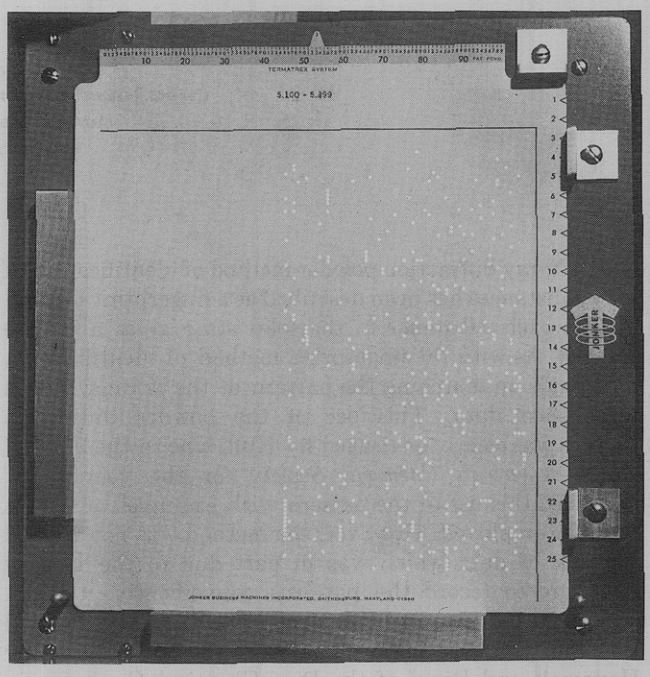

Fig. 2.—Termatrex card on which all compounds having a strong line in the range d = 5.100 to d = 5.399 have been recorded by drilling a hole in the space on the card allocated to that compound.

A number of decisions had to be made in setting up the index for which only limited experience was available as a guide. The range of d values, from 1.5 to 10 Å., was divided into 50 segments, based on the earlier experiments of X-ray powder patterns. They were chosen to give an approximately equal distribution of substances in each segment to have equal loading on each "descriptor" card used in the index. This will be recognized as a major advantage which cannot usually be arranged in information retrieval. It makes more effective an analysis of the system based on probability theory. The extent to which our first attempt accomplished this is shown in Table I. Using these data, a second edition can be prepared which will give a more even loading of the descriptors, but in choosing these the probable error in measurement of d has to be taken into account.

Table I. The Coordinate Index for X-Ray Powder Data
Distribution of Substances by d group

d values	No. of substances	d values	No. of substances
<1.5	110		
1.500–1.599	850	3.440–3.519	525
1.600–1.699	1000	3.520–3.599	410
1.700–1.799	970	3.600–3.679	335
1.800–1.899	1095	3.680–3.759	325
1.900–1.999	955	3.760–3.839	255
2.000–2.079	885	3.840–3.919	315
2.080–2.149	900	3.920–3.999	230
2.150–2.219	810	4.000–4.099	286
2.220–2.289	645	4.100–4.199	199
2.290–2.359	665	4.200–4.299	269
2.360–2.429	695	4.300–4.399	266
2.430–2.499	785	4.400–4.499	237
2.500–2.569	840	4.500–4.649	232
2.570–2.639	630	4.650–4.799	247
2.640–2.709	800	4.800–4.999	296
2.710–2.779	715	5.000–5.299	323
2.780–2.849	770	5.300–5.699	382
2.850–2.919	635	5.700–5.999	232
2.920–2.999	910	6.000–6.399	225
3.000–3.059	640	6.400–6.899	237
3.060–3.139	810	6.900–7.499	245
3.140–3.209	660	7.500–8.499	226
3.210–3.279	560	8.500–9.999	210
3.280–3.359	545	10.000+	216
3.360–3.439	520		

A critical decision was how many lines or "descriptors" should be used for each pattern. The more descriptors per pattern in the index, the higher the probability that a line chosen from the pattern of the unknown will be recorded in the index. However, it is also true that the more lines per pattern the less effective any given line is in searching the index.

Probability theory can be applied in the following manner. Given a system with 10,000 compounds and using 50 descriptor cards, let h equal the number of holes per compound, then the number of holes is $10{,}000h$ and the average number of holes/card, $(10{,}000h/50) = 200h$. The probability that a given hole will be punched in a given card is

$$p = \frac{200h}{10{,}000} = \frac{h}{20}$$

The probability that n of the cards will all have the same hole punched is

$$p_n = \left(\frac{h}{50}\right)^n$$

and the mean minimum number of coincidences is

$$c = 10{,}000\left(\frac{h}{50}\right)^n$$

If we wish c to equal one, *i.e.*, to obtain on the average at least one answer then

$$\frac{1}{10{,}000} = \left(\frac{h}{50}\right)^n$$

$$n = \frac{4}{\log 50 - \log h}$$

The plot of n, the number of cards to give on the average at least a single answer, against h, the number of descriptors/compound, is shown as an "S" curve in Fig. 3.

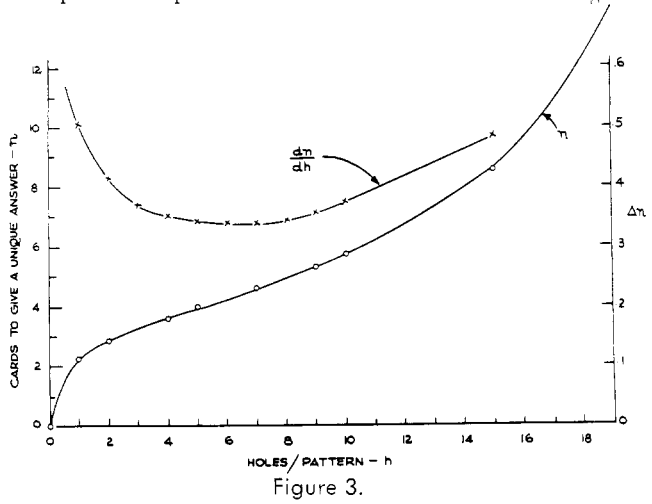

Figure 3.

This gives infinitely small answers when $h = 0$ where no cards have any holes, and infinitely large when $h = 50$ where all holes in all cards are punched. In between is an inflection point in the range of five holes per compound. This is the region of practical importance. By plotting the first derivative

$$\frac{dn}{dh} = \frac{4}{n(\log 50 - \log h)^2}$$

the second curve in Fig. 3 is obtained. This shows a flat region around the minimum. By equating the second derivative to zero the minimum value of $h = 6.8$ is found. The significance of this inflection point is not certain, but it may indicate that under these circumstances the optimum number of holes per compound is in the region of 6.8. In this index it was decided to use five to six lines per compound, without the aid of probability theory. These lines were chosen as the most intense between the range of 1.5 to 10 Å. A more general treatment of probability theory as applied to this problem is in preparation with the assistance of S. H. Storey and F. van Zeggeren who were mathematical consultants for this work.

The question of search on either side of the measured *d* value was effectively answered by use of overlapping descriptors as shown in Fig. 4. The range of values included in each card in the first set A was shifted by half in the second set B so that a descriptor is always available in which values above and below the measured values are included. As shown in Fig. 4, using both the A and B sets, ranges having half the coverage of the original sets are presented.

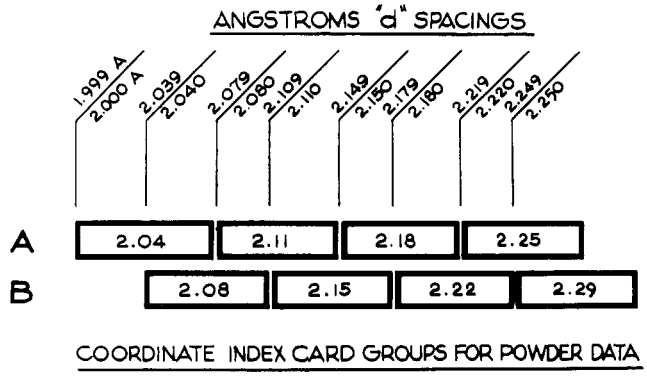

Figure 4.

Using the elements as descriptors for chemical composition can be most helpful in guiding a search by combining chemical information with the numerical data. Similarly negative data, *e.g.*, "there is no calcium or sodium in the unknown" can be applied very effectively using negative cards for the more common elements. Also, negative cards for *d* values indicating the absence of lines in a given range can be prepared. The chemical composition cards and their negative counterparts are most useful in identifying components of mixtures. Using "random access" colored tabs the refiling of cards is eliminated and the use of the index is thereby greatly facilitated.

Holes in colored transparent sheets are used to record the presence of oxygen (yellow sheets) and hydrogen (blue). When these are superimposed, white spots indicate that both oxygen and hydrogen are present; blue spots indicate oxygen but no hydrogen is present; yellow spots indicate hydrogen but no oxygen; and green spots indicate that neither is present. The green color results from superposition of the yellow and blue transparent sheets.

X-Ray diffraction powder data have proved of particular value in the identification of alloys and of minerals. Hence, cards separating these data have been included. A card for hydrates also has been prepared.

A study of the application of this technique to infrared absorption data is already underway,[2] and a system of this type is now available in Europe.[5]

REFERENCES

(1) *Chem. Eng. News*, **40**, 84, No. 22 (1962).
(2) American Society for Testing Materials, 1916 Race Street, Philadelphia, Pa.
(3) W. E. Batten, Report of the Proceedings of the 22nd Conference, ASLIB, London, 1947.
(4) M. G. Cordonnier, Revue Mensuelle de l'Organization, Avril-Juillett, 57 rue de la Babylone, Paris, 1951.
(5) "Documentation of Molecular Spectroscopy," Butterworth Scientific Publications, London, 1962.
(6) J. D. H. Donnay, *Am. Mineralogist*, **23**, 91 (1938).
(7) C. J. Gray, *Trans. Geol. Soc., S. Africa*, **23**, 114 (1921).
(8) J. D. Hanawalt, H. W. Rinn, and L. K. Frevel, *Ind. Eng. Chem. Anal. Ed.*, **10**, 457 (1938).
(9) J. D. Hanawalt and H. W. Rinn, *ibid.*, **8**, 244 (1936).
(10) A. W. Hull, *J. Am. Chem. Soc.*, **41**, 1168 (1919).
(11) Jonker Business Machines, Gaithersburg, Md.
(12) L. E. Kuentzel, Wyandotte Chemical Corp., Wyandotte, Mich., 1951.
(13) H. P. Luhn, U. S. Patent 2,011,722 (1933).
(14) F. W. Matthews, *Anal. Chem.*, **21**, 1172 (1949).
(15) F. W. Matthews, *Mater. Res. and Std.*, **2**, 643 (1962).
(16) H. Taylor, U. S. Patent 1,165,465 (1915).
(17) W. A. Wildhack and J. Stern, "Information Systems in Documentation," Vol. II, Interscience Publishers, New York, N. Y., 1957, p. 209.

Government Services for Technical Information*

By HAROLD WOOSTER
Air Force Office of Scientific Research, Office of Aerospace Research, Washington 25, D. C.
Received April 11, 1963

The theme of the following four papers might well be: "Do not ask what you can do for your government. Ask what your government can do for you." Last year the Federal government spent some $3000 million on scientific information activities concerned with the billions of dollars spent on research and development. Federal agencies are both producers and consumers of scientific information. By definition, at least in the Air Force, anything having to do with basic research is unclassified. The same holds true for almost all of our applied research program. The primary responsibility of any Federal agency is to spend its resources wisely in accomplishment of its mission; certainly a strong secondary responsibility is to ensure that useful information paid for by public funds is available to that selfsame public.

There are certain difficulties in reducing this principle to practice. One of these is certainly the simple effect

* Presented before the Division of Chemical Literature, 142nd National Meeting of the American Chemical Society, Atlantic City, N. J., September 12, 1962.