

Computer-aided X-ray diffraction identification of minerals in mixtures

ALLEN F. GLAZNER¹ AND DONALD B. MCINTYRE

*Geology Department, Seaver Laboratory, Pomona College
Claremont, California 91711*

Abstract

Minerals are commonly identified by searching a file of standard X-ray powder diffraction patterns for a match with the pattern of the unknown. Such searching is generally done by hand, but the APL language, now available on most computers, allows direct expression and execution of these search procedures. Two features of APL make solution of the search problem particularly simple. The membership function, ϵ , permits immediate comparison between an entire file of standards and the unknown; consequently mixtures are matched as easily as single phases. APL also allows the analyst to control the tolerance used by the computer in making comparisons. We have implemented a search program on a desk-top IBM 5100 computer, which students use routinely in an elementary mineralogy class at Pomona College. Although the entire JCPDS file can be searched, we illustrate the logic by use of a file small enough that each step can be performed by hand. The method demonstrated its effectiveness by successful identification of the minerals in a test mixture whose 2θ values were supplied by the Computer Subcommittee of JCPDS.

Introduction

The identification of mineral mixtures by X-ray powder diffraction has long been important to geologists. Identification is essentially a 2-step process. The first step is the "search," in which a list of possible phases is selected by comparing the pattern of the unknown to a large set of reference mineral patterns. The second step is the "match," in which a final mineral assemblage is chosen through considerations of peak intensities, sample chemistry, mineral associations, and so on. The computer method presented here performs a "search," leaving the analyst to make the final "match."

When comparing an unknown peak with a standard peak, we must decide how far the two can diverge (in both angle and intensity) and still be equated. Fortunately, APL allows us to set the comparison tolerance used by the computer in comparing two numbers (see appendix). Thus, if we set the comparison tolerance for 2θ to 0.1 degrees, two peaks will be judged equal if the absolute value of their difference does not exceed 0.1 degrees. The existence of this feature greatly simplifies the search method. In the

examples that follow, the comparison tolerance is set to 0.1 degrees.

The notation of symbolic logic allows us to express concisely and unambiguously the operations we wish to perform, and the APL language (Iverson, 1973; McIntyre, 1969; Polivka and Pakin, 1975) permits execution of these expressions. The result is an effective search program of great simplicity, which can be easily adapted to meet special requirements. Search programs written in other languages are likely to be much larger and more complex, requiring considerable knowledge of computer systems even to use, and still more to understand and modify. For example, the very comprehensive Fortran search/match program written by Gerald G. Johnson, Jr., and distributed by the Joint Committee on X-ray Powder Diffraction Standards (JCPDS), requires a very large machine, such as an IBM 360/370 under OS with a Fortran IV H-level compiler and a work space of 220,000 bytes.

Method

The file of standards may be the complete file of 27,000 patterns published by JCPDS, the subfile of 3,000 minerals, or any collection of patterns chosen because of their special interest to the investigator.

¹ Present address: Department of Earth and Space Sciences, University of California, Los Angeles, California 90024.

Although we commonly use the entire JCPDS mineral file, to illustrate the method it is sufficient to consider a file consisting of six minerals, each with the 2 θ values of its five strongest peaks ordered in decreasing intensity:

ANDESINE	27.79	28.05	22.00	23.66	24.38
BIOTITE	8.75	26.44	33.69	36.67	41.41
HORNBLENDE	28.88	26.36	27.09	10.46	9.79
OLIGOCLASE	28.05	22.05	27.88	23.66	30.50
ORTHOCLASE	28.05	22.10	23.40	26.76	29.77
QUARTZ	26.66	20.84	50.20	36.55	39.48

Suppose that peaks at the following values of 2 θ have been recorded for the sample, there being no significance to the order in which they are given: 28.1 22.1 23.4 26.8 29.8 23.7 30.5 26.4 8.8 19.9 26.5. Let the numeric part of the file of standard minerals be named F , so that F is a matrix (table) with 6 rows and 5 columns. Let the peaks of the unknown sample be named X , so that X is a vector (string) with 11 elements.

Which peaks of the file F are present in the sample X ? To answer the question we must compare the sample with each pattern in the file. The question is stated formally as $F \epsilon X$. The symbol ϵ , denoting membership, is an abbreviation of the Greek $\epsilon\sigma\tau\iota$ for "is a member of." This notation, long used in formal logic, has been incorporated into the extended algebra known as the APL language and can be executed on most computers.

The function ϵ checks to see if the elements of its left argument are present in its right argument. For example, to see if the strongest peak of andesine (27.79) is present (within tolerance) in the unknown, we execute

27.79 ϵX

0

The computer's response, 0, indicates that the strongest andesine peak is not present. Checking for the second strongest peak:

28.05 ϵX

1

we find it present. We can also check for both simultaneously:

27.79 28.05 ϵX

0 1

In fact the arguments of ϵ can be any shape. The result will be an array of the same shape as the left

argument, with a 1 in a given position if the corresponding element is present in the right argument, and a 0 if it is not. To illustrate:

$F \epsilon X$					
0	1	1	1	0	
1	1	0	0	0	
0	1	0	0	0	
1	1	0	1	1	
1	1	1	1	1	
0	0	0	0	0	

The entire search procedure is accomplished by the simple expression $F \epsilon X$, which compares every peak in the file with each peak in the sample. For example, the above matrix shows that no quartz peak (last row) is recognized in the sample, but peaks corresponding to all 5 of the strongest peaks of orthoclase (second last row) are present.

An obvious criterion for goodness of match is given by the sum of the values in each row of $F \epsilon X$. First assign the name A to the matrix $F \epsilon X$; then assign the name B to the sum along the rows of A and display the result:

$A \leftarrow F \epsilon X$

$B \leftarrow +/A$

B

3 2 1 4 5 0

To find those minerals in the file that match at least two peaks in the unknown, we create the logical vector

$B \geq 2$

1 1 0 1 1 0

and use it to compress $NAMES$, the table of mineral names corresponding to the file F . The symbol \neq indicates that certain rows, rather than columns, are to be selected by the logical vector.

$(B \geq 2) \neq NAMES$

ANDESINE

BIOTITE

OLIGOCLASE

ORTHOCLASE

These are the minerals that pass the first criterion for a successful match. Another criterion, by which the list could be further reduced, is to require that the strongest peaks of a standard be matched before that standard qualifies. This is achieved by creating a new matrix from A :

			$\wedge \setminus A$	
0	0	0	0	0
1	1	0	0	0
0	0	0	0	0
1	1	0	0	0
1	1	1	1	1
0	0	0	0	0

The symbols $\wedge \setminus$, or *and scan*, denote that each row is scanned for the appearance of the first 0, after which all remaining values are set to zero. If we now sum along the rows of this new matrix, we find for each standard the number of matched peaks that occur before the first miss.

			$C \leftarrow +/\wedge \setminus A$		
			C		
0	2	0	2	5	0

The selection vector for those standards whose two strongest peaks match the sample is

			$C \geq 2$		
0	1	0	1	1	0

To find those minerals that have the right total number of matched peaks *and* at the same time match the requisite number of strong peaks before the first miss, we use the logical *and* (\wedge) to combine the two selection vectors:

			$D \leftarrow (C \geq 2) \wedge (B \geq 2)$		
			D		
0	1	0	1	1	0

$D \neq \text{NAMES}$

BIOTITE
OLIGOCLASE
ORTHOCLASE

Andesine is eliminated because, although 3 of its peaks are matched in the sample, its strongest peak is not found.

We now gather these separate steps into the formal definition (Iverson, 1976, p. 146-148) of a single executable function, named *SEARCH*:

$\text{SEARCH: } ((\alpha[2] \leq +/\wedge \setminus A) \wedge \alpha[1] \leq +/A \leftarrow F \epsilon \omega) \neq \text{NAMES}$

The left argument, α , of this function is a vector of 2 elements defining respectively the minimum number of matches and the "strong peaks" criteria. The right argument, ω , is the vector of peaks recorded in the sample to be identified. For example, to search the file for all minerals that match at least 3 peaks, including the two strongest, we execute

3 2 *SEARCH X*
OLIGOCLASE
ORTHOCLASE

These minerals are clearly strong candidates for acceptable matches, but what about observed peaks not accounted for by them? Because this simple example is restricted to 5 peaks for each standard, it is possible that the unaccounted peaks belong to the matched minerals but are those of lower intensity, but in practice a larger number of peaks would be used to define the standards. To get a list of minerals, each containing at least one of the unaccounted peaks, we first create the subset of peaks corresponding to the minerals that have met the criteria:

$$Q \leftarrow ((C \geq 2) \wedge B \geq 3) \neq F$$

The peaks unaccounted for are then

$$(\sim X \epsilon Q)/X$$

(where \sim is the symbol for the logical *not*), and the inclusive *or* (\vee) of logic forms the vector that selects the names of minerals with at least one of these unaccounted peaks:

$$(\vee / F \epsilon (\sim X \epsilon Q)/X) \neq \text{NAMES}$$

BIOTITE
HORNBLLENDE

Any remaining peaks are not in the file, and it is a simple matter to isolate them

$$(\sim X \epsilon F)/X$$

19.9

This is the only peak observed in the sample that cannot be matched in the file. Either the file is incomplete or an error has been made in recording this peak. The judgment of the analyst is needed to resolve the problem.

In this discussion we have emphasized the formal description of a matching algorithm, without regard to efficiency of machine execution. But if no further refinements are included, the computer will make many unnecessary comparisons. For a file of the 10 strongest peaks for 100 reference minerals and an unknown defined by 25 peaks, *SEARCH* must make $10 \times 100 \times 25 = 25,000$ comparisons. If we choose to consider only minerals whose strongest peak is matched, then most of these comparisons are unnecessary. For example, if we require that the strongest peak be matched, then $G \leftarrow (F[;1] \epsilon X) \neq X$ will produce a reduced file G containing only minerals meeting this criterion. This process can be applied to

Table 1. Sample output for search NBS mixture #4A

PDF #	Name	Formula	N	S	Matches	GOF
8-449	Smithsonite	ZnCO ₃	10	10	*****	100
5-453	Aragonite	CaCO ₃	8	8	*****	99
13-192	Vaterite	CaCO ₃	8	6	***** *	98
5-586	Calcite	CaCO ₃	7	6	***** *	98
16-707	Schreibersite	(Fe,Ni) ₃ P	6	4	**** **	93
11-692	Spherocobaltite	CoCO ₃	9	3	*** *****	92

N: total number of peaks matched
 S: number of strong peaks matched before the first miss
 GOF: peaks are assigned exponentially decreasing weights;
 GOF (for goodness of fit) is the sum of the weights
 of matched peaks

succeeding columns (*i.e.* $G \leftarrow (G[;2] \epsilon X) \neq G$) as many times as the "strong peaks" criterion requires. The resulting reduced file G can then be used by *SEARCH* in the usual manner. This simple modification can cut the number of comparisons by a factor as large as 10.

Test of the method

Through the courtesy of Ron Jenkins, chairman of the Computer Subcommittee of JCPDS, we were able to test our search method as participants in the second round-robin test of computer search/match procedures conducted by JCPDS², the results of which will be published by the National Bureau of Standards. Mixture #4A was described by the positions and intensities of 59 peaks. Our file consisted of the 10 strongest peaks of 1,500 minerals abstracted from the JCPDS file, sets 1-16. With the 2θ comparison tolerance set to 0.1 degrees, we searched the file using the method described in this paper, which considers peak positions only. Six minerals satisfied the conditions that at least 6 of their 10 strongest peaks, including their 3 strongest peaks, be present in the sample (Table 1). From these 6 minerals the analyst can make a choice based upon his knowledge of chemistry, optics, provenance, or other characteristics of the sample. For example, schreibersite occurs only in meteorites and slags, and simple spectrographic tests

² For information write to Dr. Camden R. Hubbard, A221 MATL, National Bureau of Standards, Washington, D.C. 20234.

Appendix

On the IBM 5100, comparison tolerance is controlled by the system variable $\square CT$. Two numbers α and ω will be judged unequal if the absolute value of their difference exceeds the relative "fuzz" defined by $RF: \square CT \times 16 * \lfloor 16 * (\alpha) \rfloor \omega$. Relative fuzz is, therefore, constant between successive powers of 16. If we transform our data by adding an arbitrary number, say 20, so that all data lie between $16 * 1$ and $16 * 2$ then the above expression simplifies to $RF: \square CT \times 16$. Thus the fuzz can be set to an appropriate value (e.g. 0.1) by executing $\square CT + 0.1 \div 16$.

There is presently a controversy over implementation of $\square CT$ in different systems. Larry Breed of IBM Corporation (personal communication, 1978) has recommended that a defined membership function be used in place of ϵ . We suggest either

$$EPS: \vee / (\omega \circ . \leq \alpha + \bar{\epsilon}) \wedge (\omega + , \omega) \circ . \geq \alpha - \bar{\epsilon}$$

or $EPS: \vee / \bar{\epsilon} \geq | \alpha \circ . - , \omega$

where $\bar{\epsilon}$ is a global variable defining the fuzz.

can detect the presence of Zn, Ca, Fe, P, Ni, or Co. The National Bureau of Standards has reported that sample #4A is a mixture of calcite, aragonite, vaterite, and smithsonite, thus showing the power of the method.

Acknowledgments

Ervin Frazier prepared an early version of the mineral subfile from the JCPDS tape. Eric Dowty, C. R. Hubbard, K. E. Iverson, Ron Jenkins, G. G. Johnson, Jr., D. L. Orth, A. O. Woodford, and T. L. Wright read and commented on a draft of the manuscript. Ron Jenkins and C. R. Hubbard made possible our participation in the second round-robin test.

References

- Iverson, K. E. (1973) An introduction to APL for scientists and engineers. *IBM Corporation Tech. Rep. No. 320-3019*; available from APL Press, Pleasantville, New York.
 — (1976) *Elementary Analysis*. APL Press, Pleasantville, New York.
 McIntyre, D. B. (1969) Introduction to the study of data matrices. In P. Fenner, Ed., *Models of Geologic Processes—An Introduction to Mathematical Geology*. American Geological Institute, Washington, D.C.
 Polivka, R. P. and S. Pakin (1975) *APL: The Language and its Usage*. Prentice-Hall, Englewood Cliffs, New Jersey.

Manuscript received, September 1, 1978;
 accepted for publication, October 19, 1978.