FICHTNER, K. (1983). *Cryst. Res. Technol.* **18**, 77-84.
FICHTNER, K. (1984). *Cryst. Res. Technol.* **19**, 49-54.
JAIN, P. C. & TRIGUNAYAT, G. C. (1977). *Acta Cryst.* A**33**, 257-260.
MARUMO, F. & SAITO, Y. (1972). *Acta Cryst.* B**28**, 867-870.
OHSUMI, K. & NOWACKI, W. (1981). *Bull. Mineral.* **104**, 211-214.
PAŁOSZ, B. (1983). *Acta Cryst.* C**39**, 521-528.
PAŁOSZ, B. & GIERLOTKA, S. (1984). *Z. Kristallogr.* **166**, 53-64.
PATTERSON, A. L. (1944). *Phys. Rev.* **65**, 195-201.
WAHAB, M. A. & TRIGUNAYAT, G. C. (1980). *Acta Cryst.* A**36**, 1013-1016.

Acta Cryst. (1986). A**42**, 101-105

Compound Identification and Characterization using Lattice-Formula Matching Techniques

BY ALAN D. MIGHELL AND VICKY L. HIMES*

Reactor Radiation Division, National Bureau of Standards, Gaithersburg, MD 20899, *USA*

(*Received* 20 *May* 1985; *accepted* 4 *September* 1985)

Abstract

A lattice-formula matching technique has been developed to be used in conjunction with the *NBS Crystal Data File* [(1982), National Bureau of Standards, Gaithersburg, MD] for the identification and characterization of crystalline materials. This technique is reliable, efficient and highly selective. In the first step of the identification/characterization procedure, a unit cell defining the lattice is determined. The cell is reduced and derivative supercells and subcells are calculated. These cells are then checked against the *NBS Crystal Data File* in which all lattices have been represented by standard reduced cells. By routinely calculating derivative supercells and subcells and matching against the file of known compounds, it is possible to find related materials and/or to make an identification in spite of certain types of errors made by the experimentalists (*e.g.* missing rows of spots on diffraction photographs or the diffractometer *etc.*). Finally, the identification obtained by lattice matching is verified using known chemical data. Practical experience and an analysis of the data in the *NBS Crystal Data File* have proved that the lattice-formula combination is highly characteristic of a crystalline material. Since the method is subject to precise mathematical techniques, the entire procedure can be highly automated. Both the unit-cell determination and the identification/characterization procedure can be carried out in the same instrument. A Fortran program and the *NBS Crystal Data File* are available.

Introduction

We have designed a lattice-formula matching technique to be used in conjunction with the *NBS Crystal Data File* (1982) for the characterization and identification of crystalline materials. The NBS Crystal Data Center maintains a data base that contains evaluated crystallographic and chemical data on approximately 60 000 materials. The data fall into the following categories: organics, organometallics, metals, intermetallics, inorganics and minerals. There are two fundamental ways that large crystallographic data bases can be used. As a source of critically evaluated data, the data base can be used as a basis for scientific research, or as an aid to scientific research (*e.g.* to identify unknown compounds, to locate certain molecules, to obtain bibliographic data *etc.*). The type of data that can be obtained through search and retrieval programs includes chemical name and formula, cell parameters and cell volume, crystal system, space-group symbol and number, density, bibliographic data, plus additional data. Since the data base is formatted, many of these data items may be searched readily using systems software available at a particular institution. However, general systems software will not be adequate for certain types of information search and retrieval operations. One such example is the identification of unknown compounds by matching unit-cell parameters and, if available, some chemical data. Although simple in principle, lattice-formula matching is a complex operation that requires a specialized scientific background in order to design a practical computer search algorithm.

Three relatively recent developments have given the lattice-formula method for compound identification great potential as a routine analytical tool. Firstly, automated methods to determine a unit cell

* Work carried out as a National Bureau of Standards–National Research Council Postdoctoral Research Fellow and a JCPDS–International Centre for Diffraction Data Research Associate.

0108-7673/86/020101-05$01.50 © 1986 International Union of Crystallography

(and crystal structure) are in widespread use. For the same reason, the data base of known crystalline compounds is large and is rapidly growing. Secondly, new mathematical theories have practical applications in lattice-matching procedures. These procedures permit fast and effective identification in spite of certain experimental errors made in unit-cell determinations. Thirdly, advancements in computer technology have greatly increased the efficiency of search strategies through direct access of computer file records and increased mass storage capability.

Background and theory

The identification of materials by powder diffraction is a well established analytical technique. With the powder method, identification is based on matching the many diffraction positions and intensities of an unknown with those of known materials. In contrast, the method described herein identifies materials based on their crystalline lattices as defined by the unit-cell parameters. Matching unit-cell parameters may be compared to matching the positions of the diffraction lines (d spacings) in a powder pattern. The two are not theoretically equivalent, however, since a unit cell uniquely defines a lattice whereas a set of d spacings may be consistent with more than one lattice. In addition, there are many practical advantages to the lattice-matching approach, mainly due to the very compact nature of the unit cell as compared to a powder pattern. Firstly, it is far easier to treat the experimental errors when matching unit-cell parameters than it is to evaluate the errors in all the observed (and unobserved) diffraction positions in a powder pattern. Secondly, the computer search times and mass storage requirements are significantly decreased since there are fewer search parameters for the lattice-matching method. Finally, with a given unit cell, mathematical procedures allow one to calculate derivative unit cells that could result from certain errors made by the experimentalists. Since the lattice-matching method is efficient, one is able to search routinely the data base of known materials for the experimentally determined unit cell as well as for its calculated derivative cells.

There are many unit cells that can be chosen to define a crystalline lattice. When identifying a material using the powder method, unique sets of data are matched since any unit cell defining the lattice gives the same calculated d spacings. With the lattice-matching method, this uniqueness is guaranteed by always comparing reduced cells. The reduced cell is a unique primitive cell based on the three shortest lattice translations and satisfying a specified set of mathematical conditions (*International Tables for Crystallography*, 1983).

The NBS Crystal Data Center has prepared a data base that can be used for compound identification and characterization. In the *NBS Crystal Data File*, all known cells have been transformed to a reduced cell and the file has been sorted first on increasing magnitudes of the unit-cell parameters. With this classification scheme, metrically similar lattices are located near each other in the file, making it possible to design efficient lattice-search algorithms.

A summary of the lattice-formula matching procedure is presented in Fig. 1. From the unknown crystal, one determines a unit cell in reciprocal space. Next, a unit cell in direct space is determined and reduced. The reduced cell is then checked against a file of known materials. If desired, one calculates derivative lattices (Santoro & Mighell, 1973), which are also reduced and checked against the file of knowns. Finally, the identification obtained through lattice matching is verified using known chemical data. Practical examples of this identification procedure are illustrated in Figs. 2 and 3.

In Fig. 2, only one single crystal of the sample was available; thus, a chemical analysis was not possible. The initial cell was determined on an automated four-circle diffractometer. This C-centered cell was transformed to a primitive cell and reduced. The

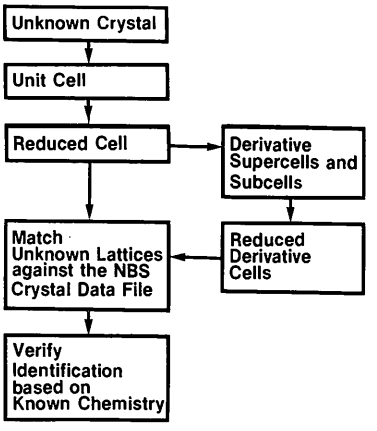


Fig. 1. Identification by lattice-formula matching.

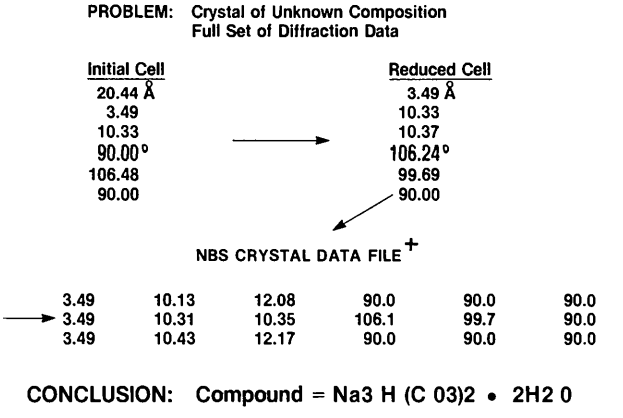


Fig. 2. Identification: direct cell match. +The cell parameters shown represent three adjacent file entries, not three separate matches.

reduced cell was then checked against the *NBS Crystal Data File*. The sample was found to be sodium sesquicarbonate dihydrate by a direct match of the reduced-cell parameters. A full structure determination by neutron diffraction confirmed the identification (Choi & Mighell, 1982).

In cases similar to the example described above, a Gandolfi camera might be used to obtain a powder pattern in order to identify the compound based on matching observed d spacings and intensities. However, in view of the many advantages of the lattice-matching method, a unit cell should be determined whenever possible. The identification may then be carried out by lattice matching, followed by comparing powder intensity data and any available chemical information. Identification procedures based on matching d spacings and intensities should be reserved for those cases in which a lattice match was not obtained, for mixtures, and for those cases in which a unit cell could not be determined.

The example given in Fig. 3 shows how an identification can be made by using a derivative lattice procedure. In diffraction work, one may determine a supercell or subcell rather than the correct cell of the lattice. For example, when using powder methods, a supercell in direct space may be determined if one does not find the smallest cell consistent with the set of d spacings. In single-crystal work, a subcell in direct space is determined if reciprocal-lattice nodes are missed on a diffraction photograph or a diffractometer. In either case, it is possible to calculate derivative cells systematically and identify the material. Here, the unknown cell was correct and the known cell was a supercell. The initial cell and space group were determined on a single-crystal diffractometer and a full set of diffraction data was collected. The empirical formula was known. (Since the lattice is rhombohedral, the initial cell edges and angles were averaged prior to carrying out the identification procedure.) When checked against the file of known compounds, there were no matches. Next, the seven supercells of twice the volume were calculated; only three were unique because the initial cell is metrically rhombohedral. One of the three calculated supercells matched a cell in the *NBS Crystal Data File*. The known cell, which was a supercell of the correct cell, had been determined from limited powder data only. Nevertheless, it was possible to verify the correct composition and to find appropriate literature references.

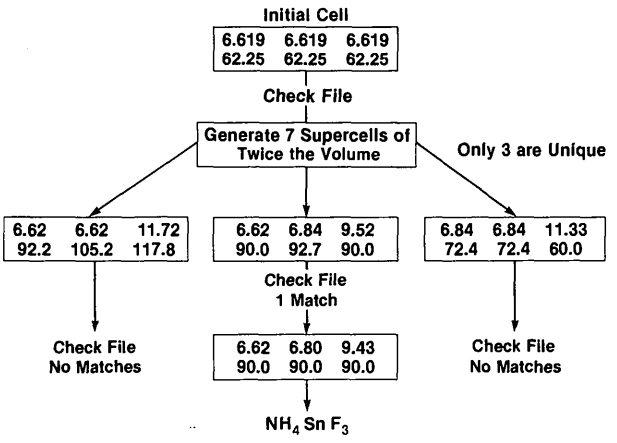
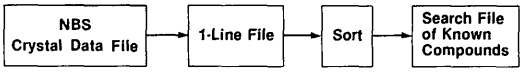


Fig. 3. Identification: match of a derivative cell.

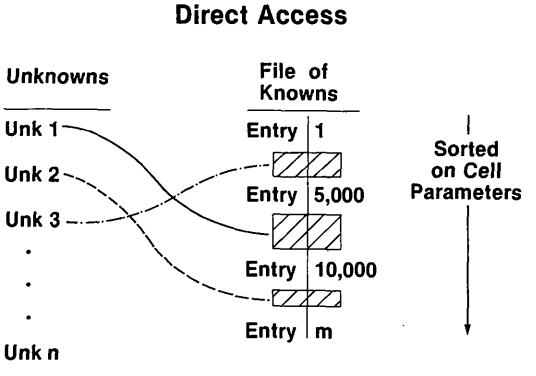
Algorithm

The theoretical basis for identification by lattice-formula matching was discussed in the previous section. Here, emphasis is placed on the strategy used to match the unknown lattice(s) against the *NBS Crystal Data File*.



The first step in the NBS lattice-matching strategy is to prepare a compact search file of known compounds. This is done to save computer mass storage, to increase the efficiency of a search and to simplify the program's input and output operations. Selected data from the *NBS Crystal Data File* are used to create a search file in which each line corresponds to the data for one crystalline compound. For a given computer, the one-line search file is prepared only once.

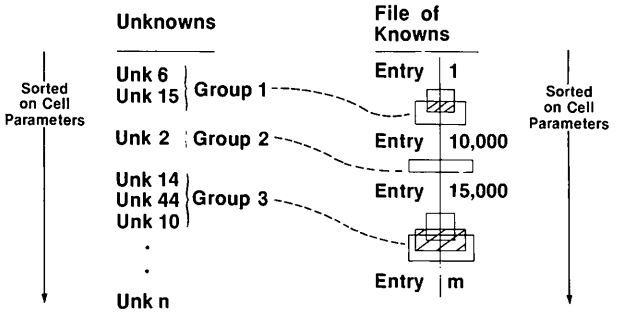
In the one-line search file, each known lattice, regardless of its original centering, is represented by its primitive reduced cell. The entries are ordered by increasing magnitudes of the reduced-cell parameters. The space group, literature reference, chemical formula and other information are included in order to facilitate identification.



The one-line search file used for identification is a direct-access computer file. Unlike sequential access files in which one may read only from the beginning to the end in a forward manner, direct access enables one to go directly to a specified point in the file. One

can access the file in any order, 'forwards' or 'backwards'. The direct access of file records is efficient as it can significantly decrease the number of file reads. In this direct-access code, a knowledge of the data in the file of knowns is used to set up a system of pointers that allows the program to start searching in about the right region of the data file. A sequential search is used only for the regions of the file in which a match is possible. In the diagram, the search regions are designated by boxes.

Grouped-Entry Direct Access



A 'grouped-entry' direct-access search strategy is actually used in the computer program. To carry out a grouped-entry direct-access search, the unknowns are sorted on increasing values of the reduced-cell parameters. Next, unknown cells with similar values of the *a* cell parameter are searched as a group against the appropriate region of the data file. In the diagram, the regions that overlap are searched only once. This approach is very efficient when one searches a number of unknowns against the one-line search file. Thus, each unknown cell as well as derivative supercells and subcells may be matched against the file of knowns in a single run.

Lattice matching based on *a*, *b*, *c*, *V* and future modifications of the computer program

The identification of unknown materials by lattice matching is possible because all input cells and all cells in the *NBS Crystal Data File* are transformed to standard reduced cells. Reduction is a mathematical procedure that leads to a unique cell in all cases, provided there is no experimental error in the unit-cell parameters. In practice, two experimentally determined cells defining the same lattice will always give the same *a*, *b*, *c* (within experimental error) for the reduced cell. Occasionally, however, the reduced-cell angles may differ because of the interactions of the experimental errors with the special conditions for reduction. For this reason, the program matches only the cell edges and cell volume of each unknown reduced cell against the known cells in the *NBS Crystal Data File*. It is possible that every cell match obtained in this way does not define the same lattice. In the next version of the program, a mathematical procedure (Santoro, Mighell & Rodgers, 1980) will be used to distinguish between lattices that have only *a*, *b*, *c*, *V* in common and lattices that are identical. For those cases in which more than one match of *a*, *b*, *c*, *V* occurs, knowledge of the empirical formula or some other chemical information is almost always sufficient to eliminate unwanted matches. This chemical screening will be automated in a later version of the computer program, completing the identification procedure by lattice-formula matching.

NBS Crystal Data File: selectivity and nature of the data

The *NBS Crystal Data File* contains evaluated crystallographic and chemical data on approximately 60 000 materials. For convenience, these data have been divided among two separate one-line search files, one composed of 'organic' data and the other containing 'inorganic' data. Each search file contains approximately 30 000 entries. Certain differences have been noted in the space-group frequency distributions of the organic and inorganic data (Mighell, Himes & Rodgers, 1983; Mighell & Rodgers, 1980). The compounds in the organic search file usually crystallize in the lower-symmetry crystal systems; while the space groups for materials in the inorganic search file are skewed towards the higher-symmetry crystal systems. This has practical implications in the identification of compounds. For organic compounds, the cell alone (*a*, *b*, *c*, *V*) is usually sufficient to characterize the material fully. However, because the inorganic search file generally contains more compounds from the higher-symmetry crystal systems and a larger percentage of isostructural data, some chemical information is often needed to characterize these materials fully.

Experience has shown that the lattice (as defined by a unit cell) is highly characteristic of a compound and, like a powder pattern, may be used for identification. On the average, only ~500 entries in either the organic or inorganic search file will separate two entries when the *a* values of the reduced cells differ by ~0·20 to ~0·25 Å. In the middle region of each search file, ~500 file entries will separate two reduced cells whose *a* values differ by ~0·11 Å. Thus, the lattice-matching procedure is relatively selective based on the *a* cell parameter alone. When values of *b*, *c* and *V* for the reduced cell are considered, the selectivity is dramatically increased. (The selectivity of the algorithm will be further enhanced by the addition of a mathematical technique to distinguish between lattices that are identical and lattices having only *a*, *b*, *c* and *V* of the reduced cell in common.)

As part of a practical test of the lattice-matching method, the entire one-line organic search file of

~30 000 compounds was considered as a file of unknowns to be matched against itself. When tolerances of ±0·1 Å for *a*, *b*, *c* and ±10% for *V* were used, most 'unknowns' gave only one or two matches. In the entire analysis, a maximum of 17 matches of *a*, *b*, *c*, *V* were found for a single 'unknown'. These matches contained data for the same or isostructural compounds. As expected, a similar analysis of the inorganic search file yielded more matches on the average. These two analyses demonstrated the efficiency of the grouped-entry direct-access search strategy as well as the selectivity of the lattice-matching method.

Concluding remarks

The NBS lattice-formula matching procedure is reliable as one is commonly working with well characterized materials. Usually, the cells have been refined by least-squares techniques and the chemistry is known. For most crystals, the lattice is unique. This means that the cell or the cell plus some chemical data is sufficient for identification.

The lattice-formula method of identification is a general procedure that may be applied to any cell regardless of the technique used for its determination. Cell parameters may be measured by X-ray, neutron or electron diffraction using single-crystal or powder diffraction methods. For example, the method can be applied to a compound analyzed by powder diffraction. One can index the powder pattern and search the *NBS Crystal Data File*. If no matches are found, one can then check a powder diffraction file using a more complicated search technique based on matching *d* spacings and diffraction intensities.

Lattice-matching methods are subject to precise mathematical procedures and a high degree of computerization. Experience with practical problems has shown that identification by matching reduced cells is very straightforward and reliable when the correct cell of the lattice has been determined. However, a material can often be identified even if an error has been made in which a derivative cell of the unknown has been found. For example, if the cell centering is missed when indexing a powder pattern or, when using single-crystal methods, if rows of spots are missed on the diffraction photographs, this would result in the determination of a derivative cell of the correct lattice. To make an identification, one systematically calculates the derivative supercells and subcells and looks for matches in the *NBS Crystal Data File*. Checks for the 55 supercells of two, three and four times the volume of the unknown reduced cell and the 55 subcells of 1/2, 1/3 and 1/4 times the volume are made routinely. The ability to check for matches of 110 extra derivative cells for each unknown is only possible because of the efficiency of the search algorithm and the selectivity of the method. Finally, the entire identification procedure can be highly automated including the measurement of the unit-cell parameters. Both the unit-cell determination and the identification process can be carried out in one instrument.

Data file and program availability

A magnetic tape of the *NBS Crystal Data File* is available for lease. A Fortran program, *NBS*LATTICE*, has been written to analyze lattice relationships and is available for distribution by the NBS Crystal Data Center. The present version of *NBS*LATTICE* performs several functions including the identification of unknown materials using lattice-formula matching techniques, the calculation of the reduced cell of the lattice, the calculation and reduction of specified derivative supercells and/or subcells, and the determination of metric lattice symmetry. Further information concerning the *NBS Crystal Data File* and requests for the *NBS*LATTICE* program should be directed to Drs Vicky L. Himes and Alan D. Mighell, Reactor Radiation Division, National Bureau of Standards, Gaithersburg, MD 20899, USA.

We thank the Office of Standard Reference Data at the National Bureau of Standards for its support of this work.

References

CHOI, C. S. & MIGHELL, A. D. (1982). *Acta Cryst.* B**38**, 2874–2876.

International Tables for Crystallography (1983). Vol. A, edited by TH. HAHN, pp. 734–743. Dordrecht: Reidel.

MIGHELL, A. D., HIMES, V. L. & RODGERS, J. R. (1983). *Acta Cryst.* A**39**, 737–740.

MIGHELL, A. D. & RODGERS, J. R. (1980). *Acta Cryst.* A**36**, 321–326.

NBS Crystal Data File (1982). A magnetic tape of crystallographic and chemical data compiled and evaluated by the NBS Crystal Data Center, National Bureau of Standards, Gaithersburg, MD 20899, USA.

SANTORO, A. & MIGHELL, A. D. (1973). *Acta Cryst.* A**29**, 169–175.

SANTORO, A., MIGHELL, A. D. & RODGERS, J. R. (1980). *Acta Cryst.* A**36**, 796–800.