Aust. J. Phys., 1988, **41**, 145–53

Profile Data Acquisition for the JCPDS–ICDD Database*

Ron Jenkins

International Center for Diffraction Data,
1601 Park Lane, Swarthmore, PA 19081, U.S.A.

Abstract

The principal advantage offered by a fully digitised diffraction pattern is the retention of all features of the experimental pattern, including the line width and shape, the form and distribution of the background, etc. A file containing this type of reference data would in the future allow the use of techniques yet to be developed and of data processing, such as peak location, background subtraction and α_2 stripping. The availability of digitised reference patterns would also allow the use of pattern-recognition techniques for qualitative phase analysis, as well as offering interesting possibilities for quantitative work. Until recently most commercially available automated powder diffractometers were limited to 10–20 Mbytes of disc storage and since a single fully digitised pattern requires about 10 kbytes, the provision of a file for thousands of digitised single phase reference patterns has not been possible. The recent advent of compact disc–read only memory (CD–ROM) systems providing in excess of 500 Mbytes now offers a low cost data storage capability. Plans are now in place for a new version of the Powder Diffraction File consisting of fully digitised patterns. Because of the need to maintain the database for years to come, it is most important that the stored data be as accurate and complete as possible.

1. Introduction

The technique of using X-ray powder diffraction patterns as a means of phase identification has been utilised for more than 50 years. The means of archiving and retrieving patterns proposed by Hanawalt and Rinn (1936) still provides the basis of search–match methods used today. The traditional method for storage of data is to reduce the experimental rate meter recording, or the X-ray film data, to a table of d/I values, often referred to as a 'stick pattern'. In the present paper such a pattern is referred to as a 'reduced pattern' because the process reduces the analog pattern to a reduced digital form. Unfortunately, during this conversion process information on the line shape and intensity distribution is lost. In addition, problems due to inadequate data treatment in finding the angular positions of the peak maxima, plus uncertainties in the value of the experimental wavelength(s), all add errors to the experimental d values. The intensities can be similarly distorted due to problems of partially resolved diffraction wavelength multiplets, and to line broadening due to particle size and/or strain considerations. Many of the problems of data reduction should be solved in the future and, indeed, considerable improvements have already

* Paper presented at the International Symposium on X-ray Powder Diffractometry, held at Fremantle, Australia, 20–23 August 1987.

0004-9506/88/020145$03.00

occurred during the last several years. The use of 'profile fitting' techniques (Parrish *et al.* 1976) has done much to give a better measure of integrated peak intensities and profile maxima. A possible solution to the short-term data-handling problem might be to develop a completely new database in which the powder diffraction data are stored as fully digitised patterns. Such a database would allow alternative data treatments to be applied in future years when new techniques become available.

There are several other technical reasons for establishing a digitised pattern database, referred to as the PDF-3 database, and these reasons are discussed in the next section. However, the problem remains of how such a database should be set up. Because the data to be stored are essentially raw and untreated, it is clear that the data should be free from significant systematic errors. Where such a guarantee cannot be given, the systematic errors should at least be defined, so that corrections can be applied later.

2. Potential Uses of the Digitised Pattern Database

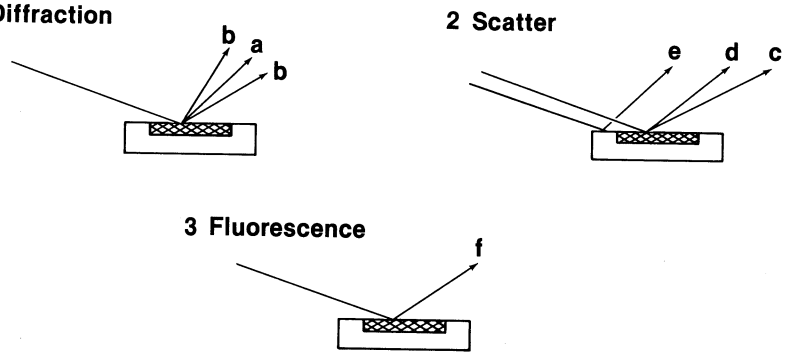
A major advantage in establishing the PDF-3 is that all original features of the experimental pattern are retained. This includes information on line width and shape, the form and distribution of the background, the occurrence of variations in the background, etc. There are several potential uses for the PDF-3 database and the most important are: (1) the future use of better data treatment methods; (2) pattern recognition techniques in qualitative analysis; (3) pattern matching techniques in quantitative analysis; and (4) analysis of standard mixtures.

Use of the full pattern for qualitative analysis has advantages over the use of a reduced pattern. The lines in a diffraction pattern have a definite structure and their shapes may be affected by a number of instrumental and sample dependent factors. In some cases, for example in the case of clay mineral (00l) lines, the line shape may be a better clue to the nature of the phase than the actual d-spacings. Unfortunately, line shape information is lost in the reduction of a line profile to a single 2θ value. While it is true that the second differential peak location method (Schreiner and Jenkins 1980) does give some information about the peak width, considerably more information than this is desirable. The PDF-3 would be especially useful for quantitative analysis, and data have recently been reported (Smith *et al.* 1987) using this method for the analysis of polyphase mixtures. This method offers many advantages where line broadening, line overlap and intensity variations due to preferred orientation make the use of single reduced line(s) a poor measure of the diffraction intensity. Another possible use of the PDF-3 is in the analysis of 'standard mixtures'. Often specimens are analysed where combinations of phases are present; for example, corrosion products and steam boiler tube deposits. In addition, there are cases where phases are only stable in the presence of other phases. A third category might be certain naturally occurring mixtures of minerals. In each case, the physical separation of the mixture into its component phases to provide reference patterns is difficult, and in some cases unwise. Better results in qualitative characterisation would probably be gained by matching the unknown pattern with patterns of the likely occurring mixtures.

3. Composition of an X-ray Diffractogram

If the PDF-3 is to be successful, it is vital that the data be reliable and free both from systematic errors and from unwanted lines and artifacts in the patterns.

It is useful to review the make-up of a typical diffractogram. The components shown in Fig. 1 include three phenomena which contribute to the observed intensity: diffraction, scatter and fluorescence. The diffractogram itself is made up of diffraction peaks superimposed on the background. The desired peaks arise from diffraction at the assumed experimental wavelength, whereas unwanted peaks arise from other wavelengths. These other wavelengths may arise because the experimental wavelength may not be completely monochromatic, as discussed in the next section. The background arises from coherent and incoherent scattering from the sample and the beam path and, especially at low angles, this may include air scatter and scatter from the sample support (Jenkins and Squires 1982). When the X-rays incident on the specimen are sufficiently energetic to excite characteristic radiation from the sample (fluorescence), and when this excited radiation falls within the energy acceptance range of the detection system, additional background may occur.



Type	Source	Gives
a	Diffraction of required wavelength	Wanted peaks
b	Diffraction of other wavelengths	Unwanted peaks
c	Coherent scatter from sample	General background
d	Incoherent scatter from sample	General background
e	Scatter from sample support	Extra background at low angles
f	Fluorescence from sample	General background

Fig. 1. Composition of a diffraction pattern.

At present most qualitative work is carried out using a powder diffractometer in the Bragg–Brentano para-focusing geometry, which may introduce additional aberrations and/or features to the diffractogram. The five most important are:

(a) Systematic errors in the observed 2θ maxima due to inherent aberrations, including flat specimen error, axial divergence and specimen transparency. These errors are angle dependent and manifest themselves as an increasingly large relative error in the d spacing as the diffraction angle is decreased.

(b) Systematic errors in the observed 2θ maxima due to the displacement of the specimen surface from the goniometer focusing circle. Additional errors may accrue due to inadequate alignment or incorrect calibration of the goniometer.

(c) Line profile shapes may be asymmetrically distorted at lower 2θ angles because of axial divergence allowed by the Soller slits. The amount of distortion increases as the axial divergence of the collimators increases. The collimator characteristics may vary with the design of the diffractometer.

(d) Intensities are highly dependent on specimen orientation because the optical arrangement of the goniometer requires that the receiving slit scans at a constant rate through two-dimensional space, intersecting diffraction rings on the Ewald sphere. Because the distribution of intensity around these rings is itself orientation dependent, the portion of the ring intersected by the receiving slit may not be a good measure of the integrated intensity distribution around the ring.

(e) Intensities vary with divergence slit aperture. If the slit aperture is fixed and the irradiation length of the sample is less than the sample length, the relative intensities are approximately constant. Where a variable divergence slit is used (Jenkins and Paolini 1974), the observed intensities, relative to those obtained with a fixed slit, increase at higher angles by as much as a factor of three.

An additional factor to consider is the possibility that other diffractometer or Guinier geometries may be employed in the future. This is an additional reason to ensure that all systematic aberrations in standard digitised patterns are properly quantified.

4. Problem of Defining the Wavelength

A major source of additional weak lines and artifacts in a diffractogram is the polychromatic nature of the source. Probably in excess of 90% of all powder studies carried out at present in the United States is with Cu Kα radiation. While it is generally the intent to diffract just the Cu K$\alpha_{1,2}$ doublet, the radiation actually diffracted and detected may be more than simply the Kα doublet. Fig. 2 shows the spectral distribution in the energy region around Cu Kα radiation. The Kα and β doublets are shown along with other lines which could fall within the acceptance range of the monochromatisation–detection device. There are three methods that are commonly employed to render the radiation 'monochromatic' (in fact 'bichromatic' because the α doublet is generally employed):

(a) use of a β filter, generally in association with a proportional type detector and a pulse height selector (PHS);

(b) use of a diffracted beam monochromator, typically based on a monocrystal of pyrolytic graphite (PG);

(c) use of a solid state proportional detector, typically Si(Li) or Ge(Li).

Also shown in Fig. 2 are the relative bandpass regions for these three monochromatisation systems. The filter is a single bandpass device which is mainly used to improve the ratio of Cu Kα to Kβ to about 50 : 1. It is generally supplemented by energy discrimination to remove high energy white radiation also coming from the X-ray tube. The effectiveness of this removal depends upon the resolution of the detector. For Cu Kα, the resolution of the scintillation detector is about 2000 eV, which will not allow the removal of radiation in the immediate vicinity of the Kα lines. While the angular dispersion of the pyrolytic graphite crystal is less than 100 eV, because its mosaic structure allows diffraction over a fairly wide angular

range, it has an effective bandpass of around 500 to 1000 eV. The actual bandpass for a given monochromator will depend upon the widths and positions of limiting apertures placed somewhere on the focusing circle of the monochromator. It should also be understood that a slightly wrong setting of the monochromator can displace its energy acceptance window in either direction. Finally, the Si(Li) detector has a resolution of about 200 eV for Cu Kα radiation. Thus, it is clear that, even with the monochromatising devices correctly set, additional radiation can be diffracted and pass to the counting circuits. When the monochromatising device is incorrectly set, even more radiation may pass. Table 1 summarises the potential interfering lines.

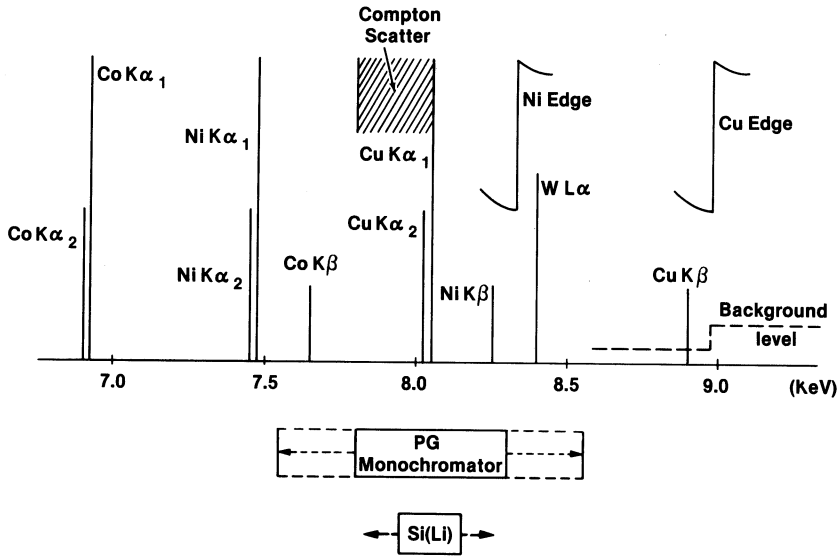


Fig. 2. Bandpass range of a graphite monochromator and a Si(Li) detector.

Table 1. Potential line interferences with monochromatisation devices

Line	Energy (keV)	β filter/PHS	PG monochromator	Si (Li) detector
Cu Kβ	8·90	Reduces to <2% of Kα	Removes	Removes
W Lα	8·40	Reduces	Reduces	Removes
Ni Kβ	8·26	Little effect	Little effect	Reduces
Cu Kβ (Compton)	7·80	No effect	Little effect	Partly removes
Co Kβ	7·65	Little effect	Reduces	Removes
Ni Kα	7·47	Little effect	Removes	Removes

A final point to consider in the wavelength problem is the increasing use made of synchrotrons for powder diffraction measurements. Though the fraction of diffraction work carried out by this means is still small, the high degree of monochromaticity and wavelength tunability make it very different from conventional X-ray tube source systems. It would seem prudent, therefore, to carefully consider the impact of the inclusion of synchrotron source derived patterns in the PDF-3.

5. Data Treatment Procedures

In the design of the PDF-3 consideration should be given to the way data will be processed in the future. Essentially two methods are in use at the present time—profile fitting and a combination of smoothing, stripping and peak location. The latter option is by far the most common and Fig. 3 lists the typical steps employed in this process. Following collection of the digitised data, the data are smoothed and the background subtracted. The α_2 wavelength can then be stripped and a peak location method applied. This peak location method is typically done using a second-differential method in which the negative space is fitted to a parabola and the minimum calculated. A 2θ calibration may then be applied based either on an internal standard or an external calibration curve. Finally, the peak data are converted to d-spacings and stored to an appropriate significance level. While different data-handling systems may apply these steps in a slightly different order, most software schemes include all these steps.

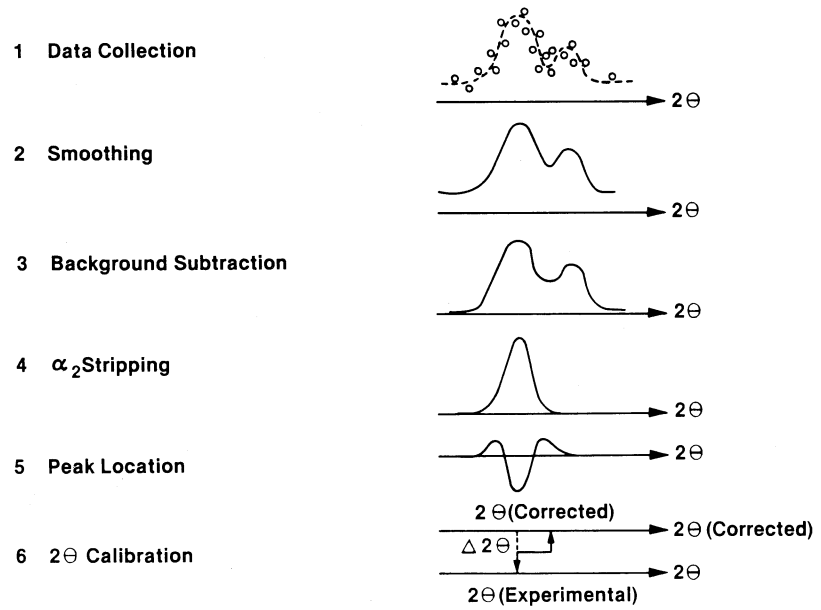


Fig. 3. Steps involved in the collection and treatment of peak data.

6. Defining a Digitised Pattern Database

The three major problems in establishing a PDF-3 are, first, to select the storage medium; second, to define the format of the stored data; and third, to ensure that the information in the database is correct. For the storage medium, the issue is one of finding a low-cost high data density medium with a reasonably fast access time. To record a diffraction pattern over a 90°2θ angular range, with a step size of 0·01°, and

to include an adequate definition of the experimental conditions would require about 10 kbytes per pattern. If the current JCPDS–ICDD Powder Diffraction File of about 48 000 patterns were to be stored in this way, almost 500 Mbytes would be required. This storage capacity is already being provided by current CD–ROM technology. ICDD has recently announced (Jenkins and Holomany 1987) support for the Powder Diffraction File (PDF-2) on CD–ROM, and an extension of this project to include other databases seems feasible. There is little doubt that optical disc technology will soon allow low-cost data storage at the gigabyte level, which is more than sufficient for the size of the database envisaged.

The problem of defining the PDF-3 is much more complex. As already stated, while most commercially available X-ray diffractometers are similar in their geometry, this may not always be the case. What is an unimportant parameter at present may be critical in the future. The wisdom of carefully defining a database is borne out by the experience obtained with the National Bureau of Standards diffraction database project NBS*AIDS83 (Mighell *et al.* 1981). This work provided a common database format for both the PDF-2 and the Crystal Data File (CDF). A further possible complication is that more commonality may begin to appear in scientific databases. Because inexpensive mass storage media are now becoming available, it is reasonable to assume that several different databases might be present on the same disc system. The problems of peak location and integration are common to many branches of analytical science, and thus it also seems reasonable to predict that a degree of similarity in data treatment software will be required. The usefulness of such an approach will clearly be predicated by commonality of database format.*

One example of the complications involved in providing a suitable definition for digitised patterns is the selection of angular step size. Fig. 4 illustrates this problem in a simplified way. While one generally attempts to produce a single phase 'standard' material of suitable particle size, to reduce orientation and heterogeneity problems and to be free from strain, this is not always possible. Often lines in a pattern may be broad and, as shown, the choice of step size and any subsequent data smoothing is predicted on how much of the profile should be sampled. A rule-of-thumb is that between 10 and 20 data points should be collected above the peak width at half maximum. Too few points causes loss suppression of the peak maximum and can also lead to small peak shifts. Too many points can lead to problems in peak maximum definition, again introducing shifts in the reported peak maximum. The situation is clearly different where profile fitting techniques are employed because, in this case, many more of the profile data are being used.

Another point to consider in the definition of the PDF-3 is to ensure that the quality of data in the database is acceptable, not just at present but also in the future. The database format of the PDF and CDF has remained unchanged for many years and there has been an ongoing effort to continually review and upgrade the quality of data (Jenkins *et al.* 1987). The likely source of data for the PDF-3 is the general

* This question of spectral database commonality is being addressed by the Joint Committee on Atomic and Molecular Properties [JCAMP-DX. A standard form for exchange of infrared spectra in computer readable form (Revision 4·23 February 1987)], a multi-discipline group seeking a suitable general purpose database format. As well as being active on this committee, the ICDD is also investigating database format definition through several of its subcommittees, including the Instrument Data Collection Task Group, the Database Subcommittee, plus a recently formed Database Format Review Committee.

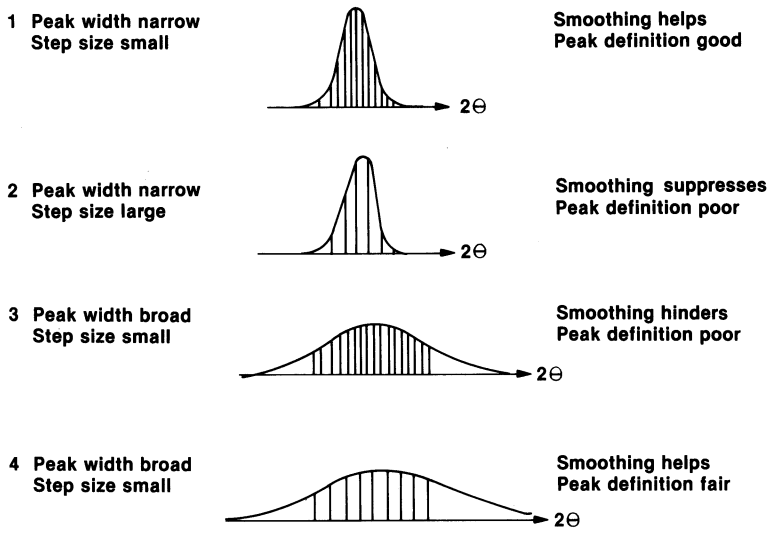


Fig. 4. Choice of step size and smoothing for narrow and broad peaks.

diffraction community, including a significant proportion of laboratories which are, at least in part, supported by the ICDD Grants-in-Aid program. Round-robin tests conducted by the ICDD have indicated that the quality of powder diffraction data is highly variable, so that careful calibration schemes will have to be introduced. Three specific calibration problems are now being considered:

(a) calibration for 2θ;

(b) calibration for intensities; and

(c) calibration for experimental wavelength(s).

The 2θ calibration is required mainly to compensate for inherent aberrations due to the diffractometer geometry, and any residual misalignment. If an internal standard were to be used, it would also correct for specimen displacement. The 2θ calibration is probably the easiest of the three to carry out, and good angle reference standards are available through the NBS Standard Reference Material program (Dragoo 1986). The intensity calibration is mainly required to correct for divergence slit aperture differences and possible dead-time effects. This calibration is harder because of the difficulty of finding stable materials essentially free from orientation problems. Preliminary results of a recent ICDD round-robin have indicated (Schreiner and Jenkins 1988) that beta-spodumene might be a useful candidate for this purpose. The need for experimental wavelength calibration has already been discussed at length. This calibration is probably the most difficult of the three (partly because it is a test not generally made in most X-ray laboratories). While a suitable procedure is still under consideration, one possibility is the use of a single crystal of silicon as the spectral dispersion medium. The Si(Li) detector will also prove useful in this area, though it is likely that some information will be lost due to the relatively broad bandpass of the system.

7. Conclusions

A good case can be made for the provision of a full pattern database, one which would provide many advantages over the present reduced database. Preliminary data reported from experiments using such a database for quantitative work are very encouraging. The storage problems of the PDF-3 are likely to be trivial with the growth of optical disc technology, but major problems remain in the area of database definition.

References

Dragoo, A. L. (1986). *Powder Diffraction* **1**, 294.
Hanawalt, J. D., and Rinn, H. W. (1936). *Ind. Eng. Chem. Anal. Ed.* **8**, 244.
Jenkins, R., and Holomany, M. (1987). *Powder Diffraction* **2**, 215.
Jenkins, R., Holomany, M., and Wong-Ng, W. (1987). *Powder Diffraction* **2**, 84.
Jenkins, R., and Paolini, F. R. (1974). *Norelco Reporter* **21**, 9.
Jenkins, R., and Squires, B. (1982). *Norelco Reporter* **29**, 26.
Mighell, A., Hubbard, C. R., and Stalick, J. K. (1981). NBS*AIDS80. A Fortran program for crystallographic data evaluation. National Bureau of Standards, Technical Note 1141 (NBS: Washington, D.C.).
Parrish, W., Huang, T. C., and Ayers, G. L. (1976). *Trans. Am. Crystallogr. Assoc.* **12**, 55.
Schreiner, W. N., and Jenkins, R. (1980). *Adv. X-ray Anal.* **23**, 287.
Schreiner, W. N., and Jenkins, R. (1988). *Adv. X-ray Anal.* **31** (in press).
Smith, D. K., Johnson, G. G., Jr, Scheible, A., Wims, A. M., Johnson, J. J., and Ullman, G. (1987). *Powder Diffraction* **2**, 73.

Manuscript received 21 August, accepted 5 November 1987