Acta Cryst. (1993). A**49**, 116-126

Description and Peak-Position Determination of a Single X-ray Diffraction Profile for High-Accuracy Lattice-Parameter Measurements by the Bond Method. II. Testing and Choice of Description

BY EWA GAŁDECKA

Institute of Low-Temperature and Structure Research, Polish Academy of Sciences, ul. Okólna 2, 50-950 Wrocław, Poland

(*Received* 20 *October* 1991; *accepted* 26 *June* 1992)

Abstract

The reliability of the results of the peak-position determination by means of various approximating functions - the polynomials and shape functions considered in paper I [Gałdecka (1993). *Acta Cryst.* A**49**, 106-115] - was examined based on test calculations in which both measurement data and computer-simulated data were used. It was found that accurate peak-position determination must be based on a reasonable physical and statistical model of the diffraction profile. The statistical model is required to provide an objective criterion of the goodness of fit. The goodness of fit is a necessary but not sufficient requirement for obtaining accurate results. To provide unbiased and stable enough (independent of scanning range) results, the function to be used must, in addition, be continuous and give a good approximation to known physical models of the diffraction profile. It was proved that the use of a reasonable shape function - here, a pseudo-Voigt function with a linear (or exponential) asymmetric factor, the best of the functions considered - leads to accurate enough and highly stable results. Results obtained using polynomials, even when the goodness of fit is being carefully checked, are no more precise and are less stable and so less accurate than those obtained using the carefully selected shape function. Only in the case of a parabola - the simplest and the most preferred among polynomials - there is a possibility of reducing the bias of outcomes by an extrapolation of results obtained for various scanning ranges. However, the improvement also requires a reasonable model (here, the shape function mentioned above).

1. Introduction

In paper I (Gałdecka, 1993), in view of some disadvantages of the polynomial approximation, the use of some other analytic descriptions of the diffraction profile was considered. The descriptions available were reviewed, analysed and discussed in all aspects. In the present paper (paper II), results of test calculations are reported, as a basis for a discussion of the effect of the best model of the diffraction profile on the accuracy of the peak-position determination. The outcomes obtained using selected shape functions are compared with those obtained using polynomials and the first group of test calculations reported relates to the polynomial approximation. Reasonable statistical criteria are used for testing the goodness of fit.

2. Data and criteria used in test calculations reported

2.1. *Data sets used in the tests*

To find out the accuracy of the peak-position determination that can be achieved in practice and which of the approximating functions considered will give the most accurate results, test calculations were performed by approximating a series of data with various functions by least-squares fitting. In the tests two kinds of data were used: (i) real measurement data; (ii) data obtained using computer modelling.

The first kind of data [(i)] was used to show how various methods of approximation are applicable to an arbitrary set of real measurement data. The series of data consisted of $M = 10$ collections of $n = 61$ intensity values, $h_1, \ldots, h_n$, recorded within a diffraction profile of the 444 interference of a silicon single crystal with a Bond-system diffractometer (Łukaszewicz, Kucharczyk, Malinowski & Pietraszko, 1978). The peak intensity was $H \simeq 10\,000$ counts and the half-width of the profile was $\omega_h \simeq 480''$ (0.133°). The intensities were recorded at the truncation level

$$(h_1 + h_n)/2H = 0.16. \qquad (1a)$$

A constant scanning step, $\Delta\omega = 20''$, was used and the angular values covered the range

$$2\Omega = (n-1)\Delta_\omega. \qquad (1b)$$

Repeated measurements were used for estimation of the mean values (sample mean) and the standard deviations of the peak positions determined. To estimate the expected values of the goodness-of-fit parameters, $\bar{R}$ and $\hat{\sigma}$, for various scanning ranges, the model of variances and covariances derived by Gałdecka (1985) (see also paper I, § 2.6) was used with the parameters $\sigma(\omega)_R = 1''$, $\sigma(\omega)_S = 2''$ and $\sigma(I)/I = 0.0055$. The corresponding calculated

0108-7673/93/010116-11$06.00

values, R and σ, resulting from the approximation by least-squares fitting reported here (§§ 3 and 4) are the mean values over $M = 10$ repeated measurements.

The collections of real measurement data were modified by rejecting exterior points to test the effect of the scanning range, 2Ω, on the peak position determined. It was, however, impossible to evaluate objectively the accuracy of the peak-position determination since the true peak position, ω_P, of the underlying 'pure' profile (not affected by statistical errors) was unknown. Ultimately, it was assumed that the 'true' peak position was that of the approximating function giving the best results (§ 4.4). Since, as mentioned in paper I, § 1.1, the Bragg angle in the Bond method is determined from the difference between two angular positions, $\omega_{0,1}$ and $\omega_{0,2}$, so the position of the origin of the ω scale is not important, the ω_0 values reported will be related to an arbitrary origin (here, $\omega_0 \simeq 50''$).

In case (ii), to check the accuracy in a more objective way, a series of computer-simulated data was used. With the values retained for basic parameters characterizing the real measurement described in point (i), $M = 10$, $\Delta_\omega = 20''$, $H \simeq 10\,000$, $\omega_h = 480''$ and with the assumption that the 'recorded' counts obeyed Poisson statistics and that the true peak position was at $\omega_P = 50''$ (as for real data), one of the distributions derived by Johnson (1949) [see paper I, (24a), (24d)],

$$v(u) = (1+u^2)^{-1/2} \times \exp\left(-0.5b\{1+\varepsilon \ln[u+(1+u^2)^{1/2}]\}^2\right), \quad (2)$$

with parameters $b = 2.0$, $\varepsilon = 0.1$, was used to model the pseudoexperimental data.

Additional test calculations were performed to compare selected approximating functions and known models of the diffraction profile. The tests will be described in § 4.2, together with the data used in them.

2.2. χ^2 *test for the goodness of fit*

As mentioned in paper I (§ 2.6), to treat a given approximating function as a model of the diffraction profile within a given scanning range 2Ω, the mean deviance, σ_h, obtained by least-squares fitting must be equal to the mean (over 2Ω) standard deviation of observations, $\hat{\sigma}_h$. The equality or otherwise can be tested using the χ^2 test. As results from statistical considerations [for more details see, for example, Hamilton (1964, § 2-8)], to accept a description considered at a given significance level α (say $\alpha = 0.1$), the following condition must be satisfied:

$$[\chi^2_{\nu,1-\alpha/2}/(n-m)]^{1/2} \le \sigma_h/\hat{\sigma}_h \le [\chi^2_{\nu,\alpha/2}/(n-m)]^{1/2}, \quad (3a)$$

where $\chi^2_{\nu,1-\alpha/2}$ and $\chi^2_{\nu,\alpha/2}$ are critical values of the χ^2 distribution for $\nu = n - m$ degrees of freedom. For large ν values ($\nu \ge 100$), an approximate formula may be used (Abrahams, 1969):

$$\sigma_h/\hat{\sigma}_h \le \{1 - 2/9(n-m) + u_\alpha[2/9(n-m)]^{1/2}\}^{3/2}, \quad (3b)$$

where u_α is the value of the normal distribution corresponding to the probability α. Since, as mentioned above, the σ_h values reported in the paper are the mean values based on $M = 10$ repeated measurements, the number of degrees of freedom, $\nu = Mn - m$, is so large in this case that an almost 'exact equality' is required,

$$0.9 < \sigma_h/\hat{\sigma}_h < 1.1. \quad (3c)$$

As can easily be shown, the ratio $\sigma_h/\hat{\sigma}_h$ in (3a), (3b), (3c) may be replaced by the ratio $R/\hat{R}$ of corresponding discrepancy factors. Since many authors prefer the latter criterion, both goodness-of-fit parameters, σ_h and R, as compared with their expected values, $\hat{\sigma}_h$ and $\hat{R}$, will be reported in §§ 3 and 4.

2.3. *Measures of accuracy, precision and stability*

A measure of accuracy of a single peak-position determination is the actual error in the peak position, $\mathrm{d}\omega_P$, *i.e.* the difference between the peak position ω_P determined and the 'true' peak position ω_0,

$$\mathrm{d}\omega_P = \omega_P - \omega_0. \quad (4)$$

Normally, the true peak position is unknown. However, we shall assume that the 'true' peak position is that of the best model function describing the measured diffraction profile (§ 4.3).

Since the actual peak position ω_P is affected by both the systematic error (the bias) $\Delta\omega_P$ and the statistical error $\delta\omega_P$, the difference defined by (4) will consist of two components,

$$\mathrm{d}\omega_P = \Delta\omega_P + \delta\omega_P, \quad (5)$$

where

$$E(\delta\omega_P) = 0, \quad \text{so} \quad \Delta\omega_P = E(\mathrm{d}\omega_P)$$

(E denotes the expected value).

To evaluate the actual error, its two components and their total effect on the accuracy of a single peak position, let us consider limits of the confidence interval, calculated for M values ω_P determined from repeated measurements. Using Student's t distribution for a chosen significance level α (say $\alpha = 0.10$) and $\nu = M - 1$ degrees of freedom, one can assess with the probability $1 - \alpha$ (0.90) that the result of a single peak-position determination, ω_P, will lie inside the interval

$$\bar{\omega}_P - s(\omega_P)t_{\alpha,\nu}M^{-1/2} \le \omega_P \le \bar{\omega}_P + s(\omega_P)t_{\alpha,\nu}M^{-1/2}, \quad (6)$$

where $\bar{\omega}_P$ is the sample mean and $s(\omega_P)$ is the sample standard deviation. By comparing (4) and (6) we

obtain an estimator of a likely actual error of a single peak-position determination,

$$\bar{\omega}_P - \omega_0 - s(\omega_P) t_{\alpha,\nu} M^{-1/2} \leq d\omega_P \leq \bar{\omega}_P - \omega_0 + s(\omega_P) t_{\alpha,\nu} M^{-1/2}. \quad (7)$$

As a measure of the accuracy of a single peak-position determination, we shall use the maximum absolute value of the actual error described by (7). For $M = 10$ ($\nu = 9$) and $\alpha = 0.1$ we obtain $t_{0.1,9} = 1.833$, so [*cf.* (7)]

$$d\omega_{P,\max} = |\bar{\omega}_P - \omega_0| + 0.58 s(\omega_P). \quad (8a)$$

For M large enough, it is possible to separate the bias and the statistical error from the actual error, since

$$\Delta\omega_P \simeq \bar{\omega}_P - \omega_0 \quad (8b)$$

and

$$\delta(\omega_P) \simeq s(\omega_P) t_{\alpha,\nu} M^{-1/2}. \quad (8c)$$

Since the data used for the peak-position determination may be collected within different (narrower or wider) scanning ranges, it is required that the results being obtained should be stable, *i.e.* independent of the scanning range. As a measure of the stability, we propose to use the ratio of the maximum difference of results obtained within the range of scanning ranges 2Ω permitted by the statistical criterion [(3*a*)] to the length of the range,

$$\mathrm{st} = (\omega_{0,\max} - \omega_{0,\min})/(2\Omega_2 - 2\Omega_1), \quad (9)$$

where $\omega_{0,\max}$ and $\omega_{0,\min}$ are extreme values of the peak position, $2\Omega_2$ is the widest scanning range and $2\Omega_1$ is the narrowest.

3. Polynomial approximation

As mentioned in paper I, § 1.2, there are two basic problems of peak-position determination by use of polynomial approximation:

(i) a limited range of arguments for which a given polynomial might be treated as a 'model' of the diffraction profile and for which the results obtained were free from any bias;

(ii) dispersion of results that increases with decreasing scanning range.

These problems are illustrated in Figs. 1 and 2. The figures present the use of a parabola and a sixth-degree polynomial. In fact, many more test calculations were performed using polynomials of all degrees from $m = 2$ to $m = 12$. The examples shown in Figs. 1 and 2 are representative for all the tests and illustrate well the substance of the problem; it is thus unnecessary to show more examples. In the tests reported, both 'ordinary' polynomials and the Gram orthogonal polynomials were used. Results obtained did not vary for $m \leq 12$, but calculations by means of the Gram polynomials were much faster.

As shown in Figs. 1(*a*) and 2(*a*), the actual error in a single peak position defined by the maximum distance between limits of a given confidence interval and the 'true peak position' [*cf.* (8*a*)] exceeds 1″ (3×10^{-4}°), so is greater than that permitted for the accuracy of 1 part in 10^6 (see paper I, § 1.1). Considering the range of scanning ranges 2Ω permitted for a given polynomial by the statistical criterion [(3*a*); see Figs. 1 and 2, (*b*) and (*c*)], the actual error, in the case of measurement data, ranged from 2.4 to 3.7″ for a least-squares parabola ($m = 2$) and from 2 to 6″

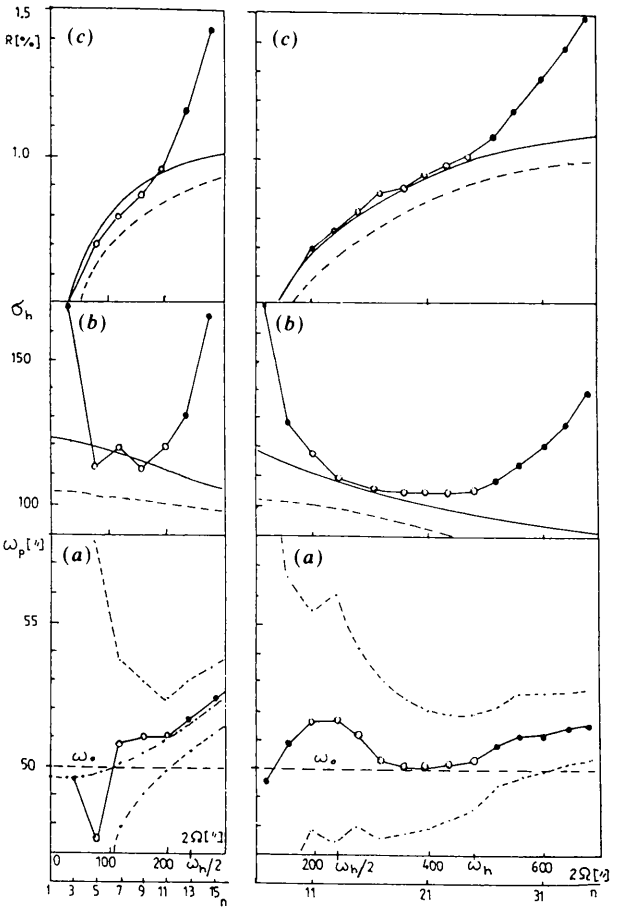

Fig. 1. Peak-position determination by a polynomial approximation, as dependent on the scanning range 2Ω, for real measurement data. A constant scanning step, $\Delta_\omega = 20''$, and hence a changeable number of points, n, were used. The left parts of the figure relate to a least-squares parabola ($m = 2$); the right parts to a sixth-degree polynomial ($m = 6$). (*a*) The mean values of the peak positions determined (o—o), limits of successive confidence intervals (·---·), and an approximate value of the 'true' peak position ($\omega_P = 49.9''$; *cf.* § 4.4). For $m = 2$ the mean values have been approximated with a parabola (·--·) (as described in § 5). (*b*), (*c*) Values of statistical parameters of the goodness of fit, σ_h and R, respectively. The values obtained in the course of calculations by least-squares fitting are marked ◦ while the corresponding expected values are drawn with a continuous line (real counting statistics) and broken line (Poisson counting statistics). The filled circles • relate to cases in which the statistical criterion [(3)] is not satisfied.

for a sixth-degree polynomial ($m = 6$). In the case of computer-simulated data (Fig. 2; the lower level of statistical errors of observations), the actual error ranged from 1.2 to 3.5″ for $m = 2$ and from 1.3 to 10.6″ for $m = 6$. Thus, the actual error is no less for a sixth-degree polynomial (in general, a higher-degree polynomial) than for a parabola.

The question arises of whether it is possible to achieve the accuracy desired. How can the total error be reduced? A comparison of (*a*), (*b*) and (*c*) of Fig. 1 with the corresponding parts of Fig. 2 reveals that the equality of σ_h and $\hat{\sigma}_h$ (or R and $\hat{R}$) [(3)] is a necessary but not precise enough indicator to show how to select the scanning range to minimize the actual error. It might be added that neither minimum σ_h nor minimum R, criteria commonly used in practice, can serve as such a precise indicator. In particular, the discrepancy factor R decreases when $n - m$ decreases [see paper I, (30*c*) and (34)], reaching zero when the approximation becomes an interpolation. This manifests itself in an enormous increase of the actual error.

Let us consider now two components of the actual error: the statistical error (here defined by the sample variance or the sample standard deviation) and the systematic error (bias), *i.e.* the difference between the expected value (in practice, the sample mean) and the 'true' (here, known) value. For real measurement data the standard deviation of the peak position determined varied from 2.3 to 4.8″ for $m = 2$ and from 2.9 to 7.7″ for $m = 6$; for computer-simulated data it varied from 1.6 to 5.2″ for $m = 2$ and from 1.5 to 7″ for $m = 6$. The minimum values are thus comparable for $m = 2$ and $m = 6$.

The variance of the peak position of a least-squares parabola is described by the formula (Wilson, 1965*a*; notation of the present paper)

$$\sigma^2(\omega_P) = (\omega_h^2/Hn)0.75/X^2[v''(u_P)]^2, \qquad (10)$$

where H is the peak intensity, ω_h is the half-width, $X = 2\Omega/\omega_h$ is the standardized scanning range and $v''(u_P)$ is the second derivative at the peak of the underlying shape function. Thus, the dispersion of results is particularly remarkable for a small number of points, a narrow scanning range and a flat-topped profile [small value of $v''(u_P)$]. The latter two effects can be observed in Fig. 3, in which results of an additional test based on computer-simulated data are shown. A large enough and constant number of points ($n = 21$) and a changeable scanning step were used in the test. As shown in Fig. 3, the statistical error increases rapidly at narrower scanning ranges (here, at $2\Omega = 50''$). That is because [*cf.* (10)] X decreases and because the 'observed' $v''(u_P)$ tends to

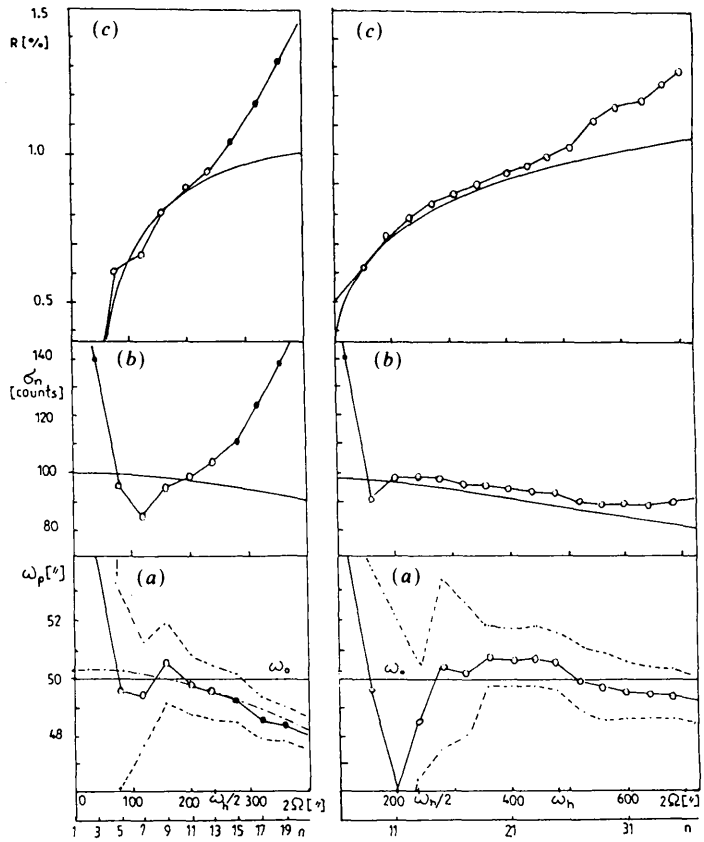

Fig. 2. Peak-position determination by a polynomial approximation, as dependent on the scanning range 2Ω, for computer-simulated data. The figure legend is the same as that of Fig. 1. An exception is that the true peak position, $\omega_P = 50.0''$, is in this case accurately known.

zero (intensities near the peak are almost equal). Therefore, the 'flat-top effect' causes the variance of the peak position for a narrow scanning range to be greater than that estimated from (10) for a fixed shape of the profile. The variance of the peak position of polynomials of higher (and even) degrees ($m = 4, 6, 8$) is described by the semi-empirical dependence (Thomsen, 1974; notation of the present paper)

$$\sigma^2(\omega_P) \simeq (\omega_h^2/Hn)0.017m^3(\tan^{-1} X)/X^3, \quad (11)$$

derived for $0.5 \leq X \leq 1$, under the assumption that the underlying profile is of Cauchy shape. Furthermore, the variance $\sigma^2(\omega_P)$ for given 2Ω is greater for polynomials of odd degrees ($m = 2k+1$, k is an integer) than for neighbouring polynomials of even degree ($m = 2k$). For example, for $m = 3$ it is about six times as big as for $m = 2$ (see Kirk & Caulfield, 1977).

With regard to the statistical error, it is therefore suggested that the scanning range is chosen to be as large as possible among those permitted by the statistical criterion [(3*a*)] and to avoid polynomials of odd degrees. Moreover, there is no special need for selecting polynomials of higher (and even) degrees, since the results are no more precise than those for a parabola, even if each polynomial is used in its proper range and a constant scanning step is used (so n increases for increasing m values). Equations (10) and (11) show some other possibilities of reducing the variance (see paper I, § 1.2) and so of reducing the actual error up to the limit defined by the systematic error (bias) [*cf.* (7), (8*b*)]. The basic problem is then to evaluate and reduce the bias. As shown in Figs. 1, 2 and 3, even within carefully selected scanning ranges there is an appreciable dependence of the mean value of the peak position on the scanning range, so it is difficult to avoid the likely bias. For real measurement data (Fig. 1) the largest bias [(8*b*)] and the stability were as follows [the stability, defined by (9), is given in parentheses]: 1.6″ (0.012) for $m = 2$ and 1.6″ (0.006) for $m = 6$; those obtained for computer-simulated data were: 0.7″ (0.007) for $m = 2$ and 4.6″ (0.010) for $m = 6$. Thus, although sixth-degree polynomials and, in general, higher-degree polynomials provide acceptable (in a statistical sense) results within a wide range of scanning ranges they do not give any supposed advantages over a parabola, such as less biased and more stable outcomes, even if each polynomial is used in its best range 2Ω. So it is difficult to achieve the accuracy of 1″ using polynomial approximation since the bias alone is likely to exceed the limit. Only in the simplest case of a parabola can the bias of the peak position be estimated, and thence reduced by extrapolation, by use of the formula (Wilson, 1965*a*)

$$\Delta\omega_P = \bar{\omega}_P - \omega_0 = 2(\Omega^2/\omega_h)[v'''(u_P)/v''(u_P)], \quad (12)$$

where $v'''(u_P)$ is the third derivative of the shape function at the peak. However, the underlying diffraction profile must be known to obey the formula. Therefore, we shall return to the problem (§ 5) after an extensive discussion of various possible shape functions (§ 4).

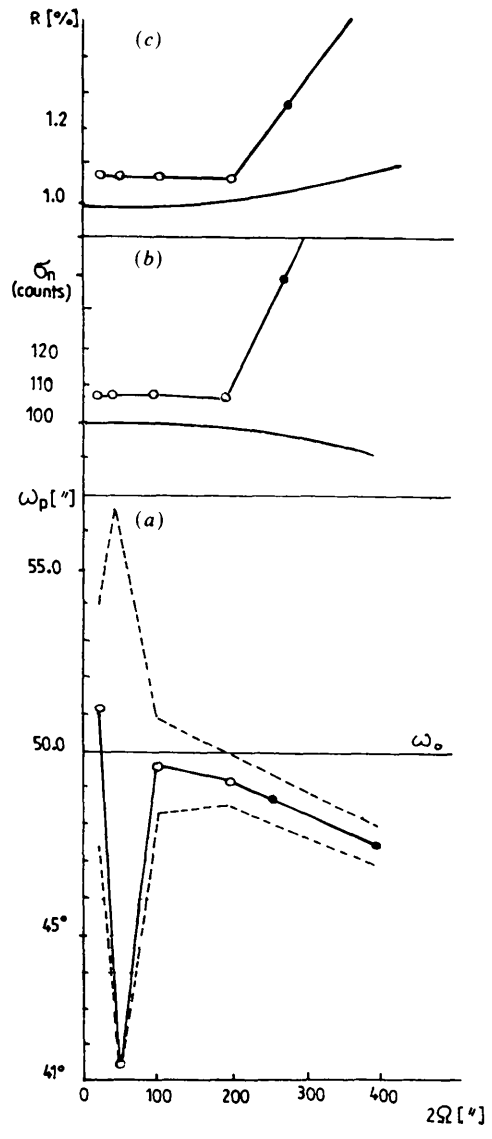

Fig. 3. Peak-position determination by an approximation with a parabola, as dependent on the scanning range 2Ω, for computer-simulated data. A constant number of points, $n = 21$, and so a changeable scanning step Δ_ω have been used. The symbols and the arrangement of the figure are analogous to those of Figs. 1 and 2.

4. The use of shape functions

4.1. *Preliminary remarks*

From previous considerations (§ 3), shape functions, in view of the problem in question, could be

Table 1. *Relations between the parameters u_1, u_2 of the collimator function and the shape parameters of approximating functions*

$m = 5 =$ constant, $n = 501 =$ constant, $\alpha = 0.1$, $[\chi^2_{n-m,\alpha/2}/(n-m)]^{1/2} = 1.052$. Numbers given in parentheses denote standard deviations of parameters.

The collimator function		The pseudo-Voigt function		The Pearson VII function		The Johnson distribution [(2)]	
u_1	u_2	c	$\sigma_h/\hat{\sigma}_h$	m	$\sigma_h/\hat{\sigma}_h$	b	$\sigma_h/\hat{\sigma}_h$
0.05	0.00	0.996 (4)	0.97	1.001 (3)	0.97	0.688 (3)	1.29
0.10	0.00	0.994 (4)	0.96	1.003 (3)	0.96	0.690 (3)	1.29
0.20	0.00	0.986 (4)	0.96	1.008 (3)	0.97	0.694 (3)	1.32
0.50	0.00	0.935 (4)	0.99	1.046 (4)	1.05	0.724 (4)	1.52
0.50	0.20	0.926 (4)	1.00	1.053 (4)	1.08	0.730 (4)	1.56
1.00	0.00	0.804 (4)	1.12	1.166 (7)	1.48	0.815 (6)	2.14
1.00	0.50	0.758 (5)	1.28	1.218 (9)	1.78	0.854 (8)	2.37
2.00	0.00	0.561 (5)	1.35	1.562 (20)	2.42	1.084 (14)	2.90
2.00	1.00	0.461 (8)	2.46	1.888 (43)	3.60	1.285 (25)	3.96

applied not only as approximating functions (*i.e.* for a direct approximation of the measurement data) but also as models of the shape of the diffraction profile [see (12)] to improve the results obtained using a parabola.

However, the use of shape functions as approximating functions and 'models' of the diffraction profile creates new problems. Although the statistical criterion [(3*a*)] based on real counting statistics (see paper I, § 2.6) is the only one needed to select the best model of the diffraction profile in the case of polynomial approximation, where the angular range and/or the degree of polynomial is to be chosen, it may be insufficient in the case of more complex and specific descriptions. In particular, it is required that the function to be used should be continuous at its peak (see paper I, § 2.5) and should give a good approximation to well founded models of the diffraction profile. The above requirements and their meaning for accurate peak-position determination will be considered in §§ 4.2, 4.3, 4.4 and 5.

4.2. *Shape functions and physical models*

To test the extent to which shape functions mentioned in § 4.1 are capable of providing a satisfactory approximation to an arbitrary collection of data, one can check either the agreement between various sets of experimental data and approximating shape functions or the agreement between a model (or models) of the diffraction profile, representing various possible measurement data, and approximating shape functions. The latter approach was applied in test calculations reported below. In the tests, two functions were used as models of the diffraction profile, a function describing the effect of in-plane collimation [paper I, (14*a*), (14*b*), (14*c*)] and a Voigt function [paper I, (16*a*), (16*b*)], within large ranges of parameters defining their shapes. The functions discussed in paper I, § 2.3 describe a great variety of experimental profiles. Three approximating functions were selected to be fitted to the models: a pseudo-Voigt function [paper I, (17*a*)], a Pearson VII function [paper I, (18*a*), (18*b*)] and one of the Johnson distributions [(2)]. In the tests, all the functions mentioned were used in their basic symmetric form to compare their shapes separately from the problem of asymmetry. A random component was added to the calculated values with the assumption of Poisson statistics for 'recorded' counts and the peak intensity, with $H = 10\,000$ counts. To minimize the error due to a finite truncation level, wide enough ranges of arguments were used, so the intensities at the ends of the ranges did not exceed 0.5% of the maximum intensity. The goodness of fit was tested by comparing the mean standard deviations σ_h obtained in the course of calculations by least-squares fitting with their expected values $\hat{\sigma}_h$ [*cf.* paper I, (30*a*), (32)], using χ^2 statistics (as described in § 2.2 above). Since the number of parameters was $m = 5$ for all approximating functions, and the number of observations ranged from $n = 457$ to $n = 605$, the highest values of the ratio $\sigma_h/\hat{\sigma}_h$ permissible for a model at the significance level $\alpha = 0.1$ were estimated from (3*b*).

Results of the approximation of the first function (describing the in-plane collimation) by the shape functions selected are given in Table 1; those that relate to the Voigt function are given in Table 2. With regard to the results given in Tables 1 and 2, the following conclusions might be drawn:

1. The best of the approximating functions considered is the pseudo-Voigt function; it gives the lowest values of the ratio $\sigma_h/\hat{\sigma}_h$. If the broadening of the diffraction profile, in relation to the original Cauchy-shaped profile, is not very large, *i.e.* $u_1 \leq 0.5$ (Table 1) and $r < 0.5$ (Table 2), the goodness of fit is fully satisfactory and the form of the model function can be replaced by the shape function. For larger broadenings, the ratios $\sigma_h/\hat{\sigma}_h$ are greater than expected, so the approximation is unsatisfactory.

Table 2. *Relations between the parameter r defining the shape of the Voigt function and shape parameters of approximating functions*

Numbers given in parentheses denote standard deviations of parameters.

	The Voigt function		The pseudo-Voigt function		The Pearson VII function		The Johnson distribution [equation (2)]	
r	n	$(\sigma_h/\hat{\sigma}_h)_{max}$	c	$\sigma_h/\hat{\sigma}_h$	m	$\sigma_h/\hat{\sigma}_h$	b	$\sigma_h/\hat{\sigma}_h$
0.05	457	1.054	1.000 (4)	0.97	1.000 (3)	0.97	0.688 (4)	1.34
0.10	459	1.054	0.992 (4)	0.99	1.005 (3)	0.99	0.687 (4)	1.39
0.20	465	1.054	0.976 (4)	1.01	1.016 (3)	1.01	0.620 (4)	1.45
0.33	471	1.053	0.948 (4)	0.99	1.036 (4)	1.02	0.712 (4)	1.49
0.50	477	1.053	0.888 (4)	1.07	1.084 (4)	1.19	0.752 (5)	1.77
0.67	485	1.053	0.827 (4)	1.07	1.142 (6)	1.35	0.796 (6)	1.96
1.00	501	1.052	0.701 (4)	1.15	1.294 (10)	1.80	0.908 (8)	2.41
2.00	545	1.050	0.474 (4)	1.22	1.827 (24)	2.29	1.234 (15)	2.67
5.00	681	1.045	0.220 (4)	1.22	3.904 (88)	1.81	2.086 (34)	1.91
10.00	605	1.047	0.089 (3)	1.19	8.433 (168)	1.16	3.305 (48)	1.17

2. Somewhat worse results are obtained using the Pearson VII function. The ranges of parameter values of respective model functions for which a satisfactory agreement can be achieved are narrower than in the former case, $u_1 \leq 0.5$ ($u_2 = 0$) or $r \leq 0.33$.

3. For the Johnson distribution [(2)] and the model functions considered, the ratio $\sigma_h/\hat{\sigma}_h$ always exceeds its permissible value, irrespective of values of the parameters r, u_1 and u_2. Thus, the distribution can be considered to be 'different' from the given models.

One may doubt whether the conclusions concerning the approximation of the selected models of the diffraction profile by given shape functions are valid for all possible forms of the complete physical model of the diffraction profile, which is not available in analytical form. In the face of the lack of further and 'general enough' information on the model, this question remains open. Otherwise, similar tests to those described in the paper might be performed for other models and other shape functions.

4.3. *Preliminary selection of shape functions*

The functions used in the tests reported in § 4.2 were taken within large ranges of arguments. Since for narrower ranges, say no larger than two half-widths, a better agreement between a model (or a set of measurement data) and an approximating function is possible, it was decided to perform test calculations on all the shape functions mentioned in § 4.2.

The shape functions selected, a Pearson VII function and a pseudo-Voigt function, in contrast to the asymmetric distributions of Johnson, are defined in an ideal symmetric form and so need a correction to allow for asymmetry. Various forms of such corrections were discussed in paper I, § 2.5. Some of them, *viz* the simplest and most convenient for application - a linear multiplier, a quadratic multiplier modified by the function 'sign', an exponential multiplier and a pair of split functions - were used in the present tests.

To examine the effect of the corrections for asymmetry on the peak-position determined, further test calculations were performed in which a small and constant number of points ($n = 11$), and therefore a changeable scanning step, were used. As shown in Fig. 4, the highest accuracy and stability are obtained with a linear multiplier; a modified quadratic multiplier gives worse results. In the case of split functions, there is a remarkable dependence of the peak position on the scanning range. In consequence, the latter description has been rejected as useless for the purpose of the paper. Moreover, it is doubtful whether an approximation with the split functions, leading to such a large error in the peak position (and so in the Bragg angle), could be used for other purposes. It should be added that the statistical criterion [(3a)] is not sensitive enough to such formal defects. As a consequence (see Fig. 4; black points denote results rejected), some highly biased results may be accepted while other more correct results may be rejected by the goodness of fit test.

4.4. *Peak-position determination using selected shape functions*

Outcomes of test calculations, more detailed than those in § 4.3, in which selected shape functions [Johnson's distribution given by (2), a Pearson VII and a pseudo-Voigt function] and selected asymmetric factors (linear or modified quadratic) were examined, are shown in Figs. 5 and 6. With regard to the form of the correction for asymmetry to be used with the basic symmetric function (here a Pearson VII or a pseudo-Voigt function), present tests confirm the superiority of the linear factor, which ensures the continuity of the resultant description, over the modified quadratic factor, which leads to a discontinuity at the peak.

As shown in Figs. 5 and 6, the mean (over 2Ω) value of the peak position determined for each approximating function, even when the statistical

criterion [(3a)] is satisfied, is different, while the stability of results is in each case much higher (of the order of 0.001) than for polynomials. For real measurement data (Fig. 5) the range of the mean values of the peak position and the stabilities obtained [the latter, defined by (9), given in parentheses] were as follows: 47.4-48.5″ (0.003) for the Johnson distribution [(2)]; 49.2-49.6″ (0.002) for the Pearson VII function with the modified quadratic multiplier $1-\text{sign}\,(u)\varphi u^2$ and 48.9-49.8″ (0.002) for the function with the linear multiplier $1+\xi u$; 48.7-49.6″ (0.002) for the pseudo-Voigt function with the modified quadratic multiplier and 49.8-50.5″ (0.001) for the function with the linear multiplier.

Thus, while determining the 'true' peak position, one has a choice between taking the average value of all the results, *i.e.* of all the mean values (for different functions) permitted by the statistical criterion [here, 49.20 (64)″] or taking the mean (over 2Ω) value of the results obtained by using one selected well founded description. With regard to results of the test calculations and some preliminary information on the descriptions used (gathered in §§ 4.1, 4.2, 4.3; see also paper I, §§ 2.3, 2.4, 2.5), the latter approach seems completely justified. The only description, from among all those considered, that is simple, continuous at its peak and closely related to physical models of the diffraction profile - a

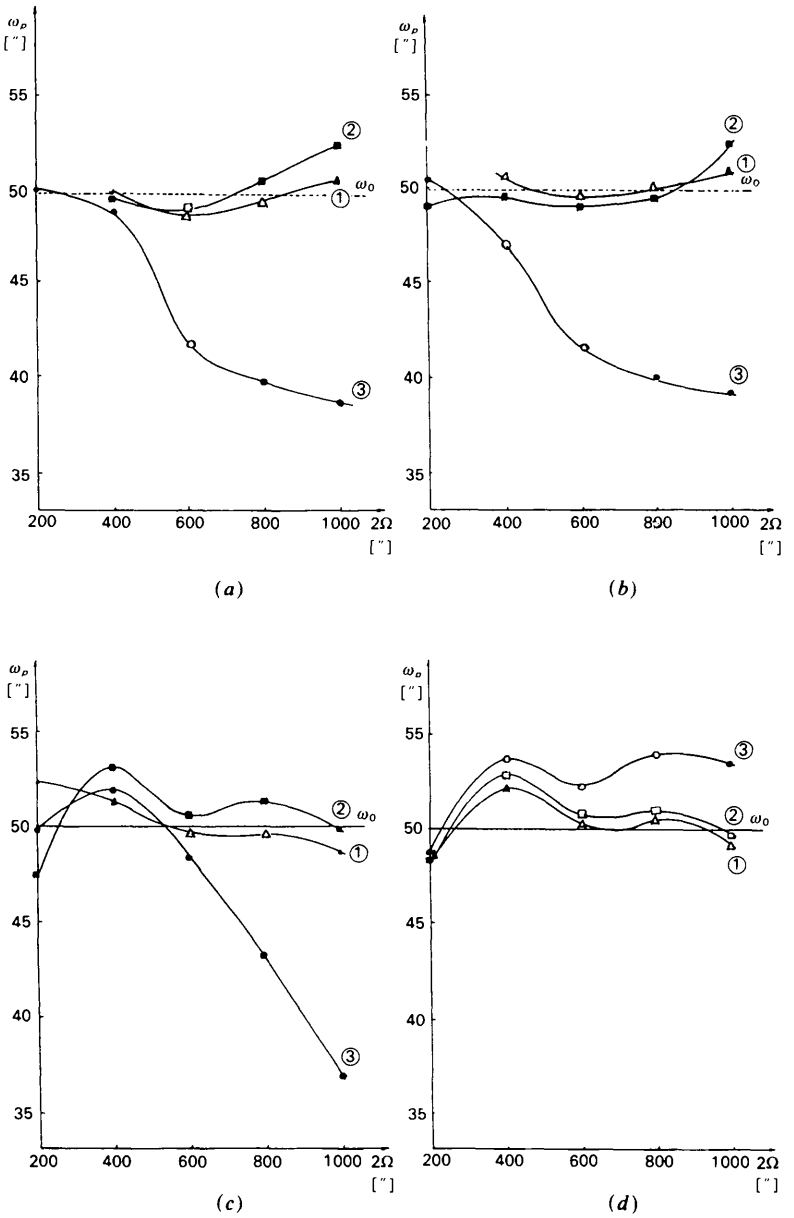

Fig. 4. The dependence of the peak position determined on the scanning range for various descriptions of asymmetry: ① a linear factor, ② a quadratic factor, ③ a pair of split functions. (*a*), (*c*) Pearson VII function; (*b*), (*d*) pseudo-Voigt function. Results shown in (*a*) and (*b*) are based on real measurement data while those in (*c*) and (*d*) are based on computer-simulated data. A constant number of measurement points, $n=11$, was used. The open points △ and □ relate respectively to a linear asymmetric factor and a modified quadratic factor. The filled points (▲, ■, ●) relate to cases in which the statistical criterion [(3a)] is not satisfied.

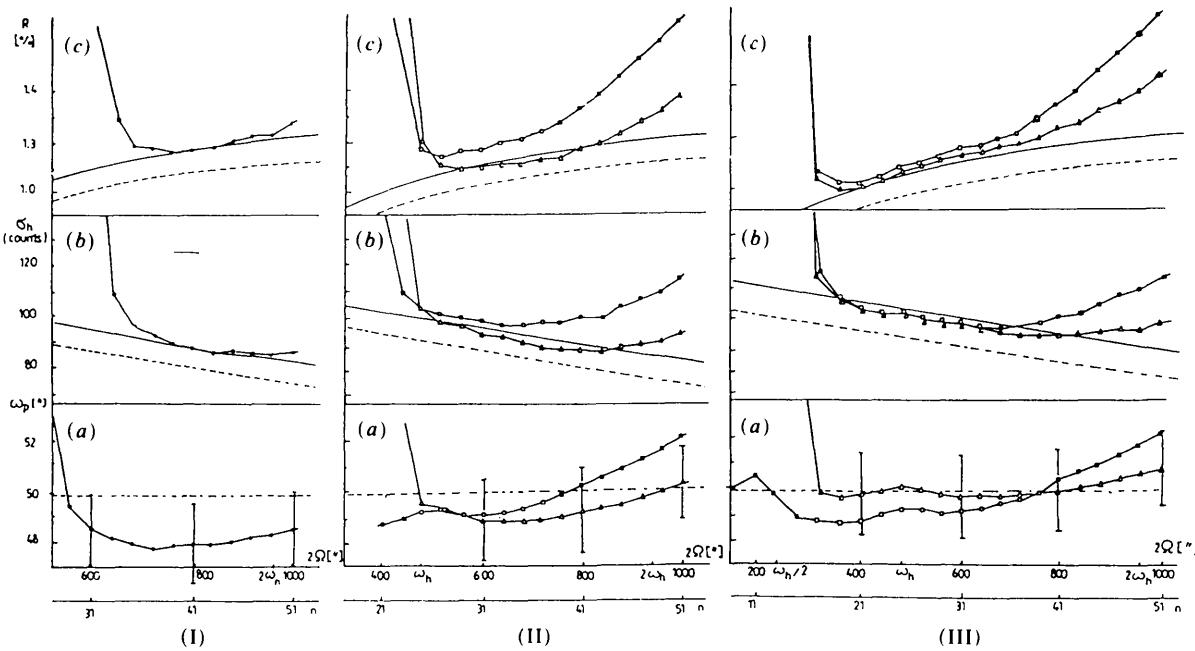

Fig. 5. Peak-position determination based on the approximation of real measurement data by means of various shape functions dependent on the scanning range. Diagrams (I) relate to the Johnson (1949) distribution [(2)]; diagrams (II) relate to the Pearson VII function; diagrams (III) relate to the pseudo-Voigt function. (*a*) The dependences of the mean peak position on the scanning range (broken lines). Limits of the confidence intervals ($\alpha = 0.1$, $\gamma = M - 1 = 9$) are marked on perpendicular sectors driven for selected scanning ranges ($2\Omega = 200, 400, \ldots, 1000''$). The horizontal dashed line (---) at $\omega_P = 49.9''$ shows the approximate 'true' peak position. (*b*) The standard deviation σ_h. (*c*) The *R* factor. Values resulting from calculations by least-squares fitting are marked with points; corresponding estimated values are marked with a continuous line (real counting statistics) and broken line (Poisson counting statistics). The open points △ and □ relate respectively to a linear asymmetric factor and a modified quadratic factor. The filled points (●, ▲, ■) relate to the cases in which the statistical criterion [(3*a*)] is not satisfied. A constant scanning step, $\Delta_\omega = 20''$, was used.

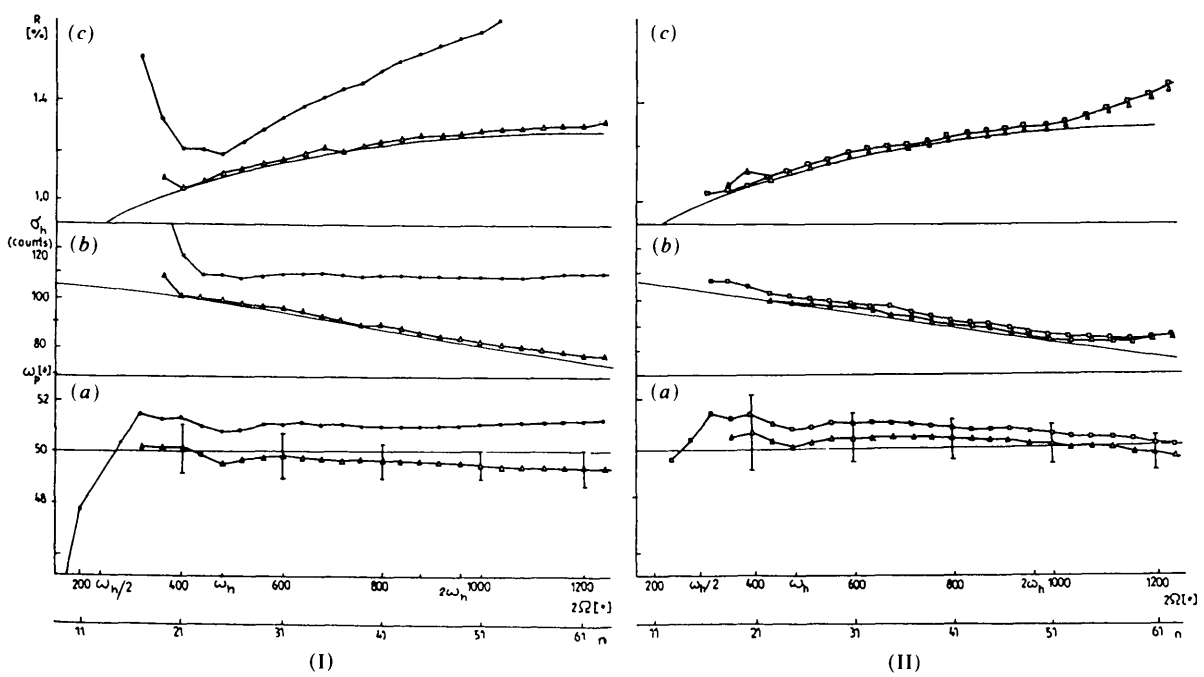

Fig. 6. Peak-position determination based on the approximation by means of a shape function dependent on the scanning range. Computer simulated data were used. Diagrams (I) relate to the Pearson VII function; diagrams (II) relate to the pseudo-Voigt function. Further details of the figure legend are as in Fig. 5. An exception is that the true peak position, $\omega_P = 50.0°$, is accurately known in this case.

pseudo-Voigt function with a linear multiplier allowing for asymmetry - gives at the same time the most stable results. Therefore, the mean value (over 2Ω) of the peak positions obtained using the above approximating function, $\omega_0 = 49.96\ (23)''$, was accepted as a definition of the 'true' peak position. The maximum bias in this case was 0.54″.

With regard to tests based on computer-simulated data, the effect of the basic form of the approximating function (a pseudo-Voigt or a Pearson VII function) on the peak positions determined was less evident than in the case of real measurement data, whereas the form of correction for asymmetry was of primary importance. Results obtained using a Pearson VII function and a pseudo-Voigt function, with a linear factor in each case (Fig. 6), were comparable: the maximum bias [*cf.* (8*b*)] and the stability [the latter, defined by (9), is given in parentheses] were 0.6″ (0.0009) and 0.8″ (0.0009), respectively; those obtained using a pseudo-Voigt function with a modified quadratic multiplier were evidently worse, 1.5″ (0.0013), while the ones obtained using a Pearson VII function with the quadratic multiplier (strongly biased, though stable) were rejected because of unsatisfactory goodness of fit. It is interesting to note that the average value of all the results permitted by statistical criteria was 49.24″, so the mean bias for such an approach was −0.76″ and thus twice as big as in the case of the individual functions used with a linear multiplier (−0.34″ for a Pearson VII function, +0.36″ for a pseudo-Voigt function). Therefore, the former approach, based entirely on statistical criteria, has no advantages over the latter, in which all criteria considered are taken into account.

The standard deviation of the peak position determined using shape functions ranged from 2.6 to 3.6″ for real data and from 1 to 2″ for computer-simulated data and the Poisson counting statistics, so was, on average, not greater than in the case of polynomial approximation. Although the likely actual error in peak position for a single set of data and a single peak-position determination (*i.e.* for one scanning range) is greater than permitted, it is possible to reduce it by repeated measurements or by other means, as mentioned in paper I [§ 1.1, points (i), (ii), (iii), (iv)], up to the limit defined by the bias (here $\simeq 0.5''$).

5. Refinement of results obtained using a least-squares parabola based on the proper model of the measured profile

As one may conclude from results of the test calculations reported in § 3, related to a parabola, and from (12), which explains them, the best approach to the accurate peak-position determination in this case is not to avoid bias by selecting a very narrow scanning range (that would lead to a rapid increase of the peak variance) but to reduce both the bias and the statistical error by using all information available concerning the method and the model. As will be shown below, the best form of the model of the diffraction profile is of primary importance for such a task.

Let us assume that the diffraction profile can be described by

$$v(u) = (1 + \xi u) v_s(u), \tag{13}$$

where $v_s(u)$ is a symmetric function, such that $v_s(0) = 1$, $v_s'(0) = 0$, $v_s''(0) = V_2$ ($-2 \le V_2 \le -1.4$ for well founded model functions). For moderate asymmetry, $\xi^2 \ll 1$ (usually $\xi = 0.1$, 0.2), the peak position of $v(u)$ is $u_P \simeq -\xi / V_2$, so $u_P^2 \ll 1$, $v''(u_P) \simeq V_2$, $v'''(u_P) \simeq 3\xi V_2$. As a consequence, (12) will take the form

$$\omega_0 = 3\xi X\Omega + \omega_P = 6(\xi/\omega_h)\Omega^2 + \omega_P = E\Omega^2 + F, \tag{14}$$

where $E = 6\xi/\omega_h$ and $F = \omega_P$ are parameters that are independent of the scanning range and of the shape of the underlying symmetric function $v_s(u)$. These can be estimated from a set of ω_0 values determined for various scanning ranges. The extrapolation of results to $\Omega = 0$ leads, theoretically, to the best peak position, $\omega_0 = F = \omega_P$. In practice, in the case of the data used in the present paper (described in § 2.1), the residual bias after the extrapolation of results obtained using a parabola [(14)] was less than 0.25″ [see Figs. 1(*a*) and 2(*a*)].

As far as the class of models assumed above leads to an interpolation of results with a parabola (14)], and so to final results comparable to those obtained using the well founded shape functions [see § 4], models in which some other descriptions of asymmetry are used may lead to false results.

For the shape functions in the form

$$v(u) = (1 - s\varphi u^2) v_s(u) \tag{15}$$

and for split functions described by the formula

$$v(u) = v_s(u^*), \tag{16}$$

where $u^* = u/(1 - s\varphi)$, $s = \mathrm{sign}(u)$ and $v_s(u)$ is as above, we have $u_P = 0$, $v''(u_P) \simeq V_2$ and $v'''(u_P) = 0$. Equation (12) then takes the form

$$\Delta\omega_P = 0, \tag{17}$$

so does not allow for any bias and is not capable of expressing the observed dependence of the peak position on the scanning range.

The above results serve as an additional argument for using a linear multiplier [(13)], rather than a modified quadratic multiplier or a pair of split functions, to model the asymmetry.

6. Concluding remarks

It was shown that the high-accuracy peak-position determination must be based on a reasonable physical and statistical model of the measured diffraction profile.

Knowledge of variances of recorded counts (their mean level, at least) is necessary for providing an objective criterion of the goodness of fit [(3*a*)]. However, the satisfactory agreement, in a statistical sense, between the measurement data and the approximating function is no guarantee of accurate enough results. The differences obtained for different approximating functions, even if well fitted to the data, and/or for different ranges of arguments may be greater than the accuracy desired (here 1″, *i.e.* 3×10^{-4}°).

In particular, the results obtained using polynomial approximation are strongly dependent on the scanning range, while those obtained for a reasonable (continuous at its peak) shape function show much higher stability. Therefore, it is suggested that the calculations be based on a well founded shape function rather than on a polynomial. The precision of all the results obtained is comparable, if each function is used in an appropriate scanning range, and is dependent on the statistics of recorded counts.

As results from the tests performed, the best shape function considered is the pseudo-Voigt function with a linear (or exponential) asymmetric factor. The function may be applied as an approximating function in direct calculations or may serve as a model for reducing the bias of results obtained for different scanning ranges using a least-squares parabola as an approximating function.

With regard to data analysis in the Bond method, in general, two basic approaches are possible.

(i) Determine the peak positions from diffraction profiles distorted by aberrations, calculate the Bragg angle and then introduce all necessary corrections using suitable formulae, mentioned in paper I, § 1.1.

(ii) Separate the underlying original profiles from the measured profiles by a deconvolution in relation to all the physical and apparatus functions causing distortions and then determine the peak positions of the pure profiles and hence the Bragg angle, free from systematic errors.

The first approach [(i)] was used here, as in the original paper by Bond (1960) and, consequently, in numerous following papers (reviewed by Gałdecka, 1992). Method (i) is simpler than its alternative, (ii), because it does not require detailed information on individual distributions as factors of the complete convolution model of the measured profile and does not involve additional complex calculations (deconvolution). There are, however, justifiable doubts as to whether the corrections being introduced in (i) have been estimated properly. The possible errors connected with the corrections might result, in this case, from the following reasons: (*a*) uncertainty of the levels of the factors causing distortions (such as the inclination of the crystal or the collimator in relation to the plane of diffraction, for example); (*b*) possible differences between distributions used for the theoretical evaluation of the aberrations and corresponding real distributions; and (*c*) nonadditive (in general) corrections calculated for the peak position, in contrast to those related to the centroid (see Wilson, 1963, 1965*b*). Recently, some of the corrections have been verified and recalculated (Berger, 1986; Härtwig & Grosswig, 1989).

The second method [(ii)] might offer an advantage over the former method [(i)] but only in the case when the respective individual distributions were accurately known. Moreover, one could foresee some computational problems, not occurring in (i), connected with the deconvolution and resulting from finite truncation limits, for instance. Some examples of the practical use of (ii), although not in connection with the Bond method, have recently been reported by Berger (1990) and Berti (1991). It would be interesting to compare the possibilities of the second approach [(ii)] with the Bond (1960) method.

However, in both case (i) and case (ii), the proper method of the peak-position determination is required and the present conclusions are valid.

The author is indebted to Professor A. J. C. Wilson FRS for stimulating suggestions, a critical reading of the manuscript and valuable discussions. The author is also grateful to the referees for helpful suggestions on the presentation. This work is supported by grant 2 0415 91 1 from the Polish State Committee for Scientific Research.

References

ABRAHAMS, S. C. (1969). *Acta Cryst.* A**25**, 165-173.
BERGER, H. (1986). *J. Appl. Cryst.* **19**, 34-38.
BERGER, H. (1990). 14th Conference on Applied Crystallography, Cieszyn, Poland, 5-8 August 1991. Poster session.
BERTI, G. (1991). Proc. First Eur. Powder Diffr. Conf. Munich, 14-16 March 1991. Abstract 01.2.
BOND, W. L. (1960). *Acta Cryst.* **13**, 814-818; erratum: (1975), A**31**, 698.
GAŁDECKA, E. (1985). *Structure and Statistics in Crystallography*, edited by A. J. C. WILSON, pp. 137-149. New York: Adenine Press.
GAŁDECKA, E. (1992). *International Tables for Crystallography*, Vol. C, edited by A. J. C. WILSON, ch. 5.3. Dordrecht: Kluwer Academic Publishers.
GAŁDECKA, E. (1993). *Acta Cryst.* A**49**, 106-115.
HAMILTON, W. C. (1964). *Statistics in Physical Science.* New York: Ronald Press.
HÄRTWIG, J. & GROSSWIG, S. (1989). *Phys. Status Solidi A*, **115**, 369-382.
JOHNSON, N. L. (1949). *Biometrika*, **36**, 149-169.
KIRK, D. & CAULFIELD, P. B. (1977). *Adv. X-ray Anal.* **20**, 283-289.
ŁUKASZEWICZ, K., KUCHARCZYK, D., MALINOWSKI, M. & PIETRASZKO, A. (1978). *Krist. Tech.* **13**, 561-567.
THOMSEN, J. S. (1974). *X-ray Spectroscopy*, edited by L. V. AZÁROFF, ch. 2. New York: McGraw-Hill.
WILSON, A. J. C. (1963). *Mathematical Theory of X-ray Powder Diffractometry.* Phillips Technical Library. Eindhoven: Centrex.
WILSON, A. J. C. (1965*a*). *Br. J. Appl. Phys.* **16**, 665-674.
WILSON, A. J. C. (1965*b*). *Röntgenstrahl-Pulverdiffraktometrie. Mathematische Theorie.* Eindhoven: Centrex.