Acta Cryst. (1949). **2**, 65

The Accuracy of Electron-Density Maps in X-ray Analysis with Special Reference to Dibenzyl

By D. W. J. Cruickshank

Chemistry Department, The University, Leeds 2, England

(*Received* 22 *July* 1948 *and in revised form* 25 *October* 1948)

Methods of correcting the systematic errors in electron-density maps due to using finite series of structure factors are discussed. Random errors in the maps are caused by experimental and computational rounding-off errors and by errors in the correction for finite series. Expressions are derived for the standard deviations of these random errors, which have a Gaussian probability distribution. It is shown how statistical significance tests can be used in the comparison of different bond-lengths, etc. These methods are illustrated by a detailed examination of the accuracy of the dibenzyl structure, which confirms Jeffrey's conclusions on shortened central bonds.

Other topics include discussions of the shape of the atomic peaks in maps and the variation of errors with limiting reciprocal radii.

Notation

α; $\alpha(h,k,l)$	The phase angle for the plane (h,k,l).
θ	Bragg reflexion angle $=\sin^{-1}(n\lambda/2d)$; or $2\pi(hx/a+ky/b+lz/c)$.
$\theta_{\text{max.}}$	The maximum reflexion angle.
θ_r	$2\pi(hx_r/a+ky_r/b+lz_r/c)$.
λ	Wave-length of radiation in A.
$\rho(x,y,z)$	Electron density at (x,y,z).
$\rho(r)$	Electron density at r.
σ	Standard deviation of an error.
$\sum_3$	$\sum_{-h}^{+h}\sum_{-k}^{+k}\sum_{-l}^{+l}$.
$\sum_3^\infty$	$\sum_{-\infty}^{+\infty}\sum_{-\infty}^{+\infty}\sum_{-\infty}^{+\infty}$.
a, b, c	Unit-cell axes in A.
A_h	$\partial\rho/\partial x$.
A_{hk}	$\partial^2\rho/\partial x\,\partial y$.
$f(s)$	Scattering factor at reciprocal radius s.
F; F_{hkl}	The structure factor.
F_c; $F_{\text{calc.}}(hkl)$	The calculated structure factor.
F_o; $F_{\text{obs.}}(hkl)$	The observed structure factor.
h,k,l	Plane indices.
m	Structure-factor multiplicity.
M	Number of atoms in the unit cell.
N	Atomic number.
P	Probability (see § 3).
r	Distance from the origin or a peak in A.
s	Distance from the origin of reciprocal space $=2\sin\theta/\lambda$.
V	Unit-cell volume.

1. Introduction

It was not until about thirteen years after W. H. Bragg, in 1915, had shown that the electron density in a crystal could be represented by a Fourier series of structure factors that the experimental technique was sufficiently advanced to permit of the application of the method, except in simple one-dimensional cases. It was then soon recognized (Bragg & West, 1930) that electron-density maps produced in this way differ from the true electron density in much the same way as an optical microscopic image differs from the object from which it is formed: (i) peaks in the maps (corresponding to atoms in the crystal) are surrounded by diffraction 'ripples'; and (ii) atoms which are closer together than about $0{\cdot}6\lambda/2\sin\theta_{\text{max.}}$ cannot be resolved as separate peaks. Although, owing to confusion of thought, it has sometimes been suggested that the *resolving power* as expressed above sets a limit to the *accuracy* with which the position of any atom can be determined, the condition (ii) does not, in fact, impose any serious limitation on crystal analysis since, by the use of three-dimensional methods, a peak separation of more than 1 A. can be ensured, and with the usual experimental conditions $0{\cdot}6\lambda/2\sin\theta_{\text{max.}}$ is considerably less than this.

The effect of the diffraction ripples in displacing peaks from their true positions was first examined by Booth (1945*a*, 1946*a*, 1947*a*), who at the same time studied the effects of experimental inaccuracies on peak positions. Booth gave methods by which the errors in a structure could be estimated, and his conclusions confirmed the possibility of obtaining the order of accuracy of atomic positions ($\pm 0{\cdot}01$ to $\pm 0{\cdot}05$ A.) claimed by various workers from the consistency of their results.

The methods suggested by Booth depend in part on assumptions which are not generally valid or on information not readily available; the present study, while following the same general mode of analysis, attempts to avoid these difficulties and to extend the scope of Booth's discussion.

As practical problems of accuracy at the present time arise most acutely in connexion with the determination of bond lengths in molecular substances, it seems desirable to apply the general methods obtained to a structure of this class. In the present case dibenzyl has been chosen because there are two sets of experimental data (Robertson, 1935; Jeffrey, 1947) and because of the interest of the results, and in the belief that so far as can be judged it is a typical organic structure; for our purpose it also has the advantage of having all seven (carbon) atoms in the asymmetric unit as nearly as possible equivalent in their electron content.

The whole of the following analysis is based on the assumption that the errors are small in comparison with the inter-atomic distances. No account has therefore been taken of 'homometric sets' (Patterson, 1944) of structures having the same amplitudes but with different phase relationships; the large errors due to this possibility are not likely to arise in practice and in any case their elimination is a matter of physics and chemistry rather than of mathematics.

A preliminary account of this work has been published already (Cox & Cruickshank, 1948).

2. Plan of this paper

The valid interpretation of fine detail in electron-density maps depends, fundamentally, on the proper use of statistical assessments of significance on the basis of the estimated errors (a matter which is quite separate from the use of statistical methods to estimate the errors). Since without a proper appreciation of this point detailed study of the various possible errors is almost valueless, it is discussed first in § 3.

In comparison with the true electron density, given by

$$\rho(x,y,z)=\frac{1}{V}\sum_{3}^{\infty}|F|\cos(\theta-\alpha), \qquad (2\cdot1)$$

electron-density maps are subject to three sources of errors, namely:

(i) errors in the experimental $|F|$ values;
(ii) effects of termination of the series at a finite (hkl) value; and
(iii) computational errors.

Part of Booth's discussion of (i) and (ii) was based on the assumption of a particular shape of atomic density distribution. This is examined with especial reference to dibenzyl in an Appendix and the conclusion is reached that the analysis cannot be reliably based on such an assumption. Though in the development from Booth's work this discussion is prior to the remainder of this paper, it has been given as an Appendix so as not to interrupt the main sequence of the present approach.

After a preliminary discussion of finite-series effects in § 5, various methods of treatment are considered in detail in §§ 6–10; the semi-analytical method of § 8 gives results in agreement with the diffraction effects observed by Bragg & West (1930) and by Robertson & Woodward (1940).

Formulae for assessing random errors are derived in § 11 and applied to dibenzyl in § 12, where comparisons are made between the random errors arising in two- and three-dimensional syntheses. Computational errors, which can be assessed by methods similar to those used for experimental errors, are examined in § 13; the effects of two- or three-figure accuracy in computation are dealt with, and the results of a complete analysis of the errors in computing the dibenzyl structure with Beevers & Lipson strips (1936) are given.

§ 14 examines the merits and defects of the analyses developed earlier in the paper, and the methods of assessing and correcting errors are applied in detail to dibenzyl in § 15. § 16 contains a brief discussion of the nature of the variation in errors likely to arise when structure analyses are carried out by different methods or by the same method to different limits, and finally § 17 summarizes the results of the paper and gives a set of rules for assessing the accuracy of any structure and for drawing significant conclusions about the results.

The reader primarily interested in assessing the accuracy of a given structure will probably find that §§ 3, 5, 10, 11 and 17 contain all that he requires.

3. Application of statistical significance tests to the results of structure analysis

Before any justifiable conclusions can be drawn from comparisons of the results of structure analyses, it is necessary to have an estimate of the errors in each case. This is especially important in the solution of questions relating to bond lengths in molecular substances, and unless a definite procedure is followed there are possibilities either that conclusions may be based on insufficient evidence or even that real differences may be dismissed as due to experimental errors.

We shall for the moment presuppose our general result that in electron-density maps the systematic errors (due to finite series) can be corrected for (except in certain details mentioned in § 14), and that the random errors (due to experimental and computational errors and to the method of correction for finite series) can be estimated. The probability distribution for these random errors is normal (or Gaussian), i.e. the probability that the error lies between x and $x+dx$ is

$$\frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}}e^{-x^2/2\sigma^2}\,dx, \qquad (3\cdot1)$$

where σ is the standard deviation of the error and is the quantity given by the formulae in succeeding parts of this paper.

With a distribution of this kind we can never be absolutely certain that the actual error is less than any finite quantity. (Apart from physico-chemical evidence setting an upper limit, this is not strictly true as there is an upper limit of error dependent in our case on the number of structure factors, but this number is sufficiently large not to affect the argument.) We have, therefore, to accept some standard less than certitude as a reasonable basis of argument.

An example will pose the problem more explicitly: a bond length A, having a standard deviation σ_A, is determined as greater than another bond length B, standard deviation σ_B, by an amount δl. Is there a real difference between A and B or might the observed difference easily occur by chance due to random errors? Common statistical practice (e.g. Brownlee, 1947) suggests the following procedure. Let P represent the probability that A could be observed as greater than B by chance although really equal to B; then

if $P \geqslant 5\ \%$ δl is not significant;
if $5\ \% > P > 1\ \%$ δl is of possible significance; and
if $P \leqslant 1\ \%$ δl is significant.

A further level of $P \leqslant 0{\cdot}1\ \%$ is sometimes used; δl then being said to be highly significant. The choice of these (or any other) levels is entirely arbitrary, but the above set is likely to be suitable for crystallographic purposes.

The following table gives δl in terms of $\sigma = (\sigma_A^2 + \sigma_B^2)^{\frac{1}{2}}$ for the various values of P for a Gaussian distribution (if $\delta l/\sigma = (\surd 2)\, x$,

$$P = \tfrac{1}{2} - \frac{1}{\surd \pi} \int_0^x e^{-t^2}\, dt = \tfrac{1}{2} - \tfrac{1}{2}\operatorname{erf} x;$$

by this definition of P the values include only the area under *one* tail of the probability distribution curve).

$P = 5\ \%$ $\delta l = 1{\cdot}645\, \sigma$,
$P = 1\ \%$ $\delta l = 2{\cdot}327\, \sigma$,
$P = 0{\cdot}1\ \%$ $\delta l = 3{\cdot}090\, \sigma$.

Thus for all practical purposes a difference of more than three times the standard deviation may be taken as real.

Until recently, in the absence of a quantitative treatment of the problem, means have been lacking to assess errors in structural analyses, and the frequently occurring statements of the type 'We regard the bond lengths given in this paper as reliable to ± 0·03 A.' have been based on judgements which, while perhaps not without justification, have lacked a common basis on which comparison might reliably be made. It is much to be hoped that in all precision analyses authors will adopt a definite convention of giving standard deviations of interatomic distances and will use numerically specified levels of significance in comparing them.

4. Summary of previous work on dibenzyl

As the data in the application of the theory of errors elaborated in this paper are taken chiefly from work on dibenzyl it will be convenient here to summarize previous work, in so far as it bears on our discussion.

The crystal structure of dibenzyl,

$$C_6H_5 . CH_2 . CH_2 . C_6H_5,$$

was first examined in detail by Robertson (1935), who found a unit cell of dimensions $a = 12{\cdot}77$ A., $b = 6{\cdot}12$ A., $c = 7{\cdot}70$ A.; $\beta = 116°$, containing two centro-symmetric molecules related by the space-group symmetry $P2_1/a$. He measured photometrically the absolute intensities of about 150 planes, and found the atomic co-ordinates by Fourier projections on the principal planes.

Jeffrey (1947) measured the intensities from 746 planes visually, and these, after conversion to an absolute scale by comparison with Robertson's results, were used in three-dimensional Fourier syntheses to determine the atomic co-ordinates. The bond lengths found by Jeffrey are shown in Fig. 1, in which the molecule is projected parallel to the c axis. The results of particular interest are the lengths of the three acyclic bonds, which are considerably shorter than the normal C–C single-bond length of 1·54 A. In conjunction with results from the crystal-structure analysis of geranylamine hydrochloride (Jeffrey, 1945) these are taken as indicating special structural properties associated with the system of two unsaturated groups separated by three single carbon bonds. From the internal consistency of the molecular dimensions Jeffrey concluded that the bond lengths in dibenzyl could be regarded as reliable to ± 0·01 A.

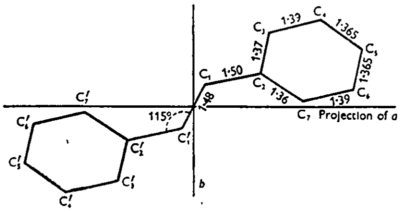

Fig. 1. Jeffrey's dibenzyl bond lengths.

5. Effects of finite series

Booth (1946*a*) has shown that the whole summation

$$\rho(x, y, z) = \frac{1}{V} \sum_3 |\, F \,| \cos (\theta - \alpha)$$

can be regarded as composed of M parts, each of which is completely in phase at one atom, however many terms there are in the summation. The position of one peak is therefore affected by disturbances from the other $(M-1)$ peaks. This idea of the allocation of the electron density to M parts is the key to the treatment of finite series, but it entails certain approximations which must be noted before we consider the character of the disturbances. A similar mode of analysis is adopted for hypothetical models in structure-factor calculations.

First, we must say what we mean by an atomic position in bond-length investigations. Atomic positions are defined to be those points where, in the electron density at the absolute zero of temperature,

$$\left.\begin{aligned} &\frac{\partial \rho}{\partial x} = \frac{\partial \rho}{\partial y} = \frac{\partial \rho}{\partial z} = 0, \\ \text{and}\quad &\frac{\partial^2 \rho}{\partial x^2},\ \frac{\partial^2 \rho}{\partial y^2} \text{ and } \frac{\partial^2 \rho}{\partial z^2} \text{ are negative.} \end{aligned}\right\} \qquad (5{\cdot}1)$$

This definition applies alike to postulated structures and electron densities in a crystal. However, no crystal is at zero temperature, and consequently the maxima of the electron density at temperatures other than zero are not necessarily at the same points as the atomic positions.

Calculations on postulated structures introduce a further complication. Strictly, the postulated structure should include allowances for asymmetry of electron distributions, electrons in bonds, etc., but for convenience it is taken as consisting of M distinct atoms, each of which is spherically symmetrical. This conventional process of splitting a structure into M parts will not, in general, give the original density when the M parts are superposed; in particular it does not necessarily ensure, even at absolute zero temperature, that the atomic-density distributions will be zero at inter-atomic-centre distances, i.e. the electron density resulting from the superposition of M distinct atoms does not necessarily have its maxima at the atomic positions. If this occurs at zero temperature, the peak shifts are spurious and are due to the conventional procedure. At zero temperature this approximation, if carefully carried out, will give substantially correct densities at the atomic centres and zero densities at great distances from any atoms. It will be in error at intermediate positions. At non-zero temperatures, part, at any rate, of the peak shifts on superposition are due to thermal motion, as we discussed above.

In the ensuing discussion on the correction of finite series we shall assume that the peak shifts on superposition at zero temperature are negligible, and that at other temperatures the peak shifts on superposition are the same as those in the actual electron density due to thermal motion.

Though the effect is automatically included in the various treatments for finite series, it is informative to calculate the displacement of the maxima of the electron density due to thermal motion for the dibenzyl CH group discussed in the Appendix. Using the data from Table A2 we find for one group that at

$$r = 1{\cdot}386\,\text{A.}\quad \partial\rho/\partial r = -0{\cdot}358\,\text{e.A.}^{-4},$$

and at $$r = 1{\cdot}540\,\text{A.}\quad \partial\rho/\partial r = -0{\cdot}196\,\text{e.A.}^{-4}.$$

Hence, since $\partial^2\rho/\partial r^2 = -80{\cdot}2$, two groups 1·386 A. apart will have their maxima shifted inwards

$$2\times 0{\cdot}358/80{\cdot}2 = 0{\cdot}009\,\text{A.},$$

and the distance between two groups 1·540 A. apart will apparently be shortened by $2\times 0{\cdot}196/80{\cdot}2 = 0{\cdot}005$ A. The thermal vibration in dibenzyl is large, so that perhaps these results can be regarded as indicating an upper limit to the effect.

The problem of correction of finite series is to eliminate the disturbances upon one peak from the other $(M-1)$. We shall examine three methods of calculating these disturbances, which, since the series is finite, have a periodic or rippling character. These methods are:

(i) semi-analytically from the f curve (§§ 6, 7, 8);
(ii) f_c synthesis (§ 9);
(iii) F_c correction synthesis (§ 10).

6. Semi-analytical calculation of rippling

To calculate the rippling due to observing only within the reciprocal sphere of radius s_0 we require to integrate

$$\rho(r) = \int_0^{s_0} f(s)\,4\pi s^2\,\frac{\sin 2\pi sr}{2\pi sr}\,ds. \qquad (6{\cdot}1)$$

Booth (1946*a*, 1947*a*), by taking $f(s) = N\exp.[-\pi^2 s^2/p]$ (A2·1), was able to calculate the maximum slopes in $\rho(r)$ and consequently to set an upper limit to the displacement of one peak by another. We shall comment in the Appendix on the use of such a scattering-factor curve, and conclude that it is not generally applicable.

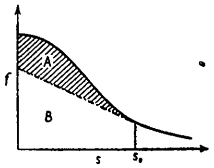

Fig. 2. Full line, $f_o(s)$; broken line, $f_e(s)$.

We can gain some insight into the effect of finite series by dividing up the area under the f curve into two parts (see Fig. 2). Let the dividing line be

$$f_e(s) = a + bs, \qquad (6{\cdot}2)$$

such that it touches $f_o(s)$ at the limiting reciprocal radius s_o. The whole integration of $f_o(s)$ is to be performed in two parts, zone A numerically and zone B analytically; since the integration is in three dimensions zone B will generally make the larger contribution to $\rho(r)$, since at large s the ordinates of A are small.

We require to integrate

$$I_B = \int_0^{s_0} (a+bs)\,4\pi s^2\,\frac{\sin 2\pi sr}{2\pi sr}\,ds. \qquad (6{\cdot}3)$$

Putting the ordinate at limiting reciprocal radius

$$f_o = a + bs_o,$$

and the slope at s_o $f'_o = b$, we find

$$I_B = \rho_B(r) = \cos 2\pi s_o r\left\{\frac{-s_o f_o}{\pi r^2} + \frac{f'_o}{2\pi^3 r^4}\right\} - \frac{f'_o}{2\pi^3 r^4} + \sin 2\pi s_o r\left\{\frac{f_o + sf'_o}{2\pi^2 r^3}\right\}. \qquad (6{\cdot}4)$$

With $f'_o = 0$ this is identical with the result obtained by James (1948, equation (8)).

The slope of I_B is

$$\rho'_B(r) = \cos 2\pi s_o r\left\{\frac{s_o}{\pi r^3}(3f_o + s_o f'_o) - \frac{2f'_o}{\pi^3 r^5}\right\} + \frac{2f'_o}{\pi^3 r^5} + \sin 2\pi s_o r\left\{\frac{2f_o s_o^2}{r^2} - \frac{1}{2\pi^2 r^4}(3f_o + 5s_o f'_o)\right\}. \qquad (6{\cdot}5)$$

It is apparent that $\rho_B(r)$ is approximately periodic with period $s_o r$ and that its amplitude steadily decreases with increasing r. For large r

$$\rho_B(r) \simeq -\frac{s_o f_o}{\pi r^2} \cos 2\pi s_o r. \tag{6·6}$$

The complete $\rho(r)$ is obtained by adding $\rho_A(r)$, calculated numerically, to $\rho_B(r)$.

The displacement δr of one peak by the disturbance from another is then given by

$$\rho'(r) = -\frac{\partial^2 \rho}{\partial r^2} \delta r, \tag{6·7}$$

where $\partial^2\rho/\partial r^2$ is the central curvature of the peak whose displacement is being investigated. Clearly the displacement is along the line of centres.

7. Application of the semi-analytical method

We can illustrate the general principles of the method by considering two atoms separated by a distance of 1·54 A., and also a benzene ring, using the data from dibenzyl.

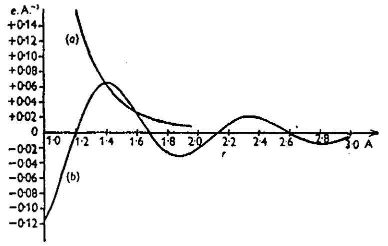

Fig. 3. (a) I_A for $2 \sin\theta = 1·65$ obtained by numerical integration. (b) I_B for $2 \sin\theta = 1·65$ calculated from equation (6·4).

There is no sudden termination of the series in dibenzyl, so that the results obtained may not be wholly applicable to the dibenzyl structure. It is also difficult to estimate what the equivalent s_o is: we shall assume a limiting value of $2 \sin\theta = 1·65$. The other relevant data for dibenzyl are

$$\left.\begin{array}{ll} \text{mean side of hexagon} & 1·375\text{A.}; \\ \text{central peak } \partial^2\rho/\partial r^2 & -34·5; \\ f_o \text{ at } 2\sin\theta = 1·65 & 0·317; \\ f_o' \text{ at } 2\sin\theta = 1·65 & -1·164. \end{array}\right\} \tag{7·1}$$

Fig. 3 shows $\rho_B(r)$ calculated from (6·4). The value of $\rho_A(r)$ for $2 \sin\theta = 1·65$ obtained by numerical integration is also plotted, and it is clear that $\rho_A(r)$ influences only neighbouring peaks.

For the diatomic system with $d = 1·540$ A. we have the following slopes at one peak due to the tail of the other peak

$$\begin{array}{rl} & -0·284\,\text{e.A.}^{-4} \text{ due to } \rho_B(r), \\ & -0·155\,\text{e.A.}^{-4} \text{ due to } \rho_A(r). \\ \text{Total} & -0·439\,\text{e.A.}^{-4}. \end{array}$$

Hence one peak is shifted 0·439/34·5 = 0·0127 A. inwards. The two peaks will therefore be observed as 1·5146 A. apart, an appreciable reduction in distance. Without detailed knowledge of the scattering factors and limiting reciprocal radii for many compounds one cannot assert a general rule that C–C single bonds will be observed short, but the example serves to illustrate the large effect that finite series can have.

Referring to Fig. 4 the symmetry of the system shows that the benzene ring will expand or contract as a whole. It will be sufficient to calculate the movement of peak A along the diameter AD due to the disturbances from the other atoms considered in turn.

Density changes at A:

Atoms B and F (each)	0·064	Due to ρ_B
C and E (each)	0·020	
D	−0·012	
B and F (each)	0·070	Due to ρ_A
Total	0·296 e.A.$^{-3}$	

Slope changes at A:

	Slope at A	Peak shift outwards along AD	
Atoms B and F (each)	0·086	0·0012	Due to ρ_B
C and E (each)	−0·037	−0·0009	
D	−0·039	−0·0011	
B and F (each)	−0·346	−0·0050	Due to ρ_A
Total		−0·0105 A.	

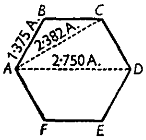

Fig. 4. Dimensions of benzene ring in dibenzyl.

We have obtained these shifts on the assumption that the true side of the ring was 1·375 A., but this was the observed size (with diffraction); the true size is thus the observed size minus the corrections calculated above. As the corrections are small compared with the period of the diffraction waves we need not recalculate the changes for the corrected positions. Making these corrections we obtain the mean side of hexagon equal to 1·385 A.

The mean observed peak height in the ring was 6·00 e.A.$^{-3}$. It should have been 6·00 − 0·30 = 5·70 e.A.$^{-3}$. This alteration in the peak density due to the diffraction effect is considerable, and must be properly taken into account when discussing the possibilities of detecting real density differences between various atoms of a molecule.

8. The semi-analytical method in two dimensions

The analysis can easily be extended to diffraction effects in plane projections. The Fourier integral for the representation of a circularly symmetrical atom in two dimensions is

$$\rho(r) = \int_0^{s_o} 2\pi s f(s)\, J_o(2\pi s r)\, ds. \tag{8·1}$$

Setting $f_e(s) = a + bs$, we get

$$\rho_B(r) = \frac{as_o}{r} J_1(2\pi s_o r) + \frac{b}{4\pi^2 r^3} \int_0^{2\pi s_o r} x^2 J_o(x)\, dx. \quad (8{\cdot}2)$$

In discussing the diffraction effects in NaCl Bragg & West (1930) pointed out the analogy with optical diffraction rings. They arrived at an expression

$$\rho(r) = \frac{cs_o}{r} J_1(2\pi s_o r), \quad (8{\cdot}3)$$

which may be obtained from (8·2) by putting $b = 0$. This expression gives the positions of the maxima and minima of the diffraction pattern almost correctly, as the maxima and minima of $J_1(2\pi sr)/r$ are close to the maxima and minima of $\frac{1}{r^3}\int_0^{2\pi s_o r} x^2 J_0(x)\, dx$.

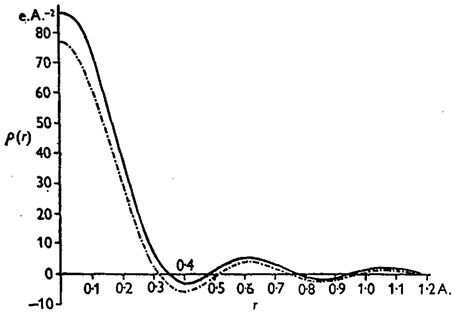

Fig. 5. Electron density for chlorine. Full line, Bragg & West synthesis. Broken line, density calculated from equation (8·2).

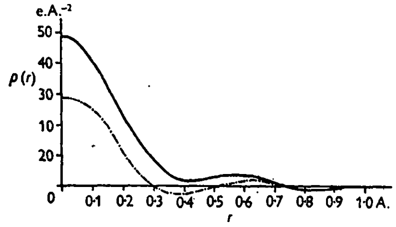

Fig. 6. Electron density for sodium. Full line, Bragg & West synthesis. Broken line, density calculated from equation (8·2).

Bragg & West calculated the electron densities of sodium and chlorine for a synthesis, based on the Hartree atom models, terminated at a particular reciprocal radius. The density plots they gave in Fig. 8 of their paper are reproduced here in Figs. 5 and 6, which also show the densities calculated from (8·2) using the data on f values given by Bragg & West. (They state that $\theta_{\text{max.}}$ was 30°, $\lambda = 0{\cdot}615$ A., but their results and tables are consistent with $\theta_{\text{max.}} = 42\frac{1}{2}°$.) The densities calculated from (8·2) closely follow those obtained in the synthesis, the difference decreasing as r increases, indicating a monotonic transform of zone A.

Fig. 7 shows the density of platinum given by Robertson & Woodward (1940), and that calculated from (8·2) using $2 \sin\theta = 1{\cdot}7$, $\lambda = 1{\cdot}54$ A., $f_o = 26$, $f'_o = -30{\cdot}8$. In this case, from the second zero outwards, the agreement between the curves is complete within the limits of graphical error.

9. f_c synthesis

As an alternative to the numerical integration of the f curve, the characteristic diffraction effects can be obtained by evaluating a synthesis, whose coefficients are the scattering factors for one atom, having the same terms as the F synthesis. The superposition of M f syntheses, the origin of each being at an atomic peak, will then give the required density changes and peak shifts.

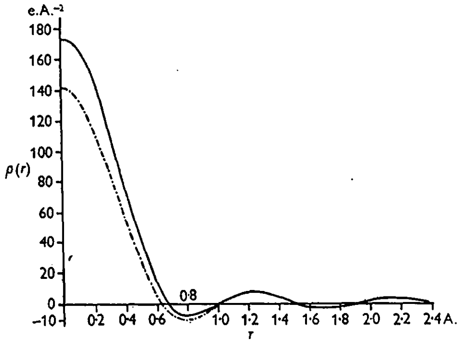

Fig. 7. Electron density for platinum. Full line, Robertson & Woodward synthesis. Broken line, density calculated from equation (8·2).

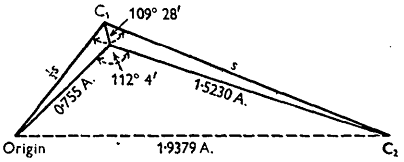

Fig. 8. Positions of central peaks in dibenzyl.

If the synthesis is terminated sharply at some limiting reciprocal radius, each f synthesis will be nearly spherically symmetrical and it will be sufficient to evaluate the densities along one line. However, if the synthesis is terminated because the reflexions become too weak to stand out from the general background scattering, so that the cessation of observations is irregular with unobserved planes even from small Bragg angles, then the calculations are laborious as the synthesis is not symmetrical. The density plots then obtained show ripples which do not have the periodicity and damping of the usual diffraction effects. In these cases the F_c syntheses discussed in the next section are preferable.

10. F_c synthesis

Booth (1946a) has suggested that, when the F_o syntheses have been finished, a synthesis of F_c should be evaluated, containing the same terms as the F_o syntheses, each F_c having been calculated for the peak positions given by the last F_o synthesis. Owing to the diffraction effects, this F_c synthesis will have peaks slightly displaced from the original positions. Since the peaks of the F_o synthesis are shown on a map with diffraction effects, their real positions, without diffraction, are obtained by reversing the displacements obtained in the F_c synthesis. This mode of correcting for finite series is justifiable if

(i) the shifts are small compared with the periodicity of the diffraction waves, and

(ii) there is good agreement between F_o and F_c for all planes.

The first condition will be fulfilled almost invariably, though a further second-order correction can be made by computing an F_c synthesis based on the corrected positions, and by then adjusting these positions by the difference between the peaks of this second F_c synthesis and the F_o synthesis. The experimental errors will rarely be so small as to make this additional refinement worthwhile.

If condition (ii) is not fulfilled, the amplitudes of the diffraction waves will be wrong, and consequently the corrections will be in error. As an example of a case where care is needed in applying the method we instance geranylamine hydrochloride (Jeffrey, 1945). The experimental syntheses seem to show a considerable variation in the thermal motion between the polar and non-polar ends of the molecule. If an F_c synthesis (based on identical spherically symmetrical carbon atoms) shows that this broadening of the contours is not due to some geometrical fluke, and thus that the thermal motion differences are real, the diffraction effects will vary considerably down the chain and corrections based on identical atoms will not be accurate. For the method to be applied with confidence in such complex cases it may be necessary to vary the temperature factor for each atom, and possibly even to allow for asymmetry of thermal motion.

As in equation (A 3·2), we can extend these correction syntheses by constructing a synthesis whose first terms are F_o, and whose last terms are F_c, where no observations were possible. This gives a density

$$\rho(x,y,z)=\sum_3 F_o\cos(\theta-\alpha)+\sum_{\pm h}^{\pm\infty}\sum_{\pm k}^{\pm\infty}\sum_{\pm l}^{\pm\infty}F_c\cos(\theta-\alpha) \tag{10·1}$$

$$=\sum_3 F_o\cos(\theta-\alpha)+\rho_{\text{calc.}}(x,y,z)-\sum_3 F_c\cos(\theta-\alpha)$$

$$=\rho_{\text{calc.}}(x,y,z)+\sum_3 (F_o-F_c)\cos(\theta-\alpha). \tag{10·2}$$

The peaks of this synthesis will lie between those found by the method above and the peaks of the F_o synthesis, since the correction displacement is inversely proportional to the central curvature and this is greater for the infinite series of (10·1) than for either of the finite series.

Bonding electrons and hydrogen atoms have their biggest proportionate effect on structure factors at small reciprocal radii, so that they do not affect the extrapolation of the F values very much. Consequently, if the assessment of temperature factors has been carefully made, we may regard (10·2) as giving the most faithful general representation of the electron density we can obtain, though it does not give quite the most accurate atomic positions. As Fankuchen (1947) has pointed out, owing to its lack of diffraction effects it will give a good opportunity of detecting hydrogen atoms and of studying electron distributions in regions of low density, though in all such studies the effects of experimental errors must be very carefully considered.

11. The effect of random errors

If the F values were accurate, the electron density represented by an incomplete series would be

$$\rho(x,y,z)=\frac{1}{V}\sum_3 |F|\cos(\theta-\alpha). \tag{11·1}$$

The first differentials are of the type

$$\frac{\partial\rho}{\partial x}=-\frac{1}{V}\sum_3 2\pi\frac{h}{a}|F|\sin(\theta-\alpha), \tag{11·2}$$

and the second differentials are

$$\frac{\partial^2\rho}{\partial x\,\partial y}=-\frac{1}{V}\sum_3 4\pi^2\frac{hk}{ab}|F|\cos(\theta-\alpha). \tag{11·3}$$

We will use Booth's (1946*b*) convenient notation

$$\frac{\partial\rho}{\partial x}=A_h \quad\text{and}\quad \frac{\partial^2\rho}{\partial x\,\partial y}=A_{hk},\ \text{etc.}$$

If the experimental $|F|$ values are in error by amounts ΔF, then we have for the errors in the incomplete series

$$\Delta(\rho)=\frac{1}{V}\sum_3 \Delta F\cos(\theta-\alpha), \tag{11·4}$$

$$\Delta(A_h)=-\frac{1}{V}\sum_3 2\pi\frac{h}{a}\Delta F\sin(\theta-\alpha), \tag{11·5}$$

$$\Delta(A_{hk})=-\frac{1}{V}\sum_3 4\pi^2\frac{hk}{ab}\Delta F\cos(\theta-\alpha). \tag{11·6}$$

Consider the evaluation of $\Delta(A_h)$ as typical. The individual ΔF's are in principle unknowable, so that $\Delta(A_h)$ cannot be evaluated exactly. We have therefore to resort to statistical methods, for which we require a probability distribution of ΔF, and we must be satisfied by estimating from this the probability distribution of $\Delta(A_h)$.

It is apparent that a probability distribution of ΔF estimated on a basis of the difference between two equally reliable sets of experimental $|F|$ values will underestimate the true ΔF, since no allowance will be made for systematic errors common to both sets of observations; while the true ΔF will be overestimated on a basis of the difference between a set of experimental F values and a set of F values calculated for a postulated structure ($F_{\text{calc.}}(hkl)=\sum_1^M f_r\cos\theta_r$; see also § 5). We shall adopt the latter procedure for the following reasons:

(i) In general, there will not be two sets of experimental values for comparison.

(ii) We shall be on the safe side in estimating the experimental errors from the difference between observed and calculated values.

(iii) The corrections required for finite series are estimated by methods involving calculated F's for a postulated structure. These corrections may themselves be in error, for which some estimate must be made; usually in fact the total random error will be correctly estimated by taking ΔF as the difference between observed and calculated values. The matter is further discussed in § 14.

The statistical analysis which follows is based on well-known theorems given by Whittaker & Robinson (1944). The relevant theorems are:

(A) When (in a large class of cases) a deviation ϵ is constituted by the summation of a very large number of independent deviations $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ the probability that ϵ lies between x and $x+dx$ is

$$\frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} e^{-x^2/2\sigma^2} dx, \tag{3·1}$$

where σ is independent of x and is called the standard deviation. This is the normal or Gaussian probability distribution.

(B) The standard deviation of a sum of deviations $y=(\lambda_1\epsilon_1+\lambda_2\epsilon_2+\ldots+\lambda_n\epsilon_n)$ is

$$\sigma(y)=\left\{\sum_1^n \lambda_r^2\sigma_r^2\right\}^{\frac{1}{2}},$$

where σ_r is the standard deviation of ϵ_r.

Assuming that the distribution laws of the ΔF's are such that, in conjunction with the large number of ΔF's, the qualifying phrase of theorem A 'in a large class of cases' is not violated, joint application of these theorems shows that the probability distribution of $\Delta(A_h)$, for example, is normal and gives the standard deviation of $\Delta(A_h)$ in terms of the standard deviations of the ΔF's. We shall show in the next section that it is not permissible to take a constant standard deviation for ΔF. The only estimate available is

$$\sigma(\Delta F_{hkl})=\Delta F=F_o-F_c.$$

On the supposition of a large number of ΔF's this will lead to reliable estimates of $\sigma(A_h)$, etc.

Hence we have

$$\sigma(A_h)=\frac{2\pi}{aV}\left\{\sum_3 \lambda_{hkl}^2 h^2 \Delta F_{hkl}^2\right\}^{\frac{1}{2}}, \tag{11·7}$$

where $\sigma(A_h)$ is the standard deviation of $\Delta(A_h)$ and

$$\lambda_{hkl}=[\sin\theta_1+\sin(-\theta_1)+\ldots]. \tag{11·8}$$

θ_1 involves the co-ordinates at which $\sigma(A_h)$ is being evaluated; the number of terms in the expression for λ_{hkl} depends on the space group and is in any case a multiple of 2, since invariably $\Delta F_{hkl}\equiv\Delta F_{\bar{h}\bar{k}\bar{l}}$, while additionally we might have, e.g., $\Delta F_{hkl}\equiv\Delta F_{\bar{h}k\bar{l}}\equiv\Delta F_{h\bar{k}l}$. Let the number of planes related in this way be m, which may be different for sets of general (hkl) planes than for sets of special, e.g. $h0l$, planes.

Unless (11·7) is being evaluated at or near a special point such as (0, 0, 0) it is reasonable to extend the statistical argument by taking the R.M.S. value of (11·8) for any point (x, y, z).

As the sine terms are linked in pairs this gives

$$\lambda_{hkl}=\frac{1}{\sqrt{2}}\left\{2^2\frac{m}{2}\right\}^{\frac{1}{2}},$$

the R.M.S. value of $\sin\theta$ being $1/\sqrt{2}$.

We now have the general result

$$\sigma(A_h)=\frac{2\pi}{aV}\left\{\sum_3 h^2\Delta F^2\right\}^{\frac{1}{2}}, \tag{11·9}$$

the m dropping out as $\sum\limits_3$ indicates a summation over each observed plane.

Similarly, the R.M.S. values of the other errors are

$$\sigma(\rho)=\frac{1}{V}\left\{\sum_3 \Delta F^2\right\}^{\frac{1}{2}}, \tag{11·10}$$

$$\sigma(A_{hk})=\frac{4\pi^2}{abV}\left\{\sum_3 h^2k^2\Delta F^2\right\}^{\frac{1}{2}}. \tag{11·11}$$

In deriving these expressions we have made no assumptions regarding ΔF, such as that ΔF is proportional to the plane index. We might have expressed ΔF as a function of the plane index and magnitude of F by analysing the ΔF's for a set of observed and calculated structure factors; but the above method achieves the final result without the trouble of finding ΔF in a functional form.

With regard to those planes which were within the limiting reciprocal sphere of observation, but for which no F's were observed, there are two possibilities depending upon the method used for correcting for termination of series. If the semi-analytical method of §§ 6–8 or the f syntheses of § 9, both of which depend on a 'sharp cut-off', are used, then put $\Delta F=F_c$ in the expression above. If the F_c method of § 10 is used, then omit ΔF for all unobserved planes.

For orthorhombic or monoclinic cells with β nearly 90°, the standard deviations of a peak position are

$$\sigma(x)=\sigma(A_h)/A_{hh};\quad \sigma(y)=\sigma(A_k)/A_{kk};\quad \sigma(z)=\sigma(A_l)/A_{ll}. \tag{11·12}$$

For general cells Booth (1946*b*) has derived equations for the precise location of a peak, when the series A_h, A_k, A_l, A_{hk}, etc., are evaluated at some nearby point. If ϵ_x, ϵ_y, ϵ_z are the peak co-ordinates relative to the nearby point, then

$$\left.\begin{aligned} A_{hh}\epsilon_x+A_{hk}\epsilon_y+A_{hl}\epsilon_z+A_h&=0,\\ A_{hk}\epsilon_x+A_{kk}\epsilon_y+A_{kl}\epsilon_z+A_k&=0,\\ A_{hl}\epsilon_x+A_{kl}\epsilon_y+A_{ll}\epsilon_z+A_l&=0.\end{aligned}\right\} \tag{11·13}$$

These equations will give the R.M.S. peak errors if we insert $\sigma(A_h)$, $\sigma(A_k)$, $\sigma(A_l)$ from (11·9) in place of A_h, A_k, A_l. In solving these equations, since $\sigma(A_h)$, etc., are standard deviations, $\sigma(A_h)$ plus $\sigma(A_k)$ is $\{\sigma(A_h)^2+\sigma(A_k)^2\}^{\frac{1}{2}}$.

The solution of (11·12) for spherically symmetrical peaks in a monoclinic cell reduces to

$$\left.\begin{aligned}\sigma(x)&=\frac{\{\sigma(A_h)^2-\cos^2\beta\,\sigma(A_l)^2\}^{\frac{1}{2}}}{A_{hh}\sin^2\beta},\\ \sigma(y)&=\sigma(A_k)/A_{hh},\\ \sigma(z)&=\frac{\{\sigma(A_l)^2-\cos^2\beta\,\sigma(A_h)^2\}^{\frac{1}{2}}}{A_{hh}\sin^2\beta},\end{aligned}\right\}\qquad(11\cdot14)$$

where $\sigma(x)$, $\sigma(y)$, $\sigma(z)$ are the R.M.S. peak errors.

We may expect, in general, that

$$\sigma(x)\simeq\sigma(y)\simeq\sigma(z).\qquad(11\cdot15)$$

If so, the R.M.S. radial error of position will be

$$\sigma(r)=(\sqrt{3})\,\sigma(x).\qquad(11\cdot16)$$

The standard deviation in the distance between two symmetrically independent atoms with the same A_{hh} will be

$$\sigma(\text{dist})=(\sqrt{2})\,\sigma(x),\qquad(11\cdot17)$$

since errors in directions at right angles to the interatomic line do not change its length.

It can be shown that the error in the difference between the slopes at two points is

$$\frac{2\pi}{aV}\left\{\sum_3 4h^2\Delta F^2\sin^2\frac{\delta\theta}{2}\right\}^{\frac{1}{2}},\qquad(11\cdot18)$$

where $\delta\theta=\theta_1-\theta_2$.

If the two points are close together, this difference in slope will be less than indicated by (11·17), which might reduce the error in the distance between two peaks, but this will only be if all the terms $\delta\theta$ in (11·18) are small, say less than $\frac{1}{4}\pi$. With Cu $K\alpha$ radiation this requires the peaks to be closer than 0·1 A.; consequently (11·18) effects no reduction in the error between two peaks and (11·17) stands.

The formulae above have assumed no phase-angle error, and are for use with centro-symmetric structures, but they will underestimate the errors if any appreciable number of planes have the wrong phase and they must be applied cautiously to 'poor' structures.

For non-centrosymmetric structures the phase-angle errors must be taken into account. Corresponding to (11·4) we obtain

$$\Delta(\rho)=\frac{1}{V}\sum_3\{F_t\cos(\theta-\alpha_t)-F_o\cos(\theta-\alpha_c)\},\quad(11\cdot19)$$

where the suffix t denotes true values.

Let $\alpha_t-\alpha_c=\Delta\alpha$ and $F_t-F_o=\Delta F'$. If $\Delta\alpha$ is small, so that $\cos\Delta\alpha=1$ and $\sin\Delta\alpha=\Delta\alpha$, then

$$\Delta(\rho)=\frac{1}{V}\sum_3\{\Delta F'\cos(\theta-\alpha_t)+\Delta\alpha\,.\,F_o\sin(\theta-\alpha_t)\}.\qquad(11\cdot20)$$

When we include the contribution of the error in the finite series correction to the total random error, we find additional contributions of the type $(F_t-F_c)\cos(\theta-\alpha_t)$ and negative ones $\Delta\alpha F_c\sin(\theta-\alpha_t)$, where this $\Delta\alpha$ is identical with that in (11·20). Hence, following the same arguments as above, we obtain, corresponding to (11·10),

$$\sigma(\rho)=\frac{1}{V}\left\{\sum_3[\Delta F^2+\Delta\alpha^2\,.\,\Delta F^2]\right\}^{\frac{1}{2}}.\qquad(11\cdot21)$$

If we neglect the small contributions from the terms $\Delta\alpha^2\,.\,\Delta F^2$, we see that the errors in non-centrosymmetric structures are given by the same formulae as for centrosymmetric structures. It was incorrectly stated previously (Cox & Cruickshank, 1948) that a factor $\sqrt{2}$ was involved.

12. The effect of random errors in dibenzyl

Taking the two-figure values of F_o and F_c given by Jeffrey (1947), applying the formula of the last section, and omitting ΔF for all planes for which there was no F_o, we obtain

$\sigma(\rho)=0{\cdot}125$ e.A.$^{-3}$,

$\sigma(A_h)=0{\cdot}262$; $\sigma(A_k)=0{\cdot}246$; $\sigma(A_l)=0{\cdot}245$ e.A.$^{-4}$,

$\sigma(A_{kk})=0{\cdot}992$ e.A.$^{-5}$.

Taking the mean value of $\partial^2\rho/\partial r^2=-34{\cdot}1$ obtained in § A1, and substituting the values in equations (11·14), we get

$$\sigma(x)=0{\cdot}0087\text{ A.};\quad\sigma(y)=0{\cdot}0072\text{ A.};\quad\sigma(z)=0{\cdot}0079\text{ A.}\qquad(12\cdot1)$$

These have an R.M.S. value $\sigma=0{\cdot}0080$ A.

Booth (1946*a*) has suggested that the experimental errors could be estimated by assuming the same most probable experimental error for each F_o, obtained by using two sets of experimental data to estimate ΔF. The standard deviation for all planes observed by both Robertson and Jeffrey is 1·12 (omitting the 200 plane); however, this must be divided by $\sqrt{2}$ since both sets of observations are subject to error. Substituting this in (11·9), again omitting ΔF for all unobserved planes, and solving as before, we get the positional standard deviations as

$$\sigma(x)=0{\cdot}0061\text{ A.};\quad\sigma(y)=0{\cdot}0058\text{ A.};\quad\sigma(z)=0{\cdot}0060\text{ A.}\qquad(12\cdot2)$$

These values are somewhat below those of (12·1), though since (12·1) is calculated from the differences between observed and calculated F's we should have expected them to be considerably smaller. The reason would seem to be that ΔF depends on the reciprocal radius of the plane, and that the common planes of Robertson and Jeffrey contain a larger proportion of small index planes with bigger errors than the complete set of planes used in the three-dimensional synthesis.

Table 1 gives the R.M.S. values of ΔF for different groups. This table shows clearly the variation of ΔF with plane index. Investigation did not reveal any simple relation between f and ΔF. Within each subgroup of the table the distribution of the errors is roughly Gaussian.

It is interesting to compare the errors (12·1) of the three-dimensional synthesis with those of the two-dimensional syntheses. Table 2 gives the standard deviations of the positional errors for the two-dimensional syntheses and has been calculated by the methods of the last section for the difference between Jeffrey's F_o and F_c, taking $\partial^2\rho/\partial x^2 = -37{\cdot}0$. Comparison with (12·1) shows the superiority of the three-dimensional synthesis, especially when the possible lack of resolution in two dimensions is remembered.

Table 1. R.M.S. *values of* ΔF

	(*a*)	(*b*)	(*c*)	(*d*)
$f>2$	1·38	2·91	2·70	2·82
$2\geqslant f>1$	0·85	1·36	1·05	1·13
$1\geqslant f>0{\cdot}5$	0·95	1·14	1·01	1·02
$0{\cdot}5\geqslant f$	—	—	0·80	0·80
All f's	1·15	2·16	1·23	1·45

(*a*) Difference between Robertson's and Jeffrey's experimental results omitting $\Delta F_{200} = 9$.
(*b*) Difference between Jeffrey's observed and calculated results for planes common to (*a*), omitting $\Delta F_{20\bar{2}} = 19$.
(*c*) Difference between Jeffrey's observed and calculated results for planes not included in (*b*).
(*d*) Aggregate of (*b*) and (*c*).

Table 2. *Standard deviation of positional error for two-dimensional syntheses*

	$\sigma(x)$ (A.)	$\sigma(y)$ (A.)	$\sigma(z)$ (A.)
Projection on 100	—	0·0186	0·0170
Projection on 010	0·0168	—	0·0208
Projection on 001	0·0137	0·0120	—

R.M.S. value $\sigma = 0{\cdot}0167$ A.

Table 3 gives the errors taking ΔF as the difference between Robertson's and Jeffrey's F_o's, divided by $\sqrt{2}$, thus giving the effect of the random errors of intensity estimation. As would be expected from columns (*a*) and (*b*) of Table 1, these errors are appreciably smaller than those of Table 2.

We shall defer the discussion of errors in particular bond lengths until § 15.

Table 3. *Effect of random errors of intensity estimation*

	$\sigma(x)$ (A.)	$\sigma(y)$ (A.)	$\sigma(z)$ (A.)
Projection on 100	—	0·0068	0·0068
Projection on 010	0·0062	—	0·0044
Projection on 001	0·0062	0·0072	—

R.M.S. value $\sigma = 0{\cdot}0063$ A.

The results obtained in Tables 2 and 3 appear to be at variance with some of the conclusions reached by Robertson & White (1947) in their investigation of the two-dimensional errors in a hypothetical hydrocarbon structure. They concluded that the convergence of the Fourier series is of much greater importance than extreme accuracy of intensity measurement. The hypothetical structure used had 10 atoms in well-resolved positions. Employing a two-dimensional projection, syntheses were evaluated for four cases:

I. Accurate F values to $2\sin\theta = 2{\cdot}0$.
II. Accurate F values to $2\sin\theta = 1{\cdot}5$.
III. $2\sin\theta = 2{\cdot}0$ and random errors in the F's.
IV. $2\sin\theta = 2{\cdot}0$ and systematic errors in the F's.

The average and maximum errors in the atomic positions were those shown in Table 4. We have added the R.M.S. errors to put these results on a comparable basis with those of Tables 2 and 3.

Table 4. *Errors in atomic positions from two-dimensional syntheses*

	I	II	III	IV
Average error (A.)	0·014	0·021	0·018	0·021
Max. error (A.)	0·020	0·034	0·029	0·030
R.M.S. error (A.)	0·015	0·0235	0·0195	0·021

Statistically the component of III due to random errors only is $=(0{\cdot}0195^2-0{\cdot}015^2)^{\frac{1}{2}}=0{\cdot}0125$ A. Even by comparison with II, this is a substantial source of error. The matter can be further checked by taking the difference in the positions of the peaks of III and I as the measure of the random errors. We can also get the effect of the systematic errors alone by taking the difference between the peaks of IV and I. We find

	Random errors Difference III–I	Systematic errors Difference IV–I
Average error (A.)	0·025	0·0175
Max. error (A.)	0·039	0·040
R.M.S. error (A.)	0·027	0·0205

The result for difference III–I is very surprising, as it suggests that the effect of the random errors alone is greater than when they are combined with the termination-of-series errors. Such a result is absurd and the explanation would seem to be that 10 atoms are too small a sample to provide statistically reliable results. By an unlikely chance there has been a tendency for the random errors to be partially corrected by the termination-of-series errors.

We can also calculate the R.M.S. random errors by the formulae of the last section. After setting

$$\partial^2\rho/\partial r^2 = -62{\cdot}0,$$

obtained approximately from Fig. 2 of Robertson & White's paper, we find

$$\sigma(x) = 0{\cdot}0144\,\text{A.},\quad \sigma(z) = 0{\cdot}0145\,\text{A.}$$

These are intermediate between the results of Tables 2 and 3.

The standard deviation of the two-dimensional radial error is $\sigma(r) = [\sigma(x)^2 + \sigma(z)^2]^{\frac{1}{2}} = 0{\cdot}0204$ A., which lies between the two estimates of the random errors in the hypothetical structure already obtained above.

On the evidence from dibenzyl and from this hypothetical structure we might hazard the generalization that for a wide range of organic crystals with careful intensity estimations the radial standard deviation of carbon peak positions due to random errors in two-dimensional projections is approximately 0·020 A.

13. Computational errors

It may seem superfluous to emphasize the necessity of not making computational mistakes, but if a high degree of accuracy is claimed for a structure it is quite essential that there are no mistakes. Quite a small one may easily shift the peaks 0·02 A., a distance which is sometimes a major reason for the structure analysis in question. In such cases proper caution would indicate an entirely separate check, say recalculation of the peak positions by differential syntheses (Booth, 1946*b*).

Computational errors due to rounding off may occur at various stages of the calculations, though they can be virtually eliminated by the use of suitable techniques (e.g. Booth, 1946*b*; Cox, Gross & Jeffrey, 1947*a*, *b*). The total effect of the many small rounding-off errors will be similar to the effect of the many small experimental errors and will produce an error in the density or density slope with normal probability distribution.

The standard deviation of the error in rounding off to the nearest whole number is

$$\sigma=\left\{\int_{-\frac{1}{2}}^{+\frac{1}{2}} x^2\,dx\right\}^{\frac{1}{2}}=\frac{1}{2\sqrt{3}}. \qquad (13{\cdot}1)$$

Obviously the error in rounding off to the first decimal place is one-tenth of this.

We shall now investigate the differences between using structure factors rounded off to the nearest whole number and those rounded off to the nearest decimal place.

The rounding-off error ΔF is independent of the order or magnitude of the plane, and hence from (11·9) and (11·10) we have

$$\sigma(\rho)=\frac{1}{V}\left\{\sum_{3}\sigma^2\right\}^{\frac{1}{2}}, \qquad (13{\cdot}2)$$

$$\sigma(A_h)=\frac{2\pi}{aV}\left\{\sum_{3}h^2\sigma^2\right\}^{\frac{1}{2}}. \qquad (13{\cdot}3)$$

Solving these for dibenzyl we get

$$\sigma(\rho)=0{\cdot}082\,\sigma\ \text{e.A.}^{-3}, \qquad (13{\cdot}4)$$

$$\sigma(A_h)=0{\cdot}237\,\sigma\ \text{e.A.}^{-4}. \qquad (13{\cdot}5)$$

For rounding off to the nearest whole number we get

$$\sigma(\rho)=0{\cdot}024\ \text{e.A.}^{-3}, \qquad (13{\cdot}6)$$

$$\sigma(x)=0{\cdot}0022\ \text{A.} \qquad (13{\cdot}7)$$

For the random errors (taken as the difference between the two-figure F_o and F_c) we had

$$\sigma(\rho)=0{\cdot}125\ \text{e.A.}^{-3};\quad \sigma=0{\cdot}0080\ \text{A.} \qquad (12{\cdot}1)$$

If there had been no rounding off, the random error $\sigma(\rho)$ would have been

$$\sigma(\rho)=[0{\cdot}125^2-2\times0{\cdot}024^2]^{\frac{1}{2}}=0{\cdot}120\ \text{e.A.}^{-3},$$

and in any direction

$$\sigma=[0{\cdot}0080^2-2\times0{\cdot}0022^2]^{\frac{1}{2}}=0{\cdot}0074\ \text{A.} \qquad (13{\cdot}8)$$

The effect of rounding off the F's to the nearest whole number is apparently small, and the error would be quite negligible if the F's were rounded off to the first decimal place.

The calculation of the effect of rounding-off errors in using Beevers & Lipson (1936) strips to find peak positions is rather involved and as it does not lead to any simple general formulae we shall not give the details of the analysis. It does, however, appear from this analysis that the errors in the directions of the two axes of the plane are different because of summing over one index before the other.

In using Beevers & Lipson strips for dibenzyl, Jeffrey (1947) obtained x and z peak co-ordinates from syntheses of plane sections of the cell (his totals being summed over l before being summed over h), and y co-ordinates from line syntheses. The analysis of the errors in using the strips gave the following results:

$$\sigma(x)=0{\cdot}0041\ \text{A.},\quad \sigma(y)=0{\cdot}0020\ \text{A.},\quad \sigma(z)=0{\cdot}0057\ \text{A.} \qquad (13{\cdot}9)$$

14. Random and systematic errors

Our mode of analysis has treated the errors due to finite series as systematic and correctable; and the experimental and computational errors, and the errors in the method of correction for finite series, as random. The former have been estimated, using assumed scattering-factor curves, and the latter on a basis of $\Delta F=F_o-F_c$. We must now examine the justification of this procedure more closely.

If there are no systematic errors in either F_o or F_c, the correction for finite series will be exact, except for the errors of the conventional approximation mentioned in § 5. So far as bond lengths are concerned, but not necessarily for other details, if the series has a very large number of terms, this correction will be almost exactly right, and the random errors will be overestimated by taking $\Delta F=F_o-F_c$. Speaking in round terms, if the corrections for finite series are smaller than the standard deviations of the random errors, then it is likely that the random errors are being overestimated by taking $\Delta F=F_o-F_c$.

Without instancing the causes, there may be systematic errors in F_o depending both on the reciprocal radius and the direction in reciprocal space. The effect of these systematic errors will be equivalent to the use of a set of distorted scattering factors; consequently the largest absolute density errors will be in regions around the peaks, and these errors may be larger than would be estimated as random errors. If, as might happen when the values of f_c are obtained directly from the experimental data as for dibenzyl in § A 2, f_c is subject to the same systematic errors as f_o, then the experimental errors will probably be underestimated in most regions of the unit cell. Errors of scale are in a peculiar category; they do not affect the internal details of a structure, as Booth (1946*a*) has pointed out, but in comparison with other structures their effect on the absolute electron density may be much larger than estimated.

Errors in f_c, relative to f_o not to the true f, have a twofold effect: they increase the errors $\Delta F=F_o-F_c$, and they cause wrong corrections to be made for finite series. To some extent the one effect allows for the other (in fact taking the R.M.S. over the whole unit cell they cancel out) but there may be systematic build-ups in some regions. We must distinguish between two types of error in the f_c curves, wrong values at and near the limiting reciprocal radius, and wrong values at other reciprocal radii. Both cause systematic errors in the densities near the centres of peaks which may be underestimated if treated as random errors. The former will cause wrong estimation of both peak position and also of low-density-region corrections, which may be underestimated if they are treated as random errors.

If the scales of F_o and F_c are different, the random errors will be overestimated, since the finite-series correction does not depend on the scale of F_c. Hence before estimating the random errors care should be taken to ensure that F_o and F_c are on the same scale.

One further point remains to be examined. If we use the method of § 10 for correcting finite series, F_c will have to be calculated for peak positions before correction for finite series. To save labour it is undesirable to have to recalculate F_c from the corrected positions in order to substitute in $\Delta F = F_o - F_c$ to obtain the random errors. Does this procedure considerably overestimate the random errors?

From a result obtained by Booth (1945*b*) it can be shown that if the peaks in an electron-density map can be represented by the expressions (A 1·1), then

$$\sum \Delta F^2 = V \sum_{1}^{M} N^2 \frac{p^{\frac{5}{2}}}{(2\pi)^{\frac{3}{2}}} \delta r^2, \qquad (14\cdot1)$$

where ΔF is the change in structure factor due to shifting the peaks an R.M.S. radial distance δr.

We have confirmed that the peaks in dibenzyl can be approximately represented by (A 1·1), and we shall show in the next section that the R.M.S. radial shift of the peaks on correction for finite series is

$$\sqrt{3} \times 0\cdot0142 = 0\cdot0246\,\text{A}.$$

Substituting this for δr in (14·1), and setting $V = 527\,\text{A.}^3$, $N = 5\cdot44$ and $p = 3\cdot35$, as obtained in § A1, we find $\Sigma \Delta F^2 = 344$.

For $\Delta F = F_o - F_c$, as in § 12, we find $\Sigma \Delta F^2 = 4316$. Thus there would be no substantial reduction for dibenzyl in the values of $\Delta F = F_o - F_c$ by recalculating the F_c's for the positions obtained after the correction for finite series, and by examining the values of the ΔF's in (14·1) we can confirm that there will be no substantial reduction in summations like $\Sigma h^2 \Delta F^2$ which are used to calculate standard deviations of peak positions due to random errors. The reason is that the ΔF's due to a slight change of peak position are systematic not random, and their magnitudes are correspondingly small.

We shall assume that this result is true generally.

We may summarize these remarks as follows:

When correction has been made for finite series (i) the errors in the density near the centres of peaks are likely to be larger than would be estimated as random errors, especially if in making corrections for finite series bad f_c values are used; (ii) the estimate of the errors in peak positions and regions of low electron density will be substantially correct if treated as random with values given by the formulae of §§ 11 and 13, except that they may be underestimated if the f_c values at and near the limiting reciprocal radius used in making corrections for finite series are in poor agreement with the f_o, or alternatively they may be overestimated if there is good agreement between the f_o and f_c values and if there is a very large number of terms in the series.

It is hoped that the rather general remarks above indicate some of the limitations of the present methods and some of the difficulties that must be overcome if better estimates are required, but also that the present methods, if applied with due caution, are satisfactory for many purposes.

15. Detailed analysis of the accuracy of the dibenzyl structure

As there is no sharp cut-off in Jeffrey's F_o's in dibenzyl, the effect of finite series can be estimated only by the use of F_c syntheses, as described in § 10. The peak positions were found, using differential syntheses (Booth, 1946*b*), the whole set of calculations being performed by punched-card methods which are being described elsewhere. This enabled every stage of the calculation to be made at three-figure accuracy or better. Consequently, rounding-off errors should be negligible. The opportunity was also taken to recalculate the peaks of the experimental synthesis starting from three-figure structure factors. The same planes were used in both the F_o and F_c syntheses. No direct comparison is possible between these results and those obtained by Jeffrey, to estimate the rounding-off errors in the latter, as a small number of mistakes in the original list of structure factors were found and were corrected in these calculations.

The new peak positions and central electron densities found for the experimental synthesis are shown in Table 5, and the corresponding bond lengths in Table 6.

Table 5. *Peak positions and central electron densities from experimental syntheses of dibenzyl*

Co-ordinates as fractions of the cell sides

Atom	x	y	z	$\rho(0)$ e.A.$^{-3}$
C_1	0·0266	0·0959	− 0·0267	5·75
C_2	0·1458	0·1534	0·1200	6·73
C_3	0·1654	0·3348	0·2332	5·94
C_4	0·2785	0·3911	0·3658	5·48
C_5	0·3703	0·2584	0·3860	5·73
C_6	0·3513	0·0787	0·2764	5·99
C_7	0·2412	0·0256	0·1468	6·48

Table 6. *Bond lengths from co-ordinates of Table* 5

	(A.)		(A.)
C_1'–C_1	1·501	C_4–C_5	1·376
C_1–C_2	1·486	C_5–C_6	1·342
C_2–C_3	1·366	C_6–C_7	1·360
C_3–C_4	1·398	C_7–C_2	1·385
Angle $C_1'C_1C_2$	114° 23′	Mean side of ring	1·371

After correcting for finite series by F_c syntheses the peak positions and central densities are those given in Table 7. The corrected central densities were obtained by subtracting the difference between the density given by the F_c synthesis and the mean calculated density (6·25 e.A.$^{-3}$).

Table 7. *Peak positions and central electron densities after correction for finite series by F_c syntheses*

Atom	x	y	z	$\rho(0)$ e.A.$^{-3}$
C_1	0·0249	0·0970	−0·0298	6·04
C_2	0·1472	0·1542	0·1213	6·24
C_3	0·1661	0·3378	0·2337	6·10
C_4	0·2782	0·3933	0·3656	5·60
C_5	0·3706	0·2631	0·3859	5·89
C_6	0·3531	0·0790	0·2763	5·97
C_7	0·2399	0·0229	0·1458	6·23

The bond lengths after correction are given in Table 8.

Table 8. *Bond lengths after correction*

	(A.)		(A.)
C_1'–C_1	1·510	C_4–C_5	1·376
C_1–C_2	1·523	C_5–C_6	1·366
C_2–C_3	1·374	C_6–C_7	1·394
C_3–C_4	1·385	C_7–C_2	1·374
Angle $C_1'C_1C_2$	112° 4′	Mean side of ring	1·378

The general effect of the corrections for finite series follows the results suggested by § 7, viz. that the size of the ring and the central bonds are increased. However, there are as well various irregular corrections, which must be due to the irregular manner of 'cut-off' of observations; these are certainly not negligible. The R.M.S. peak corrections in the various directions are

$$x,\ 0{\cdot}0160\,\text{A.};\quad y,\ 0{\cdot}0104\,\text{A.};\quad z,\ 0{\cdot}0154\,\text{A.};$$

which have an R.M.S. value 0·0142 A.

As the scattering factors used in the F_c synthesis were obtained from the experimental results (see § A 2), we shall be justified in regarding the co-ordinates and bond lengths of Tables 7 and 8 as final, and as having random errors given by (13·8). We shall now investigate the significance of some aspects of these results.

The chief interests of Jeffrey's results were the shortened central bonds C_1'–C_1 and C_1–C_2. The correction for finite series has increased these lengths, but they remain shorter than the C–C single bond length of 1·5445 A. in diamond (Lonsdale, 1947) which has an error negligible for our purposes. We can test for the significance of the differences.

The difference, δ, between C_1'–C_1 and the C–C diamond single bond is $1{\cdot}5445-1{\cdot}5100=0{\cdot}0345$ A. By (13·8) the standard deviation of position in any direction is 0·0074 A. Since the errors in the positions of C_1' and C_1 must be the same by symmetry, the standard deviation of the bond length is $2\times0{\cdot}0074=0{\cdot}0148$ A. Hence

$$\delta/\sigma=0{\cdot}0345/0{\cdot}0148=2{\cdot}33.$$

This is in the significance zone defined in § 3; the actual probability P of C_1'–C_1, if a true single bond of 1·5445 A., being observed as 1·510 A. or less, due to experimental error, is 0·00990.

The difference δ between C_1–C_2 and the diamond C–C bond is 0·0215 A. The standard deviation of the length C_1–C_2 is $\surd 2\times0{\cdot}0074=0{\cdot}0105$ A., since C_1 and C_2 are independent. In this case

$$\delta/\sigma=0{\cdot}0215/0{\cdot}0105=2{\cdot}05,$$

and is in the zone of possible significance. The actual probability P of such a difference happening by chance when the bond is really single is 0·0202.

The simple analysis above has treated the central bond lengths as independent, whereas in fact they have a common error at C_1. An interesting point is that the distance from C_2 to the origin (1·9379 A.) is very close to the distance C_2 would be from the origin (1·9438 A.) if C_1'–C_1 and C_1–C_2 were both single bonds with the normal tetrahedral angle $C_1'C_1C_2$ of 109° 28′. This indicates that the crux of the matter is the displacement of C_1 from the position to be expected in a single-bond system. This we can calculate to be 0·0279 A., the situation being as shown in Fig. 8 (see p. 70). (For convenience in calculation we have reduced the 'single distance bond' from C_2 to the origin to 1·9379 A.; this has reduced the apparent displacement slightly.) We have now

$$\delta/\sigma=0{\cdot}0279/\surd 2\times0{\cdot}0074=2{\cdot}66.$$

By § 3 C_1 is significantly displaced from the position to be anticipated in a single-bond system, the actual probability P of such a result or greater being observed in a true single-bond system being 0·00391. From these considerations we confirm Jeffrey's conclusions that the central bonds C_1'–C_1 and C_1–C_2 in dibenzyl are significantly shorter than the C–C single bonds in diamond, 'significantly' here being understood to possess the quantitative meaning defined in § 3.

A direct (though slightly involved) analysis of the errors in the angle $C_1'C_1C_2$ shows that there is a very significant difference between it and the normal tetrahedral angle, with $P=0{\cdot}00024$.

In his paper Jeffrey suggested that the bonds C_3–C_4 and C_6–C_7 might be longer than the other four bonds of the ring. It is apparent that there is nothing in the bond lengths of Table 8 to disprove symmetry of the ring about the line C_2–C_5. We can thus average the bond lengths about this line obtaining

average of C_2–C_3 and C_7–$C_2=1{\cdot}3740$ A. $=l_1$, (say),

average of C_3–C_4 and C_6–$C_7=1{\cdot}3895$ A. $=l_2$,

average of C_4–C_5 and C_5–$C_6=1{\cdot}3710$ A. $=l_3$.

We shall now test successively for the differences of l_1 and l_3 from l_2. Unfortunately, the standard deviation σ included in the ratio δ/σ is not obtained as simply as in the cases above, as the errors in l_1 and l_2, and l_2 and l_3, all of which are means, are not independent. We can show that the error ϵ, R.M. squared for all directions of the component errors, in the difference $l_2 - l_3$ is

$$\epsilon = \tfrac{1}{2}\epsilon_3 + \epsilon_4 \sin(\tfrac{1}{2}\alpha) + \epsilon_5 \cos\tfrac{1}{2}\alpha + \epsilon_6 \sin\tfrac{1}{2}\alpha + \tfrac{1}{2}\epsilon_7,$$

where $\alpha = 120°$. Whence, taking the R.M.S. and putting $\sigma_3 = \sigma_4 = \sigma_5 = \sigma_6 = \sigma_7 = 0{\cdot}0074$ A., we have

$$\sigma = \sqrt{(9/4)} \times 0{\cdot}0074 = 0{\cdot}0111 \text{ A}.$$

Hence $\quad (l_2 - l_3)/\sigma = 0{\cdot}0185/0{\cdot}0111 = 1{\cdot}666,$

which is in the zone of possible significance, $P = 0{\cdot}0478$. Similarly,

$$(l_2 - l_1)/\sigma = 0{\cdot}0155/0{\cdot}0111 = 1{\cdot}40,$$

giving $P = 0{\cdot}0808$.

We conclude that the evidence in favour of the differences in the bond lengths l_2 from l_1 and l_3 is not very definite; further evidence would be required to establish any differences with confidence.

The mean ring size is 1·378 A., for which the standard deviation is $0{\cdot}0074/\sqrt{6} = 0{\cdot}0030$ A. However, this estimate of the standard deviation takes no account of systematic errors of the type referred to in § 14, which, while they might safely be included in the standard error $\sqrt{2} \times 0{\cdot}0074$ A. of a single-bond length, may render 0·0030 A. an underestimate of the error in the mean ring size. Nevertheless, there would seem to be strong evidence that the ring in dibenzyl is smaller than 1·39 A., the usually accepted value for the benzene ring (Brockway & Pauling, 1934).

The values of the central peak densities given in Table 5 from the experimental synthesis show great variation from C_4 with density 5·48 e.A.$^{-3}$ to C_2 with density 6·73 e.A.$^{-3}$. The correction for finite series in Table 7 reduces this difference considerably to C_4 density 5·60 e.A.$^{-3}$ and C_2 density 6·24 e.A.$^{-3}$. C_4 is actually much lower than any other peak, the next lowest is $C_5 = 5{\cdot}89$ e.A.$^{-3}$, its counterpart across the central line C_2–C_5, C_6, having a density 5·97 e.A.$^{-3}$. By (13·8) the standard deviation of the density is 0·120 e.A.$^{-3}$. The discrepancy of C_4–$C_6 = 0{\cdot}37$ e.A.$^{-3}$ is large, and could be interpreted as an instance of the errors at the peak centres being greater than estimated by the formulae of § 11.

We can test the difference between the mean density of C_1, C_2, C_3 and of C_4, C_5, C_6. This difference is 0·370 e.A.$^{-3}$, and the standard deviation of the difference is $0{\cdot}120 \times \sqrt{(2/3)} = 0{\cdot}098$ e.A.$^{-3}$.

$$\delta/\sigma = 0{\cdot}370/0{\cdot}098 = 3{\cdot}77,$$

which by § 3 is very significant, though, in view of the remarks above on the difficulties mentioned in § 14, the argument is less certain. However, there seems some evidence for greater central densities in the ring peaks nearer the centre of the molecule than at the ends. Their interpretation, especially after the discussion of bond lengths above, is problematic though they might be taken as indicating that the thermal vibration of the molecule includes a component of angular motion about the molecular centre of symmetry.

16. Variation of reciprocal radii, etc.

In the preceeding sections very little attention has been given to the variation of errors with different reciprocal radii or to the relative errors in different experimental procedures or modes of structure determination. Without exploring this subject of the 'economics' of structure analysis, much of which would have to be tackled from a rather different view-point from the present, some results follow quickly from the previous data.

We can make some estimation of the variation of errors with limiting reciprocal radii for dibenzyl. The problem is complicated by the irregular cessation of observations in the reciprocal sphere, and consequently the estimated equivalent reciprocal radii of sharp cut-off are different for random and termination-of-series errors. The results below are based on the following assumptions:

Equivalent reciprocal radius of cut-off for the structure (obtained from Jeffrey's observations): $2\sin\theta_0 = 1{\cdot}65$; termination-of-series error for this structure in any direction: 0·0142 A. (§ 15); random error: 0·0080 A. (12·1); random errors of ΔF in the range $2\sin\theta_0 = 1{\cdot}2 - 2{\cdot}0$ (assumed constant): 0·80 (Table 1*d*) multiplied by a factor to allow for difference of equivalent reciprocal radii; $\partial^2\rho/\partial r^2$: the values obtained from integration of the scattering factor given by Table A 2 and extrapolated by (A 3·3).

It is difficult to estimate accurately the termination-of-series effect for different reciprocal radii, but we can set upper and lower limits to its magnitude. We may assume either the disturbing ripple to be constant, in which case the error is inversely proportional to $\partial^2\rho/\partial r^2$ for the peak being shifted, or, by using (6·6) for a given configuration of atoms, $\rho'_B(r)$ is proportional to $f_o s_o^2$ and the error is proportional to $f_o s_o^2/(\partial^2\rho/\partial r^2)$. The true error lies somewhere between these extremes.

Table 9 shows a steady reduction of the termination-of-series error within the range considered, while the random error decreases very slowly to a minimum at $2\sin\theta_0 = 1{\cdot}8$. The increase beyond 1·8 is due to the assumed errors being of the same magnitude as the *F*'s.

Table 9. *Variation of errors in dibenzyl with limiting reciprocal radii*

$2\sin\theta_0$	1·2	1·4	1·6	1·65	1·8	2·0
Random error (A.)	0·0106	0·0089	0·0081	0·0080	0·0079	0·0081
Termination-of-series error estimated as:						
$\propto 1/(\partial^2\rho/\partial r^2)$	0·0321	0·0211	0·0152	0·0142	0·0119	0·0098
$\propto f_0 s_0^2/(\partial^2\rho/\partial r^2)$	0·0476	0·0272	0·0158	0·0142	0·0097	0·0056

We infer, since the f values in dibenzyl are so small at large θ, that in general the random error will be reduced if the reciprocal radius is extended to the limit of observation.

By comparison the errors in any direction in Robertson & White's (1947) two-dimensional hypothetical structure are

	$2 \sin \theta_0 = 1{\cdot}5$ (A.)	$2 \sin \theta_0 = 2{\cdot}0$ (A.)
R.M.S. random error	0·0205	0·0144
R.M.S. termination error	0·0166	0·0105

Both errors are less at the larger θ.

Booth (1947*b*) has recently suggested that structures may be determined by minimizing the quantity $R_2 = \Sigma(F_o^2 - F_c^2)^2$, or by other least-squares procedures. Assuming accurate computation, and that there are no systematic misrepresentations of the electron density by these methods, we can compute the random errors of their results. (They automatically correct for finite series.) The R.M.S. difference between the densities of the observed and calculated structures will be given by the formulae of § 11 with $\Delta F = F_o - F_c$. To get the error between the true and calculated structures an allowance must be made for the experimental errors in F_o. These are unlikely to exceed $(F_o - F_c)/\sqrt{2}$. Hence the total error of the calculated structure will be given by the formulae of § 11 with ΔF lying between $(F_o - F_c)$ and $(1 + \frac{1}{2})^{\frac{1}{2}} (F_o - F_c) = 1{\cdot}225(F_o - F_c)$.

Assuming the agreement between sets of observed and calculated F's is not substantially better than for sets derived from the Fourier process (as seems likely from the results on dibenzyl quoted at the end of § 14) the errors in results obtained from minimization processes are likely to be about the same or slightly greater than those obtained by the normal Fourier methods with correction for finite series.

17. Summary

In attempting to achieve our aim of producing a reliable and workable set of rules for assessing the accuracy of structures, a number of incidental results have been obtained either in reaching the rules or as examples of their application.

The peaks of dibenzyl can be approximately represented by the exponential density distribution suggested by Booth, but the peaks formed from an infinite series cannot be so represented and consequently analyses based on scattering factors which give such peaks are inaccurate (Appendix).

The standard deviations of the random errors in any direction in the dibenzyl peak positions found from two-figure $|F|$ values are 0·0080 A. for three-dimensional syntheses, and 0·0167 A. for two-dimensional syntheses.

The two-dimensional errors are thus approximately twice the three-dimensional errors, and, with a re-examination of the results from Robertson & White's hypothetical structure, suggest that two-dimensional random errors of this magnitude may be typical of a large class of organic compounds (§ 12).

The computational rounding-off errors, using Beevers & Lipson strips to find the dibenzyl peaks either in a three-dimensional section or in a plane projection, have a standard deviation in any direction of approximately 0·005 A. (§ 13).

Application of corrections for finite series in dibenzyl increases both the length of the central bonds and the size of the benzene ring, and reduces the variation in peak central densities. Part of the correction is due to spherical diffraction effects and part to effects caused by the irregular manner of cessation of observed reflexions. The R.M.S. value of the peak correction in any direction is 0·0142 A.

After correction for finite series, the central C–C bonds in dibenzyl are shown to be significantly shorter than the C–C bond in diamond (§ 15).

The random errors in non-centrosymmetric structures are given by the same formulae as for centrosymmetric structures.

Rules for assessing the accuracy of electron-density maps

(1) Correct the errors due to finite series. If there is no sharp limiting reciprocal radius, this must be done by F_c syntheses (§ 10). If there is a definite limiting reciprocal radius, much trouble can be saved by using the semi-analytical method (§§ 6–8), or the f_c synthesis (§ 9).

(2) Estimate the standard deviation of any computational rounding-off errors (§ 13).

(3) Estimate on a basis of $\Delta F = F_o - F_c$ the standard deviation of the experimental and finite-series correction errors (§ 11).

(4) Except in certain respects mentioned in § 14 the errors of the electron-density maps corrected by (1) are the statistical sum of the errors (2) and (3), and have a Gaussian probability distribution.

(5) All comparisons, either between results obtained in different parts of the same electron-density map or between results in different maps, made in trying to establish differences or similarities must involve the use of significance tests (§ 3). In expressing the conclusions of such comparisons, words such as 'significant' should be used with a defined quantitative meaning.

It is a great pleasure to express my thanks to Prof. E. G. Cox for much helpful criticism and for his continuous interest in this work; to Dr G. A. Jeffrey for access to his original data on dibenzyl and for many useful discussions; to Mr L. Gross for his help in devising and carrying out the punched-card methods to compute the dibenzyl differential syntheses; to Mr R. M. Gabriel for several enlightening discussions on Fourier transforms; to Dr B. L. Welch, and to Mr A. P. Robertson for discussions on statistical matters.

APPENDIX

A1. The peak shapes in dibenzyl

Costain (1941) and Booth (1946*a*) have suggested that the electron-density distributions of the peaks obtained in Fourier syntheses may be closely represented by expressions of the type

$$\rho(r)=Ae^{-pr^2}. \qquad (\text{A}1\cdot1)$$

By examination of the peaks for carbon, nitrogen and oxygen in pentaerythritol tetranitrate (Booth & Llewellyn, 1947), chlorine in geranylamine hydrochloride (Jeffrey, 1945) and sulphur in β-isoprene sulphone (Cox & Jeffrey, 1942), Booth obtained a value of $p=4\cdot69$. This value gave a close approximation to the electron densities around the centres of the peaks in these different compounds.

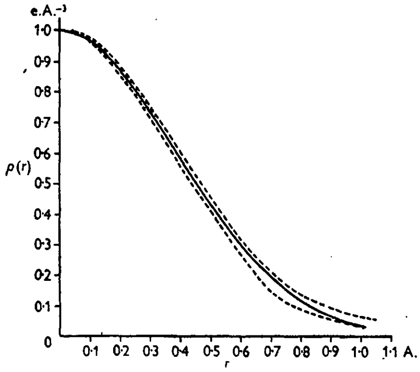

Fig. A1. Dibenzyl. Broken line, envelope of $\rho(r)$ for ring atoms in directions away from C–C bonds. Full line, $e^{-3\cdot35r^2}$.

By normalizing to include the correct number of electrons in the peak (A1·1) becomes

$$\rho(r)=N(p/\pi)^{\frac{3}{2}}\,e^{-pr^2}, \qquad (\text{A}1\cdot2)$$

where N is the atomic number.

At the centre of the peak

$$(\partial^2\rho/\partial r^2)_{r=0}=-2pN(p/\pi)^{\frac{3}{2}}. \qquad (\text{A}1\cdot3)$$

The applicability of this representation has been tested for the carbon atoms in dibenzyl as follows:

(*a*) Fig. A1 shows the variation of $\rho(r)$ with r for the atoms of the ring, the density values having been taken from the Fourier syntheses along directions not lying in the C–C bonds and scaled down to reduce all central densities ($r=0$) to unity. The dotted curves envelop the experimental densities, and the full curve represents Ae^{-pr^2} with $A=1$ and $p=3\cdot35$. $\rho(r)=e^{-3.35r^2}$ gives a good approximation to the reduced density, except for values of $r>0\cdot9$ A. where it is smaller than the experimental density.

(*b*) Table A1 has been calculated from the results of Jeffrey's syntheses, and shows the mean central electron density, central p value, and central curvature for the peaks. The p values were calculated by using a three-point parabola method at the centre; they thus exactly represent the densities at $r=0$, whereas $p=3\cdot35$ gives the best fit for $0<r<0\cdot9$. The curvature was obtained by substitution in (A1·3). As the map density for atom C_1 was rather different from those of the ring atoms, and as it is the only atom of the seven with two hydrogen atoms attached, we have averaged the six ring atoms separately, and used these values in the following discussion.

Table A1. *Details of the dibenzyl peaks*

	Electron density (e.A.$^{-3}$)	Mean p	$\partial^2\rho/\partial r^2$
Mean for all atoms	5·93	2·87	−34·1
Mean without C_1	6·00	2·86	−34·5

If an equation of type (A1·1) is to be valid, it must represent the results of Table A1 as well as those of Fig. A1, and consequently with $p=3\cdot35$ (from Fig. A1) we must put $A=6\cdot00$ in order to give the correct peak height from Table A1. This leads to (A1·2 and A1·3)

$$N=5\cdot44, \quad \text{and} \quad (\partial^2\rho/\partial r^2)_{r=0}=-40\cdot2.$$

We thus see that the use of (A1·1), adjusted to give the observed peak density, gives too high a curvature at the centre and too low a value of the total electron content of the peak. (N should be 7 for CH or 6 for C.) This accords with the observation from Fig. A1 that $p=3\cdot35$ gives low densities at $r>0\cdot9$ A. On the other hand, if we put $A=6\cdot00$ and use the observed mean $p=2\cdot86$ from Table A1 we find $N=6\cdot87$ and we get poor agreement with observed densities in the region of $r=0\cdot6$ A.

A2. Scattering-factor curve for CH in dibenzyl

Each molecule of dibenzyl has 14 carbon atoms and 14 hydrogen atoms, and, except for C_1 and C_1' (each of which has two) and C_2 and C_2' (none), each carbon is bound to one hydrogen. We assume, therefore, that without serious error we can deduce a scattering-factor curve for CH from the average results for the whole dibenzyl molecule. CH scattering factors were obtained by dividing the observed structure factors (expressed to a tenth of a unit) by the calculated geometrical form factors, i.e. by means of the relation

$$f=F_{\text{obs.}}\Big/\sum_1^M \cos\theta_r.$$

The errors in the atomic positions discussed elsewhere in this paper are too small to affect the accuracy of this procedure appreciably. The f values were then plotted against $2\sin\theta$ and a mean f curve was constructed (Fig. A2). The spread of the points about the mean is not shown but may be inferred by recalling (Jeffrey, 1947) that

$$\frac{\Sigma\,|\,|F_{\text{obs.}}|-|F_{\text{calc.}}|\,|}{\Sigma\,|F_{\text{obs.}}|}\sim 0\cdot15.$$

Now if $\quad \rho(r)=N(p/\pi)^{\frac{3}{2}}\,e^{-pr^2}, \qquad$ (A1·2)

then $\quad f(s)=Ne^{-\pi^2s^2/p} \qquad$ (A2·1)

(e.g. Booth, 1946*a*, equation (6·12)).

Hence we can calculate the scattering-factor curve for a density of the type (A1·2) with $N=5{\cdot}44$ and $p=3{\cdot}35$, as obtained in the last section from the synthesis peaks. For comparison this also is shown in Fig. A2. The transforms of both these curves, the former terminated at the limiting reciprocal radius of experiment and the latter extended to infinity, both approximately fit the observed density distribution. Their differences indicate that the real electron density is somewhat different from that in the map peaks, and that the use of a scattering factor of the type (A2·1) is unsatisfactory for dibenzyl.

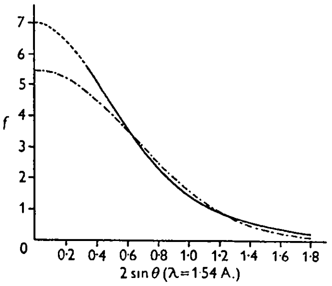

Fig. A2. Full line, CH f curve for dibenzyl. Broken line, $f=5{\cdot}44\, e^{-\pi^2 s^2/p}$, $p=3{\cdot}35$.

A3. The shape of the 'real' density distributions

Van Reijen (1942) has shown how charge distributions without diffraction (termination-of-series) effects may be calculated by smoothly extrapolating the f curve. We shall assume that extrapolation does remove the diffraction effects, although a smooth fit is not a sufficient condition mathematically. Assuming spherical symmetry

$$\rho(r)=\int_0^\infty 4\pi s^2 f(s)\frac{\sin 2\pi sr}{2\pi sr}\,ds. \qquad (A3{\cdot}1)$$

This may be written

$$\rho(r)=\int_0^{s_0} 4\pi s^2 f_o(s)\frac{\sin 2\pi sr}{2\pi sr}\,ds + \int_{s_0}^\infty 4\pi s^2 f_e(s)\frac{\sin 2\pi sr}{2\pi sr}\,ds, \qquad (A3{\cdot}2)$$

where s_0 is the limiting reciprocal radius of experiment, $f_o(s)$ the observed scattering-factor function, and $f(s)$ the extrapolated scattering-factor function.

We use $f_o(s)$ from Fig. A1 to the limit $2\sin\theta=1{\cdot}7$ and then a suitable extrapolated f curve. Such a curve is

$$f(s)=2{\cdot}26\, e^{-\pi^2 s^2/5\cdot 75}. \qquad (A3{\cdot}3)$$

Hence, if we perform the integration of (A3·2), we get a value of $\rho(r)$ for dibenzyl without any termination-of-series effect. This will be a spherically averaged $\rho(r)$, since by taking a single $f(s)$ curve we cannot allow for the directional effects of the bonds. Fig. A3 shows this 'real' $\rho(r)$ and compares it with the density, shown in Fig. A1, obtained from the terminated synthesis. The numerical values of the 'real' $\rho(r)$ are given in Table A2.

Table A2. *The averaged 'real'* CH *'group'*

r (A.)	$\rho(r)$ (e.A.$^{-3}$)	r (A.)	$\rho(r)$ (e.A.$^{-3}$)
0	7·23	0·924	0·467
0·154	6·46	1·078	0·270
0·308	4·65	1·232	0·150
0·462	2·81	1·386	0·079
0·616	1·54	1·540	0·0396
0·770	0·83	1·694	0·0185

$(\partial^2\rho/\partial r^2)_{r=0}=-80{\cdot}2.$

The peak of the 'real' $\rho(r)$ is both denser and sharper than those in the maps, the cause of the difference being the finiteness of the series which generated the map.

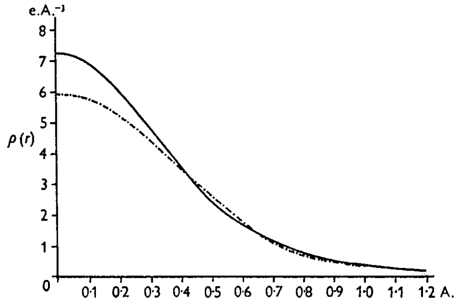

Fig. A3. Full line, $\rho(r)$ by integrating extrapolated f curve. Broken line, $\rho(r)$ from synthesis.

The 'real' peak cannot be represented by an expression of the type (A1·1). This was to be expected, since no atom whose electron density is not of the Gaussian form without heat motion, can attain a Gaussian form with any reasonable form of thermal motion. The fact that some map peaks may be represented approximately by this form is, in part, a consequence of the finite series.

The value $N=2{\cdot}26$ obtained for the extrapolated f curve (A3·3) suggests that the principal contributors to the f curve at the limiting reciprocal radius are the two non-bonding electrons in each carbon atom. If this were the case, one would expect that the central portion of the 'real' atom would rise head and shoulders above the outer regions. Such an effect may be seen in the line section of the projected charge distribution in diamond on a (110) plane, with the diffraction effects eliminated by using an extrapolating f curve, given in Fig. 8 of Van Reijen's paper. In that case a sharp peak, of approximate integrated volume two electrons, rises above the density in the outer regions. However, examination of Fig. A3 shows no such 'super peak', and the value $N=2{\cdot}26$ is without special significance.

References

Beevers, C. A. & Lipson, H. (1936). *Proc. Phys. Soc., Lond.*, **48**, 772.

Booth, A. D. (1945*a*). *Nature, Lond.*, **156**, 51.

Booth, A. D. (1945*b*). *Phil. Mag.* **36**, 609.

Booth, A. D. (1946*a*). *Proc. Roy. Soc.* A, **188**, 77.

Booth, A. D. (1946*b*). *Trans. Faraday Soc.* **42**, 444.

Booth, A. D. (1947*a*). *Proc. Roy. Soc.* A, **190**, 482.

Booth, A. D. (1947*b*). *Nature, Lond.*, **160**, 196.

BOOTH, A. D. & LLEWELLYN, F. J. (1947). *J. Chem. Soc.* p. 837.
BRAGG, W. H. (1915). *Philos. Trans.* **215**, 283.
BRAGG, W. L. & WEST, J. (1930). *Phil. Mag.* **10**, 823.
BROCKWAY, L. O. & PAULING, L. (1934). *J. Chem. Phys.* **2**, 867.
BROWNLEE, K. A. (1947). *Industrial Experimentation.* London: His Majesty's Stationery Office.
COSTAIN, W. (1941). Ph.D. Thesis, University of Birmingham.
COX, E. G. & CRUICKSHANK, D. W. J. (1948). *Acta Cryst.* **1**, 92.
COX, E. G., GROSS, L. & JEFFREY, G. A. (1947*a*). *Nature, Lond.*, **159**, 433.
COX, E. G., GROSS, L. & JEFFREY, G. A. (1947*b*). *Proc. Leeds Phil. Lit. Soc.*, Science Section, **5**, 1.
COX, E. G. & JEFFREY, G. A. (1942). *Trans. Faraday Soc.* **37**, 241.
FANKUCHEN, I. (1947). *J. Sci. Instrum.* **24**, 19.
JAMES, R. W. (1948). *Acta Cryst.* **1**, 132.
JEFFREY, G. A. (1945). *Proc. Roy. Soc.* A, **183**, 388.
JEFFREY, G. A. (1947). *Proc. Roy. Soc.* A, **188**, 222.
LONSDALE, K. (1947). *Philos. Trans.* **240**, 219.
PATTERSON, A. L. (1944). *Phys. Rev.* **65**, 195.
ROBERTSON, J. M. (1935). *Proc. Roy. Soc.* A, **150**, 348.
ROBERTSON, J. M. & WHITE, J. G. (1947). *Proc. Roy. Soc.* A, **190**, 329.
ROBERTSON, J. M. & WOODWARD, I. (1940). *J. Chem. Soc.* p. 36.
VAN REIJEN, L. L. (1942). *Physica, 's Grav.* **5**, 461.
WHITTAKER, E. T. & ROBINSON, G. (1944). *The Calculus of Observations.* London and Glasgow: Blackie.

Acta Cryst. (1949). **2**, 82

The Experimental Determination of the Soundness of Crystals from X-ray and Density Measurements*

BY M. E. STRAUMANIS

Department of Metallurgy, University of Missouri, School of Mines, Rolla, Mo., U.S.A.

(*Received* 9 *August* 1948 *and in revised form* 7 *December* 1948)

A method is described which makes possible the detection of imperfections in crystals of elements, compounds and solid solutions by the determination of molecular weights from X-ray and density measurements. Precision determinations of density and lattice constants are necessary for the calculations.

It is shown that the elements Ag, Cu, Fe, Pb, Te and the compounds LiF, $Pb(NO_3)_2$ and calcite may be considered as having a normal structure, as the number of imperfect sites found does not exceed the possible error of determination of these imperfections. A normal structure is also shown by the tetragonal mixed crystals of the general formula (Zn, Cd, Co) $[Hg(CNS)_4]$ and by the α phase in solid solutions of the Ag–Zn system.

The problem

To determine whether a crystal is sound or whether it has vacant sites or interstitial atoms it is necessary to know the lattice constants and the density with high precision. It is well known from the publications of Barrett, Bunn, Buerger, Foote & Jette, Hume-Rothery, Laves, Owen and others how to distinguish the different types of imperfections in crystals. The method developed, however, is not unambiguous, e.g. concerning Avogadro's number, for which different values have been used by different investigators. A modification of the method which avoids such pitfalls may be proposed as follows.

The method

The idea of this method is to use the molecular (or atomic) weights determined by means of X-rays (M_X) and to compare them with the chemical data (M). The M_X can be calculated by the formula

$$M_X = k N_s v d / n, \tag{1}$$

where v is the volume of the unit cell in kX^3, d is the density, and n is the number of molecules per unit cell. The difference between this formula and other similar ones (for example, the one for the determination of density by X-rays) lies in two respects: (1) N_s represents the Avogadro number used by Siegbahn in his wave-length determinations ($N_s = 6{\cdot}0594 \times 10^{23}$), and (2) $k = 1{\cdot}0002$ is a factor representing the raising of the molecular weight of calcite from 100·075, as used in the time of Siegbahn's X-ray wave-length determinations, to 100·095, the most recent figure including the effect of mixed-crystal formation. The reason why Siegbahn's Avogadro number and the factor k must be used is shown in an article due to appear in the *Zeitschrift für Physik*. If it is desirable to use Ångström units, the Siegbahn Avogadro number must be replaced by

---

* Presented to the First Congress of the International Union of Crystallography at Harvard University, Cambridge, Massachusetts, in July 1948 under the title, 'Some Further Examples of Ideal Structure of Crystals'.