Acta Cryst. (1958). **11**, 129

A Least-Squares Analysis of the Diffuse X-ray Scattering from Carbons

BY R. DIAMOND*

Crystallographic Laboratory, Cavendish Laboratory, Cambridge, England

(*Received* 15 *May* 1957)

In an earlier paper, the theoretical X-ray intensities scattered from perfectly regular, condensed, aromatic molecules of various sizes were presented in numerical form. The present paper describes a matrix method whereby a linear combination of such theoretical curves may be fitted to the observed intensities with minimum total square error; the coefficients in the combination so found at once provide an estimate of the distribution of molecular sizes within the material. Some features of the method are briefly discussed and some experimental examples are given.

Introduction

Hitherto, the sizes of the flat, aromatic, graphite-like molecules (or 'layers') which occur in all 'amorphous' carbons, have been estimated from X-ray measurements by one or other of two methods: (i) a direct comparison of the observed diffraction bands with the line profiles calculated by Warren (1941), Wilson (1949) and Brindley & Méring (1951), a method which has been used by Franklin (1950, 1951*a*, *b*), Hirsch (1954) and others; and (ii) the inversion of the observed intensities by the Debye radial distribution function, as was done by Franklin (1950), Nelson (1954) and others. Both the above techniques are subject to serious limitations: the first technique is not strictly applicable when a wide distribution of molecular sizes is present in the specimen, since the large molecules contribute heavily to sharp central maxima, and are principally responsible for the height of each peak, whereas the smaller molecules produce what appears as a broadening of each peak at its base. No *single* curve of the Warren type will fit such an observed curve, and what is judged to be the best fit of a single curve in these circumstances usually leads to an over-estimate of the mean layer-size, since the small molecules are not properly accounted for.

The second method is beset with even greater difficulties if layer sizes are to be determined. First, if atomic rather than electronic distributions are of interest, it is necessary to subtract the incoherent (Compton) scattering from the observations, and to divide the remainder by the square of the atomic scattering factor for carbon, f_C^2, before inversion. The observations must therefore be placed on an absolute scale, which is usually done by scaling at high angles. This procedure is itself subject to error, especially as the subtraction, important at medium angles, depends on data at high angles and therefore calls for long-range consistency in both theoretical and observational data. Furthermore, since the incoherent radiation forms the greater part of the scattering, even at medium angles, a small error in scaling can produce a large error in the remainder, which has yet to be divided by f_C^2 before the data are ready for inversion.

Secondly, in the case of carbons it is usually difficult to cut off the intensity curve satisfactorily for Fourier inversion.

Thirdly, the radial distribution of atomic centres, when obtained, does not directly give information concerning the sizes of the layers, but rather, the information serves to demonstrate the fact that the atoms are arranged, as in graphite, on a hexagonal net. Information concerning the sizes of the layers can, in principle, be obtained from the envelope of the peaks within the radial distribution curve, as this should have a definite shape (for a single size group), and should fall to zero when the interatomic vector, r, becomes equal to the diameter of the layers present. If a distribution of sizes is present, the problem becomes more complicated, and the envelope becomes concave upwards, with the cut-off again indicating the size of the larger molecules.

The method of analysis now to be described makes the assumptions, based on the earlier work of others, that the layers present in amorphous carbons are graphite-like in their internal co-ordination, are generally randomly orientated, but with a systematic tendency to align themselves parallel to their closest neighbours, and that they are perfectly regular in their internal structure. This last assumption is the only one so far not adequately substantiated, and there is reason to believe that it is only partially true. Imperfections, if such exist, cause the apparent size of the molecules to be smaller than the true size.

The method then has the following points to commend it:

1. It directly yields a distribution of molecular sizes, usually presented in histogram form, the terms in

* Present address: Crystallographic Laboratory, Department of Physics, The Pennsylvania State University, University Park, Pennsylvania, U.S.A.

which represent fractions by weight of material in particular size groups, and which are usually reliable to a few per cent in favourable cases.

2. The pre-treatment of the observational data (in the Debye–Scherrer case) is limited to the removal of the polarization factor, which depends on the nature of the monochromator employed, and which may vary from one application to another.

3. After allowing for such pre-treatment, it attaches equal weight to all observations, and guarantees the best possible fitting between observed and calculated data.

4. It places minimum reliance on the theoretical data for the incoherent scattering, the effect of which falls out entirely when the first eigenfunction is extracted (see below).

5. It works within an angular range chosen by the author as containing useful information unmasked by irrelevant effects, and is completely independent of the behaviour of the observed curve outside this range, so that uncertainties in the values of the Compton scattering at high angles are immaterial. The inversion is dependent on eigenfunctions which are perfectly orthogonal within this range, and, unlike a Fourier inversion over an arbitrary interval, has no cut-off problem.

6. As the mathematics is entirely linear, a weighted mean determination of the size distribution can be obtained from two or more sets of data, by adding the intensity measurements together before the inversion is done.

7. It is very rapid in its application; the inversion process requiring under an hour with an electric desk calculator.

In its present form the method makes no allowance for the presence of non-carbon atoms. Such 'foreign' atoms are assumed to scatter X-rays with the same angular dependence as does carbon within the working range, except for a constant factor approximately equal to the square of the ratio of the atomic number of the element concerned to that of carbon. All such atoms are assumed to contribute only as single atom scatterers, having no systematic phase relationship to their neighbouring atoms. In a later paper, which will deal with the results obtained from a series of carbonization experiments, these assumptions will be critically re-examined and a treatment of the limiting effects of oxygen and hydrogen will be given; and their influence on the results obtained will be illustrated. At the same time, allowance will be made for the revision of the figures for Compton scattering occasioned by the recent publication by Keating & Vineyard (1956) of accurate figures for this function. For the purposes of the present paper, however, these revisions may be ignored, though any accurate application of the method described herein should take them into account.

The mathematical basis of the method

If we assume that an amorphous carbon consists of an assembly of graphite-like layers of various sizes, together with some disorganized material which may be regarded as producing gas-like (i.e. single-atom) scattering, then, in the absence of any phase coherence between the various scattering units so defined, the intensity of X-rays scattered from such an assembly is a linear combination of the intensity functions associated with each layer-size group. The coefficients in such a combination are proportional to the fraction, by weight, of material in each size group, provided that the intensity functions are expressed on a basis of scattering per atom. On such a basis, the intensity function for a given size group may be expressed as

$$B_i(s) = f^2 J_i(s) + C(s) , \qquad (1)$$

in which $B_i(s)$ is the intensity scattered by size group i as a function of s, ($s = 2 \sin \theta/\lambda$), f is the atomic scattering factor for carbon, $J_i(s)$ is the scattering in atomic units for size group i, as defined and calculated in a previous paper (Diamond, 1957) and $C(s)$ is the incoherent (Compton) scattering. We then have

$$I(s) = \sum_i \lambda_i B_i(s) + v(s) , \qquad (2)$$

in which $I(s)$ is the observed intensity, $v(s)$ is a residual error and the λ_i's are the coefficients of each size group, such that $\sum_i \lambda_i$ is equal to the ratio of absolute to experimental intensity units.

In the circumstances in which the present work was developed, only one type of phase coherence between the scattering units is encountered—that due to the 'turbostratic' packing of the layers parallel to one another. Such packing produces local phase coherence along a line normal to any given layer, so that the intensity in reciprocal space is that due to the layers taken independently (i.e. in accordance with equation (2)), together with a modulation along the line (00l) in reciprocal space. All other regions of reciprocal space are unaffected by such packing, and it happens that in the applications of which the author has experience the (00l) variation has decayed to negligible amplitude for $l > 4$ so that the interpretation of the (11) and (20) bands (which occur at higher s values than that corresponding to (005)) is not complicated by the packing of the layers. If, however, the packing is good enough for a (006) reflection to be present, it will confuse the interpretation unless it is accurately subtracted out. If the packing is sufficiently poor, equation (2) may be taken as a complete representation of the intensity (but for geometrical factors) for regions of $s > 0{\cdot}65$ Å^{-1}.

We now select 31 points on the s scale, evenly spaced in the range $0{\cdot}66 \leq s \leq 0{\cdot}96$ Å^{-1}, and measure the observed intensity at each point, divide each by the polarization factor, and arrange the resulting

readings in order in a column matrix, **I**, of 31 elements, with the value for $s = 0{\cdot}66$ Å^{-1} heading the column. We also define a matrix **B** having 31 columns corresponding to the 31 s values, and 8 rows corresponding to 8 selected size groups, and whose elements are the appropriate B values, in accordance with equation (1). We also define a column matrix, $\boldsymbol{\lambda}$, whose 8 elements are the coefficients in the linear combination, and a column matrix **v** containing the 31 residuals. Equation (2) may then be rewritten as

$$\mathbf{I} = \mathbf{B}'\boldsymbol{\lambda} + \mathbf{v}\,, \tag{3}$$

in which the prime denotes a transpose (i.e. rows and columns interchanged). It is now our purpose to discover $\boldsymbol{\lambda}$ such that the product $\mathbf{v}'\mathbf{v}$ (sum of the squares of the errors) is a minimum.

It may be said at once that the residuals, **v**, may not contain any linear combination of the theoretical functions, B, (provided these are linearly independent) since, if they did, a reduction in the sum of their squares would be effected by transference of that combination to the combination $\mathbf{B}'\boldsymbol{\lambda}$. Since the residuals are to be void of any linear combination of the quantities B, they must contain no component of any of the quantities B taken singly; thus the matrix **v** must be orthogonal to every row of **B**, i.e.

$$\mathbf{B}\mathbf{v} = \mathbf{0}\,. \tag{4}$$

Premultiplication of equation (3) by **B** then yields at once

$$\mathbf{B}\mathbf{I} = \mathbf{B}\mathbf{B}'\boldsymbol{\lambda}\,, \tag{5}$$

which is an equation (the 'normal equations' in matrix form) soluble for $\boldsymbol{\lambda}$. The same result may be obtained by setting $\frac{\partial}{\partial \lambda_i} \Sigma v^2 = 0$ for all i.

Now **BB'** (being equal to its transpose) is a symmetrical square matrix (8×8), and as such it may be considered as a symmetrical, second-rank tensor relating the two 8-dimensional vectors $\boldsymbol{\lambda}$ and **BI**, and may therefore be represented by an 8-dimensional ellipsoid. **BB'** contains theoretical quantities only, and so may be treated without reference to any one set of observations, all the experimental information lying in the two vectors.

The shape and orientation of the representational ellipsoid are of prime importance in determining the characteristics of the problem, and are worthy of some examination. For any given set of intensities, we have from equation (3) that

$$\boldsymbol{\delta}\mathbf{v} = -\mathbf{B}'\boldsymbol{\delta\lambda}\,, \tag{6}$$

and if σ is the sum of the squares of the errors, then $\sigma = \mathbf{v}'\mathbf{v}$ and

$$\delta\sigma = \mathbf{v}'\boldsymbol{\delta}\mathbf{v} + \boldsymbol{\delta}\mathbf{v}'\mathbf{v} + \boldsymbol{\delta}\mathbf{v}'\boldsymbol{\delta}\mathbf{v}\,, \tag{7}$$

which, with equations (6) and (4), gives

$$\delta\sigma = \boldsymbol{\delta\lambda}'\mathbf{B}\mathbf{B}'\boldsymbol{\delta\lambda}; \tag{8}$$

i.e. a given departure, $\boldsymbol{\delta\lambda}$, from the least-squares solution in λ-space increases the sum of the squares of the errors by $\delta\sigma$, given above. Or, conversely, equation (8) may be taken to define a contour of constant error in the neighbourhood of the solution. Such a contour is clearly the representative ellipsoid of **BB'** in λ-space, being a quadratic in $\boldsymbol{\delta\lambda}$. Note that since equation (8) involves $\boldsymbol{\delta\lambda}$ but not $\boldsymbol{\lambda}$ itself, such a contour of constant error is independent of the location of the solution.

A further appreciation of the shape and orientation of the ellipsoid may be obtained from the following discussion. Suppose that it is desired to fit to an observed intensity curve a linear combination of only two, closely similar, theoretical functions B. Fig. 1

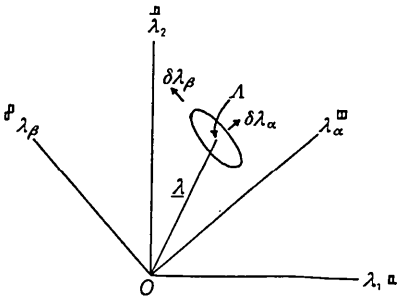

Fig. 1. Two-variable example in λ-space.

then represents the situation in two-dimensional λ-space, in which the point A represents the least-squares solution. It is then clear that a departure from A in the direction of $\delta\lambda_\alpha$, by increasing both the coefficients λ_1 and λ_2, will cause the fitted theoretical curve to be raised over-all, with consequent rapid increase in σ. The solution A is therefore closely defined in this direction.

If, on the other hand, a departure is made from A in the direction of $\delta\lambda_\beta$, the calculated curve will remain close to the observed curve, but will change a little in shape, depending on the details of the two theoretical curves employed. Comparatively large departures from A can therefore be tolerated in the β direction.

In the extreme case when the two theoretical curves are identical, the solution A will be replaced by a line running in the β direction. In these and similar circumstances the normal equations are said to be ill-conditioned (Hartree, 1952), and it is clearly preferable to specify A in terms of coordinates λ_α and λ_β, as the α component is quite unaffected by indeterminacy in the β component.

Since the theoretical curves employed (B values) do resemble one another, and approach linear combinations of one another, it has been found necessary to determine the 8-dimensional analogues of the directions α and β.

To obtain the principal axes of the ellipsoid defined by **BB'**, we seek a unitary orthogonal matrix **A**, such

that the matrix **ABB'A'** is in diagonal form. The rows of the matrix **A** are then the components of the unit principal axes of the ellipsoid **BB'** in λ-space expressed on the original axes, and are known as the eigenvectors of the problem. These are shown in Fig. 3.

The matrix **A** has been found by a procedure of iterative rotation, using the computer EDSAC and a sub-routine designed by Dr D. J. Wheeler of the University Mathematical Laboratory, Cambridge.

A is evidently the poly-dimensional analogue of the array of direction cosines familiar in three-dimensional transformations, and as such (and by definition) it has the property that

$$\mathbf{A'A} = \mathbf{E}\,, \tag{9}$$

in which **E** is the unit diagonal matrix. Premultiplying equation (5) by **A** and interposing **A'A** gives

$$\mathbf{ABI} = \mathbf{ABB'.A'A.\lambda}\,. \tag{10}$$

Now **ABB'A'** is in diagonal form, whence its reciprocal is obtained by inverting its diagonal elements, whence

$$\boldsymbol{\mu} = \mathbf{A\lambda} = (\mathbf{ABB'A'})^{-1}\mathbf{ABI} \tag{11}$$

and

$$\boldsymbol{\lambda} = \mathbf{A'}(\mathbf{ABB'A'})^{-1}\mathbf{ABI} \equiv \mathbf{HI}\,. \tag{12}$$

Thus **H** is a matrix, reciprocal to **B**, which contains theoretical quantities only, such that the distribution $\boldsymbol{\lambda}$ is obtained by premultiplying **I** by **H**.

Now we may define a matrix **C**, given by

$$\mathbf{AB} = \mathbf{C}\,, \tag{13}$$

in terms of which

$$\boldsymbol{\mu} = (\mathbf{CC'})^{-1}\mathbf{CI}\,. \tag{14}$$

By equation (13), and from the diagonal nature of **ABB'A'**, we see that the rows of **C** (the eigenfunctions) form a set of orthogonal linear combinations of the functions which form the rows of **B**; and from equation (14) we see that we obtain the solution, $\boldsymbol{\mu}$, by multiplying the function I by each row of **C**, summing the products, and dividing the result by the appropriate element (eigenvalue) on the diagonal of **CC'**. This may be recognized as parallel to the procedure of Fourier analysis of a function of one variable, in which the eigenfunctions are sines and cosines and the eigenvalues are all the same. The eigenfunctions of the present problem (actually the quantities $(\mathbf{CC'})^{-\frac{1}{2}}\mathbf{C}$) are shown in Fig. 2; their resemblance to cosines is quite marked.

Now it has been pointed out that the elements of $\boldsymbol{\mu}$ (i.e. the coordinates of the solution Λ expressed on the principal axes) are independent of one another, and that some of them may be quite inaccurately determined if the theoretical curves employed approximate to linear combinations of one another. Since a molecular size group cannot be negatively populated, it is clear that in the two-dimensional example of Fig. 1, the solution, Λ, must lie in the first quadrant, so that a constraint is placed on the β component, ($|\lambda_\beta| \leq \lambda_\alpha \geq 0$). In the polydimensional case, the least accurate component, the θ component,

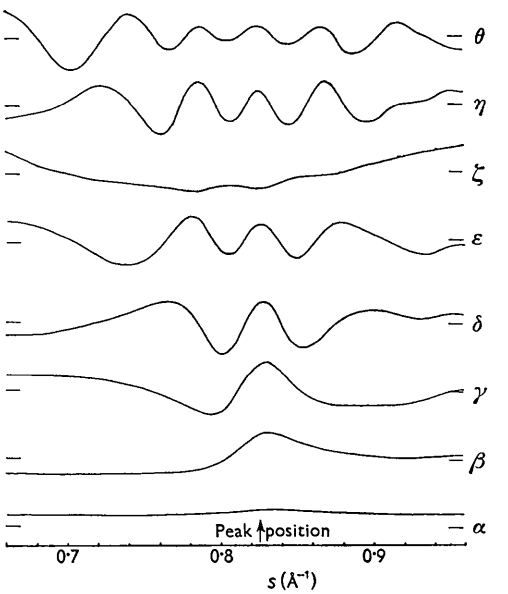

Fig. 2. The eigenfunctions of the problem. Each function is a linear combination of the theoretical intensity functions employed, and each one is orthogonal to all the others.

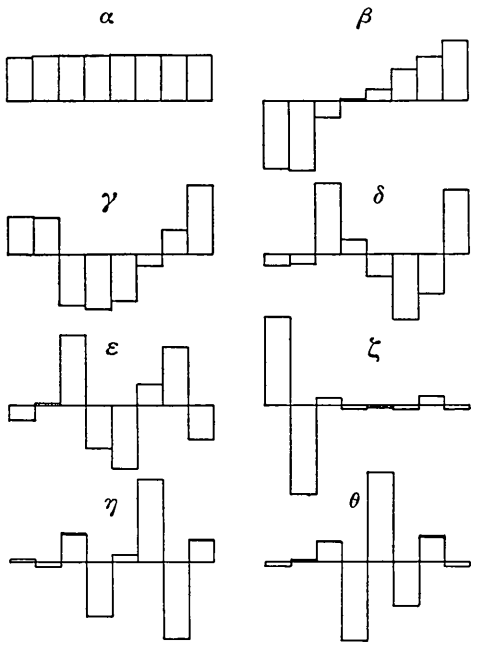

Fig. 3. The eigenvectors of the problem. The terms in each of these vectors are the coefficients which yield the orthogonal combinations of Fig. 2. They are also normalized and orthogonal to one another. Each experimental histogram is made up from a combination of these in the manner of Fig. 4. The eight size groups appearing in each eigenvector are, from left to right, single atoms, pairs of atoms, and aromatic layers of nominal diameters 5·8, 8·4, 10, 15, 20 and 30 Å respectively.

say, is again subject to a similar constraint for physical reasons. The least accurately determined component is always the one which (see Fig. 3) has alternating terms (cf. β component in Fig. 1); thus a molecular size distribution containing a large component of this type will be polymodal, whereas one would expect a distribution to be unimodal or bimodal but probably not more. A large θ component is therefore improbable and subject to definite limits, and since it is not possible to measure it accurately it is considered best to constrain its coefficient to zero, and to bear in mind when considering a molecular size histogram that the number of degrees of freedom allowed to it may be less (by one or two) than the number of terms in the histogram.

The elimination of such immeasurable components may conveniently be effected by incorporating a filter matrix $\mathbf{Z}_f$ in $\mathbf{H}$. The matrix $\mathbf{Z}_f$ is a diagonal square with f1's on the diagonal (f = number of degrees of freedom allowed), and $(n-f)$0's (n = number of size-groups used in the histograms). We then employ in equation (12) a matrix $\mathbf{H}_f$ given by

$$\mathbf{H}_f = \mathbf{A}'\mathbf{Z}_f(\mathbf{ABB}'\mathbf{A}')^{-1}\mathbf{AB}\,, \qquad (15)$$

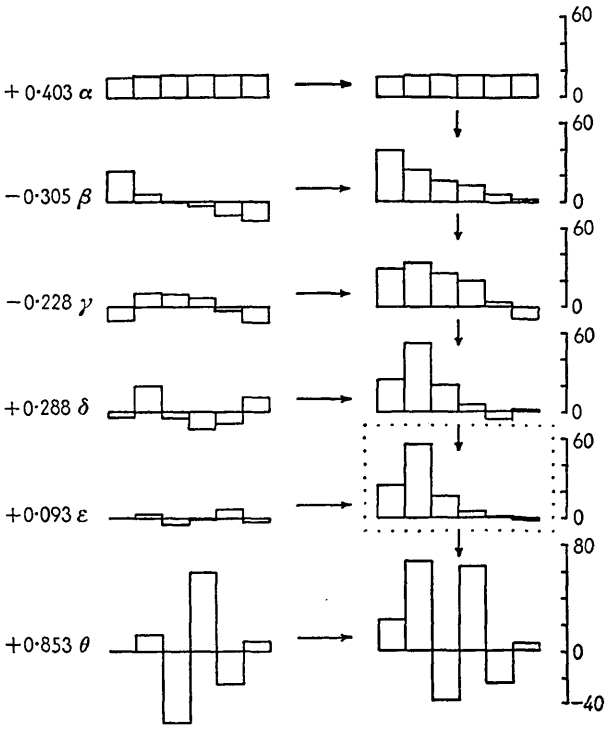

Fig. 4. The development of an experimental histogram from its components, showing the effect of filtering. The example was obtained from an earlier set of matrices than those now presented, wherefore the histograms shown have only six terms corresponding (from left to right) to amorphous material (single-atom scattering) and to aromatic layers of nominal diameters 5·8, 8·4, 10, 15 and 20 Å. The vertical scale is in weight per cent. The example is of a coal (initially 84% C d.a.f. basis) carbonized for 3 hr. at 300° C., having an estimated carbon content of 81%.

which extracts a maximum of information from the observations whilst preventing the perturbation of the results by effects which are beyond resolution.

Fig. 4 illustrates the development of a histogram from its components. The coefficients on the left of the diagram are the elements of $\boldsymbol{\mu}$. The first component, α, having all positive terms, sets the overall height of the fitted theoretical curve. It is the analogue of the constant term that generally heads a Fourier cosine series. It is also the only term for which the sum of the coefficients, λ, relating to the size groups, differs significantly from zero, so that the coefficients of the remaining eigenfunctions are derived from the coherent part of the scattering only.

The β term then determines whether the layers present are predominantly large (β positive) or small (β negative). The β eigenfunction depends for its shape almost entirely on differences between the accurate and smooth functions $J(s)$, since the same f^2 values are used for each $B(s)$ and the Compton scattering drops out. The method therefore makes the best possible use of the theoretical data.

The remaining eigenvectors fill in the form of the histogram with increasing detail and decreasing accuracy. In particular, the θ component, which is beyond resolution, has an absurdly large coefficient, the effect of which is to mask the good information obtained from the first five eigenfunctions. The histogram shown in the dotted box is therefore adopted as the final result; it may be obtained by a single application of the matrix $\mathbf{H}_5$, five degrees of freedom being considered optimum.

Specification and mode of use

The matrices now presented employ six nominal layer sizes (L values as previously defined (Diamond, 1957)) of 5·8, 8·4, 10, 15, 20 and 30 Å, and in addition they allow for single atoms and for pairs of atoms 1·54 Å apart. The characteristics described below indicate, however, that it is not possible to distinguish between single atoms and pairs of atoms, so that many of the following histograms show only seven terms, the coefficients for pairs and singles having been added together. In cases where only six terms are shown, the results are obtained from an earlier set of matrices (not published), in which provision was not made for layers of 30 Å diameter, nor for pairs of atoms.

The present matrices are based on an assumed C–C bond length of 1·405 Å for all six aromatic size groups, and they therefore use $J(s)$ figures derived from, but not identical with, those previously published (Diamond, 1957). The use of a single bond length, different from either the benzene (1·39 Å) or graphite (1·42 Å) values, has the advantage that an accurate mean bond length for the material investigated may be obtained by measuring the disregistry on the s scale between the observed curve and the fitted theoretical curve.

The fractional change in bond length relative to the assumed value is then given by

$$\varkappa = \Sigma vsI'/\Sigma(sI')^2 , \qquad (16)$$

in which I' is $\partial I/\partial s$ and the summation is over the fitted range. Allocation of assumed bond lengths thought to be appropriate to each L value results in confusion if an interpretation of $\varkappa$ is to be sought.

In cases where the experimentally determined bond length lies outside the range $1{\cdot}400 < l < 1{\cdot}410$ Å, it is advisable to perform two analyses, using the calculated curve from the first analysis to determine $\varkappa$, which is then used in a three-term interpolation formula to side-shift the observed intensity curve to a position corresponding to $l = 1{\cdot}405$ Å. The shifted observed curve is then used for the final analysis. This method of determining the mean bond length generally leads to values having a random error of $\sim 0{\cdot}002$ Å.

It is noteworthy that a slight horizontal disregistry between observed and calculated intensity curves has very little effect on the layer-size histograms obtained, since such a disregistry introduces a nearly odd or sine-like function to the residuals, v, with respect to the peak position, whereas the eigenfunctions are nearly even or cosine-like. Thus a considerable contribution to v from this source may be tolerated whilst still maintaining the orthogonality of **v** to **B**. The empirical limits to this tolerance are indicated above.

Further remarks on the measurement of l have been given by Diamond (1956, chap. 9) in which a scheme to include $\partial I/\partial s$ and sine-like eigenfunctions in the matrices is described.

In addition, to computing the histograms and bond lengths, it is profitable to calculate a mean layer size given by

$$\bar{L} = \Sigma'\lambda L/\Sigma'\lambda , \qquad (17)$$

in which the prime denotes omission of the terms for single atoms and pairs. In computing $\bar{L}$ we drop much of the information contained in a histogram, but obtain a single parameter of high accuracy, since $\bar{L}$ is very little affected by errors in the coefficients of the least accurate eigenvectors. Thus successive determinations from a given sample may yield histograms of varying appearance, yet having closely similar $\bar{L}$ values. It may also be shown that the accuracy of $\bar{L}$ decreases with increasing $\bar{L}$. Random errors in $\bar{L}$, for cokes formed below 1000° C., have generally been found to be of the order of 0·2–0·3 Å. These points are all exemplified in Fig. 7, in which $\bar{L}$ values are shown alongside the histograms from which they were obtained.

The figures employed for the atomic scattering factor are those given by Hoerni & Ibers (1954) for carbon in the valency state. Figures for f^2 at the desired fine interval were obtained by careful graphical interpolation, followed by smoothing to bring second differences to nearly constant values. The least-squares analysis, is, however, quite insensitive to the choice of figures for f^2 provided the same figures are used throughout. Tests have shown that if the figures given by McWeeny (1951) are to be preferred to those of Hoerni & Ibers, then results obtained from these matrices will not be in error from this cause by more than 4% in any one histogram term, even in the least favourable cases.

The figures for the incoherent (Compton) scattering were taken from Compton & Allison (1935) and were treated in the same way as those for f^2. It is regrettable that at the time the present matrices were computed and employed, no more accurate figures than these were available; the effects of errors from this source are, however, minimized by this interpretive technique, as indicated above.

In common with Hirsch and J. D. Watt (private communications), it was found that agreement between observed and calculated curves in the range $2{\cdot}0 < s < 2{\cdot}2$ Å^{-1} was improved by omitting the Breit–Dirac relativistic recoil factor, B^{-3}, from the figures for the Compton scattering (B = ratio of X-ray frequencies in the primary and scattered beams), and this has been done in computing the matrices. This factor has very little effect in the working range ($\sim 3\%$), so that the uncertainty attaching to this factor is of little consequence in this method, unlike methods which involve scaling at high angles.

Some characteristics of the method, with examples

Before any critical appreciation of results can be attempted, it is necessary to consider some of the characteristics of the matrices used, especially those associated with the filtering, and with certain types of systematic error.

It is fruitful to consider the results obtained from the self-analysis of the theoretical curves used, i.e. the product $\mathbf{H_5B'}$ (which is equal to $\mathbf{A'Z_5A}$). Fig. 5(*a–h*) shows eight such self-analyses. In each case, the broken line represents the distribution to which the distribution shown with a full line is calculated to correspond. First, as the eigenvectors of Fig. 3 show, the first five eigenvectors treat the single atoms and pairs as almost the same, discrimination between them falling entirely to the 6th (ζ) vector, which is beyond resolution and excluded from $\mathbf{H_5}$. The effect of this is obvious in Fig. 5(*a*, *b*) and entitles us to recombine the terms for single atoms and pairs, and to extend the interpretation of this term to small aliphatic groupings of about one to three atoms. It is also noteworthy that, since only the first three eigenvectors (Fig. 3) contain appreciable terms for the amorphous scattering, the amorphous term is generally the most accurate one in any experimentally determined histogram.

Next we note that narrow distributions at the

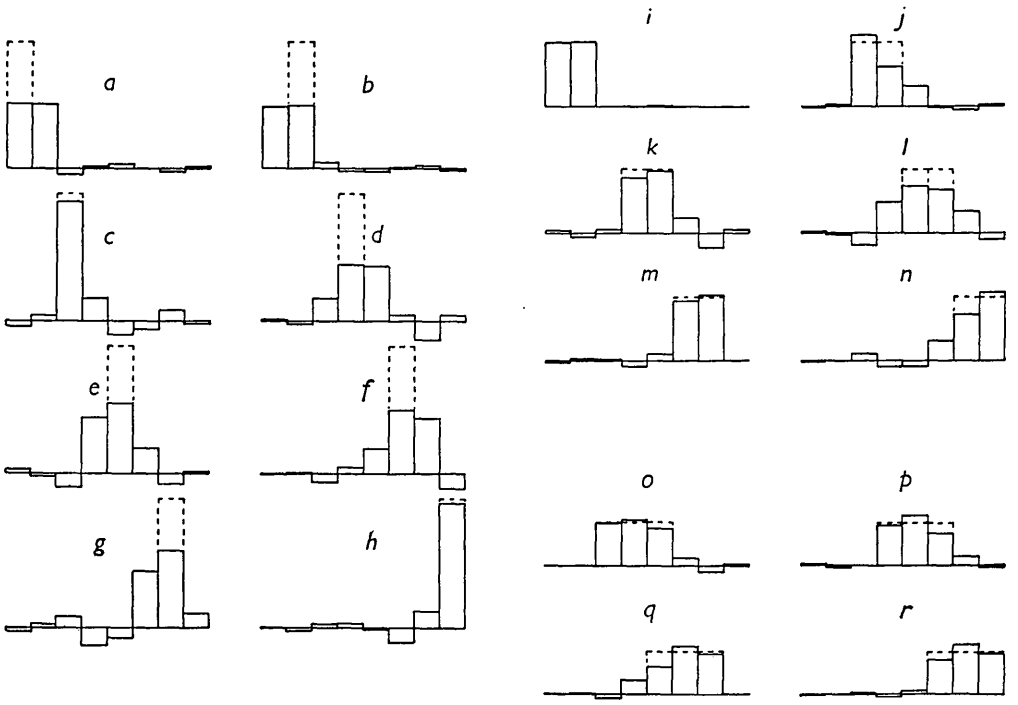

Fig. 5. Self-analysis of the theoretical data.

extremes of the range (Fig. 5(*c*, *h*)) are better approached than narrow distributions in the middle of the range (Fig. 5(*d*, *e*, *f*)). The apparent broadening in the latter cases is due entirely to the suppression of the η and θ vectors, which have larger terms in the middle of the range than at the extremes. From inspection of these figures, we conclude that an experimental histogram may be broader than the true histogram but cannot be narrower; from which we infer that a narrow experimental histogram probably implies an even narrower true distribution.

Fig. 5(*i*–*r*) shows analyses of distributions with 50% of material in each of two, or with 33% in each of three neighbouring size groups. From an inspection of these we may conclude that the broader and more smoothly varying a given distribution is, the more faithfully it will be reproduced; thus the broad distributions which may be expected to occur in practice should be interpreted with greater fidelity than that pertaining in the artificial circumstances of Fig. 5(*a*–*h*).

Thus the effect of the exclusion of the ζ, η and θ vectors is a genuine up-grading of the numerical accuracy of the terms in the resulting histograms, at the expense of down-grading their interpretive significance, thus yielding information of optimum value. The histograms obtained are a convolution of the true distribution with the distributions of Fig. 5(*a*–*h*), so that any term in such a histogram no longer has the significance of material in that size group alone, but of sizes in the close neighbourhood of the nominal size.

In addition to the above characteristics, which are purely properties of the matrices themselves, there are many others which involve the nature of the carbons investigated and the verity of the assumed model upon which the interpretation is based. A number of such characteristics are to be described in another paper, in which the results of an investigation into the structures of carbonized coals will be presented. However, as an example of the manner in which such aberrations may be studied, we show here the effect produced by an error in the positioning of the origin of intensity, or by fluorescence in the specimen.

Since the mathematics of the interpretation is entirely linear, it is clear that the effect of adding to the observed intensities any function, T, is to add to the results obtained a matrix $\mathbf{H_5T}$; and in cases such as background error, where T is a known function, it may be easier to subtract its effect from the histograms obtained than from the initial intensity readings. Fig.

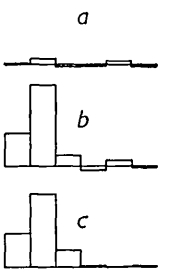

Fig. 6. The effect of fluorescence or an error in positioning the origin of intensity.

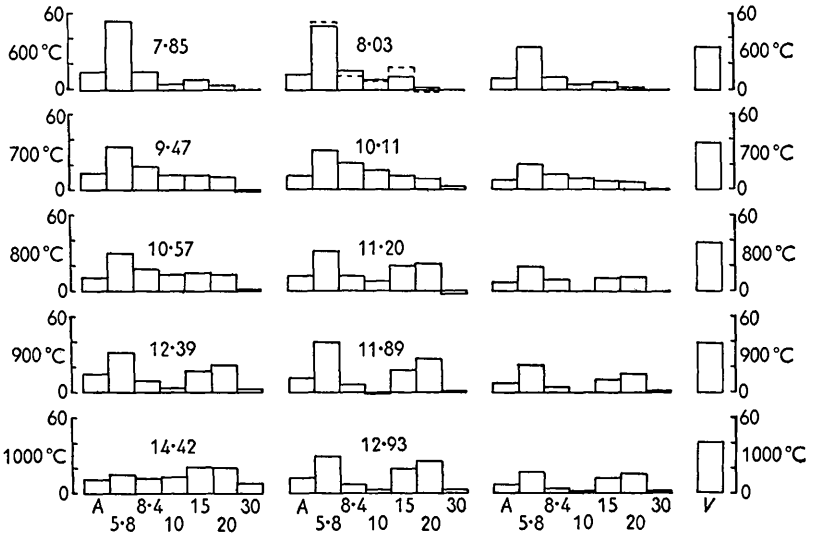

Fig. 7. Experimental examples from a vacuum-carbonized coal of initially 84 % C. In this temperature range the carbon content rises steadily from 86 % to 95 %.

6(*a*) shows the type of term which may be attributed to fluorescence, or to an error in locating the origin of intensity, whilst Fig. 6(*b*) shows an experimental histogram which may be affected by fluorescence, and Fig. 6(*c*) shows the same after allowing for 4 % fluorescent radiation and renormalizing.

It so happens that in this example the correction term (Fig. 6(*a*)) consists mainly of an ε component, which is the least accurate component admitted to the histograms. Results similar to Fig. 6(*b*) may arise from a random error in the ε component, which is on the verge of resolution. When this occurs, it is considered justified to attenuate the ε vector artificially, as in the example at the top of Fig. 7, in which the result before attenuation is shown in broken lines. It should be emphasized that the increment in the total square error brought about by this attenuation is extremely slight, owing to the smallness of the ε eigenvalue.

Fig. 7 shows a series of results obtained from a caking coal, initially 84% carbon, carbonized 3 hr. *in vacuo* at temperatures indicated. In each case, two independent determinations of the layer-size distribution are shown in the left and central columns, the scales at the sides indicating weight per cent. In the third column, the first two columns are averaged, and scaled down in such a way that the sum of the terms in the histogram, plus the amount of material evaporated, designated V, comes to 100 %. The figure shown above each determination is $\overline{L}$, the mean layer size, in Ångström units. The figure illustrates the general standard of consistency and reliability to be expected. The two determinations at 1000° C. also illustrate a type of error to which the system is particularly sensitive, namely that due to an error in measurement of the (20) band. This band occurs on the high-angle side of the (11) peak, and is difficult to measure accurately, since it appears as little more than an inflexion on a 'tail'. If the (20) band is over-estimated, the effect is to emphasize the bi-modal nature of the result, as in the 1000° C. example, centre column. If the (20) band is underestimated, the centre of the resulting histogram is enhanced. Now if photometric resolution is poor, such an inflexion is more likely to be underestimated than over-estimated, wherefore it may be argued that the true distributions should be at least as bi-modal as those obtained, so that a tendency to bi-modality in the specimens may be taken as established by these results.

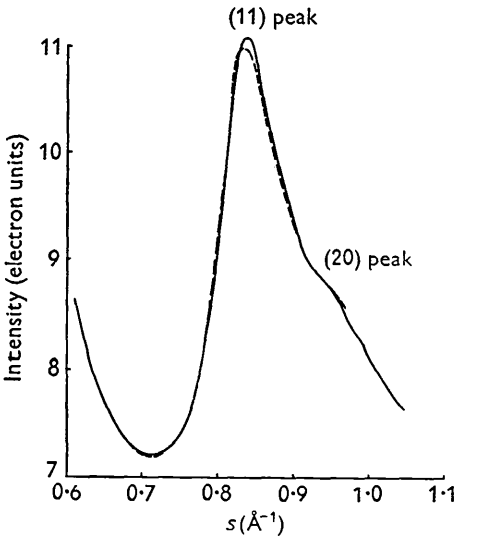

Fig. 8. Comparison of observed and calculated intensities in the fitted range. Full line: observed intensity; broken line: calculated intensity.

Fitting is effected in the range $0{\cdot}66 \leq s \leq 0{\cdot}96$ Å^{-1}. The example is the 800° C. case of Fig. 7.

Fig. 8 shows a typical example of the agreement obtained between the side-shifted observed curve and the fitted calculated curve within the working range. The r.m.s. error between the two curves is about 0·05 electron units of intensity, or approximately 0·5%. Error theory shows that, with five degrees of freedom, readings of this order of accuracy are required to obtain the degree of self-consistency shown in Fig. 7; and this may be obtained using a self-balancing and automatic recording photometer, if a smooth mean line is drawn through the records obtained.

Accuracy

Finally we quote the results of theoretical treatments of random errors.

Since the errors in the components of μ are independent of one another, whilst those of λ are heavily coupled, only errors in μ may readily be investigated. We find

$$\sigma(\mu_\alpha) = \tau_\alpha \sqrt{\frac{f\Sigma v^2}{(m-f)(\mathbf{CC}')_{\alpha\alpha}}}, \qquad (18)$$

in which $\sigma(\mu_\alpha)$ is the standard deviation of μ_α, f is the number of degrees of freedom allowed (generally 5), m is the number of observations employed (31), and

$$\sum_\alpha \tau_\alpha^2 \simeq 1, \qquad (19)$$

the result being normalized if the errors, v, are expressed in electron units.

We also find that

$$\sigma(\bar{L}) = \frac{e}{a}\sqrt{(11{\cdot}8+0{\cdot}047\bar{L}^2)}, \qquad (20)$$

in which e is the r.m.s. error between observed and calculated curves at the 31 fitted points expressed in electron units, and a is the fraction of material in aromatic form, $\bar{L}$ being expressed in Ångström units.

The matrix **B′**

s	A	P	5·8 Å	8·4 Å	10 Å	15 Å	20 Å	30 Å
0·66	9·630	9·71533	7·47338	7·32407	7·23026	7·04386	6·98496	6·85280
0·67	9·505	9·66290	7·37160	7·19868	7·11478	6·92407	6·85425	6·71860
0·68	9·390	9·61333	7·32010	7·08334	7·01246	6·82290	6·75048	6·60340
0·69	9·285	9·56683	7·30735	6·98231	6·92044	6·72655	6·66465	6·15080
0·70	9·180	9·51291	7·34006	6·89962	6·83151	6·64105	6·57491	6·43050
0·71	9·085	9·46221	7·42000	6·85844	6·75675	6·58580	6·49939	6·36120
0·72	9·000	9·41541	7·54631	6·87387	6·70885	6·56493	6·44610	6·29780
0·73	8·920	9·36720	7·70810	6·95706	6·70876	6·55836	6·43588	6·25290
0·74	8·840	9·31256	7·90434	7·12523	6·78912	6·54603	6·42531	6·22640
0·75	8·750	9·24077	8·11820	7·37300	6·96768	6·52749	6·41802	6·21860
0·76	8·665	9·16785	8·34791	7·70625	7·27354	6·57446	6·41612	6·22450
0·77	8·590	9·09964	8·59561	8·12215	7·72117	6·80465	6·44297	6·26890
0·78	8·520	9·03152	8·84432	8·58658	8·28566	7·33094	6·63659	6·33300
0·79	8·460	8·96840	9·08916	9·07942	8·93370	8·21038	7·34483	6·47250
0·80	8·400	8·90076	9·31334	9·55097	9·59542	9·36699	8·75485	7·45850
0·81	8·340	8·82871	9·50478	9·97165	10·20199	10·60752	10·67578	10·28910
0·82	8·285	8·75788	9·66713	10·31226	10·70277	11·68062	12·51061	13·85640
0·83	8·225	8·67862	9·78281	10·54494	11·03419	12·34323	13·56627	15·60460
0·84	8·170	8·60098	9·85987	10·66435	11·18319	12·49817	13·55881	14·72310
0·85	8·125	8·53118	9·90244	10·68167	11·15307	12·19054	12·73615	12·81630
0·86	8·080	8·45877	9·89908	10·59448	10·96193	11·56518	11·62128	11·40270
0·87	8·040	8·38939	9·86220	10·43163	10·66381	10·84789	10·66692	10·61860
0·88	8·000	8·31796	9·78682	10·20834	10·30365	10·16352	10·00292	10·06590
0·89	7·965	8·24993	9·68447	9·95365	9·93595	9·64437	9·55529	9·59610
0·90	7·930	8·18108	9·55331	9·69030	9·60245	9·29166	9·21484	9·20770
0·91	7·895	8·11148	9·40383	9·44184	9·32815	9·08305	8·96308	8·89290
0·92	7·860	8·04176	9·23985	9·21617	9·11825	8·97554	8·83862	8·65860
0·93	7·830	7·97688	9·07428	9·02507	8·97355	8·94515	8·86775	8·65590
0·94	7·795	7·90700	8·89758	8·85504	8·85413	8·93977	8·98354	8·98250
0·95	7·760	7·83806	8·72260	8·70578	8·75625	8·92325	9·07883	9·36660
0·96	7·730	7·77482	8·55785	8·57642	8·65606	8·86288	9·05532	9·37340

The matrix **A′**

	ζ	α	ε	β	θ	η	γ	δ
Amorphous	0·7059421	0·3353978	−0·1140262	−0·5285068	−0·0341473	0·0295038	0·2938581	−0·0918873
Pairs	−0·6991435	0·3489017	0·0180540	−0·5458069	0·0191565	−0·0392034	0·2887727	−0·0770779
5·8 Å	0·0614555	0·3533529	0·5608557	−0·1364764	0·1543683	0·1952877	−0·4059305	0·5580621
8·4 Å	−0·0278902	0·3553148	−0·3362153	0·0128989	−0·6112744	−0·4255372	−0·4367605	0·1194496
10 Å	−0·0223263	0·3562098	−0·4950958	0·0889445	0·6691305	0·0604024	−0·3692149	−0·1785529
15 Å	−0·0320483	0·3581366	0·1699643	0·2439186	−0·3447466	0·6178506	−0·0926505	−0·5226403
20 Å	0·0750120	0·3594012	0·4612799	0·3439538	0·1819970	−0·6042000	0·1902401	−0·3158299
30 Å	−0·0337179	0·3610433	−0·2688589	0·4671951	−0·0356657	0·1671159	0·5418700	0·5051053

The matrix $\mathbf{H}_5'$

s	A	P	5·8 Å	8·4 Å	10 Å	15 Å	20 Å	30 Å
0·66	0·066339 8	0·018195 4	−0·264990 6	0·092619 9	0·179436 4	−0·014693 4	−0·129654 1	0·058771 1
0·67	0·064146 8	0·018619 2	−0·253803 7	0·085976 0	0·169594 4	−0·011454 7	−0·120634 4	0·053461 0
0·68	0·055631 9	0·019392 7	−0·211421 8	0·063380 0	0·134213 1	−0·002566 9	−0·090396 5	0·037334 2
0·69	0·043523 9	0·019958 1	−0·150840 8	0·033449 5	0·085293 7	0·006601 6	−0·051151 0	0·018271 2
0·70	0·027360 5	0·020036 5	−0·070149 1	−0·003743 5	0·022132 5	0·015326 3	−0·003056 5	−0·003401 9
0·71	0·007807 5	0·019926 2	0·026564 3	−0·047985 5	−0·053146 1	0·026023 1	0·054498 9	−0·029897 8
0·72	−0·013761 3	0·019521 2	0·131848 3	−0·095758 0	−0·134524 0	0·038259 8	0·117249 3	−0·059820 2
0·73	−0·033966 0	0·018183 3	0·229040 3	−0·136719 2	−0·207058 0	0·046415 9	0·170880 8	−0·084480 2
0·74	−0·044313 5	0·013695 1	0·276280 8	−0·142172 7	−0·231081 9	0·033181 0	0·175354 8	−0·079007 0
0·75	−0·041632 0	0·005850 5	0·257308 0	−0·104103 8	−0·193419 5	−0·003643 9	0·120324 0	−0·038639 5
0·76	−0·023716 5	−0·004396 2	0·159202 3	−0·021617 9	−0·087306 4	−0·057149 1	0·007101 4	0·030587 6
0·77	0·007309 2	−0·015019 8	−0·009870 0	0·091671 2	0·073980 3	−0·111700 2	−0·142759 1	0·110240 5
0·78	0·034076 5	−0·019547 9	−0·168223 9	0·168531 8	0·205438 4	−0·115697 1	−0·231466 7	0·131760 5
0·79	0·024648 0	−0·008797 7	−0·160389 9	0·097045 3	0·155972 9	0·003026 5	−0·101274 6	−0·005593 5
0·80	−0·009414 9	0·008785 0	−0·018708 2	−0·060248 9	−0·019966 8	0·153629 5	0·138242 8	−0·188792 4
0·81	−0·011144 5	0·009504 6	0·013736 4	−0·072112 7	−0·053035 8	0·115075 0	0·130701 6	−0·129213 1
0·82	0·018665 1	−0·004528 8	−0·079910 9	0·051375 8	0·062766 1	−0·067444 3	−0·091617 9	0·115269 8
0·83	0·024220 4	−0·008106 3	−0·088615 2	0·076403 4	0·078199 0	−0·122949 7	−0·145726 1	0·191358 0
0·84	−0·006812 9	0·004852 3	0·029424 1	−0·050710 1	−0·059764 2	0·018991 5	0·060128 2	0·007608 0
0·85	−0·024668 4	0·012878 3	0·073917 8	−0·125271 6	−0·122810 8	0·141556 8	0·200295 3	−0·152796 6
0·86	−0·006349 2	0·004121 6	−0·031495 4	−0·044870 7	−0·011337 6	0·108382 2	0·099066 6	−0·113799 6
0·87	0·017140 8	−0·010409 9	−0·140283 2	0·072390 0	0·123572 1	0·006983 8	−0·073062 2	0·008157 3
0·88	0·020737 1	−0·019544 7	−0·146065 9	0·121462 3	0·161222 1	−0·060636 4	−0·153889 0	0·081260 5
0·89	0·007395 4	−0·021466 4	−0·071333 8	0·101054 9	0·112345 7	−0·074504 2	−0·134042 9	0·084560 0
0·90	−0·007953 7	−0·019799 1	0·010172 3	0·060491 3	0·045959 4	−0·063286 8	−0·081996 4	0·059819 0
0·91	−0·021009 7	−0·016013 2	0·077120 5	0·015763 2	−0·016659 9	−0·037439 2	−0·020210 0	0·021341 4
0·92	−0·033320 9	−0·010258 7	0·137873 6	−0·036003 5	−0·081479 7	0·002062 8	0·054389 9	−0·030849 6
0·93	−0·040696 4	−0·005118 4	0·176723 2	−0·074234 5	−0·127545 1	0·031017 7	0·108354 0	−0·066385 7
0·94	−0·036077 4	−0·004032 8	0·164172 3	−0·068053 0	−0·120187 5	0·018441 0	0·093690 0	−0·045718 2
0·95	−0·023675 2	−0·005313 9	0·114888 5	−0·034122 2	−0·076523 8	−0·014314 7	0·038851 8	0·002839 1
0·96	−0·017264 5	−0·003862 8	0·085926 2	−0·025386 7	−0·058005 4	−0·014155 8	0·027122 6	0·008448 9

The matrices

The matrices $\mathbf{B}'$, $\mathbf{A}'$ and $\mathbf{H}_5'$ are presented, together with the elements on the diagonal of the matrix $\mathbf{CC}'$ which govern the accuracy with which the elements of the matrix $\boldsymbol{\mu}$ are determined. These latter are as follows:

α	β	γ	δ
19108·81666	270·47898	32·87454	3·11407
ε	ζ	η	θ
0·55526	0·30882	0·05548	0·00665

The writer is very grateful to Professors W. L. Bragg and N. F. Mott and to Dr W. H. Taylor for the provision of facilities and for much help and encouragement, to Dr P. B. Hirsch for much advice and many helpful discussions, to Dr M. V. Wilkes and the staff of the University Mathematical Laboratory, Cambridge, for the use of the computor EDSAC, and to the National Coal Board for the financial support for this work.

References

Brindley, G. W. & Méring, J. (1951). *Acta Cryst.* **4**, 441.

Compton, A. H. & Allison, S. K. (1935). *X-rays in Theory and Experiment.* London: Macmillan.

Diamond, R. (1956). Ph.D. Dissertation, University of Cambridge.

Diamond, R. (1957). *Acta Cryst.* **10**, 359.

Franklin, R. E. (1950). *Acta Cryst.* **3**, 107.

Franklin, R. E. (1951*a*). *Proc. Roy. Soc.* A, **209**, 196.

Franklin, R. E. (1951*b*). *Acta Cryst.* **4**, 253.

Hartree, D. R. (1952). *Numerical Analysis.* Oxford: Clarendon Press.

Hirsch, P. B. (1954). *Proc. Roy. Soc.* A, **226**, 143.

Hoerni, J. A. & Ibers, J. A. (1954). *Acta Cryst.* **7**, 744.

James, R. W. & Brindley, G. W. (1931). *Phil. Mag.* **12**, 81.

Keating, D. T. & Vineyard, G. H. (1956). *Acta Cryst.* **9**, 895.

McWeeny, R. (1951). *Acta Cryst.* **4**, 513.

Nelson, J. B. (1954). *Fuel, Lond.* **33**, 381.

Warren, B. E. (1941). *Phys. Rev.* **59**, 693.

Wilson, A. J. C. (1949). *Acta Cryst.* **2**, 245.