ACTA CHEMICA SCANDINAVICA 3 (1949) 560—568

The Accuracy of Inter-atomic Distances Obtained in Electron Diffraction Investigations of Molecular Structures

D. W. J. CRUICKSHANK and H. VIERVOLL

St. John's College, Cambridge, England,
and
Department of Physics, University of Oslo, Blindern — Oslo, Norway

A large and growing mass of data from several different experimental fields has been obtained about the dimensional details of molecular structures. To co-ordinate and to make the best possible use of this information, it is necessary to free the results of any systematic errors, to estimate the random errors, and to subject any metrical interpretations or comparisons to standard statistical tests of significance. The present paper describes how to treat the results of electron diffraction investigations; its approach to the problem of accuracy has had two sources. A similar theory of errors has been discussed in X-Ray Crystallography by Cox and Cruickshank[1], and Cruickshank[2]; these papers contain a full discussion of the details of the problem, and give the references to those authors who have contributed to its solution in X-Ray Crystallography. On the other hand Finbak, Hassel and co-workers in a number of papers[3] have applied Fourier Analysis in the sector method of electron diffraction, and Viervoll[4] has shown the use of the method of differencing experimental and theoretical $\sigma_m(r)$ curves in eliminating diffraction effects, searching for light atoms and guarding against anomalies.

In the electron diffraction of gaseous molecules interatomic distances may be obtained from the radial distribution functions $\sigma_m(r)$, $\frac{\sigma_m(r)}{r}$ or $D_m(r)$ which are related by $\sigma_m(r) = 4\pi(r)^2 D_m(r)$. The functions $\sigma(r)$ and $\frac{\sigma(r)}{r}$ defined by equation (17) of Viervoll's paper[4] are also used. For definiteness our discussion refers to $\sigma_m(r)$; similar remarks apply, however, to the other radial distribution functions. $\sigma_m(r)$ is obtained from the function $\frac{I_m(s)}{K}$ which is found experimentally, $I_m(s)$ being an intensity varying with $s = \frac{4\pi}{\lambda} \sin \Theta$. ($\lambda$ = wave length of electron beam; 2Θ = scattering angle.)

$$\sigma_m(r) = \frac{2r}{\pi} \int_0^\infty \frac{I_m(s)}{K} s^5 \sin sr \, ds \qquad (1)$$

POSITIONS OF THE MAXIMA

In numerical calculations as an alternative to using $\sigma_m(r)$ the positions of the maxima may be found from the zeros of the function which is the slope of $\sigma_m(r)$ (as after Booth[5] in X-Ray Crystallography). As slope functions are also needed in estimating the accuracy of the positions of the maxima, it may sometimes be convenient to calculate them directly *e. g.*

$$\frac{d}{dr}\left(\frac{\sigma_m(r)}{r}\right) = \frac{d}{dr}\int_0^\infty \frac{2}{\pi}\frac{I_m(s)}{K} s^5 \sin sr \, ds = \frac{2}{\pi}\int_0^\infty \frac{I_m(s)}{K} s^6 \cos sr \, ds \qquad (2)$$

The curvatures of the peaks at the maxima, which are also needed in accuracy estimations, may be found from the slopes of the slope functions.

CORRECTION FOR FINITE INTEGRATION

The finite range of integration (s_1 to s_2 instead of 0 to ∞) gives rise to systematic errors known as diffraction effects, after the name proposed by Bragg and West[6] for the similar effects occuring in the Fourier maps of X-Ray Crystallography. The diffraction effects take the form of ripples which create false peaks and shift the positions of the real maxima. By the use of 'normal curves' which represent the distribution function for a single inter-atomic distance with diffraction effects, Viervoll[4] has developed a method for detecting the spurious maxima and for correcting the positions of the displaced maxima.

When provisional interatomic distances have been obtained directly from the experimental radial distribution curve, the correction for finite intergration may be made in the following manner. By taking a complete set of 'normal curves' we may obtain a theoretical $\sigma_m(r)$ for a molecular model based on these provisional distances; this theoretical $\sigma_m(r)$ will be found to have its maxima slightly displaced from the assumed positions due to diffraction effects. Since the maxima of the experimental $\sigma_m(r)$ are obtained from a curve with diffraction effects, their correct positions, without diffraction, are obtained by reversing the displacements found in the theoretical $\sigma_m(r)$. Provided these shifts are small it is unneccesary to repeat the process with the theoretical $\sigma_m(r)$ based on the (once) corrected positions.

As appears from the discussion on p. 129 of Viervoll's paper [4], even with an infinite range of integration the inter-atomic distances are not exactly given by the maxima of $\sigma_m(r)$. The preceeding method automatically allows for this also. When the maxima corresponding to two different inter-atomic distances are not resolved, the distances must be estimated by »fitting» the normal curves to give the best agreement. A discussion of the conditions of validity of the similar procedure in X-Ray Crystallography, first proposed by Booth [7, 8] is given in § 10 of Cruickshank's paper [2].

RANDOM ERRORS

The distances now obtained are the final estimates, and are in error due to:

1) Inaccuracies in correction for finite integration.
2) Errors in $I_m(s)$.
3) Approximation errors in calculation.

There may also be an additional error due to an error in the linear scale of the intensity diagram. (In X-Ray Crystallography this corresponds to errors in the cell dimensions.) This error is usually negligible in comparison with those mentioned above, but if it is not, an estimate must be added statistically to the other errors.

On the assumption that the effect of the errors 1)—3) is that of many small random errors (as must for instance, occur in the determination of $I_m(s)$ over a large range of s), the errors in the inter-atomic distances have a normal (Gaussian) probability distribution. We shall now show how to estimate the standard deviation of these errors.

The problem cannot be treated in the same way as the corresponding one in X-Ray Crystallography, discussed in § 11 of Cruickshank's paper [2]. In that case estimates were sought of the errors in the electron density of a finite unit cell due to small errors in a set of discrete intensities; we are now seeking to estimate the errors at different points of a radial distribution function of infinite extent due to small errors in a continuous intensity function. Two points may be noticed at once. The errors in $\sigma_m(r)$ depend on r and they must tend to zero as r tends to infinity. The usual range of observation of s is sufficiently large for the root mean square difference of the theoretical and experimental intensities to be a reliable estimate of their r. m. s. error, and hence in a sufficiently large range of r the r. m. s. difference of the experimental and theoretical slopes of $\sigma_m(r)$ is a reliable estimate of the r. m. s. error of the slope in that range of r.

This suggests the following procedure. Let $\gamma(r)$ be the difference between the theoretical and experimental slopes of $\sigma_m(r)$. $\gamma^2(r)$ will be found to oscillate irregularly as in Fig. 1. Draw a smooth curve $p^2(r)$ (indicated by the dotted line) such that in any sufficiently large range of r its mean value is equal to the mean value of $\gamma^2(r)$ in that range. Then the estimate of the standard deviation of the slope of $\sigma_m(r)$ is $p(r)$.

One obvious difficulty is to decide on the length of the ranges of r to take in finding $p(r)$. They must be long enough to give reliable estimates of the r. m. s. error in that range, yet short enough to show variation with r. We suggest (arbitrarily) that ranges of length $\frac{4\pi}{s_2}$ should be used, so that $p^2(r)$ should be close to the mean value of $\gamma^2(r)$ in the range length $\frac{4\pi}{s_2}$ centred at this r, and that no value of $\gamma^2(r)$ should exceed three times the corresponding value of $p^2(r)$ (a position of very large $\gamma^2(r)$ either indicates a position of large error, or, more likely, an anomaly *e. g.* the effects of hydrogens atoms not considered in the theoretical $\sigma_m(r)$.) If there is any doubt as to the value of $p^2(r)$ it is wise to err on the safe side and to use an upper estimate.

By this consideration of the difference between the slopes of the theoretical and experimental $\sigma_m(r)$ we have estimated directly the effects of the experimental errors in $I_m(s)$, the approximation errors in calculation and any inaccuracies in the correction for finite integration.

The estimated standard deviation of an inter-atomic distance is $\frac{p(r)}{A}$ where A is the curvature of the peak of the $\sigma_m(r)$ diagram which gives the distance. When by the symmetry of the molecule several inter-atomic distances are related to one parameter, the standard deviation d of the weighted (*i. e.* inversely as the squares of the deviations) mean estimate of the parameter is given by

$$\frac{1}{d^2} = \frac{1}{d_1^2} + \cdots\cdots + \frac{1}{d_n^2}$$

where $d_1, \ldots d_n$ are the standard deviations of the various determinations of the parameter.

AN EXAMPLE OF THE ESTIMATION OF STANDARD DEVIATIONS

The procedure will be illustrated by the structure determination of CBr_4, Fig. 1. The intensity $I_m(s)$ was in this case measured from $s_1 = 5$ to $s_2 = 16$; the corresponding experimental $\sigma_m(r)$ curve being shown in the upper part of

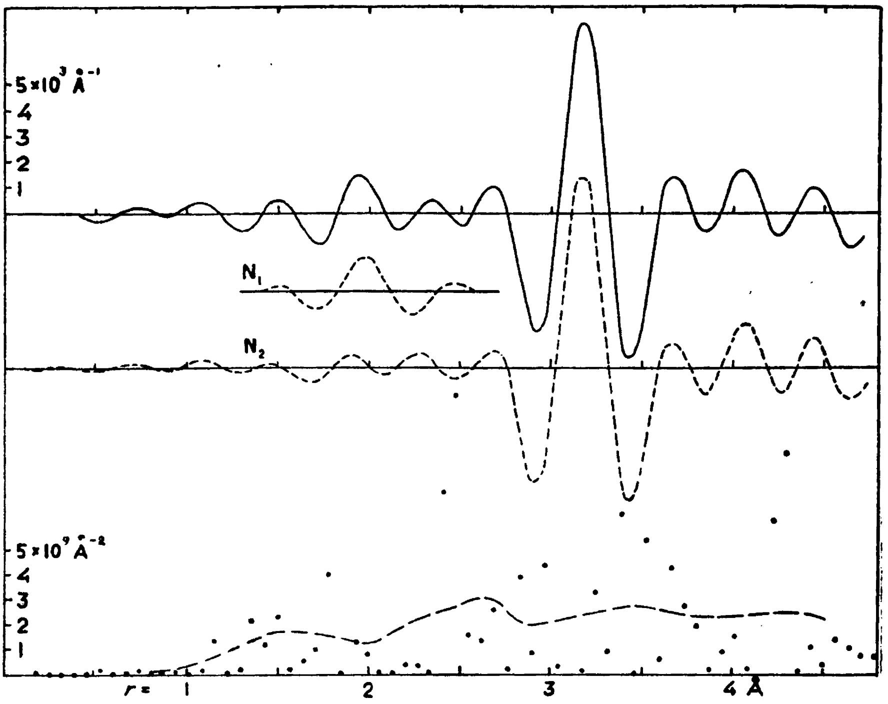

Fig. 1. Determination of standard deviation from the experimental $\sigma_m(r)$ curve of CBr_4. ——— *exp.* $\sigma_m(r)$, - - - *»normal curves»*, ● $\gamma^2(r)$, — — $p^2(r)$.

Fig. 1. The dominating peak at $r = 3.174$ Å gives the Br--Br distance. The 'normal curve' of this distance is shown by the dotted curve, N_2. By subtracting the function N_2 from the experimental $\sigma_m(r)$ function we obtain a difference function in which the diffraction ripples of the Br—Br distance have been removed. This difference function has its most pronounced maximum at $r = 1.960$ Å which gives the C—Br distance. The corresponding 'normal curve' is shown by the dotted curve N_1.

By subtracting the functions N_1 and N_2 from the experimental $\sigma_m(r)$ function, we obtain a difference curve the slope of which is $\gamma(r)$. The irregular oscillation of $\gamma^2(r)$ is plotted as dots in the lower part of Fig. 1. The curve $p^2(r)$ has been drawn as the mean value of $\gamma^2(r)$ in a range length approximate to $\frac{2\pi}{16}$.

From these curves we obtain the following estimation of the standard deviations of the two distances:

C—Br: $p^2(r) = 1.4 \times 10^9$ Å^{-2} $\quad d_1 = \dfrac{3.741}{204} = 0.0183$ Å

$A = 2.04 \times 10^4$ Å^{-2}

Br—Br: $p^2(r) = 2.3 \times 10^9$ Å^{-2} $\quad d'_2 = \dfrac{4.796}{734} = 0.0065$ Å

$A = 7.34 \times 10^4$ Å^{-2}

If we now assume the molecule to have tetrahedral symmetry, we obtain from the Br—Br distance the following value of the C—Br distance:

$$3.174 \times \sqrt{\tfrac{3}{8}} = 1.944 \text{ Å} \qquad d_2 = 0.0065 \times \sqrt{\tfrac{3}{8}} = 0.0040 \text{ Å}$$

The ratio $\left(\dfrac{d_1}{d_2}\right)^2 = 21$; by weighting the results by this factor we get the final estimate:

C—Br: 1.945 Å; $\quad d = 0.0039$ Å

TEMPERATURE FACTORS AND THE METHOD OF LEAST SQUARES

It is often necessary to multiply the theoretical intensity by a temperature factor to get agreement with the experimental data, and thus to make the heights and forms of the experimental and theoretical $\sigma_m(r)$ peaks the same. When this is done the corrections for diffraction should be estimated from normal curves which allow for the temperature factor; the same normal curves should be used to find $\gamma(r)$.

Similarly if an artificial temperature factor is applied to the experimental $I_m(s)$ to reduce the diffraction effects, it should also be applied to the 'normal curves'. It is sometimes found that the use of an artificial temperature factor not only reduces the diffraction effects but also the estimated random errors in the inter-atomic distance. To understand this we consider the method of least squares.

Let $I_{exp.}$ and I_{th} denote the experimental and theoretical intensities, and let $w(s)$ be the weight given to the intensity at any s. According to the method of least squares the best estimates of the inter-atomic parameters R_{ij} are those for which the function L is a minimum

$$L = \int_{s_1}^{s_2} \omega \, (I_{exp.} - I_{th})^2 \, ds \tag{3}$$

I_{th} being calculated for the parameters R_{ij}. When L is a minimum $\frac{\delta L}{\delta R_{ij}} = 0$. Hence

$$\int_{s_1}^{s_2} \omega \; (I_{exp.} - I_{th}) \frac{\mathrm{d}I_{th}}{\mathrm{d}R_{ij}} = 0 \tag{4}$$

Now

$$\frac{\mathrm{d}I_{th}}{\mathrm{d}R_{ij}} = \frac{mK \; (Z_i - F_i)(Z_j - F_j)}{s^4} \frac{\mathrm{d}}{\mathrm{d}R_{ij}} \frac{\sin sR_{ij}}{sR_{ij}} \tag{5}$$

m being a multiplicity; hence, dividing by mK^2

$$\int_{s_1}^{s_2} \frac{\omega \; (Z_i - F_i)(Z_j - F_j)}{Ks^4} \frac{\mathrm{d}}{\mathrm{d}R_{ij}} \left[\frac{\sin sR_{ij}}{sR_{ij}}\right] I_{exp.} \; \mathrm{d}s = \\ = \int_{s_1}^{s_2} \frac{\omega \; (Z_i - F_i)(Z_j - F_j)}{Ks^4} \frac{\mathrm{d}}{\mathrm{d}R_{ij}} \left[\frac{\sin sR_{ij}}{sR_{ij}}\right] I_{th} \; \mathrm{d}s \tag{6}$$

i. e. the slopes at R_{ij} of

$$W_{exp.}(r) = \frac{2}{\pi} \int_{s_1}^{s_2} \frac{\omega \; (Z_i - F_i)(Z_j - F_j)}{Ks^4} \frac{\sin sr}{sr} I_{exp.} \; \mathrm{d}s \tag{7}$$

and

$$W_{th}(r) = \frac{2}{\pi} \int_{s_1}^{s_2} \frac{\omega \; (Z_i - F_i)(Z_j - F_j)}{Ks^4} \frac{\sin sr}{sr} I_{th} \; \mathrm{d}s \tag{8}$$

are equal. We may regard $W(r)$ as the $D_m(r)$ obtained by using an artificial temperature factor $\alpha = \frac{\omega}{s^{10}} [(Z_i - F_i) \; (Z_j - F_j)]$. The function $\omega(s)$ must be determined from a study of the differences in $I_{exp.}$ and I_{th} and cannot be predicted entirely on theoretical grounds. By the very purpose of the least squares method α is the temperature factor which gives the least error in the determination of R_{ij}; it does this because it weights the intensity according to its reliability for different s. The use of other artificial temperature factors is quite valid, but they will lead to less accurate results, their accuracy depending on how closely they resemble α. In each case the errors will be those estimated by the methods described above. If the F's of the different atoms are not proportional several different $W(r)$'s will be needed to get the best estimate of all the parameters, though probably little accuracy will be lost by using the same a throughout.

We may notice that when nearly correct parameters have been obtained from $\sigma_m(r)$ it is possible to express approximately $I_{th}(s)$ as a linear function of the parameters. By doing this the usual 'normal equations' of the least squares method can be found and solved. When the maxima are well resolved the least squares method leads to the same results as those given by $W_{exp.}(r)$, corrected for diffraction effects. But when some of the maxima are not resolved it may be preferable to use least squares. (A similar connection in X-Ray Crystallography between the Fourier series and the least squares method has been pointed out by Cochran[9].)

SIGNIFICANCE TESTS

For valid comparisons to be made of experimental determinations of inter-atomic distances in molecular structures, or between experimental and theoretical results it is necessary to use statistical assessments of significance based on the estimated errors. Details of the application of significance tests to the comparison of bond-lengths will be found in Cruickshank's paper[2]; a general outline of the problem is given in § 3, and an example in § 15, where the complications which occur when the errors in two bond-lengths are dependent, are also discussed.

SUMMARY

Methods of correcting the systematic errors of finite integration in the radial distribution functions of electron diffraction are discussed. A method is given for estimating the standard deviations in inter-atomic distances, due to random errors caused by experimental errors, approximation errors in calculation, and any inaccuracies in the correction for finite integration. As an example this method is applied to the CBr_4 structure. The effect of artificial temperature factors on accuracy is discussed by considering the method of least squares.

REFERENCES

1. Cox, E. G., and Cruickshank, D. W. J. *Acta Cryst.* 1 (1948) 92.
2. Cruickshank, D. W. J. *Acta Cryst.* 2 (1949) 65.
3. Hassel, O., and Viervoll, H. *Acta Chem. Scand.* 1 (1947) 149.
4. Viervoll, H. *Acta Chem. Scand.* 1 (1947) 120.
5. Booth, A. D. *Trans. Faraday Soc.* 42 (1946) 444.

6. Bragg, W. L., and West, J. *Phil.Mag.* **10** (1930) 283.
7. Booth, A. D. *Proc. Roy. Soc.* **A 188** (1946) 77.
8. Booth, A. D. *Nature* **156** (1945) 51.
9. Cochran, W. *Acta Cryst.* **1** (1948) 138.
10. Brownlee. *Industrial experimentation.* London (1947).

Received April 11, 1949.