Errors in Electron Density Maps and the Use of Integrated Densities in the Study of Charge Density Distributions*

By Philip Coppens and Walter C. Hamilton

Chemistry Department, Brookhaven National Laboratory, Upton, New York 11973, *U.S.A.*

Expected mean-square errors in the experimental electron density at a point in a crystal are discussed and compared with errors in a mean electron density or electron count obtained by integration over a parallelepiped in the crystal. That the mean density is more dependent on the low-order reflections emphasizes the importance of accurate low-order data for obtaining precise electron counts. The mean density is rather insensitive to changes in the temperature factors. It is therefore especially attractive for the calculation of difference maps where errors in the temperature parameters used in calculating F_c may introduce false detail. Numerical examples for a one-dimensional model system are given.

Introduction

The recent interest in deviations from spherical symmetry of the atoms in a molecule due to bonding and the resulting controversy between so-called 'electron seers' and 'non-electron seers' have indicated the need for a precise assessment of the significance of residual features in electron density maps.

The average expected mean square error in the electron density in the unit cell has been given by Cruickshank (1949) as

$$\sigma^2(\varrho) = \frac{1}{V^2} \sum_{\mathbf{h}} \sigma^2(F_{\mathbf{h}}) \tag{1a}$$

where $\sigma^2(F_{\mathbf{h}})$ is the variance of the observation $F_{\mathbf{h}}$. The volume of the unit cell is V. For testing the significance of a model for the structure by calculating a difference density $\Delta\varrho = \varrho_{\text{obs}} - \varrho_{\text{calc}}$ we have similarly:

$$\sigma^2(\Delta\varrho) = \frac{1}{V^2} \sum_{\mathbf{h}} \sigma^2(F_{\mathbf{h}}) . \tag{1b}$$

The summations extend over all reciprocal vectors $\mathbf{h}$, and the terms in the series beyond the data termination must have $\sigma^2(F_{\mathbf{h}})$ replaced by the expected value of $F_{\mathbf{h}}^2$.† In many applications in the crystallographic literature, only terms for the observed data have been included in equations (1); this is of course incorrect in that it leads to the false notion that the greater the number of data points the greater is the error in the electron density.

A more useful error function describes the error at specific points in the unit cell (Cruickshank, 1965). Such a function can be used to estimate the significance of features in difference maps attributed to overlap populations and lone-pair electrons.

More interesting than the electron density at a certain point is the density integrated over a small volume. How many electrons, for example, are contained in a peak observed in the center of a C–C bond in a difference map? A formula for such an electron count and a formula for its estimated error are given below. This electron count may be divided by the volume of integration to obtain a mean electron density and it will be shown that the mean or average electron density can be evaluated with about the same accuracy as the electron density itself.

The error in the electron density as a function of position

The electron density in the general case may be evaluated as

$$\varrho(\mathbf{r}) = \frac{2}{V} \sum_{\mathbf{h} \geq 0} |F_{\mathbf{h}}| \cos 2\pi(\mathbf{h} \cdot \mathbf{r} - \varphi_{\mathbf{h}})^* \tag{2}$$

where $\mathbf{r}$ indicates a vector in the unit cell and $\varphi_{\mathbf{h}}$ is the phase of the structure factor $F_{\mathbf{h}}$, whose real and imaginary parts will be denoted $A_{\mathbf{h}}$ and $B_{\mathbf{h}}$. If we define

$$\Delta\varrho = \varrho_{\text{obs}} - \varrho_{\text{true}} \tag{3}$$

and

$$\Delta|F| = |F|_{\text{obs}} - |F|_{\text{true}} \tag{3}$$

we may write

$$\Delta\varrho(\mathbf{r}) = \frac{2}{V} \sum_{\mathbf{h} \geq 0} \Delta|F_{\mathbf{h}}| \cos 2\pi(\mathbf{h} \cdot \mathbf{r} - \varphi_{\mathbf{h}}) , \tag{4}$$

where we have assumed that the error in phase angle is negligible (see Cruickshank, 1949). The mean square error $\sigma^2(\varrho)$ in the electron density is defined as the expected value of $(\Delta\varrho)^2$; thus

$$\sigma^2(\varrho) \equiv E\{(\Delta\varrho)^2\} = \frac{4}{V^2} \sum_{\mathbf{h} \geq 0} \sum_{\mathbf{h}' \geq 0} E\{\Delta|F_{\mathbf{h}}|\Delta|F_{\mathbf{h}'}|\}$$
$$\times \cos 2\pi(\mathbf{h} \cdot \mathbf{r} - \varphi_{\mathbf{h}}) \cos 2\pi(\mathbf{h}' \cdot \mathbf{r} - \varphi'_{\mathbf{h}}) . \tag{5}$$

---

* Research performed under the auspices of the U.S. Atomic Energy Commission.

† Alternatively the calculated values of $F_{\mathbf{h}}$ may be used for these terms in the Fourier series. This requires an estimate of $\sigma(F_{\text{calc}})$ for application of (1b).

* By $\mathbf{h} \geq 0$ we mean that the summation is taken over one half of reciprocal space, that is, only one member of each Friedel pair is included and $F(0,0,0)$ is given weight $\frac{1}{2}$.

If the reflections in a whole hemisphere of reciprocal space have been collected, we may assume a probability distribution of the errors such that ΔF_h and $\Delta F_{h'}$ are statistically independent:

$$E\{\Delta F_h \Delta F_{h'}\} = 0 \quad \text{if } h \neq h' \tag{6}$$

$$E\{\Delta F_h \Delta F_{h'}\} = \sigma^2(F_h) \quad \text{if } h = h'$$

and we obtain

$$\sigma^2\{\varrho(\mathbf{r})\} = \frac{4}{V^2} \sum_{h \geq 0} \sigma^2(F_h) \cos^2 2\pi(\mathbf{h} \cdot \mathbf{r} - \varphi_h) . \tag{7a}$$

For computational convenience this may be rewritten as:

$$\sigma^2\{\varrho(\mathbf{r})\} = \frac{4}{V^2} \sum_{h \geq 0} \sigma^2(F_h) \left(\frac{A_h^2}{|F_h|^2} \cos^2 2\pi h \cdot \mathbf{r} \right.$$
$$\left. + \frac{B_h^2}{|F_h|^2} \sin^2 2\pi h \cdot \mathbf{r} + \frac{2A_h B_h}{|F_h|^2} \cos 2\pi h \cdot \mathbf{r} \sin 2\pi h \cdot \mathbf{r} \right) \tag{7b}$$

or

$$\frac{4}{V^2} \sum_{h \geq 0} \left\{ \sigma(F_h) \frac{A_h \cos 2\pi h \cdot \mathbf{r} + B_h \sin 2\pi h \cdot \mathbf{r}}{|F_h|} \right\}^2 , * \tag{7c}$$

which in the centrosymmetric case reduces to:

$$\sigma^2\{\varrho(\mathbf{r})\} = \frac{4}{V^2} \sum_{h \geq 0} \sigma^2(F_h) \cos^2 2\pi h \cdot \mathbf{r} . \tag{7d}$$

The effect of symmetry on the error in ϱ

The presence of symmetry elements other than a center of symmetry introduces well known relationships between structure factors. The Fourier summation can then be obtained with less than a hemisphere of data. Cruickshank & Rollett (1953) pointed out that this results in larger errors at special positions (which are not necessarily positions of lower multiplicity). In space group Pm, for example, $F_{hkl} = F_{h\bar{k}l}$, and therefore, if only one reflection of this pair has been measured $\Delta(F_{hkl}) = \Delta(F_{h\bar{k}l})$. This correlation gives rise to a cross term in expression (5), which averages to zero at all but the special positions. It is worthwhile to note that this increase in the standard deviation at the special positions is eliminated when a full hemisphere of data is collected.

The integrated electron density

A pertinent expression for the density averaged over a cubic volume has been given by Weiss (1966). For a cube centered at the point (x,y,z) with edge δ the mean electron density is

$$\bar{\rho}(x,y,z,\delta) = \frac{1}{V} \sum_h \sum_k \sum_l F_{hkl} \exp(-2\pi i \mathbf{h} \cdot \mathbf{r})$$
$$\times \frac{\sin \pi h\delta/a}{\pi h\delta/a} \frac{\sin \pi k\delta/b}{\pi k\delta/b} \frac{\sin \pi l\delta/c}{\pi l\delta/c} . \tag{8}$$

* Dr E. N. Maslen, University of Western Australia, has pointed out that this expression excludes errors in the scale factor k, which add a term $k^{-2}\sigma^2(k)\varrho(\mathbf{r})^2$.

This expression may be generalized as follows: suppose we are interested in the average density in a parallelepiped centered at the point (x,y,z) with edges of length $\delta_1, \delta_2, \delta_3$ in Å parallel to the unit vectors $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$, such that $\mathbf{u}_i = a_{i1}\mathbf{a} + a_{i2}\mathbf{b} + a_{i3}\mathbf{c}$, where the vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}$ are the crystal axes of length a, b, and c respectively.

Triple integration of the usual expression for the Fourier summation leads to the following result for the electron count in the parallelepiped:

$$\varrho_{\text{int}}(x,y,z,\delta_1,\delta_2,\delta_3)$$
$$= \frac{\Delta V}{V} \sum_h \sum_k \sum_l F_{hkl} \exp(-2\pi i \mathbf{h} \cdot \mathbf{r}) S(\alpha_1) S(\alpha_2) S(\alpha_3) \tag{9a}$$

with

$$S(\alpha_i) = \frac{\sin \alpha_i}{\alpha_i}, \quad \alpha_i = (ha_{i1} + ka_{i2} + la_{i3})\pi\delta_i \tag{9b}$$

and ΔV is the volume of the parallelepiped. Dividing by ΔV we find for the mean density in the parallelepiped

$$\bar{\rho}(x,y,z,\delta_1,\delta_2,\delta_3)$$
$$= \frac{1}{V} \sum_h \sum_k \sum_l F_{hkl} \exp(-2\pi i \mathbf{h} \cdot \mathbf{r}) S_1 S_2 S_3 . * \tag{10}$$

Thus the contribution of each term in the conventional summation for the electron density is multiplied by $S_1 S_2 S_3$. The function $S_i = \sin \alpha_i / \alpha_i$ is plotted in Fig. 1.

* Kurki-Suonio (1959) has considered the structure amplitude $f_T(h)$ of a volume T centered at the origin. Expression (10) equals $f_T(0)$ in the case where T is a parallelepiped at $\mathbf{r} = 0$. It follows from the expression derived by Kurki-Suonio that the mean density in an ellipsoid centered at (x,y,z) with principal axes $\delta_i u_i$ is:

$$\bar{\rho} = 3/V \sum_h \sum_k \sum_l F_{hkl} \exp(-2\pi i \mathbf{h} \cdot \mathbf{r}) (\sin x - x \cos x)/x^3$$
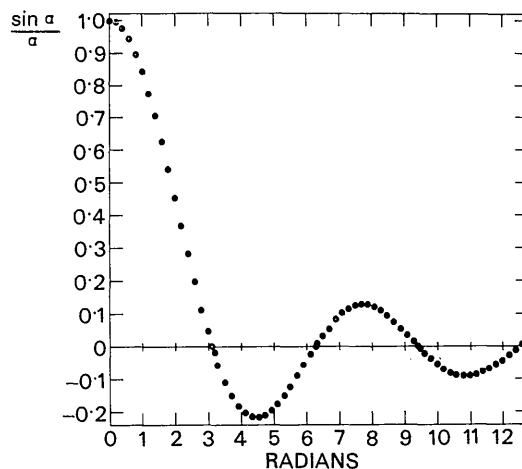
in which

$$x = 2\sqrt{\sum_i \alpha_i^2} .$$

Fig. 1. The function $\sin \alpha / \alpha$.

For $\alpha_t = \pi$, which corresponds to $(ha_{i1} + ka_{i2} + la_{i3})\delta_t = 1$, S_i is equal to zero. Physically this means that the period of the wave hkl in the direction $\mathbf{u}_t$ is the same as the dimension of the parallelepiped in this direction.
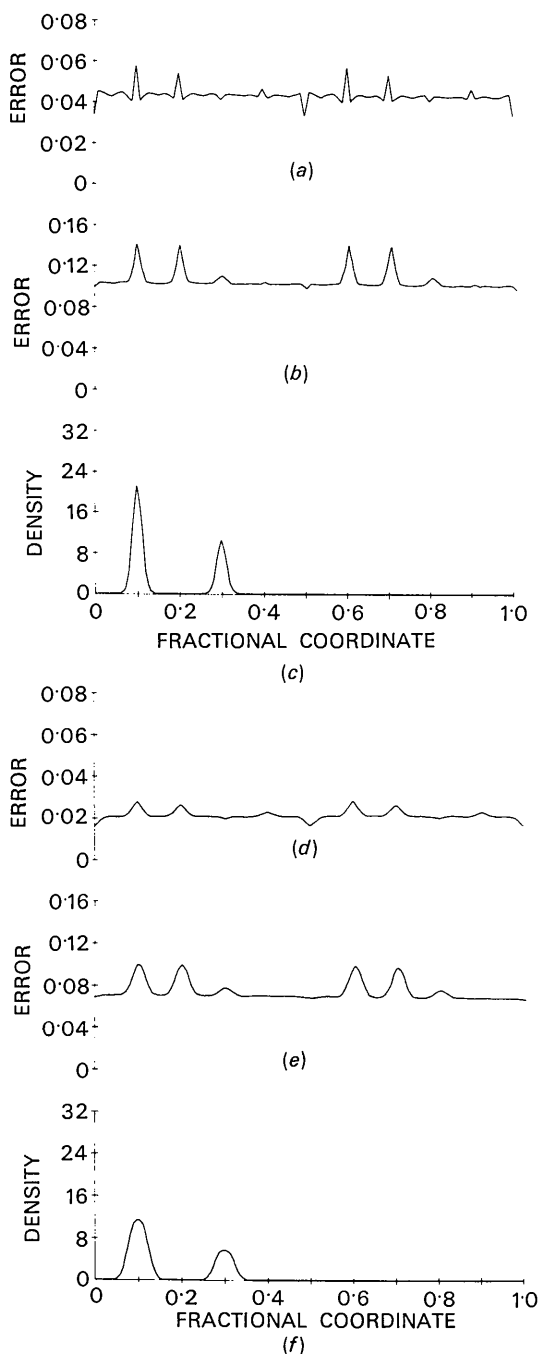

Fig. 2. Electron density, mean density and associated errors in a one-dimensional unit cell of 10 Å length containing two atoms ($B=1\cdot0$ Å^2). (a) error curve for $c_1=0$, $c_2=0\cdot05$. (b) error curve for $c_1=0\cdot03$, $c_2=0\cdot05$. (c) Electron density. (d), (e), (f) Corresponding curves for the density averaged over boxes of 0·5 Å length.

The contribution of this wave to the integrated density is therefore zero. It can be shown easily that for a parallelepiped with maximum dimension δ_{max}, α will always be smaller than π if $2\delta_{max}$ is smaller than the wavelength of the radiation employed.

The shape of the function S indicates that generally less weight is given to higher order reflections in the summation (10), although the weight increases again for high order reflections for which α exceeds π. This occurs when a large parallelepiped is chosen or a very short wavelength is used.

The error in the integrated electron density

The error in the function (10) can be derived by the methods discussed in the second paragraph. With the same assumptions we obtain, analogous to (7),

$$\sigma^2\{\bar{\rho}(\mathbf{r})\} = \frac{4}{V^2} \sum_{\mathbf{h} \geq 0} \sigma^2(F_\mathbf{h}) S_1^2 S_2^2 S_3^2$$

$$\times \left[\frac{A_\mathbf{h}\cos 2\pi\mathbf{h}\cdot\mathbf{r} + B_\mathbf{h}\sin 2\pi\mathbf{h}\cdot\mathbf{r}}{|F_\mathbf{h}|} \right]^2, \quad (11)$$

which reduces to

$$\sigma^2\{\bar{\rho}(\mathbf{r})\} = \sum_{\mathbf{h} \geq 0} \sigma^2(F_\mathbf{h}) S_1^2 S_2^2 S_3^2 \cos^2 2\pi\mathbf{h}\cdot\mathbf{r}$$

for the centrosymmetric case.

For $\delta_1 = \delta_2 = \delta_3 = 0$, $S_1^2 S_2^2 S_3^2$ equals 1 and (11) reduces to (7). For $\delta_i \neq 0$, however, each term in the summation is multiplied by the corresponding value of $S_1^2 S_2^2 S_3^2$, which decreases rapidly with increasing δ. Thus, the error in the mean electron density will be less dependent on the accuracy of the high order reflections. In addition, since every term in (10) is multiplied by $S_1 S_2 S_3$ series termination errors in $\bar{\rho}$ will be relatively small. This is of considerable importance.

The error in the observations

The application of the error formulas (11) and (7) hinges on the estimate of the errors $\sigma^2(F)$ in the observations. For diffractometer data, it is often found that expressions like

$$\sigma^2(I-B) = I + B + c^2(I-B)^2 \quad (12)$$

are suitable. (See e.g. Coppens & Schmidt, 1965.) I is the measured intensity and B is the background. The term $I+B$ is due to the counting statistics, while the second term is an empirical one which includes most other errors. An additional term representing uncertainty in absorption and extinction corrections may also be appropriate. The constant c may be estimated from measurements of symmetry related reflections or by analysis-of-variance at the conclusion of the refinement. It is typically found to be in the range 0·02–0·10.

$\sigma^2(F)$ can now be derived from $\sigma^2(I-B)$. As stated earlier, terms in the series beyond termination should be taken into account. Possible extinction should be

corrected either by calculation or by collecting data on crystals of different sizes. The assumption that $\sigma(F_h)$ is statistically independent of $\sigma(F_{h'})$ used in the derivation of (7) and (11) is not valid if extinction or other systematic errors are not properly allowed for.

Numerical calculations

Calculations were performed on a one-dimensional model system with point atoms having non-zero temperature factors. The decrease of the atomic form factor with $(\sin\theta/\lambda)^2$ was thus approximated by an exponential.

Structure factors were calculated and used in calculating the electron density ϱ, the mean density $\bar{p}$ and the corresponding errors. The error in F was obtained from:

$$\sigma(F) = (c_1^2 F^2 + c_2^2)^{1/2} . \qquad (13)$$

This expression is analogous to (12) from which it may be derived by replacing $(I+B)$ and $(I-B)$ by F^2 and noting that $2F\sigma(F) \simeq \sigma(F^2)$. The following examples illustrate some of the results. The unit cell has a length of 10 Å while the maximum h index is 39 (complete 'sphere' for $\lambda = 0.5$ Å). However, the examples with large temperature factors have very small high order reflections; they are therefore representative for experiments in which more common X-ray wavelengths > 0.5 Å are used. It should be noted that inclusion of 39 terms for the 10 Å cell makes series termination effects very small in all examples discussed below.

1. Influence of the distribution of the errors in F

Two atoms with 6 and 3 electrons respectively are placed in the unit cell 2 Å apart.

Case A: $c_1 = 0.00$, $c_2 = 0.05$ (constant errors); $B = 1.0$ Å²; box size for integration 0.5 Å

Case B: as A but with $c_1 = 0.03$ and $c_2 = 0.05$.

The results shown in Fig. 2 indicate that the errors in both ϱ and $\bar{p}$ are peaked at certain positions in the unit cell, especially at the atomic centers and at the midpoint of the two atoms. The peaks are about 40% above the average error, indicating that the errors are fairly constant. In more complicated three-dimensional structures with more data the errors can be expected to vary even less. Therefore average formulae like $(1a)$ seem to be adequate for routine structure work.

At first sight it seems curious that the periodicity of π in the error function is equal to half the periodicity in ϱ. The reason becomes apparent on noting that the squares of the trigonometric function appear in equations (7) and (11).

Case A. For constant values of $\sigma(F)$, $\bar{p}$ is considerably more accurate than ϱ. As was stated above, the relative contribution of the strong low order terms is larger in $\bar{p}$. If the constant error assignment is correct, the strong reflections are relatively more accurate, and accordingly $\bar{p}$ can be evaluated more precisely than

can ϱ. This once again illustrates the need for accurate, extinction-free, low order data when electron distributions are to be studied.

Case B. The error in $\bar{p}$ is generally smaller than the error in ϱ, but the 'signal to noise ratio' $\bar{p}/\sigma(\bar{p})$ is at the peak positions smaller than the corresponding value for the conventional unintegrated map. This was found to be true for values of c_1 ranging from 0.01 to 0.1 and for the somewhat more complicated model of example 4.

2. Influence of the temperature factor

The calculations described in the preceding paragraph were repeated with $B = 1.1$ Å² and $B = 3.0$ Å² for both atoms (Fig. 3). At the atomic positions $d\bar{p}/dB$ is much smaller than $d\varrho/dB$. The peak heights for the heaviest atom are, for example, 21.2, 20.2 and 12.3 e.Å⁻¹ for $B = 1.0$, 1.1 and 3.0 Å² respectively. The cor-
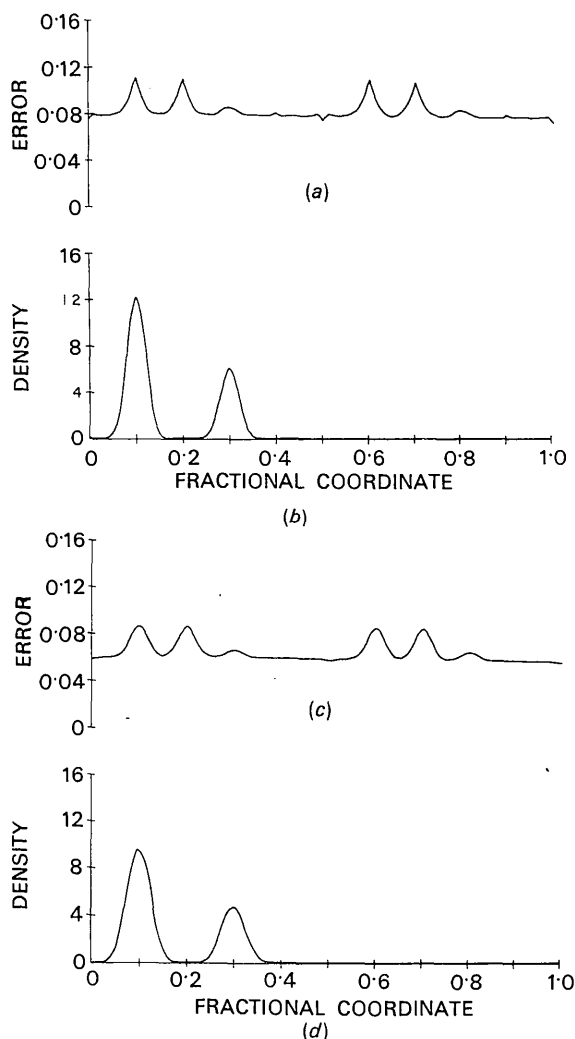

Fig. 3. As in Fig. 2(b), (c), (e), (f) but with $B = 3.0$ Å⁻². (a) Error in the electron density. (b) Electron density. (c) Error in the mean density. (d) Mean density.

responding values for $\bar{p}$ are 11·7, 11·6 and 9·6 e.Å^{-1}. The integrated electron count differs from $\bar{p}$ by a factor equal to the size of the box, and is therefore equally
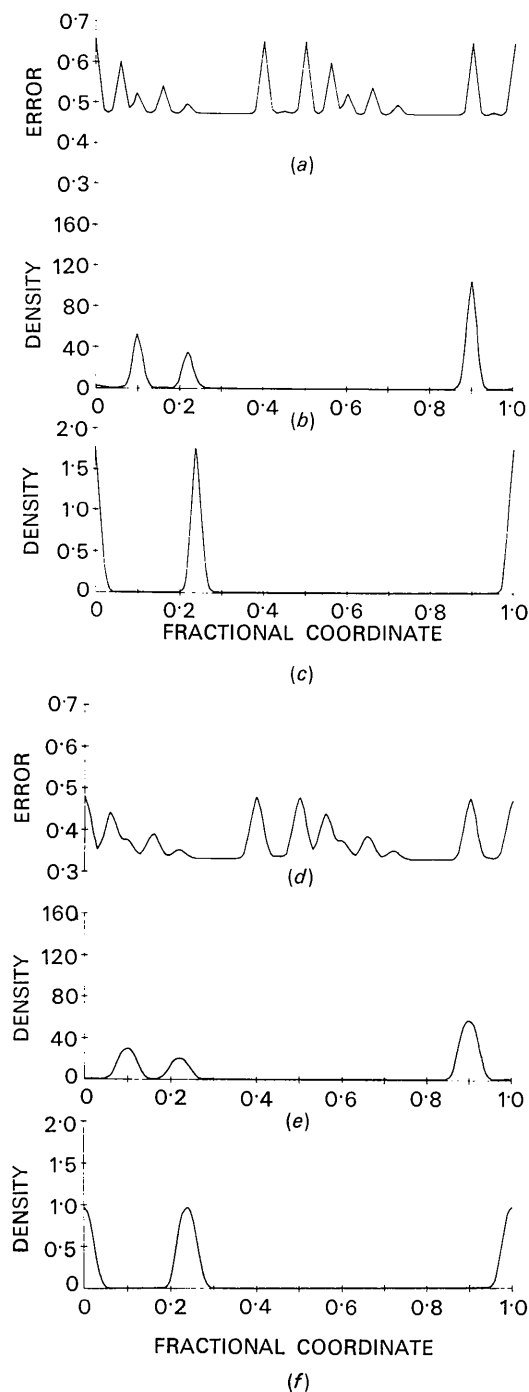

Fig. 4. One-dimensional unit cell with 3 heavy atoms and 2 light atoms ($B=1\cdot0$ Å^2), $c_1=0\cdot02$, $c_2=0\cdot05$. (a) Error in electron density and difference density. (b) Electron density. (c) Difference electron density with light atoms only. (d), (e), (f) Corresponding curves for the density averaged over boxes of 0·5 Å length.

insensitive to variations in B. This is important in difference maps in general, and in particular in combined X-ray and neutron difference maps. In the latter case the electron density calculated with the neutron parameters may be subtracted from the observed X-ray density (Coppens, 1967). Such an 'F_X-F_N synthesis' is subject to both errors in the X-ray structure factors and errors in the neutron parameters used in the calculation of F_N. Obviously, errors in the neutron temperature factors are less effective in the integrated difference map than in the conventional difference map.

3. Variation in box size

The calculations described under 1 and 2 were repeated for a box of 0·2 Å length. As expected the results are intermediate between those obtained without integration and those obtained by integration over 0·5 Å.

4. Errors in a difference map

In these calculations two light atoms with 0·5 electron are placed at $x=0\cdot0$ and $x=0\cdot24$. $B=1\cdot0$ Å^2. Heavier atoms are at 0·90 (30 electrons); 0·10 (15 electrons) and 0·22 (10 electrons). Both an electron density map and a difference density map are reproduced in Fig. 4, $c_1=0\cdot02$, $c_2=0\cdot05$. As indicated by (1b) the errors in a difference density are the same as the errors in ϱ.* There is no definite relationship between the amplitudes ΔF and their standard deviations $\sigma(F)$, as the latter are derived from F rather than from ΔF. The error curve is therefore peaked at positions which are not related to the position of the peaks in the difference map.

Conclusion

The expressions for the mean electron density (10) and its error (11) can be readily evaluated. It is recommended that they replace the conventional summation whenever the interest is in the quantitative electron count. A modified version of the A. Zalkin Fourier program FØRDAP based on (10) has been written and is presently in use in our laboratory.

References

COPPENS, P. (1967). Science. **158**, 1577.
COPPENS, P. & SCHMIDT, G. M. J. (1965). Acta Cryst. **18**, 654.
CRUICKSHANK, D. W. J. (1949). Acta Cryst. **2**, 65.
CRUICKSHANK, D. W. J. (1965). In Computing Methods in Crystallography, p. 107. Edited by J. S. ROLLETT. Oxford: Pergamon Press.
CRUICKSHANK, D. W. J. & ROLLETT, J. S. (1953). Acta Cryst. **6**, 705.
KURKI-SUONIO, K. (1959). Ann. Acad. Sci. Fennicae, Series A**31**, 1.
WEISS, R. J. (1966). X-ray Determination of Electron Distributions, p. 77. Amsterdam: North-Holland Publishing Co.

---

* Not considering errors in the model used to calculate the F_c's.