Fig. 1. Thianthrene bond lengths in Ångström units.

in the crystal. The closest approach between molecules occurs between two carbon atoms separated by 3·61 Å. The shortest distance between sulfur atoms in adjacent molecules is 3·77 Å; the shortest intermolecular S–C contact is 3·83 Å.

A comparison of additional results obtained in this work with those of Lynton & Cox is given in Table 4. In general the results of the two determinations agree very well; the most significant difference between the two concerns the C–C bond length. Lynton & Cox report apparently significant variations among C–C bonds. No

significant variations among these bonds are indicated by our results. Pending further work, this aspect of the molecular structure of thianthrene must be considered unresolved. A schematic diagram of thianthrene, giving bond lengths determined in this investigation, is shown in Fig. 1.

### Bibliography

ABRAHAMS, S. C. (1955). *Acta Cryst.* 8, 661.
CRUICKSHANK, D. W. J. (1949). *Acta Cryst.* 2, 65.
LYNTON, H. & COX, E. G. (1956). *J. Chem. Soc.* p. 488.
ROWE, I. & POST, B. (1956). *Acta Cryst.* 9, 827.
WOOD, R. G. & CRACKSTON, J. E. (1941). *Phil. Mag.* 31, 62.

## Scale-factor adjustment for optimum comparison of observed and calculated structure factors.*

By EPHRAIM SEGERMAN,† *Polytechnic Institute of Brooklyn, 333 Jay Street, Brooklyn 1, N. Y., U.S.A.*

In most of the structures reported to date, the observed structure factors have been scaled to the calculated ones by using the ratio of the sums as a scaling factor. This procedure is justified only if the differences between the observed and calculated structure factors are random over a large body of data. With the accuracy and sensitivity of modern techniques for collecting data, and assuming a 'correct' structure has been proposed, this is usually the case. Thus a more accurate method for deriving the scale factor, as is proposed here, will be of little general value. Nevertheless, if a structure is based on only a few reflections and the discrepancies are large, it is quite possible to give a marked improvement in the reported accuracy of a structure determination by adjusting this scale factor. The derived atomic coordinates are unaffected by this adjustment; but absolute electron-density measurements could be improved.

The optimum scale factor can be directly derived by including the scale factor as a variable in the expression for the residual being used as a measure of the structure determination accuracy, and setting the partial derivative of the residual with respect to the scale factor equal to zero.

In the following examples we shall define $x$ as the scale factor, $k$ the index $(hkl)$, $F_k$ cal. and $F_k$ obs. given

calculated and observed structure factors, and $R$ the residual. In the case of the general least-squares approach (Hughes, 1941; Booth, 1947), where

$$R_1 = \sum_k W_k (|F_k \text{ cal.}| - |xF_k \text{ obs.}|)^2$$

the scale factor for minimum $R$ becomes

$$x_m = \frac{\sum_k W_k |F_k \text{ obs.}||F_k \text{ cal.}|}{\sum_k W_k |F_k \text{ obs.}|^2} .$$

In the case of the least-squares method used by Shoemaker *et al.* (1950), where

$$R_2 = \sum_k W_k^{\frac{1}{2}} (|F_k \text{ cal.}|^2 - |xF_k \text{ obs.}|^2)^2 ,$$

the scale factor for minimum $R_2$ becomes

$$x_m = \left[ \frac{\sum_k W_k^{\frac{1}{2}} |F_k \text{ cal.}|^2 |F_k \text{ obs.}|^2}{\sum_k W_k^{\frac{1}{2}} |F_k \text{ obs.}|^4} \right]^{\frac{1}{2}} .$$

In the case of the iterative Fourier technique, where the residual usually employed is the 'reliability index', which for the moment may be defined as

$$R_3 = \frac{\sum_k \left| |F_k \text{ cal.}| - |xF_k \text{ obs.}| \right|}{\sum |F_k \text{ cal.}|} ,$$

the calculation of the optimum scale factor is not as straightforward as in the above examples; it follows:

Let

$$y = R_3 \sum_k |F_k \, \text{cal.}|, \quad i = k \quad \text{when} \quad |F_k \, \text{cal.}| > x|F_k \, \text{obs.}|,$$

and

$$j = k \quad \text{when} \quad |F_k \, \text{cal.}| < x|F_k \, \text{obs.}|.$$

Then

$$y = \sum_k \left||F_k \, \text{cal.}| - |xF_k \, \text{obs.}|\right|$$

$$= \sum_i |F_i \, \text{cal.}| - \sum_j |F_j \, \text{cal.}| + x\left(\sum_j |F_j \, \text{obs.}| - \sum_i |F_i \, \text{obs.}|\right)$$

and

$$\partial y/\partial x = \sum_j |F_j \, \text{obs.}| - \sum_i |F_i \, \text{obs.}| . \qquad (a)$$

Since

$$\sum_k |F_k \, \text{obs.}| = \sum_j |F_j \, \text{obs.}| + \sum_i |F_i \, \text{obs.}|,$$

then

$$\sum_j |F_j \, \text{obs.}| = \tfrac{1}{2} \sum_k |F_k \, \text{obs.}| + \tfrac{1}{2} (\partial y/\partial x). \qquad (b)$$

As $x$ increases, the $k$'s, one by one, change from being $i$'s to $j$'s. From equation $(a)$, in the range of $x$'s between each change $\partial y/\partial x$ is a constant, and at each change $\partial y/\partial x$ discontinuously increases, thus forming on the whole an increasing step function. The function $y$ is then a polygon whose minimum in general will be at a corner, i.e., at the value of $x$ at which the discontinuity in $\partial y/\partial x$ crosses $\partial y/\partial x = 0$. These discontinuities occur when $|F_k \, \text{cal.}| = x|F_k \, \text{obs.}|$, and so the desired $x$ is the ratio of a particular pair of observed and calculated $F$'s: $x_m = |F_m \, \text{cal.}| \div |F_m \, \text{obs.}|$. Equation $(b)$ suggests the method for finding this pair. If we have $\sum_k |F_k \, \text{obs.}|$ calculated and we keep adding terms to $\sum_j |F_j \, \text{obs.}|$, the first term added which brings the latter sum to greater than half the former one is the desired term. Thus the procedure is as follows: List $|F_k \, \text{obs.}|$'s, $|F_k \, \text{cal.}|$'s, and their ratios. Order the terms according to increasing ratios. Sum the $|F_k \, \text{obs.}|$'s. Start summing them again and stop when half the total sum is exceeded. The listed ratio of the last term added is the required scale factor $x_m$ to be used for the whole set.

When the usual scaling method is employed it is immaterial whether the denominator of $R_3$ is defined as $\sum_k |F_k \, \text{cal.}|$ or $\sum_k |xF_k \, \text{obs.}|$. In the method proposed here these values are different, so the alternative should be discussed:

If we define

$$R_3' = \frac{\sum_k \left||F_k \, \text{cal.}| - x|F_k \, \text{obs.}|\right|}{\sum_k x|F_k \, \text{obs.}|},$$

the solution is quite similar: $y'$ becomes $R_3' \sum_k |F_k \, \text{obs.}|$ so

$$y' = \frac{\sum_k \left||F_k \, \text{cal.}| - |xF_k \, \text{obs.}|\right|}{x}$$

and

$$\frac{\partial y'}{\partial x} = \frac{\sum_j |F_j \, \text{cal.}| - \sum_i |F_i \, \text{cal.}|}{x^2}$$

and

$$\sum_j |F_j \, \text{cal.}| = \tfrac{1}{2} \sum_k F_k \, \text{cal.} + \frac{x^2}{2} \frac{\partial y'}{\partial x}. \qquad (c)$$

It is obvious according to equation $(c)$ that the same summing procedure as in the previous case gives the appropriate result, except that the $|F_k \, \text{cal.}|$'s are summed rather than the $|F_k \, \text{obs.}|$'s.

The function $R_3'$ would be preferred over $R_3$, for the experimental data is what we are trying to match.

The value of the 'reliability index' (or any of the other residuals) as a criterion for judging the accuracy of a structure determination has been much discussed (Donohue & Trueblood, 1956; Lipson & Cochran, 1953; Evans, 1956). The consensus of opinion seems to be that it is inadequate as a sole criterion but has value as one of several criteria to be applied concurrently. Examples of the others to be applied are: unusual differences between observed and calculated structure factors of individual reflections, shapes and relative heights of Fourier peaks and the extent of background variations between them, reasonableness of bond lengths and angles and of the packing arrangements, etc.

Under rare circumstances would differences in the 'reliability index' of less than 1% be considered significant. A guess, based on only limited experience with the method for deriving the scale factor reported here, would be that it would not significantly improve the value of $R_3$ if the number of independent reflections observed is over 50 and $R_3$ is under or in the neighborhood of 20%. The cases in which this method would give significant improvement in $R_3$ are just those in which unusual differences between the observed and calculated structure factors of particular reflections exist, and thus the proposed structure should be questioned in spite of the value of $R_3$.

Nevertheless it is conceivable that experimental conditions in particular instances might unavoidably be so poor that large discrepancies are inevitable, and in such instances application of this method might indicate that the experimenter was not quite as bad off as he thought he was.

On the other hand a more accurate scaling factor results in more accurate electron-density measurements, and it is conceivable in particular instances that the accuracy dictated by the data is so great that the slightly improved scaling factor would significantly improve the results.

## References

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

Donohue, J. & Trueblood, K. N. (1956). *Acta Cryst.* **9**, 615.

Evans, H. (1956). 'Point Group' Colloquium. New York: Polytechnic Institute of Brooklyn.

Hughes, E. W. (1941). *J. Amer. Chem. Soc.* **63**, 66.

Lipson, H. & Cochran, W. (1953). *The Determination of Crystal Structures*, pp. 144–9. London: Bell.

Shoemaker, D. P., Donohue, J., Schomaker, V. & Corey, R. B. (1950). *J. Amer. Chem. Soc.* **72**, 2334.