Acta Cryst. (1958). **11**, 833

An Application of Electronic Computing to X-ray Crystallography

By J. Gillis

Department of Applied Mathematics, The Weizmann Institute of Science, Rehovot, Israel

(*Received* 26 *March* 1958)

A programme for examining observed structure factors for significant Harker–Kasper inequalities has been run successfully on the digital electronic computer at the Weizmann Institute (WEIZAC). As an incidental aid in the preparation of the programme a polynomial approximation has been calculated to an atomic scattering factor of carbon.

1. We describe below a programme for the examination of observed intensities from centro-symmetric structures for sets to which the Harker–Kasper inequalities can be usefully applied.

The input data required are the set of $F_{\text{obs.}}$, the lattice constants, and a mean atomic scattering function $f(\sin\theta/\lambda)$. In the original version of the programme $f(\sin\theta/\lambda)$ was inserted as a table. However, it was subsequently found that the programming could be simplified by means of a sub-routine to calculate f directly. This was achieved by using the polynomial

$$\begin{aligned} f(x) = {} & 6{\cdot}02+3{\cdot}59x-319{\cdot}82x^2+2{,}014{\cdot}94x^3 \\ & -6{,}280{\cdot}07x^4+11{,}593{\cdot}61x^5-13{,}245{\cdot}89x^6 \\ & +9{,}196{\cdot}58x^7-3{,}552{\cdot}93x^8+585{\cdot}10x^9 , \end{aligned} \qquad (1)$$

where $x = \sin\theta/\lambda$. The polynomial (1) is in good agreement with the atomic scattering factor for carbon as given by Viervoll & Ögrim (1949). In fact the error is never more than 1 per cent over $0 \le x \le 0{\cdot}8$, nor more than 5 per cent over $0{\cdot}8 \le x \le 1{\cdot}3$, and this certainly covers the entire useful range for most practical purposes. The polynomial was obtained in stages. A first approximation was constructed by means of Tchebycheff interpolation and its deviations from the true valucs plotted. These appeared to constitute something like a quartic curve, and so a suitable fourth degree polynomial was added to the original. The effect was to leave only small patches of relatively small error, and these were again reduced by the addition of further small polynomials. It is not claimed that $f(x)$ is the simplest or most economical polynomial to satisfy the accuracy requirements. In fact, no serious attempt has been made to simplify it, since this could not have resulted in any substantial saving in the sub-routine. Since this work was completed, Vand, Eiland & Pepinsky (1957) have announced more elegant approximations using only two Gaussian terms. The advantage of our polynomial is that it is valid over the entire Cu range and most of the Mo range and with a much smaller maximum relative error than seems to be obtainable from two Gaussians.

2. The first step is the calculation of $|U_{hkl}|$, after doing which the machine stores up in single words the following information:

$$h, k, l; \quad |U_{hkl}|; \quad \exp(-\sin^2\theta) .$$

As a check on large errors of observation and also on possible indexing errors, the machine takes note of all cases in which $|U_{hkl}|$, as computed, comes out ≥ 1. It prints the relevant (h, k, l) and replaces $|U_{hkl}|$ by the arbitrary value 0·8. The list of 'delinquent' (h, k, l)'s is useful subsequently as a warning not to take too seriously, or without independent confirmation, results obtained from inequalities involving these terms. In addition, the appearance of several such items is taken as an indication of the need to adjust the scaling of $F_{\text{obs.}}$. Any sets (h, k, l) for which $\sin^2\theta > 1$ are listed separately and the corresponding terms are ignored throughout the remainder of the programme.

3. We now have to consider the inequalities themselves. For this purpose we shall, as usual, use capitals such as H to represent sets (h, k, l). We consider each pair of terms in the list of data, H_1, H_2, (say) and test them to check whether the parity properties of H_1, H_2 are such that we can find H, H' with $H_1 = H+H'$, $H_2 = H-H'$. If the H_1, H_2 pair fail this test we pass on to another pair. Whenever we have a pair which pases the test then, following with very slight modifications Grison's formulation, (1951), we calculate

$$\begin{aligned} L &= \text{Max}\,(|U_{H_1}|, |U_{H_2}|) \\ M &= \text{Min}\,(|U_{H_1}|, |U_{H_2}|) \\ A &= (1+L)(1+M) \\ B &= (1+L)(1-M) \\ C &= (1-L)(1+M) \\ D &= (1-L)(1-M) \\ E &= (|U_H|+|U_{H'}|)^2 \\ F &= (|U_H|-|U_{H'}|)^2 . \end{aligned}$$

These quantities are then tested to see which, if any, of the following inequalities they satisfy:

$$E > B > D > F \qquad (1)$$
$$E > B > F > D \qquad (A)$$
$$B > E > C > D > F \qquad (2)$$
$$B > E > C > F > D \qquad (3)$$
$$B > E > F > C > D \qquad (B)$$
$$B > C > E > D > F \qquad (4)$$
$$B > C > E > F > D\,. \qquad (5)$$

If any of these seven inequalities is satisfied the machine proceeds to print H_1, H_2 and the indicator of the inequality in question, i.e., one of the numbers 1 to 5 or one of the letters A, B, as the case may be.

The inequalities (A), (B) are theoretically impossible. Their appearance is therefore taken as a danger signal. If they can all be cancelled by reducing a small number of $F_{\text{obs.}}$, then it is reasonable either to do so or, better still, to try to work without these F's altogether. If several (A)'s and/or (B)'s appear without it being possible to attribute them to a small number of $F_{\text{obs.}}$, then one is led to one of the following explanations: either that the $F_{\text{obs.}}$ are not on a correct absolute scale and should be reduced accordingly or that the temperature correction to f has been assumed too steep. The programme permits trying out either or both corrections very simply.

Since, at this stage of a structure determination, the true temperature correction can only be guessed at, the programme has been written on the basis that, if no (A)'s or (B)'s appear at all, the temperature factor can safely be assumed not to be too large. The machine then, on completing the examination, proceeds to multiply each $|U_H|$ by the corresponding $\exp(\frac{1}{4}\sin^2\theta)$ and then re-examines the modified $|U_H|$'s for significant inequalities. It goes on doing this until, at some stage, either too many (A)'s or (B)'s appear or some $|U_H|$ comes out ≥ 1. At this point the machine stops.

The subsequent analysis of the inequalities for actual information about signs is fairly simple and does not take a great deal of time. For this, as well as for other reasons, it has not been thought necessary as yet to mechanize the process. However, the question is at present under consideration.

4. The programme was applied to the $hk0$ terms from α-trans cinammic acid data kindly provided by Dr G. Schmidt by whom, in fact, the structure had already been refined. The total running time on the machine was rather less than ten minutes and the subsequent analysis of the results for sign information took slightly under one hour. The outcome was the determination of the signs of 13 terms. The two-dimensional Fourier sum of these terms was then evaluated on WEIZAC and the result is shown in Fig. 1. The actual positions of the atoms as obtained from the ultimate refinement of the structure are shown in the same figure. The total time from the input of data into WEIZAC to the plotting of the Fourier sum was about two hours.

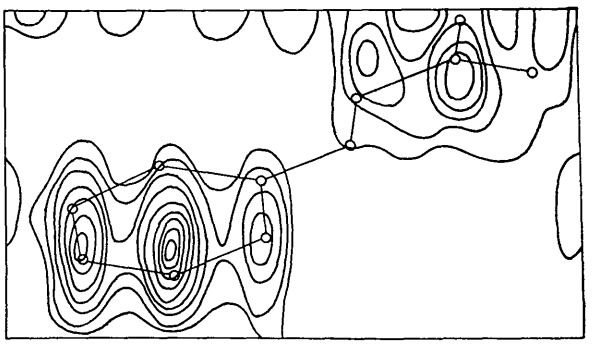

Fig. 1. α-*trans* cinammic acid.

It transpired later that, owing to a scaling error, the $F_{\text{obs.}}$ had all been over-estimated by about 30 per cent. It must be accounted a fortunate accident that this did not lead to serious error. However, there was always the safeguard that, were errors being caused, the appearance of inequalities (A) and (B) would have provided the necessary warning.

A second application was to β-3:4:5:6-dibenzophenanthrene. This was a much more complicated problem with well over 300 structure factors in the $hk0$ projection and 22 atoms in each asymmetric unit. The machine search for inequalities led, in under half an hour, to the elucidation of sufficient signs for what appeared to be a first approximation. This was subsequently refined and the results will be published shortly.

References

GRISON, E. (1951). *Acta Cryst.* **4**, 489.

VAND, V., EILAND, P. F. & PEPINSKY, R. (1957). *Acta Cryst.* **10**, 303.

VIERVOLL, H. & ÖGRIM, O. (1949). *Acta Cryst.* **2**, 277.