

Crystallographic Calculations on the Manchester University Electronic Digital Computer (Mark II)

BY F. R. AHMED AND D. W. J. CRUICKSHANK

Chemistry Department, The University, Leeds 2, England

(Received 31 March 1953 and in revised form 6 May 1953)

The use of the Manchester University Electronic Digital Computer (Mark II) for crystallographic calculations is described. Programmes have been devised for the calculation of structure factors, the computation of observed and calculated differential syntheses with solution for the coordinate changes, and the estimation of errors. One cycle of the three-dimensional refinement process for a structure with 7 independent atoms and 667 structure factors can be carried out in about 1½ hr. electronic computer time.

1. Introduction

The use of a general-purpose electronic digital computer for crystallographic calculations has already been reported by Bennett & Kendrew (1952), who have used the EDSAC at Cambridge for computing two- and three-dimensional Fourier syntheses on networks of points throughout the crystallographically independent part of the unit cell. The Manchester University Electronic Digital Computer (Mark II) came into operation in the summer of 1951, and, through the kindness of Prof. F. C. Williams and the Computing Machine Laboratory, time on the machine for crystallographic calculations has been made available to Prof. E. G. Cox's crystallographic group at Leeds.

Experience in the past few years in the three-dimensional refinement of structure analyses beyond the satisfactory trial-structure stage, and in the extensive calculations involved, hitherto performed by punched-card methods (see e.g. Cox, Gross & Jeffrey, 1949), has shown the especial usefulness of differential syntheses (Booth, 1946*a*) for finding atomic coordinates. On the other hand, much unused information is provided when Fourier syntheses are evaluated throughout the unit cell. Accordingly, during the summer of 1952 we produced a prototype programme for three-dimensional differential syntheses in the space group $P2_1/a$, which was used in the refinement of seven structures. Three of these have already been reported (Ahmed & Cruickshank, 1952; Wheatley, 1953). This first programme showed us the very considerable increase in speed which the electronic computer could give, and gave us the experience with which to plan a more efficient and extensive series of programmes.

This paper describes the essential programmes for a three-dimensional refinement cycle. The stages in this cycle are: calculation of structure factors, computation of observed and calculated differential syntheses with solution for the coordinate changes, and estimation

of errors. For efficient machine computation different programmes will be needed for each of the 230 space groups. We have started with programmes for the two centrosymmetric space groups, $P\bar{1}$ and $P2_1/a$, which most commonly occur in organic structure analyses. These programmes are now in routine use, and some care was taken in their design to arrange that they can be altered easily to other centrosymmetric space groups. Programmes for non-centrosymmetric space groups will require rather larger changes. Other types of programme will be added as time permits. This paper describes the variable parts of the programmes for $P2_1/a$ only, and quotes the times for calculations on anthracene using the experimental data of Mathieson, Robertson & Sinclair (1950).

A full account of the Manchester computer would be out of place in this paper. Details of its construction have been given by Kilburn, Tootill, Edwards & Pollard (1953). Apart from the fact that the Manchester computer has a large magnetic auxiliary store which EDSAC was not provided with at the time of Bennett & Kendrew's paper, from the programmer's point of view the two machines are of the same fundamental type, with serial operation and a single address code, though their construction and coding details are quite different. Some of the general remarks given by Bennett & Kendrew about the operation of EDSAC apply also to the Manchester computer and may be used to supplement the following short description.

The high-speed storage of the Manchester computer consists of 8 'pages' (actually cathode ray storage tubes), each containing 64 'short lines' of 20 binary digits. These short lines may represent either instructions or numbers. Most arithmetical instructions use numbers represented in 'long lines' of 40 digits, formed from two successive short lines. The actual computing is done in the arithmetical unit, consisting of an 80-digit accumulator and a 40-digit multiplicand register together with some special facilities, the numbers and instructions being taken from the high-speed electronic

store. Addition, subtraction and logical instructions take 1.20 msec. and multiplication 2.16 msec. Provision is made for testing whether the number in the accumulator is positive or negative, and for using the result of this test to determine whether the machine obeys the next order in the sequence or transfers control to some other part of the programme; this facility is used for selecting alternative paths in the computation. Control transfers, whether conditional or direct, take 0.96 msec. The magnetic store at present has 128 'tracks', each holding the contents of two pages, giving a total capacity of 327680 digits. Either one or two pages can be written on to or read out of a track by instructions taking 94 msec. for writing and 38 msec. for reading or checking.

Information is fed into the machine by data punched on teleprinter tape, which is read by a photoelectric reading unit at a maximum speed of 200 characters per sec. There are 32 different characters, each one of the five-figure binary integers, which may alternatively represent a code letter. The usual operating procedure is that by a master input programme permanently available in the machine, the programmes for a given calculation are assembled in the electronic store as they are read off the tape, and transferred page by page to the magnetic store. When all the programmes and required data have been read in, the calculation proper starts. Results from the machine are either punched out on tape at a maximum speed of 15 characters per second, or printed by a teleprinter unit at a maximum speed of 6 characters per second, or they can be both punched and printed at the printing speed. The output programmes are usually arranged to print numbers in decimal form. Slower input and output speeds are obtained if the computing times between individual characters are slower than these rates.

The operating sequence in our work is that by means of the master input programme preliminary input programmes are stored, which are then used to store the structure factors, coordinates and scattering factors. The main programmes are then stored by the master input, and the computation proper is begun.

2. Structure factor input and storage

All the programmes work sequentially through the F 's, which are arranged, with the aid of code symbols, in an order such that the (hkl) indices of the plane being used at any stage of the computation are known. The F 's are stored in successive 20-digit short lines, though some short lines may contain code symbols only. The first two digits of a short line are not used in the storage. The digits 3–15 hold $|F|$, so that values from $(2^{13}-1) = 8191$ to 0 can be stored (the scale is usually chosen so that $F_{\max.} \leq 819.1$). The digits 16–20 hold the code combination, which includes the sign of the plane.

The F_o and F_c values are stored independently. For $P2_1/a$ each set of F 's is split in two groups, the first with $(h+k)$ even and the second with $(h+k)$ odd. The planes of the first group are arranged starting with $h = k = l = 0$, then

- (i) $|l|$ is increased by 1 successively, while h, k are kept constant, $F(hkl)$, if present, coming before $F(hk\bar{l})$;
- (ii) h is increased by 2, while k is kept constant, and (i) is repeated, always starting with $l = 0$;
- (iii) k is increased by 1, and (i) and (ii) are repeated, starting with $h = 0$ or 1 as the new value of k is even or odd.

The planes of the second group are arranged in the same order, except that stage (iii) starts with $h = 0$ or 1 as the new value of k is odd or even.

The tape is punched as follows:

Code symbol P (binary number 22) at the beginning of each of the two groups;

$|F(hkl)|$ as decimal figures followed by $L(18)$ or $W(19)$ for positive or negative amplitude;

$|F(hk\bar{l})|$ followed by $T(16)$ or $Z(17)$ for positive or negative amplitude;

$H(20)$ for the end of the $|l|$ runs, and change of h ;

$Y(21)$ for the end of the h runs, and change of k ;

$Q(23)$ for the end of the data.

Any missing $F(hkl)$'s are omitted, but the symbols T occur for any missing $F(hk\bar{l})$, as the programmes record the $|l|$ values by counting the T symbols. Similarly H and Y symbols occur for missing h and k values.

The F 's are stored in the machine from the tape by a two-page input programme, which assembles the $|F|$'s and code symbols in the form described above on successive short lines of an electronic page. When 128 code symbols have been read in, two pages are full and their contents are transferred to the magnetic store. Succeeding F 's are assembled on the cleared electronic pages. When the whole F tape has been read in, it is passed through the tape reader again, and the newly assembled electronic pages are checked against the material already in the magnetic store.

The hand punching and visual checking of the tape for the 667 independent F_o of anthracene takes about $2\frac{1}{2}$ hr. The input and checking of the tape into the computer takes about $1\frac{1}{2}$ min.

3. Calculation of structure factors

The programme computes structure factors

$$F_c(hkl) = \sum_r f_r \cos(hx_r + ky_r + lz_r), \quad (1)$$

where x_r, y_r, z_r are the coordinates of atom r , f_r its scattering factor, and due account is taken of the space-group symmetry. F_c 's are computed for all planes having $F_o \neq 0$. The f 's for each plane are found by interpolation in tables of f against $\sin^2 \theta$, θ being the Bragg angle.

Two computing schemes suggest themselves as possibilities, either to treat each plane separately by computing and adding the contributions of the different atoms to that plane and then repeating the process for the next plane, or to compute the contribution of the first atom to all the planes and then to repeat for the other atoms, adding contributions to the running totals of each plane. The latter scheme is possible only on a machine with a large auxiliary store, as space for the equivalent of at least three sets of F 's is needed. Simple considerations show that this scheme is much faster, the time spent on the extra magnetic transfers being small in comparison with the extra instructions and recalculations in the other scheme.

In $P2_1/a$ the computations are carried out according to the formulae:

$$h+k = 2n;$$

$$F(hkl) = 4 \sum_{\substack{\text{atom} \\ \text{kinds}}} f_r \sum_{\substack{\text{atoms} \\ \text{of kind } r}} \{(\cos hx_s \cos ky_s) \cos lz_s - (\sin hx_s \cos ky_s) \sin lz_s\}, \quad (2)$$

$$h+k = 2n+1;$$

$$F(hkl) = 4 \sum_{\substack{\text{atom} \\ \text{kinds}}} f_r \sum_{\substack{\text{atoms} \\ \text{of kind } r}} \{-(\sin hx_s \sin ky_s) \cos lz_s - (\cos hx_s \sin ky_s) \sin lz_s\}. \quad (3)$$

The computation is programmed to proceed as follows:

(i) A preliminary programme computes for atom 1 all the $\cos hx_1$, $\cos ky_1$, $\cos lz_1$ values and stores them on an electronic page. The sines are stored on another page. ($\cos x_1$ and $\sin x_1$ are computed by a standard sub-routine according to an expansion formula, and the $\cos hx_1$ and $\sin hx_1$ are generated from these by the addition formulae.)

(ii) The F_o values are read down from the magnetic to the electronic store, one page at a time, and a programme works down the list of code symbols, interpreting each and recording the (hkl) values. The trigonometric contributions of atom 1 to each plane are computed using the tables prepared in stage (i), together with sub-products ($\cos hx_1 \cos ky_1$) and ($\sin hx_1 \cos ky_1$), for the $h+k = 2n$ group of planes, which change only when h changes. These contributions are stored in corresponding short lines of another electronic page. When full this page is written in the magnetic store, and a new page of F_o values is read down.

(iii) Stages (i) and (ii) are repeated for all atoms with the same f 's as atom 1, and their contributions to each plane are added to those of atom 1.

(iv) When the trigonometric contributions of all atoms of the same kind as atom 1 have been dealt with, the trigonometric sub-totals are multiplied by the appropriate f values. The programme computes $\sin^2 \theta$ as a function of h , k , l , and then finds the f values by interpolation in the f tables.

(v) These totals for the first kind of atom are transferred to other magnetic tracks. Later the contri-

butions of other kinds of atom are added to them.

(vi) Stages (i)–(v) are repeated for the other kinds of atom.

(vii) Finally the pages of F_c values are standardized to the same pattern as the F_o values, each F_c being given the code symbol appropriate to its sign, and the code symbols referring to index changes only are inserted from the F_o pages.

Apart from checking the sine and cosine tables as they are computed, no detail checks on the calculations can be usefully inserted, so that the only fundamental check on the calculations is to repeat them. After they have been repeated a checking programme compares the two sets of F_c values, printing out the number of F_c pages which disagree. If the check succeeds completely, the programme gives the F_c phases to the F_o , and the differential syntheses may be started immediately. If there are disagreements, a third set of F_c 's is computed, after which the checking programme compares the pages of all three sets and if possible produces a set, each page of which agrees with at least two of the three corresponding pages. If this can be done the F_o 's receive the F_c phases; if not further calculations are necessary.

If the F_c values are required for study, a two-page programme prints and/or punches them out in the decimal scale with their code symbols. The output tape is in the same form as the F input tape, and may be used for input on a later occasion.

As there is no need for great accuracy, the internal calculations are arranged to be correct to the equivalent of two decimal places, though the output is only to one decimal place, with a maximum value of 819.1. The trigonometric functions are of six-decimal-figure accuracy, and the scattering factors are found by linear interpolation in tables of f 's given to two decimal places, the increments of $\sin^2 \theta$ being graded in groups to give roughly constant increments in the f 's. The largest f value which can be used is 163.84.

The whole structure-factor programme comprises 25 pages of instructions. The present programme can be used without modification to compute up to about 1200 F 's for up to 21 atoms of not more than 7 kinds.

A single computation of the set of 667 F 's for anthracene takes 6 min. 45 sec.; of this time 36 sec. are needed for stages (i) and (ii) of each atom, and 1 min. 35 sec. for stage (iv). Experience shows that two computations of sets of this length usually check completely. The printing out of the F_c 's with or without punching takes 10 min. 45 sec.; punching alone is three times as fast and takes 3 min. 25 sec.

The largest structure-factor calculations we have done so far are for β -ionylidene crotonic acid (MacGillavry, Kreuger & Eichhorn, 1951). This involved the computation in the space group $P\bar{1}$ of 1199 F 's for a total of 19 atoms of 3 kinds. Perfect agreement was obtained from the first two computations, each of which took 31 min. The printing out took 16 min., and the punching alone $4\frac{1}{2}$ min.

4. Differential syntheses

The basic programme computes

$$\rho(x, y, z) = \Sigma \Sigma \Sigma F(hkl) \cos(hx + ky + lz), \quad (4)$$

the three first derivatives $\partial\rho/\partial x$, $\partial\rho/\partial y$, $\partial\rho/\partial z$, and the six second derivatives $\partial^2\rho/\partial x^2$, $\partial^2\rho/\partial x\partial y$, $\partial^2\rho/\partial x\partial z$, $\partial^2\rho/\partial y^2$, $\partial^2\rho/\partial y\partial z$, $\partial^2\rho/\partial z^2$, for a given point (x, y, z) , due account being taken of the space-group symmetry. The three equations (Booth, 1946a)

$$\varepsilon_x \left(\frac{\partial^2 \rho}{\partial x^2} \right) + \varepsilon_y \left(\frac{\partial^2 \rho}{\partial x \partial y} \right) + \varepsilon_z \left(\frac{\partial^2 \rho}{\partial x \partial z} \right) = - \left(\frac{\partial \rho}{\partial x} \right), \text{ etc.}, \quad (5)$$

are then solved for the shifts $(\varepsilon_x, \varepsilon_y, \varepsilon_z)$, giving the coordinates $(x + \varepsilon_x, y + \varepsilon_y, z + \varepsilon_z)$ of the adjacent maximum of $\rho(x, y, z)$.

In $P2_1/a$ the computation of a derivative, such as $\partial^2\rho/\partial y\partial z$, is carried out according to the formula

$$\begin{aligned} \frac{\partial^2 \rho}{\partial y \partial z} = & 4 \sum_{h+k=2n} k \sin ky \{ \cos hx [(F(hkl) + F(hk\bar{l})) l \sin lz] \} \\ & + 4 \sum_{h+k=2n+1} k \cos ky \{ \sin hx [(F(hkl) + F(hk\bar{l})) l \sin lz] \} \\ & + 4 \sum_{h+k=2n} k \sin ky \{ \sin hx [(F(hkl) - F(hk\bar{l})) l \cos lz] \} \\ & - 4 \sum_{h+k=2n+1} k \cos ky \{ \cos hx [(F(hkl) - F(hk\bar{l})) l \cos lz] \}. \end{aligned} \quad (6)$$

The computation of the differential syntheses is programmed to proceed as follows:

- (i) The $\cos hx$ etc., values are computed and stored as at stage (i) in the structure-factor calculations.
- (ii) The F values are worked through page by page. During sequences of fixed (h, k) three sub-totals are formed. In the first summation for planes with $h+k=2n$, these are

$$\left. \begin{aligned} A(hk) &= \sum_l [F(hkl) + F(hk\bar{l})] \cos lz, \\ B(hk) &= \sum_l [F(hkl) + F(hk\bar{l})] (-l^2 \cos lz), \\ C(hk) &= \sum_l [F(hkl) + F(hk\bar{l})] (-l \sin lz), \end{aligned} \right\} \quad (7)$$

corresponding totals being formed during the other three parts of the summation.

- (iii) At the end of the sequence of fixed (h, k) , as h changes, the three sums (7) are used in six sub-totals, which are, in the first part of the summation,

$$\left. \begin{aligned} a(k) &= \sum_h A(hk) \cos hx, \\ b(k) &= \sum_h A(hk) (-h^2 \cos hx), \\ c(k) &= \sum_h A(hk) (-h \sin hx), \\ d(k) &= \sum_h B(hk) \cos hx, \\ e(k) &= \sum_h C(hk) \cos hx, \\ f(k) &= \sum_h C(hk) (-h \sin hx). \end{aligned} \right\} \quad (8)$$

(iv) When k changes the six sums (8) are similarly used in running totals for the ten required sums ρ , $\partial\rho/\partial x$, $\partial^2\rho/\partial x\partial y$ etc. The contribution, for instance, to $\partial^2\rho/\partial y\partial z$, (6), in the first part of the summation, is $e(k)(-k \sin ky)$.

(v) The F values have to be worked through twice, as only the totals corresponding to the first two summations in (6) are evaluated in the first passage.

This basic programme is usually incorporated in a larger programme which runs through all the atoms in sequence, computing both observed and calculated differential syntheses. The sequence of this master programme is as follows:

- (i) Print the coordinates of atom 1.
- (ii) Compute the ten sums for atom 1 with the F_o data, and store.
- (iii) Repeat stages (i) and (ii), until the ten sums check.
- (iv) Print the ten sums.
- (v) Solve the equations (5), print the shifts, and check.
- (vi) Repeat stages (i)–(v) with the F_c data for atom 1.
- (vii) Repeat (i)–(vi) with the other atoms.

The resultant coordinate shifts, corrected for finite-series effects by Booth's back-correction method (1946b), are obtained by subtracting the calculated from the observed shifts. This is easily done mentally.

The totals computed by the machine are accurate to just over 2 decimal places, and are printed out with 8 figures before the decimal point and 2 after. The coordinates and shifts are printed to 5 decimal places.

The whole differential-synthesis programme comprises 19 pages, of which 5 are common to the structure-factor programme.

The time taken for the observed or calculated differential syntheses for one atom of anthracene is 4 min. 30 sec., made up of 1 min. 59 sec. for each computation of the ten sums, 22 sec. for printing them, and 10 sec. to print, check, and solve for the shifts. Thus if we allow two checking failures at stage (iii) in the master programme (an average figure), the observed and calculated differential syntheses for the seven carbon atoms of anthracene can be computed in 67 min.

5. Accuracy calculations

A five-page programme computes and prints

$$\Sigma |F_o|, \Sigma |F_c|, \Sigma |\Delta F|$$

and

$$\Sigma |\Delta F|^2, \Sigma h^2 |\Delta F|^2, \Sigma k^2 |\Delta F|^2, \Sigma l^2 |\Delta F|^2,$$

where $\Delta F = |F_o - F_c|$, the summations allowing for multiplicity. The first three totals may be used to obtain the agreement index, and to check on the scaling of the F 's. The other totals may be used to estimate the electron density and slope errors (Cruick-

shank, 1949). The coordinate standard deviations can then be estimated from the slope errors and the second derivatives given by the differential-synthesis programme.

For anthracene the computation, including printing time, takes 1 min. 8 sec. It must then be repeated as a check.

6. Discussion

When the preceding programmes are used in succession one cycle of the three-dimensional refinement process for anthracene can be completed in about $1\frac{1}{2}$ hr. electronic computer time. This is a very great increase in speed in comparison with older methods; in fact the longest job is getting the data ready for the calculations. In particular, the F tapes have to be carefully prepared and checked, as false symbols will cause succeeding planes to be misinterpreted.

During our work on the refinement of anthracene and naphthalene (Ahmed & Cruickshank, 1952) we computed the differential syntheses both by the electronic computer and by 'Hollerith' punched-card equipment, comprising a Senior Rolling Total Tabulator (852/0) with 6 counters, a sorter and a reproducer (200/0). The preparation of the F_o and F_c data for anthracene took about the same time in both methods, 5 hr. to punch the computer tapes and 5 hr. 40 min. to punch and check the data cards. With the latest computer programmes the anthracene observed and calculated differential syntheses can be done in about 73 min., including the time to store the programmes and data. The punched-card methods required 63 hr., to which had to be added about 5 hr. hand computing for solving the equations for the shifts. Thus, there is a factor of the order of 50:1 between the two methods as far as machine computing time is concerned.

It must be stressed that this comparison of the two methods applies to a particular type of crystallographic calculation and to a particular Hollerith installation. If the cross derivatives $\partial^2 Q/\partial x \partial y$, $\partial^2 Q/\partial x \partial z$, $\partial^2 Q/\partial y \partial z$ had been estimated and not computed, so that the shifts were found only approximately, $25\frac{1}{2}$ hr. Hollerith time would have been saved, as against a smaller proportionate saving on the computer. The punched-card methods would also show up better in the computation of Fourier syntheses at large numbers of points on networks throughout the unit cell.

Nothing has been said so far about the work of programming and the task of getting a programme to work. As an instance, the programming of the structure-factor calculations in $P2_1/a$ (excluding the checking programmes) took about $3\frac{1}{2}$ weeks initial work, followed by another $4\frac{1}{2}$ weeks part-time work, elimi-

nating the faults found when on the machine; about 18 hr. machine time were needed in this development period. Fault finding can be a slow and tedious business, especially when the programme has to be worked through on the machine one instruction at a time following the details of the computation. This means that it is not worth programming small computations which are going to be used only once. Conversely, crystallographic work has the great merit that only standard types of calculation are involved, so that once appropriate programmes have been prepared they are ready for use in any structure analysis which needs them. New programmes are needed for each new space group, but we find that the changes can be made and checked quite quickly. The checking is done without devising special hand-tested calculations for each space group, by using the fact that calculations for any centrosymmetric space group can be expressed in terms appropriate to $P\bar{1}$, so that test calculations can be checked using the $P\bar{1}$ programmes.

The credit for this great speed-up of three-dimensional structure refinement must go primarily to the engineers who designed and constructed the Manchester computer. Our thanks are due to Prof. F. C. Williams for making time on the machine available to us, to Mr R. A. Brooker of the Manchester University Computing Machine Laboratory for much help and advice, to Dr Mary B. Hesse for assisting in the first differential-synthesis programme, and to Prof. E. G. Cox for continuous support and encouragement of this work. One of us (F.R.A.) is indebted to the Egyptian government for a scholarship held during the course of the work.

References

- AHMED, F. R. & CRUICKSHANK, D. W. J. (1952). *Acta Cryst.* **5**, 852.
 BENNETT, J. M. & KENDREW, J. C. (1952). *Acta Cryst.* **5**, 109.
 BOOTH, A. D. (1946a). *Trans. Faraday Soc.* **42**, 444.
 BOOTH, A. D. (1946b). *Proc. Roy. Soc. A*, **188**, 77.
 COX, E. G., GROSS, L. & JEFFREY, G. A. (1949). *Acta Cryst.* **2**, 351.
 CRUICKSHANK, D. W. J. (1949). *Acta Cryst.* **2**, 65.
 KILBURN, T., TOOTILL, G. C., EDWARDS, D. B. G. & POLLARD, B. W. (1953). *Proc. Instn. Elect. Engrs.* In the Press.
 MACGILLAVRY, C. H., KREUGER, A. & EICHHORN, E. L. (1951). *Proc. K. Ned. Akad. Wet. B*, **54**, 449.
 MATHIESON, A. McL., ROBERTSON, J. M. & SINCLAIR, V. C. (1950). *Acta Cryst.* **3**, 245.
 WHEATLEY, P. J. (1953). *Acta Cryst.* **6**, 369.