

## Crystallographic Calculations on the High-Speed Digital Computer SWAC

BY R. A. SPARKS, R. J. PROSEN,\* F. H. KRUSE AND K. N. TRUEBLOOD

*Department of Chemistry, University of California, Los Angeles 24, California, U.S.A.*

(Received 15 September 1955)

The high-speed digital computer SWAC has been used extensively for calculations needed in the determination and refinement of crystal structures. Programs for calculation of structure factors, normal Fourier summations, differential Fourier summations and least-squares refinement are described briefly, and some practical experience with them is discussed. In general these programs may be used with only minor specified changes for almost any crystal of any symmetry. All programs include provision for use of individual anisotropic atomic temperature factors. A few other programs for more trivial, but nevertheless tedious, calculations are also described.

### Introduction

Several applications of high-speed digital computers to the determination and refinement of crystal structures have recently been reported (Ordway, 1952; Bennett & Kendrew, 1952; Ahmed & Cruickshank, 1953; Mayer & Trueblood, 1953; Mayer, 1953; Thompson, Caminer, Fantl, Wright & King, 1954; Cochran & Douglas, 1955). The last of these reports is concerned with attempts at direct determination of structures through systematic examination of sign relationships; all of the others deal with the sort of calculations which arise during the refinement stages. Almost all of these reported procedures suffer from certain limitations of generality; for example, most were designed for particular centrosymmetric space groups, and extension to other space groups, particularly non-centrosymmetric ones, involved additional complications and considerably more coding and computing time. In addition, determination and refinement of individual anisotropic atomic temperature-factor parameters have not generally been considered. During the past two years we have written and applied programs for almost all of the calculations normally needed in refinement procedures for any space group, with allowance, when desired, for individual anisotropic atomic temperature factors. These procedures have been applied to data from more than thirty different crystalline substances, representing at least fifteen different space groups, including triclinic, monoclinic, orthorhombic, tetragonal, hexagonal and cubic symmetry. The present report consists of descriptions of the organization of the different types of calculations, together with some general comments on desirable features of routines for such calculations.

### Notation

The following special notation will be used:

\* Present address: National Bureau of Standards, Washington 25, D.C., U.S.A.

### Subscripts

- $g$  — one of the crystallographically independent atoms in an arbitrary asymmetric unit.  
 $i$  — one of the direct axes or reciprocal axes.  
 $j$  — any atom in the unit cell.  
 $m$  — an atom equivalent by symmetry to one of type  $v$ .  
 $v$  — the (arbitrarily chosen) representative atom of a given set of atoms related by symmetry in a least-squares refinement.

### Other symbols

- ${}_jB_{ii'}$  — the coefficient of  $h_i h_{i'}$  in the expression for a general anisotropic temperature factor for atom  $j$ .  
 $\beta_j$  — the exponent in the temperature factor for atom  $j$ .  
 $\varphi_j$  —  $\sum_i h_i x_{ij}$ .  
 $L$  — the Lorentz-polarization factor.  
 $S_k$  — the scale factor in the  $k$ th stage of a Fourier summation.  
 $s_{mi}$  — the sign of  $x_{ivg}$  in the  $m$ th equivalent position.  
 $T_i$  — Tunell correction for a plane on a Weissenberg photograph.  
 $\tau_{mii'}$  — sign of  ${}_{vg}B_{ii'}$  for the  $m$ th equivalent position.

### The computer

The National Bureau of Standards Western Automatic Computer (SWAC) (Huskey, Thorensen, Ambrosio & Yowell, 1953) is a four-address machine with a high-speed electrostatic memory of the Williams-tube type which can hold 256 36-digit binary numbers ('words') and a magnetic drum which at present can store 8192 such numbers. All of our routines utilize the drum for storage of subroutines, tables, data and intermediate answers. For numbers in the high-speed memory, addition and subtraction are performed in 64 microsec., multiplication in 368 microsec., and division in a few milliseconds. Transfer of information to or from the drum requires 17 millisecc. for each group of 64 or

fewer words; consequently such transfers are minimized. Standard IBM cards are used for input into the SWAC with a modified collator and output from the SWAC through a modified reproducing punch. The use of punched cards for input and output is an extremely convenient feature since it greatly facilitates the preparation, checking and processing of data and answers. Direct output to an IBM 402 tabulator may also be used, with or without simultaneous punched-card output.

Information may be input in either decimal or binary form; if the former method is used, a subroutine for conversion to binary form within the computer must be incorporated in the program. The time used for conversion is a negligible fraction of the total input time. Binary input (with 10 words per card) is normally preferred for information which is to be used more than once because it is more economical of cards, and hence of input time. Thus all routines are converted to binary cards and used in binary form thereafter, and any data which will be used repeatedly in a calculation, such as observed structure-factor amplitudes with their associated indices, are normally converted to binary form. Input and output operations are kept to a minimum because they are slow relative to the speed of operations within the machine. The effective time for binary card input is about 25 millisecond per word. In most of our calculations, 10–18 decimal answers, with 2–6 decimal digits each, are punched on each answer card; the effective time for output is of the order of 50–100 millisecond for each answer.

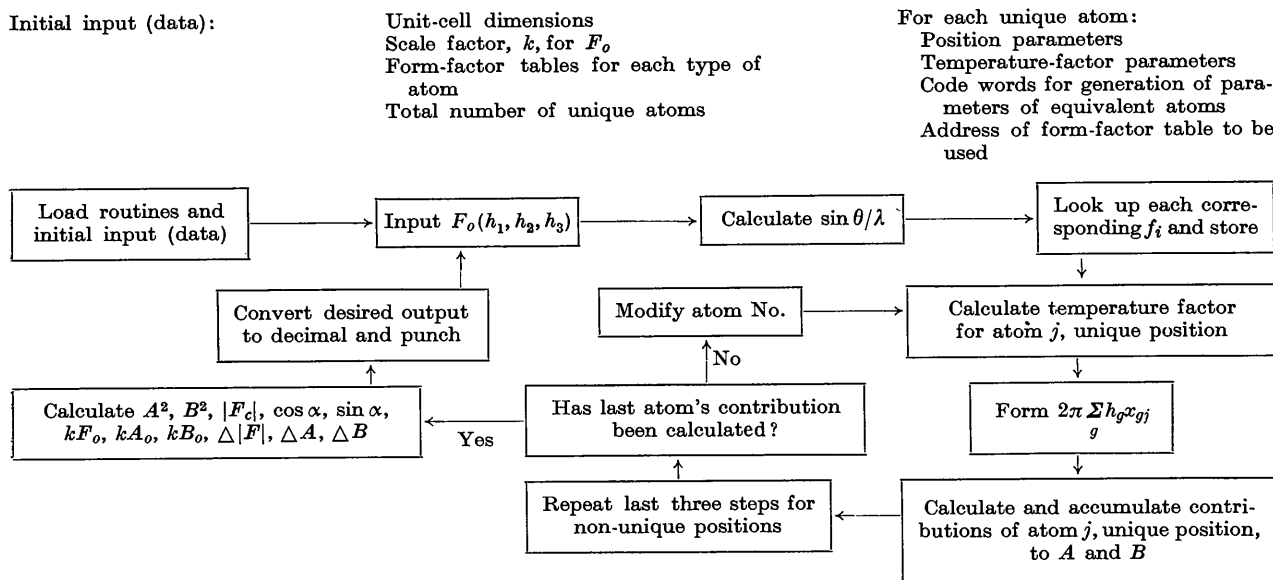
### General discussion

In all of our coding, flexibility of the final codes has been given primary emphasis. All of our basic codes may be used for any space group and for any number of atoms likely to be encountered at the present stage of crystal-structure analysis. Initially, separate programs were written for structure-factor calculations, Fourier summations, and least-squares refinement; the first two of these have now been combined with a routine for differential Fourier syntheses into one master routine which consists of approximately 1200 commands. Only about 50 of these commands, which occur on eight specified 'variable cards' in the routine, need be changed whenever a new crystal is being studied, and these changes are made in a specified manner which depends upon the lattice constants, the symmetry, the number of independent atoms and other similar unique features of the particular substance. The least-squares program, which also includes a structure-factor calculation, consists of about 700 commands, of which about 20 are normally varied from one crystal to the next. For simplicity in the present discussion each of these types of calculation will be considered separately.

Because our general programs are so flexible they

do not perform the calculations as rapidly or conveniently as would sometimes be possible with specific programs written for particular space groups. This is especially true for three-dimensional Fourier summations, since the general routine requires input for each structure factor except those related by Friedel's law. Consequently, appropriate modifications, which involved the alteration of about 60 commands, have been made in the Fourier routine in order to create a routine specific for the common space group  $P2_12_12_1$ . This routine, which requires only the unique reflections to be input, is about 25% faster than the general routine, and also far more convenient, since it obviates the need for reproduction and sorting of cards corresponding to the non-unique reflections. This latter advantage is important in work with crystals with large asymmetric units, such as the vitamin B12 fragment (Hodgkin, Pickworth, Robertson, White, Trueblood & Prosen, 1955) for which there are more than 3300 independent observed reflections. No appreciable further gain in computing time can be made as long as card input and output are used, since these operations now consume nearly one-half of the total elapsed time during a Fourier summation. Because there can be so little gain in computing time, it is questionable whether it is worthwhile to write special codes for more than the most common space groups.

One advantage of general routines, with specified variable commands, is the simplification of code-checking; once the initial routine has been code-checked logical errors are rare, since they can arise only from failure of the coder to write the variable commands correctly. Several different procedures are employed routinely to check on the calculations: (a) Hand-calculation is used whenever feasible to check specific points or stages in a complex problem. (b) In all Fourier calculations, summations are made over a slightly larger volume of the unit cell than corresponds to the asymmetric unit; this permits a check on the symmetry. (c) Because of the possibility of loss of a digit in transfer to or from the drum, all such transfers of commands and data are accompanied at least by a memory-sum check, which compares the sum of all information transferred before and after the transfer and repeats the process if an exact check is not obtained. (d) A variety of routine checks is made; for example, to verify that all of the atomic parameters have been prepared correctly, read correctly by the machine and identified with the proper atoms, all of the drum channels on which these parameters are stored may be punched out and proof-read before a complex calculation is undertaken. Because of the persistence of human errors as well as the possibility of occasional machine errors, such double-checks are highly desirable before a calculation which will require many hours of computer time. Similarly, whenever any calculation is interrupted (all routines are arranged so that the calculations need not be continuous) it is standard practice to repeat a small portion of that

Table 1. *Flow diagram for structure-factor calculation*

already done to insure that the machine is performing identically.

In actual practice random machine errors have vitiated only a very small fraction of our computing time, probably no more than 1 or 2%. When there are machine difficulties it is almost always manifest at once in a 'blow-up' of the calculation; that is, a command is altered in such a way that the normal cycle of the calculation is interrupted and the machine then proceeds in a manner obviously unrelated to the problem at hand.

### Structure-factor calculations

Structure factors, defined in equations (1)–(4), are calculated by a routine which presupposes no symmetry and which, consequently, is suitable for any space group after specified modifications in the few variable cards.

$$F(h_1, h_2, h_3) = A(h_1, h_2, h_3) + iB(h_1, h_2, h_3). \quad (1)$$

$$A = \sum_j f_j \exp[-\beta_j(h_1, h_2, h_3)] \cos 2\pi\varphi_j. \quad (2)$$

$$B = \sum_j f_j \exp[-\beta_j(h_1, h_2, h_3)] \sin 2\pi\varphi_j. \quad (3)$$

The exponential term in these equations is the temperature factor for atom  $j$ , which may be either isotropic (one parameter for each atom) or anisotropic (as many as six parameters for each atom). In the latter case  $\beta_j$  is given by the general expression

$$\beta_j = {}_jB_{11}h_1^2 + {}_jB_{22}h_2^2 + {}_jB_{33}h_3^2 + {}_jB_{12}h_1h_2 + {}_jB_{23}h_2h_3 + {}_jB_{31}h_3h_1. \quad (4)$$

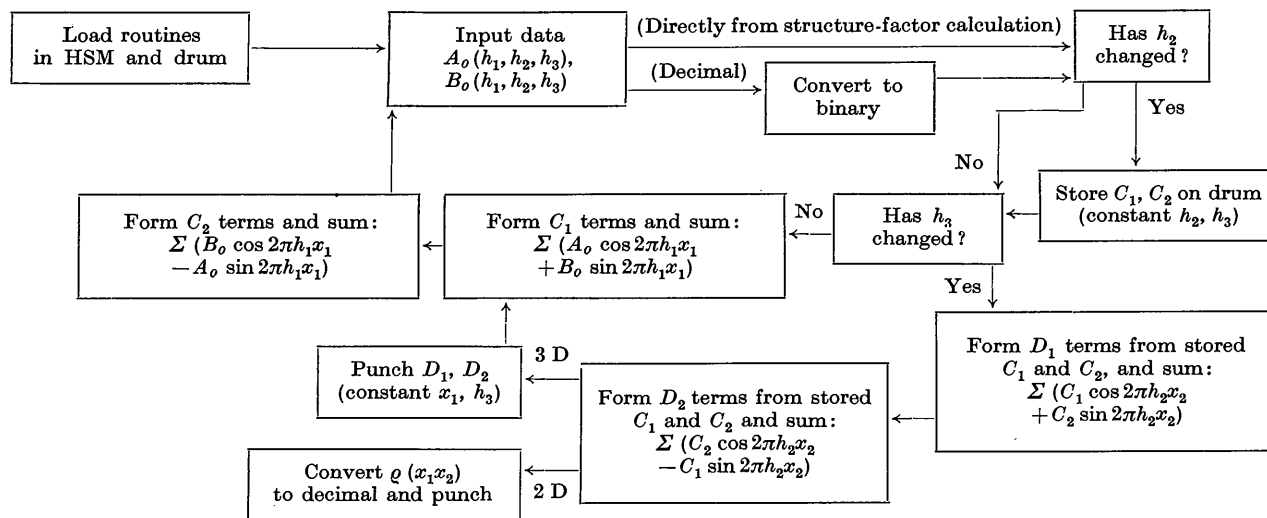
The general organization of the calculation is indicated in the accompanying flow diagram (Table 1).

Each plane is calculated independently, in contrast with some reported methods (Ahmed & Cruickshank, 1953) in which the contribution of one atom to many planes is calculated at one time. The latter method can be faster but has the disadvantage that the calculation cannot usually be interrupted, either intentionally or because of machine failure, until all contributions to all planes have been summed. In our procedure, each answer is punched as soon as it is obtained and the calculation may be interrupted at any stage.

Trials of polynomials to represent form factors (Mayer, 1953) showed that they are not significantly faster than table searching and are appreciably less convenient unless the problem of storage room is critical. Tables of form factors for each type of atom are stored with a tabular interval of 0.006 in  $\sin \theta/\lambda$  and values are selected from these tables without interpolation. The maximum error made thus corresponds to an error in  $\sin \theta/\lambda$  of  $\pm 0.003$ , or about 1% in a form factor. When a new type of atom, or a modified form factor, is encountered, less than one hour is needed to prepare the needed table, punch it decimally, and convert it for use in the routine. Normally storage room is provided for five different  $f$ -tables, corresponding to five chemically distinct types of atoms; to date we have needed no more, but simple modifications would permit extending this storage space almost indefinitely.

With a routine written for P1 it is of course necessary to consider explicitly all of the atoms in the unit cell. As indicated in the flow diagram, special code words are utilized for generation of the parameters of the atoms equivalent to the arbitrarily chosen 'unique' atoms in the asymmetric unit. These code words are examined by a sequence of commands which suc-

Table 2. Flow diagram for first and second stages of Fourier summations



cessively doubles them and tests for 'overflow'. Thus if a binary digit is placed in the extreme left (most significant) position of a word, doubling of that word moves the digit out of the word to the left, a situation which SWAC can recognize as overflow and which can cause the machine to execute further commands which it would not obey in the absence of overflow. These commands may say, 'replace  $x_1$  by  $-x_1$ ' or 'add 0.5 to  $x_2$ '; one can thus, by the presence or absence of digits in appropriate positions, readily represent all possible relations among position parameters which may result from the various crystallographic symmetry operations. When a center of symmetry is present, the routine automatically bypasses the summation indicated by equation (3).

Isotropic temperature factors do not vary with symmetry operations; the way in which general anisotropic temperature factors transform for monoclinic substances has been reported (Rollett & Davies, 1955) and we have worked out the corresponding transformations for all other possible symmetry operations (Trueblood, 1956). For all but hexagonal and trigonal crystals, these transformations may be represented by fairly simple and readily specified code words which are already normally included in our routines; the transformations for temperature factors in the hexagonal system are somewhat more complex and those for trigonal crystals are almost hopelessly so for this sort of automatic generation. Our current practice with crystals of these symmetries is to include with the input data the parameters of each atom in the unit cell related to the arbitrary 'unique' atoms by any axis of order 3 or 6. The other symmetry operations needed are then represented in the usual fashion by code words.

Structure-factor calculations have been performed for crystals with from 3 to more than 100 atoms in the asymmetric unit; in the latter case, individual iso-

tropic temperature factors were used. The routine can normally accommodate about 200 different atoms if each is permitted to be generally anisotropic, and about 1000 if each is isotropic, as of course would be assumed in any practical calculation on such a scale. The times required for the calculations for one plane in the crystals so far studied have varied between about 2 and 7 sec., of which about 1 sec. is required to punch the decimal answer card. The normal output consists of

$$h_1, h_2, h_3, |F_o|, |F_c|, \Delta|F|, A_o, B_o, \Delta A, \Delta B, \cos \alpha \text{ and } \sin \theta/\lambda.$$

Summations of  $|F_o|$ ,  $|F_c|$  and  $\Delta|F|$  are automatic when the answers are tabulated, providing an approximate check of the scale factor and the reliability index.

### Fourier summations

In the general form of the Fourier routine, summations are made in three successive stages according to equation (5) as expanded in equations (6)–(10). In these equations it is assumed that the first summation index is  $h_1$ , the second  $h_2$ , and the third  $h_3$ . The accompanying flow diagram (Table 2) describes the process for the first two stages.

$$\rho(x_1, x_2, x_3) = \frac{F(0, 0, 0)}{V} + \frac{2}{V} \sum_0^{\infty} \sum_{-\infty}^{\infty} \sum_{-\infty}^{\infty} |F(h_1, h_2, h_3)| \times \cos 2\pi[h_1 x_1 + h_2 x_2 + h_3 x_3 - \alpha(h_1, h_2, h_3)]. \quad (5)$$

Summation over  $h_1$  gives a function  $C(x_2, h_2, h_3)$  which may for our purpose most conveniently be considered as the sum of two functions,  $C_1$  and  $C_2$ :

$$C_1(x_2, h_2, h_3) = S_1 \sum_{-\infty}^{\infty} |F(h_1, h_2, h_3)| (\cos \alpha \cos 2\pi h_1 x_1 + \sin \alpha \sin 2\pi h_1 x_1), \quad (6)$$

$$C_2(x_1, h_2, h_3) = S_1 \sum_{-\infty}^{\infty} |F(h_1, h_2, h_3)| (\sin \alpha \cos 2\pi h_1 x_1 - \cos \alpha \sin 2\pi h_1 x_1). \quad (7)$$

The scale factor  $S_1$  is so chosen that the sums are on an appropriate scale. These summations are performed within the high-speed memory and the answers are stored on the drum whenever  $h_2$  changes. When  $h_3$  changes, the answers from the first stage are used in the calculation of the second stage, yielding a function  $D(x_1, x_2, h_3)$  which is again conveniently considered as the sum of two functions  $D_1$  and  $D_2$ :

$$D_1(x_1, x_2, h_3) = S_2 \sum_{-\infty}^{\infty} (C_1 \cos 2\pi h_2 x_2 + C_2 \sin 2\pi h_2 x_2), \quad (8)$$

$$D_2(x_1, x_2, h_3) = S_2 \sum_{-\infty}^{\infty} (C_2 \cos 2\pi h_2 x_2 - C_1 \sin 2\pi h_2 x_2). \quad (9)$$

The advantage of expansions in this form is that equations (6) and (8) are exactly parallel in form, as are also equations (7) and (9); thus essentially the same routine may be used for Stage 1 as for Stage 2.

In calculation of a two-dimensional Fourier summation, only  $D_1$  is evaluated and it is converted to decimal form and punched directly, since  $D_1(x_1, x_2, 0)$  is identically (with proper scale factors) the two-dimensional electron density.

In evaluation of a three-dimensional summation,  $D_1$  and  $D_2$  are punched in binary form with appropriate decimal identification indices, sorted if necessary, and then used directly as input in the final summation

$$E(x_1, x_2, x_3) = S_3 \sum_0^{\infty} (D_1 \cos 2\pi h_3 x_3 + D_2 \sin 2\pi h_3 x_3). \quad (10)$$

Since equation (10) is again precisely parallel to equations (6) and (8) the routine for performing the summation is very nearly identical with those for calculation of  $C_1$  and  $D_1$ .

The final answers are punched decimally; sixteen values of the electron density, corresponding to constant  $x_1$  and  $x_2$  and sixteen values of  $x_3$ , are punched on each card. The scale factors  $S_1$ ,  $S_2$  and  $S_3$  are chosen so that the punched answers are some convenient multiple, usually 100 times, the electron density in  $\text{e.}\text{\AA}^{-3}$ . The precise relation of  $\rho$  and  $E$  is

$$\rho(x_1, x_2, x_3) = K + \frac{2}{V} \frac{GH}{S_1 S_2 S_3} E(x_1, x_2, x_3), \quad (11)$$

in which  $K$  is  $F(0, 0, 0)/V$  if this term is omitted in the first stage of the summation (as is common), and zero otherwise;  $G$  is the ratio of the data on an absolute scale to the actual input data; and  $H$  is a factor arising in the conversion of binary answers to decimal form. Since the initial answers may be tabulated while the computation of the remainder of the asymmetric unit is taking place, the printing of the entire summation may be completed within a few minutes after a computation is finished.

Input into the Fourier routine may be in binary form directly from a structure-factor calculation when the combined master routine is used, with no intermediate punchout, or from the usual decimal structure-factor answer cards, which contain all the requisite data needed for either an observed Fourier or a difference Fourier. The input data are arranged in such order that the index summed over first varies most rapidly, that summed next varies second most rapidly, and that summed third varies least rapidly. With the exception of the special routine written for  $P2_12_12_1$ , which will serve with minor modification for any crystal of the same point group, every summation requires input of  $A$  and  $B$  for every plane except those related by Friedel's law. Consequently, binary input directly from the structure-factor calculation, although very convenient, is not the most economical of SWAC computing time because structure factors must be calculated explicitly for all of these planes. However, it is not as inefficient as might at first be supposed because, at most, only the structure factors of the independent reflections need be punched, and the punching time is, as noted, an appreciable fraction of the total computing time. In actual practice two-dimensional Fourier projections are always calculated with input directly from the structure-factor calculation; on the average between 15 and 20 min. is required for an entire structure-factor calculation and two-dimensional summation, including all desired input and output.

When decimal Fourier input is to be used, the appropriate cards for the non-unique planes may be prepared from those for the unique planes by simple reproduction, with appropriate sign changes as needed. This operation is convenient for most space groups, although it becomes excessively complex and liable to error for some, for example certain hexagonal ones for which equivalent indices cannot readily be identified with a card sorter. In such cases, it is usually safer and more efficient to calculate all structure factors with the computer, and use direct binary input into the Fourier summation.

The variable cards permit specification of the starting point of the summation, which may be anywhere in the unit cell, and the desired number of intervals along each axis. The interval is normally 1/60th; however one may also by appropriate specification on one of the variable cards calculate at intervals of 1/120th. Because the trigonometric functions needed always involve arguments which are integral multiples of a particular value, it is most convenient and economical of time to store tables of these values; in the structure-factor routine, on the other hand, because the arguments cannot be readily specified in advance if high precision is desired, a power series is used.

On the variable cards one may also specify the scale factors, the type of input to be used, whether the summation is two- or three-dimensional, and whether

certain of the summation terms represented by equations (6)–(9) should be omitted. In certain space groups some of these terms may be omitted with a resultant saving of time. When the input is directly from a structure-factor calculation it is also possible to specify whether one wishes an observed, a calculated, or a difference Fourier summation.

The time required for a three-dimensional summation is chiefly a function of the quantity of punched card output, with only minor dependence on the volume of initial input. The time required for the third stage of the calculation is approximately one hour for each eighth of the unit cell when the calculation is made at intervals of 1/60th; the time for the combined first and second stages, although less easy to specify precisely, usually is between one-third and two-thirds of that for the third pass. The longest summation we have yet made involved about 1800 independent reflections summed at intervals of 1/120th along one axis and 1/60th along each of the other two throughout one-fourth of the unit cell. The total time required was slightly more than 6 hr.

Ordinary Patterson functions are of course calculated with this routine, with appropriate bypassing of unneeded terms in equations (6)–(9). We plan to write a routine for calculation of sharpened Patterson functions, with the peak at the origin removed; the initial input will consist merely of the pack of binary cards containing  $F_o$ ,  $h_1$ ,  $h_2$  and  $h_3$  which is already used for structure-factor input.

### Differential Fourier summations

Differential Fourier syntheses (Booth, 1946) provide a particularly efficient method for refinement of atomic coordinates. Their calculation involves little more than a structure-factor calculation, with formation of certain additional products for each reflection, and accumulation of these separately for each atom. These products are those needed in the calculation of the electron density and its three first derivatives and six second derivatives at the position of each atom in the asymmetric unit; they are given by equations such as (12)–(14):

$$\rho_j(x_1, x_2, x_3) = \frac{1}{V} \sum_3 (A \cos \varphi_j + B \sin \varphi_j), \quad (12)$$

$$\partial \rho_j / \partial x_{ij} = \frac{2\pi}{V} \sum_3 h_i (-A \sin \varphi_j + B \cos \varphi_j), \quad (13)$$

$$\partial^2 \rho_j / \partial x_{ij} \partial x_{i'j} = -\frac{4\pi^2}{V} \sum_3 h_i h_{i'} (A \cos \varphi_j + B \sin \varphi_j). \quad (14)$$

The computation of these terms does not materially increase the complexity of the calculation beyond that required for the structure factors alone, because most of the needed individual terms have already been calculated in the structure-factor routine. The chief additional needs are to provide certain additional

product and accumulation commands, and to perform the calculation for the non-unique as well as the unique reflections. The master routine is so arranged that a differential synthesis, a normal Fourier synthesis, or both, may be computed in conjunction with a structure-factor calculation. The sums represented by equations (12)–(14) are punched out at regular intervals as the calculation progresses so that it is not necessary to start over from the beginning if the calculation is interrupted for any reason.

In practice, both the observed and calculated electron densities, and the nine derivatives of each, are evaluated. The calculation begins with the evaluation of the structure factor, which is punched out if desired. Then the products and sums indicated in equations (12)–(14) are formed and the twenty sums for each atom in the asymmetric unit are accumulated progressively. At the conclusion of the summation, a sub-routine solves for the shifts corresponding to both the observed and calculated syntheses, takes their difference for each parameter, and multiplies these differences by the parameter of the 'n-shift rule' (Shoemaker, Donohue, Schomaker & Corey, 1950), which is specified on a variable card. The resulting corrected shifts are next added to the initial parameters on the appropriate drum channels, and finally the revised parameters are punched for record. A new cycle of the calculation may then be started if desired.

The time involved in a differential synthesis can best be expressed relative to that required for a structure-factor calculation alone; even then it is a function of the relative numbers of unique and non-unique reflections, not counting those related by Friedel's law. The calculations for the non-unique reflections involve no punch output and consequently are significantly faster than those for the unique reflections when the structure factors are punched. For orthorhombic space groups, the combined structure-factor calculation and differential synthesis usually takes a little more than twice as long as the structure-factor calculation alone. For a typical hexagonal crystal, lithium perchlorate trihydrate (Prosen & Trueblood, 1956), the combined calculation takes between three and four times as long as the structure-factor calculation alone; there are about seven times as many non-unique as unique reflections.

We do not yet use differential syntheses for semi-automatic analytical refinement of temperature factors; this is done by least-squares methods. Differential syntheses do, however, provide excellent criteria for convergence, since, when all of the position and temperature factor (shape) parameters for an atom are correct, the difference density and its first and second derivatives should vanish.

### Least-squares refinement

Refinement of the scale factor and of atomic-position and temperature-factor parameters by least squares

is a particularly powerful procedure and has the advantage that it utilizes only the independent reflections. Just as with a differential synthesis, most of the terms needed in the calculation are those normally evaluated in a structure-factor calculation; it is necessary only to add appropriate accumulation and product commands, and of course to solve the normal equations which result.

The equations of condition for the least-squares refinement are of the form

$$\sqrt{W} \cdot \left[ \sum_g \sum_i \left\{ \frac{\partial |F|}{\partial x_{ig}} \Delta x_{ig} + \sum_{i'} \frac{\partial |F|}{\partial_{vg} B_{ii'}} \Delta ({}_{vg} B_{ii'}) \right\} + \frac{\partial |F|}{\partial k} \Delta k \right] = \sqrt{W} \cdot \Delta |F|. \quad (15)$$

There are as many of these equations as there are unique reflections. The derivatives which occur in these observational equations, and in the normal equations calculated from them, have the following form:

$$\frac{\partial |F|}{\partial x_{ig}} = \frac{\partial A}{\partial x_{ig}} \cos \alpha + \frac{\partial B}{\partial x_{ig}} \sin \alpha, \quad (16)$$

with an exactly analogous equation for  $\partial |F| / \partial_{vg} B_{ii'}$ . It is not convenient to write general equations for the individual derivatives which are valid for any space group because the relation of the temperature factors of symmetry-equivalent atoms becomes very complex for certain space groups of high symmetry. On the other hand when the symmetry is orthorhombic or lower, the following equations suffice:

$$\frac{\partial A}{\partial x_{ig}} = -2\pi h_i f_g \sum_m s_{mi} \exp[-\beta_{mg}] \sin 2\pi \varphi_{mg}, \quad (17)$$

$$\frac{\partial A}{\partial_{vg} B_{ii'}} = -f_g h_i h_{i'} \sum_m \tau_{mii'} \exp[-\beta_{mg}] \cos 2\pi \varphi_{mg}. \quad (18)$$

Similar expressions occur for the derivatives of  $B$ . Since  $s_{mi}$  and  $\tau_{mii'}$  are merely signs, they can conveniently be represented in the code words used for generation of the parameters of the atoms equivalent by symmetry to those in the asymmetric unit.

The equations of condition are reduced to the normal equations, equal in number to the number of independent parameters to be refined. In principle, these normal equations involve the sums over all  $h_i$  of the squares and products with one another of all of the derivatives of equation (15). It is customary in three-dimensional least-squares refinement of atomic positions in orthogonal crystals to assume that the matrix of the coefficients of the unknown terms is diagonal, that is, that the only terms significantly different from zero in the normal equations are those involving the squares of derivatives. This assumption, which becomes less reasonable if the axial angles depart significantly from  $90^\circ$ , means that only as many sums as there are independent parameters need be accumulated; for example with  $N$  atoms in general positions, one need

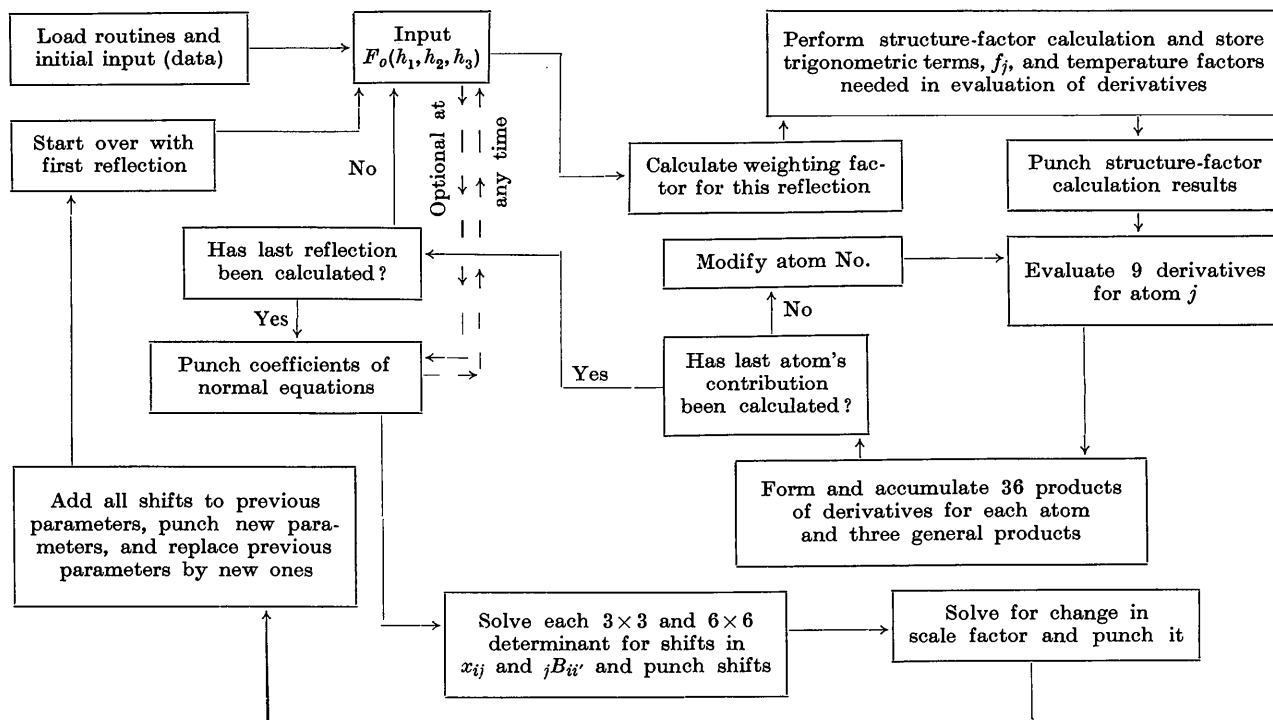
only accumulate the  $3N$  square terms, rather than the  $3N(3N-1)/2$  products which are needed if an exact solution is to be reached. With a large unit cell this represents an enormous saving in computational effort. On the other hand, for non-orthogonal crystals, cross-terms between the different parameters of a given atom may become significant. Since it is possible to calculate some cross-terms without an appreciable increase in computing time, the routine is written to calculate all of the cross-terms among the three position parameters for each atom and all of the cross-terms among the six temperature-factor parameters for each atom; no cross-terms involving interaction of position parameters with temperature-factor parameters, or interaction of any of the parameters of different atoms, are calculated. Even though  $(36N+3)$  products are calculated and accumulated for every reflection, an entire structure-factor least-squares calculation takes only about twice as long as a structure-factor calculation alone.

A description of the course of the calculation is given in the accompanying flow diagram (Table 3). The dashed lines indicate an optional punching of the accumulated products of derivatives which, when the summation is complete, are the coefficients of the normal equations. This punching may be done as often as desired without interfering with the course of the calculation; the purpose of this option is to permit interruption of the calculation at any time. In a short calculation, it is not normally needed; on the other hand in a calculation requiring more than about an hour it is a highly desirable feature, since it is frequently impossible to perform a lengthy calculation continuously, because of either machine failures or scheduling difficulties. If such partial sums are accumulated in several groups, they are loaded into the machine and summed together by a special sub-routine; the resulting normal equations are then solved in the usual fashion. The exact solution of the determinants to give the shifts in the nine parameters for each atom requires only a few seconds per atom, including the punching of the resulting shifts and the revised parameters.

Another option, not specifically indicated in the flow diagram, permits calculation of structure factors for more atoms than are to be refined by least squares. This feature is desirable when, for example, one wishes to include hydrogen atoms in a structure-factor calculation for an organic crystal without refining their positions; it was used in the lengthiest least-squares calculation which we have yet performed, that for dibenzylphosphoric acid (Dunitz & Rollett, 1956). This substance gave more than 2800 independent reflections and had 34 atoms in the asymmetric unit, of which 19 were refined by least squares; one cycle of refinement required between 7 and 8 hr. Actually this refinement was not undertaken until the structure had been extensively refined by other methods. Our more usual practice (with shorter calculations) is to

Table 3. *Flow diagram for least-squares calculations (including recycling)*

Initial input (data): Same as for structure factor calculation



use a number of cycles of refinement by least squares to proceed from a trial structure until the reliability index ( $R$ ) and  $\Sigma W(\Delta|F|)^2$ , which is essentially what is being minimized in the least-squares calculation, have reached or approached a minimum value. In our experience this has required as few as three or as many as ten cycles, depending chiefly upon the adequacy of the initial scale factor and position and temperature-factor parameters. When a refinement is continued until  $R$  and the sum of the residuals approach constancy, the parameter shifts indicated at that stage are always only a small fraction of the corresponding standard deviations.

The least-squares routine has been used extensively for both three-dimensional and two-dimensional refinement. However, it is less desirable for the latter because no provision is made for calculation of any cross terms between atoms, and these become important in unresolved projections. It would be desirable to have a routine which would calculate all the cross terms between two-dimensional position parameters for perhaps ten different atoms, with the more usual treatment of an indefinite number of others. One could specify on a variable card which sort of projection was being refined so that only a single routine would be needed. Such a routine would suffice for most projections and would not involve much more computing time than the present routine. The special  $20 \times 20$  matrix could be inverted and the shifts obtained in less than 15 min.

### Miscellaneous calculations

We have found it desirable to use SWAC for some of the other calculations normally encountered in crystal-structure analysis and have written short routines for some of these special purposes. Each of these routines has one variable card on which the significant particulars of the calculation are specified, and each is arranged so that it may be punched and listed with the standard plugboards which we have wired for the punching and tabulating of the output of our chief routines. The routines are consequently convenient and efficient to use; three of the more useful of them are described here.

#### *Correction factors for Weissenberg intensity data*

The variables for this routine include the wavelength of the radiation used, the lattice constants of the crystal, and whether the crystal is triclinic, monoclinic, or of orthorhombic or higher symmetry. The routine generates the indices of every plane within the positive octant of the sphere of reflection and also, when the crystal is triclinic or monoclinic, of the unique reflections with some indices negative. It then calculates the combined Lorentz-polarization factor,  $1/L$ , and the correction factors,  $T_i$ , for equi-inclination Weissenberg photographs about each axis (Tunell, 1939). The following equations are convenient for calculating these quantities with a high-speed computer:



$$\frac{1}{L} = \sqrt{\left\{ \frac{\sin^2 \theta - \sin^4 \theta}{1 - 2(\sin^2 \theta - \sin^4 \theta)} \right\}}, \quad (19)$$

$$T_i = \sqrt{\left\{ \frac{\cos^2 \mu - \cos^2 \theta}{\sin^2 \theta} \right\}} = \sqrt{\left\{ 1 - \left( \frac{h_i \lambda}{2a_i \sin \theta} \right)^2 \right\}}. \quad (20)$$

The second form of (20) follows from the first because  $\sin \mu = h_i \lambda / 2a_i$ . The punched output consists of one card per plane, and includes the indices of the plane,  $\sin \theta$ ,  $\sin \theta / \lambda$ ,  $1/L$ , and the value of  $T_i/L$  for rotation about each of the axes. About 90 cards are produced per minute, which is very nearly the maximum possible punching speed, since the computing time is a negligible fraction of the punching time. A printed list is then prepared for convenience in processing the intensity data.

#### Interatomic distances and angles

Calculation of interatomic distances and angles becomes very tedious when one is dealing with large molecules; for example there are more than 200 bond angles of interest in the molecule of vitamin B12. A routine was therefore written for this purpose. The variables include the lattice constants, the number of atoms, the trial number, the type of parameter input (which may be decimal or binary, and either in fractions or sixtieths), and the type of output desired, either punch, direct tabulation, or both. The routine and the parameters of the atoms are first stored in the machine; calculation is then effected by the input of a card containing the numbers of either two or three atoms. If there are two numbers on the card, the routine calculates the distance between the atoms specified by those numbers and punches a card containing the atom numbers and the corresponding distance, and also the trial number for later identification. If there are three atom numbers on the card read in, the routine calculates the angle subtended by the first and third atoms at the second, and punches out the atom numbers, the angle, the distance from each of the extreme atoms to the apical atom, and the trial number. The calculation, including input and output, takes only a few seconds for each distance or angle needed. The parameters used may be as large as  $\pm 1.9999$ ; consequently intermolecular distances involving neighboring unit cells may be calculated.

#### Location of maxima on Fourier syntheses

The positions of maximum density of peaks on Fourier syntheses in which the density is normally only sampled at discrete intervals may be obtained reliably on the assumption that the peaks can be represented by general Gaussian ellipsoids (Shoemaker *et al.*, 1950). The 19 points nearest each maximum (Donohue & Trueblood, 1952) are used in a least-squares treatment. The input into the routine consists of cards containing

the values of the density at the 19 sampling points. The output includes the position parameters of the maximum, the density at the maximum, and the values of the density calculated for each of the 19 sampling points from the parameters obtained by least squares for the Gaussian ellipsoid. Comparison of these values with those actually observed gives a measure of the adequacy of representation in this form; in general it is excellent. The computing time is negligible.

It is a pleasure to acknowledge our indebtedness to the staff of Numerical Analysis Research on this campus and in particular to Mr F. H. Hollander and Dr C. B. Tompkins for their cooperation and assistance in making this work possible. We are also grateful for the partial support of this research and of the SWAC by the United States Air Force through the Office of Scientific Research of the Air Research and Development Command, by the United States Navy through the Office of Naval Research, and by the United States Army through the Office of Ordnance Research.

#### References

- AHMED, F. R. & CRUICKSHANK, D. W. J. (1953). *Acta Cryst.* **6**, 765.  
 BENNETT, J. M. & KENDREW, J. C. (1952). *Acta Cryst.* **5**, 109.  
 BOOTH, A. D. (1946). *Trans. Faraday Soc.* **42**, 444.  
 COCHRAN, W. & DOUGLAS, A. S. (1955). *Proc. Roy. Soc. A*, **227**, 486.  
 CRUICKSHANK, D. W. J. (1949). *Acta Cryst.* **2**, 65.  
 DONOHUE, J. & TRUEBLOOD, K. N. (1952). *Acta Cryst.* **5**, 419.  
 DUNITZ, J. D. & ROLLETT, J. S. (1956). *Acta Cryst.* **9**, 327.  
 HODGKIN, D. C., PICKWORTH, J., ROBERTSON, J. H., WHITE, J. G., TRUEBLOOD, K. N. & PROSEN, R. (1955). *Nature, Lond.* **176**, 325.  
 HUSKEY, H. D., THORENSEN, R., AMBROSIO, B. F. & YOWELL, E. C. (1953). *Proc. Inst. Radio Engrs., N.Y.* **41**, 1294.  
 MAYER, S. W. (1953). Ph.D. Thesis, University of California, Los Angeles.  
 MAYER, S. W. & TRUEBLOOD, K. N. (1953). *Acta Cryst.* **6**, 427.  
 ORDWAY, F. D. (1952). *Computing Methods and the Phase Problem in X-ray Crystal Analysis*, pp. 148-165. State College: Department of Physics, Pennsylvania State University.  
 PROSEN, R. J. & TRUEBLOOD, K. N. (1956). To be published.  
 ROLLETT, J. S. & DAVIES, D. R. (1955). *Acta Cryst.* **8**, 125.  
 SHOEMAKER, D. P., DONOHUE, J., SCHOMAKER, V. & COREY, R. B. (1950). *J. Amer. Chem. Soc.* **72**, 2328.  
 THOMPSON, T. R., CAMINER, D. T., FANTL, L., WRIGHT, W. B. & KING, G. S. D. (1954). *Acta Cryst.* **7**, 260.  
 TRUEBLOOD, K. N. (1956). *Acta Cryst.* **9**, 359.  
 TUNELL, G. (1939). *Amer. Min.* **24**, 448.