

the variation in side-chain or main-chain configuration.

The author wishes to express his gratitude to Dr Mark A. Stahmann of the College of Agriculture at the University of Wisconsin for making available a sample of poly- ϵ -carbobenzoxy-L-lysine for use in this investigation and to Profs. Linus Pauling and Robert B. Corey of the California Institute of Technology for their many helpful criticisms.

Acta Cryst. (1953). 6, 727

A Simple but Versatile Strip Technique for Calculating Structure Factors

BY LEROY ALEXANDER

Mellon Institute of Industrial Research, Pittsburgh 13, Pennsylvania, U.S.A.

(Received 6 February 1953 and in revised form 25 March 1953)

Two new sets of strips especially designed for the rapid calculation of structure factors are described. The smaller set, consisting of but 150 basic strips, requires that the parameters be expressed with two-place accuracy, as in the earlier stages of a structure analysis. The second set consists of 1500 basic strips and provides for the calculation of exact structure factors for parameters accurate to 0.001. These new strips overcome several weaknesses inherent in the application of Beevers-Lipson 3° Fourier strips to structure-factor calculations. Either set permits the calculation of structure factors involving terms of the important product type, $\sin 2\pi hx \cdot \sin 2\pi ky$, etc., at the rate of 20–25 per hour in typical cases.

1. Introduction

Beevers & Lipson (1952) have recently shown how a set of standard Fourier strips at 3° intervals (Beevers, 1952) can be applied to the calculation of structure factors. Although their procedure represents a definite step forward in reducing the labor of such calculations, it nevertheless lacks the desired degree of versatility in several respects. The method is chiefly of value in evaluating simple structure factor formulas consisting of terms of the types $\sin 2\pi hx$ or $\cos 2\pi hx$. The frequently occurring formulas involving product terms, such as $\sin 2\pi hx \cdot \sin 2\pi ky$ for example, cannot be handled in such a straightforward manner. Instead, they must be solved by first computing numerically and accurately the sine or cosine functions corresponding to the larger of the two cell dimensions involved, say, the function of y , after which these factors are made the amplitudes of a set of strips used to compute the final sum by way of a summation in x , the parameter corresponding to the shorter cell dimension. This hybrid method of computation is not only cumbersome and inherently unsystematic in character, but it still leads to undesirably large errors if the shorter dimension is much larger than 6 Å. In view of the large unit-cell dimensions encountered in many present-day structural investigations, particularly among organic compounds, this represents a rather severe limitation of the method.

References

- BAMFORD, C. H., BROWN, L., ELLIOTT, A., HANBY, W. E. & TROTTER, I. F. Private communication.
 BAMFORD, C. H., HANBY, W. E. & HAPPEY, F. (1951). *Proc. Roy. Soc. A*, **205**, 47.
 EDSALL, J. T. (1952). *Nature, Lond.* **170**, 53.
 PAULING, L. & COREY, R. B. (1951). *Proc. Nat. Acad. Sci., Wash.* **37**, 241.
 YAKEL, H. L., JR., PAULING, L. & COREY, R. B. (1952). *Nature, Lond.* **169**, 920.

Beevers & Lipson (1952) also describe a modified procedure for increasing the accuracy, but it more than doubles the computational work besides increasing the likelihood of errors. A less serious drawback of both these methods for calculating structure factors is that no strips are available for direct calculations at parameters exceeding 0.25, with the result that special rules must be observed governing the choice of strips and changes in sign in the range 0.25–1.00. The several objections just cited combine to make the computational procedures far from routine. Surely a high degree of routineness should be a feature of any really valuable aid to structure factor calculations, because the probability of errors occurring is more or less proportional to the extent to which the human element must be reckoned with in the operations.

2. New structure-factor strips

Most of the above objections stem from the fact that the Beevers-Lipson method seeks to apply standard Fourier strips to the solution of a problem for which they were not primarily designed. To be specific, Fourier strips must bear values of $A \sin n\theta$ and $A \cos n\theta$ corresponding to variations in $n\theta$ of from 0 to $\pi/2$ radians and to variations in h of from 1 to at least 30, which means that a net range in the argument $n\theta$ of about 0 to 15π radians is needed. For 6°

strips the minimum angular increment between numbers on any given strip occurs on a strip for which $h = 1$ and amounts to $\pi/(2 \times 15) = 0.033\pi$; for 3° strips it is $\pi/(2 \times 30) = 0.016\pi$. On the other hand, for directness and convenience in the calculation of structure factors the strips should bear values of $\sin 2\pi hx$ and $\cos 2\pi hx$ corresponding to variations in x of from 0 to 1 and in h of from 0 to at least 20, which means that a net range in the argument $2\pi hx$ of at least 0 to 40π radians is required. Furthermore, the attainment of the necessary accuracy requires that the angular increment between successive numbers on a strip should be no larger than that corresponding to $\Delta x = 0.001$, or consequently $\Delta(2\pi x) = 0.002\pi$, which is much smaller than the smallest angular steps of 0.033π and 0.016π on the 6° and 3° Fourier strips respectively. A further point of difference in the requirements of the two applications is that separate sets of Fourier strips are required for amplitudes ranging from at least -99 to $+99$, whereas only a single common amplitude is needful in structure-factor calculations according to the method to be described. It is evident, then, that less than 1% of the strips in a standard Fourier set would be utilized in the calculation of structure factors.

It seemed to the present author that the several important points of difference between the requirements of Fourier summations on the one hand and structure-factor calculations on the other were sufficient grounds for the construction of a new set of strips based on a numerical scheme similar to that of the Beevers-Lipson Fourier strips but specifically designed to meet the altered requirements. Such strips have been constructed in this laboratory, the first trial set being designed for approximate calculations and consisting basically of but 100 sine and 50 cosine strips (not allowing for the occasional need for two or more identical strips) for parameters from 0.01 to 1.00 by steps of 0.01. Such a set can be made up by hand in a few hours' time, and, with the aid of any modern calculating machine capable of accumulating products, it enables the user to compute trial structure factors involving sums of products of the types $\cos 2\pi hx \cdot \cos 2\pi ky$, $\sin 2\pi hx \cdot \sin 2\pi ky$, or $\cos 2\pi hx \cdot \sin 2\pi ky$ at the rate of 20-25 per hour even when two or three kinds of atoms are included. Structure factors involving sums of terms of the type $\sin 2\pi hx \cdot \sin 2\pi ky \cdot \sin 2\pi lz$ can evidently not be computed with equal speed; not only are more mathematical operations involved, but the triple products cannot be accumulated automatically by most calculating machines.* This set of 150 basic

strips will be hereafter designated the *small* set to distinguish it from the *large* set of 1500 basic strips for accurate calculations, to be described below.

The small set of strips can be of surprising utility in the early stages of an analysis when trial structures are being tested for approximate agreement between F_o and F_c . The rounding-off of parameters to two places, which is mandatory in applying this set of strips, results in the frequent creation of serious errors in F_c at the larger indices; hence it is not desirable to extend the number of values of $\sin 2\pi hx$ or $\cos 2\pi hx$ on each strip beyond $h = 8$. Furthermore, two-place accuracy in these numbers is adequate for either the small or large sets of strips for the reason that this permits the attainment of a precision in F_c considerably greater than the accuracy with which the scattering factors are known or that with which F_o can be measured.

For the refinement of lattice parameters, however, it is required that the increment in x be not larger than 0.001, as pointed out earlier. This necessitates the preparation of 1000 sine and 500 cosine strips for the basic set, or possibly 4500 strips in a working set if allowance is made for the need of as many as three strips of a kind. The strips of this large set must be longer in order to provide space for numbers corresponding to h values from 1 to 20. Extending the range to $h = 30$ is not justifiable without a simultaneous improvement of the discrimination in x , say, from 0.001 to 0.0005. Such longer strips are, in addition, more awkward to handle and store.

3. Constructional details

The construction of the strips and a typical structure-factor calculation will be described in terms of the small set of strips. The numerical arrangement is the same as on a Beevers-Lipson Fourier strip except that both the odd and even values of n in $\sin nh\theta$ or

C $\frac{10}{90}$	81	31	(31)	(81)	(1)	(81)	(31)	31
C $\frac{25}{75}$	0	(1)	0	1	0	(1)	0	1
C $\frac{46}{54}$	(97)	88	(73)	54	(31)	06	19	(43)
S 10	59	95	95	59	0	(59)	(95)	(95)
S 25	1	0	(1)	0	1	0	(1)	0
S 46	25	(46)	69	(84)	95	(1)	98	(91)

Fig. 1.

$\cos nh\theta$ appear on one side of the strip. There is a different sine strip for each value of x from 0.01 to

$$\frac{\sin 2\pi hx \cdot \sin 2\pi ky}{1/\sin 2\pi lz},$$

which requires only three entries. The amount of time saved by this device varies with the particular calculating machine used.

* Some calculating machines permit the computation of a three-factor product with only three entries into the machine, while others require four. The referee has pointed out that with machines of the latter type the number of entries can be reduced to three if the backs of the strips bear the reciprocals of the quantities appearing on the front sides. Then, for example, the product $\sin 2\pi hx \cdot \sin 2\pi ky \cdot \sin 2\pi lz$ can be computed by means of the operation

	$k = 1 \ 2 \ 3 \ 4 \ 5 \ 6 \ 7 \ 8$								$l = 1 \ 2 \ 3 \ 4 \ 5 \ 6 \ 7 \ 8$									
C_2	S·49	06	(13)	19	(25)	31	(37)	43	(48)	S·08	48	84	1	91	59	13	(37)	(77)
C_4	S·50	0	0	0	0	0	0	0	0	S·11	64	98	88	37	(31)	(84)	(99)	(69)
C_5	S·67	(88)	84	06	(91)	81	13	(93)	77	S·02	13	25	37	48	59	69	77	84
C_6	S·75	(1)	0	1	0	(1)	0	1	0	S·96	(25)	(48)	(69)	(84)	(95)	(1)	(98)	(91)
O_2	S·41	54	(91)	99	(77)	31	25	(73)	98	S·11	64	98	88	37	(31)	(84)	(99)	(69)
O_4	S·43	43	(77)	97	(98)	81	(48)	06	37	S·17	88	84	(06)	(91)	(81)	13	93	77
O_5	S·75	(1)	0	1	0	(1)	0	1	0	S1·00	0	0	0	0	0	0	0	0
O_6	S·90	(59)	(95)	(95)	(59)	0	59	95	95	S·87	(73)	(1)	(64)	13	81	98	54	(25)
O_7	S·63	(73)	1	(64)	(13)	81	(98)	54	25	S·19	93	69	(43)	(1)	(31)	77	88	(13)
N_1	S·66	(84)	91	(13)	(77)	95	(25)	(69)	98	S·99	(06)	(13)	(19)	(25)	(31)	(37)	(43)	(48)
N_3	S·41	54	(91)	99	(77)	31	25	(73)	98	S·14	77	98	48	(37)	(95)	(84)	(13)	69

Fig. 2.

1·00, and the successive numbers from left to right on a strip are the two-place values of $\sin 2\pi hx$ corresponding to $h = 1, 2, 3, 4, 5, 6, 7$ and 8 . The cosine strips are made up similarly except that only 50 are required in the basic set because of the symmetry of $\cos 2\pi hx$ about $x = 0$. Thus one strip serves for $x = 0\cdot01$ and $0\cdot99$, another for $0\cdot02$ and $0\cdot98$, and so on. Fortunately it is not necessary to compute $\sin 2\pi hx$ and $\cos 2\pi hx$ for the many combinations of h and x values; these quantities can be transcribed directly from tables published some years ago by Buerger (1941). Typical sine and cosine strips are shown in Fig. 1. The letter at the left end of each strip identifies it as sine (S) or cosine (C), and the small figures following the letter specify the parameter. Circled quantities are negative. A decimal point is understood to precede each number with the exception of 1 and $\bar{1}$, which represent 100 and $\bar{100}$ and are written in the abbreviated form in the interests of neatness and space economy. Possible confusion with $0\cdot01$ and $0\cdot0\bar{1}$ is avoided since these quantities are written as 01, respectively uncircled and circled. Convenient dimensions for the strips of the small set, which may be cut from light cardboard, are $\frac{1}{2} \times 5\frac{1}{2}$ in.

4. Method of calculation and example

The calculative procedure can be best understood by referring to an actual example. For this purpose an organic structure now under investigation in this laboratory will be used. The space-group symmetry is $C_{2h}^5-P2_1/c$, and the unit cell contains four molecules comprised of 4 carbon, 5 oxygen, 2 nitrogen and 6 hydrogen atoms. Disregarding the hydrogen atoms, all the other atoms occupy sets of general 4(e) positions, making 33 parameters to be determined. The general structure-factor expression for this space group symmetry involves a trigonometric component,

$$A = 4 \cos 2\pi \{hx + lz + \frac{1}{4}(k+l)\} \cos 2\pi \{ky - \frac{1}{4}(k+l)\}, \quad (1)$$

which, for $(0kl)$ reflections with $k+l$ odd, simplifies to the form

$$A = -4 \sin 2\pi ky \cdot \sin 2\pi lz. \quad (2)$$

The final structure-factor formula for this case is then

$$F(0kl) = \sum_n f_n A_n = -4 \sum_n f_n \sin 2\pi ky_n \cdot \sin 2\pi lz_n, \quad (3)$$

$k+l$ being odd. Similarly for $(hk0)$ reflections with k odd, the formula is

$$F(hk0) = -4 \sum_n f_n \sin 2\pi hx_n \cdot \sin 2\pi ky_n. \quad (4)$$

Suppose the parameters to be tested are those given in Table 1. Consider the computation of the structure factors of $(0kl)$ reflections with $k+l = 2n+1$. The small set of strips is used by first rounding off each y and z parameter to two places and then selecting the proper sine strips and assembling them in two banks placed side by side, one for y and one for z parameters, as illustrated in Fig. 2. To compute $F(032)$, for example, one employs the calculating machine to multiply the first number appearing in the

Table 1. Trial parameters

Atom	x	y	z
C_2	0·456	0·492	0·078
C_4	0·322	0·500	0·111
C_5	0·210	0·668	0·022
C_6	0·216	0·745	0·962
O_2	0·587	0·410	0·106
O_4	0·305	0·430	0·166
O_5	0·075	0·750	0·995
O_6	0·110	0·896	0·867
O_7	0·120	0·625	0·188
N_1	0·348	0·656	0·992
N_3	0·447	0·410	0·140

No. 3 column of the y strips, 0·19, by the first number appearing in the No. 2 column of the z strips, 0·84. Next one multiplies the second pair of numbers in the same two columns and adds this product to the

first, and so on until one has accumulated the four products of the carbon atoms. This partial total is written down, and then the oxygen and nitrogen cross-products are summed and the two totals written down. The calculator is next used to multiply the trigonometric sum for each kind of atom by the proper scattering factor, the total of these three products being taken. This grand sum for all the atoms is finally multiplied by -4 to yield the structure factor, $F(032)$.

An approximate but speedier way of allowing for the differences in scattering power of the three groups of atoms is to take advantage of the fact that the ratios f_C/f_O and f_N/f_O are nearly constant over a considerable angular range and equal respectively to about 0.6 and 0.8. When this plan is used, the trigonometric sum for carbon is multiplied by 0.6 rather than f_C and the sum for nitrogen by 0.8 instead of f_N . These two quantities are added to the trigonometric sum for oxygen, and this grand sum is multiplied by the factor $-4f_O$ to give the structure factor. This procedure saves time in the early stages of a structure analysis involving carbon, nitrogen and oxygen by eliminating the need for taking account of the variations in f_C and f_N with $(\sin \theta)/\lambda$. It should be observed, however, that this approximate method is less satisfactory when the structure comprises both light and heavy atoms.

The order in which the computations are performed on the calculating machine is now illustrated by means of the planes (032). Preliminary data available:

$$(\sin \theta)/\lambda = 0.410. \quad f_C = 1.87, f_O = 2.82, f_N = 2.25.$$

Step 1. Carbon, $(.19 \times .84) + (0 \times .98) + (.06 \times .25) + (1 \times .48) = -0.31$.

Step 2. Oxygen, $(.99 \times .98) + (.97 \times .84) + (1 \times 0) + (.95 \times 1) + (.64 \times .69) = +2.29$.

Step 3. Nitrogen, $(.13 \times .13) + (.99 \times .98) = +0.99$.

Step 4. $(-0.31 \times 1.87) + (2.29 \times 2.82) + (0.99 \times 2.25) = +8.11$.

Step 5. $F(032) = -4 \times 8.11 = -32.4$.

In case the simplification is adopted that $f_C/f_O = 0.6$ and $f_N/f_O = 0.8$, steps 4 and 5 are replaced by:

Step 4'. $(-0.31 \times 0.6) + (2.29 \times 1.0) + (0.99 \times 0.8) = +2.89$.

Step 5'. $F(032) = -4 \times 2.82 \times 2.89 = -32.6$.

5. Evaluation of the method

Once the necessary strips have been selected and arranged and the values of f_C , f_O and f_N obtained from a graph or table, the calculations for a given reflection of the type shown above can be performed within two minutes. It should be emphasized that in the present method a minimum of time is consumed in doing mental and paper arithmetic; of the arithmetic given above, only the quantities 1.87, 2.82, 2.25, -0.31 , $+2.29$, $+0.99$ and -32.4 need be written down. The calculating machine does the rest.

The large set of strips can be handled in exactly

the same way and with nearly the same speed, except that the greater length of the strips makes them somewhat less convenient to handle. For this reason it is advantageous to make them somewhat narrower than the strips of the small set, which also has the effect of reducing the needed length. Dimensions of $\frac{3}{8} \times 8$ in. are suggested. The author has used a rack of the type described by Patterson & Tunell (1942) for holding the strips of the small set in position during the computations, but a smaller more easily handled mounting frame would be a distinct advantage. The most flexible scheme would make use of six such frames, three for $\sin 2\pi hx$, $\sin 2\pi ky$ and $\sin 2\pi lz$, and three for the corresponding cosines. The strips could be initially set up on the six frames for a set of parameters to be tested, after which the structure factors for all types of reflections could be calculated by choosing the proper pairs or trios of frames. After the two or three sets of strips have been arranged side-by-side for the calculations (see Fig. 2), it has been found helpful to cover all the numbers except the particular columns represented in the index triplet by means of vertically slotted cardboard masks, and to employ a horizontal straight-edge to avoid errors in selecting the proper pair or trio of numbers to be multiplied from the two or three sets of strips. When using the longer strips of the large set this latter expedient can hardly be dispensed with. Storage of the strips in a manner providing for utmost convenience in their selection and replacement after use is absolutely necessary to the success of this or any other strip method. Boxes with sloping sides and containing a separate hole for strips of each parameter can be

Table 2. Values of 18 structure factors as computed by different methods

Method A. Strips for parameters to 0.01;
 $f_C/f_O = 0.6$ and $f_N/f_O = 0.8$ assumed.
Method B. Strips for parameters to 0.01;
Hartree values of f_C , f_N , and f_O used.
Method C. Strips for parameters to 0.001;
Hartree values of f_C , f_N , and f_O used.

<i>(hkl)</i>	Calculated values of <i>F</i>		
	<i>A</i>	<i>B</i>	<i>C</i>
210	47.0	46.6	46.7
310	35.4	36.6	37.0
410	6.3	7.1	7.4
510	-5.0	-4.8	-4.0
610	16.9	17.0	18.6
810	-22.0	-24.4	-24.5
230	7.5	7.3	7.3
530	-18.2	-19.8	-21.9
630	-1.7	-2.4	-4.8
830	13.2	16.0	10.9
012	-41.3	-41.0	-41.9
014	-15.9	-17.8	-19.0
021	3.7	3.4	3.2
023	15.5	15.4	15.7
032	-32.6	-32.4	-31.0
036	32.7	33.9	36.3
041	13.1	13.9	12.8
047	4.4	5.1	3.2

constructed according to the general designs suggested by Lipson & Beevers (1936), Beevers & Lipson (1952) or Patterson & Tunell (1942).

Table 2 compares the numerical values of 18 structure factors as computed with strips accurate to 0.01 and to 0.001. The parameters are again those given in Table 1, and the formulas for $F(0kl)$ and $F(hk0)$ with $k+l$ odd are given by equations (3) and (4). Columns *B* and *C* compare directly the approximate numerical values derived with the small set of strips with the exact values computed using the large set for parameters accurate to three places. In view of the anticipated discrepancies at larger indices, the quality of the agreement is rather surprising. Even for the planes (630), (830) and (047) the errors need not cause concern save in the final stages of the parameter-refining process. The only logical explanation for this unexpectedly good agreement at the higher indices is that the errors in the trigonometric products for the

various atoms, being random in sign and magnitude, tend to cancel each other on the average. For this reason the small set of strips may have a special value in solving complex structures involving many parameters. The lack of significant differences between the values in columns *A* and *B* demonstrates that in the earlier stages of parameter refinement one may assume the constancy of f_c/f_0 and f_N/f_0 with impunity.

References

- BEEVERS, C. A. (1952). *Acta Cryst.* **5**, 670.
 BEEVERS, C. A. & LIPSON, H. (1952). *Acta Cryst.* **5**, 673.
 BUERGER, M. J. (1941). *Numerical Structure Factor Tables*. Geological Society of America, Special Paper No. 33.
 LIPSON, H. & BEEVERS, C. A. (1936). *Proc. Phys. Soc.* **48**, 772.
 PATTERSON, A. L. & TUNELL, G. (1942). *Amer. Min.* **27**, 655.

Acta Cryst. (1953). **6**, 731

The Meaning of the Average of $|F|^2$ for Large Values of the Interplanar Spacing*

BY DAVID HARKER

Polytechnic Institute of Brooklyn, Brooklyn 1, N.Y., U.S.A.

(Received 13 January 1953)

It is shown that A. J. C. Wilson's equation, $\langle I \rangle = \Sigma f^2$, is valid only if $d(hkl)$ is smaller than the smallest interatomic distance in the crystal. The correct equation for $\langle I \rangle$ for use at large d 's is derived. It is shown that the equation, $\langle I \rangle = \Sigma F_g^2$, where F_g is the scattering power of a glob of atoms, is nearly correct for d 's larger than twice the shortest interatomic distance and smaller than the shortest distance between centers of globs. This last statement is most nearly true for globs that consist of 'spherical tops of scattering matter', i.e. whose second moment of scattering is the same for all directions.

Introduction

In 1942, Wilson (1942) announced the discovery of a relation between the average of the squared magnitudes of the structure factors of a crystal and the atomic scattering factors. This relationship has the form:

$$\langle I \rangle = \sum_{j=1}^N f_j^2(s). \quad (1)$$

In (1), $\langle I \rangle$ is the average value of $|F(hkl)|^2$ for all values of h , k and l for which $s = 1/d(hkl)$ lies between s and $s+ds$, N is the number of atoms in one unit cell of the crystal and $f_j(s)$ is the scattering factor of the j th atom. In a later paper Wilson (1949) pointed out that in practical cases relation (1) is not valid at small values of s (large values of $d(hkl)$) and suggested a limiting value of s beyond which (1) could be safely

applied. In the present paper, the interpretation of $\langle I \rangle$ is discussed for the small values of s which can occur in the data from crystals with large unit cells.

In the following paragraphs, the exact equation corresponding to (1) will be derived and will be found to contain additional terms which become larger as the product sr_{ij} decreases (where r_{ij} is the distance between the i th and j th atoms), but which are unimportant for large values of sr_{ij} . It will then be shown—qualitatively—that equation (1) is generally valid only if the minimum value of sr_{ij} is unity. Consequently, equation (1) is useful only if, for each of the structure factors used in the averaging, $d(hkl)$ is less than about 1.5 Å for organic crystals, or about 2 Å for most others; for larger values of $d(hkl)$ equation (1) can be quite inaccurate and is therefore useless. *In particular, most crystalline proteins give diffraction effects only for $d(hkl)$ greater than 1.5 Å and data from such crystals should never be interpreted by the use of equation (1).*

* Contribution No. 1 from the Protein Structure Project.