interfaces. Various tensor quantities such as distortion, torsion and curvature have been developed in terms of the Burgers circuit and are shown to give what is perhaps the most complete description available until now of the Burgers vector and dislocation density associated with such two-phase interfaces.

The present research effort was carried out at the Institut für Theoretische und Angewandte Physik der Universität Stuttgart, Germany (BRD) under a Senior US Scientist Award presented to the author by The Alexander von Humboldt Stiftung in conjunction with a one-year sabbatical leave. Financial support for the present study was also provided in part by the National Science Foundation under Grant No. DMR-7202944. The author is indebted to Professor Ekkehart Kröner for his kind assistance during the course of this investigation.

References

BILBY, B. A., GARDNER, L. R. T., GRINBERG, A. & ZORAWSKI, M. (1966). *Proc. Roy. Soc.* A**292**, 105–121.

FUNG, Y. C. (1965). *Foundations of Solid Mechanics.* New Jersey: Prentice Hall.

KONDO, K. (1955). *Memoirs of The Unifying Study of The Basic Problems in Engineering Sciences by Means of Geometry* pp. 458–469. Tokyo: Gakujutsu Bunken Fukyukai.

KONDO, K. (1962). *RAAG Memoirs*, Vol. III, edited by K. KONDO pp. 173–184. Tokyo: Gakujutsu Bunken Fukyukai.

KRÖNER, E. (1958). *Kontinuumstheorie Der Versetzungen Und Eigenspannungen.* Berlin: Springer.

KRÖNER, E. (1959). *Arch. Rational Mech. Anal.* **4**, 273–334.

MARCINKOWSKI, M. J., SADANANDA, K. & CULLEN, W. H. JR (1975). *Acta Cryst.* A**31**, 292–299.

SCHOUTEN, J. A. (1951). *Tensor Analysis For Physicists.* Oxford: Clarendon Press.

SCHOUTEN, J. A. (1954). *Ricci-Calculus.* Berlin: Springer.

SCHRÖDINGER, E. (1954). *Space-Time Structure.* Cambridge Univ. Press.

WIT, R. DE (1973). *J. Res. Natl. Bur. Stand., A Phys. Chem.* **77A**, 607–658.

ZORAWSKI, M. (1967). *Théorie Mathématique des Dislocations.* Paris: Dunod.

Acta Cryst. (1977). A**33**, 486–492

Efficient Structure-Factor Calculation for Large Molecules by the Fast Fourier Transform

BY LYNN F. TEN EYCK

Institute of Molecular Biology, University of Oregon, Eugene, Oregon 97403, *USA*

(*Received* 2 *August* 1976; *accepted* 10 *November* 1976)

A method is presented for calculating structure factors by Fourier inversion of a model electron density map. The cost of this method and of the standard methods are analyzed as a function of number of atoms, resolution, and complexity of space group. The cost functions were scaled together by timing both methods on the same problem, with the same computer. The FFT method is $3\frac{1}{2}$ to 7 times less expensive than conventional methods for non-centrosymmetric space groups.

Structure factor calculation is one of the major expenses in refinement of macromolecular structures. The purpose of this paper is to show how the fast Fourier transform (FFT) method can be used to calculate structure factors, and to compare FFT structure-factor calculation with conventional methods in terms of convenience, speed, and economy. A specific comparison of the two methods for the space group $P6_122$ has been carried out as part of a project in this laboratory to refine the structure of thermolysin, a proteolytic enzyme from *Bacillus thermoproteolyticus* (Matthews, Weaver & Kester, 1974). It will be shown that the FFT structure-factor calculation is considerably less expensive than any of the direct-summation methods.

A. Calculation of structure factors by the FFT method

The fast Fourier transform algorithm can only compute finite discrete Fourier transforms. Therefore we must construct a model electron density map sampled on a grid and invert it to obtain our structure factors. The calculation of such a model map presents two problems which must be solved in order to use this method. First, a procedure must be devised to solve the sampling problem; otherwise the model electron density map must be constructed on a very fine grid, which greatly increases the computer storage requirements and the cost of the calculation. Second, a convenient functional representation must be found for the electron density distribution of a single atom. Sayre (1951) discussed these problems in some detail. In the following sections we describe ways of solving these problems, and also discuss programming strategy.

A.1. *The sampling problem*

The FFT algorithm can only be used to calculate the *finite discrete* Fourier transform of a function sampled at regular intervals on a grid. The act of sampling a continuous function implies potential loss of information. Specifically, if a function is sampled at N evenly spaced grid points there is only enough information present to determine N Fourier coefficients.* The Fourier series representation of a continuous function is in general an infinite series. In the particular case where that series terminates after N or fewer terms a finite discrete Fourier transform based on N evenly spaced samples will give the same results as the integral Fourier transform. Unfortunately the Fourier series representation of a crystal structure does not terminate. (The fact that the observed data are limited in resolution is a separate problem.)

When the Fourier series representation of a function requires more coefficients than are provided by the sampling grid, a finite discrete Fourier transform will produce N Fourier coefficients which will exactly reproduce the sampled function at the grid points but will not in general reproduce the function at other points in space. It is well known that structure factors calculated by an FFT program will be the sum of the desired structure factors and those spaced at multiples of N away from the desired structure factors. [For discussion see, for example, Sayre (1951) or Brillouin (1962).] The proof that follows is intended primarily to introduce notation and definitions.

Let $\varrho(x)$ be a one-dimensional electron density distribution which repeats with a period $\mathbf{a}$. Then $\varrho(x)$ has a complex Fourier series expansion

$$\varrho(x)=\frac{1}{\mathbf{a}}\sum_{h=-\infty}^{+\infty}\mathbf{F}(h)\exp\left(-2\pi ihx/\mathbf{a}\right) \quad (1)$$

where

$$\mathbf{F}(h)=\int_0^{\mathbf{a}}\varrho(x)\exp\left(2\pi ihx/\mathbf{a}\right)\mathrm{d}x\,. \quad (2)$$

If we now sample $\varrho(x)$ at N equally spaced points over one complete period and calculate the finite discrete Fourier transform we obtain $\mathbf{F}'(h)$, as follows:

$$\begin{aligned}\mathbf{F}'(h)&=\frac{\mathbf{a}}{N}\sum_{j=0}^{N-1}\varrho(j\mathbf{a}/N)\exp\left(2\pi ihj/N\right)\\&=\frac{\mathbf{a}}{N}\sum_{j=0}^{N-1}\left[\frac{1}{\mathbf{a}}\sum_{k=-\infty}^{+\infty}\mathbf{F}(k)\right.\\&\qquad\left.\times\exp\left(-2\pi ikj/N\right)\right]\exp\left(2\pi ihj/N\right)\\&=\frac{1}{N}\sum_{k=-\infty}^{+\infty}\mathbf{F}(k)\sum_{j=0}^{N-1}\exp\left[2\pi i(h-k)j/N\right]. \quad (3)\end{aligned}$$

The inner sum in (3) is clearly zero unless $h-k=pN$ with p an integer; in that case it is N. Therefore

$$F'(h)=\sum_{p=-\infty}^{+\infty}\mathbf{F}(h+pN)\,. \quad (4)$$

For most practical crystallographic purposes the structure factors fall off with resolution rapidly enough that the only significant contamination is from $\mathbf{F}(h\pm N)$, but this contamination can be serious. The following procedure can be used to reduce the contamination by higher-order structure factors to any desired level.

Since a crystal is composed of atoms the structure factor can be written as

$$\mathbf{F}(h)=\sum_{j=0}^{n}f_j\exp\left[-B_j(\sin^2\theta)/\lambda^2\right]\exp\left(2\pi ihx/\mathbf{a}\right) \quad (5)$$

where f_j is the atomic scattering factor for atom j, B_j is the isotropic thermal factor for atom j, θ is the scattering angle, and λ is the wavelength of the X-rays. Now let a constant B° be added to all of the thermal factors and calculate $\mathbf{F}_B(h)$, noting that $(\sin\theta)/\lambda=h/2\mathbf{a}$.

$$\mathbf{F}_B(h)=\exp\left(-B^\circ h^2/4\mathbf{a}^2\right)\mathbf{F}(h) \quad (6)$$

$$\mathbf{F}_B(h+N)=\exp\left[-B^\circ(h+N)^2/4\mathbf{a}^2\right]\mathbf{F}(h+N) \quad (7)$$

$$\mathbf{F}_B(h-N)=\exp\left[-B^\circ(h-N)^2/4\mathbf{a}^2\right]\mathbf{F}(h-N)\,. \quad (8)$$

Adding (6), (7) and (8) and multiplying by $\exp\left(B^\circ h^2/4\mathbf{a}^2\right)$ gives, from (4),

$$\begin{aligned}\mathbf{F}'_B(h)\exp\left(B^\circ h^2/4\mathbf{a}^2\right)&\simeq\mathbf{F}(h)\\&+\exp\left[-B^\circ N(N+2h)/4\mathbf{a}^2\right]\mathbf{F}(h+N)\\&+\exp\left[-B^\circ N(N-2h)/4\mathbf{a}^2\right]\mathbf{F}(h-N) \quad (9)\end{aligned}$$

where both of the exponentials on the right-hand side can be made as small as desired by an appropriate choice of B°.

As a numerical example let us take $\mathbf{a}=50$ Å, $N=50$, $h=20$ (2·5 Å resolution, typical of much protein work), and $B^\circ=15$ Å^2. The structure factor $\mathbf{F}(20)$ will be contaminated by $\mathbf{F}(70)$, which is at 0·7 Å resolution and probably is very small, and by $\mathbf{F}(-30)$, which is at 1·67 Å resolution and could be a problem if high accuracy is required. From equation (9) we have

$$1{\cdot}822\mathbf{F}'_B(20)=\mathbf{F}(20)+0{\cdot}472\mathbf{F}(-30)+0{\cdot}001\mathbf{F}(70)\,.$$

Since the mean structure amplitude at 1·67 Å would be expected to be about 50–60% of the mean amplitude at 2·5 Å (assuming a mean atomic B of 10 Å^2) the error in $\mathbf{F}(20)$ would be of the order of 25%, which is probably too large to tolerate. If we take a finer grid the situation looks much better, and we do not have to use such a large B°. With $N=75$ the worst contamination is from $\mathbf{F}(-55)$, which is at 0·9 Å resolution and would be expected to be about 7% of $\mathbf{F}(20)$. A B° value of 5 Å^2 then gives

$$1{\cdot}221\mathbf{F}'_B(20)=\mathbf{F}(20)+0{\cdot}270\mathbf{F}(-55)+0{\cdot}013\mathbf{F}(95)$$

which reduces the error in the 2·5 Å data to about 2%.

For most protein work a grid size of roughly three times the maximum index of the data in each direction seems appropriate for structure-factor calculation, as recommended by Lipson & Cochran (1966).

---

* If the function is real instead of complex the Fourier transform will contain $N/2$ *unique* complex numbers, but a complex number carrys twice as much information as a real number.

This is not a hard and fast rule. In particular there is *no* reasonable value of B° which would let one calculate 6 Å resolution structure factors with a grid spacing near 2 Å. In that case the first contaminant would be a 3 Å structure factor and the damping factor in (9) would be only $\exp(-B^\circ/48)$.

Sayre (1951) considered the possibility of using atoms with large temperature factors in constructing the model density function and concluded that it had two serious disadvantages. They were (*a*) that the structure factors so calculated cannot be directly compared with the observed structure factors, and (*b*) that the high-resolution structure factors so calculated are so small that they are rendered uncertain by rounding-off errors. However, if the *same* large B° is added to the temperature factors of all of the atoms it is a simple matter to re-scale the results, which eliminates the first objection. The second problem can be more serious, depending on the values of B° and N. An error analysis by Gentleman & Sande (1966) shows that the FFT method is extremely stable with respect to round-off errors. It is very unlikely that round-off error would be a problem on any modern computer, but in case of doubt the formulae of Gentleman & Sande will give limiting values for that source of error for any computer word length.

The method proposed (and used in an example) by Sayre for overcoming the sampling problem was to construct the model electron density map from a representation of the atoms which contained no Fourier coefficients of order higher than $\frac{1}{2}N$. This method has the advantage that the finite discrete Fourier transform gives analytically correct results without the necessity for modifying the model in any way. The disadvantage of this method is the necessity for calculating the profile of each type of atom in the structure for each distinct value of the atomic temperature factor.

A.2. *Electron density distributions for model atoms*

For medium-resolution work, say out to 2·5 Å, a simple Gaussian approximation for the electron density distribution of an individual atom is adequate. The form factor for C in this range is roughly equivalent to a point atom with a thermal factor of about 10 Å^2. The real thermal factor should be added to this pseudo-thermal factor to obtain a B', and the electron density can be calculated as

$$\varrho(x)=\frac{Z}{\sigma^3}\exp(-\pi r^2/\sigma^2) \qquad (10)$$

where $\sigma^2=B'/4\pi$, r is the distance in Å from the center of the atom to the grid point x, and Z is the number of electrons in the atom.

At resolution greater than 2 Å the atomic scattering factors deviate substantially from Gaussian form. An appropriate model electron density function can be obtained in several ways. One approach would be to use the electronic distributions calculated from quantum mechanical treatments of the various atom types, but this is laborious and presents problems with the thermal parameters. (The model electron density is the convolution of the electron density for an atom at rest with a Gaussian function which describes the thermal motion.) A second approach would be to calculate the Fourier transforms of the scattering factor curves and interpolate as was done by Sayre (1951), but this procedure also presents problems with respect to individual atomic thermal factors.

The approach suggested here is to use an appropriate analytical approximation to the scattering factor curves and calculate the analytical Fourier transform of the approximation. Vand, Eiland & Pepinsky (1957) noted that the scattering factors can be approximated quite well by a function of the form

$$f(x)=P\exp(-px^2)+Q\exp(-qx^2)+R \qquad (11)$$

where $x=\sin\theta$. Forsyth & Wells (1959) modified (11) by taking x as $(\sin\theta)/\lambda$ and calculated tables of the coefficients for most types of atoms. The r.m.s. error in these approximations is generally a fraction of 1% of the number of electrons in the atom.

Equation (11) is in a particularly useful form for incorporating thermal factors since multiplication throughout by $\exp(-Bx^2)$ does not change the functional form of either the approximation or its Fourier transform. The final form for our model electron density function is a sum of three Gaussian functions,

$$\varrho(r)=\frac{P}{\sigma_1^3}\exp(-\pi r^2/\sigma_1^2)+\frac{Q}{\sigma_2^3}\exp(-\pi r^2/\sigma_2^2)$$
$$+\frac{R}{\sigma_3^3}\exp(-\pi r^2/\sigma_3^2) \qquad (12)$$

where $\sigma_1^2=(p+B)/4\pi$, $\sigma_2^2=(q+B)/4\pi$, and $\sigma_3^2=B/4\pi$.

A.3. *Organization of the calculation*

This section of the paper describes the design of a program which has been used to calculate structure factors for thermolysin, a protein containing 2437 non-hydrogen atoms in the asymmetric unit. The space group is $P6_122$, which has 12 equivalent positions and very complicated structure-factor formulae. Calculation of $P6_122$ structure factors by FFT methods takes only slightly longer than a $P1$ FFT structure-factor calculation for the same asymmetric unit.

The first step is to select all of those atoms in the structure which will contribute to the electron density in the crystallographic asymmetric unit. (This list is not the same as the list of all unique atoms in the structure because atoms near the edge of the asymmetric unit will contribute density in more than one place.) For this purpose a maximum atomic radius, defined as the radius at which the electron density is $\exp(-4\pi)$ (*i.e.* $3{\cdot}5\times10^{-6}$) times its maximum value, is calculated for each atom as it is read into the computer. The atom is then placed in all of the equivalent positions and all of the neighboring unit cells, and any position which will contribute electron density to the chosen asymmetric volume is retained.

The second step is to organize the retained atoms by electron density section number. One pass through the list of retained atoms is sufficient to construct a set of lists of pointers; there is a list for each section, containing pointers to all of the atoms which will contribute electron density to that plane. The list of atoms is then written on a scratch file by section number – that is, all atoms which are in the list for the first section are written out, then those for the second section, and so on. Each atom will be written several times, but this procedure frees all of the computer memory previously used to hold the atoms and lists for use in constructing the model electron density sections.

Finally the scratch file is rewound and the model density sections are calculated. [It should be noted that Sayre's (1951) comment about multiplicity corrections at cell edges and symmetry elements does not apply in this case.] The model map is saved on magnetic tape for input to a space-group specific FFT Fourier inversion program, which was written with the use of techniques described previously (Ten Eyck, 1973). It should be noted that in principle it is possible to write optimum space-group specific FFT programs for all of the space groups except those in the cubic system and those which contain pure threefold or fourfold rotation axes. The execution time for such a program depends only on the number of grid points in the crystallographic asymmetric unit. In practice, such programs can be difficult to write for some space groups, particularly those involving centered lattices.

B. Calculation of structure factors by direct summation

The formula for calculation of structure factors by direct summation is

$$\mathbf{F}(hkl) = \sum_{\substack{\text{atoms}\\ r}} \{f_r \exp[-B_r(\sin^2\theta)/\lambda^2] \times \sum_{\substack{\text{positions}\\ s}} \exp[2\pi i(hx_s+ky_s+lz_s)]\} \quad (13)$$

where f_r is the atomic scattering factor for atom r at scattering angle θ, B_r is the isotropic thermal factor for atom r, and the sum over s is over all of the equivalent positions in the unit cell. The sum over equivalent positions can usually be reduced in cost by using the symmetry relations for each particular space group.

There are two general approaches which can be taken in applying (13) to the calculation of a set of structure factors. The most commonly used strategy seems to be to evaluate each $\mathbf{F}(hkl)$ separately. Recently Burnett & Nordman (1974) have shown that there are often advantages to evaluating the contribution of each atom to all of the structure factors before proceeding to the next atom, and present a data organization scheme which facilitates this strategy. In particular, the arguments of the sines and cosines needed for the complex exponentials in (13) advance by fixed increments as the indices (hkl) advance; this fact can be used to speed the calculation. The data organization and general program flow chart suggested by Burnett & Nordman will be assumed throughout the rest of this paper.

B.1. *Space-group symmetry*

It was mentioned previously that the space-group symmetry is generally used to reduce the complexity of the complex exponential terms in (13). Formulae for all of the space groups are given in *International Tables for X-ray Crystallography* (1968). However, these formulae are not always in the most useful form for machine computation – particularly for some of the more complicated space groups. The objective is to factor out the dependence on at least one index. This can easily be done for all space groups except those in the cubic system, which fortunately are rare except for very simple structures.

The factoring is easiest if the structure-factor formulae are left in complex exponential form. There are two possible forms,

$$\sum_s \exp[2\pi i(hx_s+ky_s+lz_s)] = E_0 \exp(2\pi ilz) \quad (14)$$

or

$$\sum_s \exp[2\pi i(hx_s+ky_s+lz_s)] = E_1 \exp(2\pi ilz) + E_2 \exp(-2\pi ilz) \quad (15)$$

where the particular values E_0, E_1 and E_2 can depend on h, k, l, x and y, but not on z. Furthermore the dependence on l (if any) will be such that for any particular value of h and k there will be only a small number of possible values of E_0, E_1, or E_2. The quantities E may be real, imaginary or complex, and E_1 may or may not be equal to E_2 or its complex conjugate. Analysis of these possibilities shows that for all non-cubic space groups *there are only four levels of complexity* in terms of the number of arithmetic operations required inside the inner loop of the calculation. Allowing for the multiplications by the atomic scattering factor and the additions required to accumulate $\mathbf{F}(hkl)$, these levels of complexity are $2m+1a$, $3m+2a$, $4m+2a$, and $6m+4a$, where m represents a multiplication and a represents an addition.

As a specific example the space group $P6_122$ gives

$$\begin{aligned} E_1 = \{&\exp[2\pi i(hx+ky)] \\ &+\exp(\pi il)\exp[-2\pi i(hx+ky)]\} \\ &+\exp(2\pi il/3)\{\exp[2\pi i(kx+iy)] \\ &+\exp(\pi il)\exp[-2\pi i(kx+iy)]\} \\ &+\exp(4\pi il/3)\{\exp[2\pi i(ix+hy)] \\ &+\exp(\pi il)\exp[-2\pi i(ix+hy)]\} \end{aligned}$$

and
$$\begin{aligned}E_2=&\{\exp[2\pi i(hx+iy)]\\&+\exp(\pi il)\exp[-2\pi i(hx+iy)]\}\\&+\exp(2\pi il/3)\{\exp[2\pi i(kx+hy)]\\&+\exp(\pi il)\exp[-2\pi i(kx+hy)]\}\\&+\exp(4\pi il/3)\{\exp[2\pi i(ix+ky)]\\&+\exp(\pi il)\exp[-2\pi i(ix+ky)]\}\end{aligned}$$

where $i=-(h+k)$. At first sight these expressions seem very complicated. However, $\exp(\pi il)=(-1)^l$, the expressions $\exp(2\pi il/3)$ and $\exp(4\pi il/3)$ can only take on three distinct values, and the terms grouped by curly brackets involve complex conjugate pairs. Therefore the whole calculation can be reduced to six cases, depending only on l, of the form

$$\mathbf{F}(hkl)=\sum_{\substack{\text{atoms}\\r}} f_r\exp[-B_r(\sin^2\theta)/\lambda^2]\times\{(a_r+ib_r)\cos 2\pi lz+(c_r+id_r)\sin 2\pi lz\}\quad(16)$$

where the six different values of a_r, b_r, c_r and d_r for each atom can be computed outside the inner loop of the calculation. The space group $P6_122$ therefore belongs to the $6m+4a$ category, as would be expected. Interestingly enough, the geometrically simplest possible space group, $P1$, also belongs to this category.

Burnett & Nordman (1974) suggest exploiting the fact that the arguments of the complex exponentials in (13) advance by fixed increments for each atom as the indices *hkl* advance. In the unfactored version of this method the number of operations inside the inner loop of the calculation is $2m+2sa$ for non-centrosymmetric space groups and $1m+\frac{1}{2}sa$ for centrosymmetric space groups, where s is the number of equivalent positions; there are also either s or $\frac{1}{2}s$ add-shift-mask operations to determine the indices for the sine/cosine table. This method is comparable in cost to the factored method described previously, and is particularly advantageous for the cases $s=1$ and $s=3$. It is of course possible to incorporate the factored expressions in the Burnett & Nordman method. In that case one has the same number of arithmetic operations as in the ordinary factored method, plus an add-shift-mask to determine the sine/cosine value required. Since all of the values of $\exp(2\pi ilz)$ needed in the fully factored method can be precalculated for each atom there appears to be a slight disadvantage to using the factored version of the Burnett & Nordman procedure compared with a factored procedure as will be described in the next section. However, it should be remembered that there are space groups for which the *unfactored* Burnett & Nordman procedure will beat *any* factored procedure.

B.2. Programming considerations

A program was written with the basic logic of Burnett & Nordman (1974), but with a factored structure-factor expression in the inner loop. Since all dependence on $\exp(2\pi ilz)$ has been factored out of the structure factor formulae, all required values of $\cos 2\pi lz$ and $\sin 2\pi lz$ can be tabulated as a function of l for each atom. Therefore the inner loop contains *no* sine or cosine calculations, and the table look-up index is also the loop index. The factors $f_r\exp[-B_r(\sin^2\theta)/\lambda^2]$ must be tabulated as a function of $(\sin\theta)/\lambda$; it was found that a table of length 200 produced changes in the third digit of the results as compared with use of analytic values of f_r. This is below the level of the measurement error in the experimental structure amplitudes, but the size of the table was increased to 500 anyway. The longer table gave results which differed from true in the fourth digit instead of the third. Finally, it was found worthwhile to treat the three centrosymmetric zones as special cases because about 20% of the thermolysin data are centrosymmetric.

C. Comparison of the methods

This portion of the paper will give a direct comparison of the performance of both methods on the same problem, with the same computer, and then consider how the methods may be expected to work on other problems. We have used both techniques to calculate structure factors for thermolysin, a protein containing 2437 non-hydrogen atoms, at 2·3 Å resolution – some 15 000 structure factors. The space group is $P6_122$, which has 12 equivalent positions; the unit-cell constants are $a=94{\cdot}2$ and $c=131{\cdot}4$ Å. The computer used in this work was a PDP-10 with a KA-10 central processing unit.

*C.*1. *Performance of the FFT method*

The model electron density map was constructed on a grid with $N_x=N_y=120$, $N_z=156$. The values of B' and B° were both set to 15 Å^2. This choice of parameters should lead to a maximum error of the order of 13% of the **F**(80,0,0) structure factor in the **F**(40,0,0) structure factor. Since the **F**(80,0,0) structure factor is at 1·15 Å resolution this is certainly an acceptable level of accuracy.

The asymmetric unit was chosen as $0\le x<1$, $0\le y<1$, and $0\le z\le 1/12$. This volume contains 201 600 grid points. 10 min of central processor time were required to calculate the model electron density function on this grid.

The cost of constructing the model electron density is roughly linear in the number of grid points; the same asymmetric unit on a grid $150\times150\times180$ (449 100 grid points) takes about 20 min. Use of the triple Gaussian model for the electron density would approximately double these times, the exact factor depending on the speed with which the Gaussian functions are calculated.

The Fourier inversion required 6·5 min of central processor time. The cost of calculating the Fourier transform is also roughly proportional to the size of the grid. The arithmetic cost is proportional to

$N \log N$, with N of the order of 10^5, and all of the program overhead (such as input/output costs) is directly proportional to N.

The total time required to calculate the structure factors by the FFT method is 16·5 min, and depends almost entirely on the size of the sampling grid. In turn, the size of the sampling grid depends on the required resolution. The cost as a function of the number of atoms n and the resolution in ångströms d is approximately given by

$$\underset{\mathrm{FFT}}{\mathrm{cost}} = (K_1 n + K_2)/d^3 \tag{17}$$

where K_1 and K_2 are constants which will depend on the particular computer and programs.

C.2. Performance of the direct-summation method

Test timings of the direct-summation program indicated that it would require about 4 h to calculate a set of thermolysin structure factors at 2·3 Å resolution. Therefore the complete cycle over all indices hkl for each atom was re-written in machine code for maximum speed. (The FFT structure factor programs are all written in Fortran IV.) With this change the program ran in 115 min. The mean absolute difference between the structure factors calculated by the two methods was 0·16% of the mean structure amplitude.

The obvious scale factor to relate this time to a time for other problems is the product of the number of atoms and the number of structure factors. However, there is also a 'complexity factor' which depends on the amount of arithmetic inside the inner loop of the program. (The cost of any direct-summation program depends almost exclusively on the number of operations inside the inner loop of the program.) The space group $P6_122$ requires six multiplications and four additions per structure factor per atom. (The program logic used here does not require any arithmetic to calculate the trigonometric functions needed.) This is the minimum number of operations for all of the non-centrosymmetric trigonal and hexagonal space groups, as well as for $P1$ and the non-centrosymmetric tetragonal space groups with $4/mmm$ Laue symmetry. The remaining non-centrosymmetric tetragonal space groups and the non-centrosymmetric space groups in the monoclinic and orthorhombic systems require three or four multiplications and two additions inside the inner loop. Centrosymmetric space groups seem to require at best two multiplications and one addition, and at worst three multiplications and two additions. On many computers (including the PDP-10) a floating point multiplication takes twice as long as a floating point addition. Therefore we have a set of 'complexity factors' ranging from 5 to 16. (Complexity factors for cubic space groups will be slightly less than three times the complexity factors of their orthorhombic or tetragonal subgroups.) Thus we have a cost function of the form

$$\mathrm{cost}_{DS} \propto kn/d^3 \tag{18}$$

where k is the space group complexity factor, n is the number of atoms and d is the resolution in ångströms.

Subject to the same assumptions as above concerning the relative costs of multiplication and addition, the complexity factor for the method of Burnett & Nordman is $2s+4$ for a non-centrosymmetric space group with s equivalent positions. Therefore

$$\mathrm{cost}_{BN} \propto (2s+4)n/d^3\ . \tag{19}$$

D. Discussion

The results of the preceding section are summarized in Tables 1 and 2, which show the relative cost of a fast Fourier transform structure-factor calculation and the two direct-calculation techniques discussed in this paper. The cost factors are scaled together based on the actual results for the two $P6_122$ structure-factor calculations, which indicate that the FFT method is similar in cost to a direct calculation with a complexity factor of 2·3. The FFT method is superior in all cases.

All three techniques discussed here retain the same relative cost factors as resolution is increased, with one important exception. When the resolution is better than 2 Å the FFT structure factor calculation

Table 1. *Comparison of FFT and factored structure-factor calculation*

Type of space group*	Simplest centrosymmetric	Other centrosymmetric, simplest non-centrosymmetric	Simple non-centrosymmetric	Other non-centrosymmetric (not cubic)
Arithmetic operations inside inner loop	$2m+1a$	$3m+2a$	$4m+2a$	$6m+4a$
Complexity factor†	5	8	10	16
Cost relative to FFT‡	2·2	3·5	4·3	7·0

* The space groups are classified according to arithmetic operations inside the inner loop, not geometric complexity. $P1$ is not a 'simple' space group in this context.

† Estimates of the complexity factor assume one multiplication is as expensive as two additions ($m=2a$).

‡ The costs in this row will be somewhat lower for space groups with threefold and fourfold pure rotation axes (*e.g.* $R3$ or $P4$, but not $P3_1$, or $P4_1$). For these space groups the FFT cost formulae will be either $(K_1n+2K_2)/d^3$ or $(K_1n+3K_2)d^3$.

becomes more expensive if high precision (exact scattering factors) is required. It is doubtful, however, whether it would ever become more expensive than either of the direct methods.

Table 2. *Comparison of unfactored Burnett & Nordman* (1974) *method with others for non-centrosymmetric space groups*

Number of equivalent positions	1	2	3	4	6	8	12
Complexity factor, Burnett & Nordman (1974)	6	8	10	12	16	20	28
Cost relative to factored method*	0·4	1·0	0·6	1·2	1·0	1·3	1·8
Cost relative to FFT method†	2·6	3·5	4·3	5·2	7·0	8·7	12·2

* This row is approximate because different space groups can have different complexity factors while having the same number of equivalent positions. For example, *P*222 has a complexity factor of 10, but *P*4 has a complexity factor of 8.

† The relative costs listed in this row will be somewhat lower for space groups with threefold or fourfold pure rotation axes.

High-resolution calculations may also involve anisotropic temperature factors. Anisotropic temperature factors completely destroy the relations used in factoring the complex exponentials of (13), so the calculation becomes triclinic. In addition we will have two more multiplications and an exponential function inside the inner loop. With tabulated exponentials assumed the cost of the 'factored' method becomes proportional to $20sn/d^3$. The Burnett & Nordman method also requires two extra multiplications and an exponential function per atom for anisotropic thermal factors; therefore the cost becomes proportional to $(6s+4)n/d^3$, which is much cheaper than the factored expression for all values of s. Anisotropic thermal factors can readily be accomodated in the FFT method by using elliptical atoms rather than spherical atoms when constructing the model electron density function. Elliptical atoms complicate the process of determining which atoms in the unit cell will contribute electron density to the asymmetric unit but do not complicate the actual computation of the electron density function. (The reason is that the equation for a circle in a skew coordinate system is the same as the equation for an ellipse in an orthogonal coordinate system.) Therefore anisotropic thermal factors should cause a moderate increase in the cost of an FFT structure-factor calculation.

The considerable labor involved in writing space-group specific FFT structure-factor programs is only justified at present for large problems for which the calculation will be repeated many times. Refinement of biological macromolecules is a case in point. Such problems typically involve several thousand atoms and more than ten thousand structure factors, and the projects last several years. For the typical small-molecule crystallographic laboratory, where many space groups are encountered and the number of atoms is not large, the FFT structure-factor calculation is not worth the labor required to program it.

This work was supported in part by grants from The National Institutes of Health (GM20066) and The National Science Foundation (BMS74-18407).

References

Brillouin, L. (1962). *Science and Information Theory*, 2nd ed., pp. 93–97. New York: Academic Press.

Burnett, R. M. & Nordman, C. E. (1974). *J. Appl. Cryst.* **7**, 625–627.

Forsyth, J. B. & Wells, M. (1959). *Acta Cryst.* **12**, 412–415.

Gentleman, W. M. & Sande, G. (1966). *Proc. Fall Joint Computer Conf.* pp. 563–578.

International Tables for X-ray Crystallography (1968). Vol. I, 2nd ed. Birmingham: Kynoch Press.

Lipson, H. & Cochran W. (1966). *The Determination of Crystal Structures*, 3rd ed., pp. 94–102. Ithaca: Cornell Univ. Press.

Matthews, B. W., Weaver, L. H. & Kester, W. R. (1974). *J. Biol. Chem.* **249**, 8030–8044.

Sayre, D. (1951). *Acta Cryst.* **4**, 362–367.

Ten Eyck, L. F. (1973). *Acta Cryst.* A**29**, 183–191.

Vand, V., Eiland, P. F. & Pepinsky, R. (1957). *Acta Cryst.* **10**, 303–306.