Acta Cryst. (1976). A**32**, 832

Methods and Programs for Direct-Space Exploitation of Geometric Redundancies

By G. Bricogne

M.R.C. Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, England

(*Received* 19 *January* 1976; *accepted* 1 *March* 1976)

Phases can be determined for geometrically redundant amplitudes by iteration of the following procedure: compute an electron density map from the currently available phases; average the electron densities of all the crystallographically independent molecules; rebuild the crystal(s) from this averaged subunit, setting the density outside the molecular boundaries to its average value; obtain phase information from the resulting structure, and combine it with that given by isomorphous replacement to produce the phases to be used in the next iteration. This algorithm converges very rapidly, and has proved to be a powerful tool in the solution of two large unknown protein structures. This paper describes the computational techniques developed to implement it, which include a swift and general method for real-space averaging of electron density maps.

Introduction

Proteins frequently crystallize with several identical subunits in the asymmetric unit, or in several crystal forms which contain the same molecule in different arrangements. Rossmann & Blow (1963) recognized that intensity data collected from such structures are redundant, and that their redundancy could be a source of phase information. A variety of techniques have been developed to implement these ideas, which have become known collectively under the broad designation of the 'molecular replacement method' (Rossmann, 1972). This paper will be concerned almost exclusively with the last stage of this method, namely the generation of phase information once the geometrical transformations relating the independent subunits have been determined.

The phase constraints implied by the consistency of geometrically redundant intensities were first derived by Rossmann & Blow (1963), and generalized by Main & Rossmann (1966). Crowther (1967, 1969) reformulated them as linear eigenvalue equations between structure factors. On this basis, he proposed the first practical procedure to solve them: the iterative '**H**-matrix method', which was used by Jack (1973) to determine 840 signs for a 17-fold symmetric projection of TMV coat protein discs (the abbreviations used in this paper are listed in Table 1). However, any such reciprocal-space method was bound to generate amounts of computation proportional to N^2 for N independent reflexions. This limited their usefulness quite severely, and N could not exceed a few thousands (see Jack, 1972, 1973).

In a previous paper [Bricogne (1974), hereafter referred to as (I)], the theory was reformulated in real space, and the equations obtained were proved to be equivalent to the most general molecular-replacement equations previously derived in reciprocal space, as had been suggested by Main (1967) and Rossmann (1972). This showed that the most costly step in Crowther's procedure could be carried out much more economically by the following set of operations in real space: average the electron densities of all crystallographically independent subunits; rebuild the crystals from this averaged subunit, setting the density outside the molecular boundaries to its average value; the resulting structure can then be used to obtain phase information. This involves amounts of computation proportional only to N.

Table 1. *List of abbreviations used*

MRM	molecular replacement method
I.R.	isomorphous replacement
SIR	single isomorphous replacement
DIR	double isomorphous replacement
MIR	multiple isomorphous replacement
EM	electron microscopy
a.u.	asymmetric unit
n.c.s.	non-crystallographic symmetry
r.m.s.	root mean square
e.d.	electron density
B. St.	*Bacillus stearothermophilus*
GPD	D-glyceraldehyde-3-phosphate dehydrogenase
TMV	tobacco mosaic virus
TBSV	tomato bushy stunt virus
SBMV	southern bean mosaic virus

The averaging of independent copies of the same molecule has been frequently used to improve e.d. maps (Birktoft & Blow, 1972; Muirhead, Cox, Mazzarella & Perutz, 1967; Buehner, Ford, Moras, Olsen & Rossmann, 1974). However, early attempts to derive phase information from the averaged maps remained inconclusive (Muirhead *et al.*, 1967). This was probably because only two independent molecules were available. Recent work has been more successful. Harrison & Jack (1975) have used real-space averaging to improve some 1300 phases for an icosahedral virus. In a low-resolution study on a known protein structure, Argos, Ford & Rossmann (1975) showed that the same method, if provided with a high-resolution molecular envelope, could extend 6·0 Å SIR phases to

4·9 Å resolution, or even generate phases *ab initio*; although the errors were rather large, the phases obtained were good enough to locate heavy atoms in isomorphous derivatives. However, serious computational difficulties were encountered when more than about 7000 reflexions had to be handled, even with a rather coarse (2 Å) sampling of the e.d. maps; the method was concluded to be still 'beyond reasonable computer limits at high resolution'.

In this paper, I describe a set of computational techniques which have been used to solve the phase problem for two unknown protein structures by a combined use of isomorphous replacement and non-crystallographic symmetry. The main emphasis has been put on the possibility of coping with large problems. Indeed, the two structures mentioned (GPD from *B.St.* and the disc of TMV coat protein) are among the very largest phased to this date, in terms of the amount of data involved (42 000 and 38 000 independent reflexions respectively). A general and efficient system for averaging densities at points related by n.c.s., and for reconstructing the e.d. for an averaged crystal, has been devised to operate on maps which may contain up to 10^7 points or more. This method is based on a double-sorting technique. The averaged structure is then used as a 'known part' in the sense of Sim (1959) to introduce symmetry-based phase information into the weighting scheme proposed by Blow & Crick (1959) for I.R. data. The phase probability density generated for each reflexion by Sim's formula is combined with any previously available phase information, to give the best phase and figure of merit used in the next iteration. This algorithm converges very rapidly (typically, in two or three cycles).

After presenting the programming techniques used, I shall discuss some general principles of their practical implementation. An account of their use for structure determination may be found in the recent paper of Champness, Bloomer, Bricogne, Butler & Klug (1976) on the structure of TMV protein.*

1. Survey of program requirements

Geometric redundancies will very seldom be able to generate phases by themselves. I shall examine later the question of the minimum phase information which may provide an adequate starting point.

For the moment, let us consider as an example the case of a structure for which we have incomplete phase information (such as might be given by a single isomorphous derivative) but whose asymmetric unit contains n copies of molecule M, related by *known* local symmetry. The solution of this structure is performed in outline as follows:

(1) Encode the preliminary phase information given by I.R.

(2) Define the envelope of the basic aggregate of n molecules.

(3) Iterate the following procedure: (*a*) compute an e.d. map from the observed moduli and the current phases and figures of merit; (*b*) average the n independent copies of M, calculate the average density in solvent regions; (*c*) rebuild the asymmetric unit from this averaged aggregate, set solvent density to its average value; (*d*) compute structure factors from this e.d. map; (*e*) combine the (single isomorphous) phase information with the information derived from the calculated structure factors by the use of Sim's formula; obtain new phases and figures of merit.

Fast Fourier transform programs (Ten Eyck, 1973) are used in steps 3(*a*) and 3(*d*). I shall now describe the methods used in the other steps.

2. Definition of molecular envelopes

The molecular boundaries are determined by calculating an e.d. map with the initially available phases, and averaging it by the local symmetry elements of a particular molecular aggregate. Because of the local character of this symmetry, the chosen aggregate is preserved by the averaging, whereas the neighbouring ones are smeared. Comparing the unaveraged and averaged maps usually gives a good idea of the boundaries of the aggregate (Muirhead *et al.*, 1967; Buehner *et al.*, 1974).

To define them in a form usable by a computer, previous workers used to write down, for each line of the grid on which the map had been averaged, the coordinates of the end points of those segments belonging to the envelope.

I chose to take advantage of a tracing system developed in this laboratory by Dr J. White for the study of nematode neuroanatomy. The envelope is traced section by section onto a graphical input tablet linked to a Modular One computer. Each profile is coded as a string of pen movements in eight possible directions on a 256 × 256 grid. This type of description is very compact indeed, but for our purposes a representation by arrays of logical variables was more convenient. A processing program was written in IBM Fortran IV to carry out this conversion, producing consecutive sections of the envelope as arrays of '1-byte integers'; this allows for the necessity of labelling the envelopes of the various subunits when dealing with improper local symmetries (see § 3.3). To keep all the sections in register, three fiducial marks are used: the grid on which the array is to be produced is defined by specifying the grid coordinates corresponding to these marks.

The e.d. handling system described below can be used to 'move' envelopes between different grids (see § 3.5). The problem of the overlap of neighbouring envelopes will be dealt with at that stage.

* *Note added in proof*: these programs have now been used to solve the structure of TBSV to 5·5 Å resolution, in collaboration with Dr S. C. Harrison.

3. Handling of electron densities

The averaging and rebuilding operations of steps 3(*b*) and 3(*c*) pose serious computational problems as soon as the initial map becomes too large to be fitted into the core memory of the computer. The difficulty lies in the fact that, in order to calculate the average density at a given point under a symmetry operation of order n, densities have to be fetched at $(n-1)$ other points; these can be anywhere in an oversampled map which may contain several millions of points. The problem is thus basically an *addressing* problem in which, at first sight, an enormous file has to be made randomly accessible.

An approach towards the solution of this problem has been made by Dr G. Ford (see Buehner *et al.*, 1974), using a pagination technique similar to the virtual memory systems of modern computers.

The approach used here is much more radical, and eliminates the initial random access requirement by means of two sorting operations.

3.1. *The double-sorting technique*

The nature of the addressing problem, and of the solution proposed here, are best understood by first analysing a simple process: the interpolation of the electron density of a crystal from a domain $\mathscr{D}_1$ of a grid $\mathscr{G}_1$ to a domain $\mathscr{D}_2$ of a grid $\mathscr{G}_2$ 'skew' relative to $\mathscr{G}_1$. Usually, $\mathscr{D}_2$ will be a rectangular box; $\mathscr{D}_1$ will be the asymmetric unit of the crystal, and grid $\mathscr{G}_1$ will be 'crystallographic' (*i.e.* based on the crystal axes). Programs have long been available to do this when the initial map can be accommodated in core (Cullis, Muirhead, Perutz, Rossmann & North, 1962).

Let the grid coordinates in $\mathscr{G}_1$ be $X_1=(x_1,y_1,z_1)$; they need not be integral. A bar will denote the operation of reduction to the asymmetric unit $\mathscr{D}_1$, *i.e.* of finding the point $\bar{X}_1$ in $\mathscr{D}_1$ equivalent to X_1 by a suitable space group transformation. In grid $\mathscr{G}_2$, we shall be interested only in integral points, with coordinates $I_2=(i_2,j_2,k_2)$. Let us denote by $X_1(I_2)$ the coordinates in $\mathscr{G}_1$ of the point with coordinates I_2 in $\mathscr{G}_2$.

With this notation, it is clear that the density wanted at point I_2 of $\mathscr{D}_2$ is to be fetched at point $\bar{X}_1(I_2)$ of $\mathscr{D}_1$. We can therefore represent the whole interpolation as a succession of elementary operations of the form:

'Get density at $\bar{X}_1(I_2)$; put it at I_2' repeated for I_2 running through $\mathscr{D}_2$. If domain $\mathscr{D}_2$ is scanned in an ordered fashion relative to the coordinates in $\mathscr{G}_2$, the resulting sequence of points $\bar{X}_1(I_2)$ will keep jumping back and forth all through $\mathscr{D}_1$ in an unpredictable way, whence the apparent need for a random access to the initial map.

However, we can proceed more thoughtfully. Scanning $\mathscr{D}_2$ in the usual way, we write out for each of its points I_2 a record of the form:

$$I_2, \bar{X}_1(I_2)\,.$$

This generates a file $(\mathscr{I})$ containing a complete description of the tasks to be performed. But instead of executing these get–put requests in the order in which they have been issued, we shall *reorder them beforehand* so as to minimize the labour involved in collecting the densities out of the initial map.

Suppose the map is given in $\mathscr{D}_1$ as z sections, written sequentially in ascending z order. Let us sort the records of file $(\mathscr{I})$ on z_1, into another file $(\mathscr{I}_s)$. We can now perform the whole interpolation in *one pass* through the map and through this sorted file, with only two sections in core at any one time.

> Read the first two sections of the map into core. For each record read from $(\mathscr{I}_s)$, check z_1 against the upper section number z_u. If $z_1<z_u$, interpolate the density at $\bar{X}_1(I_2)$ using sections z_u and z_u-1 which are currently in core, overwrite $\bar{X}_1(I_2)$ by ϱ and output a record of the form: I_2, ϱ on a file $(\mathscr{O})$. If $z_1\ge z_u$, overwrite the (now useless) lower section by the next section, check z_1 against the new z_u; repeat this process until $z_u-1\le z_1<z_u$; then interpolate as above.

Upon completion of this procedure, file $(\mathscr{O})$ contains all the densities needed to reconstitute $\mathscr{D}_2$. They are in a random order, but have their addresses attached to them. If, for example, we want the final map as a sequential file of y sections, we sort the records in $(\mathscr{O})$ on j_2, to get a file $(\mathscr{O}_s)$. We can now reconstitute this map in *one pass* through the sorted file, with only one section in core at any one time.

> For each record read from $(\mathscr{O}_s)$, check j_2 against the current section number j_c. If $j_2=j_c$, put density ϱ at location described by (i_2,k_2) in current section. If $j_2>j_c$, write out the (now complete) current section, increment j_c by 1 and check j_2 against it; repeat the process until $j_c=j_2$; then pursue the reconstitution as above.

This five step procedure (generation/sort/interpolation/sort/reconstitution) does therefore succeed in producing the desired domain of density, without needing random access to the initial map.

3.2. *Solution of the general averaging problem.*

The operations of truncation, averaging, crystal rebuilding and background setting will now be incorporated into the scheme outlined above. For simplicity, we shall first deal with 'proper' n.c.s.'s; in this case, the local symmetries of the aggregate form a group, so that there is essentially one envelope: the envelope of the whole aggregate [see Rossmann (1972) and (I) § 3.3.2].

3.2.1. *Treatment of the protein density: first method*

Suppose the electron density in the asymmetric unit $\mathscr{D}_1$ of crystal 1 (space group S_1) is given on a crystallographic grid $\mathscr{G}_1$. A very general requirement will be as follows: extract a molecular aggregate $\mathscr{A}$ from this map; average it by its local symmetries; then build from this averaged aggregate the asymmetric unit $\mathscr{D}_2$ of a different crystal (crystal 2, with space group S_2), on a crystallographic grid invariant by S_2. This change of lattice and space group will be necessary in studies involving several crystal forms.

Let a bar and a double bar denote the reductions to the asymmetric units of the two crystals ($\bar{X}_1$ reduces to $\bar{X}_1 \in \mathscr{D}_1$; I_2 to $\bar{\bar{I}}_2 \in \mathscr{D}_2$). Let $T_1, \ldots, T_n$ be the local symmetry operations of aggregate $\mathscr{A}$, including the identity, given as grid-coordinate transformations in $\mathscr{G}_2$. Let $\mathscr{U}$ be the envelope of $\mathscr{A}$, given as an array of logical variables in $\mathscr{G}_2$; $\bar{\bar{\mathscr{U}}}$ will denote $\mathscr{U}$ 'folded back' into $\mathscr{D}_2$ by S_2.

The operation to be performed can then be phrased: for all points I_2 in $\mathscr{U}$, obtain the average of the densities at points $T_1I_2, \ldots, T_nI_2$, and put it at point $\bar{\bar{I}}_2$; this will extract $\mathscr{A}$ from $\mathscr{D}_1$, average it, and fold it back into $\mathscr{D}_2$.* Since the densities needed are to be fetched in $\mathscr{D}_1$ at points $\bar{X}_1(T_1I_2), \ldots, \bar{X}_1(T_nI_2)$, the whole task is a succession of elementary operations of the form:

'Get density at $\bar{X}_1(T_mI_2)$; accumulate it at $\bar{\bar{I}}_2$', repeated for I_2 running through $\mathscr{U}$ and m running from 1 to n.

Let us generate a file ($\mathscr{F}$) by writing out, for each point I_2 in $\mathscr{U}$, n records of the form:

$$\bar{\bar{I}}_2, \bar{X}_1(T_mI_2) \quad (m=1, \ldots, n) .$$

If we processed this list as described in the previous section, each point $\bar{\bar{I}}_2$ in $\bar{\bar{\mathscr{U}}}$ would receive, at reconstitution time, n densities coming from the n points symmetry-related to the initial point I_2. Let us then slightly alter this last step. To perform the averaging, we replace the 'put' operations by accumulations in an initially cleared array, keep counts of how many densities are received at each location of the array, and divide each sum by the corresponding count when the section is complete. This is slightly more expensive than uniformly dividing all the sums by n. But it will cure the problems due to the unavoidable overlap of the envelopes at the contacts between aggregates, since in these regions it will average the densities belonging to the two neighbouring aggregates instead of adding them. Also, these individual counts will be needed to deal with the background.

With this modified reconstitution program, we can average and rebuild simultaneously (*i.e.* using only *one* interpolation), with an amount of computation proportional to $n \times \mathscr{N}_2$ if $\mathscr{N}_2$ is the number of points of grid $\mathscr{G}_2$ enclosed in $\mathscr{U}$.

3.2.2. *Treatment of the protein density*: *second method*

Averaging and rebuilding may also be carried out *consecutively*. To do this most economically, we shall take advantage of a slight non-equivalence between the roles of S_1 and S_2. Indeed, S_2 must be a true space group: it must map $\mathscr{G}_2$ onto itself since $\bar{\bar{I}}_2$ has to belong to $\mathscr{G}_2$ if I_2 does. On the other hand, S_1 is not thus restricted, since it handles non-integral coordinates anyway; its only function is to enable us to retrieve from $\mathscr{D}_1$ all the electron densities we may ever need. In this role as an 'e.d. supply', the pair ($\mathscr{D}_1$, S_1) can be replaced by any pair ($\mathscr{D}, S$) such that every point at which the e.d. will be needed in the interpolation be equivalent to a point in the initial domain $\mathscr{D}$ by a transformation of S.

Let us take for $\mathscr{D}$ a domain of an intermediate Cartesian grid $\mathscr{G}$ (with coordinates noted I or X) containing a suitably chosen $1/n$th of the aggregate, defined by its envelope in $\mathscr{G}$ noted $\mathscr{U}/n$; S will consist of the local symmetries $T_1, \ldots . T_n$, expressed as grid coordinate transformations in $\mathscr{G}$.

The averaged density in $\mathscr{D}$ is first computed from the initial map $\mathscr{D}_1$, using a list of records of the form:

$$I, \bar{X}_1(T_mI) \text{ for } I \text{ in } \mathscr{U}/n, \text{ and } m=1, \ldots, n .$$

The whole final asymmetric unit $\mathscr{D}_2$ is then rebuilt from $\mathscr{D}$, using a list of records of the form:

$$\bar{\bar{I}}_2, \hat{X}(I_2) \quad \text{for } I_2 \text{ in } \mathscr{U},$$

where the ^ on $\hat{X}$ means a reduction to $\mathscr{D}$ by S. The only restriction imposed on the choice of $\mathscr{U}/n$ is that such a reduction should be possible, *i.e.* that the logical union of T_1 ($\mathscr{U}/n$), $\ldots$, $T_n(\mathscr{U}/n)$ should contain $X(\mathscr{U})$; in particular, $\mathscr{U}/n$ does not have to coincide with the boundary of a subunit.

The total amount of computation involved is now proportional to $\mathscr{N}+\mathscr{N}_2$ (if $\mathscr{U}$ contains $\mathscr{N}$ points of grid $\mathscr{G}$), but a *double* interpolation is required. The two methods will be compared in § 5.

3.2.3. *Treatment of the background*

Some further modifications are needed in order to deal with the density outside the molecular boundaries.

Note that if we want to set it to a uniform value, this value should be the average of the density *outside* the molecules [and not *inside* as in Argos *et al.* (1975)] if the correct relative scale between protein and solvent is to be preserved.

To estimate the average density in the solvent regions of crystal 1, we keep track of which points of $\mathscr{D}_1$ are used during the interpolation, and evaluate the background density as the average of the densities at all unused points; this value can then be passed on to the reconstitution program for $\mathscr{D}_2$, and put wherever no density has been received in a completed section.

If crystals 1 and 2 are the same, we may wish to retain the background structure of the initial map; for this purpose, we reconstitute each section into the

* We could also get densities directly for all points I_2 in $\bar{\bar{\mathscr{U}}}$. However, $\bar{\bar{\mathscr{U}}}$ usually consists of several disconnected pieces, and the local symmetries of each piece are not the T_m's, but the $R_iT_mR_i^{-1}$'s, where R_i is the operation of S_2 used to produce piece no. i from the original aggregate $\mathscr{A}$. The other procedure avoids this inessential complication.

corresponding section of the initial map (instead of a cleared array), letting the first density received at each point overwrite the initial density.

3.3. *The case of improper local symmetries*

Thus far, we have considered only proper n.c.s.'s. 'Improper' local symmetries, in which the group property is lost, can be handled within the same general framework.

Let T_{pq} be the transformation bringing molecule p onto molecule q. Let us label the envelopes of the n molecules as U_p ($p=1,\ldots,n$) by defining them by means of integral (instead of purely logical) variables. We can then use these labels to select the transformations to be applied to each point at which an averaged density has been requested.

To average and rebuild simultaneously, scan a domain of $\mathscr{G}_2$ containing all the U_p's, generating the list:

$$I_2, X_1(T_{pq}I_2) \text{ for } I_2 \text{ in } U_p, \text{ and } q=1,\ldots,n\,,$$

then proceed as usual.

To rebuild an asymmetric unit possessing the required improper local symmetry from an already averaged subunit, given in a reference position labelled 0, scan the same domain, generating the list:

$$I_2,\ X(T_{p0}I_2) \quad \text{for } I_2 \text{ in } U_p,$$

then proceed as usual.

The determination of the transformations T_{pq} or T_{p0} entails special difficulties. In the case of proper symmetries, the position of one heavy atom per molecule (assumed to occupy an intramolecular site) suffices to determine uniquely the local symmetries of the aggregate; this is not true for improper symmetries, as each molecule still possesses three rotational degrees of freedom. A possible solution is to compute an e.d. map using the phases given by the single isomorphous derivative, then try and correlate the regions of density surrounding the heavy-atom sites. This method was used by Jack (1972) for the enzyme barnase. Fletterick & Steitz (1975) refined in a similar fashion the transformations relating the different subunits present in two crystal forms of yeast hexokinase. The implementation of this correlation technique within the double-sort framework is discussed in the next section.

3.4. *Correlation of electron densities*

In their study of the effect of oxygen binding on the quaternary structure of haemoglobin, Muirhead *et al.* (1967) used a least-squares superposition method: for each pair of corresponding oxy- and deoxyhaemoglobin chains, six parameters (Euler angles and three translation components) were refined in order to produce the best fit between the two electron densities. The logic of the program is as follows (J. M. Baldwin, personal communication): keep one of the maps in core; read successive sections of the second map into an array; for each point within the boundaries of the chain under consideration, compute the coordinates of the corresponding point of the first map, using the current values of the parameters; calculate the electron density and its gradient at this point, and increment the sums needed for the least-squares parameter refinement.

Clearly, if the densities of map no. 2 were sorted on the coordinate labelling the sections of map no. 1, we could perform the calculation without needing to keep map no. 1 in core.

Using the notation of § 3.2, we want to correlate the density of the region of crystal 2, contained in envelope $\mathscr{U}$ with a corresponding domain of crystal 1; we have approximate starting values for the parameters relating them.

We first obtain electron densities in a domain $\mathscr{D}_1'$ of crystal 1 (on grid $\mathscr{G}_1$) containing all the points which will be needed in the refinement. This is done most economically with a list of records:

$$\bar{I}_1, I_1 \text{ for all points } I_1 \text{ in } \mathscr{D}_1'$$

containing only integral grid coordinates; in this way we can replace the interpolation step by a simpler 'pick-up' step, which needs only have one section of the map $\mathscr{D}_1$ in core at any one time.

We then generate a file ($\mathscr{I}$) of records:

$$\bar{\bar{I}}_2, X_1(I_2) \text{ for all points } I_2 \text{ in } \mathscr{U},$$

the correspondence between I_2's and X_1's being that defined by the initial values of the parameters. We sort these records on $\bar{\bar{I}}_2$ so that we can pick densities out of $\mathscr{D}_2$ and obtain a file ($\mathscr{O}$) of records:

$$\varrho_2, X_1(I_2) \quad \text{where } \varrho_2 \text{ has been picked up at } I_2.$$

Finally, we sort these records on X_1 (*e.g.* z_1 if the map in $\mathscr{D}_1'$ has been reconstituted as z sections) and obtain a file ($\mathscr{O}_s$).

The correlation program can now work from map $\mathscr{D}_1'$ and the sorted list of densities ($\mathscr{O}_s$). The logic of this procedure is similar to that of the program used by Muirhead *et al.*, but it can handle much larger problems since only two sections of $\mathscr{D}_1'$ have to be present in core at a time.

3.5. *Handling of molecular envelopes*

The tracing system of § 2 codes molecular envelopes as arrays of logical (or 1-byte integral) variables, corresponding to the sections of the map from which the boundaries were inferred. Usually, these sections will be skew relative to any crystallographic grid (*e.g.* perpendicular to a local n-fold axis). On the other hand, in §§ 3.2 to 3.4, envelopes were used as arrays defined on other grids. Therefore, the facilty must be provided to move envelopes between different grids.

We shall use the notation of the preceding sections.

In the case of proper symmetries, the envelope is described by purely logical variables. To move envelope $\mathscr{U}$, defined on grid $\mathscr{G}_2$, onto grid $\mathscr{G}_1$, with the option of imposing local symmetry by transformations

$T_1,\ldots,T_n$, generate a list ($\mathscr{I}$) of records:

$X_1(T_mI_2)$ for all points I_2 in $\mathscr{U}$, and $m=1,\ldots,n$;

sort these records (*e.g.* on x_1 if the envelope in $\mathscr{G}_1$ is wanted as x sections) into a file ($\mathscr{I}_s$); then reconstitute the envelope in one pass through this sorted file, with only two sections in core at any one time, as follows:

> Clear the arrays for two sections and initialize their numbers. For each record read from ($\mathscr{I}_s$), check x_1 against the upper section number x_u. If $x_1<x_u$, put the value 'true' at the four nearest grid points in each of the two sections x_u and x_u-1 which are currently in core. If $x_1\geq x_u$, write out the (now complete) bottom section x_u-1; clear the corresponding array, increment x_u by 1 and check x_1 against it; repeat until $x_u-1\leq x_1<x_u$, then pursue reconstitution as above.

In other words, the presence in ($\mathscr{I}_s$) of a record for point (x_1,y_1,z_1) is interpreted as meaning that its eight nearest neighbours in $\mathscr{G}_1$ are within the envelope.

In the case of improper symmetries, the envelopes have to be labelled (§ 3.3). To move a set of labelled envelopes U_p $(p=1,\ldots,n)$ from grid $\mathscr{G}_2$ to grid $\mathscr{G}_1$, we use a list of records:

$X_1(I_2),p$ for all I_2's in U_p and all p's,

and proceed as above, except that in the reconstitution we put the value p (instead of 'true') at the eight nearest grid points. Ambiguities due to overlaps are resolved by giving priority to the highest p values.

The generation of a set of labelled envelopes U_p $(p=1,\ldots,n)$ from a reference envelope U_0 proceeds likewise, using a list of records:

$X_1(T_{0p}I_2),p$ for all points I_2 in U_0, and $p=1,\ldots,n$.

Note that there are no reductions to the asymmetric unit, since envelopes are used 'unfolded' (see footnote to § 3.2).

3.6. *Scope of this map handling system*

The handling of e.d. maps by the double-sorting technique possesses several advantageous features:

– A large variety of operations can be carried out by generating the appropriate file ($\mathscr{I}$), then proceeding in an invariable way. Thus, each type of operation corresponds to an option in the generating program. Table 2` summarizes those available in the existing program.

– When the averaging/rebuilding has to be iterated, as in the procedure of § 1, the same file ($\mathscr{I}_s$) can be used as long as the envelope and the symmetry elements are not modified.

– The space-group specificity is confined to the subroutines of the generating program which perform the reductions to the initial domain ($\mathscr{D}_1$ or $\mathscr{D}$) and to the final asymmetric unit ($\mathscr{D}_2$).

Table 2. *Options of the generating program*

Option number	Records on ($\mathscr{I}$) Types	Ranges	Use of this option
0	$\bar{I}_1,I_1$	All I_1 in $\mathscr{D}_1'$	Generation of an arbitrary crystallographic domain from a crystal a.u.
1	$I_2,\bar{X}_1(T_mI_2)$	All I_2 in $\mathscr{D}_2$ $m=1,\ldots,n$	Skew section calculation, with optional averaging
2	$I_2,\bar{X}_1(T_mI_2)$	All I_2 in $\mathscr{U}$ $m=1,\ldots,n$	*id.*, with truncation by an envelope
3	$X_1(T_mI_2)$	All I_2 in $\mathscr{U}$ $m=1,\ldots,n$	Transport of an unlabelled envelope with optional symmetrization
4	$X_1(I_2),l$	All I_2 in U_l $l=1,\ldots,n$	Transport of a set of labelled envelopes
5	$\bar{\bar{I}}_2,\bar{X}_1(T_mI_2)$	All I_2 in $\mathscr{U}$ $m=1,\ldots,n$	Averaging by proper n.c.s.
6	$\bar{\bar{I}}_2,\bar{X}_1(T_{lm}I_2)$	All I_2 in U_l $l=1,\ldots,n$ $m=1,\ldots,n$	Averaging by improper n.c.s.
7	$X_1(T_{0m}I_2),m$	All I_2 in U_0	Generation of a set of labelled envelopes from a reference envelope
8	$\bar{\bar{I}}_2,\bar{X}_1(T_{m0}I_2)$	All I_2 in U_m $m=1,\ldots,n$	Generation of an a.u. possessing an improper n.c.s. from a subunit in reference position

– Any mode of sectioning maps and envelopes may be adopted, provided the corresponding sorting steps are made consistent with them.

– Main storage and data transfer facilities are used with maximum efficiency. Any method using a random access to the initial map is equivalent to an inefficient partial sorting of the densities needed by the operations we list in file ($\mathscr{I}$).

This system has been programmed in Fortran IV for IBM 370 machines, on which a highly efficient *SORT/MERGE* package is available. The half-word arithmetic facility is used to make the interpolation program work from a half-word e.d. map: the whole procedure of § 1 then requires only *one* full-word section of the initial map. This limitation is the same as in most fast Fourier transform programs, so that if an e.d. map can be calculated at all, it can be handled by this system. To deal efficiently with the large number of short records contained in the intermediate work files, I used a special input/output assembler routine written in this laboratory by Dr R. C. Ladner. The speed of execution of the three programs varies with the options chosen, but lies between 10 000 and 30 000 records per second on an IBM 370/165.

A domain of density can easily be extracted from a crystal and used to build a different crystal. This is of special interest in the solution of structures by analogy, when only an e.d. map (but no atomic coordinates) is available for the known structure. A mutant of human haemoglobin has been studied using the present map handling system (Anderson, 1975).

4. Handling of phase information

4.1. *Usefulness of a weighting scheme*

Having obtained the averaged e.d. map, we could simply calculate structure factors from it, and use their phases in the next iteration. This method is equivalent to that proposed by Crowther (1969) in his reciprocal space approach, and was used by Muirhead *et al.* (1967), Harrison & Jack (1975) and Argos *et al.* (1975). The convergence properties of the resulting algorithm were examined in (I),* and it was shown that it should converge for perfect data if the crystal asymmetric unit contains at least three identical molecules, and if the starting phases are good enough.

It seems desirable, however, to weight each structure factor according to the accuracy of its phase. Not only will the noise in the final e.d. map be reduced, but the use of weights *during the iteration itself* will make the algorithm less noise-sensitive, and hence widen its domain of convergence. Indeed, it will allow the well phased reflexions to phase the poorly phased ones, while minimizing the detrimental effect of the initially poorly determined phases.

Blow & Crick (1959) introduced such a weighting scheme for the combination of phase information obtained from different isomorphous derivatives. Their analysis of I.R. data produces, for each reflexion, a phase probability density $P_{\rm iso}(\alpha)$ which is used to compute the weighted Fourier coefficient:

$$|F|m_{\rm iso} \exp{(i\varphi_{\rm iso})} = |F| \int_0^{2\pi} P_{\rm iso}(\alpha) \exp{(i\alpha)} {\rm d}\alpha\,.$$

These structure factoıs yield the 'best' e.d. map, with the minimum expected mean-square error.

Symmetry-based phase information will now be cast in a similar probabilistic form, and combined with that given by I.R. so as to provide weights at all stages of the phase determination. A similar procedure was sketched by Buehner *et al.* (1974).

4.2. *The averaged structure as a source of phase information*

The simplest and most useful way of expressing the initial hypotheses of n.c.s. is to say that the structure should be invariant by the averaging/rebuilding operations of steps 3(*b*) and 3(*c*) (§ 1). In the early iterations, or even later if the symmetry is not exact, this invariance will be only approximate; nevertheless, the averaged electron density will still contain a 'substantial part' of the final solution.

Sim (1959) has shown how to use such a 'known part' of a structure, with Fourier coefficients $|F_K| \exp{(i\alpha_K)}$, to derive a phase probability density $P_{\rm Sim}(\alpha)$ for each Fourier coefficient $|F| \exp{(i\alpha)}$ of the total structure. If $\langle I_U \rangle$ is the mean intensity contributed by the unknown part of the structure, his formula can be written:

$$P_{\rm Sim}(\alpha) \propto \exp{\{[2|F|\,.\,|F_K|\,.\,\cos{(\alpha-\alpha_K)}]/\langle I_U \rangle\}}\,.$$

The larger the 'known' fraction, the smaller $\langle I_U \rangle$, hence the stronger the phase indications given. Sim's formula thus provides a probabilistic solution to the structure factor equations derived by Rossmann (1967) from the same hypotheses.

In complete analogy, the use of Sim's formula with the average e.d. as a known part gives the desired probabilistic solution to the equations implied by n.c.s. Estimating the quantity $\langle I_U \rangle$ by the mean discrepancy between observed and calculated intensities will automatically correlate the overall weight given to the symmetry with the closeness of fit after symmetrization.

4.3. *Combination of phase information*

Rossmann & Blow (1961) examined the problem of combining phase information from I.R. and from a partial knowledge of the structure, and proposed to multiply the corresponding phase probability densities, *i.e.* to use

$$P(\alpha) = P_{\rm iso}(\alpha)\,.\,P_{\rm Sim}(\alpha)\,.$$

It may not always be legitimate to consider $P_{\rm Sim}(\alpha)$ as independent from $P_{\rm iso}(\alpha)$. Indeed, the 'known part' will usually be either an interpreted and refined portion of an e.d. map phased by I.R. (Sweet, Wright, Janin, Chothia & Blow, 1974), or, in our case, the averaged version of a map computed with I.R. phase information. However, this dependence is very complex, relating globally all the $P_{\rm Sim}$'s to all the $P_{\rm iso}$'s. The only practical solution to this problem seems to be an inspired relative weighting of the two sources of phase information, as in

$$P(\alpha) = P_{\rm iso}(\alpha)^u P_{\rm Sim}(\alpha)^v.$$

This point will not be discussed any further.*

Rossmann & Blow pointed out that phase combination could be greatly simplified by an appropriate encoding of the functions involved. Indeed, $P_{\rm Sim}(\alpha)$ is completely described by the two numbers C_H and φ_H such that:

$$P_{\rm Sim}(\alpha) \propto \exp{[C_H \cos{(\alpha-\varphi_H)}]}\,.$$

Similarly, I.R. produces probability densities which can, with good accuracy, be cast in the form:

$$P_{\rm iso}(\alpha) \propto \exp{[C_{\rm is} \cos{(\alpha-\Phi_1)} - D_{\rm is} \cos{(\alpha-\Phi_2)}]}$$

* A factor of two was overlooked in the final expression relating the quadratic errors at cycles i and $i+1$; line 13 on page 405 of (I) should read:

$$\varepsilon_{i+1}^2 \le [\tfrac{3}{2} + (2\kappa)^{1/2}]\varepsilon_i^2\,.\,\kappa\,.$$

As $\kappa[\frac{3}{2} + (2\kappa)^{1/2}] < 1$ for $\kappa < 0{\cdot}4$ (instead of 0·5), the conclusions drawn in (I) remain practically unaffected.

* This problem is also present to some extent in the I.R. method: the refinement of the heavy-atom parameters of the different derivatives can meet with severe difficulties, due to biases (Blow & Matthews, 1973).

and can be coded by C_{is}, D_{is}, Φ_1 and Φ_2.

This analysis was extended by Hendrickson & Lattman (1970), who used standard functions of the form:

$$P_{ABCD}(\alpha) = \exp\,(A\cos\alpha + B\sin\alpha + C\cos 2\alpha + D\sin 2\alpha)$$

(overlooking normalization factors). The coefficients A, B, C, D are easier to handle than those chosen by Rossmann & Blow: combination of phase information from different sources amounts to a simple addition of their homologous coefficients. Such functions can also represent phase information from anomalous scattering and direct methods; they can even be made rigorously accurate for I.R. if the Blow & Crick error model is slightly altered, so as to use a lack of closure on intensities rather than on moduli. This complete system has been used by Hendrickson *et al.* (1973) in a 2 Å resolution study of sea lamprey haemoglobin, which includes a test application of the tangent formula.

When the original Blow & Crick error model is used, it is necessary to determine, for each reflexion, the values of A, B, C and D (hereafter called the 'phase coefficients') giving the best fit between $P_{\text{iso}}(\alpha)$ and $P_{ABCD}(\alpha)$. Hendrickson (1971) proposed using a least-squares fitting of their logarithms, *i.e.* to minimize the quantity:

$$\int_0^{2\pi} w(\alpha)[\ln P_{\text{iso}}(\alpha) - (A\cos\alpha + B\sin\alpha + C\cos 2\alpha + D\sin 2\alpha)]^2 \mathrm{d}\alpha\ ,$$

with a weight $w(\alpha)$ which would optimize the fit between the probabilities themselves. The weight $w(\alpha) = P_{\text{iso}}(\alpha)$ was found to be excellent. However, $w(\alpha) \equiv 1$ is still quite good and computationally cheaper since the phase coefficients are then simply the first four real Fourier coefficients of $\ln P_{\text{iso}}(\alpha)$.

4.4. *The phase combination program*

At the outset (step 1 of §1), we save in a coded form the initial probability densities $P_{\text{iso}}(\alpha)$ in a reference list. The phase coefficients are computed as described above, with weight $w(\alpha) \equiv 1$, in a routine grafted onto a program for the refinement of heavy-atom parameters which originated at Purdue (Adams *et al.*, 1969).

Before estimating the quantity $\langle I_U \rangle$ by comparing calculated and observed intensities, it is necessary to apply a temperature factor to the calculated structure factors. This correction allows for the following two effects:

(1) The structure factors used in the calculation of the e.d. map [step 3(*a*) of § 1] contain the observed moduli weighted by figures of merit. These fall off with increasing resolution, and their mean value usually behaves like an artificial temperature factor (Dickerson, Kendrew & Strandberg, 1961).

(2) The use of linear interpolation causes a loss of spectral power, which also increases with resolution. This effect is correctable by a temperature factor only if relatively fine grids are used; for coarse grids, a special profile correction is needed. This will be discussed at length in § 5.3.

The phase combination [step 3(*e*), §1] therefore proceeds in three passes:

1st pass. The reflexions common to the reference list and the calculated structure factor list are selected. A rephasing flag is set for each of those which are within the prescribed resolution limits, and whose figure of merit is less than a given maximum value (above this value, the phases are to be kept fixed). A scale factor and a temperature factor are determined by a Wilson plot to produce the best fit between the two intensity distributions.

2nd pass. The scale and temperature corrections are applied to the calculated structure factors. Statistics are made on the comparison of their phases with the current phases, and of their (corrected) moduli with the observed moduli. The quantity $\langle I_U \rangle$ is computed in resolution shells as $\langle |I_{\text{obs}} - \theta I_{\text{calc}}| \rangle$ where θ is the fraction of the total structure which should ideally be present in the 'known part'. When using an averaged structure, θ will be taken as 1, but for other applications [*e.g.* use of a model for part of the structure, as in Sweet *et al.* (1974)], θ may be less than 1.

3rd pass. Phase combination is performed, unless a simple use of the calculated phases has been requested. The phase coefficients taken from the reference list are incremented by those calculated from Sim's formula, to compute the new phases and figures of merit. These are used to update the reference list, and to produce Fourier coefficients of the form:

$$(w_{\text{obs}}|F|_{\text{obs}} - w_{\text{calc}}|F_{\text{calc}}|)\exp\,(i\alpha_{\text{combined}})\ ,$$

where w_{obs} and w_{calc} can be specified at will; in this way, any desired type of Fourier map may be obtained (typically, $F_o, F_o - F_c$ or $2F_o - F_c$ maps). Statistics are calculated on the phase shifts upon combination, both from the previous current phase and from the calculated phase, and on the changes of figure of merit.

The various statistics produced by this program provide a thorough check on the progress of the refinement. All the statistics are compiled as a function of resolution and current figure of merit. Those on moduli include an R index $(\sum |F|_{\text{obs}} - |F_{\text{calc}}| / \sum |F|_{\text{obs}})$, a correlation coefficient between $|F|_{\text{obs}}$ and $|F_{\text{calc}}|$, and the lack of closure defined above; those on phase shifts are made both on absolute shifts ($|\Delta\alpha|$) and on relative shifts ($|\Delta\alpha|/\cos^{-1} m$). The program is written in IBM Fortran IV; using a 5° interval in the calculation of the centroid phases, it can process 20 000 reflexions a minute.

5. Discussion

5.1. *Minimum starting phase information*

To what extent can the MRM be considered as autonomous (*i.e.* independent from I.R.) in the study of an unknown structure?

Rossmann (1972) distinguishes three successive problems to be solved in its application: (1) the rotation problem; (2) the translation problem; (3) the actual phase determination.

I would like to add a problem 2^A: the molecular envelope problem. Indeed, the role of the envelope is far from negligible: Argos *et al.* found that a true *ab initio* phasing of lobster GPD, starting with a spherical envelope, did not yield any useful result.

As mentioned previously, the availability of a single (intramolecular) isomorphous derivative will provide a solution to all four problems. Rotations and translations are determined from heavy-atom positions, either directly (for proper symmetiies) or after an e.d. correlation search on the SIR map (for improper symmetries). Averaging of the SIR map gives a reasonably accurate envelope. Refinement of the SIR phases by direct-space averaging and phase combination will usually solve the phase problem. It seems that information of comparable accuracy can be obtained without I.R. in the sole case of some spherical viruses, such as TBSV (Harrison, 1971) or SBMV (Johnson, Rossmann, Smiley & Wagner, 1974), which occupy a special position of a crystal lattice in which they express part of their icosahedral symmetry.

In the absence of any I.R. data, the rotation problem can be tackled using the method proposed by Rossmann & Blow (1962) and made more practical by Crowther's fast rotation function (Crowther, 1972). This method has had numerous successes, but its application is not free from pitfalls (see for example, Åkervall *et al.*, 1971). The determination of translations is a much more serious problem; the technique available (Rossmann, Blow, Harding & Coller, 1964) has been used to position local diads in α-chymotrypsin (Blow, Rossmann & Jeffery, 1964) and insulin (Dodson, Harding, Hodgkin & Rossmann, 1966), but is not of general applicability. On lobster GPD, the rotation function was able to determine correctly the non-crystallographic rotations (Rossmann, Ford, Watson & Banaszak, 1972) but the position of the molecular centre could only be found with the help of I.R. (Buehner *et al.*, 1974). Finally, apart from 'packing considerations', no method exists for the determination of protein molecular envelopes without phase information. For virus structures however, EM can provide a valuable help (see *e.g.* Finch, Gilbert, Klug & Leberman, 1974; Jack, Harrison & Crowther, 1975).

The phase determination stage may also be unmanageable without I.R. phase information, even if the other problems have been solved without it. A possible snag lies in enantiomorph ambiguity. Crowther (1969) showed that if the arrangement of envelopes was centrosymmetric, only the real parts of the structure factors could be determined by the exploitation of non-crystallographic symmetry alone. Therefore, phases generated from such information may not even be immediately useful to locate heavy atoms in isomorphous derivatives. This problem has not yet been encountered in practice: Jack (1973) solved a *centrosymmetric* projection of TMV; Argos *et al.* (1975) and Jack *et al.* (1975) circumvented it by using envelopes determined from I.R. and EM respectively. It is likely to be most frequent in the case of viruses, where the envelopes obtained 'for free' will usually be centrosymmetric.

Another source of difficulties may be the inability of the local rotations to provide a *homogeneous* interaction between all reflexions. An extreme case in this respect is the TMV disc, in which the 17-fold axis is almost parallel to the c^* axis. At low resolution, the phase constraints implied by this symmetry couple reflexions hkl and $h'k'l'$ very strongly for $l=l'$, but very loosely for $l\neq l'$. If an *ab initio* phasing is attempted, a segregation may occur among the planes of reciprocal space with different l values: phases will rapidly reach consistent values *within* each plane, while enantiomorph consistency *between* planes may take much longer to set in or even fail to do so, especially across planes of systematically weak reflexions. Phase determination or extension on TMV in three dimensions should therefore start with SIR phases for some non-centrosymmetric reflexions, at least within a slab containing the c^* axis, to ensure *global* enantiomorph consistency.

It seems therefore (at least in the present 'state of the art') that some data on a single isomorphous derivative will most often be a necessary prerequisite to the exploitation of n.c.s. in the solution of an unknown structure.

5.2. *Global and relative weighting in the phase combination*

The purpose of the internal weighting (§ 4.1) is to 'rectify' the interactions among phases, by allowing reliably phased reflexions to influence poorly phased ones while inhibiting the converse. Fig. 1 is a plot of the mean phase errors on *B.St.* GPD (relative to the final phases obtained by DIR and two cycles of averaging), as a function of starting figures of merit at various stages of the refinement. It shows that the weighting scheme does have the desired effect: the more accurate a phase is initially, the less it is altered during the refinement.

The relative weight given to MIR and MRM phase information is automatically determined, in our program, by the global agreement between observed and calculated moduli (§ 4.4). Recent work in this laboratory suggests that, when the two sources of phase information are truly independent, this relative weighting is essentially correct. In his study of human foetal haemoglobin (HbF), J. Frier (private communication, 1975) used data on the native protein and one isomorphous derivative to generate SIR phases; these were then combined with phases calculated from a human adult haemoglobin (HbA) molecule placed in the HbF unit cell. HbF differs from HbA by 39 residues out of 146 in the β-chain. Although these discrepant side chains were not deleted from the atomic coordinate list used to calculate model structure factors, the final HbF e.d. map shows unambiguous HbF side chains for most of them, with almost no residual HbA features. As the mean figure of merit rose from 0·47 to 0·77 during phase combination, it appears that the weight given to the calculated phases is strong enough to resolve the SIR phase ambiguity satisfactorily, but not so strong as to swamp the differences between the model and the structure under study. Encoding the complete SIR probability curve $P_{\rm SIR}(\alpha)$ by its phase coefficients certainly plays a crucial role in this case, in that it is able to discriminate between two classes of reflexions with low initial figures of merit $m_{\rm SIR}$: (1) those for which $P_{\rm SIR}$ is flat (no information present); (2) those for which $P_{\rm SIR}$ is bimodal (two well defined possibilities between which no choice is yet possible).

These two categories behave quite differently when combined with calculated phases by the Rossman & Blow method, and retaining this distinction is essential.

A simplified procedure has been used by Fletterick & Steitz (1975) to combine phases from two crystal forms of yeast hexokinase. The reliability of the I.R. phase is determined by its figure of merit; that of the calculated phase is inferred from local (rather than global) agreement between observed and calculated moduli. This method was found satisfactory when used with good MIR phases, in which bimodal probabilities are rare; $m_{\rm MIR}$ is then indeed a good indicator of the phase error. In a case such as HbF, however, the consideration of $m_{\rm SIR}$ alone would have led to an underestimation of the amount of phase information provided by SIR, and resulted in an excessive weight being given to the model phases.

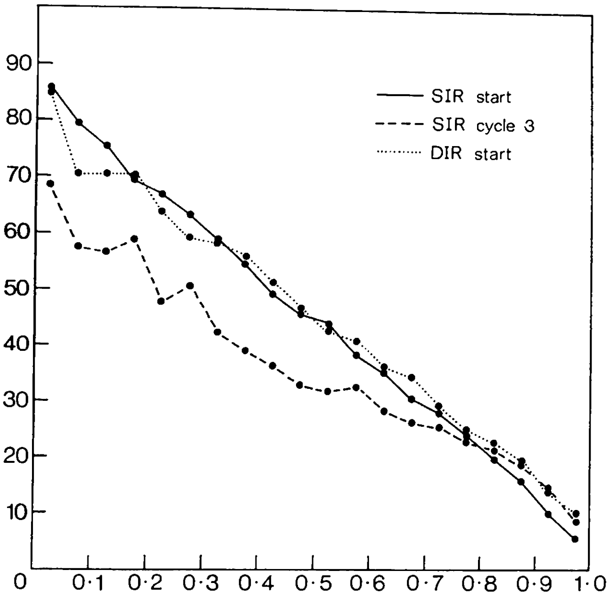

Fig. 1. Phase discrepancies at various stages of the phase refinement for *B. St.* GPD, as a function of the figure of merit *at these same stages.* The 'standard of truth' is the set of phases obtained after two cycles of refinement from DIR phase information (DIR cycle 2).

5.3. *Optimization of averaging computations*

To make the molecular replacement equations soluble in practice, a compromise has to be found between accuracy and size of the computations. This final paragraph is devoted to a study of the relation between gain in computational speed and loss of accuracy, which should be useful in this respect.

In the approximation thus far used in reciprocal space, both of these quantities are determined by a single parameter (the cut-off of the interference function), so that the dependence on N^2 mentioned in the introduction cannot be overcome.

The direct-space averaging method is more flexible. By proper choice of the sampling intervals in the initial and final e.d. maps, and application of a profile correction to the calculated structure factors, it is possible to perform fast and accurate calculations using linear interpolation. The computation time then varies as N, and the core requirements as $N^{2/3}$. With the program system described in §§ 2 to 4, such calculations are manageable in the 100 000 range.

5.3.1. *N^2-dependence of exact calculations*

A rigorous solution of the equations expressing non-crystallographic symmetry, whether in direct or in reciprocal space, is intrinsically a computation of size proportional to N^2 for N independent reflexions. This is due to the existence of confinement constraints in both spaces, with symmetrical roles: (1) in reciprocal space, data are available only within the resolution sphere S_Δ of radius Δ^{-1}(Å^{-1}); this causes the non-independence of e.d.'s sampled on a grid finer than $\Delta/2$, and yields an exact interpolation formula for them (Shannon, 1949); (2) in direct space, the protein density is bounded by the molecular envelope U; this is what causes each structure factor to interact with its neighbours and those related by noncrystallographic rotations, in a way formally identical to a Shannon interpolation [see (I), § 3.3.2].

These constraints (in principle incompatible) are most easily expressed multiplicatively in each space, as

$$\chi_{S_\Delta} \,.\, F=F \quad \text{and} \quad \chi_U \,.\, \varrho^0=\varrho^0 \,.$$

using the notation developed in (I). However, as we perform the averaging in one of these spaces only, the confinement constraint in the other space has to be treated by convolution, typically in a Shannon interpolation, which is the source of the N^2-dependence. Attempts to reduce the amount of computation involve an approximation of this convolution operation.

5.3.2. *Approximation in reciprocal space*

In the treatment of reciprocal-space equations, computation can be saved by truncating the interference function $G=1/U\mathscr{F}[\chi_U]$. This replaces the envelope indicator χ_U by its convolution $\chi_U*\mathscr{F}[\chi_T]$ with the

Fourier transform of the truncating function χ_T, thus defining the envelope U at a resolution lower than that at which phases are to be determined. The errors introduced are quite considerable, as discussed by Main (1967) and by Jack (1973) who sought to minimize them by overestimating the envelope dimensions. However, it is untenable to keep this cut-off fixed as the resolution is increased, so that the computational requirements still vary as N^2.

Another possibility would be to compensate the distortion of χ_U by multiplying the electron density in real space by

$$\chi_U/(\chi_U*\mathscr{F}[\chi_T])\,.$$

This correction would be difficult at the contacts between neighbouring aggregates, and could not prevent some smearing of fine details of the envelope. It bears a formal analogy to the direct-space method described below, but is less flexible.

5.3.3. *Approximation in direct space*

Shannon's interpolation can be replaced by linear interpolation. This amounts to convoluting the sampled density with a triangular wedge function, instead of the Fourier transform of the limiting resolution sphere. A detailed derivation of the effects of this approximation is given in the Appendix, which may be summarized as follows: the recomputed structure factors suffer an attenuation which increases with resolution, corresponding to the shape of the central peak of the transform of the wedge, and the signal lost reappears as random noise.

The attenuation can be compensated by a suitable profile adjustment of the calculated structure factors, applied before temperature scaling in the first pass of the phase combination (§ 4.4). The noise level can be reduced by the use of a fine enough sampling interval α in the initial e.d. map calculation [step 3(*a*), § 1]. On the other hand, the size ω of the grid on which the averaged map is computed need only be slightly finer than the coarsest grid allowed by Shannon's sampling criterion [see Appendix, equation (*A*6)], which is of size $\varDelta/2$ for calculations out to resolution $\varDelta$ (Å). As shown in Table 3, the amount of computation depends mainly on this final sampling interval ω, whereas α determines the maximum core requirements. The situation is thus more favourable than in reciprocal space, since accuracy can be increased while keeping the length of the calculations almost unchanged. Once the ratios $\alpha/\varDelta$ and $\omega/\varDelta$ have been chosen, the calculations can be carried out with constant relative accuracy.as the number N of independent reflexions is increased; their length varies as N, and the core requirements as $N^{2/3}$. This makes the direct-space averaging method workable in practice.

5.3.4. *Parameter optimization in direct-space averaging*

How oversampled should the initial e.d. map be? Let Σ be the total r.m.s. error due to linear interpolation, as estimated in the Appendix. Obviously, if it is larger than the errors in the observed moduli, the convergence will be limited by this systematic error rather than by the accuracy of the data. The desired criterion is thus that Σ *should be less than the standard deviation of the data at the outer resolution limit.*

Table 3. *Costs of the different steps of direct-space averaging, using simple interpolation* (§3.2) *for a proper n.c.s. of order n.*

The symbols used in this table are defined as: $\mathscr{N}_1$ = number of grid points in map no. 1; $\mathscr{N}_2$ = number of grid points in map no. 2 within the aggregate boundaries; $\mathscr{S}_i$ $(i=1,2)$ = number of grid points in one section of map no. *i*.

	Step name	Computation	Core size
Done only once	GENERATE	$n\mathscr{N}_2$	Fixed
	SORT 1	$n\mathscr{N}_2$	Adjustable
Done at each cycle	FOURIER SYNTHESIS	$\mathscr{N}_1$	$\mathscr{S}_1$
	INTERPOLATION	$n\mathscr{N}_2$	$\mathscr{S}_1$
	SORT 2	$n\mathscr{N}_2$	Adjustable
	RECONSTITUTION	$n\mathscr{N}_2$	$\mathscr{S}_2$
	FOURIER ANALYSIS	$\mathscr{N}_2$	$\mathscr{S}_2$

The degree of oversampling required is often underestimated. For example, Argos *et al.* (1975) did their calculations on lobster GPD with a grid interval $\alpha=2$Å out to resolution $\varDelta=4{\cdot}9$ Å. As double interpolation was used, the r.m.s. noise level, as given by equation (*A*8) of the Appendix ($n=4$, $n_I=4$, $\kappa=0{\cdot}15$), was 115% of the r.m.s. values of high-resolution moduli. Although this estimate is an upper bound, it accounts for the slow convergence and large final phase errors obtained. It also suggests that these results could be considerably improved by using a finer grid and simple interpolation. Similar considerations apply to the calculations of Muirhead *et al.* (1967).

The same formula (*A*8), applied to the calculations done on TMV ($\varDelta=5{\cdot}0$, $\alpha=0{\cdot}9$, $n=17$, $n_I=16$, $\kappa=0{\cdot}018$) gives an r.m.s. noise level of at most 11·8% at 5·0 Å resolution. On *B.St.* GPD, simple interpolation was used on a 0·45 Å grid at 2·7 Å resolution; equation (*A*7) of the Appendix, with $n=4$, $n_I=3$, $\kappa=0{\cdot}15$, gives an r.m.s. noise level of at most 4·2%. In both of these cases, the r.m.s. error due to linear interpolation was about a third of the mean standard deviation of the highest resolution data.

As a general rule, $\varDelta/\alpha$ should lie between 5 and 6, unless a very high symmetry is handled; $\varDelta/\omega$ need not exceed 2·5. These typical values are useful in choosing the averaging method (§ 3.2.1 and 3.2.2). As shown in Table 3, averaging by simple interpolation involves an amount of computing proportional to n times $\mathscr{N}_2$. If double interpolation is used, the intermediate grid $\mathscr{G}$ (§ 3.2.2) should be as fine as grid $\mathscr{G}_1$, since it also limits the accuracy; the amount of computation is then proportional to $\mathscr{N}+\mathscr{N}_2$, with $\mathscr{N}=(\omega/\alpha)^3\mathscr{N}_2$. As $(\omega/\alpha)^3$ is typically of the order of 15, double interpolation should be used only if $n \gg 16$. For $n \sim 16$, it would roughly double the noise level while saving little computation.

5.3.5. *Capabilities of the present programs*

With the parameters quoted above, one cycle of phase refinement by direct-space averaging takes about 12 minutes for *B.St.* GPD at 2·7 Å resolution ($N=$ 42 000) and 25 minutes for TMV at 5·0 Å resolution ($N=38$ 000) on an IBM 370/165 computer; in both cases, this is shorter than a cycle of heavy-atom parameter refinement. Early work on GPD, using $\omega=\alpha$, took 80 minutes per cycle to handle more than 10^7 density points; with ω determined as above, this amount of computation would be sufficient to average GPD at 1 Å resolution ($N=900$ 000). Phase determination to 2·8 Å on TMV is being undertaken ($N=212$ 000); one cycle will take 100 minutes on the same computer, with a maximum core size of 150K words, for a noise level of at most 5·5% at high resolution.

These figures compare favourably with previous work in reciprocal space; the calculation of the **H**-matrix for 1200 reflexions of barnase took one hour on the same computer (Jack, 1972).

A possible improvement would be to use a more sophisticated interpolation formula, involving a convolution with a truncated transform of the resolution sphere. This would decrease the rate of oversampling needed to achieve a given accuracy, and thus reduce the size of the sections of the initial e.d. map. On the other hand, the interpolation program would need to keep more sections in core at a time; its logical complexity and the amount of computation per point would be increased. The optimization of this choice will be very much dependent upon the type of computer used.

In any case, the figures quoted above show that the programs as they stand can already cope with the largest problems which present data collection techniques can handle.

I wish to thank Dr D. M. Blow for frequent discussions during the course of this work, and for his critical reading of the manuscript. Discussions with Dr A. J. Wonacott were particularly helpful during program elaboration. Drs P. Argos and T. Steitz made some useful remarks on parts of this paper. I am grateful to Dr J. White for carrying out the modifications of his tracing system needed for the drawing of molecular envelopes, and to Dr R. Ladner for the gift of his routine *BLOCKBIN*. I thank the International Business Machines Corporation for creating the IBM 370 and *SORT/MERGE*, and the Délégation Générale à la Recherche Scientifique et Technique for financial support.

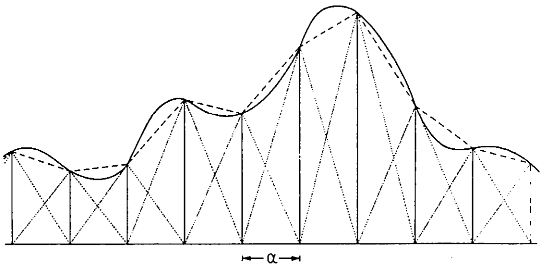

Fig. 2. Linear interpolation as a convolution of the sampled function with a triangular wedge. ——— ϱ; – – – $I_\alpha\varrho$; . . . copies of the wedge W_α.

APPENDIX

Effects of linear interpolation on calculated structure factors

This Appendix presupposes some familiarity with elementary distribution theory, such as presented in Schwartz (1967), and used in (I).

A.1. *Linear interpolation in one dimension*

Let ϱ be any complex valued function of a real variable, supposed to have an inverse Fourier transform $F=\bar{\mathscr{F}}[\varrho]$ (F may be a distribution).

Linear interpolation of ϱ on a grid of size α will produce a function $I_\alpha\varrho$, which may be written as the convolution product:†

$$I_\alpha\varrho=[\varrho\,.\sum_{l=-\infty}^{+\infty}\delta_{(l\alpha)}]*W_\alpha$$

where $\delta_{(x)}$ denotes a Dirac distribution at point x, and W_α is the triangular wedge function of half-width α, defined by:

$$\begin{cases} W_\alpha(x)=0 & \text{if } |x|>\alpha\,, \\ W_\alpha(x)=1-|x|/\alpha & \text{if } |x|<\alpha\,. \end{cases}$$

This representation is made clear by Fig. 2.

The inverse Fourier transform G_α of $I_\alpha\varrho$ is readily obtained, since

$$\bar{\mathscr{F}}[W_\alpha](\xi)=\frac{1}{\alpha}\left(\frac{\sin\pi\alpha\xi}{\pi\xi}\right)^2$$

and

$$\bar{\mathscr{F}}\left[\sum_{l=-\infty}^{+\infty}\delta_{(l\alpha)}\right]=\frac{1}{\alpha}\sum_{n=-\infty}^{+\infty}\delta_{(n/\alpha)}\,.$$

$\bar{\mathscr{F}}$ exchanges multiplication and convolution, giving the general formula:

$$G_\alpha(\xi)=\frac{1}{\alpha}\left(\frac{\sin\pi\alpha\xi}{\pi\xi}\right)^2\left[F*\frac{1}{\alpha}\sum_{n=-\infty}^{+\infty}\delta_{(n/\alpha)}\right],$$

i.e.

$$G_\alpha(\xi)=\left(\frac{\sin\pi\alpha\xi}{\pi\alpha\xi}\right)^2\sum_{n=-\infty}^{+\infty}F\left(\xi-\frac{n}{\alpha}\right).\tag{A1}$$

This relation between F and G_α is linear. As any e.d. map is a linear combination of complex plane waves, we can limit ourselves to considering this correspondence for the one-dimensional wave:

$$\varrho(x)=\exp\left[-\frac{2\pi i}{\lambda}(x-x_0)\right]=\exp\,(i\varphi)\exp\left(-\frac{2\pi i}{\lambda}x\right)$$

† This representation was pointed out to me by Dr R. Diamond.

with structure factor:

$$F=\mathscr{F}[\varrho]=\exp{(i\varphi)}\;\delta_{(1/\lambda)}\;.$$

The transform G_α of $I_\alpha\varrho$ is then easily found to be:

$$G_\alpha(\xi)=\exp{(i\varphi)}\left(\frac{\sin \pi\dfrac{\alpha}{\lambda}}{\pi\dfrac{\alpha}{\lambda}}\right)^2\sum_{n=-\infty}^{+\infty}\frac{\delta_{(1/\lambda+n/\alpha)}}{\left(1+\dfrac{n\lambda}{\alpha}\right)^2}\,.\qquad (A2)$$

This expression shows, as illustrated on Fig. 3, that the effect of linear interpolation is to multiply the initial structure factor by

$$A\left(\frac{\alpha}{\lambda}\right)=\left(\frac{\sin \pi\dfrac{\alpha}{\lambda}}{\pi\dfrac{\alpha}{\lambda}}\right)^2$$

and to create 'ghost' terms corresponding to the periodicity of the initial sampling sequence. Fig. 4 shows a plot of the attenuation factoı A as a function of the ratio α/λ.

For small enough values of this ratio ($\alpha/\lambda\le 0{\cdot}3$), the fall-off of A can be accurately represented by a temperature factor; equating the second derivatives of $A(\alpha x)$ and $\exp{(-Bx^2/4)}$ with respect to x for $x=0$ gives the relation: $B=\frac{4}{3}\pi^2\alpha^2$.

Note that the sum of the weights of all the peaks is preserved since the classical identity:

$$\sum_{n=-\infty}^{+\infty}\frac{1}{(z+n)^2}=\left(\frac{\pi}{\sin \pi z}\right)^2$$

[see for example Cartan (1961), pp. 152–153] implies that

$$\left(\frac{\sin \pi\alpha/\lambda}{\pi\alpha/\lambda}\right)^2\sum_{n=-\infty}^{+\infty}\left(1+n\,\frac{\lambda}{\alpha}\right)^{-2}=1\,.\qquad (A3)$$

Let ϱ now be a band-limited function, that is, one whose inverse Fourier transform $F(\xi)$ vanishes for $|\xi|>1/\Delta$. Then, by equation (A1), the transform G_α of $I_\alpha\varrho$ will consist of a 'main band'

$$\left(\frac{\sin \pi\alpha\xi}{\pi\alpha\xi}\right)^2 F(\xi)$$

corresponding to $n=0$, and an infinity of ghost bands ($n\ne 0$), so that G_α is no longer band-limited (see Fig. 5).

In practice, $I_\alpha\varrho$ will be computed on another grid, of size ω, with an origin shifted by $\varepsilon(0\le\varepsilon<\omega)$. The result of this computation will be $I_\alpha\varrho$ sampled on this final grid, *i.e.*

$$J_{\alpha\omega\varepsilon}=I_\alpha\varrho\,.\sum_{r=-\infty}^{+\infty}\delta_{(r\omega+\varepsilon)}\,,$$

whose inverse Fourier transform is:

$$H_{\alpha\omega\varepsilon}(\xi)=\left[\exp{(2\pi i\varepsilon\xi)}\,\frac{1}{\omega}\sum_{p=-\infty}^{+\infty}\delta_{(p/\omega)}\right]*G_\alpha\,,$$

i.e.

$$H_{\alpha\omega\varepsilon}(\xi)=(1/\omega)\sum_{p=-\infty}^{+\infty}\exp{[2\pi ip(\varepsilon/\omega)]}\;G_\alpha\left(\xi-\frac{p}{\omega}\right)\;(A4)$$

where G_α is given by (A1).

$H_{\alpha\omega\varepsilon}$ contains an infinite number of copies of G_α [$G_x^{(p)}$, labelled by index p, with a phase shift $2\pi p\varepsilon/\omega$], spaced by $1/\omega$; each $G_\alpha^{(p)}$ in turn consists of an infinite number of bands, labelled by index n, spaced by $1/\alpha$. This is shown on Fig. 6.

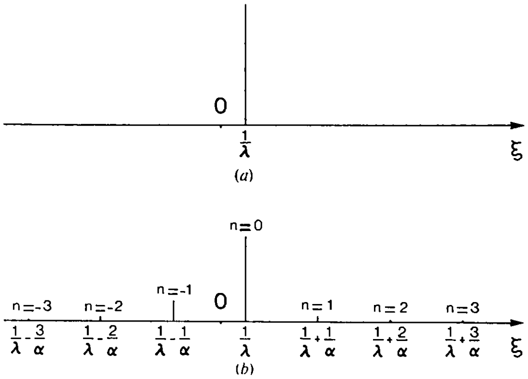
0
$\frac{1}{\lambda}$
ξ
(a)
n=0
n=-1
0
n=-3
n=-2
n=1
n=2
n=3
$\frac{1}{\lambda}-\frac{3}{\alpha}$
$\frac{1}{\lambda}-\frac{2}{\alpha}$
$\frac{1}{\lambda}-\frac{1}{\alpha}$
$\frac{1}{\lambda}$
$\frac{1}{\lambda}+\frac{1}{\alpha}$
$\frac{1}{\lambda}+\frac{2}{\alpha}$
$\frac{1}{\lambda}+\frac{3}{\alpha}$
ξ
(b)

Fig. 3. Effect of linear interpolation on the spectrum of a plane wave. (a) Transform before interpolation. (b) Transform after interpolation, showing attenuated main peak and ghost peaks.

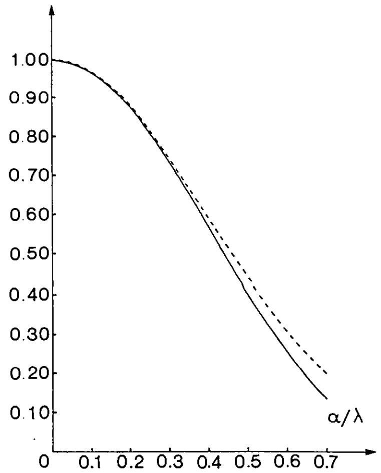

Fig. 4. Attenuation of main peak. ——— attenuation function $A(\alpha/\lambda)=\left(\dfrac{\sin \pi\alpha/\lambda}{\pi\alpha/\lambda}\right)^2$; – – – equivalent temperature effect $B(\alpha/\lambda)=\exp{\left(-\dfrac{\pi^2}{3}\,\dfrac{\alpha^2}{\lambda^2}\right)}$.

Calculation of structure factors from $J_{\alpha\omega\varepsilon}$ to resolution Δ (Å) will give $H_{\alpha\omega\varepsilon}$ in the range $-(1/\Delta)\leq\xi\leq(1/\Delta)$. If we follow Shannon's sampling criterion for ϱ and make ω less than $\Delta/2$, then $1/\omega>2/\Delta$, and bands with same n but different p do not overlap; in particular, the main bands of the various $G_\alpha^{(p)}$ will be separated. But, because $I_\alpha\varrho$ is not band-limited, no value of ω can prevent the overlap of the main band of each $G_\alpha^{(p)}$ with the ghost bands of all the $G_\alpha^{(q)}$ with $q\neq p$. Only if ω/α is integral and $\varepsilon=0$ (*i.e.* if no interpolation actually takes place) will contributions from main and ghost bands interfere so as to restore the original distribution F, as a consequence of the summation formula ($A3$). If ω/α is integral, while ε is not zero but is uniformly distributed between 0 and ω, random phase errors are introduced into the contributions with $p\neq0$. If ω/α is not integral, these contributions are out of register. Both effects will be present in a general interpolation, so that all the signal lost from the main peak will appear as random noise. The amount of noise picked up in the range $-(1/\Delta)\leq\xi\leq(1/\Delta)$ is quite insensitive to the value of ω, unless $1/\omega$ is large enough to separate patterns consisting of a main band *and some ghost bands*. But the elimination of only first-order ghosts would require that $1/\omega$ be greater than $2/\Delta+1/\alpha$, *i.e.* $\omega\geq\Delta/[2+(\Delta/\alpha)]$; this is unreasonably fine, and the same noise reduction can be achieved more economically by decreasing α (see § A.4).

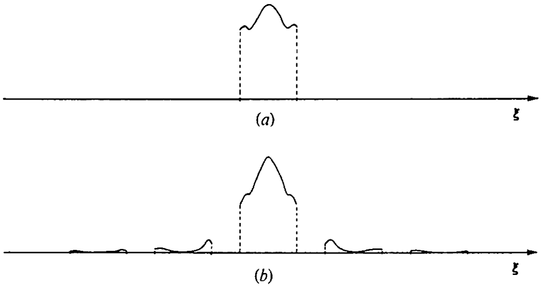

Fig. 5. Effect of linear interpolation on the spectrum of a band-limited function. (a) Transform before interpolation. (b) Transform after interpolation, showing fall-off of main band, and ghost bands.

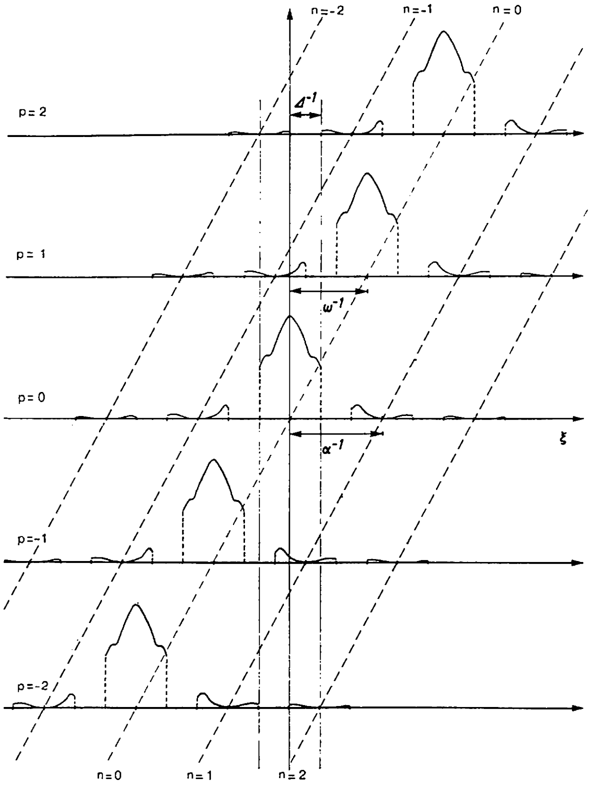

Fig. 6. Decomposition of $H_{\alpha\omega\varepsilon}$.

A.2. *Linear interpolation in three dimensions*

When this argument is generalized to three dimensions, the conclusions are exactly the same as in one dimension. The fall-off of calculated structure factors F_{hkl} is of the form:

$$A=\left[\frac{\sin(\pi h\alpha a^*)}{\pi h\alpha a^*}\right]^2\left[\frac{\sin(\pi k\beta b^*)}{\pi k\beta b^*}\right]^2\left[\frac{\sin(\pi l\gamma c^*)}{\pi l\gamma c^*}\right]^2. \qquad (A5)$$

It is anisotropic, but this can be neglected if the grid intervals α, β, and γ are almost equal. The pattern of main terms and ghost terms becomes very complex if averaging is performed, which justifies the statistical treatment given in § A.4.

A.3. *Choice of the final sampling interval* ω

The interval ω will need to be slightly larger than its minimum value $\Delta/2$ if envelope constraints are used in real space. Truncation of an e.d. map ϱ by an envelope U is equivalent to convoluting its transform F with the interference function $(1/U)\mathscr{F}[\chi_U]$, which produces some smearing of F. This effect allows some degree of phase extension to resolution higher than Δ. To separate such smeared bands, ω should be determined from

$$\frac{1}{\omega}=2\left(\frac{1}{\Delta}+\frac{m+0\cdot5}{2R}\right) \qquad (A6)$$

where R is the radius of a sphere roughly equivalent to U, and m is the number of peaks of the interference function considered as significant (typically, $m=1$).

A.4. *Estimation of errors*

It follows from §§ A.1 and A.2 that a fraction $1-A(\alpha d^*_{hkl})$ of each structure factor F_{hkl} is converted into noise by linear interpolation, where the isotropic attenuation factor A is given by:

$$A(\alpha d^*_{hkl})=\left(\frac{\sin\pi\alpha d^*_{hkl}}{\pi\alpha d^*_{hkl}}\right)^2.$$

For fixed α, the quantity $1-A(\alpha d^*_{hkl})$ varies parabolically as d^*_{hkl} goes to 0 (see Fig. 4); on the other hand, the moduli $|F_{hkl}|$ have an approximately Gaussian profile. Therefore, the product $|F_{hkl}|\ [1-A(\alpha d^*_{hkl})]$ increases with d^*_{hkl}, so that most of the noise comes from high-ıesolution terms; this is enhanced by the fact that the number of terms in a given resolution range is proportional to $(d^*)^2$.

The complex pattern of overlap between main and ghost bands will result in noise being spread over *all* terms, but we want an upper estimate of the error at high resolution. This will be obtained by considering interaction between high-resolution terms only, and neglecting the phase incoherence between the various ghost bands contributing to the noise. With these pessimistic approximations, the noise can be considered as having r.m.s. value:

$$\sigma=\langle|F_{hkl}|[1-A(\alpha d^*_{hkl})]\rangle_{\substack{\text{r.m.s.}\\ \text{hr}}}\le\langle|F|\rangle_{\text{hr}}(1-A)$$

where $\langle|F|\rangle_{\text{hr}}$ is the r.m.s. value of $|F_{hkl}|$ at high resolution, and A is the value of $A(\alpha d^*_{hkl})$ at the outer resolution limit.

The relative r.m.s. error at high resolution is thus: $\sigma/\langle|F|\rangle_{\text{hr}}=1-A$ for each interpolation. If a local symmetry of order n is handled by simple interpolation (§ 3.2.1), and if n_{I} interpolated densities are needed to compute the average density at each point, the mean square error introduced by the linear interpolations is $\sigma^2(n_{\text{I}}/n^2)$; this adds to the mean square error of the current estimates of the structure factors caused by incorrect phases. As each cycle multiplies all mean square errors by $\kappa=U/v$ [see (I)], the total mean square error introduced by interpolations as the process is iterated is at most:

$$\sigma^2(n_{\text{I}}/n^2)\ (1+\kappa+\kappa^2+\ldots)=(\sigma^2/n^2)\ \frac{n_{\text{I}}}{1-\kappa}\ .$$

The structure factors calculated from the averaged map are subsequently profile-corrected by division by $A(\alpha d^*_{hkl})$, so that the final estimate for the maximum r.m.s. error Σ is given by:

$$\frac{\Sigma}{\langle|F|\rangle_{\text{hr}}}=\frac{1-A}{nA}\sqrt{\frac{n_{\text{I}}}{1-\kappa}}\ . \qquad (A7)$$

If double interpolation is used, the second interpolation converts fraction $(1-A)$ of the intermediate structure factors (already attenuated by A) into noise, supposing that the intermediate grid is also of size α. This further noise is not correlated with the first, so that only their mean-square values add. Recycling of this error results in the same factor $(1-\kappa)^{-1/2}$ as previously. Finally, the total attenuation is A^2, so the profile correction factor is $1/A^2$, giving:

$$\Sigma=\langle|F|\rangle_{\text{hr}}\left[(1-A)^2\,\frac{n_{\text{I}}}{n^2}+A^2(1-A)^2\right]^{1/2}\frac{1}{\sqrt{1-\kappa}}\,\frac{1}{A^2}\ .$$

i.e.

$$\frac{\Sigma}{\langle|F|\rangle_{\text{hr}}}=\frac{1-A}{A^2\sqrt{1-\kappa}}\sqrt{\frac{n_{\text{I}}}{n^2}+A^2}\ . \qquad (A8)$$

References

Adams, M. J., Haas, D. J., Jeffery, B. A., McPherson, A. Jr, Mermall, H. L., Rossmann, M. G., Schevitz, R. W. & Wonacott, A. J. (1969). *J. Mol. Biol.* **41**, 159–188.

Åkervall, K., Strandberg, B., Rossmann, M. G., Bengtsson, U., Fridborg, K., Johannisen, H., Kannan, K. K., Lövgren, S., Petef, G. & Öberg, B. (1971). *Cold Spring Harbor Symp. Quant. Biol.* **36**, 469–488.

Anderson, N. L. (1975). *J. Mol. Biol.* **94**, 33–49.

Argos, P., Ford, G. C. & Rossmann, M. G. (1975). *Acta Cryst.* A**31**, 499–506.

Birktoft, J. J. & Blow, D. M. (1972). *J. Mol. Biol.* **68**, 187–240.

Blow, D. M. & Crick, F. H. C. (1959). *Acta Cryst.* **12**, 794–802.

Blow, D. M. & Matthews, B. W. (1973). *Acta Cryst.* A**29**, 56–62.

Blow, D. M., Rossmann, M. G. & Jeffery, B. A. (1964). *J. Mol. Biol.* **8**, 65–78.

Bricogne, G. (1974). *Acta Cryst.* A**30**, 395–405.

Buehner, M., Ford, G. C., Moras, D., Olsen, K. W. & Rossmann, M. G. (1974). *J. Mol. Biol.* **82**, 563–585.

Cartan, H. (1961). *Théorie des Fonctions Analytiques.* Paris: Hermann.

Champness, J. N., Bloomer, A. C., Bricogne, G., Butler, P. J. G. & Klug, A. (1976). *Nature, Lond.* **259**, 20–24.

Crowther, R. A. (1967). *Acta Cryst.* **22**, 758–764.

Crowther, R. A. (1969). *Acta Cryst.* B**25**, 2572–2580.

Crowther, R. A. (1972). In *The Molecular Replacement Method*, edited by M. G. Rossmann. New York: Gordon & Breach.

Cullis, A. F., Muirhead, H., Perutz, M. F., Rossmann, M. G. & North, A. C. T. (1962). *Proc. Roy. Soc.* A**265**, 161–187.

Dickerson, R. E., Kendrew, J. C. & Strandberg, B. E. (1961). In *Computing Methods and the Phase Problem in X-ray Crystal Analysis*, pp. 236–251. Oxford: Pergamon Press.

Dodson, E., Harding, M. M., Hodgkin, D. C. & Rossmann, M. G. (1966). *J. Mol. Biol.* **16**, 227–241.

Finch, J. C., Gilbert, P. F. C., Klug, A. & Leberman, R. (1974). *J. Mol. Biol.* **86**, 183–192.

Fletterick, R. J. & Steitz, T. A. (1975). Personal communication.

Harrison, S. C. (1971). *Cold Spring Harbor Symp. Quant. Biol.* **36**, 495–501.

Harrison, S. C. & Jack, A. (1975). *J. Mol. Biol.* **97**, 173–191.

Hendrickson, W. A. (1971). *Acta Cryst.* B**27**, 1472–1473.

Hendrickson, W. A. & Lattman, E. E. (1970). *Acta Cryst.* B**26**, 136–143.

Hendrickson, W. A., Love, W. E. & Karle, J. (1973). *J. Mol. Biol.* **74**, 331–361.

Jack, A. (1972). Ph.D. Thesis, Univ. of Cambridge.

Jack, A. (1973). *Acta Cryst.* A**29**, 545–554.

Jack, A., Harrison, S. C. & Crowther, R. A. (1975). *J. Mol. Biol.* **97**, 163–172.

Johnson, J. E., Rossmann, M. G., Smiley, I. E. & Wagner, M. A. (1974). *J. Ultrastruct. Res.* **46**, 441–451.

Main, P. (1967). *Acta Cryst.* **23**, 50–54.

Main, P. & Rossmann, M. G. (1966). *Acta Cryst.* **21**, 67–72.

Muirhead, H., Cox, J. M., Mazzarella, C. & Perutz, M. F. (1967). *J. Mol. Biol.* **28**, 117–156.

Rossmann, M. G. (1967). *Acta Cryst.* **23**, 173.

Rossmann, M. G. (1972). *The Molecular Replacement Method*. New York: Gordon & Breach.
Rossmann, M. G. & Blow, D. M. (1961). *Acta Cryst.* **14**, 641–647.
Rossmann, M. G. & Blow, D. M. (1962). *Acta Cryst.* **15**, 24–31.
Rossmann, M. G. & Blow, D. M. (1963). *Acta Cryst.* **16**, 39–45.
Rossmann, M. G., Blow, D. M., Harding, M. M. & Coller, E. (1964). *Acta Cryst.* **17**, 338–342.
Rossmann, M. G., Ford, G. C., Watson, H. C. & Banaszak, L. J. (1972). *J. Mol. Biol.* **64**, 237–249.
Schwartz, L. (1967). *Méthodes Mathématiques pour les Sciences Physiques*. Paris: Hermann.
Shannon, C. E. (1949). *Proc. Inst. Radio Engrs. New York*, **37**, 10–21.
Sim, G. A. (1959). *Acta Cryst.* **12**, 813–815.
Sweet, R. M., Wright, H. T., Janin, J., Chothia, C. H. & Blow, D. M. (1974). *Biochemistry*, **13**, 4212–4228.
Ten Eyck, L. F. (1973). *Acta Cryst.* **A29**, 183–191.

Acta Cryst. (1976). A**32**, 847

Applications of the Ewald Method. I. Calculation of Multipole Lattice Sums

By P. G. Cummins* and D. A. Dunmur
Department of Chemistry, The University, Sheffield S3 7HF, England
and R. W. Munn and R. J. Newham
Department of Chemistry, UMIST, Manchester M60 1QD, England

(*Received* 20 *February* 1976; *accepted* 18 *March* 1976)

General principles of the Ewald method for evaluating multipole lattice sums are reviewed. The method is used to derive an expression for the Lorentz-factor dipole tensor sum in a form convenient for computation, and comparisons are made with the direct and plane-wise summation methods. Expressions are also given for computing quadrupole and octopole sums by the Ewald method. The effect of crystal symmetry on lattice sums is outlined; the number of independent sums relating different pairs of equivalent sublattices does not exceed the total number of such sublattices. Numerical results are given for the dipole lattice sums of hydrogen cyanide, benzene, durene, anthracene and pyrene. Quadrupole sums are given for cuprous chloride and pyrene, and octupole sums are given for hydrogen cyanide, benzene and anthracene. For dipole lattice sums, the Ewald method converges much faster than direct summation; for higher multipole sums, the Ewald method has no special advantage in speed, but may prove convenient, especially when sums are required for strained lattices.

Introduction

Quantitative microscopic interpretation of many physical properties of crystals requires a knowledge of some type of lattice sum. Calculations of the internal energies of crystals involve a wide range of lattice sums (charge–charge, dipole–dipole, quadrupole–quadrupole, *etc.*), depending on the form of the potential function assumed (Born & Huang, 1954; Rae, 1969; Craig, Mason, Pauling & Santry, 1965; Aung & Strauss, 1973). Similarly, the interpretation of electronic spectra of crystals requires the evaluation of dipole–dipole and higher-order lattice sums (Craig & Walmsley, 1963; Decius, 1968; Philpott & Lee, 1973; Frech, 1973). The response of crystals to electric fields as measured by their dielectric properties (Agranovich, 1974; Sinha, Gupta & Price, 1974; Bolton, Fawcett & Gurney, 1962; Tessman, Kahn & Shockley, 1953; Koikov & Rozova, 1967) or Stark spectroscopy (Hochstrasser, 1973; Dunmur & Munn, 1975; Chen, Hanson & Fox, 1975) again requires a knowledge of appropriate lattice sums for its microscopic interpretation. The effect of static or dynamic strain on all these properties is principally due to changes in the lattice sums, which in turn may be expressed in terms of higher-order lattice sums. There is therefore ample reason for the continuing interest in methods for evaluating lattice sums of various types (Hove & Krumhansl, 1953; De Wette & Schacher 1965; Bruesch & Lietz, 1970; Philpott, 1973; Aung & Strauss, 1973; Philpott & Mahan, 1973).

Two basic approaches may be followed in evaluating lattice sums: summation of the appropriate function over all points of the direct lattice, or summation after transformation from the direct to some other lattice, usually the reciprocal lattice. There are however difficulties with either approach. Values of certain direct lattice sums are only conditionally convergent, and it becomes necessary to define a summation shape outside which all lattice points are excluded (Philpott & Lee, 1973; Burrows & Kettle, 1975). This problem

[*] Present address: Department of Chemistry, University of Southern California, Los Angeles, California 90007, U.S.A.