OSAKABE, N., TANISHIRO, Y., YAGI, K. & HONJO, G. (1981). *Surf. Sci.* **109**, 353-366.
PARKER, M. A. & SINCLAIR, R. (1986). In *Electron Microscopy* 1986, edited by T. IMURA, S. MARUSE & T. SUZUKI, Vol. 2, pp. 991-992. Tokyo: Japanese Society of Electron Microscopy.
SAXTON, W. O. (1978). *Computer Techniques for Image Processing in Electron Microscopy*. New York: Academic Press.
SAXTON, W. O. (1980). *J. Microsc. Spectrosc. Electron.* **5**, 665-674.
SAXTON, W. O., HOWIE, A., MISTRY, A. & PITT, A. (1977). *Inst. Phys. Conf. Ser.* No. 36, pp. 119-122.
SAXTON, W. O. & SMITH, D. J. (1985). *Ultramicroscopy*, **18**, 39-48.
SAXTON, W. O., SMITH, D. J. & ERASMUS, S. J. (1983). *J. Microsc. (Oxford)*, **130**, 187-201.
SCHISKE, P. (1973). In *Image Processing and Computer-Aided Design in Electron Optics*, edited by P. W. HAWKES, pp. 82-90. London: Academic Press.
SELF, P. G., O'KEEFE, M. A., SPARGO, A. E. & BUSECK, P. R. (1984). *Ultramicroscopy*, **11**, 35-62.
SHUMAN, H. & KRUIT, P. (1985). *Rev. Sci. Instrum.* **56**, 231-239.
SMITH, D. J. (1985). *J. Vac. Sci. Technol.* **B3**, 1563-1566.
SMITH, D. J. (1986*a*). In *Chemistry and Physics of Solid Surfaces* VI, edited by R. VANSELOW & R. HOWE, Chap. 15. Berlin: Springer-Verlag.
SMITH, D. J. (1986*b*). *Surf. Sci.* **178**, 462-474.
SMITH, D. J., BURSILL, L. A., BOVIN, J.-O., PETFORD-LONG, A. K. & YE, H. Q. (1986). *Surf. Interface Anal.* **10**, 135-141.
SMITH, D. J., BURSILL, L. A. & WOOD, G. J. (1983). *J. Solid State Chem.* **50**, 51-69.
SMITH, D. J., BURSILL, L. A. & WOOD, G. J. (1985). *Ultramicroscopy*, **16**, 19-32.
SMITH, D. J., CAMPS, R. A. & FREEMAN, L. A. (1982). *Inst. Phys. Conf. Ser.* No. 61, pp. 381-386.
SMITH, D. J. & MARKS, L. D. (1985). *Ultramicroscopy*, **16**, 101-114.
SMITH, D. J. & O'KEEFE, M. A. (1983). *Acta Cryst.* **A39**, 139-148.
SMITH, D. J., PETFORD, A. K., WALLENBERG, L. R & BOVIN, J.-O. (1986). *Science*, **233**, 872-875.
SMITH, D. J., SAXTON, W. O., O'KEEFE, M. A., WOOD, G. J. & STOBBS, W. M. (1983). *Ultramicroscopy*, **11**, 263-281.
SPENCE, J. C. H. (1983). *Ultramicroscopy*, **11**, 117-124.
STOBBS, W. M. (1983). *J. Microsc. (Oxford)*, **131**, 47-53.
STOBBS, W. M. & SAXTON, W. O. (1987). *J. Microsc. (Oxford)*. In the press.
TAKAYANAGI, K. (1983). *Jpn. J. Appl. Phys.* **22**, L4-6.
TAKAYANAGI, K., TANISHIRO, Y., TAKAHASHI, S. & TAKAHASHI, M. (1985). *Surf. Sci.* **164**, 367-392.
TANISHIRO, Y., TAKAYANAGI, K. & YAGI, K. (1983). *Ultramicroscopy*, **11**, 95-102.
THON, F. (1971). In *Electron Microscopy in Materials Science*, edited by U. VALDRE, pp. 570-625. New York and London: Academic Press.
TONOMURA, A. (1986). In *Electron Microscopy* 1986, edited by T. IMURA, S. MARUSE & T. SUZUKI, Vol. 1, pp. 9-14. Tokyo: Japanese Society of Electron Microscopy.
TROYON, M. (1976). *Optik (Stuttgart)*, **46**, 439-450.
VENABLES, J. A., SMITH, D. J. & COWLEY, J. M. (1987). *Surf. Sci.* **181**, 235-249.
WILSON, A. R., SPARGO, A. E. & SMITH, D. J. (1981). *Optik (Stuttgart)*, **61**, 63-78.
YAGI, K., TAKAYANAGI, K., NAGAKURA, S., KOBAYASHI, K., HONJO, G., KUBOZOE, M., SHII, K., MATSUI, I., KATAGIRI, S. & TANAKA, T. (1982). In *Electron Microscopy* 1982, Vol. 1, pp. 375-376. Frankfurt: Deutsche Gesellschaft für Elektronenmikroskopie e.V.
ZEMLIN, F., WEISS, K., SCHISKE, P., KUNATH, W. & HERRMANN, K.-H. (1978). *Ultramicroscopy*, **3**, 49-60.

Acta Cryst. (1987). A**43**, 751-763

Optimal Symbolic Phase Determination

BY RENÉ PESCHAR AND HENK SCHENK

Laboratory of Crystallography, University of Amsterdam, Nieuwe Achtergracht 166, 1018 *WV Amsterdam, The Netherlands*

(*Received* 15 *October* 1986; *accepted* 21 *May* 1987)

Abstract

A systematic method of phase determination is presented inspired by the dynamic-programming principle. In this new procedure the starting set and the best phasing sequence are determined while executing the symbolic phase determination itself. Special attention is given to the way in which statistical weights for symbolic phase indications can be calculated. Test results show that the new procedure leads to considerable improvements over the phase-determination procedures based on the convergence procedure currently available.

Introduction

The phase problem is often tackled successfully by a default run of a direct-method structure-determination program, such as *SIMPEL* (Overbeek & Schenk, 1978; Schenk & Kiers, 1985), *MULTAN* (Main, 1985), *SHELX* (Sheldrick, 1985) or *GENTAN* (Hall, 1985). However, the structure is not always found immediately. Although the reasons for failure may vary, the user of a program should at least be guaranteed that under the given basic probabilistic assumptions the best possible phase determination is carried out; that is, the chance of obtaining the correct solu-

0108-7673/87/060751-13$01.50

tion should be maximized. For the current direct-method packages such a guarantee cannot be given, whether numerical phase values (*MULTAN*, *SHELX*, *GENTAN*) or symbolic ones (*SIMPEL*) are used. The phase-determination procedures of these programs consist of two distinct parts. Firstly, the convergence procedure (Germain, Main & Woolfson, 1970) is performed to yield a number of starting phases and the order in which the remaining phases may obtain their phase values. After the origin is fixed with a few of the starting phases the others are assigned numerical or symbolic phase values. Then, secondly, most other phases are determined either in the reverse convergence order or by use of acceptance criteria or weighting schemes. It will be indicated in this paper that the convergence procedure and the subsequent phase determination are not necessarily the best choices in carrying out the phasing. A more systematic method of phase determination, suggested by the dynamic-programming principle (Bellman, 1957), is presented. This phase-determination process, referred to throughout this paper as dynamic-programming phasing or DPP, results in optimal phasing given *a priori* probabilistic assumptions.

1. The convergence method

Germain, Main & Woolfson (1970) developed the convergence procedure in direct methods to obtain a starting set of phases from which all others may be obtained in successive steps. The procedure uses a large set of the strongest reflections and all their triplet relations. The phase of the reflection which is probabilistically worst determinable by all other phases is removed from the set and at the same time all its relationships are deleted. This process is repeated until all reflections and relations have been removed. Every now and then a reflection is removed which has no relationship left; the phase of such a reflection is a starting-set phase, *i.e.* a phase the value of which has to be chosen. So, finally, the convergence procedure results in: (1) a starting set of reflections, the phases of which will be used as a starting point in the successive phase extension; and (2) a sequence S of removed reflections, which in reverse order (S^{-1}) may guide the phase extension. The criterion function for removing reflections from the set and transferring them to the sequence S is based on the joint probability distribution of the phase Φ_H taking part in n triplets, given the magnitude $|E_H|$ and all other magnitudes and phases involved.

$$P(\Phi_H|\Phi_{K_1}, \Phi_{H-K_1}, \ldots, \Phi_{K_n}, \Phi_{H-K_n}, |E_H|, |E_{K_1}|, \ldots, |E_{H-K_n}|) = L^{-1} \exp\left[2 \sum_{j=1}^{n} E_{3j} \cos(-\Phi_H + \Phi_{K_j} + \Phi_{H-K_j})\right]. \quad (1)$$

L^{-1} is a normalization constant and

$$E_{3j} = w_j Z_3 |E_H E_{K_j} E_{H-K_j}| \quad \text{for a general triplet}$$
$$E_{3j} = 0{\cdot}5 w_j Z_3 |E_H E_{K_j} E_{H-K_j}| \quad \text{for a restricted triplet.} \quad (2)$$

E_{3j} is a modified form of the usual triplet argument E_3 (Cochran, 1955; Cochran & Woolfson, 1955). Z_3 depends on the atomic numbers,

$$Z_3 = \sum_{j_1=1}^{N} z_{j_1}^3 \Big/ \left[\sum_{j_1=1}^{N} z_{j_1}^2\right]^{3/2}, \quad (3)$$

and the weight w_j depends on the space-group multiplicities of the reciprocal vectors H, K and $H-K$ (Giacovazzo, 1974). If we introduce

$$\alpha_H \exp(i\Xi_H) = \sum_{j=1}^{n} E_{3j} \exp[i(\Phi_{K_j} + \Phi_{H-K_j})], \quad (4)$$

(1) can be written as

$$P(\Phi_H|\ldots) = L^{-1} \exp[2\alpha_H \cos(-\Phi_H + \Xi_H)] \quad (5)$$

with

$$\alpha_H^2 = \sum_{j,j'} E_{3j} E_{3j'} \cos[(\Phi_{K_j} + \Phi_{H-K_j}) - (\Phi_{K_{j'}} + \Phi_{H-K_{j'}})]. \quad (6)$$

The actual criterion used by Germain *et al.* (1970) is the square root of the averaged value of α_H^2, $\langle\alpha_H^2\rangle^{1/2}$, with

$$\langle\alpha_H^2\rangle = \sum_{j=1}^{n} E_{3j}^2 + 2 \sum_{j<j'} E_{3j} E_{3j'} \times \langle\cos[(\Phi_{K_j} + \Phi_{H-K_j}) - (\Phi_{K_{j'}} + \Phi_{H-K_{j'}})]\rangle. \quad (7)$$

This expression cannot be evaluated, because no phases have been determined yet; therefore Germain *et al.* approximate the expectation values on the right-hand side of (7) by the product of the expectation values for each single triplet,

$$\langle\cos(\Phi_{K_j} + \Phi_{H-K_j} - \Phi_H)\rangle = B_q(2E_{3j}), \quad (8)$$

with B_q the quotient of two modified Bessel functions (Watson, 1952),

$$B_q(2E_{3j}) = I_1(2E_{3j})/I_0(2E_{3j}). \quad (9)$$

The convergence part of the *SIMPEL* system is applied to triplets and (optionally) quartets with an even simpler criterion,

$$\langle\alpha_H^2\rangle^{1/2} = \sum_{j=1}^{n} E_{3j}^2 + \sum_{j=1}^{n'} (E_{4j}^*)^2 \quad (10)$$

with E_4^* associated with the quartet relation (Schenk, 1973).

In *MULTAN* the reverse sequence of S, S^{-1}, is used to guide the determination of the phases. Since the calculated statistical weights and phase values propagate throughout the phasing process, it cannot

be expected that S^{-1}, calculated necessarily without this actual phasing information, will be the best possible phasing sequence.

In *SIMPEL* S^{-1} is not used; instead, a set of strict criteria is used for accepting new phase indications.

The order in which the phases are determined in both *MULTAN* and *SIMPEL* does not guarantee that the most reliable set of phases is obtained. A preferable procedure is one which calculates from scratch both the starting set and the order in which the remaining phases should be determined, optimizing at any moment the statistical weights of the phases.

2. Optimal phase determination

2.1. *Requirements for a phase-determination process*

We shall first formulate the requirements for a phase-determination process:

(1) As many phases as possible with a unique phase indication are to be determined. If symbolic phases are used, as is assumed throughout this paper, phases may acquire different symbolic phase indications. These phases cannot be used to determine other phases unless suitable assumptions are made.

(2) The final set of phases should be as reliable and as large as possible.

(3) The number of reflections the phases of which must be chosen, the starting-set reflections, should be as small as possible.

(4) Neither the starting reflections nor the order in which the remaining phases should preferably be determined must be defined *a priori.*

2.2. *The dynamic-programming approach*

The phase-determination procedure to be discussed proceeds along the lines of the dynamic-programming method introduced by Bellman (1957). Dynamic programming is a technique for handling so-called multistage decision processes in order to obtain an optimal sequence of decisions with respect to a specified criterion function.

In the case of the phase determination an optimal sequence of phases to be determined is required through a set of M reflections. The first thing to do is to select a criterion function for this problem. An obvious choice is to base the criterion function on weights to be assigned to the reflections as they are successively added to form a sequence. (For other criterion functions see § 4.) Consider the generation of a sequence of M reflections. The $(j+1)$th reflection that is going to be combined with the sequence of the j preceding reflections may form triplets with reflections in this j-sequence (we shall call a sequence of n reflections an n-sequence). The strengths and the number of the triplets are a measure of the affinity of the $(j+1)$th reflection with the j-sequence, and this can be expressed in a weight to be assigned to the $(j+1)$th reflection in the sequence. (The calculation of the weights will be discussed in § 3). In this way a weight has been assigned to each of the reflections in the sequence. The total weight of the M-sequence is a measure of its quality and can be used as a criterion function for the M-sequences.

The total number of possible sequences of M given reflections is $M!$ It is quite clear that, in general, one cannot possibly handle all possible sequences in order to arrive at the optimal one ($10! = 3{\cdot}63 \times 10^6$; $20! = 2{\cdot}43 \times 10^{18}$; $30! = 2{\cdot}65 \times 10^{32}$). A systematic procedure for getting at the optimal sequence, involving a small fraction of the effort required for the above approach, is provided by the dynamic-programming method.

The procedure will now be explained in some detail. It is executed in a number of stages. In each stage j $(1 \leq j < M)$ optimal j-sequences are selected, combining each reflection with the optimal $(j-1)$-sequences which result from the previous stage. In the zero or initial state no sequences are present yet: the procedure starts from scratch. In the first stage optimal 1-sequences are selected. Obviously, no phases can be determined since the 0-sequence contains no reflections but any reflection i_1, $i_1 = 1, \ldots, M$, can be selected as a starting-set reflection so that M optimal 1-sequences result. In the second stage, optimal 2-sequences are to be found. If only triplets amongst three different reflections are considered, any reflection i_2 can be chosen to be a starting-set reflection, in combination with any optimal 1-sequence reflection i_1, provided $i_2 \neq i_1$. Hence, after the second stage $M \times (M-1)$ optimal 2-sequences result. The first stage to be performed in practice is the third stage, concerned with the selection of optimal 3-sequences. A particular reflection i may form a triplet with the optimal 2-sequences. The triplet with the largest E_3 is the optimal 3-sequence ending at reflection i. This optimal 3-sequence and its associated weight are assigned to reflection i. In this way each reflection i, $i = 1, \ldots, M$, is provided with an optimal 3-sequence ${}^oS_i^3$, ending at reflection i, and an associated weight ${}^ow_i^3$. In the next stage each of the M reflections in the set is combined with each of the optimal 3-sequences, in which the reflection is not represented, to form 4-sequences. For each of the 4-sequences ending at reflection i the weight of reflection i, w_i^4, can be calculated. If this weight is added to the weight of the 3-sequence j the total weight $W_i^4 = {}^ow_j^3 + w_i^4$ for the 4-sequence is obtained. The largest weight of all the 4-sequences ending at reflection i indicates the optimal 4-sequence ending at this reflection. In this way for each of the reflections $1, \ldots, M$ an optimal 4-sequence ${}^oS_i^4$, with an associated weight ${}^oW_i^4$, can be obtained. So at the end of the fourth stage we are left with M optimal 4-sequences, one for each i.

This procedure is repeated until finally the Mth reflections are combined with the optimal $(M-1)$-

sequences, resulting in optimal M-sequences, ${}^{o}S_i^M$, with weights ${}^{o}W_i^M$. Of these sequences, the one with largest ${}^{o}W_i^M$ is the optimal M-sequence looked for.

Remarks

(1) The weights assigned to the individual reflections lie in the interval 0–1 (see § 3).

(2) If a reflection is combined with an optimal sequence with which it does not form triplets no weight can be calculated for this reflection. Consequently, it is assigned a starting-set reflection with a weight 1·0.

(3) It is possible that at a certain stage there are several optimal sequences with the same number of starting-set phases ending at a particular reflection. These are all included in the procedure.

(4) At each stage several optimal sequences ending with the same reflection, but differing in the number of starting-set reflections, can be present (see § 4).

(5) Towards the end of the procedure the number of sequences to be combined with an additional reflection drops, because the probability that the new reflection is already present in a sequence increases.

3. Calculation of weights for symbolic phase indications

In the numerical direct methods like *MULTAN* the calculation of a phase is based on the tangent formula (Karle & Hauptman, 1956; Karle & Karle, 1966), which can be obtained by equating to zero the derivative of (1) with respect to Φ_H. In practice, weighted tangent expressions are employed, including weights w_{K_j} and w_{H-K_j} for the known phases,

$$\tan(\Phi_H) = \frac{\sum_{j=1}^{n} E_{3j} w_{K_j} w_{H-K_j} \sin(\Phi_{K_j} + \Phi_{H-K_j})}{\sum_{j=1}^{n} E_{3j} w_{K_j} w_{H-K_j} \cos(\Phi_{K_j} + \Phi_{H-K_j})}. \tag{11}$$

The weight w_H for the phase indication (11) is not unique and may be defined in various ways (Germain, Main & Woolfson, 1971; Hull & Irwin, 1978; Giacovazzo, 1980).

In the *SIMPEL* program only phases for which a unique (symbolic) phase indication is obtained are accepted provided the (sum of) E_3's of the relations involved meet a strict reliability threshold value. No other weighting is employed for the accepted phases, which in turn may determine new ones.

We shall now explain how a symbolic phase indication and a weight for the reliability of the phase indication can be calculated simultaneously. The starting point is (5) from which the expectation value $\langle \exp(i\Phi_H) \rangle$ can be obtained in a standard way, either by integration or by summation, depending on whether H is a general or a restricted reflection respectively. These two cases will now be considered.

For a general reflection H, $\langle \exp(\Phi_H) \rangle$ can be expressed as

$$\begin{aligned}\langle \exp(i\Phi_H) \rangle &= [I_1(2\alpha_H)/I_0(2\alpha_H)] \exp(i\Xi_H) \\ &= [I_1(2\alpha_H)/I_0(2\alpha_H)] \alpha_H^{-1} \\ &\quad \times \sum_{j=1}^{n} E_{3j} \exp[i(\Phi_{K_j} + \Phi_{H-K_j})]. \end{aligned} \tag{12}$$

Expression (12) enables one to estimate $\langle \exp(i\Phi_H) \rangle$ when a number of phases Φ_K and Φ_{H-K} are known.

In the case of a single triplet indication and with Φ_K and Φ_{H-K} as starting-set phases the exponentials in the right-hand side of (12) are known exactly and the estimate can be expressed as

$$\begin{aligned}\langle \exp(i\Phi_H) \rangle &= B_q(2E_3) \exp[i(\Phi_K + \Phi_{H-K})] \\ &= g_H \exp[i(A+x)]. \end{aligned} \tag{13}$$

$A+x$ is the symbolic phase indication for Φ_H with A the symbolic and x the numerical part. g_H, which can be considered as a reliability indicator for $A+x$, lies between 0 and 1. In all other cases, the exponentials on the right-hand side of (12), assumed to be known, are not known exactly. The only available quantities are the $\langle \exp(i\Phi_H) \rangle$. Hence, after substitution of the expectation values for the unknown exponentials in (12), representation of each sum of phases $\Phi_K + \Phi_{H-K}$ in a symbolic part A_j and a numerical part x_{t_j} and multiplication of the weights g_K and g_{H-K} into g_{t_j}, the resulting expression can be noted as a sum over m different symbolic parts, for each of which m_j different numerical parts are present:

$$\langle \exp(i\Phi_H) \rangle = \frac{I_1(2\beta_H)}{I_0(2\beta_H)} \beta_H^{-1} \sum_{j=1}^{m} \exp(iA_j) \times \left[\sum_{t_j=1}^{m_j} E_{3t_j} g_{t_j} \exp(ix_{t_j}) \right]$$

with

$$\begin{aligned}\beta_H^2 = \sum_{j=1}^{m} \left(\sum_{t_j=1}^{m_j} E_{3t_j}^2 \right) &+ 2 \sum_{\substack{j,j'=1 \\ j \neq j' \text{ and } t_j < t_{j'} \\ \text{for } j = j'}}^{m} \sum_{t_j, t_{j'}=1}^{m_j, m_{j'}} E_{3t_j} E_{3t_{j'}} g_{t_j} g_{t_{j'}} \\ &\times \cos[(A_j + x_{t_j}) - (A_{j'} + x_{t_{j'}})] \Bigg\}. \end{aligned} \tag{14}$$

Expression (14) can be further simplified by application of standard goniometric expressions and with

the definition for each different A_j of the functions (15)-(22),

$$E_{3k}(j) = \sum_{t_j=1}^{m_j} E_{3t_j}^2 \tag{15}$$

$$\mathrm{wgc}(j) = \sum_{t_j=1}^{m_j} E_{3t_j} g_{t_j} \cos(x_{t_j}) \tag{16}$$

$$\mathrm{wgs}(j) = \sum_{t_j=1}^{m_j} E_{3t_j} g_{t_j} \sin(x_{t_j}) \tag{17}$$

$$\mathrm{scw}(j) = \sum_{t_j=1}^{m_j} [E_{3t_j} g_{t_j} \cos(x_{t_j})]^2 \tag{18}$$

$$\mathrm{ssw}(j) = \sum_{t_j=1}^{m_j} [E_{3t_j} g_{t_j} \sin(x_{t_j})]^2 \tag{19}$$

$$r_j^2 = [\mathrm{wgc}](j)]^2 + [\mathrm{wgs}(j)]^2 \tag{20}$$

$$\tan(y_j) = \mathrm{wgs}(j)/\mathrm{wgc}(j) \tag{21}$$

and

$$D_j^2 = E_{3k}(j) + r_j^2 - \mathrm{scw}(j) - \mathrm{ssw}(j). \tag{22}$$

For example, if only one symbolic phase part A_1 is present, (14) reduces to

$$\langle \exp(i\Phi_H) \rangle = \frac{I_1(2D_1)}{I_0(2D_1)} \frac{r_1}{D_1} \exp[i(A_1 + y_1)]$$
$$= g_H \exp[i(A_1 + y_1)]. \tag{23}$$

Expression (23) is similar to (13), the weight g_H, $0 \le g_H < 1$, serving as the indicator for the reliability of the symbolic phase indication $A_1 + y_1$.

If, owing to symmetry, the phase Φ_H is restricted in its value to v_1 and $v_1 + \pi$ with $0 \le v_1 < \pi$, the calculation of $\langle \exp(i\Phi_H) \rangle$ results in

$$\langle \exp(i\Phi_H) \rangle = \exp(iv_1) \tanh \left[\sum_{j=1}^{m} \sum_{t_j=1}^{m_j} E_{3t_j} g_{t_j} \times \cos(-v_1 + A_j + x_{t_j}) \right]. \tag{24}$$

If only one symbolic phase part A_1, restricted to the values Δ_1 and $\Delta_1 + \pi$, is present, A_1 can be written as the sum of a new symbol A_1', restricted to $0/\pi$, and the numerical part Δ_1.

If we define $x'_{t_1} = x_{t_1} + \Delta_1 - v_1$, (24) can be rearranged to give

$$\langle \exp(i\Phi_H) \rangle = \exp(iv_1) \tanh \left[\cos(A_1') \sum_{t_1=1}^{m_1} E_{3t_1} g_{t_1} \cos(x'_{t_1}) \right]. \tag{25}$$

From definition (16), the sum in (25) is $\mathrm{wgc}'(1) = s|\mathrm{wgc}(1)|$ with $s = \pm 1$. For $s = \pm 1$, $\tanh(sx) = s \tanh(x)$. Therefore

$$\langle \exp(i\Phi_H) \rangle = \exp\{i[v_1 + A_1' + \tfrac{1}{2}(1-s)\pi]\} \tanh(|\mathrm{wgc}'(1)|)$$
$$= \exp\{i[A_1 + v_1 - \Delta_1 + \tfrac{1}{2}(1-s)\pi]\} \tanh(|\mathrm{wgc}'(1)|). \tag{26}$$

Once again, $\langle \exp(i\Phi_H) \rangle$ has been expressed in the form (13), as an exponential phase part containing the symbolic phase indication times a weight factor, in this case the hyperbolic tangent term, which lies between 0 and 1. It will be clear that this similarity of expressions greatly facilitates the mixing of general and restricted reflections in the actual phase determination.

So far only the case of one unique symbolic part A_1 for a phase indication has been considered with different numerical phase parts x. When two or more different symbolic phase parts A_j are present, $\langle \exp(i\Phi_H) \rangle$ cannot be expressed as a unique phase as in (23) or (26). In the absence of any information about the most probable value for the relation between the A_j's the only reasonable choice is a uniform distribution. In the Appendix* it is shown how these calculations are performed when two different symbolic phase parts are involved. The result, which can be expressed in the form of (13), again consists of one phase indication, the most probable one, and a weight, which is lowered appreciably by the fact that another phase indication is present. It is expected that including these phases in the DPP procedure will, in general, not lead to unreasonable phase propagations because the corresponding weight is treated like all other weights; thus, only if a weight leads to maximization of the criterion function will the most probable phase indication be accepted.

Complete disregard of these phases, as in *SIMPEL*, has the disadvantage that fewer phases will be reached.

4. The criterion function

In view of the goal of DPP to phase as many reflections as possible with as few starting reflections as possible the number of starting reflections is allowed to vary within certain limits. This implies that for a particular stage the criterion function is a function of the number of starting-set phases present. This will be indicated by using for the criterion function the extended form $W_i^j(\mathrm{Nst})$, with Nst the number of starting phases present amongst the j reflections in the sequences ending with reflection i. The choice of

* The Appendix has been deposited with the British Library Document Supply Centre as Supplementary Publication No. SUP 44018 (4 pp). Copies may be obtained through The Executive Secretary, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.

the criterion function W is of the utmost importance, since the DPP aims at maximizing W, and as a result all phasing sequences are optimal only with respect to W. Several functional forms are possible which incorporate the above ideas, and three of them have been tested:

(i) Criterion function $W1$. This criterion function consists of the sum of weights of the non-starting phases. In the determination of each phase i in stage j of the DPP procedure,

$$W1_i^j(\text{Nst}) = \sum_H g_H \quad \text{maximal.} \tag{27}$$

(ii) Criterion function $W2$. This function is based on the assumption that the integral of the squared difference between the true E map ρ, calculated with measured $|E|$'s and the true Φ_H's, and the estimated density map ρ_e, calculated with the measured $|E|$'s and the estimated phases $\langle \exp(i\Phi_{\mathbf{H}})\rangle$, should be as small as possible [see also Hasek (1984)],

$$W2_i^j(\text{Nst}) = -\int |\rho(\mathbf{r}) - \rho_e(\mathbf{r})|^2 \, d\mathbf{r} \quad \text{maximal.} \tag{28}$$

By insertion of

$$\rho(\mathbf{r}) = \sum_H |E_H| \exp(i\Phi_H) \exp(2\pi i \mathbf{H} \cdot \mathbf{r}) \tag{29}$$

and

$$\begin{aligned}\rho_e(\mathbf{r}) &= \sum_H |E_H| \langle \exp(i\Phi_H)\rangle \exp(2\pi i \mathbf{H} \cdot \mathbf{r}) \\ &= \sum_H |E_H| g_H \exp(i\Phi_H^e) \exp(2\pi i \mathbf{H} \cdot \mathbf{r})\end{aligned} \tag{30}$$

and by use of $|x| = (x^* x)^{1/2}$, (28) is converted into

$$\begin{aligned}&\int \sum_{H,H'} |E_H||E_{H'}| \\ &\times (\{\exp[i(\Phi_H - \Phi_{H'})] + g_H g_{H'} \exp[i(\Phi_H^e - \Phi_{H'}^e)]\} \\ &- g_{H'}\{\exp[i(\Phi_H - \Phi_{H'}^e)] + \exp[-i(\Phi_H - \Phi_{H'}^e)]\}) \\ &\times \exp[2\pi i(\mathbf{H} - \mathbf{H}') \cdot \mathbf{r}] \, d\mathbf{r}.\end{aligned} \tag{31}$$

The only non-zero contributions are for $H = H'$ so that

$$\begin{aligned}W2_i^j(\text{Nst}) &= -\sum_H |E_H|^2 \{g_H^2 + 1 - 2g_H \cos(\Phi_H - \Phi_H^e)\} \\ &= \text{maximal.}\end{aligned} \tag{32}$$

The aim of the procedure is to find $\Phi_H^e = \Phi_H$, when (32) becomes

$$W2_i^j(\text{Nst}) = -\sum_H |E_H|^2 (g_H - 1)^2 = \text{maximal.} \tag{33}$$

The summation in (33) is for the non-starting phases only.

(iii) Criterion function $W3$. Another interesting criterion would be the maximization of

$$-\int |\rho(\mathbf{r}) - \rho_e(\mathbf{r})| \, d\mathbf{r}. \tag{34}$$

However, the integrations involved are too complicated and therefore, in order to have a criterion approximating (34), the following criterion has been used:

$$W3_i^j(\text{Nst}) = -\sum_H |E_H||g_H - 1| = \text{maximal.} \tag{35}$$

Again the summation runs over the non-starting phases only.

For all three criterion functions the optimization leads to at least one optimal sequence; in general, however, it leads to more since the functional dependence of the criteria on the number of starting phases complicates the procedure. Suppose that phase number i can be determined at stage number j by means of two sequences. Let the number of determined phases, including phase number i, be Nd(1) and Nd(2) respectively and the number of starting phases Nst(1) and Nst(2), with $j = \text{Nd}(1) + \text{Nst}(1) = \text{Nd}(2) + \text{Nst}(2)$, and the criteria values $R1$ and $R2$ respectively. Now different situations may occur:

(i)

$$\text{Nst}(1) = \text{Nst}(2) \quad \text{and} \quad \text{Nd}(1) = \text{Nd}(2).$$

This case is simple: the sequence with the largest criterion value is selected.

(ii)

$$\text{Nst}(1) > \text{Nst}(2) \; [\Rightarrow \text{Nd}(1) < \text{Nd}(2)].$$

(a)

$$R1/\text{Nd}(1) \leq R2/\text{Nd}(2).$$

Sequence 1 is not optimal because, in spite of the presence of more starting phases than in sequence 2, the average weight of the determined phases is lower.

(b)

$$R1/\text{Nd}(1) > R2/\text{Nd}(2).$$

In this case, the average weight of the determined phases is larger for sequence 1 but also more starting phases are present. Acceptance now of only one of the two sequences may result in non-optimal results because the aim of the DPP is twofold: (a) to determine as many phases as reliably as possible; and (b) to use as few starting phases as possible. If only sequence 1 is accepted, the rejection of the optimal sequence 2 with fewer starting phases violates the second part of our goal. If sequence 1, with a larger average weight, is rejected the first part of our goal is not fulfilled. Thus no choice can be made, and therefore both sequences are accepted. In the present implementation of DPP, sequences with up to six different number of starting phases can be assigned to each determined reflection i in stage j.

5. Test results and discussion

In order to judge the performance of the DPP procedure, referred to as procedure I, it will be compared with the current symbolic phase determination procedures. For this purpose, four alternative procedures have been selected. Procedure II consists of a default run of *SIMPEL* [application of convergence criterion (10) followed by phase extension] and procedure IV employs the convergence criterion (7) followed by a default symbolic addition run by means of *SIMPEL*. Since the *SIMPEL* symbolic addition routine employs a unit weighting scheme, only a comparison on the basis of phase differences is possible. In order to enable a comparison of weighted phase differences, the DPP program optionally allows the introduction of an *a priori* defined starting set, followed by the DPP procedure itself, in which the introduction of additional starting phases can be suppressed. This amounts to an optimal symbolic phase determination with respect to both the criterion function and the predefined fixed starting set. The resulting values of the criterion function now compare directly with those of procedure I and also the weighted phase errors can be compared directly. This procedure is used in combination with either the *SIMPEL* starting set resulting from (10), referred to as procedure III, or that from (7), referred to as procedure V.

A reasonable test for the quality of the calculated phase sets is the overall weighted absolute difference between the symbolic and the true phases:

$$\mathrm{AV} = \sum_H g_H |\Phi(\mathrm{symb}) - \Phi(\mathrm{true})| \Big/ \sum_H g_H, \quad (36)$$

with the summation running over the non-starting-set reflections only. For the DPP results it is to be expected that there will be a correlation between the differences and the weights. In general, extreme values of the criteria (27), (33) and (35) should indicate more reliable phase sequences, corresponding with lower AV values. To calculate (36) the symbols are replaced by their actual numerical values from the known structure. The weights g_H are either the calculated statistical weights (procedures I, III and V) or unit weights (procedures II and IV).

5.1. *Comparison of the criterion functions W1, W2 and W3*

In § 4 three criteria were introduced, *W*1, *W*2 and *W*3, given by the expressions (27), (33) and (35) respectively. *W*1 emphasizes the importance of a large average statistical weight, the other two include the individual $|E|$'s as well. The quality of the results of the DPP procedure using *W*1, *W*2 and *W*3 can be judged on the basis of the differences (36).

From all tests performed so far it appears that the three criteria lead to completely comparable error levels within 10 millicycles (mc) ($\simeq 3°$). Therefore, in this paper only the data for *W*1 will be listed. The equal error level might suggest that the actual form of the criterion function is not important, but some related criteria have been tested and these results show that this is not the case.

5.2. *Results for structure Kanter, P*1

As a first example, the results obtained for the 30-atom *P*1 structure Kanter (Kanters & van Veen, 1973) will be discussed. From the 40 phases with the strongest $|E|$'s, 32 were employed, interlinked *via* 23 triplet relations; the other eight phases were not connected with the rest of the set *via* triplets and were therefore omitted. In Table 1 the results obtained with procedures II, III, IV and V are listed. Both convergence criteria (7) and (10) (in the latter the quartets were excluded) came up with a starting set consisting of 11 phases, three origin-fixing ones and eight phases to which symbolic phase values were assigned, differing in only one symbolic phase. Procedure II (default *SIMPEL*) resulted in 15 uniquely determined phases but from the start onwards the average error for this procedure is rather high, up to 94 mc for the 15 determined phases (see Table 1). In the next columns of Table 1, the results for procedure III are listed, employing the same *a priori* starting set as procedure II. For each stage the largest three *W*1 values are listed only along with the corresponding average phase errors. These results show, in contrast to those of procedure II, an overall gradual increase of the overall error level as the number of determined phases (Nd) increases whilst the average weight of the determined phases (*W*1/Nd) decreases. Furthermore, starting from Nd = 10, the weighted average errors are systematically lower than the unweighted, illustrating that phases determined with a statistically lower weight will generally contain a larger phase error. The results for procedures IV and V show that the tendencies observed for procedures II and III are also present there. A comparison of procedures III and V indicates that starting at Nd = 10 the largest three *W*1 values are systematically larger for procedure V and the average errors lower.

In Tables 2, 3 and 4 test results of procedure I have been ordered in different ways. First of all, in Table 2, the optimal sequences have been listed according to the stage number (Ncy) which equals the total number of phases present in the sequence. In Table 3, the influence of the increase of the number of starting phases is shown while the number of determined phases is kept constant. Finally, Table 4 lists for various Nst how many phases can be determined in an optimal way.

Table 2 shows that in most stages optimal sequences are built up based on different numbers of starting phases. For example, the 20th stage results in optimal

Table 1. *Results of the phasing procedures* II, III, IV *and* V *applied to the structure Kanter* (*P*1; 30 *equal atoms*)

For procedure I see Tables 2, 3 and 4. The strongest 32 reflections have been employed, connected *via* 23 triplets. Nst = 11 for procedures II, III, IV and V. Procedures II and III on the one hand and IV and V on the other differ only in one starting phase.

	II		III					
			Sequence with highest *W*1		second *W*1		third *W*1	
Nd	Av	Nd	*W*1	Av	*W*1	Av	*W*1	Av
5	101	5	4·54	38	4·54	45	4·54	49
7	78	7	6·34	42	6·34	38	6·33	35
10	66	10	8·95	50 (51)	8·94	53 (54)	8·94	41
13	76	13	11·42	75 (77)	11·40	78 (80)	11·39	75 (77)
15	94	15	13·00	76 (77)	12·96	91 (95)	12·91	73 (74)
	IV		V					
			Sequence with highest *W*1		second *W*1		third *W*1	
Nd	Av	Nd	*W*1	Av	*W*1	Av	*W*1	Av
5	61	5	4·55	50	4·54	44	4·54	48
7	50	7	6·35	45	6·35	42	6·34	44
10	49	10	9·03	49	9·03	41	9·03	52
13	61	13	11·62	54	11·60	56 (57)	11·53	59 (61)
15	68	15	13·21	66 (67)	13·13	62 (64)	13·12	54
16	74	16	13·93	71 (73)	13·91	63 (64)	13·88	69 (72)

The following notation has been employed in all the tables. Procedures employed: (I) DPP procedure; (II) standard *SIMPEL* procedure: convergence criterion (10) and unit-weighting scheme in symbolic addition; (III) convergence criterion (10), followed by an optimal phase determination only (*a priori* starting set); (IV) convergence criterion (7), followed by a standard *SIMPEL* symbolic addition (unit-weighting scheme for determined phases); (V) convergence criterion (7), followed by an optimal phase determination (*a priori* starting set). In cases where the inclusion/rejection of the most-probable phase indications yield different results, each procedure is marked further with an *a*, indicating rejection, or *b*, denoting the inclusion of such indications.

Nd = number of determined phases, *i.e.* with a calculated weight; Nst = number of starting phases; Ncy = number of DPP stages, which also equals the number of reflections in the phasing sequence, Nst + Nd; Av = average phase error of the determined phases [see (36)]. The unweighted average error is listed between brackets if it is different from the weighted error. The average error(s) are given in millicycles (mc); *W*1 = criterion function value. Criterion function *W*1, defined in (27), has been employed throughout. With the exception of procedures II and IV, the largest three criterion values are listed for the combinations of Nd and Nst under discussion.

Table 2. *Results of dynamic-programming procedure* I *applied to the structure Kanter*

The same reflections and triplets have been used as for the procedures II–V, the results of which are listed in Table 1. For each combination of Nd and Nst the largest three criterion function values *W*1 are listed with the corresponding average phase errors of the Nd phases.

			Sequence with highest *W*1		second *W*1		third *W*1	
Ncy	Nd	Nst	*W*1	Av	*W*1	Av	*W*1	Av
10	5	5	4·54	27	4·52	30	4·52	36
15	7	8	6·36	39	6·35	45	6·34	43
	8	7	7·24	44	7·23	41	7·22	39
	9	6	7·80	57 (58)	7·75	62	7·67	59
20	10	10	9·04	50	9·04	42	9·04	42
	11	9	9·92	53	9·90	52	9·90	52
	12	8	10·72	51	10·67	57 (58)	10·65	47
	13	7	11·13	48 (47)	11·12	48 (47)	10·87	54
22	12	10	10·81	52	10·80	51	10·79	50
	13	9	11·62	61 (62)	11·57	51	11·57	56 (57)
	14	8	11·99	67 (68)	11·96	56	11·94	46
25	15	10	13·31	62	13·31	55		
	16	9	13·68	68 (70)	13·68	63 (65)	13·64	50
26	16	10	14·11	58 (59)	14·00	65 (67)	13·98	66 (67)
27	17	10	14·78	62 (63)	14·66	58 (60)	14·59	65 (66)
28	18	10	15·37	62 (63)	15·26	56 (57)		

For abbreviations used, see Table 1.

Table 3. *Results of dynamic-programming procedure* I *for the structure Kanter*

Results are ranked according to number of determined phases (Nd).

			Sequence with highest *W*1		second *W*1		third *W*1	
Nd	Nst	Ncy	*W*1	Av	*W*1	Av	*W*1	Av
5	4	9	4·27	60 (61)	4·11	51 (59)		
	5	10	4·54	27	4·52	30	4·52	36
	6	11	4·54	46	4·54	48	4·54	55
7	5	12	6·02	69	6·02	59	5·92	55
	6	13	6·33	36	6·32	36	6·31	31
	7	14	6·36	43	6·34	45	6·34	38
	8	15	6·36	39	6·35	45	6·34	43
10	6	16	8·33	58 (57)				
	7	17	8·97	42	8·96	46	8·93	46
	8	18	9·02	49	9·00	47	8·99	45
	9	19	9·04	47	9·03	53	9·03	46
	10	20	9·04	42	9·04	50	9·04	50
13	7	20	11·13	48	11·12	48	10·87	54
	8	21	11·40	67 (69)	11·36	58 (59)	11·35	58 (59)
	9	22	11·62	60 (61)	11·57	51	11·57	56 (57)
	10	23	11·68	49	11·66	49	11·66	55
15	9	24	13·09	64 (65)	13·09	69 (70)	13·03	52 (53)
	10	25	13·32	61 (62)	13·31	62	13·31	55
16	9	25	13·68	63 (65)	13·68	68 (70)	13·64	50
	10	26	14·10	58 (59)	14·00	65 (67)	13·98	66 (67)
17	10	27	14·78	62 (63)	14·66	58 (60)	14·59	65 (66)
18	10	28	15·37	62 (63)	15·20	56 (57)		

For abbreviations used, see Table 1.

Table 4. *Results of dynamic-programming procedure* I *for the structure Kanter*

Results are listed for nine starting phases (Nst = 9). For Nst = 5–10 the number of phases (Nd range) is listed which can be determined optimally.

			Sequence with highest *W*1		second *W*1		third *W*1	
Nst	Nd	Ncy	*W*1	Av	*W*1	Av	*W*1	Av
9	8	17	7·25	42	7·25	46	7·24	45
	9	18	8·15	45	8·14	51	8·14	45
	10	19	9·04	47	9·03	53	9·03	53
	11	20	9·92	48	9·92	53	9·90	52
	12	21	10·77	52	10·77	57	10·76	56
	13	22	11·62	61 (62)	11·57	51	11·57	56 (57)
	14	23	12·42	60	12·42	65 (66)	12·35	51
	15	24	13·09	64 (65)	13·09	69 (71)	13·03	52 (53)
	16	25	13·68	63 (65)	13·68	68 (70)	13·64	50

Nst	Nd range
5	4–7
6	5–10
7	6–13
8	7–14
9	8–16
10	9–18

For abbreviations used, see Table 1.

sequences containing seven, eight, nine and ten starting phases, respectively, and consequently 13, 12, 11 and 10 determined phases. For a fixed Ncy, the average weight of the determined phases decreases with Nst, and since a lower statistical weight implies a larger phase error, the largest average errors are expected to occur for the lowest Nst (see Table 2) in spite of a large spread of individual values. It can also be observed that, for increasing Ncy, the minimum number of starting phases necessary to obtain optimal sequences increases as well. At the end of the procedure only the (maximum) number of ten starting phases results in optimal sequences and, after stage 28, the phase-determining process stops. In order to include all 32 phases in the final phase set more than ten starting phases are necessary.

In Table 3 it is shown that a particular number of determined phases can be obtained in an optimal way with different numbers of starting phases. For constant Nd value, an increase in Nst results in a

Table 5. *Symbolic phase determination results for Diemal* (22 *atoms*, $P2_1$) *employing an a priori starting set of six phases, found by the convergence criteria* (7) *and* (10)

Procedure II therefore equals IV and III equals V. For procedure III (and V) two alternatives are distinguished: (i) procedure III*a*, where the most-probable phase indications are not calculated; and (ii) procedure III*b*, where the most-probable phase indications are included in the procedure. In total 34 phases are involved.

II		III*a*						
			Sequence with highest *W*1		second *W*1		third *W*1	
Nd	Av	Nd	*W*1	Av	*W*1	Av	*W*1	Av
5	41	5	4·80	32	4·80	36	4·79	31
8	46	8	7·67	37	7·66	37	7·66	37
10	48	10	9·60	30	9·59	30	9·59	26
12	41	12	11·50	28	11·45	28	11·45	28
13	37	13	12·41	27	12·40	28	12·38	33
		14	13·31	27 (26)	13·28	31	13·23	25
		16	15·00	29	14·95	31 (32)	14·83	36 (37)
		18	16·48	33 (34)	16·45	39 (41)	16·18	41 (46)
		19	17·27	37 (39)	17·27	37 (39)	17·13	45 (49)
		20	17·95	43 (46)				
		III*b*						
		Up to Nd = 17 the same results hold as for III*a*						
		18	16·48	33 (34)	16·45	39 (41)	16·32	41 (49)
		19	17·27	37 (39)	17·27	37 (39)	17·13	45 (49)
		20	17·95	43 (46)	17·82	48 (55)	17·82	48 (55)
		21	18·50	53 (61)				

For abbreviations used, see Table 1.

considerable increase in the maximum *W*1 value and consequently in a decrease in the average error as well. Indeed, for Nd = 13 only, the increase from Nat = 7 to Nst = 8 does not lead to a reduction of the average error. For Nd = 13, 15, 16, 17 and 18, the increase in the maximum *W*1 values does not end with Nst = 10, once more indicating that more starting phases will improve the final phase sets. Table 4 shows that each number of starting phases leads to the optimal determination of a limited number of phases only. Outside this Nd range, which is different for each Nst, determination is either not possible or not optimal any longer.

By comparing the five procedures it can be seen that the DPP procedure I is by far the best one and much better than the standard *SIMPEL* method. In general, the combination of convergence and dynamic extension is also less reliable. Although both procedures III and V employ an *a priori* starting set consisting of 11 starting phases the maximum *W*1 values do not differ much for procedures I and V. For the best comparison for each Nd value the largest *W*1 values at the largest Nst value should be selected. For Nd = 10 the *W*1 values of procedure I tend to be larger in spite of the smaller largest Nst value whilst those for III are worse. At Nd = 15 it can be seen that a further increase of *W*1 is probable when more than ten starting phases are chosen. It may be remarked that the larger *W*1 values for procedure I correspond with lower average phase errors as well, as can be inferred from Tables 1 and 3. Inclusion of the most probable phase indications, as described in the last part of § 3, resulted in identical results, as listed in Tables 2, 3 and 4.

5.3. *Results for structure Diemal,* $P2_1$

The second example for which the DPP procedure results will be discussed in a small $P2_1$ structure with code name Diemal [$N = 22$; Van der Putten (unpublished)]. This example has been selected because the structure is easily solvable with procedure II (*SIMPEL*), involving only a small number of symbols and, furthermore, because the effect of including most-probable phase indications can be illustrated. Therefore, in addition to the numbers I to V, the phasing procedure will be characterized further by an *a* or *b* denoting non-inclusion and inclusion of those phases respectively, except of course for procedures II and IV where non-acceptance is the only option.

In all procedures the 34 phases with the strongest $|E|$'s, interconnected *via* 40 triplets, have been used. Because II and IV yield identical results and similarly III and V, only the data for II and III are given in Table 5. A standard run of procedure II resulted in six starting reflections and the determination of 13 phases which are on average 36 mc in error. If one compares the results of procedure III with those of II, one sees that the systematically larger phase errors for procedure II illustrate its non-optimality. The differences between procedures III*a* and III*b* turned out to be small. Up to Nd = 17 the largest three *W*1 values are identical. Starting at Nd = 18, more optimal sequences are found with procedure III*b* but the extra

Table 6. *Optimal symbolic phase determination results for Diemal*

The same reflections and triplets as for the data in Table 5 have been used. Procedure I has been run in two modes: (i) procedure I*a*, no calculation of most-probable phase indications; (ii) procedure I*b*, inclusion of most-probable phase indications. The data listed are the same for procedures I*a* and I*b*, unless indicated otherwise (* denotes I*a* only). The corresponding data for I*b* can be found at the end of this table.

			I*a* and I*b*					
			Sequence with highest *W*1		second *W*1		third *W*1	
Nd	Nst	Ncy	*W*1	Av	*W*1	Av	*W*1	Av
5	2	7	4·76	21	4·76	13	4·76	21
	3	8	4·79	32	4·79	35	4·78	31
	4	9	4·79	34	4·79	31	4·79	28
	5	10	4·78	30	4·78	25	4·75	31
	6	11	4·75	37	4·73	41	4·73	41
10	3	13	9·59	26	9·59	26	9·44	20
	4	14	9·59	28	9·59	28	9·59	28
	5	15	9·56	27	9·56	32	9·54	28
	6*	16	9·54	37	9·53	32	9·54	28
	7*	17	9·51	39	9·51	36		
14	4	18	13·21	27	13·21	30 (31)	13·20	27
	5	19	13·26	26	13·26	26	13·25	26
	6	20	13·33	25	13·32	28	13·32	28
	7	21	13·35	30	13·35	30	13·32	28
	8	22	13·34	32	13·33	32	13·31	30
18	5*	23	16·30	36 (40)				
	6	24	16·84	30 (31)	16·84	27 (28)	16·81	30
	7	25	16·90	37 (38)	16·90	31	16·89	33 (32)
	8	26	16·96	34	16·96	34	16·96	32 (31)
	9	27	17·01	36	17·00	26	16·99	37
	10	28	17·00	26	17·00	36	16·98	27
20	6*	26	18·01	37 (41)	18·00	37 (40)		
	7	27	18·54	31 (32)	18·52	31 (32)	18·52	31 (32)
	8	28	18·65	31	18·64	33 (34)	18·64	28
	9	29	18·74	34	18·73	33	18·72	38 (39)
	10	30	18·78	38	18·75	39 (40)	18·75	34
23	7*	30	19·99	48 (56)				
	8*	31	20·56	55 (58)	20·32	48 (54)	20·32	48 (54)
	9*	32	21·16	38 (39)	21·16	38 (39)	21·10	36 (37)
	10*	33	21·30	41 (42)	21·30	41 (42)	21·18	40 (41)
24	7*	—						
	8*	32	20·83	52 (60)	20·83	52 (60)		
	9*	33	21·39	59 (62)	21·39	59 (62)		
			I*b* only					
10	6	16	9·54	37	9·53	37	9·52	27
	7	17	9·51	39	9·51	36	9·46	37
	8	18	9·50	36				
18	5	23	16·30	36 (40)	16·23	42 (50)		
20	6	26	18·34	29	18·06	28 (29)	18·01	37 (41)
23	7	30	20·03	53 (63)	20·01	53 (63)	19·99	48 (56)
	8	31	21·04	35 (36)	21·00	37 (39)	20·99	34
	9	32	21·16	38 (39)	21·16	39 (39)	21·13	36
	10	33	21·30	41 (42)	21·30	41 (42)	21·22	35
24	7	31	20·54	57 (69)	20·13	52 (60)	19·81	59 (69)
	8	32	21·74	34 (35)	21·71	36 (37)	21·56	34 (35)
	9	33	21·92	38 (39)	21·88	40 (42)	21·87	37 (38)
	10	34	21·96	40 (41)	21·92	34	21·65	38 (40)
25	7	32	20·64	56 (66)	20·63	56 (66)		
	8	33	21·39	60 (72)	21·37	60 (72)	20 97	56 (64)
	9	34	22·62	38	22·58	39 (41)	22·44	38
26	8	34	21·48	60 (70)				

For abbreviations used, see Table 1.

*W*1 values are still lower than those already present in III*a*. In the end, in III*b* one more phase has been determined.

Table 6 contains a selection of the results obtained with procedures I*a* and I*b*. From a comparison of Tables 5 and 6 it appears that up to Nd = 14 the largest *W*1 values are somewhat larger for procedure III. However, for larger Nd those from procedure I are systematically better and the average errors correspondingly lower. This slightly worse performance of procedure I at the beginning of the phase determination can be explained by inspection of the optimal results for each Nd value. Clearly, relatively low-weighted sequences in the early stages of the pro-

cedure survive and grow in the later stages to the final relatively high-weighted sequences with low average errors.

Table 6 also shows that, if Nd increases, the Nst for which the largest $W1$ values are obtained increases as well. At most 20 phases can be determined optimally with six starting phases but 20 phases, not necessarily the same, may also be determined optimally with from seven to ten starting phases with a clearly increasing average weight of the determined phases. In particular, the large increase of the $W1$ value when seven instead of six starting phases are involved suggests the use of seven starting phases. In procedure I the use of the most-probable phase indications shows the same tendencies observed for procedure III but in a more pronounced form. In the initial stages there is no difference between procedures I*a* and I*b*. This was to be expected since, in general, weights corresponding to most-probable phase indications are lower than those for unique phase indications, so for the main part of the phase-determination process the former will result in non-optimal, and thus rejected, sequences. Only at the end of the determination process, when the weights of the new uniquely determined phases are decreasing or when no phases can be determined uniquely, do the most-probable phase indications tend to be optimal. For example, for Nd = 24 the maximum $W1$ values obtained are clearly better for I*b* and the average phase errors are correspondingly lower. It can be concluded that at the moment when procedure I*a* starts to produce less desirable results, *i.e.* a small increase in $W1$ values, the results of I*b* tend to be better. Moreover, with procedure I*b* it is possible to include all 34 phases in the final phase set whilst I*a* yields at most 33 phases. The results for a number of other structures tested to far are similar and the conclusion may be drawn that the inclusion of most-probable phase indications has a positive effect on the final phase-determination results. However, in view of the lower associated weights it cannot be expected that many more phases will be determined. It should also be taken into account that inclusion of the extra phases takes about two to three times as much computer time.

5.4. *Further test results*

The same procedures have been used for a number of other structures. The results of three of these - Waltri [$N = 50$; Olthof (unpublished)], Angust ($N = 40$; Rychlewska, Bratek & Wiewiorowski (1978)] and Indian [$N = 56$; Agarwal, Rastogi, van Koningsveld, Goubitz & Olthof (1980)] - will be discussed here.

The new procedure I again leads to much better phasing. Comparison of the final results of procedures II and IV with the results of I with the same Nst and Nd values shows a considerable reduction of the average phase error, in particular for Angust (from 90 to 60 mc) and Waltri (from 90 to 50 mc). The data for Waltri show that 20–25 phases are best determinable with a maximum number of ten starting phases. Up to Nd = 24 the $W1$ value rises steadily for procedure I*a*, then it starts to drop. For procedure I*b*, the reduction in the maximum $W1$ value starts after Nd = 25. The same tendency can be observed for Indian, where after Nd = 25 in both I*a* and I*b* $W1$ starts to decrease. For Angust no such reduction is present, though the increase of the $W1$ values reduces as Nd increases. For all structures, when Nd increases, a general increase of the average phase error can be observed, accompanied by a decrease of the average weight of the phases determined.

The computing time involved in the DPP procedure is larger than for the standard *SIMPEL* procedure, since many symbolic additions are carried out simultaneously, but it is still moderate.

6. Concluding remarks

From the test results it can be concluded that the new DPP procedure is superior to the currently available techniques for symbolic phase determination. The DPP procedure incorporates some advantages of the commonly used phase-determinaton techniques: symbolic phase values are employed throughout the procedure, avoiding a premature choice of numerical phase values, while the multi-solution aspect is present as well, overcoming the disadvantage of a single starting set, but in a more systematic way. A second essential feature of the DPP procedure is the absence of an *a priori* fixed starting set. This flexibility offers enormous advantages over the current phase-determination procedures because in this way the selection of the best phasing sequence no longer depends on the outcome of the convergence procedure, but on the actual phase determination itself. Finally, the selection of the symbolic phasing sequences according to the DPP procedure, based on statistical criteria only, ensures that in each step of the process the best possible decisions are made with respect to the criterion function selected, so no better phasing sequence(s) could have been obtained without the criteria being violated.

The authors thank Dr C. H. Stam, whose valuable criticism helped to improve the quality of this paper.

References

AGARWAL, S. K., RASTOGI, R. P., VAN KONINGSVELD, H., GOUBITZ, K. & OLTHOF, G. J. (1980). *Tetrahedron*, **36**, 1435–1438.

BELLMAN, R. (1957). *Dynamic Programming*. Princeton Univ. Press.

COCHRAN, W. (1955). *Acta Cryst.* **8**, 473–478.

COCHRAN, W. & WOOLFSON, M. M. (1955). *Acta Cryst.* **8**, 1–12.

GERMAIN, G., MAIN, P. & WOOLFSON, M. M. (1970). *Acta Cryst.* B26, 274–285.
GERMAIN, G., MAIN, P. & WOOLFSON, M. M. (1971). *Acta Cryst.* A27, 368–376.
GIACOVAZZO, C. (1974). *Acta Cryst.* A30, 631–634.
GIACOVAZZO, C. (1980). *Acta Cryst.* A36, 74–82.
HALL, S. R. (1985). *J. Appl. Cryst.* 18, 263–264.
HASEK, J. (1984). *Acta Cryst.* A40, 338–340.
HULL, S. E. & IRWIN, M. J. (1978). *Acta Cryst.* A34, 863–870.
KANTERS, J. A. & VAN VEEN, A. M. (1973). *Cryst. Struct. Commun.* 2, 261–265.
KARLE, J. & HAUPTMAN, H. (1956). *Acta Cryst.* 9, 635–651.
KARLE, J. & KARLE, I. (1966). *Acta Cryst.* 21, 849–859.
MAIN, P. (1985). *Crystallographic Computing* 3, edited by G. M. SHELDRICK, C. KRÜGER & R. GODDARD, pp. 206–215. Oxford: Clarendon Press.
OVERBEEK, A. R. & SCHENK, H. (1978). *Computing in Crystallography*, edited by H. SCHENK, R. OLTHOF-HAZEKAMP, H. VAN KONINGSVELD & G. C. BASSI, pp. 108–112. Delft Univ. Press.
RYCHLEWSKA, U., BRATEK, M. D. & WIEWIOROWSKI, M. (1978). *Acta Cryst.* A34, S106.
SCHENK, H. (1973). *Acta Cryst.* A29, 77–82.
SCHENK, H. & KIERS, C. T. (1985). *Crystallographic Computing* 3, edited by G. M. SHELDRICK, C. KRÜGER & R. GODDARD, pp. 200–205. Oxford: Clarendon Press.
SHELDRICK, G. M. (1985). *Crystallographic Computing* 3, edited by G. M. SHELDRICK, C. KRÜGER & R. GODDARD, pp. 184–189. Oxford: Clarendon Press.
WATSON, G. N. (1952). *A Treatise on the Theory of Bessel Functions.* Cambridge Univ. Press.

Acta Cryst. (1987). A43, 763–769

The Electron Density of Beryllium Derived from 0·12 Å γ-ray Diffraction Data

BY N. K. HANSEN

Laboratoire de Minéralogie et Cristallographie, Université de Nancy I, *BP* 239, *F*-54506, *Nancy CEDEX, France*

J. R. SCHNEIDER

Hahn-Meitner-Institut, Glienicker Strasse 100, *D*-1000 *Berlin* 39, *Federal Republic of Germany*

AND W. B. YELON AND W. H. PEARSON

University of Missouri Research Reactor, Columbia, Missouri 65211, *USA*

(*Received* 9 *January* 1987; *accepted* 22 *May* 1987)

Abstract

A large set of structure factors for beryllium has been measured with a 0·12 Å γ-ray diffractometer. The present data are in good agreement with other recent X-ray and γ-ray diffraction measurements. Extinction was significant though not very strong. Deformation density maps confirm that the bonding goes through the tetrahedral holes of the hexagonal close-packed structure. Comparisons with band-structure calculations indicate that the local-density approximation to the density functional theory is valid for beryllium. The quality of the data, combined with the earlier experimental results, is sufficient to demonstrate that the structure factors are sensitive to the valence electron hybridization even at high momentum transfer [$(\sin\theta)/\lambda > 0{\cdot}75$ Å^{-1}]. Core electron deformation is a much smaller effect.

Introduction

Diffraction of short-wavelength electromagnetic radiation is a powerful tool for the examination of the electronic structure of crystalline solids. Most commonly the data are interpreted within the kinematical theory of diffraction, which is valid for small and so-called ideally imperfect crystals. Deviations from this theory are nevertheless more often the rule than the exception when X-rays of wavelengths 0·5–0·7 Å are used, and empirical corrections for secondary extinction must be applied to the experimental data. The method developed by Becker & Coppens (1974) is quite powerful for structure determinations and electron-density analysis of organic crystals. For inorganic materials the degree of perfection is often higher and the task of obtaining structure factors with a precision of better than one percent becomes quite difficult. The use of shorter-wavelength radiation or a decrease in the sample size [which requires a high-intensity source like a synchrotron – *e.g.* Bachmann, Kohler, Schulz & Weber (1985)] are ways of improving the situation. The former of these possibilities has been successfully applied with γ-radiation from neutron-activated sources; examples are measurements on plastically deformed copper crystals (Schneider, Hansen & Kretschmer, 1981) and on beryllium crystals of quite high perfection (Hansen,

0108-7673/87/060763-07$01.50