compared with compensation measuring techniques in physics. This way unimportant information will be eliminated while the significance of the information in question will be essentially improved. Additionally this method seems not to be sensitive to systematic errors, since both data sets are affected in the same way. The great advantage of the δ synthesis will be clearly seen in the case of more complex crystal structures than galenobismutite.

For a quantitative application in the case of occupancy deficiency and/or mixed occupancy a better knowledge of the anomalous-dispersion correction terms is necessary. Close to the absorption edges, the theoretical data according to Cromer & Liberman (1981) and Cromer (1983) are obviously incorrect, since these values do not take into account the complicated near-edge structure.

Since the distinction of lead and bismuth is an extreme example considering the small percentage difference of the atomic scattering powers, the results may be generalized for every combination of elements with similar scattering power.

The general applicability of the δ synthesis will be restricted by the limited wavelength range of the monochromators (0·3–2·2 Å). As a result, the δ synthesis can be applied for elements with an atomic number of $Z > 22$. The method can be extended to elements with a lower atomic number by use of vacuum techniques, which reduce significantly the absorption of soft X-rays by air. In this way measurements up to $\lambda \simeq 7$ Å are possible (*e.g.* $Z_{Si} = 13 \Rightarrow \lambda_K = 6{\cdot}7$ Å). However, the volume of the Ewald sphere decreases proportionally and the interpretation of the δ map may fail because only a small number of data is available. In the case of a typical orthorhombic silicate structure with lattice constants of $\sim 10 \times 10 \times 10$ Å, only 12 unique reflections can be measured in the range of $2\theta = 110°$. Additionally, the resolution of the electron density map will be too low ($d_{min} \geq 4{\cdot}09$ Å).

Nevertheless, even under the 'standard' experimental conditions a wide range of elements is covered by the δ synthesis. So this method can be applied advantageously in the scientific field of systematic structure analysis.

The author is greatly indebted to Professor Dr V. Kupčik, who initiated the present investigation, for numerous valuable discussions. Special gratitude is owed to Dipl.Min. M. Wendschuh-Josties for computational and experimental support. Thanks are also due to Professor Dr K. Bente for kindly providing the synthetic sample of galenobismutite and Professor Dr P. Süsse for reading the manuscript. The computations were carried out at the facilities of the Gesellschaft für wissenschaftliche Datenverarbeitung (GWDG) in Göttingen. This work was supported by a grant of the Bundesminister für Forschung und Technologie (BMFT) under contract no. 05 320IA B8.

References

BUSING, W. R., MARTIN, K. O. & LEVY, H. A. (1962). *ORFLS*. Report ORNL-TM-271. Oak Ridge National Laboratory, Tennessee, USA.

CROMER, D. T. (1983). *J. Appl. Cryst.* **16**, 437.

CROMER, D. T. & LIBERMAN, D. (1981). *Acta Cryst.* A**37**, 267–268.

HAMILTON, W. C. (1965). *Acta Cryst.* **18**, 502–510.

IITAKA, Y. & NOWACKI, W. (1962). *Acta Cryst.* **15**, 691–698.

International Tables for X-ray Crystallography (1974). Vol. IV. Birmingham: Kynoch Press. (Present distributor Kluwer Academic Publishers, Dordrecht.)

KARLE, J. (1984). *Acta Cryst.* A**40**, 1–4.

KUPČIK, V. (1979). Private communication.

KUPČIK, V., WULF, R., WENDSCHUH, M., WOLF, A. & PÄHLER, A. (1983). *Nucl. Instrum. Methods*, **208**, 519–522.

OHSUMI, K., TSUTSUI, K., TAKEUCHI, Y. & TOKONAMI, M. (1984). *Acta Cryst.* A**40**, C-255.

SHELDRICK, G. M. (undated). *XEMP*4. Empirical absorption correction program. Univ. of Göttingen, Federal Republic of Germany.

WULF, R. (1986). PhD thesis. Univ. of Göttingen, Federal Republic of Germany.

Acta Cryst. (1990). A**46**, 688–692

On the Construction of Karle–Hauptman Matrices

BY R. DE GELDER, R. A. G. DE GRAAFF AND H. SCHENK

X-ray Department, Gorlaeus Laboratories, PO Box 9502, 2300 *RA Leyden, The Netherlands*

(*Received* 4 *January* 1990; *accepted* 6 *April* 1990)

Abstract

The construction of a Karle–Hauptman matrix or a series of these matrices, suitable for the *ab initio* determination of phases, is presented. An algorithm is proposed which is suggested by graph theory. Maximization of the determinant of Karle–Hauptman matrices, constructed using the new algorithm, as a function of the phases yields phase sets with very low errors compared with earlier results.

0108-7673/90/080688-05$03.00

Introduction

In two preceding papers, describing the work on Karle–Hauptman matrices of our group, a set of routines has been introduced, enabling the use of Karle–Hauptman matrices for the determination of intractable small structures (Vermin & de Graaff, 1978; de Graaff & Vermin, 1982). Essentially, construction of suitably enlarged starting sets of between 20 and 30 reflections proved to be automatic. The phases could be refined to an average phase error of approximately 25° for all five problem structures tested. However, extension of these sets of phases to a set large enough to permit the calculation of E maps proved to be not automatic at all. In three of the structures extension *via* the tangent refinement, using *FASTAN* from the *MULTAN* system, produced at least a partial solution. The fourth failed in this way, but was finally solved using the multi-matrix method – the concurrent maximization of a number of interdependent determinants – proposed by de Graaff & Vermin (1982), while the fifth test structure resisted all attempts at phase extension.

The problematic extension of fairly large starting sets, which are more or less correctly phased, prompted an investigation into the possibilities of doing away with tangent refinement completely and using Karle–Hauptman matrices to find a starting set and to extend it as well. In order to optimize the multi-matrix method it was decided to try two possible lines of investigation:

(1) to improve the quality of the matrices used;

(2) to design an algorithm to find one common origin for the phase sets obtained from different independent matrices.

A first successful attempt was the solution of the fourth test structure mentioned above.

The current paper deals with point 1 and describes an alternative algorithm for the construction of Karle–Hauptman (henceforth KH) matrices. The second point will be discussed at a later date.

A few definitions are first given and then the new algorithm is described in some detail. In the final paragraph various aspects of the starting sets obtained will be compared with earlier results.

The relationship between the properties of the matrix – number of strong reflections, number of symmetry relations *versus* the number of independent phases, average $|E|$ value, number of unobserved reflections – and the size of the phase error in the starting sets will be commented upon.

Definitions

The following abbreviations and notations will be used throughout this paper:

A the KH matrix with elements $E(\mathbf{H})$
m the order of A
B the inverse of A
a_{ij} an element of A $(i,j=1,2,\ldots,m)$
α_{ij} the phase of a_{ij} $(i,j=1,2,\ldots,m)$
b_{ij} an element of B $(i,j=1,2,\ldots,m)$
β_{ij} the phase of b_{ij} $(i,j=1,2,\ldots,m)$
n the number of independent reflections in A
$\mathbf{H}_{ij}$ the reciprocal-lattice vector associated with a_{ij}
N the number of atoms in the unit cell.

Construction of the KH matrix

The reciprocal-lattice vector $\mathbf{H}_{ij}$ of KH matrix element a_{ij} is defined as $\mathbf{H}_{ij}=\mathbf{H}_{1j}-\mathbf{H}_{1i}$. In view of this relation, a KH matrix of order m is defined (Karle & Hauptman, 1950) by a top row of $m-1$ non-identical reflections [element a_{ii} is by definition $E(000)$]. This poses the question: which set of top-row reflections defines a KH matrix of optimum properties?

In terms of mathematical graph theory, a KH matrix may be regarded as a graph formed by the top-row elements. The branches of the graph linking all possible pairs of top-row elements are identical to KH matrix elements lower in the matrix. Reformulating the above question in terms of graph theory: Does a graph exist (formed by top-row elements) where all branches are reflections of the desired type (*e.g.* all branches are strong reflections)?

In the past (Taylor, Woolfson & Main, 1978), several attempts have been made to build matrices with high average $|E|$ values. Among other criteria this is one of the most important conditions for successful maximization of the determinant. Other criteria for the quality of a KH matrix are considered to be: the number of independent phases, the number of unobserved reflections and the distribution of the reflections over the parity groups (Main, 1975).

Two main approaches to the problem of construction can be distinguished: Building a very large matrix and chipping the best possible block from it or starting from a matrix of the desired order and optimizing iteratively. Formulating the problem in terms of graph theory leads directly to a method which combines both ideas. In a topological matrix representing the graph formed by a given set of top-row elements – reflections – each element corresponds to an element of a KH matrix. Replacement of the value of the elements in the topological matrix, normally either zero or one, by a value related to the $|E|$ value of the reflections leads to a modified topological matrix. The top row maximizing the sum of the elements of this matrix (henceforth the quality matrix) is considered to be the top row yielding the best KH matrix. This idea is the basis for the new algorithm in which one tries to find the best top row from a large quality matrix.

The algorithm consists of the following steps:

(1) A large quality matrix (order 500–1000) is constructed from a top row chosen from all large $|E(\mathbf{H})|$'s including symmetry-related reflections. Only

strong reflections are used in the top row with a reciprocal vector of length smaller than a given maximum, avoiding the generation of reflections outside the sphere of measurement in the rest of the KH matrix and stimulating the number of symmetry-dependent reflections in the resulting matrix.

(2) Rows (and corresponding columns) are sorted in descending order based on criteria involving the $|E|$ values of the reflections appearing in the matrix. Criteria based on the maximum values of $\langle|E|^2\rangle$, $\langle||E|^2-1|\rangle$, $\langle|E|^3\rangle$ and $\langle||E|^3-1|\rangle$ give satisfactory results.

(3) A block inside the quality matrix constructed from the first $m-1$ top-row elements (situated in the upper left-hand corner) is optimized iteratively by swapping rows (and columns) for ones further down in the quality matrix. Exchange occurs between the worst row in the block and the best row outside the block, until further improvement is impossible. During this optimization, elements inside the chosen block of order $m-1$ are compared with only those elements outside the block which enter the block by swapping. The order of the KH matrix desired (m) determines directly the course of the construction process. The value of m is chosen between $N/2$ and $N/3$ (Heinerman, Kroon & Krabbendam, 1979).

(4) Step 3 may be repeated for blocks constructed from top-row elements $m+1, m+2, \ldots, 2m$; $2m+1$, $2m+2, \ldots, 3m$ *etc.* Top-row elements from matrices previously optimized are not used again.

(5) From the collection of optimized blocks of order $m-1$ the block with the largest sum of elements is accepted as the most suitable one. In the selection of this block the same criterion is used as in steps (2) and (3). This block is moved to the upper left-hand corner of the quality matrix and is optimized again, now using top-row elements from other optimized blocks as well. This last step may again improve the best block because top-row elements previously unused may now be included.

The advantage of this method compared with earlier ones is the availability of all reflections during the entire construction process. Moreover, several different matrices of comparable quality can be built using the same large starting matrix.

Results and discussion

To test the matrices constructed using the new algorithm, a number of trials was carried out for three of the test structures given by de Graaff & Vermin (1982), *i.e.* TOX, TRIGAL and GLUCOPYR. These three had proved to be the most intractable; the other two, PYROC and ISOPYROC, could be solved easily using the program described in the references.

Rather than trying to produce starting sets suitable for extension by tangent refinement, our aim now is to create a number of large starting sets with small phase errors. The structure will then be solved by the combination of these sets into a whole (see *Introduction*). However, to obtain some basis for comparison, extension of the best sets obtained was tried, with results very similar to those reported earlier.

Table 1. *A survey of the tests of the matrix construction*

All matrices used in the preparation of this table are of the order twenty.

Name	Average $\|E\|$ value	Not observed	Symmetry equivalents	Starting set reflections	Lowest $\Delta\varphi$(°)
TOX	1·406	13	97	32	9·3
TOX	1·413	4	96	38	16·7
TOX	1·446	2	85	43	17·3
TOX	1·375	19	60	50	22·5
TOX	1·413	16	50	69	31·9
TRIGAL	1·298	41	92	22	11·6
TRIGAL	1·276	27	64	33	18·4
TRIGAL	1·385	18	50	50	20·9
TRIGAL	1·389	23	49	55	25·7
TRIGAL	1·509	5	45	62	27·6
GLUCO	1·503	4	103	32	10·5
GLUCO	1·478	5	91	36	18·1
GLUCO	1·453	16	90	43	18·5
GLUCO	1·343	7	46	48	27·5
GLUCO	1·426	12	42	59	34·3
TRIGAL	1·312	46	105	14	>40·0
GLUCO	1·263	11	134	14	>40·0

For each trial matrix 100 solutions were generated, using a random approach similar to that of Yao Jia-xing (1981): all independent phases in the matrix were assigned random values and the determinant was then maximized using the method based on the property that at the maximum the relation $\beta_{ij}-\alpha_{ij}=\pi$ should hold (de Graaff & Vermin, 1982).

In Table 1 a summary is given, listing for each structure a few of the trials only, giving details such as average $|E|$ value, the number of unobserved reflections, the number of independent phases *etc.* and the phase error of the best starting set ($|E|>1{\cdot}3$) obtained. The order m of all matrices used in the tests was 20.

Some of the criteria to steer the construction process lead to conflicting constraints on this process. A useful constraint on the reflections in the top row is to impose a limit on the length of the reciprocal vector of these reflections. From Fig. 1 there is a clear optimum range of this parameter. However, Fig. 2 shows a slow monotonic decrease of the ratio between symmetry-dependent and -independent reflections as a function of the same parameter. Clearly, a suitable compromise has to be found here. For structure TRIGAL a value between 0·5 and 0·6 is optimal.

In general, a high average $|E|$ value and/or a large number of strong reflections in the matrix lead to starting sets with low phase errors. This is illustrated clearly in Fig. 3: a concentration of good matrices is found in the upper right-hand corner (large number of strong reflections, high average $|E|$ value), whereas the majority of the worst matrices are found in the left part of the figure. Obviously, a correlation exists between the number of strong reflections and the average $|E|$ value, demonstrated by the linear trend in the figure. The combination of the number of strong

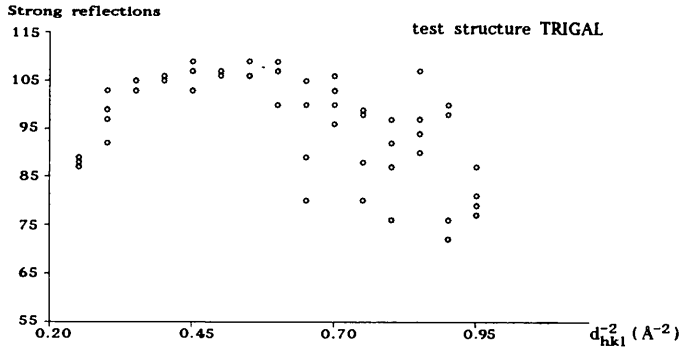

Fig. 1. Number of strong reflections in a KH matrix as a function of the maximum reciprocal-vector length (vector limit) of the reflections forming the top row of the matrix. Each point in the figure represents one KH matrix constructed using the algorithm of the preceding paragraph. For each vector limit four different matrices have been generated using the criteria $\langle|E|^2\rangle$, $\langle\||E|^2-1|\rangle$, $\langle|E|^3\rangle$ and $\langle\||E|^3-1|\rangle$, as indicated in step 2 of the description of the algorithm.

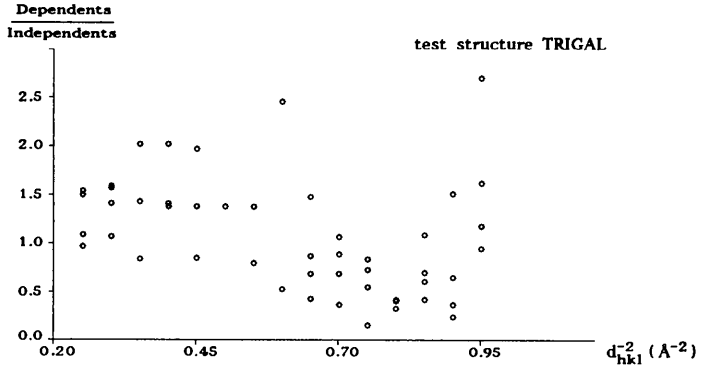

Fig. 2. Ratio of the number of symmetry-dependent reflections to the number of independent reflections in a KH matrix as a function of the maximum reciprocal-vector length (vector limit) of the reflections forming the top row of the matrix. Each point represents a matrix and these matrices are the same as those used in the construction of Fig. 1.

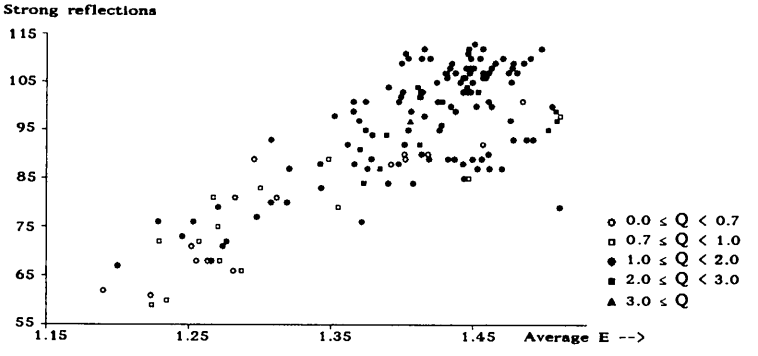

Fig. 3. Quality Q of starting sets obtained as a function of $|E|$ average and the number of strong reflections for different matrices of test structures TOX, TRIGAL and GLUCOPYR. Q is defined as the number of starting-set reflections divided by the lowest $\Delta\varphi$ from the 100 maximization trials.

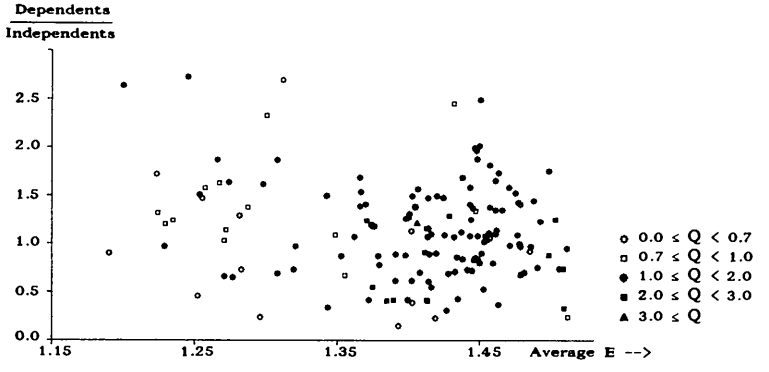

Fig. 4. Quality Q of starting sets obtained as a function of $|E|$ average and the ratio of symmetry-dependent and -independent reflections for different matrices of test structures TOX, TRIGAL and GLUCOPYR. Q is defined as the number of starting-set reflections divided by the lowest $\Delta\varphi$ from the 100 maximization trials.

reflections and the average $|E|$ value may be used as a criterion to select those matrices which on refinement are expected to yield good phases. However, Fig. 3 shows a few matrices of high average $|E|$ value not giving good starting sets at all.

Remarkably, the number of unobserved reflections does not play a significant role. The table shows very good starting sets obtained from matrices containing a large number of unobserved reflections.

The influence of the number of symmetry-equivalent reflections is not very clear. Fig. 4 contains no evidence of good matrices being found for particular values of the ratio between dependent and independent reflections only.

Clearly, the quality of the best matrices produced using the new algorithm far exceeds those from earlier attempts. For all three structures, large starting sets – over 30 reflections – could be generated with very low average phase errors (see Table 1).

Although memory requirements are considerable (2–3 Mbyte), the construction algorithm is very fast. Using a MicroVAX II, the construction takes 1–3 min only.

It is interesting to note that not until the determinant was maximized did the phases of the starting set bear any relation whatsoever to the true phases. This illustrates the validity of the generalized maximum determinant rule (Tsoucaris, 1970; Karle, 1970; Heinerman *et al.*, 1979).

References

GRAAFF, R. A. G. DE & VERMIN, W. J. (1982). *Acta Cryst.* A**38**, 464–470.

HEINERMAN, J. J. L., KROON, J. & KRABBENDAM, H. (1979). *Acta Cryst.* A**35**, 101–105.

KARLE, J. (1970). *Proc. Natl Acad. Sci. USA*, **75**, 2545–2548.

KARLE, J. & HAUPTMAN, H. (1950). *Acta Cryst.* **3**, 181–187.

MAIN, P. (1975). In *Crystallographic Computing Techniques*, edited by F. R. AHMED, pp. 165–175. Copenhagen: Munksgaard.

TAYLOR, D. J., WOOLFSON, M. M. & MAIN, P. (1978). *Acta Cryst.* A**34**, 870–883.

TSOUCARIS, G. (1970). *Acta Cryst.* A**26**, 492–499.

VERMIN, W. J. & DE GRAAFF, R. A. G. (1978). *Acta Cryst.* A**34**, 892–894.

YAO JIA-XING (1981). *Acta Cryst.* A**37**, 642–644.

Acta Cryst. (1990). A**46**, 692–711

Maximum-Likelihood Methods in Powder Diffraction Refinements

BY ANESTIS ANTONIADIS AND JACQUES BERRUYER

University of St Etienne, Department of Mathematics, 23, *Rue du Docteur Paul Michelon*, 42100 *St Etienne CEDEX, France*

AND ALAIN FILHOL

Institut Max von Laue–Paul Langevin, 156*X*, *Centre de Tri*, 38042 *Grenoble CEDEX, France*

(*Received* 15 *November* 1989; *accepted* 4 *April* 1990)

Abstract

The validity of least-squares procedures commonly used nowadays for the analysis of single-crystal, X-ray and neutron diffraction data is examined. An improved methodology that rests on sound statistical theory is proposed and turns out to be a fruitful way to consider any crystallographic refinement. A maximum-likelihood estimation procedure is developed for Poisson regression models. Measures of the goodness of fit (other than the R factor), generalized residuals and diagnostic plots are described. Confidence regions and intervals are also discussed. A set of measures of the influence of data on the fit and the parameter estimates is obtained for Poisson statistics. Finally, the effect of under or over dispersion of the data randomness with respect to a true Poisson distribution is considered and model-independent estimates of this dispersion are discussed.

General notation and symbols frequently used

$\eta, \theta, \ldots$	Lower case greek *italics* denote scalar parameters.
$\boldsymbol{\eta}, \boldsymbol{\theta}, \ldots$	Lower case greek **bold** denote column-vector parameters.
$Y, X, \ldots$	Upper case *italics* normally denote real random variables.
$y, x, \ldots$	Lower case *italics* normally denote observed values of real random variables (realizations).
$\boldsymbol{Y}, \boldsymbol{X}, \ldots$	**Bold** upper case *italics* normally denote column random vectors with corresponding components $Y_i, X_i, \ldots$.

0108-7673/90/080692-20$03.00