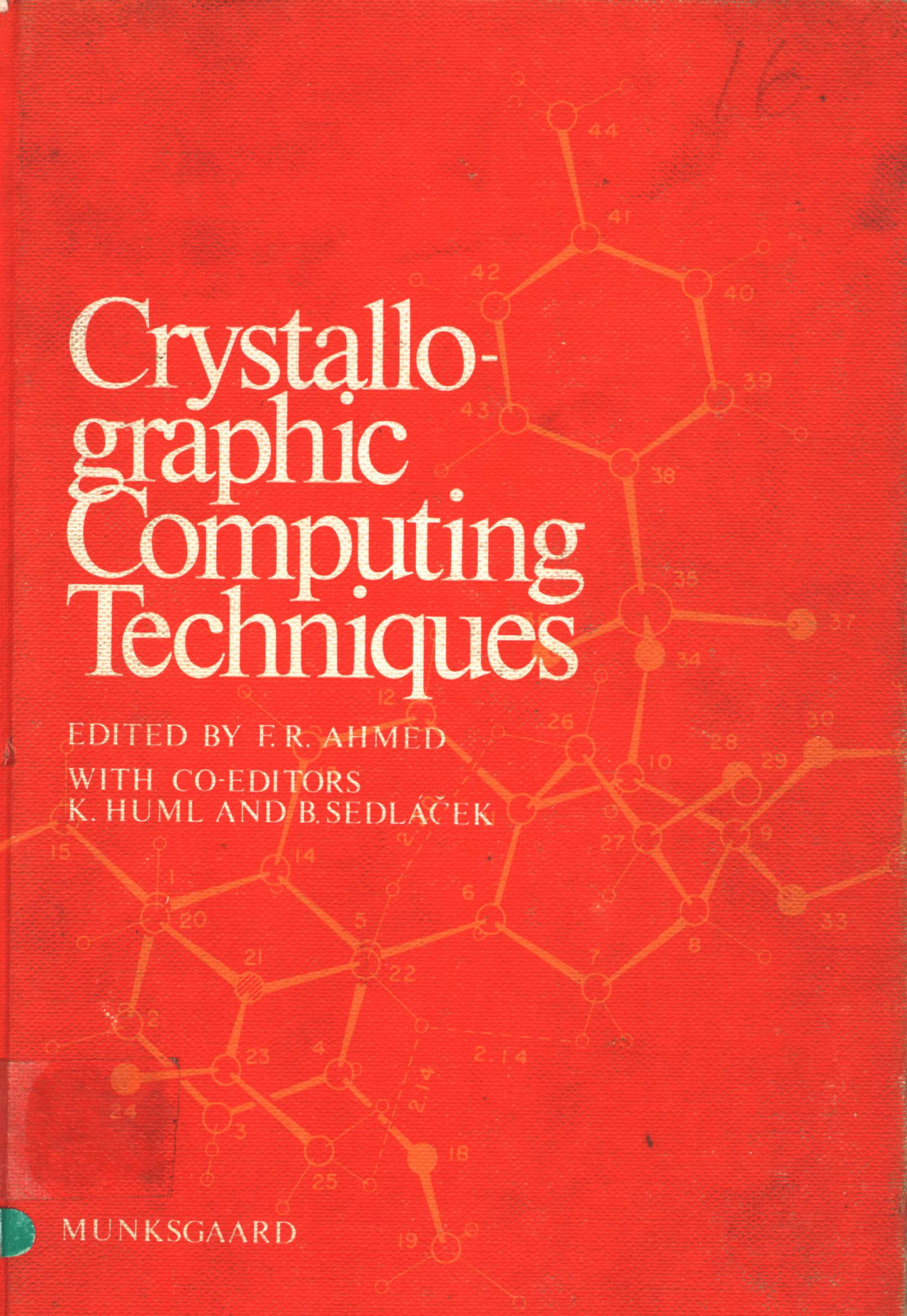
Crystallographic Computing Techniques

EDITED BY F. R. AHMED
WITH CO-EDITORS
K. HUML AND B. SEDLÁČEK



MUNKSGAARD

Crystallographic Computing Techniques

Crystallographic Computing Techniques

Proceedings of an International Summer School organized by

THE COMMISSION ON CRYSTALLOGRAPHIC COMPUTING OF THE INTERNATIONAL UNION OF CRYSTALLOGRAPHY

and held in Prague, Czechoslovakia, 28 July to 5 August 1975

Edited by
F. R. AHMED

with Co-Editors
K. HUML and B. SEDLÁČEK

MUNKSGAARD
INTERNATIONAL BOOKSELLERS and PUBLISHERS LTD., COPENHAGEN.

Published by Munksgaard, Copenhagen 1976
Cover design: F. R. Ahmed and Søren Hansen

Printed in Great Britain by William Clowes & Sons Ltd.,
London, Colchester and Beccles

ISBN 87-16-02132-0

PREFACE

At the invitation of the *Institute of Macromolecular Chemistry* of the *Czechoslovak Academy of Sciences*, the *Commission on Crystallographic Computing* of the *International Union of Crystallography* organized an *International Summer School on Crystallographic Computing* for the period 28 July–5 August 1975. The school was held at the *Agricultural University*, Prague 6-Suchdol, Czechoslovakia, and was attended by 239 crystallographers from 28 different countries. This number included 33 invited lecturers, 9 invited contributors, and the Summer School Editor.

This book contains the proceedings of the school, and covers three main topics which are considered to be of current interest to a wide segment of crystallographers, namely:

(A) Structure Solving by the Direct Method,
(B) Computational Aspects of Protein Crystallography, and
(C) Miscellaneous Crystallographic Computer Applications and Techniques.

In these lectures, you will find brief presentations of the theoretical background, description of some of the well-established techniques and details of their application, and some of the most recent developments in the field. The discussion accompanying each paper, though brief, gives a fairly good representation of the main points raised at the meetings.

The organization of the scientific programme was the responsibility of the Commission: F. R. AHMED (Chairman); S. A. ABRAHAMSSON, G. C. BASSI, A. C. LARSON, T. SAKURAI, K. SASVÁRI, V. I. SIMONOV and M. M. WOOLFSON (Members); K. HUML, J. M. STEWART and G. TSOUCARIS (Consultants), in collaboration with the following Topic Chairmen: D. M. BLOW, J. KARLE, G. KARTHA, P. W. SCHMIDT and B. K. VAINSHTEIN. The careful work of those mentioned, the advice of D. W. J. CRUICKSHANK in the early stages, and the collaboration of the invited Lecturers and Contributors are the main reasons for the high quality of the scientific programme reported in these proceedings.

The Local Organizing Committee consisted of: K. HUML (Chairman); B. SEDLÁČEK (Summer School Editor); V. PETRUS (Social Committee); S. NEŠPŮREK (Technical Committee); J. BALDRIAN, J. HAŠEK, D. HLAVATÁ, J. JEČNÝ, V. LANGER, J. PLEŠTIL and J. ŠOLER (Members); H. KONÁŠOVÁ, B. KOVÁŘÍKOVÁ and H. VORÁČKOVÁ (Summer School Secretariat). The hard work of these colleagues, their dedication and close attention to all details prior to and during the school period, were greatly appreciated by all participants. Their outstanding efforts culminated in a very successful school.

The school gratefully acknowledges the generous support given by the following sponsors: *Charles University In Prague* (Faculty of Natural Sciences, and Faculty of Mathematics and Physics), *Czech Technical University* (Faculty of Nuclear Science and Physical Engineering), *Czechoslovak Scientific and Technical Society, Institute of Chemical Technology* (Prague), *International Business Machines Corporation, International Union of Crystallography, Slovak University of Technology* (Bratislava), and *UNESCO*.

The Editors are particularly grateful to K. FRIML (Head of the Institute of Macromolecular Chemistry, Czechoslovak Academy of Sciences) and A. LÍNEK (Representative of the Czechoslovak National Committee for Crystallography) for arranging the invitation to Prague; to DOROTHY HODGKIN, S. E. RASMUSSEN and J. N. KING (Officers of the International Union of Crystallography) for administrative help and advice; and to MARGARET E. PIPPY (of the National Research Council of Canada) for transcribing the discussion from tapes and for other secretarial assistance. E. J. GABE kindly pre-edited one of the manuscripts.

It is a pleasure to acknowledge the cooperation and hard work of the various authors of the enclosed articles, and to thank them for participating in the school. Like its predecessor *Crystallographic Computing* (1970), this book is published by MUNKSGAARD, International Booksellers and Publishers, Ltd., Copenhagen, at their own financial responsibility. These publishers deserve our compliments for their interest and numerous services in the field of crystallography.

Prague, Czechoslovakia
September, 1975

F. R. AHMED
K. HUML
B. SEDLÁČEK

CONTENTS

A—Structure Solving by the Direct Methods

B—Computational Aspects of Protein Crystallography

C—Miscellaneous Crystallographic Computing Applications and Techniques

TOPIC A

STRUCTURE SOLVING BY THE DIRECT METHOD

TOPIC ORGANIZERS

J. KARLE
V. I. SIMONOV
M. M. WOOLFSON

The Direct Method for Crystal Structure Determination, Mathematical and Philosophical Concepts

JEROME KARLE

Laboratory for the Structure of Matter, Naval Research Laboratory, Washington, D.C. 20375

The field of direct crystal structure determination is marked by a kind of mathematical complexity which is not in any evident fashion amenable to the standard procedures of numerical analysis. Nevertheless, considerable success has been achieved toward the goal of making the direct method of crystal structure determination a fairly routine operation. A clear understanding of this matter as well as of the detailed procedures involved can be obtained from a proper comprehension of the mathematical character and basic philosophy of the direct method.

The term direct structure determination describes the determination of crystal and molecular structure directly from the intensities of the diffracted rays without the use of special information, for example, from the known positions of heavy atoms that may be present, or from isomorphous substitution or anomalous scattering. The Patterson function (Patterson, 1935), a Fourier map of the distribution of interatomic vectors in a crystal, can provide a basis for the direct determination of structure from the intensities of diffraction. However, in its common usage 'direct structure determination' implies the use of procedures involving direct phase determination *i.e.* the determination of structure by use of the intermediate step of evaluating the phases of the diffraction amplitudes directly from the scattered intensities. By means of appropriate Fourier series, knowledge of the phases can immediately provide structural information, thus obviating the need to analyze vector maps.

The complexity of the problem of crystal structure determination derives in part from the complexity of the contents of the unit cells in crystals and in part from the complexity of the mathematical relationship between the atomic positions and the diffracted intensities. As is generally the case for complex systems, the attack on this problem has been characterized by the discovery of special mathematical relationships and properties which could be developed into practical procedures. The identification of the particularly useful relationships and their subtle features took place over a number of years. It is important to recognize that the discovery of mathematical relationships is at best only the beginning. Much effort is often required to develop their practical aspects if, indeed, they exist at all.

Existence of solution to structure problem

There is a sound mathematical basis for expecting that a solution to the problem of determining crystal structures directly from the measured diffraction intensities should exist. This matter will now be considered.

The electron density distribution function $\rho(\mathbf{r})$ describes the structure of a crystal. The maxima of this function locate the atomic positions. Because of the three-dimensional periodicity of a crystal structure, $\rho(\mathbf{r})$ may be represented by a three-dimensional Fourier series,

$$\rho(\mathbf{r}) = V^{-1} \sum_{\mathbf{h}=-\infty}^{\infty} F_{\mathbf{h}} \exp(-2\pi i \mathbf{h}.\mathbf{r}) \qquad (1)$$

where the coefficients

$$F_{\mathbf{h}} = |F_{\mathbf{h}}| \exp(i\varphi_{\mathbf{h}}) \qquad (2)$$

are the crystal structure factors, the amplitudes of coherent scattering associated with plane segments in the unit cell labeled with the vectors **h**. The components of **h** are the integers h,k,l which are inversely proportional to intercepts of the corresponding planes on the chosen axes. The angle $\varphi_{\mathbf{h}}$ is the phase associated with $F_{\mathbf{h}}$ and **r** is the position of any point in the unit cell.

The structure of any crystal could be readily computed from (1) if the coefficients $F_{\mathbf{h}}$ were directly available from experiment. Ordinarily, only the magnitudes $|F_{\mathbf{h}}|$ of the coefficients are obtainable from experiment which are proportional to the square roots of the corresponding experimentally measured intensities. However, one of the subtle features of the mathematics of diffraction from crystals is the fact that the phase information, which seems to be lost in a diffraction experiment, is in fact contained in the measured intensities of scattering. The problem of determining crystal structures is associated with the apparent absence of phase information which, in this context, is generally referred to as the 'phase problem'. The manner in which the values of the phases of the structure factors can be extracted from the experimentally determined magnitudes of the structure factors is, in essence, the main consideration of this article.

The Fourier inversion of (1) followed by the reduction of the integral to the sum of contributions from the N discrete atoms in the unit cell gives for the Fourier coefficient,

$$|F_{\mathbf{h}}| \exp(i\varphi_{\mathbf{h}}) = \sum_{j=1}^{N} f_{j\mathbf{h}} \exp(2\pi i \mathbf{h}.\mathbf{r}_j) \quad (3)$$

where $f_{j\mathbf{h}}$ is the atomic scattering factor of the jth atom in the unit cell and $\mathbf{r}_j$ is its position vector.

The solvability of the phase problem can be established by examining the set of simultaneous equations comprising Eqs. (3). Equations (3) form a system of simultaneous equations since a large number of vectors **h** are considered. The unknown quantities are the phases $\varphi_{\mathbf{h}}$ and the atomic position vectors $\mathbf{r}_j$. The known quantities are the magnitudes of the structure factors $|F_{\mathbf{h}}|$, obtainable from experiment, and the atomic scattering factors $f_{j\mathbf{h}}$ which are tabulated. By multiplying Eqs. (3) by their complex conjugates, phases are eliminated giving

$$|F_{\mathbf{h}}|^2 = \sum_{j=1}^{N} \sum_{k=1}^{N} f_{j\mathbf{h}} f_{k\mathbf{h}} \exp 2\pi i \mathbf{h}.(\mathbf{r}_j - \mathbf{r}_k). \qquad (4)$$

The unknown quantities in Eqs. (4) are the independent atomic coordinates in an asymmetric unit of the unit cell. A copper target for X-radiation usually provides the values of the intensities, $|F_{\mathbf{h}}|^2$, for as many as 150 independent reflexions for each atom in the asymmetric unit of a centrosymmetric crystal, and 75 independent reflexions for each atom in the asymmetric unit of a noncentrosymmetric crystal. Since each atom has three positional coordinates, the simultaneous Eqs. (4) would be overdetermined by a factor of about 50 for centrosymmetric crystals and by about 25 for noncentrosymmetric ones. Although there is some inaccuracy of a few percentages in the values of the $|F_{\mathbf{h}}|^2$ and the $f_{j\mathbf{h}}$, the great overdeterminancy of Eqs. (4) makes these errors unimportant.

Equations (3) from which Eqs. (4) are obtained contain the additional unknown phases. The overdeterminancy for this set is the same as that for Eqs. (4) because each equation in set (3) involving complex quantities is actually two equations, one for the real part and one for the imaginary part. Thus, the problem of determining phases from the measured intensities is greatly overdetermined. The overdeterminancy provides the rationale and the

motivation for searching for a solution to the problem of determining structures directly from the intensities of diffraction. It is generally considered more readily feasible to determine structures by first determining phases and then computing Eq. (1) than to find the atomic coordinates directly.

Atomic coordinates may be obtained directly from an analysis of the Patterson function (Buerger, 1959; Hoppe, 1957; Nordman & Nakatsu, 1963; Huber & Hoppe, 1965)

$$P(\mathbf{r}) = \sum_{\mathbf{h}=-\infty}^{\infty} |F_{\mathbf{h}}|^2 \exp(-2\pi i \mathbf{h}.\mathbf{r}). \qquad (5)$$

The maxima of this function represent the interatomic vectors in a structure. Evidently the coefficients of the series are not dependent upon a knowledge of the phases and therefore the Patterson function can be calculated for any crystal from the measured X-ray intensities. The analysis of the Patterson function is generally very difficult because of the lack of resolution of the $N(N-1)$ interatomic vectors. However in the case that a structure possesses only a few heavy atoms, the Patterson function is quite useful because the interatomic vectors associated with the heavy atoms are readily identifiable. The positions of the heavy atoms, as obtained from the interatomic vectors, can be used in Eqs. (3) to calculate an initial set of phases. There are several methods for developing a complete structure from this information (Woolfson, 1956; Bertaut, 1957; Sim, 1960; Hoppe & Huber, 1963; Srinivasan, 1966; Karle, 1968).

Despite the limitations on the direct analysis of a Patterson function, the great overdeterminancy of the structure problem, as previously demonstrated, implies that there may well be an alternative method which extends the range of complexity of structures that can be readily handled by X-ray analysis. The direct method of phase determination has this capability.

It is of interest to examine the mathematical background and philosophic concepts which have led to the development of this method.

Basic concepts and mathematical relationships

It is often important, in attempting to solve physical problems, to take full advantage of the various special mathematical and physical properties of the system under study. The basic property leading to the practical solution of the problem of determining phases from the measured X-ray diffraction intensities is the non-negativity of the electron density distribution. This property of non-negativity was utilized to obtain the main formulas for phase determination (Karle & Hauptman, 1950).

The concept of non-negativitity in structure analysis was first developed for analysis of molecular structure utilizing electron diffraction by gaseous molecules (Karle & Karle, 1949, 1950). The problem in this field of structure determination was to find a suitable background intensity, representing the atomic coherent and incoherent scattering, so that the molecular interference intensity could be accurately separated from the total intensity of scattering. The total intensity is composed of the molecular intensity plus the background intensity. The problem was solved by recognizing that the Fourier sine transform of a function of the molecular intensity is related to the probability of finding interatomic distances in the molecule, a radial distribution function. It is apparent that the Fourier transform, representing probabilities, must be non-negative. The shape of the background intensity is determined by requiring this smooth function to have the property that the corresponding molecular intensity function have a non-negative Fourier sine transform (Karle & Karle, 1949, 1950).

The application of this mathematical requirement was successful and led to the search for other possible applications of the concept of non-negativity in the field of structure research. It is quite apparent that the electron density distribution function for crystals is a non-negative function. Relationships occur among the structure factors as a consequence of the non-negativity which take the form of an infinite set of determinantal inequalities of increasing order whose elements are the structure factors (Karle & Hauptman, 1950). In order of complexity, the first inequality statement is that F_{000} must be non-negative, the second is that $|F_{\mathbf{h}}| \leqslant F_{000}$ and the third gives a relationship among the structure factors which is of great significance to direct crystal structure analysis. The inequality may be written

$$\begin{vmatrix} F_{000} & F_{-\mathbf{k}} & F_{-\mathbf{h}} \\ F_{\mathbf{k}} & F_{000} & F_{-\mathbf{h}+\mathbf{k}} \\ F_{\mathbf{h}} & F_{\mathbf{h}-\mathbf{k}} & F_{000} \end{vmatrix} \geqslant 0. \quad (6)$$

The most significant features of this inequality cannot be readily seen when written in the form of (6). In order to understand the meaning of (6), it is important to rewrite it in the alternative form given in the 1950 paper of Karle & Hauptman.

$$\left| F_{\mathbf{h}} - \frac{F_{\mathbf{k}} F_{\mathbf{h}-\mathbf{k}}}{F_{000}} \right| \leqslant \frac{\begin{vmatrix} F_{000} & F_{-\mathbf{k}} \\ F_{\mathbf{k}} & F_{000} \end{vmatrix}^{1/2} \begin{vmatrix} F_{000} & F_{-\mathbf{h}+\mathbf{k}} \\ F_{\mathbf{h}-\mathbf{k}} & F_{000} \end{vmatrix}^{1/2}}{F_{000}}. \quad (7)$$

The interpretation of this inequality is that the structure factor $F_{\mathbf{h}}$ is bounded by a circle in the complex plane which has its center at $F_{\mathbf{k}} F_{\mathbf{h}-\mathbf{k}}/F_{000}$ and a radius given by the right side of (7). For the structure factors of unusually large magnitude, the inequality (7) becomes quite restrictive because the right side of (7) becomes rather small and it is a simple matter to conclude that $\varphi_{\mathbf{h}}$, the phase of $F_{\mathbf{h}}$, is approximately equal to $\varphi_{\mathbf{k}} + \varphi_{\mathbf{h}-\mathbf{k}}$, the phase of $F_{\mathbf{k}} F_{\mathbf{h}-\mathbf{k}}$. By defining 'sharpened' structure factors *e.g.* the unitary structure factors,

$$U_{\mathbf{h}} = F_{\mathbf{h}} \Big/ \sum_{j=1}^{N} f_{j\mathbf{h}},$$

inequality (7) may be replaced by

$$|U_{\mathbf{h}} - U_{\mathbf{k}} U_{\mathbf{h}-\mathbf{k}}| \leqslant \begin{vmatrix} 1 & U_{-\mathbf{k}} \\ U_{\mathbf{k}} & 1 \end{vmatrix}^{1/2} \begin{vmatrix} 1 & U_{-\mathbf{h}+\mathbf{k}} \\ U_{\mathbf{h}-\mathbf{k}} & 1 \end{vmatrix}^{1/2} \quad (8)$$

Such an inequality is strengthened as compared to (7) because the right side of (8) is relatively smaller in the sense that $|U_{\mathbf{k}}| > |F_{\mathbf{k}}|/F_{000}$ and $|U_{\mathbf{h}-\mathbf{k}}| > |F_{\mathbf{h}-\mathbf{k}}|/F_{000}$. It is important at this point to note, however, that the size of the bounding circle is not a satisfactory measure of the validity of the relation

$$\varphi_{\mathbf{h}} \sim \varphi_{\mathbf{k}} + \varphi_{\mathbf{h}-\mathbf{k}}. \quad (9)$$

Were this the case, inequalities (7) and (8) and their implication (9) would appear to be applicable to only the simplest crystals.

To obtain a proper understanding of the broad applicability of relations (7)–(9) it is necessary to consider another aspect of the inequalities, namely their probabilistic properties. Previous to the development of the inequalities such as (6)–(8), Harker & Kasper (1948) developed a system of inequalities based on the Schwarz and Cauchy inequalities. It was apparent after they were obtained that the complete set of inequalities of Karle & Hauptman (1950), derived from the non-negativity of the electron distribution, contained the Harker–Kasper inequalities. Non-negativity was implicitly assumed in the derivation of the latter. Although inequalities (7) or (8), having principal significance to present methods of phase determination, are not to be found explicitly stated among the Harker–Kasper inequalities, the inequalities of Harker and Kasper have