GOODMAN, D. M., JOHANSSON, E. M. & LAWRENCE, T. W. (1993). *Multivariate Analysis: Future Directions*, edited by C. R. RAO, ch. 11. Amsterdam: Elsevier.
HODEL, A., KIM, S.-H. & BRÜNGER, A. T. (1992). *Acta Cryst.* A**48**, 851–858.
HYNES, T. R. & FOX, R. O. (1991). *Proteins: Struct. Funct. Genet.* **10**, 92–105.
JONES, T. A. (1985). *Methods Enzymol.* **115**, 157–171.
LOLL, P. J. & LATTMAN, E. E. (1989). *Proteins: Struct. Funct. Genet.* **5**, 183–201.
MAALOUF, G. J., HOCH, J. C., STERN, A. S., SZÖKE, H. & SZÖKE, A. (1993). *Acta Cryst.* A**49**, 866–871.
OGATA, C. M., GORDON, P. F., DE VOS, A. M. & KIM, S.-H. (1992). *J. Mol. Biol.* **228**, 893–908.
PFLUGRATH, J. W., SAPER, M. A. & QUIOCHO, F. A. (1984). *Methods and Applications in Crystallographic Computing*, edited by S. HALL & T. ASHIAKA, p. 407. Oxford: Clarendon Press.
PRESS, W. H., FLANNERY, B. P., TEUOLSKY, S. A. & VETTERLING, W. T. (1989). *Numerical Recipes: the Art of Scientific Computing (FORTRAN version)*. Cambridge Univ. Press.
STARK, H. (1987). Editor. *Image Recovery: Theory and Application.* Orlando: Academic Press.
SZÖKE, A. (1993). *Acta Cryst.* A**49**, 853–866.
WILSON, A. J. C. (1949). *Acta Cryst.* **2**, 318–321.

*Acta Cryst.* (1995). A**51**, 708–716

# Solution of the Phase Problem in Crystallography and Application to Dynamical Electron Diffraction

BY WILLIAM F. TIVOL

*Wadsworth Center for Laboratories and Research and the School of Public Health, Empire State Plaza, PO Box* 509, *Albany, NY* 12201-0509, *USA*

(*Received* 5 *September* 1994; *accepted* 6 *March* 1995)

## Abstract

Unitarity, a fundamental principle of scattering theory, leads to the prediction of an essentially unique set of phases for the scattering amplitude from a complete knowledge of the differential cross section or, in the case of a crystal, from the diffracted intensities. The Sayre equation and all the direct methods of phasing following therefrom are derived as a special case of unitarity for zero excitation error. Dynamical and kinematical scattering are considered, and the relationship between them, $\hat{S} = \exp(i\tau n_z \hat{K})$, is obtained. Applications to the case of electron diffraction including for non-zero excitation error are discussed.

## Introduction

When diffraction was first used to calculate molecular structures, it was realized that in addition to the intensities, which were directly measured, phases had to be determined for each of the reflections. Many schemes were devised to ascertain these phases, such as comparing isomorphous crystals, one of which had one or more heavy atoms that were lacking in the other, or looking at an unknown molecule which had as part of its structure a molecule whose structure was already known (Argos & Rossmann, 1980). These methods were very successful; however, not all materials of interest could be crystallized with and without heavy atoms or described by a known part plus an unknown part.

An alternative procedure is the derivation of the phases from the values of the measured intensities. All such techniques of using the known intensities to provide information about the unknown phases are collected under the category of direct methods of phase determination. Many of these methods are based on an equation first derived by Sayre (1952), who calculated the diffraction amplitudes of an arrangement of equal non-overlapping atoms and of the same arrangement of 'squared atoms'. By comparing the Fourier expansions of these two expressions, he was able to relate one (phased) amplitude to a convolution of all other (phased) amplitudes:

$$F(\mathbf{H}) = (\theta/V) \sum_{\mathbf{K}} F(\mathbf{K}) F(\mathbf{H} - \mathbf{K}), \qquad (1)$$

where $\mathbf{H}$ and $\mathbf{K}$ are sets of Miller indices, $F$ is the complex amplitude, $V$ is the unit-cell volume and $\theta$ is a constant of proportionality.

Direct methods of phasing diffraction patterns have been used very successfully in *ab initio* structure solutions in both X-ray (Ladd & Palmer, 1980; Day & Pendry, 1993; Glusker, 1993) and electron crystallography (Dorset, 1993; Dorset, Tivol & Turner, 1991, 1992, 1993). Phase extension, where initially the low-order phases are determined by some means and direct methods are used to relate the higher-order phases to the low-order ones, have also been quite successful (Dorset, 1993; Dorset, Kopp, Fryer & Tivol, 1995).

It has been stated many times that the phases can be extracted from the measured intensities because the electron density is everywhere positive and the unit cell of a crystal consists of equal non-overlapping point-like atoms. It is also stated that the fact that the atoms are not




ISSN 0108-7673 

equal for most interesting cases is not too important from a practical point of view. It would, however, be desirable to have a firm theoretical foundation for the use of direct methods even for circumstances for which it has heretofore not been recognized that they apply.

In particular, one result of the present work is that the phase of the scattering amplitude can be derived from the differential cross section for the general scatterer, and for a crystal, even in the cases of dynamical or other multiple scattering, the unitarity of the scattering operator implies that an essentially unique set of phases for the scattering amplitudes is determined by the intensities.

It must be emphasized that the dynamical amplitudes are not related to the structure factors as simply as are the kinematical amplitudes, so that solving a structure by means of the dynamical amplitudes requires more than just the correct determination of the phases. Although the complete process of calculating the structure from dynamical diffraction information is beyond the scope of this paper, there is some discussion of the process in the last two sections.

The triplet formula is perhaps the best known direct phasing technique. To use this formula, the diffraction amplitudes are listed in order of their magnitudes (amplitudes normalized to remove the overall angular dependence of scattering are most often used at present). Those amplitudes greater than a cut-off value are examined to see whether there are triplets whose Miller indices add to zero (*e.g.* 123, $\bar{2}0\bar{4}$ and $1\bar{2}1$). Whenever such triplets are found, the sums of their phases are set to zero. In addition to these relationships, phases can be chosen for (in general) three structure factors that determine the unit-cell origin. Other triplet relationships that include up to three or so algebraic constants can also be used. A set of potential maps is constructed, one map for each choice for the algebraic constants. One of the maps must be interpretable in terms of expected atomic positions. If not, then a different cut-off is used for the amplitudes and the process is started again. Once a starting map has been selected, the phases for all the observed reflections are calculated for the atoms localized on the map. A second map is constructed from the phased amplitudes and, if the process is working, new atoms appear on the map, and the old peaks are reinforced. In this way, better maps are successively constructed, and the positions are refined from the best map using the usual Fourier techniques.

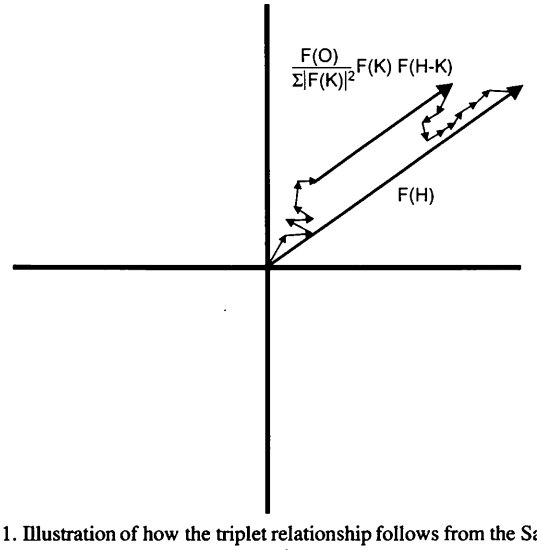


Fig. 1. Illustration of how the triplet relationship follows from the Sayre equation.

To see how the triplet formula follows from the Sayre equation, $F(\mathbf{H})$ and individual terms of the sum are shown in the complex plane in Fig. 1. It is assumed that $F(\mathbf{H})$ is large as shown in Fig. 1. If one of the terms in the convolution is also large (again as shown in Fig. 1), it is likely to be approximately parallel to $F(\mathbf{H})$, since there is not enough excess length in the other terms of the convolution to add up to $F(\mathbf{H})$ in the case where the phase of the large segment differs greatly. For centrosymmetric unit cells, the phases can only be 0 or $\pi$, so terms must be parallel or antiparallel, and the triplet relation is especially useful in these circumstances. In a few instances, when the magnitudes of $F(\mathbf{K})$ and $F(\mathbf{H}-\mathbf{K})$ are large enough, the likelihood of parallelism is 100%, but usually there is only a probabilistic relationship with a likelihood of somewhat less than 100%.

Simultaneous consideration of all the triplet relations for a particular value of $\mathbf{H}$ contained in the Sayre equation, each with its associated probability of being correct, leads to a phase prediction

$$\tan\varphi_{\mathbf{H}} = \frac{\sum w_{\mathbf{H}}|F(\mathbf{K})||F(\mathbf{H}-\mathbf{K})|\sin(\varphi_{\mathbf{K}}+\varphi_{\mathbf{H}-\mathbf{K}})}{\sum w_{\mathbf{H}}|F(\mathbf{K})||F(\mathbf{H}-\mathbf{K})|\cos(\varphi_{\mathbf{K}}+\varphi_{\mathbf{H}-\mathbf{K}})}, \tag{2}$$

where $\varphi$ is the phase of the complex amplitude and $w$ is the weighting factor; this is the tangent formula for the phases.

Simultaneous consideration of all the relationships defined by the Sayre equation for all $\mathbf{H}$ leads to other phase-determining algorithms. One particular algorithm, maximum entropy, calculates the joint probabilities of a new set of phases given a starting set, examines the likelihood of observing the moduli of the new set given the amplitudes of the starting set *versus* the hypothesis that the starting-set amplitudes are zero, and calculates the *a posteriori* probability that the starting set is correct given the new set (Bricogne, 1991). Proceeding in this way, predictions are eventually made for all phases.

The triplet formula gives unambiguous predictions but requires a set of large normalized amplitudes. The tangent formula and maximum entropy give more ambiguous results (the correct solution path must be chosen from among several possibilities at each stage, thus, they are called multisolution methods) but, because all the phase predictions are used simultaneously, either for one or for all phases, the variation in the sizes of the

normalized amplitudes, needed for the triplet method, is not required for the multisolution methods.

One of the most fundamental laws in physics is that the scattering operator, $\hat{S}$, is unitary. Unitarity means that total probability is conserved; *i.e.* if, initially, there is a distribution of states whose total probability is unity, then, after the scattering process, the total probability of all the final states is also unity.

Mishnev (1991) pointed out that the Sayre equation can be derived as a special case of unitarity. Gerber & Karplus (1972) showed that the phases can always be derived from the intensities of a scattering potential that obeys the unitary relation and that, under certain conditions, it is possible to write an algorithm that converges uniformly to a unique set of phases [excepting the known ambiguities: (1) if the operator $\hat{A}$ is a solution to the unitary equation, so is $-\hat{A}^\dagger$; (2) each phase, $\varphi$, can be replaced by $\varphi + 2n\pi$]. Newton (1968) proved explicitly the existence of an essentially unique solution for the phases (except for the ambiguities listed above) and derived an algorithm for the solution under fairly non-restrictive conditions, and he showed the existence, but not the uniqueness, of a solution for any unitary scattering process. Martin (1969) showed the uniqueness of the phase solution within a broader range of conditions and gave heuristic evidence that the phase solution is unique whenever unitarity holds. Sakurai (1967) derived the unitary equation covalently and applied it to scattering of relativistic electrons. The net result of this set of papers is that, for all but a small class of scattering potentials, an essentially unique solution for the phases is determined by the values of the differential cross section, and for that small class there is reason to believe that the phase solution is also essentially unique.

These papers must be examined closely in order to be sure that the results are applicable to the case of dynamical scattering in electron diffraction. Gerber & Karplus (1972) concentrated on scattering of electromagnetic waves with the express purpose of applying the results to X-ray crystallography. Towards the end of the paper, they considered a general scatterer, where they showed that similar equations can be used to determine the phases, but convergence of the equations and uniqueness of the solution were not proved.

The proofs in Gerber & Karplus (1972) invoke the methods used by Newton (1968) and Martin (1969) and, in Newton's paper, the differential cross section must be constrained so that it cannot be too small in any one interval relative to its value at all other points, and that it cannot be zero anywhere. Furthermore, he assumed that the differential cross section is continuous. For the case of diffraction from an ideal infinite crystal, however, the differential cross section is zero almost everywhere and is not continuous. Fortunately, the need for non-zero cross sections is only to make an integral,

$$I = \int A_{\mathbf{k}'',\mathbf{k}'} A_{\mathbf{k}'',\mathbf{k}} \mathrm{d}\mathbf{k}'' / A_{\mathbf{k}',\mathbf{k}}, \qquad (3)$$

non-singular. $A$ is the amplitude for scattering from one momentum state ($\mathbf{k}$) into another. Thus, if there is a point where the denominator is zero, it suffices to have the numerator go to zero in the same way. If two points are on a regular lattice, the point at the sums of their indices is also on the lattice. Therefore, if the point at the sums of indices is not on the lattice, then at least one of the other points must also not be on the lattice. In this case, the integral can rigorously be replaced with a sum over the lattice points, continuity is not then required, and the proof is still valid.*

The same caveats apply to the proofs in Martin's (1969) paper, since the same procedures were used. In the consideration of the general scatterer that obeys unitarity, but for which the cross sections do not fall within the restrictions that allow one to show rigorously that the solution is essentially unique, Martin considered the case where the cross section can be described exactly by a finite number of partial waves, and concluded that, once again, an essentially unique solution is determined by unitarity.

### Extension of the results of Martin

In the following, $\mathrm{d}\sigma/\mathrm{d}\Omega = 1/k^2 |A_{\mathbf{k}',\mathbf{k}}|^2$. This gives for the (unitary) scattering operator $\hat{S} = \hat{I} + (ik/2\pi)\hat{A}$. The unitary relation is just

$$\hat{S}^\dagger = \hat{S}^{-1} \quad \text{or} \quad \hat{S}\hat{S}^\dagger = \hat{I}, \qquad (4)$$

where $\dagger$ means conjugate transpose. In terms of $\hat{A}$, we have

$$\hat{A} - \hat{A}^\dagger = (ik/2\pi)\hat{A}^\dagger \hat{A}. \qquad (5)$$

Specifying the initial‡ and final states and inserting a complete set of intermediate states on the right-hand side gives

$$A_{fi} - A_{fi}^* = (ik/2\pi) \sum A_{fn}^* A_{ni}, \qquad (6)$$

where $A$ is the scattering amplitude between pairs of states. For initial momentum vector $\mathbf{k}$, final momentum vector $\mathbf{k}'$ and intermediate-state momentum vector $\mathbf{k}''$, conserving momentum and energy gives

* For the case of systematic absences, where the amplitude at a lattice point is zero, the unitary equation may not give a solution for that phase. However, since terms containing that phase are always multiplied by the amplitude, they make no contribution to the equations determining the other phases. Furthermore, the phase of a zero amplitude has no physical meaning.

‡ *N.B.* The fact that the unitary equation is an operator equation allows complete freedom to choose the initial state. Thus, in addition to the usual experimental condition where the initial state is a plane wave along the $z$ axis, the initial state can be chosen to represent a combination of experiments where the initial state is, for example, incident plane waves along several zone axes with phases randomized or an incoming modulated spherical wave that does not correspond to a realizable experimental condition.

$k = |\mathbf{k}| = |\mathbf{k}'| = |\mathbf{k}''|$, and therefore we find

$$A_{\mathbf{k}',\mathbf{k}} - A^*_{\mathbf{k}',\mathbf{k}} = (ik/2\pi) \int A^*_{\mathbf{k}',\mathbf{k}''} A_{\mathbf{k}'',\mathbf{k}} \mathrm{d}\Omega_{\mathbf{k}''}. \quad (7)$$

If $A_{\mathbf{k}'',\mathbf{k}}$ is expressed as a partial-wave series, $A^L_{\mathbf{k}'',\mathbf{k}}$ is the partial sum for $l \leq L$,

$$A^L_{\mathbf{k}'',\mathbf{k}} = \sum_0^L \sum_{m=-l}^{l} (2l+1) \exp(i\delta_{l,m}) \sin\delta_{l,m} Y_l^m(\theta_{\mathbf{k}'',\mathbf{k}}, \varphi_{\mathbf{k}'',\mathbf{k}}) \quad (8)$$

(using the spherical harmonic functions instead of Legendre polynomials, since the scattering potential is not necessarily cylindrically symmetric). In the case where the partial wave expansion converges, *i.e.* where $A_{\mathbf{k}'',\mathbf{k}}$ is of bounded variation, then, for any $\varepsilon > 0$, there is an $L_\varepsilon$ such that, for all $L > L_\varepsilon$,

$$\int |A_{\mathbf{k}'',\mathbf{k}} - A^L_{k'',k}|^2 \, \mathrm{d}\Omega_{\mathbf{k}''} < \varepsilon, \quad (9)$$

with similar equations for $A_{\mathbf{k}',\mathbf{k}''}$ and $A_{\mathbf{k}',\mathbf{k}}$. Furthermore, since the total intensity of the scattered wave cannot exceed the intensity of the incident wave, $\sigma_{\mathrm{tot}} \leq 1$, and

$$\sigma_{\mathrm{tot}} = \int |A_{\mathbf{k}'',\mathbf{k}}|^2 \mathrm{d}\Omega_{\mathbf{k}''} = \sum_0^\infty \sum_{m=-l}^{l} |(2l+1)\exp(i\delta_{l,m})\sin\delta_{l,m}|^2; \quad (10)$$

thus, again, for $L > L_\varepsilon$, we have

$$\sum_L^\infty \sum_{m=-l}^{l} |(2l+1)\exp(i\delta_{l,m})\sin\delta_{l,m}|^2 < \varepsilon. \quad (11)$$

Taking $L_\varepsilon^{\max}$ to be the largest of the four values of $L_\varepsilon$ found, the equation for the phase (Martin, 1969),

$$4\pi |A_{\mathbf{k}',\mathbf{k}}| \sin(\varphi_{\mathbf{k}',\mathbf{k}}) = \int |A_{\mathbf{k}'',\mathbf{k}}||A_{\mathbf{k}'',\mathbf{k}'}| \cos(\varphi_{\mathbf{k}'',\mathbf{k}} - \varphi_{\mathbf{k}'',\mathbf{k}'}) \, \mathrm{d}\Omega_{\mathbf{k}''}, \quad (12)$$

has coefficients of the phases that are arbitrarily close to those of the sequence of equations with $L > L_\varepsilon^{\max}$,

$$4\pi |A^L_{\mathbf{k}',\mathbf{k}}| \sin(\varphi^L_{\mathbf{k}',\mathbf{k}}) = \int |A^L_{\mathbf{k}'',\mathbf{k}}||A^L_{\mathbf{k}'',\mathbf{k}'}| \cos(\varphi^L_{\mathbf{k}'',\mathbf{k}} - \varphi^L_{\mathbf{k}'',\mathbf{k}'}) \, \mathrm{d}\Omega_{\mathbf{k}''}, \quad (13)$$

which have the essentially unique solutions $\varphi^L_{\mathbf{k}',\mathbf{k}}$. For each successive equation, the functions $|A^L|$ are continuous and differentiable; thus, since the sequence of functions $|A^L|$ converges to $|A|$, the sequence of functions, $\varphi^L_{\mathbf{k}',\mathbf{k}}$, converges to $\varphi_{\mathbf{k}',\mathbf{k}}$ in the limit $L \to \infty$. Note that $\varphi_{\mathbf{k}',\mathbf{k}}$ can be a continuous function even when $|A_{\mathbf{k}',\mathbf{k}}|$ has discontinuities, such as for the case of an ideal crystal.

Finally, to apply the foregoing to scattering by a crystal, consider a column of $n_z$ unit cells having an axis parallel to the incident beam. The scattering from such a column is well behaved, *i.e.* the differential cross section is continuous, differentiable and never equal to zero, and the phase solution is, therefore, essentially unique. Now consider a set of columns of $n_z$ unit cells and extending $n_x$ cells in the $x$ direction and $n_y$ cells in the $y$ direction. The scattering is still well behaved and the phases can still be determined. As $n_x$ and $n_y$ increase, the amount of dynamical scattering changes somewhat, since electrons can be scattered by more than one column. However, since only the electrons near the edge can scatter through the edge of the crystal, and, therefore, fail to undergo the same extent of dynamical scattering as the rest, then, as the outer columns of the crystal become a smaller fraction of the total area, the extent of dynamical scattering converges to a limit. Furthermore, if the contribution of dynamical scattering is the same, the phase solution for $n_x \times n_y$ cells is the same for all $n_x$ and $n_y$, since it is the identity of the phases that leads to the phenomenon of diffraction in the first place. Thus, for a crystal of infinite extent in $x$ and $y$, the phases are determined essentially uniquely by unitarity. A crystal extending infinitely in the $z$ direction is not treated by scattering theory, since the $\hat{S}$ operator takes the asymptotic limit wave function at $t = -\infty$ (thus at $z = -\infty$) into the asymptotic limit wave function at $t = +\infty$ (thus at $z = +\infty$) and the asymptotic limit exists only when the potential goes to zero for large $|z|$.

Therefore, for the case of ideal crystals of finite thickness, whether of finite or infinite extent, the phases are determined essentially uniquely by the intensities.

The ambiguity that both $\hat{A}$ and $-\hat{A}^\dagger$ are solutions can be resolved by consideration of the meaning of $-\hat{A}^\dagger$. From the definitions of $\hat{A}$ and $\hat{S}$ and the expressions for the unitary relation in terms of these operators, since $\hat{S}$ is $\hat{I} + (ik/2\pi)\hat{A}$, then $\hat{S}^\dagger$ is $\hat{I} - (ik/2\pi)\hat{A}^\dagger$ and, since $\hat{S}^\dagger$ is the inverse of $\hat{S}$, $-\hat{A}^\dagger$ represents the time-reversed diffraction process, in which a set of beams of various amplitudes and phases scatters and produces a single final beam that corresponds to a beam of electrons emerging from the top of the scatterer and moving up the microscope-lens column. From consideration of the situation, it will be seen that the appropriate set of phases to accomplish this is $\pi - \varphi$ (Newton, 1968), where $\varphi$ corresponds to the non-time-reversed diffraction.

## Derivation of the Sayre equation from the unitary relation

The following is by and large stated in papers by Mishnev and his co-workers (Mishnev & Shvets, 1979; Mishnev & Belyakov, 1992). For an ideal crystal, the Bragg condition can be expressed as $\mathbf{k}' - \mathbf{k} = 2\pi\mathbf{H}$, where $\mathbf{H}$ is a reciprocal-lattice vector. In the condition of interest, both $\mathbf{k}'' - \mathbf{k}'$ and $\mathbf{k}'' - \mathbf{k}$ must also be perpendicular to lattice planes, *e.g.* $\mathbf{H} - \mathbf{K}$ and $\mathbf{K}$. This relationship of the vectors in the Bragg condition implies that the scattering amplitudes can depend only on the

differences of the momentum vectors ($\mathbf{k}'' - \mathbf{k}$ or $\mathbf{k}'' - \mathbf{k}'$), *i.e.* only on $\mathbf{H}$ and $\mathbf{K}$.

This can be seen by considering the conditions illustrated in Fig. 2. A beam is incident in the $\mathbf{k}$ direction and is diffracted in the $\mathbf{k}''$ direction. Then the beam is again diffracted in the $\mathbf{k}'$ direction. In this situation, the undiffracted beam is characterized by Miller indices $\mathbf{0}$, the once-diffracted beam by $\mathbf{K}$ and the twice-diffracted beam by $\mathbf{H}$. If, however, a beam were incident in the $\mathbf{k}''$ direction, the undiffracted beam (still in the $\mathbf{k}''$ direction) would be characterized by Miller indices $\mathbf{0}'$. The same shift, from $\mathbf{K}$ to $\mathbf{0}'$, applied to a once-diffracted beam in the $\mathbf{k}'$ direction results in this beam being characterized by Miller indices $\mathbf{H} - \mathbf{K}$. Thus, the transition from $\mathbf{K}$ to $\mathbf{H}$ is the same as the transition from $\mathbf{0}$ to $\mathbf{H} - \mathbf{K}$, or the amplitudes depend only on the vector difference between the initial and final momenta. This is true only in the event that all reflections considered are in the exact Bragg condition.

Furthermore, since the scattering amplitude is non-zero only when $\mathbf{H}$ is a lattice vector, the integral in the unitary relation becomes a sum [compare with equation (6) above]:

$$A(\mathbf{H}) - A^*(-\mathbf{H}) = (ik/2\pi)\sum A(\mathbf{H} - \mathbf{K})A(\mathbf{K}), \quad (14)$$

but, in the absence of anomalous scattering, Friedel symmetry gives $A^*(-\mathbf{H}) = A(\mathbf{H})$, so, taking the term for $\mathbf{K} = \mathbf{H}$ to the left-hand side of the equation, we have

$$A(\mathbf{H})A(\mathbf{0}) = -\sum{}' A(\mathbf{H} - \mathbf{K})A(\mathbf{K}), \quad (15)$$

where the prime means that $\mathbf{K} \neq \mathbf{H}$. Multiplying by $A(\mathbf{0})$ we get

$$A(\mathbf{H})A(\mathbf{0})A(\mathbf{0}) = -A(\mathbf{0})\sum{}' A(\mathbf{H} - \mathbf{K})A(\mathbf{K}). \quad (16)$$

Furthermore, from (15) for $\mathbf{H} = \mathbf{0}$, we derive

$$A(\mathbf{0})A(\mathbf{0}) = -\sum{}' A(-\mathbf{K})A(\mathbf{K}) = -\sum{}' |A(\mathbf{K})|^2. \quad (17)$$

Thus, substituting (17) into (16), we have

$$\begin{aligned} A(\mathbf{H}) &= [-A(\mathbf{0})/A(\mathbf{0})A(\mathbf{0})]\sum{}' A(\mathbf{H} - \mathbf{K})A(\mathbf{K}) \\ &= [A(\mathbf{0})/\sum{}' |A(\mathbf{K})|^2]\sum{}' A(\mathbf{H} - \mathbf{K})A(\mathbf{K}). \end{aligned} \quad (18)$$

Since, in (17), each value of $\mathbf{K}$ appears twice, once as $\mathbf{K}$ and once as $-(-\mathbf{K})$, (18) is also valid if the primed sums mean that $\mathbf{K} \neq \mathbf{H}$ and $\mathbf{K} \neq \mathbf{0}$ and that, in the sum of $A(\mathbf{K})A(-\mathbf{K})$, $\mathbf{K} \neq \mathbf{0}$ and each $\mathbf{K}$ occurs only once, *e.g.* the sum is over the upper half-plane.

Where Friedel symmetry is violated, the difference of the Friedel pair, $A(\mathbf{H}) - A^*(-\mathbf{H})$, must be included in the calculation. Therefore, (15) becomes

$$\begin{aligned} A(\mathbf{H})\{A(\mathbf{0}) + (2\pi i/k)[1 - A^*(-\mathbf{H})/A(\mathbf{H})]\} \\ = -\sum{}' A(\mathbf{H} - \mathbf{K})A(\mathbf{K}), \end{aligned} \quad (19)$$

equation (17) is

$$A(\mathbf{0})A(\mathbf{0}) = -\sum{}' A(\mathbf{K})A(-\mathbf{K}), \quad (20)$$

and (18) can be written

$$\begin{aligned} A(\mathbf{H}) = \big[A(\mathbf{0})\{\sum{}' A(\mathbf{K})A(-\mathbf{K})\big(1 + (2\pi i/k) \\ \times \{[A(\mathbf{H}) - A^*(-\mathbf{H})]/A(\mathbf{H})A(\mathbf{0})\}\big)\}^{-1}\big] \\ \times \sum{}' A(\mathbf{H} - \mathbf{K})A(\mathbf{K}). \end{aligned} \quad (21)$$

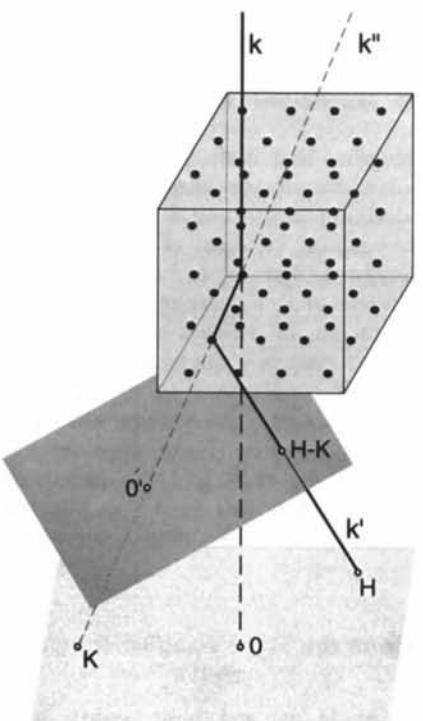


Fig. 2. Illustration of the independence of the transition amplitude of the absolute momentum directions. The solid lines represent the path of a particle scattered first in the $\mathbf{k}''$ direction then in the $\mathbf{k}'$ direction. The dashed and dotted lines are extensions of the scattering path in the directions $\mathbf{k}$ and $\mathbf{k}''$. One of the two planes is perpendicular to $\mathbf{k}$ and the other to $\mathbf{k}''$. The vector between $\mathbf{0}'$ and $\mathbf{H} - \mathbf{K}$ is the same as that beween $\mathbf{K}$ and $\mathbf{H}$, showing the dependence of the transition only on the difference between the reciprocal lattice vectors. *N.B.* This is true only if each of the reflections is in the exact Bragg condition.

## Dynamical scattering and the kinematical scattering operator

Since the structure factors are simply related to the scattering amplitudes only for the case that the scattering can be treated kinematically, it is of interest to examine the relationship between dynamical and kinematical scattering.

For unitary operator $\hat{S}$, there is a generating operator $\hat{K}$, such that $\hat{S} = \exp(i\tau n_z \hat{K})$ (Sturkey, 1962); here $\tau$ is the thickness of a unit cell and $n_z$ is the thickness of the crystal in unit cells. Let $C$ be the unit-cell dimension in the $z$ direction. Slice the crystal $N$ times perpendicular to

the $z$ axis. Let $N_\varepsilon$ be large enough such that for $N > N_\varepsilon$ the change in potential throughout any slice is less than $\varepsilon$, *i.e.* for all $z_i$ such that $|z_1 - z_2| < Cn_z/N$ we have

$$|V(x, y, z_1) - V(x, y, z_2)| < \varepsilon. \tag{22}$$

Then slice each of the $N$ slices $n'$ times. Each of these $n'$ slices from one of the $N$ slices has a potential that is the same to an arbitrary precision. The scattering operator for the $m$th of the $N$ slices, $\hat{S}_m = (\hat{S}_{s1,m})^{n'}$, and we find

$$\hat{S} = \prod \hat{S}_m = \prod[\hat{I} + (ik/2\pi)\hat{A}_m] = \prod[\hat{I} + (ik/2\pi)\hat{A}_{s1,m}]^{n'}, \tag{23}$$

where each term is raised to the power $n'$ and then the product is taken in order (this is an operator product which need not commute). As $n' \to \infty$, $\hat{A}_{s1,m}$ becomes inversely proportional to $n'$. So, letting $c$ be a constant of proportionality, we get

$$\hat{S}_m \to [\hat{I} + (ik/2\pi)(c/n')\hat{K}_m]^{n'} \to \exp[(ikc/2\pi)\hat{K}_m]. \tag{24}$$

Letting $c = 2\pi\tau n_z/k$, and remembering that the $\hat{K}_i$ can incorporate a real constant factor so that $\tau$ can be the same for each slice, we have

$$\hat{S} \to \prod \exp(i\tau n_z \hat{K}_m). \tag{25}$$

Now, if the scattering potential arises from two or more sources, the scattering from the total potential can be expressed in terms of the scatterings from each of the sources. Furthermore, the potential obeys the principle of superposition, so that the total potential is just the sum of the potentials arising from each of the sources. For the interaction Hamiltonians $\hat{H}_i'$ for the sources and $\hat{H}'$ for the total, Newton (1982) derives the expressions for the $\hat{T}$ operator

$$\hat{T}(E) = \sum_i \hat{H}_i' + \sum_i \hat{H}_i' G^+(E)\hat{T}(E) \tag{26}$$

and

$$\hat{T}(E) = \sum_i \hat{T}_i(E) + \sum_i \sum_{j \neq i} \hat{T}_i(E) G^+(E) \hat{T}_j'(E), \tag{27}$$

with

$$\begin{aligned} \hat{T}_i'(E) &= \hat{H}_i' + \hat{H}_i' G^+(E)\hat{T}(E) \\ &= \hat{T}_i(E) + \sum_{j \neq i} \hat{T}_i(E) G^+(E) \hat{T}_j'(E). \end{aligned} \tag{28}$$

Substituting (28) into (27) repeatedly gives

$$\hat{T} = \sum_i \hat{T}_i + \sum_i \sum_{j \neq i} \hat{T}_i G^+ \hat{T}_j + \sum_i \sum_{j \neq i} \sum_{k \neq j} \hat{T}_i G^+ \hat{T}_j G^+ \hat{T}_k + \ldots. \tag{29}$$

This expression contains the symmetrized forms of the products of $\hat{T}_i$, such as $(\hat{T}_1 G^+ \hat{T}_2 + \hat{T}_2 G^+ \hat{T}_1)$, so it is invariant under exchange of $\hat{T}_i$ and $\hat{T}_j$. For the case of two sources, we have

$$\begin{aligned} \hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_1 G^+ \hat{T}_2 + \hat{T}_2 G^+ \hat{T}_1 \\ + \hat{T}_1 G^+ \hat{T}_2 G^+ \hat{T}_1 + \hat{T}_2 G^+ \hat{T}_1 G^+ \hat{T}_2 + \ldots. \end{aligned} \tag{30}$$

Thus, $\hat{S}_1 \hat{S}_2 = \hat{S}_2 \hat{S}_1$, or $\hat{S}_1$ and $\hat{S}_2$ commute.

Diagonalizable operators can be expressed by the spectral representation (Deif, 1982)

$$\hat{A} = \sum \lambda_i \hat{E}_i, \tag{31}$$

where $\hat{E}$ is the projection onto the eigenvector. Since these projections are idempotent and orthogonal, *i.e.* $\hat{E}_i \hat{E}_j = \delta_{ij} \hat{E}_i$, for functions expressible in a power series, it is easy to see that

$$f(\hat{A}) = \sum f(\lambda_i) \hat{E}_i. \tag{32}$$

Two commuting operators can be simultaneously diagonalized; thus, a common set of eigenvectors can be selected for them (Merzbacher, 1961, pp. 153–154). Therefore, for $\hat{S}_1$ and $\hat{S}_2$,

$$\hat{S}_i = \sum \exp(i\tau n_z \lambda_j^{\hat{K}_i}) \hat{E}_j, \tag{33}$$

where $\lambda$ is an eigenvalue of a $\hat{K}$ operator. Whence,

$$\hat{S}_1 \hat{S}_2 = \sum \exp[i\tau n_z (\lambda_i^{\hat{K}_1} + \lambda_i^{\hat{K}_2})] \hat{E}_i. \tag{34}$$

Note, however, that, since $\exp(a + b) = \exp(a + b + 2\pi ni)$, the eigenvalues for $\hat{K}_i$ are not uniquely determined from those for $\hat{S}_i$. From (34) and (25), taking the $i$th source to be the $i$th slice, the expression for the total scattering operator is $\hat{S} = \exp(\sum i\tau \hat{K}_i)$. All the terms with the same $z$ position within the unit cell from each of the $n_z$ unit cells can be gathered together, and letting $\hat{K}$ be the operator corresponding to one unit cell, we get $\hat{S} = \exp(i\tau n_z \hat{K})$.

Thus, if $\hat{S}$ is a dynamical scattering operator, $\hat{K}$ is the corresponding kinematical scattering operator (Fujimoto, 1959). This is analogous to the generator of a finite rotation being an infinitesimal rotation. Cowley (1975*a*), working in the opposite way from that above, derives an equivalent expression for the scattering matrix from the Bloch-wave formula for the scattering amplitude.

For $\tau n_z$ sufficiently small, $\hat{K}$ can be found from $\hat{S}$ (or $\hat{A}$) by means of the series for $\ln(1 + x)$,

$$\begin{aligned} i\tau n_z \hat{K} &= -\sum[-(ik/2\pi)\hat{A}]^n/n \\ \text{or } \hat{K} &= -\sum[-(ik/2\pi)\hat{A}]^n/ni\tau n_z. \end{aligned} \tag{35}$$

Note that, since $\exp(2\pi nil) = \hat{I}$, $\hat{S}$ does not unambiguously determine $K$. However, since the operator $A$ is essentially uniquely determined by the diffracted intensities, then, as long as the dynamical scattering is not too severe, the kinematical scattering operator is also. The criterion 'not too severe' means that (1) the thickness is not so great that $\hat{S}$ is on a different sheet from $\hat{K}$, and

(2) the series for $\hat{K}$ converges rapidly enough so that errors in the measurement of the intensities do not become large as they propagate through the series. These errors can accumulate rapidly when an operator is raised to a high power (Cowley, 1975*b*). In practice, if criterion (2) is satisfied, criterion (1) will also be.

A physical argument for the previous section is that the terms of the $\hat{T}$ operator series represent multiple scatterings giving rise to a total scattering, and that, for example, the incident particle can either scatter first from the first part of the potential and then from the second or from the second part then the first. The total scattering consists of contributions from single scatterings taken in all possible permutations; thus, the total scattering is invariant if the *i*th and *j*th parts are interchanged, *i.e.* if the order of two partial scatterings is reversed.

Consideration of the Green-function formalism for the solution of the Schrödinger equation leads to the same conclusion. With the Schrödinger equation written as $(\nabla^2 + k^2)\psi = U\psi$, the Green function is the solution for a function potential, *i.e.* $(\nabla^2 + k^2)G = 4\pi\delta(\mathbf{r} - \mathbf{r}')$ (Merzbacher, 1961, p. 222), whence

$$\psi(\mathbf{r}) = (2\pi)^{-3/2}\exp(i\mathbf{k}\cdot\mathbf{r}) - (1/4\pi)\int G(\mathbf{r},\mathbf{r}')U(\mathbf{r}')\psi(\mathbf{r}')\,\mathrm{d}^3\mathbf{r}'. \quad (36)$$

This represents the incident wave plus the waves scattered from each point superposed to give the total wave. Since each point is equivalent to any other, the scattering from any collection of sets of points will be the same regardless of the order in which the sets are taken, *i.e.* the scattering operators for each of the sets commute.

Summarizing the results of this section: (1) for an infinitesimal thickness of the crystal, scattering from each part is kinematical; (2) the scattering operators for finite thickness can be expressed as exponentials of the kinematical scattering operators; (3) if we have potential source $s_1$ giving rise to scattering operator $\hat{S}_1$ and potential source $s_2$ giving rise to scattering operator $\hat{S}_2$, then for potential source $s_1 + s_2$ the total potential is the superposition of the separate potentials, the kinematical scattering operators add, and the scattering operators for finite thickness commute. This process can be repeated for $s_3$ $s_4$ *etc.*, so this is true for any number of sources.

## Practical applications

The derivation of the Sayre equation from unitarity assumes that the reflections are in the exact Bragg condition. For non-zero excitation error, the transition from **K** to **H**, illustrated in Fig. 2, must be determined by tilting the crystal so that the incident beam is parallel to the direction of the diffracted beam **K**. This can always be accomplished, but makes the measurement of all the quantities needed for phase determination much more complicated than merely taking one zone-axis electron diffraction pattern.

The excitation error is $\zeta = (1 - \cos\theta)/\lambda \simeq \theta^2/\lambda$. For a fixed resolution, $\zeta \simeq \theta/4d$, so for a resolution of 0.05 nm, $\zeta \simeq 0.2\,\mathrm{nm}^{-1}$ at 100 kV, $\zeta \simeq 0.1\,\mathrm{nm}^{-1}$ at 400 kV and $\zeta \simeq 0.04\,\mathrm{nm}^{-1}$ at 1200 kV accelerating voltages. These values are comparable with the reflection half-width for a 10 nm thick crystal (Hirsch, Howie, Nicholson, Pashley & Whelan, 1965). The effect of excitation errors of this size depends on the properties of the crystal being examined (size, imperfections *etc.*) and on the rocking curve, so generalizations are not warranted. However, a study of the voltage-dependent effects on dynamical scattering from crystals of copper perchlorophthalocyanine (Tivol, Dorset, McCourt & Turner, 1993) found that dynamical scattering accounted for the differences observed except at 200 kV and, to a much lesser extent, at 300 kV. The cases of these two voltages were explained by excitation error effects. For 400 kV and above, therefore, the excitation errors were small enough to be insignificant at 0.19 nm resolution. Thus, it is expected that, for crystals about 10 nm thick and at an accelerating voltage of 1200 kV, the effects of non-zero excitation error will be marginal to insignificant for atomic level structure determination.

The phases that are determined from the unitary equation are the phases of the scattering amplitudes, which are not necessarily the phases of the structure factors. In fact they are the 'dynamical phases' used to produce the so-called 'electron-density maps' referred to by Peng & Wang (1994). Although these maps are actually the wave functions produced by convolutions of scatterings, and not a true electron density map, they can still be reasonably accurate representations of projections of the structures. This paper shows that the solution for the dynamical phases is essentially unique, although the existence of many solutions that fit the intensity data almost as well has not been ruled out.

Phase extension is possible regardless of whether or not the scattering can be treated as kinematical, since the scattering operator is unitary for the dynamical case; thus, the success of phase extension even for crystals containing heavy atoms should not be surprising. Furthermore, the resolutions at which the multiple scattering amplitudes are comparable with the single scattering amplitude are less than 0.5 Å (Peng & Wang, 1994), which is beyond the limit to which most organic crystals diffract.

Although the phases of the amplitudes are not necessarily those of the structure factors, the resulting maps give reasonable atomic positions (Peng & Wang, 1994), which can be used as starting positions for refinement, and, if a multislice calculation is used to determine predicted intensities for comparison with those observed, the correct structure should be found.

This is in accord with the results of Cowley & Moodie (1959), who predicted that reasonable atomic positions

could be determined due to the exit wave function having peaks in the same positions as the projected potential for the situation that the incident beam is aligned with a zone axis.

The existence of a unique solution for the phases means that any solution that is consistent must, in fact, be the true solution. While it is comforting to have an algorithm that converges uniformly, as under the conditions considered in Gerber & Karplus (1972), Newton (1968) and Martin (1969), phases obtained in other ways can quite freely be used as a starting set, since uniqueness implies that they must lead to the correct solution.

The fact that the unitary equation is an operator equation means that it holds for any initial state whether experimentally realizable or not; furthermore, for a real experiment only those reflections with non-zero intensities contribute to the convolution. Therefore, the relation holds if there is only a line of spots, if there is a plane of spots, a three-dimensional array, or a 'four-dimensional' array as seen with some incommensurate structures.

Although Sayre (1952) solved the structure of hydroxyproline using his equation, in the years that followed, its use was felt to be impractical [by way of illustration, Lipson & Cochran (1966) said of the Sayre equation '... the application is of historical interest only and nowadays nobody would contemplate trying to solve a crystal structure in this way'.]

Recent work, however, has shown the use of the Sayre equation to be quite practical (Dorset, Kopp, Fryer & Tivol, 1995). Particularly, proceeding from a low-order phase set determined, for example, from an electron micrograph, and using the Sayre equation for phase extension, is mathematically straightforward and not too computationally intensive. The results are unambiguously determined, without the complications that can arise in multisolution methods.

Of course, all the foregoing has been derived with the assumption that the data are known to infinite accuracy. A situation where direct methods could fail to reach the correct solution can occur if the errors in the measured intensities and the computational errors, such as round-off, give rise to equations for the phases with significantly incorrect coefficients. In the case of a real experiment, the diffracted intensities must be measured accurately for a real crystal, which invariably has defects and may be bent or consist of a number of crystallites with different orientations. Furthermore, the incident beam is not a plane wave but has components of differing (vector) momentum. Measuring the intensities accurately is, therefore, not a trivial undertaking. However, using convergent-beam conditions and taking the intensities from corresponding locations within each diffracted disc is one method that overcomes some of the limitations of real crystals (Spence, 1995). In the centrosymmetric case, where the phases must be 0 or $\pi$, the errors have to be larger to give incorrect phases than in the non-centrosymmetric case, where small errors in some phases can propagate through the calculations. In addition, calculation of the kinematical scattering operator will be more accurate when the logarithmic series converges rapidly, *i.e.* for small deviations of the dynamical scattering from the kinematical.

The final test for the applicability of direct methods of phasing to real electron diffraction experiments must be the ability to obtain correct structures from real data. So far, there have been several successes (Dorset, 1993; Dorset, Tivol & Turner, 1991, 1992, 1993), although further experimental work is necessary to discover the limits of applicability of these methods.

Although there is no constraint on the scattering other than unitarity, so that neither a positive-semidefinite potential nor resolved atoms are needed for direct methods to find the phases, the process of refining a structure by, for example, least-squares search methods requires that the structure be described by relatively few parameters (compared with the number of data points). Therefore, even though the scattering potential for electron diffraction is the continuous shielded-Coulomb potential, the search parameters are still point locations describing the sources of the potential, *i.e.* atoms.

Once the source locations have been parameterized, however, the potentials generated by the sources can take the forms appropriate to the particular physical or chemical environment of the atom, such as $sp^3$ orbitals for tetrahedral carbon, $sp^2$ for graphitic carbon, net charges on ionic bonded atoms *etc.* This means that more realistic potentials can be used for fitting model structures to diffraction data, which should lead to better structure determination and better and more significant $r$ values.

This work was supported by Biotechnological Resource grant RR01219, awarded by the National Center for Research Resources, Department of Health and Human Services/Public Health Service, to support the Wadsworth Center's Biological Microscopy and Image Reconstruction Facility as a National Biotechnological Resource.

## References

ARGOS, P. & ROSSMANN, M. G. (1980). *Theory and Practice of Direct Methods in Crystallography*, edited by M. F. C. LADD & R. A. PALMER, pp. 361–417. New York/London: Plenum Press.

BRICOGNE, G. (1991). *Direct Methods of Solving Crystal Structures*, edited by H. SCHENK, pp. 157–175. New York/London: Plenum Press.

COWLEY, J. M. (1975*a*). *Diffraction Physics*, pp. 203–210. Amsterdam/Oxford: North-Holland; New York: American Elsevier.

COWLEY, J. M. (1975*b*). *Diffraction Physics*, p. 217. Amsterdam/Oxford: North-Holland; New York: American Elsevier.

COWLEY, J. M. & MOODIE, A. F. (1959). *Acta Cryst.* **12**, 360–367.

DAY, P. & PENDRY, J. B. (1993). *Proc. R. Soc. London*, **442**, 1–230.

DEIF, A. S. (1982). *Advanced Matrix Theory for Scientists and Engineers*, p. 113. Tunbridge Wells/London: Abacus Press; New York/Toronto: Halstead Press Division, John Wiley.
DORSET, D. L. (1993). *MSA Bull.* **23**, 99–108.
DORSET, D. L., KOPP, S., FRYER, J. R. & TIVOL, W. F. (1995). *Ultramicroscopy*, **57**, 59–89.
DORSET, D. L., TIVOL, W. F. & TURNER, J. N. (1991). *Ultramicroscopy*, **38**, 41–45.
DORSET, D. L., TIVOL, W. F. & TURNER, J. N. (1992). *Acta Cryst.* A**48**, 562–568.
DORSET, D. L., TIVOL, W. F. & TURNER, J. N. (1993). *J. Appl. Cryst.* **26**, 778–786.
FUJIMOTO, F. (1959). *J. Phys. Soc. Jpn*, **14**, 1558–1568.
GERBER, R. B. & KARPLUS, M. (1972). *J. Chem. Phys.* **56**, 1921–1936.
GLUSKER, J. P. (1993). Editor. *Acta Cryst.* D**49**, 1–222.
HIRSCH, P. B., HOWIE, A., NICHOLSON, R. B., PASHLEY, D. W. & WHELAN, M. J. (1965). *Electron Microscopy of Thin Crystals*, 1st ed. New York: Plenum Press; London: Butterworths.
LADD, M. F. C. & PALMER, R. A. (1980). *Theory and Practice of Direct Methods in Crystallography*, edited by M. F. C. LADD & R. A. PALMER, pp. 93–150. New York/London: Plenum Press.
LIPSON, H. & COCHRAN, W. (1966). *The Determination of Crystal Structures*, p. 246. London: G. Bell.
MARTIN, A. (1969). *Nuovo Cimento*, A**59**, 131–152.
MERZBACHER, E. (1961). *Quantum Mechanics*, pp. 153–154, 222. New York/London: John Wiley.
MISHNEV, A. F. (1991). *Direct Methods of Solving Crystal Structures*, edited by H. SCHENK, pp. 399–400. New York/London: Plenum Press.
MISHNEV, A. F. & BELYAKOV, S. V. (1992). *Acta Cryst.* A**48**, 260–263.
MISHNEV, A. F. & SHVETS, A. E. (1979). *Sov. Phys. Crystallogr.* **24**, 13–15.
NEWTON, R. G. (1968). *J. Math. Phys.* **9**, 2050–2055.
NEWTON, R. G. (1982). *Scattering Theory of Waves and Particles*, 2nd ed., pp. 191–195. New York/Heidelberg/Berlin: Springer-Verlag.
PENG, L.-M. & WANG, S. Q. (1994). *Acta Cryst.* A**50**, 759–771.
SAKURAI, J. J. (1967). *Advanced Quantum Mechanics*, pp. 185–188. Reading, MA/Menlo Park, CA/London/Don Mills, ON: Addison Wesley.
SAYRE, D. (1952). *Acta Cryst.* **5**, 60–65.
SPENCE, J. (1995). Personal communication.
STURKEY, L. (1962). *Proc. Phys. Soc. London*, **80**, 321–354.
TIVOL, W. F., DORSET, D. L., MCCOURT, M. P. & TURNER, J. N. (1993). *MSA Bull.* **23**, 91–98.

*Acta Cryst.* (1995). A**51**, 716–739

# Equilibrium Morphology of Incommensurately Modulated Crystals: a Superspace Description

BY M. KREMERS, H. MEEKES, P. BENNEMA AND M. A. VERHEIJEN

*RIM Laboratory of Solid State Chemistry, University of Nijmegen, Toernooiveld, 6525 ED Nijmegen, The Netherlands*

AND J. P. VAN DER EERDEN

*Interfaces and Thermodynamics, University of Utrecht, 3508 TB Utrecht, The Netherlands*

(*Received* 11 *November* 1994; *accepted* 16 *March* 1995)

## Abstract

The theory for the explanation of equilibrium morphologies of incommensurately modulated one-dimensional crystals, presented in a previous paper, is extended to the case of incommensurately modulated three-dimensional crystals. It is shown that, concerning the morphology, there exists a one-to-one correspondence between faces on the physical crystal and crystallographic hyperplanes of the embedded crystal in superspace. This holds for both main faces and satellite faces. The occurrence of the latter, however, is unique for incommensurately modulated crystals. It is shown that the stability of satellite faces, as well as main faces, can be attributed to a principle of selective cuts. The superspace approach that is developed leads to a calculation method for surface free energies that, in principle, can be applied to incommensurately modulated structures of arbitrary complexity. Equilibrium morphologies are constructed from the calculated surface free energies by means of a standard Wulff plot. The dependence of the equilibrium morphology on several structural parameters is studied for an incommensurately modulated simple cubic model crystal. This study allows for a basic understanding of the differences in morphology of $AuTe_2$ crystals and $[(CH_3)_4N]_2ZnCl_4$ crystals.

## 1. Introduction

It is well known that the morphology of crystals is often determined by flat faces. The orientation of these faces is related to directions of Fourier wave vectors of the structure; the Fourier wave vectors are parallel to the face normals. Crystal faces can be labelled perfectly by a set of three integral indices because the Fourier wave vectors of a classical crystal build a three-dimensional lattice. In incommensurately modulated crystals, it is not possible