Acta Cryst. (1954). 7, 757

A Method of Determining the Relative Positions of Molecules using Fourier-Transform Principles

BY C. A. TAYLOR

Physics Department, College of Technology, Manchester 1, *England*

(*Received* 17 *May* 1954 *and in revised form* 26 *June* 1954)

After the shape and orientation of a molecule have been determined by Patterson, Fourier-transform, or other methods its position in the unit cell must be found. Details are given of a method by which this has been done, and examples are given for projections of structures which have the plane groups *p*2 and *pgg*. Some theoretical implications and possible lines of development are indicated.

Introduction

In applying the principles of physical optics to the determination of crystal structures, Hanson, Lipson & Taylor (1953) have shown that, for certain space groups, the problem is to be treated in two stages; first the shape and orientation of the molecules should be found and then the molecules can be located in the unit cell. The optical methods described, however, deal mainly with the first stage, and the second stage has been left to trial-and-error methods, except for certain special cases. It would obviously be a great advantage if an objective method of solving the position problem were available, particularly if there happened to be some doubt about the shape of the molecule; the possibility of improving a slightly incorrect molecule by Fourier refinement would be greatly increased if it were correctly placed. Such a method is described in the present paper.

It should be pointed out immediately that, although the idea arose out of optical transform studies, it could be applied equally well in conjunction with the more usual computational methods of testing trial structures. There are also some interesting theoretical implications. These cast some light on the problem of structures which, though giving a reasonably low agreement residual, cannot be satisfactorily refined by Fourier methods. The method has already been used on the structure of triphenylene (Klug, 1950), which falls into this category; the published structure turns out to have molecules of the correct shape and orientation in completely incorrect positions. Details of the corrected structure will be published later.

Basic principles

The idea can best be introduced by referring to the optical transforms (Hanson *et al.*, 1953) of arbitrary arrangements of holes shown in Fig. 1.

Fig. 1(*a*) is the optical transform of one molecule of the hypothetical structure shown inset. If this were the sole content of the unit cell, the intensities at points in the X-ray diffraction pattern would be proportional to the intensities at the corresponding points of this figure (assuming that the change of scattering factor with angle is taken into account (Lipson & Taylor, 1951)). If, however, the effective unit cell of the projection contained two such molecules, related by a centre of symmetry, a second factor would be introduced and, for a given molecular shape, the intensity at a point would change if the relative positions of the molecules were changed. The *maximum* possible intensity at any point, regardless of relative molecular positions, would still be proportional to the intensity at the corresponding point in Fig. 1(*a*); the value actually observed would depend on the position of the point chosen relative to the wavy fringe system that crosses the transform (Fig. 1(*b*)) and which in turn depends on the relative positions of the molecules. In the particular case of centrosymmetrical molecules (which have identical projections when related by a centre of symmetry or twofold axis) the fringes are straight (Fig. 1(*c*)).

The principle of the optical method, as applied to structures containing molecules related by a centre of symmetry in the projection, is to adjust the shape and orientation of a single molecule until the intensities at points of its transform corresponding to the strongest X-ray intensities are sufficiently great; the weaker reflexions are ignored at this stage. The weak X-ray intensities may occur either because the intensity in the transform of the single molecule is low, or because of the operation of the fringe function. In applying the present method, points are chosen for which the X-ray intensity is low even though the optical transform of the single molecule allows a high maximum value. The observed values must then be governed chiefly by the fringe function, and information about the relative positions of the molecules can be derived.

It was pointed out in the earlier paper (Hanson *et al.*, 1953) that, for pyrene, the straight fringes can be observed directly in the weighted reciprocal-lattice section, and Woolfson (1953) has made use of such

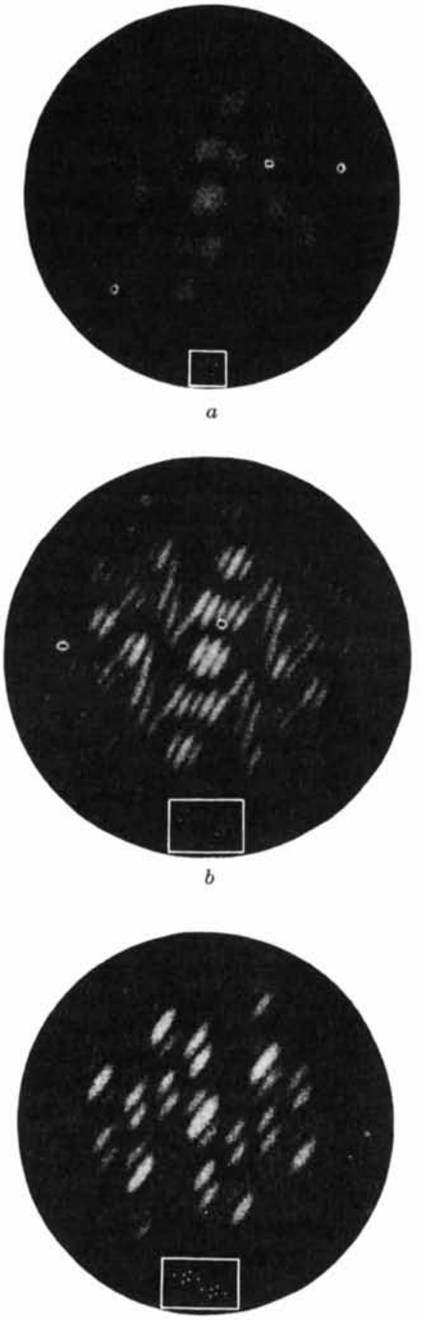

Fig. 1. (*a*) Optical transform of a hypothetical asymmetric unit of five holes. (*b*) Optical transform of two of the same units arranged centrosymmetrically. (*c*) Optical transform of two centrosymmetrical units arranged centrosymmetrically. (From Hanson, Lipson & Taylor, 1953; reproduced by permission.)

fringes in determining the structure of 1:1:6:6 tetraphenylhexapentaene. The procedure is often difficult, however, and becomes impossible with the wavy fringes associated with non-centrosymmetrical molecules.

The theory and an example for the projection of a structure containing two molecules in the unit cell in the plane group $p2$ will be considered first. Later it will be shown that the method can be extended to more complicated structures, and the theory and an example for the plane group pgg with four molecules in a unit cell will be given.

Theory of the method

The Fourier transform G_0 of a set of points represented by vectors $\mathbf{r}_n$ with respect to an origin O is a function which, at some point in reciprocal space represented by a vector $\mathbf{S}$, is given by the equation

$$G_0 = \sum_{n=1}^{N} f_n \exp\{2\pi i \mathbf{r}_n \cdot \mathbf{S}\}\,, \tag{1}$$

where f_n is the scattering factor associated with each point (see, for example, Lipson & Cochran, 1953).

If the origin is now re-chosen at a point O', where the translation vector O'–O is $\mathbf{r}$ (see Fig. 2), the transform becomes

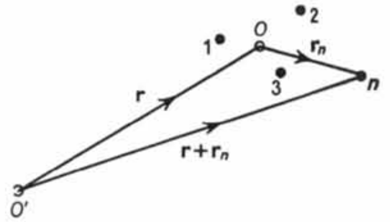

Fig. 2. Relationship between the coordinates of an atom relative to the arbitrary origin O and the coordinates relative to the true origin O'.

$$\begin{aligned} G_r &= \sum_{n=1}^{N} f_n \exp\{2\pi i(\mathbf{r}_n+\mathbf{r})\cdot\mathbf{S}\} \\ &= \sum_{n=1}^{N} f_n \exp\{2\pi i \mathbf{r}_n\cdot\mathbf{S}\}\exp\{2\pi i\mathbf{r}\cdot\mathbf{S}\} \\ &= G_0 \exp\{2\pi i \mathbf{r}\cdot\mathbf{S}\} \qquad \text{(from equation 1).} \end{aligned}$$

If another, centrosymmetrically related set of points is now added at a distance $-\mathbf{r}$ from the new origin, then the transform of this new set alone is

$$G_{-r} = G_0^* \exp\{-2\pi i \mathbf{r}\cdot\mathbf{S}\}\,,$$

where G_0^* is the complex conjugate of G_0. The total transform, G, of the two sets, relative to the origin at O' is thus

$$G = G_r + G_{-r} = G_0 \exp\{2\pi i\mathbf{r}\cdot\mathbf{S}\} + G_0^* \exp\{-2\pi i\mathbf{r}\cdot\mathbf{S}\}\,.$$

But

$$G_0 = |G_0| \exp\{i\varphi_0\} \quad\text{and}\quad G_0^* = |G_0| \exp\{-i\varphi_0\}\,,$$

where φ_0 is the phase angle at the point $\mathbf{S}$ in G_0. Therefore

$$
\begin{aligned}
G &= |G_0| \exp\{i\varphi_0\} \exp\{2\pi i \mathbf{r}\cdot\mathbf{S}\} \\
&\qquad + |G_0| \exp\{-i\varphi_0\} \exp\{-2\pi i \mathbf{r}\cdot\mathbf{S}\} \\
&= |G_0| \exp\{2\pi i \mathbf{r}\cdot\mathbf{S}+\varphi_0\} \\
&\qquad + |G_0| \exp\{-(2\pi i \mathbf{r}\cdot\mathbf{S}+\varphi_0)\} \\
&= 2|G_0| \cos(2\pi \mathbf{r}\cdot\mathbf{S}+\varphi_0)\,. \qquad (2)
\end{aligned}
$$

This result may be related to the Figs. 1(*a*), (*b*) and (*c*). Fig. 1(*a*) represents the transform G_0 with respect to some arbitrary origin O (strictly speaking it represents $|G_0|^2$, since the intensity is recorded photographically). Fig. 1(*b*) represents the transform G of two units related by a centre of symmetry at O', a point removed by a distance $\mathbf{r}$ from O.

Equation (2) shows that $|G|$ has a maximum value of $2|G_0|$; that is, the amplitude at a point in Fig. 1(*b*) can never exceed twice the amplitude at a corresponding point in Fig. 1(*a*). The quantity $\cos(2\pi i\mathbf{r}\cdot\mathbf{S}+\varphi_0)$ represents a fringe system which, because φ_0 varies from point to point in reciprocal space, is not straight. It may however have straight portions in regions over which φ_0 is almost constant; such straight sections can be seen in Fig. 1(*b*). The units from which Fig. 1(*c*) was derived are themselves centrosymmetrical and therefore φ_0 can take values of 0 or π only. Equation (2) thus reduces to

$$G = \pm 2|G_0| \cos 2\pi\mathbf{r}\cdot\mathbf{S}$$

and the fringes are straight. In order to make use of equation (2) it must be re-written

$$(\psi-\varphi_0)/2\pi = \mathbf{r}\cdot\mathbf{S} \qquad (3)$$

by making the substitution $\psi = \cos^{-1} G/2|G_0|$.

Now $\mathbf{r}\cdot\mathbf{S} = hX+kY$, where X and Y are the fractional coordinates of the arbitrary origin O with respect to the true origin O' at the centre of symmetry of a unit cell, and h and k are the indices corresponding to a point $\mathbf{S}$ in reciprocal space. Thus

$$hX+kY = (\psi-\varphi_0)/2\pi\,. \qquad (4)$$

In any particular problem the magnitudes of G at different points are known ($|G| = |F_o|$ if F_o is on an absolute scale) but the signs are unknown. $|G_0|$ and φ_0 at any point h,k can easily be calculated for the proposed molecule relative to an arbitrary origin. It is thus possible to set up equation (4) for various values of h and k with X and Y as the only unknowns.

Such a set of simultaneous equations could be solved if it were not for the fact that ψ (being derived from an inverse trigonometrical function) is multivalued and also that φ_0 may take any one of a series of values differing by 2π.

Equation (4) should, therefore, be re-written

$$hX+kY = (2n\pi\pm\psi'-\varphi_0')/2\pi \qquad (5)$$

for positive values of G ($\equiv F$ positive),

$$hX+kY = ((2n+1)\pi\pm\psi'-\varphi_0')/2\pi \qquad (6)$$

for negative values of G ($\equiv F$ negative), where ψ' is the value of $\cos^{-1}|G|/2|G_0|$ lying between 0 and $\frac{1}{2}\pi$ and φ_0' is the value of φ_0 lying between 0 and 2π. Expressions (5) and (6) could, of course, be combined in one equation, but the cross-checking of results is simplified if they are kept separate. Since the sign of G is unknown, both equations must be considered.

Solution of the equations

If the equations (5) and (6) are set up for a particular reflexion $hk0$ the choice of true origin is immediately limited to points which lie on a family of straight lines of slope $-h/k$, intercepting the X axis at points

$$X = (2n\pi\pm\psi'-\varphi_0')/2\pi h$$

and

$$X = ((2n+1)\pi\pm\psi'-\varphi_0')/2\pi h\,.$$

If the equations are set up for a second pair of values of hk, a second family of straight lines is produced and the origin can then only lie on one or other of the intersections of the two sets of lines. Symmetry considerations show that it is only necessary to consider the region of 'X–Y space' lying between $X = \pm 0{\cdot}25$ and $Y = \pm 0{\cdot}25$. It is obvious, therefore, that from two observed reflexions the origin can be limited to a finite number of possible positions. If now the equations are set up for more pairs of values of h and k, some of these possibilities may be eliminated; it is found in practice that, if suitable points are chosen, only about 6–8 reflexions are needed to fix the origin uniquely. There are a number of practical points to be noted in applying the method and these will be given after the example.

A simple application

Pyrene (Robertson & White, 1947) has four centrosymmetrical molecules in its unit cell and the space group is $P2_1/a$. The effective unit cell of the projection along [010] contains two molecules in the plane group $p2$, and therefore the method can be applied directly to determine the X and Z coordinates of the molecules. Since the molecules are centrosymmetrical, the arbitrary origin O was chosen at the centre of symmetry, thus making $\varphi_0' = 0$ or π.

To avoid confusion, all fractional coordinates were taken relative to the full unit cell, so the formula for ψ' has to be changed to $\cos^{-1} F_o/4|G_0|$, since the maximum possible value at any point is the contribution of all the four molecules. Also, because of the halving of the unit cell and because the molecule is centrosymmetrical, it is necessary to consider only one quadrant of 'X–Z' space, that is, values of X and Z from 0 to $0{\cdot}25$. The values of G_0 for the indices used were calculated for an idealized molecule.

In Fig. 3, six families of lines have been plotted and give a unique solution. The families for 001 (one line) and 200 (two lines) were plotted first and are shown as broken lines in the figure. These lines intersect at only two points, and these are therefore the only positions for the origin which would give satisfactory agreement for these two reflexions. It should be pointed out, however, that the fringes corresponding to these low-order reflexions are so broad that the agreement is not very sensitive to small changes in position of the origin. They must be regarded as indicating the *region* of the origin rather than its precise position. The 006 family (six lines) and the $12,0,\bar{5}$ family (seventeen lines) intersect in a large number of places, but, in particular, they intersect at points that are very close to the intersections of 001 and 200. Either of these two intersections, therefore, will satisfy 001, 200, 006, and $12,0,\bar{5}$. The families for $40\bar{4}$ and 10,0,2 were added next. For both these reflexions F_o is less than 3 and therefore the origin position is restricted not to the lines themselves but to the region between the line corresponding to $F_o = +3$ and $F_o = -3$. This is indicated by the shading. The four bands of $40\bar{4}$ and the six bands of 10,0,2 intersect at twelve places, but only one of these coincides with one of the two previous intersections. Thus there is only one position for the origin which gives agreement for all the six reflexions. More families could be added to confirm this choice or to fix the position with greater precision. The fractional X and Z coordinates of this intersection are 0·118 and 0·054 which are equivalent to 1·60 Å and 0·45 Å. The final coordinates found by Robertson & White, after refinement, were 1·604 Å and 0·470 Å.

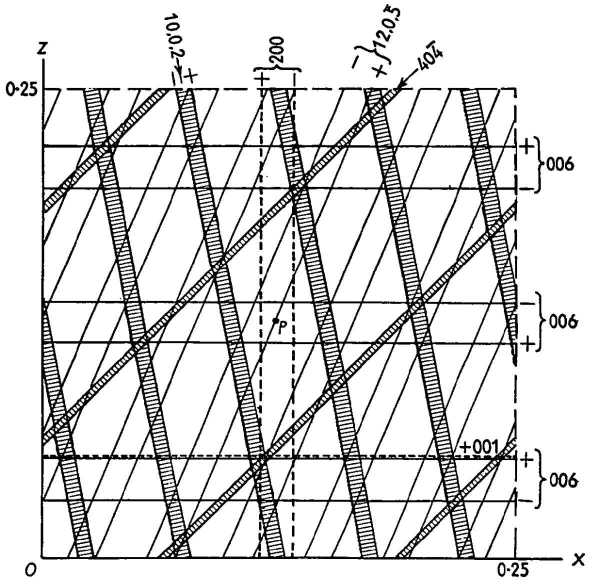

Fig. 3. The solution of equations (5) and (6) for six reflexions from pyrene. All the members of families corresponding to axial reflexions and representatives of each of the other families are labelled.

Practical points of importance

In theory all the reflexions could be used, but it is rarely necessary to use more than eight or ten. It is important, however, to use great care in selecting the reflexions to be used. The following general rules will be found useful.

First it is always desirable to include one or two low-order reflexions as these give only a few lines and restrict the possible positions of the origin much more than the higher orders. In general, it is useful to plot the low-order lines first and then gradually to add higher-order families to eliminate all but one of the alternatives and to increase the accuracy of location of the true origin.

Secondly, if the molecular shape or orientation is incorrect the intersection will not be exact and it will not be possible to satisfy the requirements of all the reflexions. If part of the molecule is thought to be more reliable than the rest, the reflexions used should be chosen in regions where the contribution of the most reliable part is predominant. For instance, for a molecule consisting of a benzene ring with a side group, the orientation of the benzene ring may be known much more precisely than the configuration of the side groups. The reflexions used should therefore be chosen from the regions corresponding to the intense parts of the benzene transform.

Thirdly, since the ratio of F_o to a calculated maximum $2|G_0|$ is involved, the scale of F_o is important. In the example just given, the structure being known, the absolute scale was also known. At the beginning of an investigation, however, the scale may not be known with any degree of accuracy. It is important, therefore, to choose reflexions which are observed very small but which have high values of $2|G_0|$—that is, which lie on high parts of the transform of the single molecule—in order to minimize this error. If only

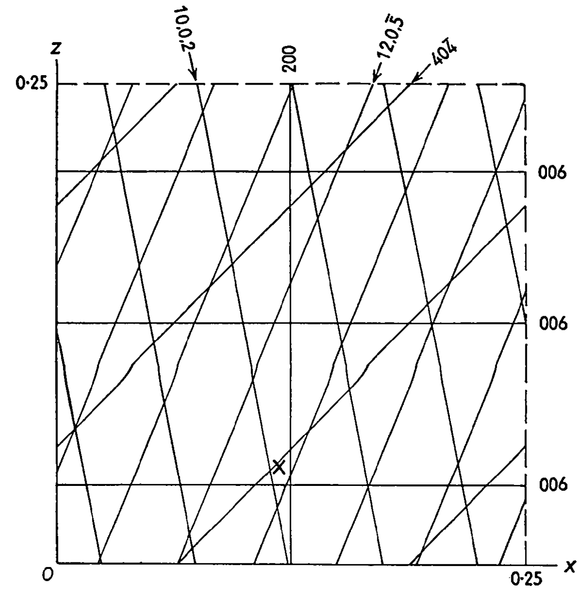

Fig. 4. An approximate solution of equations (5) and (6) for pyrene derived by using the approximation $F_o = 0$ for five reflexions. The position of the intersection from Fig. 3 is marked ×.

reflexions which are observed zero were used, then the method would be independent of scale, but usually there are not enough of these. It is sometimes adequate to choose reflexions which are observed fairly small, and which lie on intense parts of the single transform but make the approximation $F_o = 0$. Fig. 4 shows the effect of applying this approximation to five of the six reflexions for pyrene which were used before. There is, of course, no true point of intersection, but it can be seen that the closest approach to an intersection occurs very close to the point found by the more accurate method (marked ×).

Finally, it must be remembered that φ_0' must be placed in the right quadrant (for example if A and B are *both* negative $\pi < \varphi_0 < 3\pi/2$ (cf. *International Tables*, 1952)). Fortunately, a simple method exists by which the self-consistency of the method can be checked. Consideration of equations (5) and (6) shows that the diagram could be used to estimate the changes in F_c which would occur if the separation of the molecules were changed. The lines of one family represent positions of the origin which make $|F_c|$ for the corresponding reflexion equal to $|F_o|$. If, for example, the origin were moved away from the line corresponding to F positive towards that corresponding to F negative, F_c would decrease, becoming zero at a point half-way between. Thus, once an origin has been selected, whether it is at a true intersection or not, the F values for the reflexions used can be estimated and compared both in magnitude *and sign* with those calculated by conventional methods.

Thus, in Fig. 3, if the origin were placed at P, 12,0,5 would have $F_c = F_o$ and would be positive; 200 and 006 would both have $F_c =$ zero; $40\bar{4}$ and 10,0,2 would both have F_c very high and positive, as P lies half-way between two positive lines in each case.

Extension of theory to plane group *pgg*

With slight modifications, the method can be applied to the plane group *pgg*, with four molecules in the unit cell. In this plane group there are two pairs of centrosymmetrically related molecules. One pair has the centre of symmetry at 0, 0 and the other at $\frac{1}{2}$, $\frac{1}{2}$. The transform of one pair is the mirror image of the transform of the other pair and, because of the difference in position of the centres of symmetry, the transforms have phase differences which are such that the two transforms add at points for which $(h+k)$ is even and subtract where $(h+k)$ is odd. The shape can still be worked out using the transform of one molecule, but it is now only necessary to ensure that the strong reflexions lie on strong parts either of the transform *or* its mirror image. In practice it is easier to look at the problem the other way round and ensure that if $(hk0)$ is strong, either $hk0$ or $h\bar{k}0$ lies on a strong part of the transform. The method can thus be used to determine the separation of one pair of molecules if the reflexions are chosen for which the contribution of one molecule of this pair is large and the contribution of one molecule of the other pair is zero. This extra condition governing the choice of reflexions restricts the choice so much that in order to obtain enough usable reflexions it is necessary to relax the condition and make allowance for the margin of error introduced.

Suppose G_0 is the transform of one molecule relative to its arbitrary origin O as before at a point $\mathbf{S}$ in reciprocal space. Let $\bar{G}_0$ be the transform of the corresponding molecule of the other pair and $\bar{\varphi}_0$ the corresponding phase angle. Then the contribution of the pair of molecules related by a centre at 0, 0 will be given by

$$G_1 = 2|G_0| \cos (2\pi \mathbf{r} \cdot \mathbf{S} + \varphi_0)$$

as before, and the contribution of the other pair, related by the centre at $\frac{1}{2}$, $\frac{1}{2}$ will be

$$G_2 = 2|\bar{G}_0| \cos (2\pi \mathbf{r} \cdot \mathbf{S} + \bar{\varphi}_0)\ .$$

The total is thus

$$G = 2|G_0| \cos (2\pi \mathbf{r} \cdot \mathbf{S} + \varphi_0) \pm 2|\bar{G}_0| \cos (2\pi \mathbf{r} \cdot \mathbf{S} + \bar{\varphi}_0)\,,$$

the positive sign being taken if $(h+k)$ is even at the point $\mathbf{S}$ and the negative sign if $(h+k)$ is odd.

Thus if $|\bar{G}_0|$ is zero, the expression reduces to that previously obtained for $p2$, i.e. equation (2).

If, however, $|\bar{G}_0|$ is not zero we must write

$$G = 2|G_0| \cos (2\pi \mathbf{r} \cdot \mathbf{S} + \varphi_0) + \varepsilon\ ,$$

where ε is the error introduced and may take any value between $+2|\bar{G}_0|$ and $-2|\bar{G}_0|$ according to the value of the cosine function which is unknown. We must thus modify the expression for ψ' to $\psi' = \cos^{-1}(G+\varepsilon)/2|G_0|$, taking, as before, the value between 0 and $\frac{1}{2}\pi$.

For axial reflexions,

$$G_0 = \bar{G}_0,\ G = 4|G_0| \cos (2\pi \mathbf{r} \cdot \mathbf{S} + \varphi_0)$$

and

$$\psi' = \cos^{-1} G/4|G_0|\,.$$

The equations can thus be set up and solved graphically as before, except that when ε has a finite value the loci of possible positions for the origin will no longer be a family of lines but a family of bands. In practice, lines are drawn for the extreme negative and positive values of ε and the region between is shaded in some way. Obviously ε should be kept as small as possible in order to keep the bands as narrow as possible.

Application to purpurogallin

Purpurogallin (Taylor, 1952; Dunitz, 1952) has the space group $P2_12_12$ with four molecules in the unit cell. The [001] projection thus has the plane group *pgg* and the modified theory can be applied. The molecule is non-centrosymmetrical and the method has been applied to the first trial molecule found by Taylor before the Fourier refinement and the scale of F's available

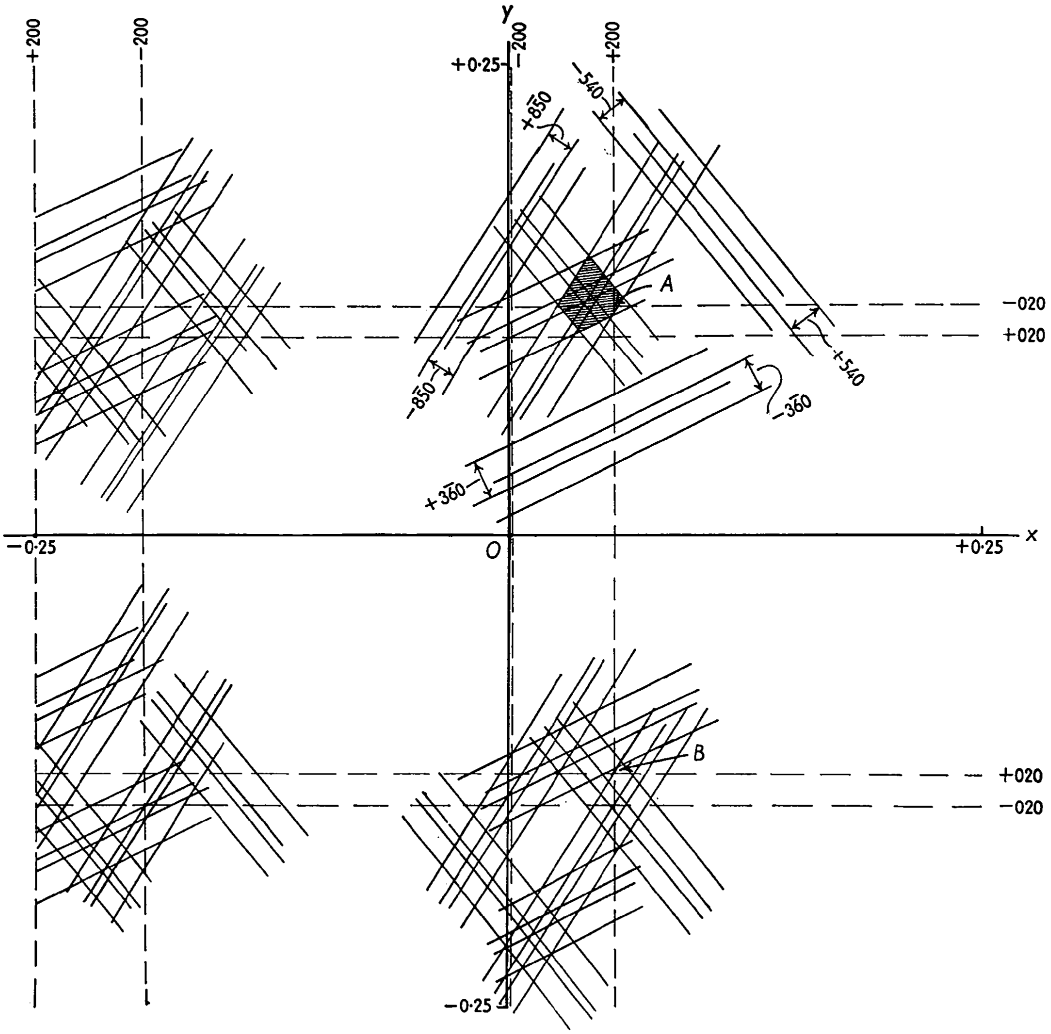

Fig. 5. First stage of the solution for purpurogallin. The complete families corresponding to the axial reflexions 200 and 020 are shown in broken lines; portions of the families for three more reflexions are shown where the bands pass close to intersections of the broken lines. The extent of the bands is indicated by the double-headed arrows on one member of each family.

at that time was used. The result would therefore be expected to be less precise than that for pyrene.

Two axial reflexions and six other reflexions were chosen according to the principles outlined above. For each of these G_0, $\bar{G}_0$ and φ_0 were calculated, and from these ε, and hence the range of values for ψ', could be found. In Fig. 5 the four vertical and four horizontal broken lines correspond to the two axial reflexions and give sixteen possible positions for the origin. These reflexions are not very small and scale errors may therefore be appreciable so that slight movements away from these precise positions may be expected, but they do fix likely regions.

On adding bands corresponding to $3\bar{6}0$, 540 and $8\bar{5}0$, only two (*A* and *B*, Fig. 5) of the sixteen first possibilities lie on a band of all the three families simultaneously, and one of these (*B*) is on the edge of all three. The shaded area in the figure is the region of overlap of all

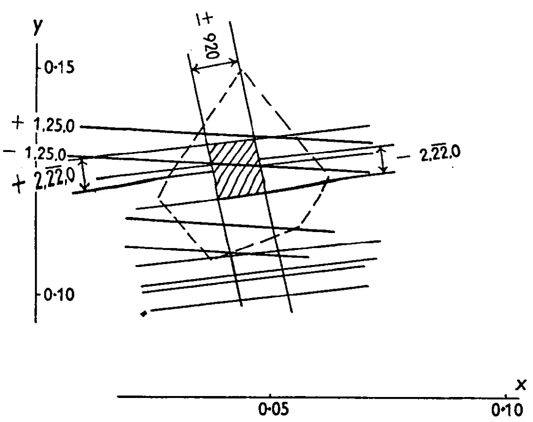

Fig. 6. Second stage of the solution for purpurogallin on a larger scale than Fig. 5. The broken lines indicate the region which was shaded on Fig. 5.

three bands which is considered the most likely position for the origin. In Fig. 6 this shaded area is drawn in broken lines on a much larger scale, and the new shaded area represents the region common to the earlier shaded area and to the bands of 920 and $2,\overline{22},0$. (Notice that as 920 happens to be observed zero the positive and negative bands coincide exactly.) When lines corresponding to 1,25,0 are added (here ε is almost zero so the bands have no appreciable thickness), one of them passes through the middle of the shaded area.

It can thus be seen how the addition of successive higher orders confirms the choice and fixes the position of the origin with increasing precision.

The arbitrary origin was chosen at the centre of the six-ring and, although the shape of the molecule changed very considerably during refinement, the final position of the centre of the six-ring is in quite good agreement with that found by this method:

	Final coordinates from synthesis	Coordinates from graphical method
X	0·036	0·041
Y	0·128	0·129

Some theoretical implications

It has been observed, during the application of this method to a number of structures, that the more highly symmetrical molecules need a larger number of lines on the diagram to define the position of the molecule uniquely. It is sometimes possible to find perhaps six or seven possible origins which will give reasonable agreement for quite a high proportion of the F's. Such origins give reasonable agreement residuals for the whole zone and Fourier syntheses are quite good, but refinement is not possible. If one of the incorrect intersections were taken and then the origin moved slightly in any direction, the agreement for these reflexions would become worse and the normal process of refinement would be discontinued. The implication is that when a structure of this type gives moderately good agreement but will not refine, the position of the molecules may be far removed from the correct one. This happened for triphenylene, as mentioned in the introduction.

Possible developments

One or two lines of possible development suggest themselves. First of all, modifications of the method might be made for other space groups. It is also possible that by suitable selection of the reflexions used it might be practicable to fix one part of a molecule in the unit cell when the rest is still unknown. The obvious example is that quoted earlier of a molecule containing a benzene ring. If only reflexions corresponding to the ring were used, its exact position in the cell might be fixed. The relative positions of two known groups in a given molecule may also be discovered by suitable modifications.

I wish to acknowledge my indebtedness to Dr H. Lipson, Head of the Department, and to my colleagues in the Department of Physics, particularly Mr P. R. Pinnock, for much helpful advice, assistance and encouragement.

References

Dunitz, J. (1952). *Nature, Lond.* **169**, 1091.
Hanson, A. W., Lipson, H. & Taylor, C. A. (1953). *Proc. Roy. Soc.* A, **218**, 371.
International Tables for X-ray Crystallography (1952), vol. 1, p. 355. Birmingham: Kynoch Press.
Klug, A. (1950). *Acta Cryst.* **3**, 165.
Lipson, H. & Cochran, W. (1953). *The Crystalline State*, vol. 3, p. 234. London: Bell.
Lipson, H. & Taylor, C. A. (1951). *Acta Cryst.* **4**, 458.
Robertson, J. M. & White, J. G. (1947). *J. Chem. Soc.* p. 358.
Taylor, C. A. (1952). *Nature, Lond.* **169**, 1088.
Woolfson, M. M. (1953). *Acta Cryst.* **6**, 838.